

Sadržaj

SPSS: priručnik za preživljavanje

Glavni urednik Olga Milanko
Redaktor Stela Spasić
Tehnički urednik Sanja Tasić
Slog Sanja Tasić
 Nataša Pavlov

Izdavač Mikro knjiga, Beograd
Direktor Dragan Tanaskoski

Štampa Greenfield studio, Beograd

Ako imate pitanja ili komentare, ili ako želite da dobijete besplatan katalog, pišite nam ili se javite:

Mikro knjiga
P. fax 20-87
11030 Beograd
tel: 011/3540-544
pisma@mikroknjiga.rs

Autorizovan prevod sa engleskog jezika knjige **SPSS Survival Manual, Third Edition.**

Copyright © 2009 Mikro knjiga. Sva prava zadržana. Nije dozvoljeno da ijedan deo ove knjige bude reprodukovani ili emitovani na bilo koji način, elektronski ili mehanički, uključujući fotokopiranje, snimanje ili bilo koji drugi sistem za beleženje, bez prethodne pismene dozvole izdavača.

Authorized translation from the English language edition, entitled **SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows (Version 15), Third Edition**, by Julie Pallant, published by Allen & Unwin, Copyright © 2007 Julie Pallant.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without prior permission in writing from the publisher. Serbian language edition published by Mikro knjiga.

СР – Каталогизација у публикацији
Народна библиотека Србије, Београд

004.42 SPSS for Windows(035)
311:004.9(035)

ПАЛАНТ, Жули

SPSS : priručnik za preživljavanje :
postupni vodič kroz analizu podataka pomoću
SPSS-a za Windows (verzija 15) / Julie
Pallant : prevod 3. izdanja Miljenko Sućur. –
Beograd : Mikro knjiga, 2009 (Beograd :
Greenfield studio). – XVII, 348 str. :
ilustr. : 24 cm

Prevod dela: SPSS Survival Manual. – Preporučena
literatura: str. 337-338. – Bibliografija: str. 339-342.
– Registar.

ISBN 978-86-7555-341-0
a) Aplikativni program "SPSS for Windows"
– Приручници b) Математичка статистика –
Апдикативни програми – Приручници
COBISS.SR – ID 170306828

SPSS/322/193094O5P56M22S109K910 5 4

Predgovor	x
Datoteke s podacima i Web lokacija	xi
Uvod i pregled	xiii
Struktura knjige	xiv
Kako koristiti ovu knjigu	xiv
Saveti za sprovođenje istraživanja	xvi
Dodatni resursi	xvii

Deo I Kako početi	1
1 Projektovanje studije	3
Planiranje studije	3
Izbor odgovarajućih skala i mera	5
Priprema upitnika	7
2 Priprema šifarnika	11
Imena promenljivih	12
Šifrovanje odgovora	13
Šifrovanje otvorenih pitanja	13
3 Upoznajte SPSS	15
Pokretanje SPSS-a	15
Otvaranje postojeće datoteke s podacima	16
Korišćenje datoteka s podacima	16
SPSS-ovi prozori	17
Meniji	21
Okviri za dijalog	22
Kako zatvoriti SPSS	23
Kako dobiti pomoć	24

Deo II Priprema datoteke s podacima	25
4 Pravljenje datoteke za podatke i unošenje podataka	27
Izmena opcija u SPSS-u	28
Definisanje promenljivih	31
Unošenje podataka	36
Modifikovanje datoteke s podacima	36
Unošenje podataka u Excelu	37
Spajanje datoteka	39
Korisne mogućnosti SPSS-a	40
Upotreba skupova	42
5 Pronalaženje i uklanjanje pogrešnih podataka	45
Korak 1: traženje grešaka	46
Korak 2: pronalaženje i ispravljanje grešaka	
u datoteci s podacima	49
Sažet prikaz slučajeva	51
Deo III Preliminarne analize	53
6 Opisni statistički pokazatelji	55
Kategorijske promenljive	56
Neprekidne promenljive	57
Nedostajući podaci	59
Procena normalnosti raspodele	60
Otkrivanje netipičnih tačaka	65
Dodatne vežbe	66
7 Upotreba dijagrama za opisivanje i analizu podataka	67
Histogrami	68
Stubičasti dijagrami	69
Linijski dijagrami	71
Dijagram rasturanja	74
Pravougaoni dijagram	77
Dorada dijagrama	79
Uvoz dijagrama u Wordove dokumente	80
Dodatne vežbe	81
8 Dorada podataka	83
Izračunavanje ukupnih rezultata na skalama	84
Transformisanje promenljivih	88
Podela neprekidne promenljive na grupe	92
Smanjenje broja kategorija kategorijske promenljive	93
Dodatne vežbe	95

9 Provera pouzdanosti merne skale	97
Objašnjenje primera	98
Tumačenje rezultata provere pouzdanosti	100
Predstavljanje rezultata provere pouzdanosti skala	101
Dodatne vežbe	101
10 Izbor prikladnih statističkih tehnika	103
Pregled raznih statističkih tehnika	104
Proces donošenja odluka	107
Osnovne osobine glavnih statističkih tehnika	113
Zbirna tabela svojstava osnovnih statističkih tehnika	118
Literatura za dalje usavršavanje	120
Deo IV Statističke tehnike za istraživanje veza između promenljivih	121
Tehnike obrađene u četvrtom delu knjige	121
Pregled osnovnih načela	122
11 Korelacija	129
Objašnjenje primera	130
Preliminarne analize za korelaciju	130
Tumačenje rezultata korelacije	134
Predstavljanje rezultata korelacije	136
Računanje korelacije između grupa promenljivih	137
Poređenje koeficijenata korelacije dve grupe	139
Provera statističke značajnosti razlike između koeficijenata korelacije	141
Dodatne vežbe	144
12 Delimična korelacija	145
Objašnjenje primera	145
Tumačenje rezultata delimične korelacije	148
Predstavljanje rezultata delimične korelacije	148
Dodatna vežba	148
13 Višestruka regresija	149
Glavne vrste višestruke regresije	150
Pretpostavke na kojima se zasniva višestruka regresija	151
Objašnjenje primera	152
Standardna višestruka regresija	154
Tumačenje rezultata standardne višestruke regresije	158
Hijerarhijska višestruka regresija	163
Tumačenje rezultata hijerarhijske višestruke regresije	166
Predstavljanje rezultata višestruke regresije	167
Dodatne vežbe	168

14	Logistička regresija	169
	Pretpostavke	170
	Objašnjenje primera	171
	Priprema podataka: šifrovanje odgovora	171
	Tumačenje rezultata logističke regresije	176
	Predstavljanje rezultata logističke regresije	180
15	Faktorska analiza	181
	Koraci od kojih se sastoji faktorska analiza	182
	Objašnjenje primera	186
	Postupak faktorske analize	187
	Tumačenje rezultata	191
	Predstavljanje rezultata faktorske analize	199
	Dodatne vežbe	202
Deo V	Statističke tehnike za poređenje grupa	203
	Tehnike obrađene u petom delu knjige	203
	Pretpostavke	205
	Greška prve vrste, greška druge vrste i moć testa	207
	Planirana poređenja/naknadne analize	208
	Veličina uticaja	209
	Nedostajući podaci	210
16	Neparametarske tehnike	213
	Kratak pregled tehnika obrađenih u ovom poglavlju	214
	Hi-kvadrat	215
	Mera slaganja kapa	221
	Man-Vitnijev U test	224
	Vilkoksonov test ranga	227
	Kruskal-Volisonov test	229
	Fridmanov test	231
	Dodatne vežbe	233
17	T-testovi	235
	T-test nezavisnih uzoraka	235
	T-test uparenih uzoraka	239
	Dodatne vežbe	244
18	Jednofaktorska analiza varijanse	245
	Jednofaktorska ANOVA različitih grupa s naknadnim testovima	246
	Jednofaktorska ANOVA različitih grupa s planiranim poređenjima	251
	Jednofaktorska ANOVA ponovljenih merenja	254
	Dodatne vežbe	259

19	Dvofaktorska analiza varijanse različitih grupa	261
	Objašnjenje primera	262
	Tumačenje rezultata dvofaktorske analize varijanse	265
	Predstavljanje rezultata dvofaktorske analize varijanse	268
	Dodatne analize kada se dobije značajan uticaj interakcije	268
	Dodatne vežbe	269
20	Kombinovana analiza varijanse	271
	Objašnjenje primera	272
	Tumačenje rezultata kombinovane analize varijanse	276
	Predstavljanje rezultata kombinovane analize varijanse	279
21	Multivarijaciona analiza varijanse	281
	Objašnjenje primera	282
	Sprovođenje MANOVA analize	288
	Tumačenje rezultata MANOVA analize	291
	Predstavljanje rezultata MANOVA analize	294
	Dodatna vežba	294
22	Analiza kovarijanse	295
	Čemu služi ANCOVA	295
	Osnovne pretpostavke analize kovarijanse	296
	Jednofaktorska ANCOVA	298
	Dvofaktorska ANCOVA	308
Dodatak	Pojedinosti o datotekama s podacima	317
	Deo A: materijali za survey3ED.sav	320
	Deo B: materijali za experim3ED.sav	327
	Deo C: materijali za staffsurvey3ED.sav	329
	Deo D: materijali za sleep3ED.sav	332
	Deo E: materijali za depress3ED.sav	335
	Preporučena literatura	337
	Bibliografija	339
	Spisak termina korišćenih u knjizi	343
	Indeks	345

Predgovor

Pomisao da moraju da uče neki statistički predmet ili da primene statistiku u svom istraživanju, za mnoge studente je izvor velikog stresa i frustracije. Prvo izdanje ove knjige (iz 2000. godine) trebalo je da posluži kao jednostavan postupni vodič kroz proces analize podataka pomoću SPSS-a. Za razliku od drugih knjiga iz statistike, on se nije bavio matematičkom osnovom statističkih tehnika, već prikladnom upotrebom SPSS-a kao alatke. Od pojavljivanja prva dva izdanja ovog priručnika, dobila sam stotine e-poruka studenata zahvalnih za pruženu pomoć (ili spas).

Isti, jednostavan pristup zadržan je i u ovom, trećem izdanju, zasnovanom na verziji 15 SPSS-a. Oduprla sam se iskušenju i pozivima mnogih studenata, predavača i recenzenata da obuhvatim mnogo novih tema, već sam umesto toga nadogradila i proširila postojeći materijal. Sva poglavlja su ažurirana tako da odgovaraju verziji 15 SPSS-a, mada je veći deo materijala podesan i za korisnike verzija 13 i 14. Na traženje brojnih korisnika prva dva izdanja, dodala sam opise još nekih mogućnosti SPSS-a, kao što su spajanje datoteka, upotreba skupova i upotreba komandnih datoteka. Dodati su odeljci o statističkom merenju veličine uticaja i o naknadnim testovima. Poglavlje o neparametarskim tehnikama prošireno je primerima njihove upotrebe u mnogim oblastima, naročito u zdravstvu, medicini i psihologiji. U nekim drugim poglavljima (npr. u onom o faktorskoj analizi), tekst je izmenjen da bi preporučeni postupak bio u skladu s literaturom i sa smernicama za objavljivanje u naučnoistraživačkim časopisima.

Ova knjiga ne obuhvata sve moguće statističke postupke koje SPSS podržava, niti odgovora na sva pitanja koja bi istraživač mogao imati u vezi sa statistikom. Umesto toga, osposobiće vas da počnete istraživanje i da steknete samopouzdanje prilikom korišćenja SPSS-a. Postoji još mnogo odličnih udžbenika iz statistike koje bi trebalo da pročitate; upućujem na njih u svakom poglavlju knjige. Dodatni materijal je na pratećoj Web lokaciji; pojedinosti o tome date su u narednom odeljku.

Datoteke s podacima i Web lokacija

Kroz celu knjigu sretaćete primere istraživanja uzete iz nekoliko datoteka s podacima (engl. *data files*) objavljenih na pratećoj Web lokaciji originalne verzije ove knjige, www.allenandunwin.com/spss.

S te lokacije možete preuzeti podatke na čvrsti disk, disketu ili fleš memoriju; biće dovoljno da sledite uputstva koja će se prikazati na ekranu. Zatim bi trebalo da pokrenete SPSS i otvorite te datoteke s podacima. One se mogu otvoriti samo u SPSS-u.

Datoteka **survey3ED.sav** sadrži podatke dobijene u istraživačkom projektu jednog od mojih odeljenja diplomaca. Da biste stekli osećaj za istraživački proces od početka do kraja, u dodatak sam stavila kopiju upitnika pomoću kojeg su ti podaci dobijeni, i šifarnik (engl. *codebook*) pomoću kojeg su ti podaci šifrovani. Zahvaljujući tome, možete i sami obaviti sve analize prikazane u ovoj knjizi i dalje eksperimentisati s drugim promenljivama.

Druga datoteka s podacima (**error3ED.sav**) sadrži isto što i **survey3ED.sav**, ali sam u nju namerno dodala greške da biste u poglavlju 5 mogli da vežbate pregledanje i ispravljanje podataka.

Treća datoteka (**experim3ED.sav**) sadrži izmišljene podatke i ilustruje upotrebu više tehnika obrađenih u petom delu knjige (npr. t-testa uparenih uzoraka, analize varijanse (ANOVA) ponovljenih merenja). U toj datoteci su i dodatne promenljive za vežbanje veština koje se uče na drugim mestima u knjizi. Samo nemojte previše da se obrađujete rezultatima koje ćete dobiti i ne pokušavajte da ih ponovite u nekom sopstvenom istraživanju!

depress3ed.sav je četvrta datoteka upotrebljena u primerima iz knjige. Nju koristimo u poglavlju 16 o neparametarskim tehnikama, kako bismo ilustrovali neke metode koje se upotrebljavaju u medicinskim istraživanjima.

Date su još dve datoteke, pa možete da uradite još ponešto s podacima iz različitih disciplina. Datoteka **sleep3ED.sav** sadrži stvarne podatke iz istraživanja uticaja poremećaja sna na život osoba koje pate od njega. Datoteka **stuffsurvey3ED.sav** potiče iz ankete o zadovoljstvu zaposlenih, sprovedene u jednoj velikoj australijskoj obrazovnoj ustanovi.

Dalje pojedinosti o tim datotekama (i prateće materijale) potražite u dodatku. Pored navedenih datoteka s podacima, Web lokacija ove knjige sadrži više stvari koje bi studentima i predavačima mogle dobro poslužiti, među kojima su:

- smernice za pripremu izveštaja o istraživanju;
- vežbe;
- izmene unete u nove verzije SPSS-a;
- hiperveze ka drugim korisnim Web lokacijama;
- spisak dodatne literature; i
- vodič za predavače.

Uvod i pregled

Ova knjiga je namenjena studentima koji pohađaju nastavu iz projektovanja i statističke obrade istraživanja, kao i onima koji samostalno planiraju i obavljaju istraživanja. Trebalo bi da vam ulije samopouzdanje da statističkim analizama pristupate spokojno i pribrano ili barem bez velikog stresa!

Dobar deo poteškoća koje studenti iskuse u statističkoj analizi potiču od anksioznosti i zbunjenosti zbog nepoznatog žargona, složenih teorija i prevelikog broja odluka koje moraju doneti. Nažalost, većina statističkih predmeta i udžbenika podstiču oba ta osećanja! Pokušala sam da u ovoj knjizi prevedem statistiku na jezik koji je lakše razumeti i svariti.

Knjiga *SPSS: priručnik za preživljavanje* veoma je pregledna zahvaljujući svojoj strukturi; u njoj se postupno, korak po korak izlaže ono što vam treba za pripremu i analizu podataka. Posmatrajte podatke kao sirove sastojke recepta. Sastojke možete pripremiti na različite načine: kao predjelo, glavno jelo ili desert. U zavisnosti od sastojaka koje imate, razne moguće obrade mogu biti prikladne ili neprikladne. (Nema smisla planirati pripremanje lovačkih šnicli ako imate samo piletinu.) Planiranje i priprema su važan deo postupka (i u kuvanju i u analizi podataka). Trebalo bi da razmislite o sledećim pitanjima:

- Imate li odgovarajuće sastojke u odgovarajućim količinama?
- Kakva priprema je potrebna da bi sastojci bili pogodni za kuvanje?
- Koju vrstu termičke obrade ćete upotrebiti (kuvanje, pečenje, pirjanje, prženje)?
- Imate li predstavu kako rezultat (recimo, kolač s čokoladnim prelivom) treba da izgleda?
- Kako ćete znati da je jelo gotovo?
- Kada kulinarska obrada bude gotova, kako ćete ga poslužiti a da izgleda privlačno?

Ista pitanja su jednako relevantna i za postupak analize podataka. Eksperiment ili anketu morate isplanirati tako da pruži potrebne informacije u odgovarajućem formatu. Datoteku morate pripremiti kako treba i podatke pažljivo uneti u nju. Morate imati jasnu predstavu o pitanjima na koje treba

dobiti odgovore i načinu na koji ćete ih formulirati. Morate znati koje statističke tehnike postoje i koje vrste podataka su za njih prikladne, a koje ne. Morate umeti da ispravno primenite odabranu statističku tehniku (npr. t-test) i da protumačite njen rezultat. Najzad, taj rezultat treba umešno da dovedete u vezu s prvobitnim istraživačkim pitanjima i predstavite u svom izveštaju. (Kada se prebacimo na kulinarstvo, tome bi odgovarala pitanja treba li kolač s čokoladnim prelivom da poslužite sa šlagom ili sa sladoledom, ili možda s jagodama/malinama i šćernom glazurom?).

Ni u kulinarstvu ni u analizi podataka ne možete samo izmešati sve sastojke, gurnuti ih u renu (odnosno SPSS) i nadati se da će rezultat biti dobar. Nadam se da će vam ova knjiga pomoći da malo bolje razumete postupak analize podataka i dati vam samopouzdanje i veštine potrebne dobrom „kuvaru“.

Struktura knjige

SPSS: priručnik za preživljavanje ima 22 poglavlja, koja obuhvataju ceo postupak istraživanja – od projektovanja studije do analize podataka i predstavljanja rezultata. Podeljena je na pet delova. Prvi deo (*Kako početi*) obuhvata ono uvodno: projektovanje studije, pripremu šifarnika (engl. *codebook*) i upoznavanje SPSS-a. U drugom delu (*Priprema datoteke s podacima*) naučićete kako se priprema datoteka, unose podaci i pronalaze greške u njoj. Preliminarne analize obrađene su u trećem delu knjige, čija su poglavlja posvećena primeni opisnih statističkih pokazatelja (engl. *descriptive statistics*) i dijagrama, doradi podataka i postupcima provere pouzdanosti mernih skala. Osim toga, provešću vas korak po korak kroz katkada težak zadatak izbora statističkih tehnika prikladnih za date podatke.

U četvrtom delu su predstavljene statističke tehnike za pronalaženje međuzavisnosti (npr. korelacija, delimična korelacija, višestruka regresija, logistička regresija i faktorska analiza). U tim poglavljima sažeto je opisano sledeće: namena svake tehnike, njene polazne pretpostavke, kako dobiti rezultate, kako tumačiti rezultate i kako ih predstaviti u naučnom radu ili izveštaju.

U petom delu se razmatraju statističke tehnike za poređenje grupa. Među njima su neparametarske tehnike, t-testovi, analiza varijanse (ANOVA), analiza varijanse više od dve promenljive (multivarijaciona analiza – MANOVA) i analiza kovarijanse (ANCOVA).

Kako koristiti ovu knjigu

Da biste ovu knjigu mogli delotvorno da upotrebite kao vodič kroz SPSS, morate posedovati neke osnovne veštine rada na računaru. U uputstvima i primerima polazim od pretpostavke da već umete da koristite PC računar, naročito Windowsove funkcije. U nastavku je spisak potrebnih veština. Ukoliko s bilo kojom od narednih radnji imate poteškoća, potražite nečiju pomoć:

- upotreba Windowsovih padajućih menija;
- upotreba levog i desnog tastera miša;
- biranje teksta tako što ga pritisnete i prevučete mišem preko njega;
- spuštanje i podizanje, to jest minimiziranje i maksimiranje prozora;
- pokretanje i zatvaranje programa iz menija Start ili iz Windows Explorera;
- prelazak iz jednog programa u drugi kada su istovremeno otvoreni;
- otvaranje, snimanje, preimenovanje, premeštanje i zatvaranje datoteka;
- rad s više datoteka istovremeno i prelazak iz jedne u drugu istovremeno otvorenu datoteku;
- kopiranje datoteka sa diskete ili fleš memorije na čvrsti disk ili obrnuto pomoću Windows Explorera;
- pravljenje direktorijuma (foldera, fascikle, omotnice) i prelazak iz jednog direktorijuma u drugi u Windows Exploreru.

Ova knjiga nije sveobuhvatna. Pretpostavlja se da ste već naučili osnove statistike i da imate odgovarajući udžbenik. Treba da imate predstavu o tome šta SPSS radi ispod površine. SPSS je izuzetno moćan softverski paket za analizu podataka, koji ume da obavlja veoma složene statističke procedure. U ovom priručniku se ne objašnjavaju sve statističke tehnike koje taj program podržava; obrađene su samo one koje se najčešće koriste. Knjiga bi trebalo da vam omogući da samostalno započnete statističku analizu i da steknete samopouzdanje u korišćenju SPSS-a.

U zavisnosti od podataka koje imate i istraživačkih pitanja (hipoteza) na koja tražite odgovor, možda ćete morati da upotrebite neku od složenijih analiza u SPSS-u. Ima mnogo dobrih knjiga o raznim statističkim tehnikama koje SPSS podržava. Čitajte sve što vam padne šaka. Prelistajte knjige na policama biblioteke u koje ste učlanjeni, potražite one koje statistiku objašnjavaju (barem delimično) razumljivim jezikom! Prikupite sav taj materijal u celinu koju ćete koristiti u svim statističkim predmetima i svom istraživačkom projektu. Umesno je prikupiti i članke iz stručnih časopisa u kojima se objašnjavaju statističke analize i predstavljaju njihovi rezultati. One vam mogu poslužiti kao uzor za pisanje sopstvenog izveštaja.

SPSS: priručnik za preživljavanje prikladan je i kao udžbenik za rad s predavačem koji objašnjava razne vidove istraživačkog postupka, i kao knjiga za samostalno učenje, za one koji treba samostalno da realizuju neki istraživački projekat. Kada budete učili, ne propustite da vežbate korišćenje SPSS-a za analizu podataka koji se mogu preuzeti s prateće Web lokacije ove knjige. (Pojediniosti o tome date su na str. xi.) Najbolje se uči uz rad, a ne samo čitajući kako to drugi rade. Pre nego što počnete da obrađujete sopstvene datoteke s podacima, poigrajte se onima iz kojih su uzeti primeri u ovoj knjizi. Tako ćete steći veće samopouzdanje i dobiti priliku da proverite da li ispravno radite analize.

Ponekad će vam se desiti da dobijete drugačiji rezultat od onoga u knjizi. To će se verovatno dogoditi ako budete upotrebljavali drugu verziju SPSS-a od one koja je korišćena u ovoj knjizi (SPSS za Windows, verzija 15). SPSS se stalno ažurira i to je odlično sa aspekta poboljšavanja programa, ali zbunjuje studente koji na ekranu svog računara ne vide rezultat naveden u knjizi. Razlika obično nije velika, pa ostanite pribrani i poigrajte se detektiva. Možda je informacija zapravo pred vama, samo u drugom obliku. Informacije o promenama SPSS-a za Windows potražite na pratećoj Web lokaciji.

Saveti za sprovođenje istraživanja

Evo još nekoliko saveta za one koji knjigu koriste kao vodič kroz sopstvene istraživačke projekte.

- **Pažljivo isplanirajte projekat.** Neka vam postojeće teorije i istraživanja posluže kao uzor prilikom planiranja sopstvenog projekta. Morate shvatiti šta pokušavate da dobijete, i zašto.
- **Razmišljajte unapred.** Očekujte i predvidite potencijalne probleme, pošto ih svaki projekat ima! Upoznajte statistički postupak koji namećete da primenite i pomoću tih informacija oblikujte materijale za prikupljanje podataka. Postarajte se da, kada dođe vreme za primenu statističkih metoda analize, imate vrstu podataka koji se njima *moгу* analizirati.
- **Organizujte se.** Pažljivo vodite beleške o svim relevantnim istraživanjima, referencama itd. Smislite delotvoran sistem arhiviranja gomile stručnih članaka koje ćete prikupiti i, kasnije, rezultata SPSS-a. Uopšte nije teško izgubiti se u gomili papira i brojki.
- **Vodite dobre beleške.** Kada pomoću SPSS-a budete analizirali podatke, pažljivo pišite šta radite. Svim studentima preporučujem da u svesci sa spiralnim povezom vode beleške o svakoj sesiji korišćenja SPSS-a. Zapišite datum, imena promenljivih koje ste napravili, analize koje ste sproveli i imena datoteka u koje ste snimili izlaz iz SPSS-a. Kada nađete na problem ili se datoteka s rezultatima na misteriozan način ošteti, instruktor će vas pomoću tih vaših beležaka spasiti!
- **Ostanite pribrani!** Ako prvi put analizirate podatke pomoću SPSS-a, dešavaće se da se osetite bespomoćni. Udahnite nekoliko puta duboko i sami sebi recite da sve ide dobro i da će sve biti dobro. Mislite samo o koraku koji neposredno predstoji – dozvolite sebi da ponekad pogrešite i da ne znate sami kako da nastavite. Kada vam bude svega dosta i previše, prekinite, prošetajte i razbistrite glavu pre nego što pokušate da nastavite rad. Većina studenata lako koristi SPSS kada ga upoznaju. Kao sa svakim drugim učenjem, treba samo prevazići početnu zbunjenost i nedostatak samopouzdanja.

- **Dajte sebi dovoljno vremena.** Istraživački postupak uvek traje duže nego što se očekivalo, pogotovo unošenje i analiza podataka, pa za te aktivnosti planirajte dosta vremena.
- **Radite s prijateljem.** Tokom analize podataka tražite emocionalnu i praktičnu pomoć od drugih studenata. Podrška drugih je odlična zaštita od stresa!

Dodatni resursi

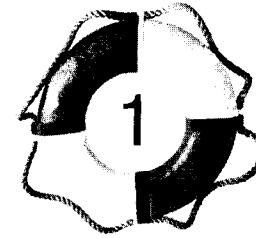
U ovoj knjizi obrađeno je više oblasti – od početnog projektovanja studije, sastavljanja upitnika, osnovnih statističkih tehnika (t-testovi, korelacija), pa sve do težih statističkih alatki (analiza varijanse više promenljivih, faktorska analiza). Po celoj knjizi posejane su preporuke za dodatnu literaturu. Pokušajte da čitate što više možete, naročito ako se budete bavili složenijim statističkim postupcima.

DEO I

Kako početi

Analiza podataka je tek jedan deo procesa istraživanja. Mnogo toga treba da se desi pre nego što budete mogli da pomoću SPSS-a analizirate podatke. Prvo, treba da isprojektujete istraživanje i odaberete odgovarajuća sredstva za prikupljanje podataka. Kada sprovedete istraživanje, dobijene podatke treba (pomoću tzv. šifarnika, engl. *codebook*) pripremiti za unošenje u SPSS. Da biste podatke mogli da unesete, morate znati kako SPSS radi i kako da s njim komunicirate. Svi ti koraci su objašnjeni u ovom, prvom delu knjige.

U poglavlju 1 su saveti i preporuke za projektovanje studije kojom se dobijaju podaci dobrog kvaliteta. U poglavlju 2 objašnjena je priprema šifarnika za prevođenje nenumeričkih podataka iz studije u format prikladan za SPSS. Poglavlje 3 je putovanje kroz SPSS; razmotrene su osnovne veštine koje će vam biti potrebne. Ukoliko vam je ovo prvi put da koristite SPSS, pročitajte poglavlje 3 pre nego što pređete na analize prikazane u nastavku knjige.



Projektovanje studije

Možda vam izgleda čudno da se u knjizi o SPSS-u razmatra projektovanje istraživanja, ali ono je nezaobilazan deo istraživačkog procesa, koji ima posledice po kvalitet podataka što će biti prikupljeni i analizirani. Podaci koje unosite u SPSS moraju odnekud doći: kao odgovori na upitnike, kao informacije prikupljene u intervjuima, kao šifrovana (brojevima izražena) opažanja stvarnog ponašanja ili kao objektivna merenja izlaza ili performansi. Podaci vrede samo koliko sredstvo kojim ste ih prikupili i istraživački okvir koji je usmeravao njihovo prikupljanje.

U ovom poglavlju razmotrićemo više vidova istraživačkog procesa koji utiču na kvalitet podataka. Prvo, razmotrićemo opšte projektovanje studije; zatim ćemo obraditi neka pitanja o kojima treba razmisliti prilikom izbora skala i mera; najzad, daćemo nekoliko smernica za pripremu upitnika.

Planiranje studije

Dobro istraživanje nije moguće bez pažljivog planiranja i sprovođenja studije. O projektovanju istraživanja napisano je mnogo odličnih knjiga, koje će vam pomoći tokom tog procesa: od pregleda literature, formulisanja hipoteza, izbora projekta studije, izbora i dodele subjekata (ispitanika), beleženja opažanja (opservacija) i prikupljanja podataka. Odluke koje se donose u svakoj od tih faza utiču na kvalitet podataka koji će se analizirati i na način obavljanja istraživanja, to jest traženja odgovora na istraživačka pitanja. Preporučila bih vam da ne štedite vreme prilikom projektovanja studije i da je

napravite što bolje možete. Koristiće vam da pročitate različite udžbenike o tome. Nekoliko dobrih naslova koji se lako čitaju jesu: Stangor (2006), Goodwin (2007) i, ako se bavite istraživanjem tržišta, Boyce (2003). Za zdravstvena i medicinska istraživanja, dobar elementarni pregled daje Peat (2001).

Za početak, razmislite o sledećim preporukama za projektovanje studije:

- Razmotrite koja vrsta istraživanja (npr. eksperiment, anketa, opservacija) omogućava najbolji pristup pitanju kojim se bavite. Svaka vrsta istraživanja ima svoje dobre i loše strane; izaberite pristup koji najbolje odgovara konkretnom pitanju na koje tražite odgovor. Prethodno morate dobro upoznati ranija istraživanja u toj oblasti.
- Ukoliko odaberete eksperiment, odlučite da li pitanju kojim se bavite bolje odgovara istraživanje različitih grupa (različiti subjekti u istoj eksperimentalnoj situaciji) ili ponovljena merenja (u različitim eksperimentalnim situacijama ispituju se isti subjekti). Oba pristupa imaju svoje prednosti i nedostatke (videti Stangor, 2006), pa ih pažljivo procenite.
- U eksperimentalnim istraživanjima, predvidite *dovoljno* nivoa nezavisne promenljive. Predvideti samo dva nivoa (ili grupe) znači da će biti potrebno manje subjekata, ali i da će zaključci koji se mogu izvesti biti ograničeni. Da li je kontrolna grupa neophodna ili uopšte poželjna? Da li će nepostojanje kontrolne grupe ograničiti zaključke koji se mogu izvesti?
- Uvek odaberite više subjekata nego što je potrebno, pogotovo kada su vam u uzroku ljudi. Ljudi su čuveni po svojoj nepouzdanosti: neće se pojaviti u dogovoreno vreme, razboleće se, odustaće ili neće popuniti upitnike kako treba! Zato računajte s tim unapred i isplanirajte istraživanje tako da mu to ne naudi. U neizvesnim situacijama pretpostavite da će ishod biti lošiji, a ne bolji.
- U eksperimentalnim istraživanjima predvidite dovoljno subjekata u svakoj grupi (i potrudite se da one budu približno jednake veličine, ako je moguće). Kada su grupe male, teško je otkriti statistički značajne tj. neslučajne razlike između njih (od čega zavisi moć istraživanja, kao što je objašnjeno u uvodu petog dela knjige). Potrebnu veličinu uzorka treba izračunati. Primera radi, pročitajte odgovarajući deo u knjizi koju je napisao Stangor (2006) ili nekom drugom udžbeniku za statistiku; potražite termin snaga ili moć (engl. *power*).
- Kad god je to moguće, nemojte koristiti postojeće grupe, nego svakoj eksperimentalnoj situaciji subjekte dodelite nasumično. Tako ćete u istraživanju različitih grupa (za razliku od ponovljenih merenja istih grupa) smanjiti problem neekvivalentnih grupa. Razmotrite i da li ćete dodatnim merenjima grupa dokazati da se one suštinski ne razlikuju. Možda ćete moći da (npr. analizom kovarijanse) statistički uračunate uticaj otkrivenih razlika između grupa.

- Izaberite prikladne zavisne promenljive koje su valjane (validne) i pouzdane; videti raspravu o tome u nastavku poglavlja. Bilo bi dobro da ispitete više različitih obeležja, pošto su neka obeležja osetljivija od drugih. Ne stavljajte sve na jednu kartu.
- Pokušajte da predvidite mogući uticaj remetilackih ili zbunjujućih promenljivih. To su promenljive pomoću kojih se rezultati mogu drugačije objasniti. Ponekad ih je teško uočiti kada čovek sam projektuje istraživanje, zato pre sprovođenja istraživanja obavezno neka neko drugi (supervizor, kolega istraživač) proveri projekat vašeg istraživanja. Učinite sve što možete da biste uračunali uticaj remetilackih promenljivih. U prepoznavanju mogućih remetilackih promenljivih, pomaže dobro poznavanje oblasti istraživanja. Ako ima dopunskih promenljivih koje ne možete da kontrolišete, možete li da ih izmerite? Time biste statistički uračunali njihov uticaj (npr. analizom kovarijanse).
- Kada radite anketu, najpre obavite probno (engl. *pilot*) istraživanje kojim ćete utvrditi da li su uputstva, pitanja i vrednosti na skalama jasni. Kad god je to moguće, probno istraživanje sprovedite među istom vrstom ljudi među kojima će i glavna studija biti sprovedena (npr. adolescentima, nezaposlenim mladim ljudima, zatvorenici). Morate obezbediti da ispitanici razumeju stavke ankete odnosno upitnika i da na njih odgovaraju na prikladan način. Probno istraživanje bi trebalo da otkrije i pitanja ili stavke koje bi mogle da uvrede potencijalne ispitanike.
- Kada sprovodite eksperiment, bilo bi dobro da održite generalnu probu i pilot-ispitivanjem proverite i eksperimentalni postupak i mere koje nameravate da upotrebite. Proverite kako radi oprema koju ćete koristiti. Ako će u vašem istraživanju učestvovati više pomoćnika u svojstvu eksperimentatora ili anketara, proverite da li su dobro obučeni i znaju li šta treba da rade. Ukoliko će više osoba posmatrati i ocenjivati ponašanja, proverite da li znaju ispravno da zabeleže svoje opservacije. Napravite probnu seriju i proverite pouzdanost njihovih ocena, tj. koliko su ocene različitih posmatrača međusobno saglasne. Pilot (ili probno) ispitivanje postupaka i merenja pomoći će vam da otkrijete šta bi moglo da zaškripi kada počne veliko istraživanje, i sve dodatne remetilacke činioce koji bi mogli da utiču na rezultate. Neke od njih nije moguće predvideti (npr. građevinske radove koji na dan početka velikog istraživanja mogu početi ispod prozora laboratorije), ali vi pokušajte da kontrolišete one činioce koje možete.

Izbor odgovarajućih skala i mera

Zavisno od prirode istraživanja, ima mnogo načina prikupljanja podataka. Mogu se meriti performanse ili učinak po nekom objektivnom kriterijumu ili ocenjivati ponašanje prema skupu specificiranih kriterijuma. Možda ćete

upotrebljavati merne skale projektovane da „operacionalizuju“ određeni pripadni osnovni konstrukt ili atribut koji se ne može izmeriti neposredno (npr. samopoštovanje). U istraživanju se može upotrebiti na hiljade overenih skala. Ponekad nije lako pronaći onu koja je prikladna za vašu svrhu. Treba početi temeljnim pregledom literature u oblasti od interesa. Koje mere su upotrebljavali drugi istraživači u toj oblasti? Katkada se u dodatku stručnog članka navode vrednosti od kojih se sastoje merne skale; inače ćete možda morati da pronađete originalan članak u kome su opisani projekat i validacija skale od interesa. Neke skale su zaštićene kao intelektualno vlasništvo, što znači da biste od izdavača morali da kupite zvanične kopije koje ćete koristiti. Ostale skale, u potpunosti objavljene u stručnim člancima, predstavljaju javno vlasništvo, što znači da ih istraživači mogu upotrebljavati bez naknade. Međutim, veoma je važno da navedete odakle potiče svaka merna skala koju ste upotrebili i sve detalje odgovarajućih referenci.

Prilikom izbora skala morate uzeti u obzir njihovu pouzdanost i valjanost (validnost). Na kvalitet dobijenih podataka mogu da utiču oba ta faktora. Tokom razmatranja mogućih skala, treba prikupiti informacije o pouzdanosti i valjanosti svake od njih. Te informacije će vam trebati za odeljak „Metodologija“ izveštaja o istraživanju. Bez obzira na to koliko su pohvalni objavljeni izveštaji o pouzdanosti i valjanosti odabranih skala, ne propustite da i njih podvrgnete probnom ispitivanju na predviđenom uzorku. Neke skale su pouzdane unutar određenih grupa (npr. odraslih iz porodica koje govore engleski), ali su potpuno nepouzdate kada se upotrebe unutar drugih grupa (npr. dece iz porodica koje ne govore engleski).

Pouzdanost

Pouzdanost (engl. *reliability*) merne skale pokazuje stepen njene otpornosti na slučajne greške. Dva često korišćena pokazatelja pouzdanosti skale jesu njena *vremenska stabilnost*, engl. *test-retest reliability* (stabilnost njenih rezultata pri ponovljenoj upotrebi na istom uzorku) i *unutrašnja saglasnost* (engl. *internal consistency*). Vremenska stabilnost se meri primenom iste skale na istim subjektima u dva navrata; zatim se izračuna korelacija između tako dobijenih rezultata. Pouzdanija je skala koja ima visoku korelaciju. Prilikom razmatranja ove vrste pouzdanosti, treba uzeti u obzir prirodu konstrukta koji skala meri. Skala projektovana da meri stanja tekućeg raspoloženja teško da može ostati stabilna tokom više sedmica. Stoga će vremenska stabilnost skale raspoloženja verovatno biti niska. Međutim, mere svojstava (obeležja) stabilnih ličnosti trebalo bi da ostanu gotovo jednake i da imaju visoku korelaciju rezultata pri ponovnoj upotrebi skale.

Drugi vid pouzdanosti skale koji se može oceniti jeste njena unutrašnja saglasnost. To je stepen do kojeg vrednosti koje čine skalu mere isti pripadni atribut (tj. do kojeg su međusobno povezane). Unutrašnja saglasnost se meri na više načina. Najčešće se upotrebljava Kronbahov (Cronbach) koeficijent

alfa, koji i SPSS ume da izračuna; videti u poglavlju 9. To je prosečna korelacija između svih vrednosti na skali. Iznos tog pokazatelja je naravno između 0 i 1, pri čemu veći broj (viša korelacija) pokazuje veću pouzdanost.

Zavisno od prirode i namene skale, zahtevaju se različiti nivoi pouzdanosti, ali Nunnally (1978) preporučuje da se ne prihvata pouzdanost manja od 0,7. Kronbahov koeficijent alfa se menja u zavisnosti od broja vrednosti na skali. Za mali broj vrednosti na skali (manji od 10), Kronbahov koeficijent alfa ponekad je veoma mali. Tada je bolje izračunati i u izveštaju navesti srednju vrednost korelacije između svakog para vrednosti. Optimalna srednja vrednost korelacije između parova vrednosti na skali iznosi između 0,2 i 0,4 (po preporuci navedenoj u Briggs & Cheek, 1986).

Validnost

Validnost skale je stepen do kojeg ona meri ono što bi trebalo da meri. Nažalost, ne postoji jasan pokazatelj validnosti skale. Validacija (provera validnosti) skale podrazumeva prikupljanje empirijskih dokaza o njenoj upotrebi. U člancima se uglavnom raspravlja o validnosti sadržaja, validnosti kriterijuma i validnosti konstrukta.

Validnost sadržaja se odnosi na adekvatnost s kojom je uzet uzorak za meru ili skalu iz predviđenog univerzalnog skupa ili domena sadržaja. *Validnost kriterijuma* je odnos između rezultata na skali i određenog merljivog kriterijuma. *Validnost konstrukta* znači ispitivanje skale ne prema jednom kriterijumu, nego prema teorijski izvedenim hipotezama o prirodi pripadne promenljive ili konstrukta. Validnost konstrukta se istražuje ispitivanjem njegovog odnosa s drugim konstruktima, kako srodnim (konvergentna validnost), tako i onim nesrodnim (diskriminantna validnost). Lako čitljiv i sažet prikaz raznih vrsta validnosti možete pročitati u udžbenicima čiji su autori Stangor (2006), odnosno Streiner i Norman (2003).

Ukoliko u svom istraživanju nameravate da koristite merne skale, bilo bi dobro da o tome još nešto pročitate: u udžbeniku koji je napisao Kline (2005) videti priču o psihološkom testiranju, a u već pomenutom udžbeniku Streinera i Normana (2003) o skalama za merenje zdravlja. I Bowling je napisao nekoliko odličnih knjiga o mernim skalama za zdravstvo i medicinu.

Priprema upitnika

U mnogim istraživanjima, informacije se prikupljaju od subjekata ili ispitanika. Ponekad to znači prikupljanje demografskih podataka od subjekata pre nego što ih se podvrgne nekom eksperimentalnom postupku. S druge strane, možda ćete projektovati obimnu anketu koju treba razdeliti odabranom uzorku stanovništva. Slabo isplaniran ili loše projektovan upitnik neće dati dobre podatke iz kojih se može izvući kvalitetan odgovor na istraživana pitanja. Prilikom pripreme upitnika razmislite o tome kako ćete upotrebiti te informacije; morate unapred znati koje statističke analize ćete primeniti. U zavisnosti od predviđene statističke tehnike, pitanje treba postaviti na

određeni način ili dati različite formate odgovora. U narednim odeljcima navedeni su neki od činilaca koje treba razmotriti prilikom projektovanja i sastavljanja upitnika.

Pošto ćemo samo malo zagrebat po površini teme projektovanje upitnika, preporučujem da o tome nešto više pročitate ukoliko sami projektujete svoje istraživanje. Za to je odlična literatura De Vaus (2002), odnosno Boyce (2003) za istraživanja poslovanja.

Vrste pitanja

Većina pitanja se može razvrstati u dve grupe: zatvorena i otvorena. Zatvoreno pitanje je ono kada ispitanicima nudite više unapred definisanih odgovora. Od njih se traži da jedan od ponuđenih odgovora odaberu zaokruživanjem, podvlačenjem, precrtavanjem ili štrikliranjem pripadajućeg polja i sl. Ponudeni odgovori mogu biti prosto *Da/Ne*, *Muško/Žensko*; ili se može ponuditi veći broj mogućih odgovora, na primer:

Koja je najviša škola koju ste završili (označite odgovarajuće polje)?

- ☐ 1. Osnovna škola
- ☐ 2. Deo srednje škole
- ☐ 3. Cela srednja škola
- ☐ 4. Viša škola ili dodatna obuka
- ☐ 5. Dodiplomske studije na univerzitetu
- ☐ 6. Postdiplomske studije na univerzitetu

Zatvorena pitanja je obično veoma lako pretvoriti u numerički format koji je potreban SPSS-u. Primera radi, *Da* se može šifrovati sa 1, *Ne* sa 2; *Muško* sa 1, *Žensko* sa 2. U prikazanom pitanju o obrazovanju, odabrani odgovor bi se šifrovaao unošenjem rednog broja označenog polja ispred odgovora. Na primer, ako ispitanik označi polje ispred odgovora *Dodiplomske studije na univerzitetu*, to bi se šifrovalo brojem 5. Postupak šifrovanja je lakši kada se svakom od mogućih odgovora dodeli neki broj. Za potrebe unošenja podataka, odaberite neko pravilo za dodeljivanje brojeva (npr. brojevi rastu sleva nadesno i odozgo naniže) i pridržavajte ga se kroz ceo upitnik.

Ponekad se ne može pretpostaviti koje bi sve odgovore ispitanici mogli dati, te je neophodno zadati otvorena pitanja. Prednost je to što ispitanici tada imaju priliku da odgovore na svoj način, jer nisu ograničeni izborom odgovora koje im pruža istraživač. Primera radi:

Koji je za vas trenutno glavni izvor stresa?

Radi unošenja u SPSS, odgovori na otvorena pitanja mogu se sažeti u nekoliko kategorija. One se obično utvrđuju nakon sagledavanja raspona svih odgovora dobijenih od ispitanika. I iz prethodnih istraživanja te oblasti mogu se izvući moguće kategorije. Svakoј kategoriji odgovora dodeljuje se druga šifra, to jest broj (npr. radno mesto=1, finansijsko stanje=2, ljubavni odnosi=3) i taj broj se unosi u SPSS. Više pojedinosti o tome sadrži odeljak o pripremi šifarnika u poglavlju 2.

Ponekad je najbolja kombinacija zatvorenih i otvorenih pitanja. To znači da se ispitanicima daje nekoliko unapred definisanih odgovora i jedna dodatna kategorija (*ostalo*) koju mogu odabrati ukoliko željeni odgovor nije među ponuđenima. Sledi jedan ili dva prazna reda sa crtom na koju mogu upisati svoj odgovor. Takva kombinacija zatvorenih i otvorenih pitanja naročito je korisna u ranim fazama istraživanja date oblasti, pošto pokazuje da li definisane kategorije odgovora adekvatno obuhvataju odgovore koje ispitanici najčešće daju.

Format odgovora

Prilikom formulisanja pitanja, morate odlučiti i koji ćete format odgovora ponuditi ispitanicima. Format odgovora može imati posledice po statističke analize koje se na te odgovore mogu primeniti. Za neke analize (npr. korelacija) potrebni su odgovori čije vrednosti čine širok, kontinualan opseg od malih do velikih brojeva. Ako za pitanje o njihovoj starosti ispitanicima ponudite kategorije odgovora koje treba da izaberu (npr. manje od 30, između 31 i 50, i preko 50 godina starosti), takvi podaci neće biti prikladni za korelacionu analizu. Stoga, ukoliko nameravate da istražite korelaciju između starosti osobe i, recimo, njenog samopoštovanja, ispitanicima bi trebalo omogućiti da upišu koliko imaju godina.

Pokušajte da ponudite što širi izbor odgovora na svoja pitanja. Odgovore kasnije uvek možete sažeti, ako treba (videti poglavlje 8). Nemojte ispitanike pitati samo da li se slažu s nekom tvrdnjom ili ne; upotrebite skalu Likertovog tipa, sa stupnjevitim odgovorima – od potpunog slaganja do potpunog neslaganja:

nimalo se ne slažem 1 2 3 4 5 6 7 8 9 10 slažem se potpuno

Ta vrsta skale omogućava širi raspon mogućih odgovora i povećava broj statističkih analiza koje se na takve odgovore mogu primeniti. Moraćete da odlučite koliko ćete stupnjeva u odgovorima (npr. od 1 do 10) koristiti. DeVellis (2003) dobro razmatra prednosti i mane raznih skala odgovora. Koji god format odgovora da izaberete, morate dati i jasna uputstva za njega. Želite li da ispitanici jedan od ponuđenih odgovora odaberu označavanjem polja (kućice), zaokruživanjem broja ili možda upisivanjem određenog

znaka na crtu? Vaš upitnik može biti prvi koji ispitanik popunjava u svom životu. Nemojte pretpostaviti da ispitanici već znaju kako se odgovara. Dajte jasna uputstva – po mogućstvu i primer – i uvek sprovedite probno ispitivanje one vrste ljudi od kojih će se vaš uzorak sastojati. Pre nego što razdelite stotine upitnika, uklonite sve moguće izvore zabune. Kada sastavljate pitanja, uvek razmislite o tome kako bi sve ispitanik mogao da shvati to pitanje i razmotrite sve moguće odgovore koje bi neko na to pitanje mogao da da. Na primer, hteli biste da saznate da li osoba puši ili ne. Pitanje bi moglo da glasi:

Da li pušite? (označite polje)

☐ Da

☐ Ne

Kada bude popunjavao upitnik, ispitanik bi se mogao zapitati da li se misli na cigarete, lulu ili možda na marihuanu? Da li je dovoljno znati samo puši li osoba? Treba li da saznate i koliko puši (dve-tri cigarete ili dve-tri paklice dnevno), koliko često puši (svaki dan ili samo u izuzetnim prilikama, u društvu)? Smisao ovoga je da razmislite o svakom pitanju: koje informacije će vam dati a koje neće, a potrebne su vam.

Sastavljanje pitanja

Za sastavljanje jasnih i dobro sročenih pitanja u upitniku zaista treba umeće. Iako ne postoje jednostavna pravila kako se to dobro radi, znaju se greške koje sigurno vode do loših upitnika. Pokušajte da izbegnete:

- dugačka, složena pitanja;
- dvostruke negacije;
- dva pitanja u jednom;
- žargon i skraćenice;
- pojmove i izraze čije se značenje menja zavisno od kulture kojoj osoba pripada;
- višeznačne reči;
- pitanja koja navode na odgovor; i
- reči koje pobuđuju jaka osećanja.

U odgovarajućim slučajevima trebalo bi da razmislite o uključivanju kategorije „Ne znam“ ili „Nije primenljivo“. Dalje preporuke o tome kako se pišu pitanja videti u knjigama koje su napisali De Vaus (2002) i Kline (2005).



Priprema šifarnika

Pre nego što počnete da unosite informacije iz upitnika, intervju a ili eksperimenta u SPSS, treba da pripremite šifarnik (engl. *codebook*). To je zbir instrukcija za pretvaranje informacija dobijenih od svakog subjekta ili slučaja (engl. *case*) u numerički format razumljiv SPSS-u, tj. u brojeve. U ovom poglavlju pokazaćemo korake od kojih se ta priprema sastoji, i to na primeru datoteke s podacima koju je osmislila grupa mojih postdiplomaca. Kopiju tog upitnika i šifarnik napravljen za njega naći ćete u dodatku na kraju knjige. Datoteku s podacima možete preuzeti s prateće Web lokacije. Kada imate sav taj materijal, možete da sagledate ceo postupak – od sastavljanja upitnika do pravljenja završne datoteke s podacima, spremne za analizu. Iako sam za ilustraciju koraka u razvoju šifarnika upotrebila upitnik, postupak je sličan i u eksperimentalnim istraživanjima.

Priprema šifarnika obuhvata donošenje odluke (i dokumentovanje) kako ćete:

- definisati i nazvati svaku promenljivu; i
- dodeliti po jedan broj svakom od mogućih odgovora.

Sve te informacije treba zabeležiti na papiru ili u datoteci. Držite to na nekom bezbednom mestu; nema ništa gore nego vratiti se datoteci s podacima koju neko vreme niste koristili i onda se pitati šta znače unete skraćenice i brojevi.

U šifarniku bi trebalo da navedete sve promenljive iz upitnika, skraćena imena promenljivih koja ćete upotrebljavati u SPSS-u i način na koji ćete šifrovati odgovore (pridružiti im brojeve koje će SPSS statistički analizirati). U ovom poglavlju dati su pojednostavljeni primeri koji ilustruju svaki korak. U prvoj koloni tabele 2.1 navedeno je ime promenljive (na srpskom, a ne računarskom jeziku). U drugoj koloni je skraćeno ime promenljive koje će se upotrebljavati u SPSS-u (videti dolenavedenu konvenciju), a u trećoj koloni pišete broj (šifru) koji ćete pridružiti svakom od dobijenih odgovora.

Imena promenljivih

Svako pitanje ili stavka u upitniku mora imati svoju promenljivu jedinstvenog imena. Neka od tih imena jasno će pokazivati o kakvoj se informaciji radi (recimo, pol, starost). Druga pitanja, kao što su vrednosti od kojih se sastoji neka merna skala, mogu biti identifikovana određenom skraćenicom (npr. op1, op2, op3 kao imena stavki od kojih se sastoji skala optimizma).

Tabela 2.1: Primer šifarnika

Promenljiva	Ime promenljive u SPSS-u	Uputstvo za šifrovanje (dodelu brojeva)
Identifikacioni broj	ID	Broj dodeljen svakoj anketi
Pol	Pol	1 = muški 2 = ženski
Bračno stanje	Brak	1 = neoženjen/neudata 2 = stalan partner 3 = oženjen/udata prvi put 4 = ponovo oženjen/udata 5 = razveden/a ili živi razdvojeno 6 = udovac/udovica
Stavke 1 do 5 skale optimizma	Od op1 do op5	Unesi broj koji je zaokružen, od 1 (nimalo se ne slažem) do 5 (potpuno se slažem)

Prilikom dodeljivanja imena promenljivima u SPSS-u morate se pridržavati više pravila. Ona su navedena u izdvojenom odeljku „Pravila za imenovanje promenljivih“. U ranijim verzijama SPSS-a (pre verzije 12), imena promenljivih su smela imati najviše 8 znakova. U verziji 12 SPSS-a dozvoljena su do 64 znaka. Ukoliko datoteke s podacima nameravate da koristite u raznim verzijama SPSS-a (npr. u računarskim laboratorijama na fakultetu), bezbednije je da promenljivima i dalje dajete imena od najviše osam znakova.

Pravila za imenovanje promenljivih

Imena promenljivih:

- moraju biti jedinstvena (tj. imena svih promenljivih datog skupa moraju se međusobno razlikovati);
- moraju početi slovom (a ne brojem);
- ne smeju sadržati tačke, razmake (beline), niti specijalne znakove (!, ?, ' " *);
- ne smeju sadržati reči koje su komande u SPSS-u (all, ne, eq, to, le, lt, by, or, gt, and, not, ge, with); i
- ne smeju sadržati više od 64 znaka (za verziju 12 SPSS-a), odnosno osam znakova za ranije verzije SPSS-a.

Prva promenljiva u svakom skupu podataka trebalo bi da bude ID, tj. jedinstven broj koji identifikuje svaki analizirani slučaj. Pre nego što postupak unošenja podataka počne, prođite kroz sve upitnike odnosno zapise podataka i svakom od njih dodelite jedinstven broj. Zapišite taj broj čitko na prednjoj korici. Kasnije, ako u skupu podataka pronađete grešku, pomoću jedinstvenih brojeva ćete lako otkriti onaj s greškom.

Šifrovanje odgovora

Pre unošenja u SPSS, svakom odgovoru mora biti dodeljen određen broj (šifra, numerički kôd). Deo informacija će već biti u numeričkom formatu (npr. starost u godinama); ostale promenljive, kao što je pol, treba pretvoriti u brojeve (npr. 1=muški, 2= ženski). Ako ste ponuđene odgovore na pitanja označili brojevima (videti npr. pitanje o nivou obrazovanja u poglavlju 1), nema razloga da upravo te brojeve ponovo ne upotrebite u SPSS-u kao šifre datih odgovora. Ukoliko to iz nekog razloga nećete, izaberite neko pravilo za dodelu šifara odgovorima i pridržavajte ga se. Primera radi, prvi ponuđeni odgovor šifrujte brojem 1, drugi brojem 2 itd.

Koje je vaše trenutno bračno stanje? (označite odgovarajuće polje)

☐ neoženjen/neudata ☐ stalan partner ☐ oženjen/udata ☐ razveden/a

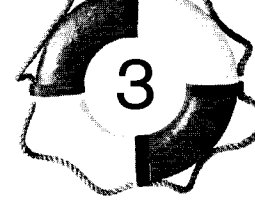
Da biste šifrovali odgovore na gornje pitanje: ako je osoba označila polje ispred odgovora *neoženjen/neudata*, šifrujte to brojem 1; odgovor *stalan partner* šifrujte brojem 2; *oženjen/udata* neka bude 3; i *razveden/a*, 4.

Šifrovanje otvorenih pitanja

Šifrovanje otvorenih pitanja (na koja ispitanici sami formulišu odgovore) malo je komplikovanije. Na primer, uzmimo pitanje: *Koji je za vas trenutno glavni izvor stresa?* Da biste šifrovali odgovore na to pitanje, pročitajte sve upitnike i potražite najčešće odgovore. Možda će mnogo ispitanika odgovoriti da im je glavni izvor stresa: radno mesto, finansijsko stanje, ljubavni odnosi, bolest ili nedostatak slobodnog vremena. U šifarniku ćete te glavne grupe odgovora navesti kao vrednosti promenljive koju ćete nazvati *stres* i svakoj ćete dodeliti po jedan broj (npr. radno mesto=1, suprug/a ili partner=2 itd.). Trebaće vam još jedan broj za one odgovore koji ne spadaju ni u jednu od navedenih kategorija (ostalo=99). Kada budete unosili podatke za svakog ispitanika, uporedite njen/njegov odgovor sa onima navedenim u šifarniku i unesite odgovarajući broj (šifru) u skup podataka za promenljivu *stres*. Nakon što napravite šifarnik, gotovo da ste spremni za unošenje podataka. Prethodno treba da uradite još dve stvari:

1. upoznate SPSS, razne njegove prozore i okvire za dijalog, te kako se otvaraju i zatvaraju datoteke.
2. formirate datoteku s podacima pomoću informacija koje ste pripremili u šifarniku.

U poglavlju 3 opisani su osnovna struktura SPSS-a i pravila koja u njemu važe, a u poglavlju 4 – postupci izrade datoteke i unošenje podataka.



Upoznajte SPSS

Pre nego što počnete da radite u SPSS-u, trebalo bi da naučite nekoliko osnovnih stvari o njemu. Prvo, SPSS razne stvari radi u različitim prozorima. Da biste mogli da pristupite tim prozorima, prethodno morate otvoriti neku od postojećih datoteka s podacima ili napraviti novu. Zato ćemo u ovom poglavlju objasniti kako se pokreće i zatvara SPSS; kako se otvaraju i zatvaraju postojeće datoteke s podacima; i kako se pravi nova (prazna) datoteka za podatke. Potom ćemo pogledati kakve prozore ima SPSS.

Pokretanje SPSS-a

SPSS se može pokrenuti na više načina:

- Najjednostavnije je potražiti ikonicu SPSS-a na radnoj površini u Windowsu. Postavite kursor iznad te ikonice i dvaput je pritisnite.
- SPSS možete pokrenuti i tako što ćete pritisnuti dugme **Start**, dovesti kursor do stavke **Programs**, i zatim preći u spisak dostupnih programa. Prelazite po tom spisku naviše i naniže dok ne nađete **SPSS for Windows**.
- SPSS će se pokrenuti i kada dvaput pritisnete neku SPSS datoteku s podacima prikazanu npr. u Windows Exploreru. Te datoteke imaju oznaku tipa (engl. *extension*) **.sav**.

Kada pokrenete SPSS, možda će vas dočekati sivi početni prozor s pitanjem „What would you like to do?“ (Šta biste hteli da uradite?) Lakše je zatvoriti taj prozor (pritiskom na dugme označeno sa x u gornjem desnom uglu) i navići se na upotrebu SPSS-ovih menija. Kada zatvorite početni prozor, prikazaće se prazna proračunska tabela (engl. *spreadsheet*). Da biste iz ovog prozora otvorili neku od postojećih SPSS datoteka s podacima, pritisnite **File** i zatim **Open** u meniju koji će se prikazati na vrhu prozora.

Otvaranje postojeće datoteke s podacima

Ukoliko hoćete da otvorite postojeću datoteku s podacima (npr. **survey3ED**), koja je, između ostalih, dostupna na pratećoj Web lokaciji – videti str. xi), pritisnite **File** u glavnom meniju na vrhu prozora, zatim **Open** u padajućem meniju koji će se otvoriti i najzad **Data**. U okviru za dijalog **Open File** pretražite direktorijume na računaru dok ne pronađete datoteku u kojoj su podaci smešteni.

Datoteke s podacima uvek treba otvarati sa čvrstog diska računara. Ako su vam podaci na disketi ili na USB disku, prebacite ih u direktorijum na čvrstom disku pre otvaranja. Potom pronađite odgovarajuću datoteku na čvrstom disku i pritisnite **Open**. Ne zaboravite da sve SPSS datoteke s podacima imaju oznaku tipa **.sav**. Datoteka će se otvoriti u prozoru **Data Editor** (više o njemu kasnije).

Savet Podatke s kojima nameravate da radite uvek najpre prebacite na čvrsti disk računara. Na kraju radne sesije, kopirajte te datoteke nazad na USB disk ili disketu.

Korišćenje datoteka s podacima

U verziji 15 SPSS-a možete imati više istovremeno otvorenih datoteka s podacima. To ume da bude korisno, ali i da zbuni. U svakom trenutku morate imati otvorenu barem jednu datoteku s podacima. Ako pokušate da zatvorite datoteku, SPSS će vas pitati želite li da je snimate pre zatvaranja. Ukoliko je ne snimate, izgubićete sve podatke koje ste u toj sesiji u nju uneli i nove promenljive kojima ste promenili šifre ili izračunali nakon poslednjeg otvaranja datoteke.

Snimanje datoteke s podacima

Trebalo bi da snimate datoteku s podacima kad god je izmenite (recimo, ako ste definisali nove promenljive) ili čim je napravite. Datoteke se ne snimaju automatski, kao u većini programa za obradu teksta. Ukoliko ne steknete naviku da povremeno snimate datoteku na kojoj radite, a nestane struja ili slučajno pritisnete pogrešan taster (i to se dešava), izgubićete sve što ste radili. Da se ne biste nervirali, redovno snimajte. Kada unosite podatke, snimajte svakih deset minuta, ili nakon svakih pet ili deset unetih upitnika.

Da biste snimili datoteku na kojoj radite, otvorite meni **File** (u gornjem levom uglu) i u njemu izaberite **Save**. Ukoliko vam je tako lakše, na paleti sa alatima gore levo pritisnite dugme sa sličicom diskete. Iako je na sličici disketa, kada je pritisnete snimićete datoteku na onaj disk na kojem trenutno radite. To bi uvek trebalo da bude čvrsti disk, zato što je rad sa diskete recept za katastrofu! Godinama mi studenti dolaze uplakani nakon što svoju datoteku s podacima upropaste radeći na disketi umesto na čvrstom disku.

Prilikom prvog snimanja nove datoteke s podacima, moraćete da zadate njeno ime i direktorijum u koji će biti snimljena. Izaberite direktorijum i upišite ime pod kojim ćete snimiti datoteku. SPSS svim datotekama s podacima koje snima automatski daje oznaku tipa **.sav**, zato što samo tako ume da ih prepozna. Ne menjajte tu oznaku, inače SPSS neće umeti da pronađe datoteku kada je sledeći put budete tražili.

Otvaranje druge datoteke s podacima

Kada završite rad na jednoj datoteci i zatreba vam druga, samo pritisnite **File** i zatim **Open**, te pronađite direktorijum u kojem je ta druga datoteka sačuvana (snimljena). Pritisnite željenu datoteku, pa onda dugme **Open**. U verziji 15 SPSS-a tako ćete otvoriti drugu datoteku s podacima, dok će prva ostati otvorena u zasebnom prozoru. Bilo bi dobro da zatvorite sve datoteke na kojima trenutno ne radite, pošto više otvorenih datoteka ume da zbuni.

Pravljenje nove datoteke za podatke

U SPSS-u je lako napraviti novu, praznu datoteku. Pritisnite **File**, pa u padajućem meniju pritisnite **New** i najzad **Data**. Od tog trenutka možete definisati nove promenljive i unositi podatke. Međutim, pre toga treba da upoznate SPSS-ove prozore i okvire za dijalog. Razmotrićemo ih u narednom odeljku.

SPSS-ovi prozori

U SPSS-u ćete uglavnom koristiti prozore **Data Editor**, **Viewer**, **Pivot Table Editor**, **Chart Editor** i **Syntax Editor**. Ovde ćemo ih prikazati sažeto, a u nastavku knjige mnogo detaljnije.

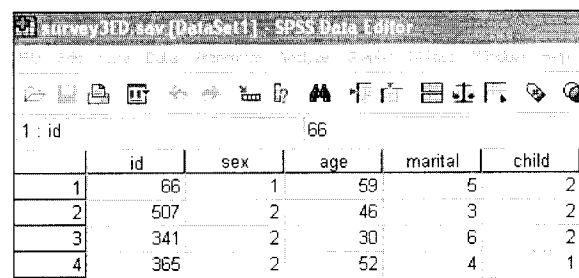
Kada počnete da analizirate podatke, imaćete više tih prozora istovremeno otvorenih. Pojedine studente to zbunjuje, ali je zapravo sasvim jednostavno kada shvatite zašto su otvoreni. Prozor **Data Editor** je uvek otvoren zato što sadrži datoteku s podacima koje analizirate. Prozor **Viewer** se otvara kada treba prikazati rezultate analize; u njemu se prikazuju rezultati svih analiza, navedeni redom kojim su analize obavljane. Molim vas, zapamtite: ovaj prozor se neće otvoriti dok ne pokrenete neku analizu.

Prozori na ekranu računara su kao parčici papira na stolu – možete ih premeštati, nekad je jedan odozgo, a nekad drugi. Svaki otvoren prozor ima svoje dugme na dnu ekrana. Kada zatreba da predete u određeni otvoren prozor, pritisnite njegovo dugme na dnu. Mogli biste i da pritisnete **Windows** u glavnom meniju, kako bi se prikazala lista svih otvorenih prozora; pritisnite u toj listi prozor koji hoćete da prikazete ispred svih ostalih.

Ponekad prozori koje SPSS prikazuje ne prekrivaju ceo ekran. Biće vam lakše kada prozor **Viewer** (sa svim rezultatima) uvećate tako da prekrije ceo ekran. Da biste to uradili, pogledajte gornji desni ugao ekrana. Tamo bi trebalo da su prikazana tri mala dugmeta. Pritisnite srednje dugme da biste prozor uvećali do kraja (tako da prekrije ceo ekran). Kada poželite da ga ponovo smanjite, opet pritisnite to srednje dugme.

Prozor Data Editor

Prozor **Data Editor** prikazuje sadržaj datoteke s podacima, i u tom prozoru otvarate, snimate i zatvarate postojeće datoteke, pravite nove datoteke, unosite nove i menjate postojeće podatke, i obavljate statističke analize (slika 3.1).



	id	sex	age	marital	child
1	66	1	59	5	2
2	507	2	46	3	2
3	341	2	30	6	2
4	365	2	52	4	1

Slika 3.1 Prozor Data Editor.

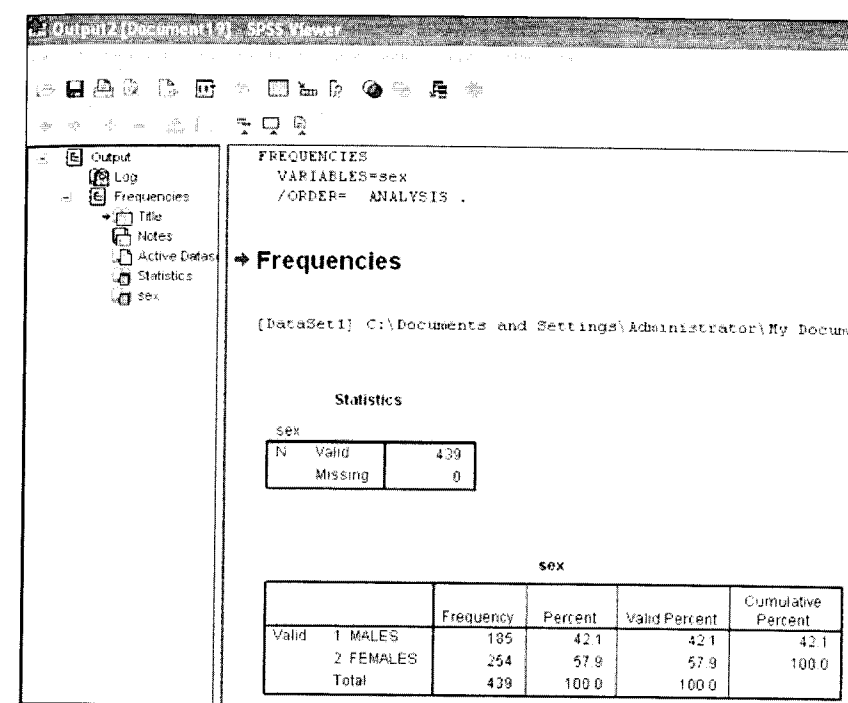
Prozor Viewer

Kada pokrenete bilo koju analizu, automatski će se otvoriti prozor **Viewer** (slika 3.2). U njemu se prikazuju rezultati svih obavljenih analiza, tabele i dijagrami. U tom prozoru rezultate možete menjati, brisati, kopirati, snimati ili prebacivati u Wordove dokumente. Kada SPSS-u zadate da snimi rezultate svojih statističkih analiza, on će ih snimiti u datoteku sa oznakom tipa **.spo** (za razliku od datoteka s podacima koje imaju oznaku tipa **.sav**).

Prozor **Viewer** se sastoji od dva dela. Levo je okno sa strukturiranim stablom (menijem) analiza, u kome je kompletan spisak svih analiza koje ste obavili. To stablo upotrebite da biste se brzo kretali kroz rezultate (kojih može biti veoma mnogo). Samo pritisnite onaj deo rezultata na koji hoćete da predete i on će se prikazati u desnom oknu prozora. Na desnoj strani prozora **Viewer** prikazuju se rezultati analiza, uključujući tabele i dijagrame (grafikone).

Snimanje rezultata

Da biste snimili rezultate svojih analiza, ispred svih ostalih prozora morate imati otvoren prozor **Viewer**. Pritisnite **File** u glavnom meniju na vrhu tog prozora.



		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 MALES	185	42.1	42.1	42.1
	2 FEMALES	254	57.9	57.9	100.0
	Total	439	100.0	100.0	

Slika 3.2 Prozor Viewer.

Pritisnite stavku **Save** u padajućem meniju koji će se otvoriti. Izaberite direktorijum u koji snimate rezultate i upišite jedinstveno ime datoteke s rezultatima. Pritisnite dugme **Save**. Kada datotekama dajem imena, koristim skraćenice koje ukazuju na datoteku s podacima čijom analizom su rezultati dobijeni i datum kada je analiza obavljena. Primera radi, datoteka *anketa-8maj2006.spo* sadrži rezultate analize podataka iz datoteke *anketa3ED.sav*, obavljene 8. maja 2006. Imam podsetnik u koji beležim imena svih svojih datoteka i pojedinosti svih obavljenih analiza, uključujući i datume. Podsetnik mi mnogo olakšava pronalaženje rezultata određene analize. Kada počnete sopstvena istraživanja, videćete da se vrlo brzo nakupi mnogo raznih datoteka s rezultatima različitih analiza. Da biste izbegli zabunu i razočaranje, organizujte se i vodite precizne beleške o obavljenim analizama i o tome gde ste snimili njihove rezultate.

Štampanje rezultata

U levom oknu prozora **Viewer** izaberite sve delove rezultata koje hoćete da štampate. To se radi tako što pritisnete prvi deo koji hoćete da štampate, zatim pritisnete i zadržite taster **Ctrl** na tastaturi, i potom proizvoljnim redosledom pritisnete sve ostale delove rezultata koje hoćete da štampate. Kada

ste izabrali sve što hoćete da štampate, pustite taster Ctrl, pritisnite meni **File** (na vrhu prozora) i izaberite u njemu stavku **Print**. SPSS će vas ipak uvek pitati želite li da štampate samo odabrani deo rezultata ili sve.

Prozor Pivot Table Editor

Tabele rezultata koje se prikazuju u prozoru **Viewer**, SPSS naziva izvedenim (engl. *pivot*) tabelama, zato što ih možete menjati i modifikovati kako god vam odgovara. Da biste modifikovali tabelu, dvaput je pritisnite, pa će je SPSS prikazati u prozoru **Pivot Table Editor**. U njemu se može promeniti izgled tabele, njena veličina, fontovi, širina kolona, a redovi i kolone tabele mogu zameniti mesta, što se naziva transponovanje.

Prozor Chart Editor

Kada od SPSS-a zatražite da napravi histogram, trakasti dijagram (engl. *bar graph*) ili dijagram rasturanja (engl. *scatterplot*), on će ih najpre prikazati u prozoru **Viewer**. Ukoliko treba da promenite vrstu grafičkog prikaza ili način na koji su podaci u njemu predstavljeni, dvaput ga pritisnite, pa će ga SPSS otvoriti u prozoru **Chart Editor**. U njemu se može promeniti izgled i format dijagrama, fontovi, boje, tekstura pozadine i markeri (slika 3.3). Postupak pravljenja dijagrama i upotrebe **Chart Editor**a detaljnije je objašnjen u poglavlju 7.



Slika 3.3 Prozor Chart Editor.

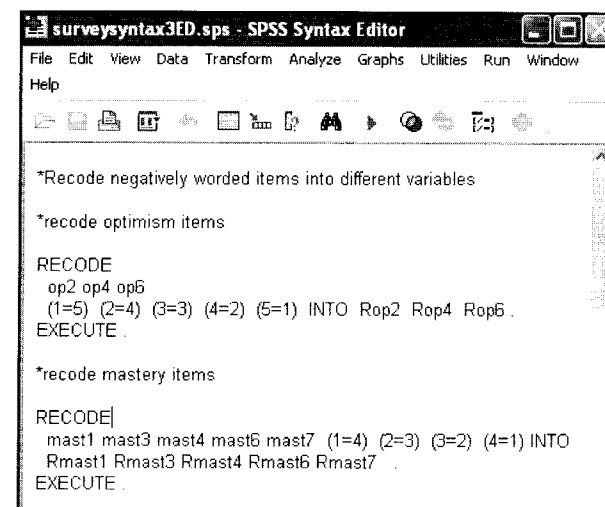
Prozor Syntax Editor

Ranije su se u SPSS-u komande zadavale posebnim komandnim jezikom (sintaksom). SPSS i dalje pravi te skupove komandi za izvršavanje svakog programa, ali se one obično ne vide, već vidite samo SPSS-ove menije koji te komande „ispisuju“ umesto vas kada izaberete neku stavku iz menija. Iako su mogućnosti dostupne u SPSS-ovim menijima dovoljne većini dodiplomaca, u nekim situacijama je korisno zaviriti iza scene i preuzeti kontrolu nad analizama koje hoćete da obavite. To se radi u prozoru **Syntax Editor** (slika 3.4).

Syntax Editor beleži sve komande koje ste upotreobili. To naročito dobro dođe kada morate da izmenite šifre većeg broja starih promenljivih ili izračunate vrednosti novih (što će biti pokazano u poglavlju 8.) **Syntax Editor** je

posebno koristan kada treba da ponovite mnogo analiza ili (pomoću SPSS-a) nacrtate više sličnih dijagrama. Upotrebite uobičajene SPSS-ove menije da zadate osnovne komande određene statističke tehnike i zatim ih prenesite u **Syntax Editor** (slika 3.4). U njemu se komande mogu kopirati i prenositi; mogu se i modifikovati komande koje je generisao SPSS. U njemu se mogu napisati veoma složene komande za sofisticirano rešifrovanje (engl. *recoding*) i obradu podataka. Više informacija o komandama potražite u SPSS-ovom meniju **Help**, stavka **Command Syntax Reference**. Sve izdate komande čuvaju se u zasebnoj tekstualnoj datoteci sa oznakom tipa **.sps**.

Komande prenete u **Syntax Editor** ne izvršavaju se automatski nego tek kada vi to zadate tako što izaberete komandu (zajedno s tačkom na njenom kraju) i pritisnete meni **Run** ili dugme sa strelicom na paleti sa alatcima. U datoteku komandi slobodno dodajte komentare; komentar je svaki red koji počinje zvezdicom (slika 3.4).



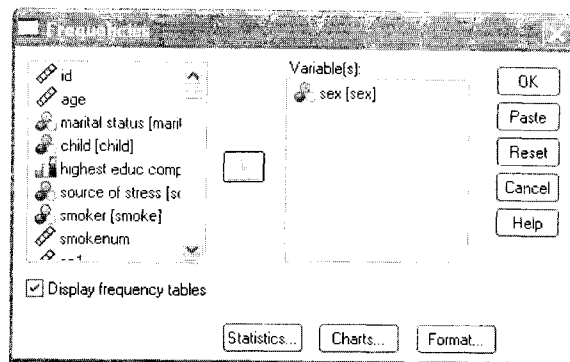
Slika 3.4 Prozor Syntax Editor.

Meniji

Svaki od prethodno opisanih prozora sadrži iznenađujuće veliki niz stavki u menijima. One se prikazuju pomoću ikonica (sličica) na paleti sa alatcima, i u padajućim menijima duž gornje ivice prozora. Ne dajte se uplašiti; prvo ćemo učiti samo osnovne alate, a sa ostalima ćete eksperimentisati kada steknete malo više samopouzdanja i iskustva.

Okviri za dijalog

Nakon što izabere stavku menija, korisnik obično mora da da još informacija. To se radi u okvirima za dijalog. Na primer, kada od SPSS-a zatražite da pokrene analizu **Frequencies**, prikazaće se okvir za dijalog u kome treba da odaberete promenljive čije učestalosti (frekvencije) hoćete da dobijete (slika 3.5).



Slika 3.5 Okvir za dijalog Frequencies.

Odavde ćete otvoriti još nekoliko okvira za dijalog, u kojima se zadaju statističke analize, dijagrami koje treba nacrtati i format u kojem treba predstaviti rezultate. Na raspolaganju su različite opcije, u zavisnosti od postupka ili analize koju treba obaviti, ali su osnovna načela upotrebe okvira za dijalog ista i sada ćemo ih razmotriti.

Izbor promenljivih u okviru za dijalog

Da biste naznačili koje promenljive hoćete da obradite, izaberite ih tako što ćete ih pritisnuti mišem u levom oknu, pa potom pritisnuti dugme sa strelicom (trougлом usmerenim udesno); izabrane promenljive će preći desno, u prazno okno **Variable(s)**. Promenljive se mogu birati jedna po jedna (što znači da svaki put treba pritisnuti strelicu) ili grupno. Ako su promenljive koje želite da izaberete susedne u listi, pritisnite samo prvu, zatim pritisnite i držite taster Shift, i potom držite taster sa strelicom nadole pritisnut sve dok ne izaberete sve potrebne promenljive. Pritisnite dugme sa strelicom (trougлом), i sve izabrane promenljive preći će u polje **Variable(s)**.

Ukoliko promenljive koje hoćete da izaberete nisu susedne u levom oknu, pritisnite prvu promenljivu koju želite, pa pritisnite i držite taster Ctrl, pomerite kursor naniže do sledeće promenljive koju želite i nju pritisnite, i tako nastavite (držeći pritisnut taster Ctrl) dok ne izaberete sve promenljive koje želite da obradite. Kada su sve potrebne promenljive izabrane, pritisnite

dugme sa strelicom (trougлом). Promenljive će preći u desno okno. Ako pogrešite ili hoćete da uklonite promenljivu iz spiska **Variable(s)**, samo obrnite postupak; promenljivu koju hoćete da uklonite izaberite u listi **Variable(s)**, pa pritisnite dugme sa strelicom (trougлом). Ona će se vratiti u prvobitnu listu levo. Smer trougla na dugmetu odgovara smeru prebacivanja promenljivih, tj. pokazuje da li se promenljive prebacuju u listu **Variable(s)** ili iz nje vraćaju u levo okno.

Dugmad u okviru za dijalog

Većina okvira za dijalog sadrži standardnu dugmad (**OK**, **Paste**, **Reset**, **Cancel**, **Help**; slika 3.5). Evo opisa namene tih dugmadi:

- **OK**: Pritisnite ovo dugme kada ste odabrali promenljive i spremni ste da pokrenete neku analizu ili postupak.
- **Paste**: Ovim dugmetom se komande koje je generisao SPSS prenose u **Syntax Editor**. To je podesno kada hoćete da zabeležite izdate komande ili više puta ponovite određenu analizu.
- **Reset**: Ovim dugmetom brišete iz okvira za dijalog sve komande izdate tokom prethodnog korišćenja date analize ili postupka. Time dobijate mogućnost da novu analizu obavite bez ikakvog nasleđenog stanja, od nule, s drugačijim promenljivama.
- **Cancel**: Pritiskom na ovo dugme zatvarate okvir za dijalog i otkazujete sve komande izdate u vezi s datom analizom ili postupkom.
- **Help**: Pritisnite ovo dugme da biste dobili informacije o datoj analizi ili postupku koji nameravate da sprovedete.

Mada sam kao primer okvira za dijalog na slici 3.5 prikazala okvir **Frequencies**, svi SPSS-ovi okviri za dijalog funkcionišu po istom osnovnom principu. Svaki okvir za dijalog ima niz dugmadi s različitim opcijama koje se odnose na konkretan postupak ili analizu. Pomoću te dugmadi otvaraju se podokviri u kojima korisnik zadaje analize koje treba obaviti ili statističke podatke koje treba prikazati.

Kako zatvoriti SPSS

Kada završite rad u SPSS-u i hoćete da zatvorite program, pritisnite meni **File** u gornjem levom uglu prozora. U padajućem meniju koji će se otvoriti pritisnite stavku **Exit**. SPSS će vas pozvati da pre zatvaranja snimite datoteku s podacima i datoteku s rezultatima. Svako od tih datoteka SPSS daje drugačiju oznaku tipa, tako da se odmah vidi o kojoj vrsti informacija je reč. Datoteka s podacima dobija oznaku tipa **.sav**, datoteka s rezultatima **.spo**, a datoteka s komandama **.sps**.

Kako dobiti pomoć

U slučaju da vam zatreba pomoć ili da ne znate na šta se odnosi neka od ponuđenih opcija, pomozite se menijem **Help**. Kada u glavnom meniju pritisnete **Help**, pojaviće se više opcija: mogli biste potražiti određenu temu, pročitati odgovarajuće uputstvo (**Tutorial**) ili pozvati u pomoć instruktora (**Statistics Coach**). Ovo poslednje je zanimljiva novost u SPSS-u, zato što istraživačima i studentima statistike koji se ne snalaze može poslužiti kao odličan vodič. Provešće vas korak po korak kroz postupak izbora statističke analize koja odgovara vašim potrebama. To ne može zameniti udžbenik iz statistike, ali može da pomogne.

U svakom od glavnih okvira za dijalog postoji i dodatni meni **Help**, sa informacijama o izabranom postupku. Potražite u njemu objašnjenje raznih opcija koje se nude u pomoćnim okvirima za dijalog. Dovedite kursor na nepoznatu opciju i pritisnite je desnim tasterom miša. Prikazaće se mali okvir s kratkim objašnjenjem te opcije.

DEO II

Priprema datoteke s podacima

Priprema datoteke s podacima za analizu obuhvata više koraka. Prvo se napravi prazna datoteka i u nju unesu podaci dobijeni u istraživanju, u obliku definisanom u šifarniku (objašnjeno u poglavlju 2). Zatim u datoteci s podacima treba pronaći eventualne greške i ukloniti ih. U drugom delu knjige razmatraju se upravo ta dva koraka. U poglavlju 4 obrađeni su postupci za pravljenje datoteke i unošenje podataka u nju. U poglavlju 5 objašnjen je postupak uklanjanja grešaka iz datoteke s podacima.



Pravljenje datoteke za podatke i unošenje podataka

Postupak izrade datoteke s podacima i njihove analize sastoji se od više koraka. Dijagram toka prikazan na sledećoj stranici sadrži osnovne potrebne korake. U ovom poglavlju provešću vas kroz postupak pravljenja datoteke i unošenja podataka u nju pomoću SPSS-a.

Objasnićemo tri ključna koraka u pripremi datoteke s podacima:

- *Korak 1.* Prvi korak je pregled i, po potrebi, izmena opcija (u prethodnim verzijama SPSS-a zvale su se preference) koje SPSS koristi za prikazivanje podataka i dobijenih rezultata.
- *Korak 2.* Sledeći korak je izgradnja strukture datoteke s podacima definisanjem promenljivih.
- *Korak 3.* Poslednji korak je unošenje podataka, tj. vrednosti koju je svakoj promenljivoj dodelio svaki učesnik ili ispitanik.

Navedene postupke ćemo ilustrovati na primeru datoteke `survey3ED.sav`, opisanoj u dodatku. U dodatku je i šifarnik pomoću kojeg su dobijeni numerički podaci za analizu rezultata.

Datoteke s podacima mogu se uvesti i iz drugih programa za tabelarne proračune (npr. iz Excela). To može umnogome da olakša postupak unošenja podataka, naročito studentima koji na svojim računarima nemaju SPSS. U Excelu korisnik može da napravi praznu datoteku i da u nju unese podatke na svom kućnom računaru. Kada to završi, datoteku uveze u SPSS i obradu i analizu podataka nastavi u njemu. Pri kraju poglavlja biće navedena uputstva za unošenje podataka u Excelu.

Dijagram toka procesa analize podataka Priprema šifarnika (poglavlje 2)

Pravljenje strukturirane datoteke za podatke (poglavlje 4)

Uklanjanje grešaka iz datoteke s podacima (poglavlje 5)

Istraživanje podataka pomoću opisnih statističkih
tehnika i dijagrama (poglavlja 6 i 7)

Modifikacija promenljivih radi daljih analiza (poglavlje 8)

Obavljanje statističkih analiza radi
istraživanja relacija (deo 4)

Sprovođenje statističkih analiza
radi poređenja grupa (deo 5)

Korelacija (poglavlje 11)
Delimična korelacija (poglavlje 12)
Višestruka regresija (poglavlje 13)
Logistička regresija (poglavlje 14)
Faktorska analiza (poglavlje 15)

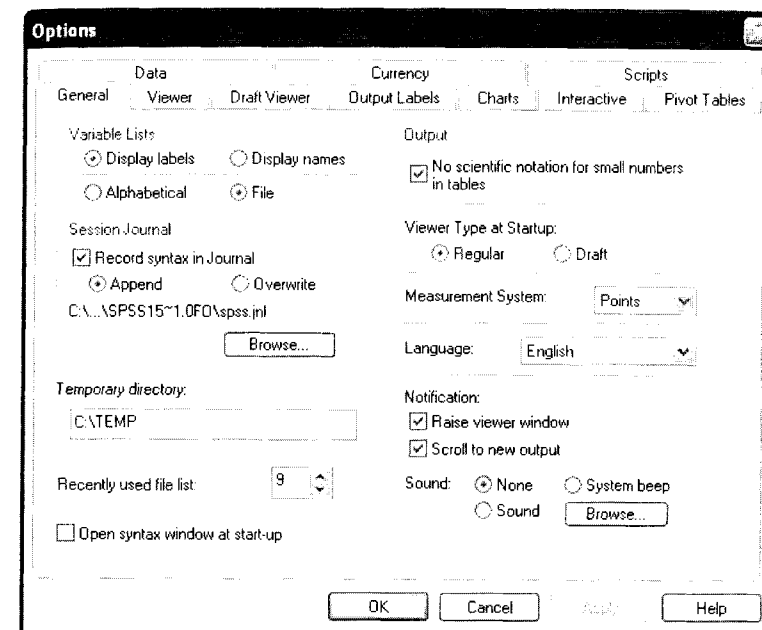
Neparametarske tehnike (poglavlje 16)
T-testovi (poglavlje 17)
Analiza varijanse (poglavlja 18, 19, 20)
Analiza varijanse više od dve promenljive
(poglavlje 21)
Analiza kovarijanse (poglavlje 22)

Izmena opcija u SPSS-u

Pre nego što napravite datoteku s podacima, bilo bi dobro da proučite opcije (postavke) SPSS-a od kojih zavisi način prikazivanja podataka i rezultata. To su opcije za definisanje načina prikazivanja promenljivih, veličine dijagrama, vrste tabela u kojima će rezultati biti prikazani i mnogih drugih aspekata programa. Možda će vam početka nešto od ovoga biti nerazumljivo, ali čim počnete da unosite i analizirate podatke u SPSS-u, možete se vratiti na ovaj odeljak.

Ukoliko računar na kome je instaliran SPSS delite s drugim korisnicima (npr. u računarskoj laboratoriji), morate znati ove postavke zato što ih drugi studenti mogu promeniti, a to će uveliko izmeniti izgled programa. Trebalo bi da znate kako da vratite opcije u staro, poželjno stanje.

Da biste otvorili prozor Options, pritisnite **Edit** u glavnom meniju na vrhu ekrana, pa izaberite stavku **Options**. Otvoriće se prozor prikazan na slici 4.1. Tu je navedeno mnogo opcija; većinu nećete morati da menjate. U nastavku ću opisati samo ključne opcije, podeljene po karticama na kojima se nalaze. (S kartice na karticu prelazite pritiskom na jezičak one na koju hoćete da pređete.) Nemojte pritisnuti **OK** dok na svim karticama ne izmenite sve opcije koje nameravate da izmenite.



Slika 4.1 Prozor Options.

Kartica General

Kada dođe vreme za obavljanje analiza, promenljive mogu biti poredane abecedno ili redom kojim se pojavljuju u datoteci. Ja uvek koristim redosled pojavljivanja u datoteci, zato što je on jednak redosledu stavki u upitniku i u šifarniku. Da bi se promenljive prikazivale po redosledu pojavljivanja u datoteci, pritisnite kružić pored opcije **File** u odeljku **Variable Lists** (gore levo). U pododeljku **Notification** odeljka **Output** (dole desno) treba da su potvrđena polja **Raise viewer window** i **Scroll to new output**. Tako će se nakon svake analize automatski otvoriti prozor **Viewer** i prikazati novi rezultati.

U odeljku **Output** (gore desno), treba da izaberete polje **No scientific notation for small numbers in tables**. Tako ćete sprečiti pojavljivanje nekih veoma čudnih brojeva među rezultatima vaših statističkih analiza.

U odeljku **Session Journal** treba da je potvrđena opcija **Append**. To obezbeđuje da se u zapisničkoj datoteci (spss.jnl) beleže svi postupci koje obavljate u SPSS-u tokom svake sesije. Kada pritisnete dugme **Browse**, možete da izaberete direktorijum u koji će ta zapisnička datoteka (engl. *journal*) biti smeštena.

Kartica Data

Na kartici **Data** su opcije koje određuju način prikazivanja datoteke s podacima. Obavezno potvrdite polje **Calculate values immediately**. Time postižete da SPSS odmah izračuna i prikaže sve rezultate.

Ako promenljive s kojima radite nisu decimalne vrednosti, mogli biste promeniti format prikazivanja svih promenljivih. U odeljku **Display format for new numeric variables**, za broj decimalnih mesta zadajte 0. To znači da se decimalna mesta neće prikazivati ni za jednu novu promenljivu. Time smanjujete datoteku s podacima i pojednostavljujete njen izgled.

Kartica Output Labels

Opcije u ovom odeljku služe za prilagođavanje načina prikazivanja običnih imena i dužih opisnih imena promenljivih u rezultatima. Na samom dnu, u odeljku **Variable values in labels are shown as**, u padajućoj listi izaberite **Values and Labels**. Tako ćete u tabelama generisanim u SPSS-ovom prozoru **Viewer** moći da vidite i numeričke vrednosti i duža, opisna imena promenljivih.

Kartica Charts

Na ovoj kartici su opcije koje određuju način prikazivanja dijagrama. Mogli biste promeniti opciju **Chart Aspect Ratio** tj. podrazumevani odnos širine i visine dijagrama. Pored toga, tu se zadaju fontovi, boje i druga svojstva dijagrama.

Kartica Pivot Tables

Većinu rezultata statističkih analiza SPSS prikazuje u izvedenim tabelama. Na kartici **Pivot Tables** može se u opsežnim listama izabrati format tih tabela. Stil koji vam najviše odgovara naći ćete eksperimentisanjem. Kada tek počnem s nekom analizom, upotrebljavam stil *smallfont.tlo*. U ovom stilu slova su sitna, pa se štedi prostor na ekranu (i papir prilikom štampanja). Međutim, taj stil nije prikladan za uvoz u dokumente koji moraju biti usaglašeni sa APA stilom (obaveznim za publikacije iz psihologije), zato što sadrži vertikalne linije. Kada budete spremni da formatirate tabele u svom izveštaju o istraživanju, moći ćete da izaberete i neki od APA stilova (bilo koji od formata *academic.tlo*).

Stilove tabela možete menjati koliko god puta hoćete, samo ne zaboravite da stil treba promeniti *pre* analize. Kada se tabele već pojave u rezultatima, više im se ne može menjati stil, ali se mnoga njihova svojstva (recimo, veličina fontova, širina kolona) ipak mogu menjati u prozoru **Pivot Table Editor**. Otvorićete ga kada dvaput pritisnete tabelu koju treba izmeniti.

Nakon što izmenite sve što želite na raznim karticama iz okvira za dijalog **Options**, pritisnite **OK**. Zatim možete preći na definisanje promenljivih i unošenje podataka.

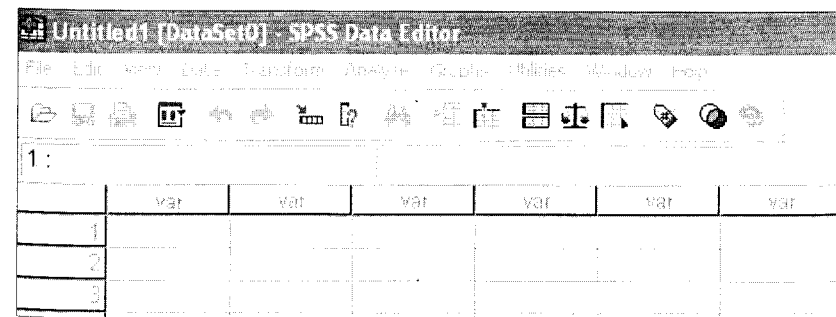
Definisanje promenljivih

Pre nego što počnete da unosite podatke, SPSS-u morate saopštiti kako vam se promenljive zovu i dati mu uputstva za šifrovanje (kodiranje), tj. pretvaranje vrednosti promenljivih u brojeve. To se naziva definisanje promenljivih, a obavlja se u prozoru **Data Editor** (slika 4.2). Od verzije 10 SPSS-a, prozor **Data Editor** se sastoji od dva različita prikaza: **Data View** i **Variable View**. Iz jednog u drugi prelazite tako što pritiskate jezičke njihovih kartica na dnu ekrana, s leve strane.

Svaka kolona u prozoru **Data View** odgovara po jednoj promenljivoj (engl. *variable*) i nosi ime *var*. Ta imena će biti zamenjena imenima promenljivih koje ste naveli u svom šifarniku (slika 4.2). Duž leve strane su brojevi 1, 2, 3 itd. To su redni brojevi analiziranih slučajeva, koje SPSS dodeljuje svakom redu s podacima. To *nisu* vaši brojevi za identifikaciju slučajeva; osim toga, brojevi redova se menjaju kada sortirate datoteku ili je podelite da biste podatke analizirali u podskupovima.

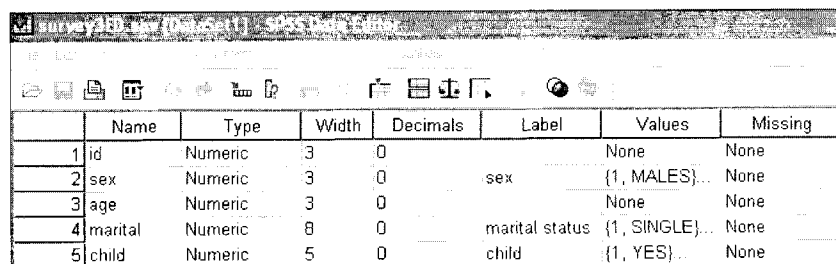
Postupak definisanja promenljivih

Da biste definisali sve promenljive od kojih se vaši podaci sastoje, najpre pritisnite jezičak kartice **Variable View** na dnu ekrana. U tom prikazu (slika 4.3), duž leve strane su navedeni redni brojevi promenljivih, a duž vrha njihovi atributi (ime, tip, širina, broj decimala, duže opisno ime itd.).



Slika 4.2 Prozor Data Editor.

Vaš je zadatak da sve promenljive definišete tako što ćete navesti potrebne informacije za svaku od njih u šifarniku. Deo tih informacija morate sami da obezbedite (npr. ime); ostale automatski dodeljuje SPSS, to su podrazumevane vrednosti. Naravno, te vrednosti se mogu menjati po potrebi. U nastavku ćemo navesti osnovne neophodne informacije. Upotrebila sam zaglavlja koja odgovaraju zaglavljinama kolona prikazanim u prozoru **Variable View**. Sledi opis jednostavnih koraka od kojih se postupak sastoji; međutim, postoje i brojne prečice koje ćete moći da koristite kada upoznate postupak. Njih navodim kasnije, u odeljku „Opcione prečice“. Najpre treba upoznati osnovne tehnike.



	Name	Type	Width	Decimals	Label	Values	Missing
1	id	Numeric	3	0		None	None
2	sex	Numeric	3	0	sex	{1, MALES}...	None
3	age	Numeric	3	0		None	None
4	marital	Numeric	8	0	marital status	{1, SINGLE}...	None
5	child	Numeric	5	0	child	{1, YES}...	None

Slika 4.3 Variable View.

Name

U ovu kolonu upišite kratko ime svake promenljive navedene u šifarniku. Pomoću imena će svaka promenljiva u datoteci s podacima biti identifikovana. Imena treba da su što kraća i da ne premašuju 64 slova znaka (od 12. verzije SPSS-a nadalje), odnosno osam slovnih znakova (ranije verzije SPSS-a). Imena moraju zadovoljiti SPSS-ovu konvenciju o imenovanju (objašnjenu u poglavlju 2). Imena moraju biti jedinstvena, početi slovom, i ne smeju sadržati razmake (beline) niti specijalne znakove. Primere dobrih imena za promenljive pročitajte u šifarniku u dodatku. Tamo su navedena imena promenljivih koje su korišćene u datotekama s podacima napravljenim za ovu knjigu. (Pojednosti o tim datotekama videti na str. 12.)

Type

Kada upišete prvo ime promenljive, u koloni **Type** će se automatski pojaviti podrazumevana vrednost tipa promenljive – **Numeric**. To će vam odgovarati za numeričke promenljive, ali za tekstualne informacije (recimo, prezime) tip treba promeniti u **String**. Tip promenljive menjate tako što pritisnete desnu stranu ćelije; trebalo bi da se pojavi okvir s tri tačke i da vam se prikažu dostupne opcije. U istom prozoru može se podešavati širina promenljive i broj njenih decimalnih mesta.

Width

Podrazumevana vrednost širine je 8. To je dovoljno za većinu numeričkih podataka. Ukoliko vaša promenljiva može imati veoma velike vrednosti ili je u pitanju tekstualna promenljiva tipa *string*, promenite podrazumevanu vrednost. U protivnom, ostavite je na 8.

Decimals

Podrazumevana vrednost za broj decimalnih mesta (koju sam zadala na kartici **Options** opisanoj u prethodnom delu poglavlja) jeste 0. Ako vaša promenljiva ima decimale, promenite ovu postavku kako god vam odgovara.

Label

Kolona **Label** (duže opisno ime) služi da promenljivoj date duži opis od onog što staje u osam slovnih znakova, koliko je iznosila maksimalna dužina imena promenljive do 12. verzije SPSS-a. Duža, opisna imena promenljivih navode se u rezultatima svih SPSS-ovih analiza. Primera radi, promenljivoj kratkog imena UMAST mogli biste dodeliti duže opisno ime Ukupno mastera.

Values

U koloni **Values** definišete značenje brojeva kojima ste šifrovali tekstualni sadržaj promenljivih, tj. zamenili ga. Pokazaću taj postupak za promenljivu pol (sex) u drugom redu tabele na prethodnoj slici.

1. Pritisnite tri tačke na desnoj strani ćelije (na preseku kolone **Values** i drugog reda tabele). Otvoriće se okvir za dijalog **Value Label**.
2. Pritisnite polje **Value** i u njega upišite 1.
3. Pritisnite polje **Value Label** i u njega upišite „muški“.
4. Pritisnite dugme **Add**. U zbirnom polju pisaće 1=muški.
5. Ponovite za ženski pol: za **Value** upišite 2, za **Value Label** upišite „ženski“. Pritisnite dugme **Add**.
6. Kada definišete sve brojeve navedene u šifarniku, pritisnite dugme **Continue**.

Missing

Ponekad istraživači koriste određene brojeve kako bi pokazali da podatak nedostaje, tj. ne postoji. To ne morate da radite, pošto SPSS svaku praznu ćeliju tumači kao nepostojanje podatka. Dakle, ukoliko nameravate da ostavite prazno mesto kada neki delić informacija nije dostupan, u kolonu **Missing** prikaza **Variable View** ne treba ništa da upisujete.

S druge strane, ako nepostojanje vrednosti nameravate da kodirate određenim brojem (npr. 99=nema vrednost), morate ga upisati u odeljak **Missing Values**, inače će ga SPSS u svim statističkim analizama smatrati za legitimnu

vrednost. U tom slučaju, izaberite stavku **Discrete missing values** i upišite taj broj (npr. 99) u predviđeno polje. Mogu se zadati do tri broja koja ukazuju na to da vrednosti ne postoje. Ukoliko upotrebljavate takve posebne brojeve (šifre), bilo bi dobro i da im dodelite odgovarajuća duža opisna imena u koloni **Values**.

Columns

Podrazumevana širina kolona obično je 8 i to je dovoljno za većinu namena. Izmenite je samo ukoliko očekujete ogromne vrednosti ili dugačka imena promenljivih.

Align

Vrednosti u kolonama obično su poravnate ulevo. Nema potrebe da to menjate.

Measure

Zaglavlje kolone **Measure** odnosi se na način merenja svake promenljive pojedinačno. Podrazumevana vrednost je **Scale**, što znači da se koristi intervalna merna skala, pogodna za neprekidne (kontinualne) podatke. Kada se promenljiva sastoji od kategorija (npr. pol, bračno stanje...), pritisnite tu ćeliju i zatim dugme sa strelicom koje će se pojaviti. Za kategorijske podatke izaberite nominalnu skalu (**Nominal**), a za one za koje je važan poredak mogućih vrednosti, izaberite ordinalnu skalu (**Ordinal**).

Opcione prečice

Kada datoteka s podacima ima mnogo promenljivih, opisani postupak ume da bude zametan. Ima više prečica pomoću kojih ga možete ubrzati. Ukoliko vise promenljivih ima iste atribute (npr. tip, širina kolone, broj decimala), definišite prvu promenljivu a zatim njene atribute kopirajte na ostale promenljive.

Kopiranje atributa iz definicije jedne promenljive na drugu promenljivu

1. U prikazu **Variable View** pritisnite ćeliju sa atributom koji treba kopirati (npr. širinu, Width).
2. Otvorite meni **Edit** i u njemu izaberite **Copy**.
3. Pritisnite ćeliju istog atributa koja pripada drugoj promenljivoj (onoj na koju treba primeniti taj atribut).
4. Otvorite meni **Edit** i u njemu izaberite **Paste**.

Kopiranje atributa iz definicije jedne promenljive na više drugih promenljivih

1. U prikazu **Variable View** pritisnite ćeliju sa atributom koji treba kopirati (npr. širinu, Width).
2. Otvorite meni **Edit** i u njemu izaberite **Copy**.
3. Pritisnite ćeliju istog atributa koja pripada prvoj promenljivoj na koju ga treba primeniti, zadržite levi taster miša pritisnut i prevucite kursor niz kolonu birajući tako sve promenljive na koje taj atribut treba primeniti.
4. Otvorite meni **Edit** i u njemu izaberite **Paste**.

Zadavanje istih atributa nizu novih promenljivih

Kada podaci obuhvataju skale koje se sastoje od više pojedinačnih stavki, od jednom napravite sve te nove promenljive i definišite njihove atribute. Kao primer, sledi opis tog postupka za šest stavki (op1 do op6) skale optimizma:

1. U prikazu **Variable View** definišite atribute prve promenljive (op1) prema prethodnim uputstvima. Trebalo bi da zadate i duža opisna imena koja odgovaraju mogućim iznosima te promenljive: 1=nimalo se ne slažem, 2=ne slažem se, 3=neutralno, 4=slažem se, 5=potpuno se slažem.
2. Dok je izabran **Variable View**, pritisnite broj reda u kome se promenljiva nalazi (trebalo bi da se time izabere ceo taj red).
3. Otvorite meni **Edit** i u njemu izaberite **Copy**.
4. Pritisnite sledeći prazan red.
5. Otvorite meni **Edit** i u njemu izaberite **Paste Variables**.
6. U okviru za dijalog koje će se pojaviti, upišite broj dodatnih promenljivih koje treba napraviti (u ovom slučaju 5), njihov prefiks (op) i redni broj kojim nove promenljive treba da počnu (u ovom slučaju, 2). Pritisnite dugme **OK**.

Dobićete pet novih promenljivih (op2, op3, op4, op5 i op6).

Vrednosti na ostalim skalama dodajte prema opisanom postupku (recimo, stavke od spost1 do spost10 skale samopoštovanja, engl. *self-esteem*). Ne zaboravite da je navedeni postupak prikladan samo kada promenljive imaju iste atribute; on ne može da se koristi kada stavke imaju različite skale odgovora, npr. kada su jedne nominalne (nazivne, kategorijske), a druge se mere u intervalima, ili kada se njihove vrednosti različito šifruju.

Unošenje podataka

Nakon što zadate imena svih promenljivih i, gde treba, duža opisna imena koja odgovaraju njihovim mogućim vrednostima, sve je spremno za unošenje podataka. Prethodno pripremite i šifarnik.

Važno Prilikom unošenja podataka ne zaboravite da datoteku redovno snimate. SPSS ništa ne snima automatski. Ukoliko to sami ne uradite, rizikujete da izgubite sve što ste uneli. Snimate tako što na paleti sa alatka pritisnete sličicu diskete ili otvorite meni **File** i u njemu izaberete stavku **Save**.

Postupak unošenja podataka

1. Podaci se unose u prikazu **Data View**. Pritisnite jezičak kartice **Data View** na dnu prozora **Data Editor**, levo. Pojaviće se proračunska tabela sa imenima novodefinisanih promenljivih duž vrha.
2. Pritisnite prvu ćeliju skupa podataka (u prvom redu prve kolone). Oko aktivne ćelije prikazaće se tamna ivica.
3. Upišite broj (ako se radi o promenljivoj ID, to bi trebalo da bude 1).
4. Pritisnite taster sa strelicom nadesno; kursor će preći u sledeću ćeliju desno, u koju treba da upišete drugi podatak za analizirani slučaj broj 1.
5. U taj red unesite sve informacije za slučaj broj 1. Pažljivo upisujte da bi svaki podatak bio u odgovarajućoj koloni.
6. Kada pritisnete taster Home, kursor se vraća na početak reda. Prelazak u drugi red ostvarićete kad pritisnete taster sa strelicom nadole. U taj red unesite podatke za slučaj 2.
7. *Kada pogrešite i poželite da ispravite grešku*, pritisnite ćeliju s greškom. Sadržaj ćelije će se prikazati iznad tabele. Tu upišite ispravnu vrednost, pa pritisnite taster sa strelicom nadesno.

Kada definišete promenljive i unesete podatke, vaš prozor **Data Editor** trebalo bi da liči na onaj sa slike 3.1.

Modifikovanje datoteke s podacima

Kada napravljenu datoteku s podacima poželite da izmenite, npr. da dodajete, brišete ili premeštate promenljive, dodajete ili uklanjate analizirane slučajeve (redove tabele), otvorite prozor **Data Editor** u prikazu **Data View**.

Uklanjanje jednog slučaja

Pronađite slučaj (red) koji treba ukloniti. Dovedite kursor na osenčeni deo s leve strane u kome je prikazan broj reda (slučaja), pa pritisnite jednom da biste izabrali ceo red. Pritisnite taster Delete na tastaturi ili otvorite meni **Edit** i u njemu izaberite **Clear**.

Umetanje jednog slučaja između postojećih

Dovedite kursor na bilo koju ćeliju u redu neposredno ispod mesta gde hoćete da umetnete nov slučaj (red). Otvorite meni **Data** i u njemu izaberite stavku **Insert Case**. Pojaviće se prazan red u koji se mogu uneti podaci za nov slučaj koji analizirate.

Uklanjanje promenljive

Dovedite kursor na osenčeni deo kolone, sa imenom promenljive koju treba ukloniti. Izaberite celu kolonu tako što ćete jednom pritisnuti to ime. Pritisnite taster Delete na tastaturi ili otvorite meni **Edit** i u njemu izaberite **Clear**.

Umetanje nove promenljive između postojećih

Dovedite kursor na bilo koju ćeliju u koloni (promenljivoj) desno od one gde biste hteli da umetnete novu promenljivu. Otvorite meni **Data** i u njemu izaberite stavku **Insert Variable**. Pojaviće se prazna kolona u koju se mogu uneti podaci za novu promenljivu.

Premeštanje postojećih promenljivih

Otvorite prozor **Data Editor** u prikazu **Variable View**. Izaberite promenljivu koju treba premestiti tako što ćete pritisnuti levu marginu (levo od broja tog reda). Pritisnite i zadržite levi taster miša, i prevucite promenljivu na novo mesto. (Dok je budete prevlačili, prikazivaće se crvena linija koja pokazuje gde ste trenutno.) Pustite levi taster miša kada stignete do željenog mesta.

Unošenje podataka u Excelu

Datoteka s podacima se može pripremiti u Microsoftovom Excelu i zatim uvesti u SPSS radi analize. To je odlično za studente koji SPSS nemaju kod kuće. Excel je obično deo Microsoftovog paketa Office. (Proverite da li ga imate tako što ćete na Windowsovoj radnoj površini pritisnuti dugme **Start**, otvoriti istoimeni meni i pritisnuti u njemu stavku **Programs**; prikazaće se spisak svih programa koji se nalaze u računaru). Sledi opis postupka pravljenja nove datoteke s podacima u Excelu i njenog uvoza u SPSS. Ukoliko nameravate da iskoristite ovu mogućnost, trebalo bi naučiti osnove rada u Excelu, pošto se to ovde neće objašnjavati.

Upozorenje: Excel može da radi s najviše 256 kolona s podacima (tj. promenljivih). Ako bi vaša datoteka s podacima trebalo da ima više promenljivih, verovatno će vam biti lakše da je napravite u SPSS-u nego da je naknadno konvertujete iz Excela. (Ako biste ipak morali da koristite Excel, to biste uradili ovako: na svaki Excelov radni list stavite do 256 promenljivih, ali tako da je ID prva promenljiva na svakom listu. Zatim svaki list zasebno prebacite u SPSS i tamo ih spojite (objedinite), prema uputstvima navedenim u sledećem odeljku.)

Korak 1: zadavanje imena promenljivih

U prvi red praznog Excelovog radnog lista upišite imena promenljivih. Ta imena treba da su usaglašena sa SPSS-ovim pravilima za imenovanje (navedenim u poglavlju 2).

Korak 2: unošenje podataka

1. Unesite podatke za prvi analizirani slučaj u neki red Excelove proračunske tabele, pazeći da oni odgovaraju kolonama u koje ih upisujete.
2. Ponovite za svaki od preostalih slučajeva. Nemojte koristiti formule niti druge Excelove funkcije. Ne zaboravite da redovno snimate datoteku tokom unošenja.
3. Pritisnite **File, Save**. U polju **Save as Type** treba da stoji **Microsoft Excel Workbook**. U odgovarajuće polje upišite ime pod kojim će datoteka biti snimljena.

Korak 3: pretvaranje u format SPSS

1. Kada završite unošenje podataka, snimite datoteku i zatvorite Excel.
2. Pokrenite SPSS i u prozoru **Data Editor**, u glavnom meniju izaberite redom **File, Open, Data**.
3. U odeljku **Files of Type** izaberite **Excel**. Excelove datoteke imaju oznaku tipa .xls. Pronađite vašu datoteku s podacima. Pritisnite je tako da se njeno ime pojavi u polju **File name**.
4. Pritisnite dugme **Open**. Pojaviće se prozor **Opening Excel Data Source**. U polju **Read variable names from the first row of data** treba da stoji kvačica. Pritisnite **OK**.

Duž vrha prozora prikazaće se imena promenljivih čiji su podaci uvezeni.

Korak 4: snimanje SPSS datoteke

1. U glavnom meniju pri vrhu prozora izaberite **File** a zatim stavku **Save As**.
2. Upišite neko prikladno ime datoteke. U polju **Save as Type** treba da stoji **SPSS (*.sav)**. Pritisnite dugme **Save**.
3. U prikazu **Variable View** prozora **Data Editor** sada treba da zadate sadržaj polja **Variable label**, **Value label**, **Values** i **Measure** (tj. duža opisna imena promenljivih i njihovih mogućih vrednosti, same te vrednosti i njihov tip; videti prethodno data uputstva).

Kada dođe vreme da analizirate te podatke u SPSS-u, treba da otvorite SPSS datoteku sa oznakom tipa .sav, a ne Excelovu datoteku sa oznakom tipa .xls.

Upozorenje Pre spajanja (objedinjavanja) datoteka, obavezno napravite njihovu rezervnu kopiju. Spojene datoteke čuvajte u drugom direktorijumu. Proverite da li je spajanje podataka uspešno, tj. da li su dobijeni podaci jednaki onima u prvobitnim datotekama.

Spajanje datoteka

Dešava se da zatreba spojiti SPSS datoteke s podacima. SPSS omogućava spajanje datoteka dodavanjem redova (tj. analiziranih slučajeva) na kraj postojeće datoteke, i dodavanjem promenljivih za svaki od postojećih slučajeva, npr. kada postanu dostupni neki novi podaci. Ova druga mogućnost je naročito podesna kada imate Excelove datoteke s podacima na više listova, koje treba spojiti na osnovu identifikatora (ID-a) slučajeva (redova).

Spajanje datoteka dodavanjem slučajeva

Ovo je postupak spajanja datoteka koje sadrže iste promenljive, ali različite analizirane slučajeve; na primer, kada se isti podaci prikupljaju na dva mesta (recimo, u dve bolnice) ili ih unose dve osobe. Obe datoteke moraju imati identična imena promenljivih, mada drugi podaci u datotekama mogu biti različiti.

Ukoliko su ID brojevi u datotekama koje treba spojiti isti (slučajevi počinju sa ID = 1, 2, 3 itd.), moraćete ih u jednoj ili drugoj datoteci izmeniti pre spajanja, tako da nakon spajanja svaki analizirani slučaj zadrži svoj jedinstven ID broj. Evo kako se to radi: otvorite jednu od datoteka, otvorite meni **Transform** i u njemu izaberite stavku **Compute Variable**. U polje **Target Variable** upišite ID, a u polje **Numeric Expression** ID + 1000 (odnosno ID + broj veći od postojećeg broja slučajeva u datoteci). Pritisnite dugme **OK** i zatim ponovo **OK** u okviru za dijalog koji će vas pitati želite li da promenite tu promenljivu. Time će novi ID brojevi u toj datoteci početi od 1001, 1002 itd. Zapišite to u svoj šifarnik da bi se znalo ubuduće. Sada je sve spremno za spajanje datoteka.

1. Otvorite prvu datoteku koju treba spojiti.
2. Otvorite meni **Data**, izaberite u njemu stavku **Merge files** i potom **Add Cases**.
3. U okviru za dijalog pritisnite **An external SPSS data file** i zatim izaberite datoteku s kojom treba spojiti onu otvorenu. (Ako je već otvorena, biće navedena u polju **An open dataset**.)
4. Pritisnite **Continue** i zatim **OK**. Novu datoteku snimite (**File, Save As**) pod nekim novim imenom.

Spajanje datoteka dodavanjem promenljivih

Ovo je postupak spajanja datoteka koje sadrže iste slučajeve, ali različite promenljive (pri čemu odgovarajući slučajevi moraju imati isti ID broj). Obe datoteke moraju početi promenljivom ID ili nekom drugom zajedničkom promenljivom, npr. prezimenom ili bolničkim brojem.

1. Sortirajte obe datoteke rastućim redosledom po polju ID. (Otvorite meni **Data**, u njemu izaberite stavku **Sort Cases** i zatim **ID**.)
2. U meniju **Data** izaberite stavku **Merge Files** i pritisnite **Add Variables**.
3. U okviru za dijalog pritisnite **An external SPSS data file** i zatim izaberite datoteku s kojom treba spojiti onu otvorenu. (Ako je već otvorena, biće navedena u polju **An open dataset**.)
4. U polju **Excluded variables** treba da je navedena promenljiva ID, zato što postoji u obe datoteke, pa je dovoljno preuzeti je iz jedne od njih, tj. ne treba je preuzimati dvaput. Ukoliko su tu navedena imena još nekih promenljivih, promenite im imena pritiskom na dugme **Rename** da bi sve promenljive u spojenoj datoteci imale jedinstvena imena.
5. Pritisnite promenljivu ID, potom polje **Match cases on key variables** i najzad dugme sa strelicom. Time će promenljiva ID preći u polje **Key Variables**. To znači da će biti objedinjeni podaci za analizirane slučajeve označene istim ID brojevima.
6. Snimite novu datoteku (**File, Save As**) pod nekim novim imenom.

Korisne mogućnosti SPSS-a

SPSS ima mnogo korisnih mogućnosti koje su podesne za analize i štede vreme i trud. U narednim odeljcima navešću nekoliko glavnih.

Sortiranje datoteke s podacima

SPSS može da sortira podatke po vrednostima određene promenljive (npr. polu, starosti).

1. Otvorite meni **Data**, izaberite u njemu stavku **Sort Cases** i zadajte promenljivu po kojoj treba sortirati. Potom zadajte da li će se podaci sortirati po rastućem (**Ascending**) ili opadajućem (**Descending**) redosledu. Pritisnite **OK**.
2. Datoteku vraćate u prvobitni redosled tako što u ponovnom postupku zadate sortiranje po ID-u.

Podela datoteke s podacima na grupe

Ponekad datoteku treba podeliti i zasebno analizirati tako dobijene grupe (recimo, muškarce i žene). Time datoteku nećete trajno izmeniti; to je opcija koja se može uključiti i isključiti. Međutim, promeniće se redosled prikazivanja analiziranih slučajeva. Datoteku vraćate u prvobitni redosled sortiranjem po ID-u, komandom **Sort Cases**.

1. U meniju **Data** izaberite stavku **Split File**.
2. Pritisnite **Compare groups** i zadajte promenljivu po kojoj treba grupisati (npr. sex). Pritisnite **OK**.

Nakon podele datoteke na grupe (u ovom slučaju, muškarce i žene), sve analize se obavljaju za svaku grupu zasebno. Kada završite analize, vratite se i isključite opciju **Split File**.

1. U meniju **Data** izaberite stavku **Split File**.
2. Pritisnite prvu tačku (**Analyze all cases, do not create groups**). Pritisnite **OK**.

Izbor slučajeva

Za neke analize moraćete da izaberete podskup celog uzorka (npr. samo muškarce).

1. Otvorite meni **Data** i u njemu izaberite stavku **Select Cases**.
2. Pritisnite dugme **if condition is satisfied**.
3. Pritisnite dugme **IF**.
4. Izaberite promenljivu koja definiše grupu od interesa (u našem primeru, pol, tj. sex).
5. Pritisnite dugme sa strelicom; izabrana promenljiva će preći u odgovarajuće polje. Pritisnite dugme sa znakom jednakosti (=) prikazano unutar tastature na ekranu.
6. Upišite broj koji odgovara izabranoj grupi (pročitajte ga u šifarniku.) Primera radi, u uzorku su muškarci kodirani brojem 1, pa biste upisali 1. Na komandnoj liniji bi trebalo da piše: sex=1.
7. Pritisnite **Continue** i zatim **OK**.

U analizama (npr. korelaciji) rađenim posle ovog postupka izbora slučajeva, biće obuhvaćena samo izabrana grupa (npr. muškarci). Kada završite analize, vratite se i isključite opciju **Select Cases**.

1. Otvorite meni **Data** i u njemu izaberite stavku **Select Cases**.
2. Pritisnite stavku **All cases**. Pritisnite **OK**.

Upotreba skupova

U velikim datotekama s podacima ponekad je zametno listati kroz mnoštvo promenljivih navedenih u SPSS-ovim okvirima za dijalog da bi se stiglo do onih potrebnih za analize. SPSS omogućava definisanje i korišćenje skupova promenljivih. To je posebno korisno u datoteci *survey3ED.sav*, gde ima mnogo pojedinačnih stavki koje se sabiraju, a zbirovi su tek na dnu datoteke. U sledećem primeru napraviću skup koji obuhvata samo demografske promenljive i zbirove.

1. Otvorite meni **Utilities** i u njemu izaberite stavku **Define Variable Sets**.
2. U listi izaberite promenljive. Među njima neka budu ID, demografske promenljive (od sex do smoke number) i svi zbirovi na dnu datoteke, od Total Optimism nadalje. Premestite te promenljive u polje **Variable in Set**.
3. U polje **Set Name** upišite neko prikladno ime skupa (npr. Totals ili Zbirovi).
4. Pritisnite dugme **Add Set** i zatim **Close**.

Skup ste napravili, ali da biste ga mogli koristiti, najpre ga morate aktivirati.

1. Otvorite meni **Utilities** i u njemu izaberite stavku **Use Variable Sets**.
2. U listi definisanih skupova promenljivih, stavite kvačicu pored imena napravljenog skupa (Zbirovi), a zatim se vratite naviše i uklonite kvačicu pored stavke **ALLVARIABLES** (jer bi se inače prikazivale sve promenljive). Ostavite kvačicu pored stavke **NEWVARIABLES**. Pritisnite **OK**.

Kada su skupovi aktivirani, samo izabrane promenljive se prikazuju u datoteci s podacima i u okvirima za dijalog potrebnim za obavljanje statističkih analiza.

Da biste isključili podelu na skupove:

1. Otvorite meni **Utilities** i u njemu izaberite **Use Variable Sets**.
2. Potvrdite opciju **ALLVARIABLES** i pritisnite **OK**.

Komentari u datoteci s podacima

U meniju **Utilities**, SPSS ima i opciju snimanja opisnih komentara u datoteci s podacima.

1. Otvorite meni **Utilities** i u njemu izaberite stavku **Data File Comments**.
2. Upišite komentare. Ako želite da se i oni prikazuju zajedno s rezultatima analiza, potvrdite polje **Display comments in output**. Komentari se snimaju zajedno s datumom kada su uneseni.

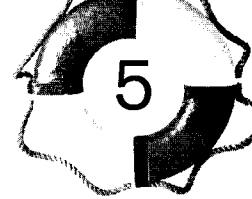
Informacije o datoteci s podacima

U poglavlju 2 pokazala sam kako se pre unošenja podataka formira šifarnik. SPSS ima i opciju **Display Data File Information**, pomoću koje se pravi sažet opis promenljivih u datoteci i njihovih vrednosti.

1. U meniju **File** izaberite **Display Data File Information**.
2. Izaberite **Working File**. Aktiviraće se prozor za rezultate (**Viewer**) i pojaviće se dve tabele: **Variable Information** i **Variable Values**. U prvoj će biti sve promenljive iz datoteke s podacima, a u drugoj brojevi kojima su kodirane kategorijske promenljive.

Prikazivanje opisnih imena promenljivih umesto njihovih numeričkih vrednosti

Kada se datoteka s podacima prikaže u prozoru **Data Editor**, obično se prikazuju i numeričke vrednosti svih promenljivih. Ukoliko hoćete da prikazete duža, opisna imena promenljivih (npr. muški, ženski) umesto njihovih numeričkih vrednosti, otvorite meni **View** i u njemu izaberite stavku **View Labels**. Ovu opciju isključujete na isti način (u meniju **View** izaberete **View Labels** da biste uklonili kvačicu ispred nje.) Drugi način da uključite ili isključite opciju **View Labels** jeste da na paleti sa alatima pritisnete ikonicu koja liči na nalepnicu ili karticu s deklaracijom.



Pronalaženje i uklanjanje pogrešnih podataka

Pre nego što počnete analizu podataka, neophodno je da iz njih uklonite greške. Veoma je lako pogrešiti prilikom unošenja podataka, a neke greške nažalost umeju da upropaste kasnije analize. Primera radi, dovoljno je upisati 35 kada je trebalo uneti 3, pa da se rezultati korelacione analize potpuno poremete. Neke analize su veoma osetljive na takozvane neregularne ili netipične tačke (engl. *outlier*), tj. vrednosti koje su mnogo veće ili mnogo manje od onih dobijenih u svim ostalim analiziranim slučajevima. Zato je bolje utrošiti vreme na početku, u traženju i otklanjanju grešaka, nego naknadno u pokušajima da se otkloni šteta koju zaostale greške prouzrokuju. Iako je taj postupak dosadan i naporan za oči kada se radi o velikim skupovima podataka, on je neophodan, a uštedeće vam mnogo nerviranja!

Postupak čišćenja podataka obuhvata više koraka:

- *Korak 1: traženje grešaka.* Prvo treba pregledati da li je vrednost neke promenljive izvan raspona mogućih vrednosti.
- *Korak 2: pronalaženje i ispravljanje grešaka u datoteci s podacima.* U datoteci s podacima treba pronaći tu grešku tj. analizirani slučaj u kome se greška javlja, i ispraviti ili obrisati tu vrednost.

Da biste imali na čemu da vežbate opisani postupak, ubacila sam greške u glavnu datoteku s podacima **survey3ED.sav** i tako dobila datoteku **error3ED.sav**, koja se može preuzeti s prateće Web lokacije ove knjige (pojedinošti videti na str. xi i u dodatku). U dodatku je i šifarnik na osnovu kojeg je vrednost svake promenljive pretvorena u odgovarajući broj tj. podatak. Zato pokrenite SPSS i otvorite **error3ED.sav**. Prorađujući navedene korake, tumačeći rezultate SPSS analiza i radeći na datoteci s podacima, upoznaćete SPSS-ove menije. Za svaki postupak sam navela i odgovarajuće SPSS komande. U poglavlju 3 je detaljnije opisana upotreba prozora **Syntax Editor** za beleženje i snimanje SPSS komandi.

Pre nego što počnete, u meniju **Edit** izaberite stavku **Options**. Na dnu kartice **Output Labels** nađite poslednje polje (**Variable values in labels shown as:**) i u njemu izaberite stavku **Values and Labels**. Time ste zadali da se za kategorijske promenljive prikazuju i tekstualni opisi (engl. *labels*) i numeričke vrednosti (tj. šifre, brojevi), što olakšava prepoznavanje grešaka.

Korak 1: Traženje grešaka

Traženje grešaka se prvenstveno svodi na traženje onih vrednosti koje su izvan raspona mogućih vrednosti date promenljive. Na primer, ako je promenljiva Pol kodirana sa 1=muški, 2=ženski, nijedna njena vrednost ne bi se smela razlikovati od 1 ili 2. Vrednosti izvan raspona mogućih vrednosti kvare rezultate statističkih analiza, pa je veoma važno ispraviti sve te greške pre nego što uopšte počnete da analizirate podatke. Tražiti greške znači pregledati učestalosti (frekvencije) *svih* promenljivih, uključujući tu i sve pojedinačne vrednosti od kojih se sastoje merne skale. Greške treba ispraviti pre izračunavanja zbirnih rezultata za te skale. Bilo bi dobro da u nekom podsetniku beležite sve pronađene greške i izmene koje ste sprovedi u datoteci s podacima.

U SPSS-u postoji više mogućnosti za traženje grešaka. Opisacu dva načina – jedan podesniji za kategorijske promenljive (npr. pol), a drugi za neprekidne (npr. starost).

Traženje grešaka u vrednostima kategorijskih promenljivih

U ovom odeljku predstaviću postupak pretraživanja kategorijskih promenljivih. U sledećem primeru pretražićemo datoteku **error3ED.sav**, koja se može preuzeti s prateće Web lokacije ove knjige (videti str. xi). Tražiću greške u vrednostima promenljivih Pol (Sex), Bračno stanje (Marital status) i Najviša završena škola (Highest education completed). U datoteku **error3ED.sav** greške su ubačene namerno, pa njih nema u glavnoj datoteci s podacima, **survey3ED.sav**.

Postupak pretraživanja kategorijskih promenljivih

1. U glavnom meniju na vrhu prozora pritisnite **Analyze**, zatim **Descriptive Statistics** i najzad **Frequencies**.
2. Izaberite promenljive koje želite da proverite (npr. pol, brak, škola).
3. Pritisnite dugme sa strelicom, čime izabrane promenljive prebacujete u polje **Variables**.
4. Pritisnite dugme **Statistics**. U odeljku **Dispersion** potvrdite polja **Minimum** i **Maximum**.
5. Pritisnite **Continue** i zatim **OK** (ili **Paste** da biste sve snimili u **Syntax Editor**).
Opisani postupak generiše sledeće komande:

```
FREQUENCIES
VARIABLES=pol brak škola
/STATISTICS=MINIMUM MAXIMUM
/ORDER= ANALYSIS
```

Sledi odabrani deo rezultata prethodnog postupka.

Statistics

	sex	marital status	highest educ completed
N	Valid 439	439	439
	Missing 0	0	0
Minimum	1	1	1
Maximum	3	8	22

sex

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1 MALES	185	42,1	42,1	42,1
2 FEMALES	253	57,6	57,6	99,8
3	1	,2	,2	100,0
Total	439	100,0	100,0	

marital status

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1 SINGLE	105	23,9	23,9	23,9
2 STEADY RELATIONSHIP	37	8,4	8,4	32,3
3 LIVING WITH PARTNER	37	8,4	8,4	40,8
4 MARRIED FIRST TIME	189	43,1	43,1	83,8
5 REMARRIED	30	6,8	6,8	90,7
6 SEPARATED	10	2,3	2,3	92,9
7 DIVORCED	24	5,5	5,5	98,4
8 WIDOWED	7	1,6	1,6	100,0
Total	439	100,0	100,0	

highest educ completed

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1 PRIMARY	2	,5	,5	,5
2 SOME SECONDARY	52	11,8	11,8	12,3
3 COMPLETED HIGH SCHOOL	85	19,4	19,4	31,7
4 SOME ADDITIONAL TRAINING	120	27,3	27,3	59,0
5 COMPLETED UNDERGRADUATE	123	28,0	28,0	87,0
6 POSTGRADUATE COMPLETED	56	12,8	12,8	99,8
22	1	,2	,2	100,0
Total	439	100,0	100,0	

Rezultati postupka su podeljeni na dva dela. U prvoj tabeli je kratak pregled *svih* traženih promenljivih. U preostalim tabelama, za svaku promenljivu pojedinačno dati su raščlanjeni rasponi odgovora. (Navedeni su i tekstualni opis kategorijske vrednosti i njemu odgovarajuća šifra (broj); takav prikaz rezultata se dobija kada postavke (**Options**) izmenite prema preporukama datim u prethodnom delu poglavlja.)

- Pogledajte najniže i najviše vrednosti. Imaju li smisla? Jesu li unutar raspona mogućih vrednosti date promenljive? U prvoj tabeli (nazvanoj **Statistics**) vidi se da je za promenljivu Pol minimum 1, a maksimum 3. Ta trojka je neispravna, pošto bi najveća vrednost trebalo da bude 2 – pogledajte šifarnik u dodatku. Maksimalna vrednost za najvišu završenu školu je 22 i to je pogrešno, pošto bi maksimum trebalo da iznosi 6.
- Proverite koliki je broj valjanih (engl. *valid*) i nedostajućih (engl. *missing*) slučajeva. Ukoliko nedostajućih slučajeva ima mnogo, zapitajte se zašto je tako? Da niste grešili prilikom unošenja podataka, npr. pobrkali kolone prilikom upisivanja? Dešava se da na dnu datoteke s podacima bude praznih redova, nastalih tako što čovek greškom pre-daleko pomeri kursor i nenamerno napravi prazne ćelije. Ako vam se to desilo, otvorite prozor **Data Editor**, pređite u red koji na sadrži podatke, pritisnite osenčenu ćeliju s brojem praznog reda, pa pritisnite taster Delete na tastaturi. Ponovite postupak **Frequencies** da biste dobili ispravne vrednosti.
- Ostale tabele s rezultatima odgovaraju po jednoj ispitivanoj promenljivoj. U njima se vidi koliko je analiziranih slučajeva upalo u svaku od legitimnih kategorija. Takođe se vidi u koliko slučajeva (redova) postoje vrednosti koje su izvan opsega. Tu je jedan slučaj u kome je šifra pola 3 i jedna osoba koja za najvišu završenu školu ima šifru 22. Te greške svakako treba pronaći i ukloniti, ali mi ćemo najpre pokazati kako se greške traže u promenljivama s neprekidnim (kontinualnim) vrednostima u datoteci s podacima.

Provera neprekidnih promenljivih

Postupak pretraživanja neprekidnih promenljivih

1. U glavnom meniju na vrhu prozora pritisnite **Analyze**, zatim **Descriptive Statistics**, pa **Descriptives**.
2. Izaberite promenljive koje hoćete da proverite (npr. starost – *age*). Zatim pritisnite dugme sa strelicom da biste izabrane promenljive prebacili u polje **Variables**.

3. Pritisnite dugme **Options**. Ovde se može izabrati više statističkih pokazatelja; glavni su srednja vrednost (engl. *mean*), standardno odstupanje, tj. standardna devijacija (engl. *standard deviation* ili skraćeno **STDDEV**), minimum i maksimum. Pritisnite pokazatelje koje SPSS treba da izračuna.
4. Pritisnite **Continue** i zatim **OK** (ili **Paste** da biste sve snimili u **Syntax Editor**). Prethodni postupak daje ove komande:

```
DESCRIPTIVES
  VARIABLES=starost
  /STATISTICS=MEAN STDDEV MIN MAX
```

Opisani postupak daje sledeće rezultate:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
age	439	2	82	37.39	13.293
Valid N (listwise)	439				

- Proverite iznose minimuma i maksimuma. Imaju li smisla? U ovom slučaju, raspon godina starosti je od 2 do 82. Minimum 2 kazuje da je došlo do greške, pošto bi uzorak trebalo da sadrži samo podatke o odraslim osobama.
- Ima li iznos srednje vrednosti (engl. *mean*) smisla? Svaki broj izvan raspona dozvoljenih vrednosti kvvari srednju vrednost. Ako se radi o promenljivoj koja se meri na nekoj skali, da li je srednja vrednost u skladu s vašim očekivanjima na osnovu prethodnih istraživanja obavljenih uz korišćenje iste merne skale? Da li je srednja vrednost na sredini raspona dozvoljenih vrednosti ili je bliža jednom kraju raspona? To se ponekad dešava kada se mere obeležja kao što su anksioznost ili potištenost.

Korak 2: pronalaženje i ispravljanje grešaka u datoteci s podacima

Šta da radimo kada pronađemo neki odgovor izvan opsega dozvoljenih (npr. šifru 3 za pol)? Prvo treba naći grešku u datoteci s podacima. Nemojte pokušavati da je sami pronađete tako što ćete čitati ceo skup podataka. Ima više dobrih načina da SPSS pronađe grešku u datoteci; ovde ću pokazati dva.

Prvi način

1. U prozoru **Data Editor** otvorite meni **Data** i u njemu izaberite stavku **Sort Cases**.
2. U okviru za dijalog koji će se pojaviti, pritisnite promenljivu (npr. pol) koja sadrži grešku i zatim strelicu kojom je premeštate u polje **Sort By**. U zavisnosti od toga želite li najveće vrednosti na dnu ili na vrhu, pritisnite **ascending** odnosno **descending**. Hoćemo da nađemo osobu za koju promenljiva pol ima šifru 3, pa bi trebalo da izaberete **descending** (kako bi se ta trojka prikazala odmah na vrhu, pošto su prihvatljive samo vrednosti 1 i 2).
3. Pritisnite **OK**.

U **Data Editoru** pritisnite jezičak kartice **Data View** tako da se prikažu vrednosti podataka. Slučaj koji sadrži pogrešnu vrednost izabrane promenljive (šifru 3 za pol) trebalo bi da je na vrhu datoteke. Pogledajte kolonu promenljive pol. U ovom primeru, videćete da promenljiva pol u prvom navedenom slučaju (ID=103) ima vrednost 3. Da su ovo vaši podaci, trebalo bi da uzmete upitnike i utvrdite da li je muškog ili ženskog pola osoba sa identifikacionim brojem 103. Zatim biste obrisali vrednost 3 i upisali ispravnu vrednost. Zabeležite taj podatak u svoj podsetnik. Ukoliko su vam originalni podaci nedostupni, uklonite pogrešnu vrednost, a SPSS će je zameniti sistemskom oznakom za nedostajući podatak. (To je tačka koju će SPSS sam upisati, pa vi ne treba da je upisujete ručno).

Nakon što otkrijete grešku u datoteci s podacima, proverite ima li grešaka u susednim kolonama (promenljivama) istog analiziranog slučaja. U ovom slučaju, osoba ID=103 ima za starost pogrešnu vrednost, 2.

Sledi prikaz drugog načina pronalaženja slučaja s pogrešno upisanim polom.

Drugi način

1. U prozoru **Data Editor** pritisnite jezičak kartice **Data View** tako da se prikažu vrednosti podataka.
2. Pritisnite ime promenljive (npr. pol) koja sadrži grešku.
3. Nakon što pritisnete ime promenljive, trebalo je da cela kolona bude izabrana.
4. Otvorite meni **Edit** i u njemu izaberite stavku **Find**.
5. U polje **Search for** upišite neispravnu vrednost koju tražite (u ovom slučaju, 3).
6. Pritisnite dugme **Search Forward**. SPSS će pretražiti datoteku i zaustaviti se na prvom slučaju koji sadrži zadatu vrednost. Zapišite negde ID broj tog slučaja (broj u prvoj koloni). Trebaće vam da biste ga pronašli među zapisima ili upitnicima i tamo pročitali ispravnu vrednost.
7. Ako treba da potražite druge slučajeve sa istom pogrešnom vrednošću, ponovo pritisnite **Search Forward**. U ovom primeru, kolona **Frequency** u tabeli pol na str. 45 kazuje nam da postoji samo jedna neispravna vrednost 3.

Kada ispravite greške, ponovite postupak pretraživanja kategorijskih promenljivih (**Analyze**, zatim **Descriptive Statistics** i najzad **Frequencies**) da biste utvrdili ima li u datoteci još grešaka. Dešava se da čovek jednu grešku ispravi, a drugu napravi. Sve skupa jeste zametno, ali nema smisla počinjati analize dok je skup podataka neispravan. Uspeh celog istraživanja zavisi od toga. Nema prečica!

Sažet prikaz slučajeva

U postupku čišćenja podataka, još jedna SPSS-ova procedura ume dobro da posluži; to je funkcija **Summarize Cases**, koja omogućava da izaberete određene promenljive i samo njih prikažete za sve slučajeve.

1. Otvorite meni **Analyze**, pritisnite **Reports**, pa **Summarize Cases**.
2. Izaberite promenljivu ID i ostale promenljive koje vas zanimaju (npr. pol, deca, pušač). Uklonite kvačicu ispred polja **Limit cases to first 100**.
3. Pritisnite dugme **Statistics** i uklonite **Number of cases** iz polja **Cell Statistics**. Pritisnite dugme **Continue**.
4. Pritisnite dugme **Options** i uklonite kvačicu ispred polja **Subheading for totals**. Pritisnite **Continue** i zatim **OK** (ili **Paste** da biste sve snimili u **Syntax Editor**).

Prethodni postupak generiše ove komande:

```
SUMMARIZE
/TABLES=id pol deca pušač
/FORMAT=VALIDLIST NOCASENUM NOTOTAL
/TITLE='Sažet prikaz slučajeva'
/MISSING=VARIABLE
/CELLS=NONE.
```

Ovako izgleda deo rezultata prethodnog postupka.

Sažet prikaz slučajeva

	id	pol	deca	pušač
1	301	1 MUŠKI	2 NE	2 NE
2	32	1 MUŠKI	2 NE	2 NE
3	79	2 ŽENSKI	1 DA	2 NE
4	357	1 MUŠKI	1 DA	2 NE
5	226	2 ŽENSKI	2 NE	2 NE
6	467	1 MUŠKI	2 NE	2 NE

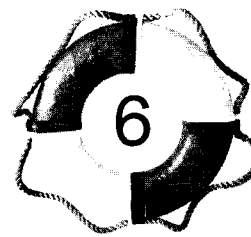
U ovom poglavlju tražili smo greške u nekoliko promenljivih u datoteci s podacima, tek da bismo pokazali kako se to radi. Naravno, u sopstvenom istraživanju proverili biste sve promenljive. Da biste vežbali pronalaženje grešaka, ponovite opisane postupke za sve promenljive u datoteci **error-3ED.sav**. U nju su namerno ubačene greške da biste zaista imali šta da tražite. Ispravne vrednosti svake promenljive navedene su u dodatku, u šifarniku za datoteku s podacima **survey3ED.sav**.

Ukoliko vam treba više informacija o pronalaženju i uklanjanju grešaka u podacima, preporučujem poglavlje 4 u knjizi koju su napisale Tabachnik i Fidell (2007).

DEO III

Preliminarne analize

Pregled datoteke s podacima i istraživanje prirode promenljivih može početi tek kada je datoteka s podacima čista, bez grešaka. Čišćenje je priprema za primenu konkretnih statističkih tehnika pomoću kojih ćete dobiti odgovore na pitanja koja bi istraživanje trebalo da rasvetli. Treći deo knjige sastoji se od pet poglavlja. U poglavlju 6 predstavljeni su postupci dobijanja opisnih statističkih pokazatelja i za kategorijske i za neprekidne promenljive. U njemu su objašnjeni i proverena normalnosti raspodele vrednosti neprekidnih promenljivih i proverena postojanja netipičnih tačaka (engl. *outliers*). Za upoznavanje podataka umeju dobro da posluže i dijagrami. U poglavlju 7 predstavljen je deo češće upotrebljivanih dijagrama koje SPSS ume da nacrtati. Katkada datoteku s podacima treba obraditi da bi postala prikladna za određene analize; recimo, sabrati iznose pojedinačnih stavki da bi se dobili zbirni rezultati na skali ili podeliti neprekidnu promenljivu na manji broj diskretnih kategorija. Te tehnike obrade podataka objašnjene su u poglavlju 8. Postupak provere pouzdanosti (unutrašnje saglasnosti) merne skale predstavljen je u poglavlju 9. To je posebno važno pri ispitivanju rezultata anketa i u studijama u kojima se pomoću skala mere karakteristike (obeležja) ličnosti, njihovi stavovi, uverenja itd. U ovom delu knjige je i poglavlje koje bi trebalo da vas uputi kako da izaberete statističke tehnike prikladne za traženje odgovora na pitanja koja su predmet istraživanja. U poglavlju 10 prikazane su neke statističke tehnike dostupne u SPSS-u i svaki korak postupka izbora statističkih tehnika prikladnih za konkretne potrebe. Istaknuti su neki važni aspekti koje treba uzeti u obzir (npr. vrsta pitanja, vrsta podataka, obeležja promenljivih).



Opisni statistički pokazatelji

Kada obezbedite da u datoteci podataka nema grešaka (ili barem da vrednosti nijedne promenljive nisu izvan raspona dozvoljenih vrednosti), možete početi opisnu (deskriptivnu) fazu analize podataka. Opisne statističke analize mogu se upotrebiti na više načina, među kojima i:

- za opisivanje uzorka u odeljku Metodologija izveštaja o obavljenom istraživanju;
- za proveru da li neka od promenljivih krši pretpostavke na kojima počivaju statističke tehnike pomoću kojih treba odgovoriti na pitanja koja bi istraživanje trebalo da rasvetli;
- za pronalaženje odgovora na konkretna istraživačka pitanja.

Informacije potrebne za opisivanje uzorka u odeljku Metodologija izveštaja o istraživanju, daju i postupci provere podataka pomenuti u poglavlju 5. U studijama koje obuhvataju ljude, dobro je prikupiti podatke o broju osoba (tj. slučajeva) u uzorku, broju i procentualnom udelu osoba muškog odnosno ženskog pola u uzorku, rasponu i srednjoj vrednosti starosti tih osoba, nivou njihovog obrazovanja i svim drugim relevantnim informacijama. Pre nego što pristupite mnogim statističkim analizama (kao što su t-test, ANOVA, korelacija), treba proveriti da li podaci krše neku od „pretpostavki“ na kojima počivaju pojedina ispitivanja. (One će biti detaljno objašnjene u četvrtom i petom delu knjige.)

Ispitivanje hipoteza (pretpostavki) obično podrazumeva da su opisni statistički pokazatelji promenljivih već izračunati. Među njima su srednja vrednost, standardno odstupanje, raspon vrednosti promenljivih, asimetrija (engl. *skewness*) i spljoštenost (engl. *kurtosis*) raspodele. U SPSS-u se opisni statistički pokazatelji mogu dobiti na više načina, iz učestalosti (*Frequencies*) pojedinih vrednosti ili pomoću funkcija *Descriptives* ili *Explore*. Sve one su u padajućem meniju *Analyze*, stavka *Descriptive Statistics*. Međutim, postupci za kategorijske promenljive razlikuju se od onih za neprekidne (kontinualne). Deo tih statističkih pokazatelja (npr. srednja vrednost, standardno odstupanje) nije primenljiv na kategorijske promenljive. U naredna dva odeljka prikazaćemo različite pristupe kategorijskim i neprekidnim promenljivama. Ako biste želeli da i sami na računaru pratite primere iz ovog poglavlja, otvorite datoteku *survey3ED.sav*.

Kategorijske promenljive

Za dobijanje opisnih statističkih pokazatelja *kategorijskih* promenljivih upotrebite učestalosti (**Frequencies**) njihovih vrednosti. Tako ćete saznati koliko je osoba dalo koji odgovor (npr. osoba muškog odnosno ženskog pola). Za kategorijske promenljive kao što su pol ili bračno stanje nema smisla izračunavati srednju vrednost, standardna odstupanja itd.

Postupak dobijanja opisnih statističkih pokazatelja kategorijskih promenljivih

1. U glavnom meniju na vrhu prozora pritisnite **Analyze**, zatim **Descriptive Statistics**, pa **Frequencies**.
2. Izaberite kategorijske promenljive koje vas zanimaju (npr. pol). Prebacite ih u polje **Variables**.
3. Pritisnite **Continue** i zatim **OK** (ili dugme **Paste** da biste komande snimili u **Syntax Editor**).

Prethodni postupak generise ove komande:

```
FREQUENCIES  
VARIABLES=pol  
/ORDER= ANALYSIS
```

Rezultat prethodnog postupka je:

Statistics

POL		
N	Valid	439
	Missing	0
Minimum		1
Maximum		2

POL

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MUSKI	185	42,1	42,1	42,1
	ŽENSKI	254	57,9	57,9	100,0
Total		439	100,0	100,0	

Tumačenje rezultata dobijenih iz učestalosti pojedinih vrednosti

Iz prikazanih rezultata vidimo da u uzorku ima 185 muškaraca (42,1 procenat) i 254 žena (57,9 procenata), koji ukupno čine 439 ispitanika. Treba zabeležiti broj ispitanika u raznim podskupovima uzorka. U nekim analizama (kao što je, recimo, ANOVA) lakše je kada su grupe (podskupovi) približno jednake veličine. Neke analize nisu prikladne za grupe čije se veličine mnogo razlikuju, naročito kada su grupe male.

Neprekidne promenljive

Za neprekidne promenljive (npr. starost) lakše je upotrebiti funkciju **Descriptives**, koja će izračunati „zbirne“ statističke pokazatelje kao što su srednja vrednost (engl. *mean*), medijana i standardno odstupanje (engl. *standard deviation*). Ne treba izlistavati sve vrednosti, posto ih neke promenljive imaju više stotina. Opisni pokazatelji svih neprekidnih promenljivih mogu se dobiti jednim potezom; ne treba raditi jednu po jednu promenljivu, nego sve promenljive od interesa zajedno prebacite u polje **Variables**. Međutim, ukoliko imate mnogo promenljivih, spisak rezultata će biti veoma dugačak; tada je lakše raditi jednu po jednu grupu promenljivih i tokom postupka beležiti šta je urađeno.

Postupak dobijanja opisnih statističkih pokazatelja za neprekidne promenljive

1. U glavnom meniju na vrhu prozora pritisnite **Analyze**, zatim **Descriptive Statistics**, pa **Descriptives**.
2. Izaberite neprekidne promenljive (npr. starost, ukupan subjektivno doživljen stres) za koje želite da dobijete opisne statističke pokazatelje. Prebacite ih u polje **Variables**.
3. Pritisnite dugme **Options**. Potvrdite polja **mean**, **standard deviation**, **minimum**, **maximum**, **skewness** i **kurtosis** (čime ste izabrali da se izračunaju srednja vrednost, standardno odstupanje, minimum, maksimum, asimetrija i spljoštenost raspodele, redom).
4. Pritisnite **Continue** i zatim **OK** (ili dugme **Paste** da biste komande snimili u **Syntax Editor**).

Prethodni postupak generise ove komande:

```
DESCRIPTIVES  
VARIABLES=starost usudstres  
/STATISTICS= MEAN STDDEV MIN MAX KURTOSIS SKEWNESS
```


Opisni statistički pokazatelji

	N	Minimum	Maximum	Mean	Std.	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
STAROST	439	18	82	37,44	13,20	,606	,117	-,203	,233
Ukupan subjektivno doživljen stres	433	12	46	26,73	5,85	,245	,117	,182	,234
Ukupan optimizam	435	7	30	22,12	4,43	-,494	,117	,214	,234
Ukupno ovladavanje	436	8	28	21,76	3,97	-,613	,117	,285	,233
Ukupno SSDOUS	431	20	88	60,60	11,99	-,395	,118	,247	,235
Valid N (listwise)	425								

Tumačenje rezultata procedure Descriptives

U prikazanim rezultatima, sažete su informacije koje su bile potrebne za svaku promjenljivu. Na primer, za promjenljivu Starost, imamo podatke o 439 ispitanika, raspon njihove starosti je od 18 do 82 godine, srednja vrednost 37,44 godine, i standardno odstupanje od te srednje vrednosti iznosi 13,20 godina. To bi trebalo navesti kao opis uzorka u odeljku *Metodologija* izveštaja o istraživanju.

Opisni pokazatelji kazuju ponešto i o raspodeli vrednosti neprekidnih promjenljivih (asimetriji i spljoštenosti njihove raspodele). Te informacije su potrebne za parametarske statističke tehnike (kao što su t-test i analiza varijanse, tj. rasturanja rezultata oko njihove srednje vrednosti). **Asimetrija** pokazuje asimetriju raspodele, a **spljoštenost** koliko je raspodela „šiljatija“ ili „spljoštenija“ od normalne. U normalnoj raspodeli, asimetrija i spljoštenost (kurtosis) jednake su 0 (sto se u društvenim naukama retko sreće).

Pozitivne vrednosti asimetrije pokazuju da je većina dobijenih rezultata levo od srednje vrednosti, među manjim vrednostima. Negativne vrednosti asimetrije pokazuju da je većina rezultata desno od srednje vrednosti, među većim vrednostima. Pozitivna vrednost spljoštenosti pokazuje da je raspodela šiljatija od normalne (ima više rezultata nagomilanih oko centra raspodele), te da ima tanke dugačke repove. Spljoštenost manja od 0 pokazuje da je raspodela pljosnatija od normalne (ima više slučajeva na repovima). Asimetričnost „nema znatnijeg uticaja na rezultate analize“ ako je uzorak razumno velik (Tabachnik i Fidell, 2007, str. 80). Spljoštenost ponekad prouzrokuje prenisku procenu varijanse (dispersije, rasturanja oko srednje vrednosti), ali se i taj rizik smanjuje kod većih uzoraka (preko 200 slučajeva; videti Tabachnik i Fidell, 2007, str. 80).

Dakle, asimetriju i spljoštenost treba obavezno izračunati kada uzorak sadrži manje od 200 analiziranih slučajeva. U velikim uzorcima, testovi za izračunavanje asimetrije i spljoštenosti previše su osetljivi. Tabachnik i Fidell (2007, str. 81) preporučuju da se oblik raspodele pogleda npr. na histogramu. Postupak daljnjeg ocenjivanja normalnosti raspodele rezultata istraživanja objasnićemo u nastavku odeljka.

Nedostajući podaci

Kada se radi istraživanje, pogotovo ono s ljudima, retko se za svaki ispitivani slučaj dobiju potpuni podaci. Uvek treba pregledati da li u datoteci nedostaju neki podaci. Pokrenite funkciju **Descriptives** i za svaku promjenljivu utvrdite koliki procenat podataka nedostaje. Kada pronađete promjenljivu kojoj neočekivano nedostaje mnogo podataka, zapitajte se koji je razlog tome i da li podaci nedostaju nasumično ili tu možete uočiti neku pravilnost (npr. mnogo žena nije odgovorilo na pitanje o svojoj starosti)?

Kada dođe vreme za statističke analize, moraćete da razmislite i odlučite šta ćete preduzeti u vezi s nedostajućim podacima. U mnogim SPSS-ovim statističkim postupcima, preko dugmeta **Options** pristupate mogućim načinima tretiranja nedostajućih podataka. Tu treba pažljivo birati, jer izbor može značajno uticati na rezultate, naročito ako prilažete spisak promjenljivih i iste analize ponavljate za sve promjenljive (npr. korelacije unutar grupe promjenljivih, t-testovi niza zavisnih promjenljivih).

- Opcija **Exclude cases listwise** znači da će biti analizirani samo slučajevi u kojima za *sve promjenljive* navedene u polju **Variables** postoje *svi* podaci. Svaki ispitivani slučaj za koji nedostaje makar i delić podataka, uopšte neće biti analiziran. To može znatno i nepotrebno ograničiti veličinu uzorka.
- Opcija **Exclude cases pairwise** znači da će slučaj (osoba) biti isključen samo iz onih analiza za koje mu nedostaje neki neophodan podatak. Dakle, i takvi slučajevi će biti analizirani kad god je to moguće, tj. kad god postoje podaci potrebni za datu analizu.
- Opcija **Replace with mean**, dostupna u nekim SPSS-ovim statističkim postupcima (npr. višestrukoj regresiji), znači da će biti izračunata srednja vrednost svih promjenljivih i da će njome biti zamenjeni podaci koji nedostaju. Ovu opciju *nikada* ne bi trebalo da koristite, pošto može znatno da iskrivi rezultate analize, naročito onda kada nedostaje mnogo podataka.

Kad god sprovedite neki statistički postupak, pritisnite dugme **Options** i proverite koja je od navedenih opcija potvrđena (jer se podrazumevana opcija menja u zavisnosti od postupka). Ako nemate jak razlog da postupite drugačije, preporučila bih vam da slučajeve isključite samo iz onih analiza za koje im nedostaju podaci. Jedina situacija kada bi vam moglo zatrebati da analize ograničite samo na slučajeve koji imaju podatke za sve promjenljive jeste ona kada treba razmotriti samo podskup slučajeva koji daje potpun skup rezultata.

Za iskusnije korisnike, SPSS ima naprednije i složenije mogućnosti procene nedostajućih podataka, npr. pripisivanjem (engl. *imputation*). One su dostupne u SPSS-ovoj analizi **Missing Value Analysis**, koja je neobavezan dodatan softverski modul. Njome biste mogli da otkrijete i pravilnost zajedničku nedostajućim podacima.

Procena normalnosti raspodele

Mnoge statističke tehnike predstavljene u četvrtom i petom delu knjige počivaju na pretpostavci da je raspodela rezultata zavisne promenljive normalna. Normalna raspodela je simetrična, zvonolika kriva s najvećim brojem rezultata u sredini i manjim brojem rezultata prema krajevima (repovima) zvona (videti Gravetter i Wallnau, 2004, str. 48). Normalnost se može donekle proceniti na osnovu izračunatih vrednosti asimetrije i spljoštenosti zvona, što je opisano u prethodnom odeljku. Međutim, SPSS u meniju **Descriptive Statistics** ima i stavku **Explore** koja nudi druge načine. Opisaću taj postupak na primeru procene normalnosti raspodele rezultata za promenljivu Ukupan subjektivno doživljen stres u celom uzorku. Na raspolaganju vam je i mogućnost da to uradite zasebno za razne grupe u uzorku, tako što ćete u spisku **Factor List** unutar okvira za dijalog **Explore** zadati još jednu dodatnu kategorijsku promenljivu (npr. pol).

Postupak procene normalnosti pomoću funkcije Explore

1. U glavnom meniju na vrhu prozora pritisnite **Analyze**, zatim **Descriptive Statistics**, i najzad **Explore**.
2. Pritisnite promenljive koje vas zanimaju (npr. Ukupan subjektivno doživljen stres) i zatim dugme sa strelicom da biste ih prebacili u polje **Dependent List**.
3. U polje **Label Cases by**: stavite promenljivu ID.
4. U odeljku **Display** treba da je potvrđeno polje **Both**.
5. Pritisnite dugme **Statistics**, zatim **Descriptives**, pa **Outliers**.
6. Pritisnite dugme **Plots**. U odeljku **Descriptive** pritisnite **Histogram**. Pritisnite opciju **Normality plots with tests**. Pritisnite **Continue**.
7. Pritisnite dugme **Options**. U odeljku **Missing Values** pritisnite **Exclude cases pairwise**. Pritisnite **Continue** i zatim **OK** (ili dugme **Paste** da biste komande snimili u **Syntax Editor**).

Navedeni postupak će generisati sledeće komande:

```
EXAMINE  
VARIABLES=usudstres  
/ID= id  
/PLOT BOXPLOT HISTOGRAM NPLOT  
/COMPARE GROUP  
/STATISTICS DESCRIPTIVES EXTREME  
/CINTERVAL 95  
/MISSING PAIRWISE  
/NOTOTAL.
```

Rezultat postupka je:

Descriptives

			Statistic	Std. Error
total perceived stress	Mean		26,73	,281
	95% Confidence Interval for Mean	Lower Bound	26,18	
		Upper Bound	27,28	
	5% Trimmed Mean		26,64	
	Median		26,00	
	Variance		34,194	
	Std. Deviation		5,848	
	Minimum		12	
	Maximum		46	
	Range		34	
	Interquartile Range		8	
	Skewness		,245	,117
	Kurtosis		,182	,234

Extreme Values

		Case Number	id	Value
total perceived stress	Highest	1	7	24
		2	262	157
		3	216	61
		4	190	6
		5	257	144
	Lowest	1	366	404
		2	189	5
		3	247	127
		4	244	119
		5	98	301

a. U tabeli ekstremnih vrednosti prikazan je samo deo slučajeva s vrednošću 42.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
total perceived stress	,069	433	,000	,992	433	,021

a. Korekcija statističke značajnosti (neslučajnosti) po Lillieforsu

Tumačenje rezultata funkcije Explore

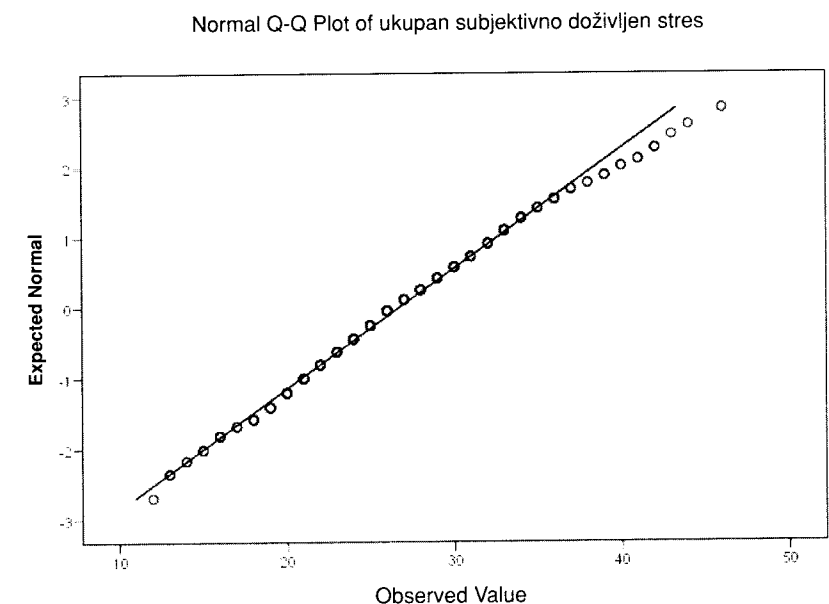
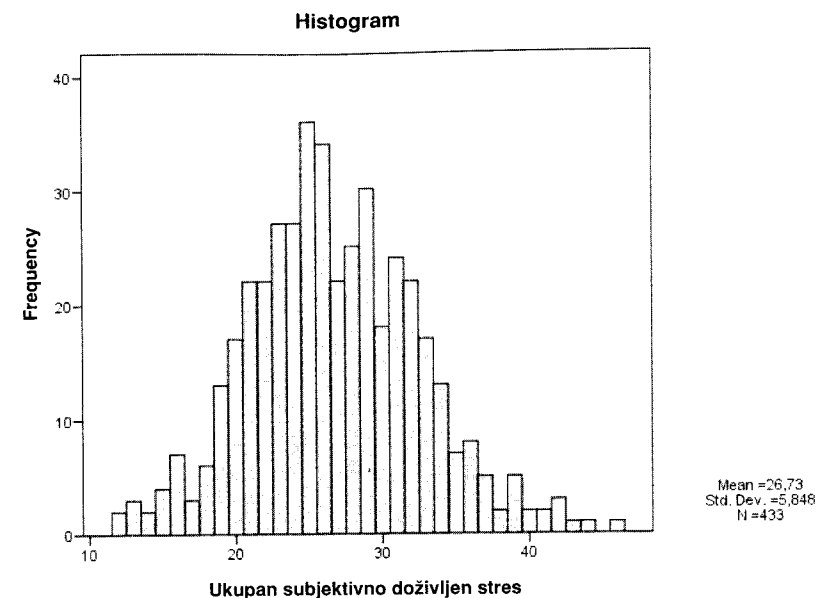
Navedeni rezultati sadrže obilje informacija. To ume da uplaši čoveka dok ne nauči šta da traži. U ovom odeljku protumačićemo te rezultate korak po korak.

- U tabeli **Descriptives** dati su opisni pokazatelji i druge informacije o analiziranim promenljivama. Kada se u spisku **Factor List** zada neka promenljiva grupisanja, informacije će biti izračunate zasebno za svaku grupu, a ne za uzorak kao celinu. Deo tih informacija ćete svakako prepoznati (mean, median, std deviation, minimum, maximum itd., tj. srednja vrednost, standardno odstupanje, minimum, maksimum itd.). Možda ne znate šta je **5% Trimmed Mean**. Da bi dobio taj broj, SPSS zanemari gornjih i donjih 5% slučajeva i bez njih izračuna novu sred-

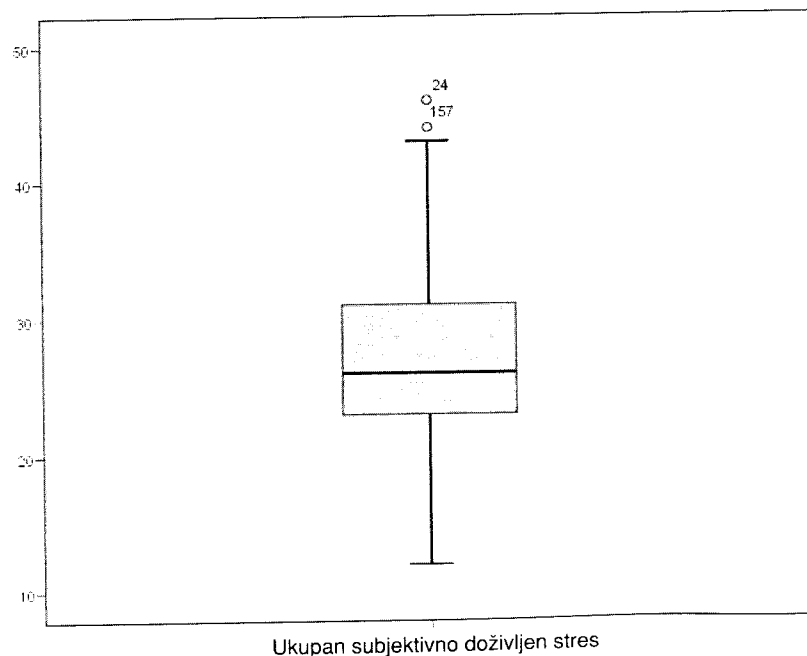
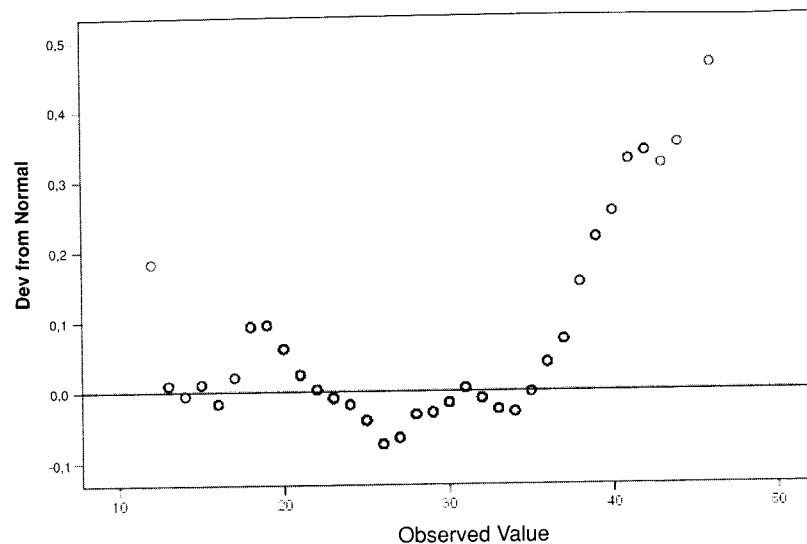
nju vrednost. Kada uporedite prvobitnu (26,73) i novu srednju vrednost (26,64), shvatićete da li ekstremne vrednosti mnogo utiču na srednju vrednost. Ako se te dve srednje vrednosti mnogo razlikuju, možda bi ekstremne tačke trebalo dodatno ispitati. ID brojevi najekstremnijih slučajeva prikazani su u tabeli **Extreme Values**.

- Među rezultatima su i asimetrija (**skewness**) i spljoštenost (**kurtosis**), koji opisuju raspodelu rezultata unutar dveju grupa (videti opis značenja tih pokazatelja dat u prethodnom odeljku).
- U tabeli **Tests of Normality** dati su rezultati ispitivanja normalnosti raspodele koje su izumeli Kolmogorov i Smirnov. Normalnost se pokazuje statistički neznačajnim (slučajnim) odstupanjem od normalnosti, tj. iznosom **Sig.** većim od 0,05. U ovom slučaju **Sig.** iznosi 0,000, što pokazuje da pretpostavka o normalnosti raspodele nije potvrđena i da je moramo odbaciti. To je sasvim uobičajeno za velike uzorke.
- Stvarni oblik raspodele rezultata svake grupe vidi se na njihovim **histogramima**. U ovom primeru izgleda kao da su rezultati prilično normalno raspodeljeni. To potvrđuje i izgled krive normalne verovatnoće, **Normal Q-Q Plot**. Na njoj se opažena vrednost rezultata crta zajedno sa očekivanom vrednošću koju bi dala normalna raspodela. Što je linija opaženih rezultata bliža pravoj liniji, to je raspodela bliža normalnoj.
- Kriva **Detrended Normal Q-Q Plot** prikazuje stvarno odstupanje opaženih rezultata od horizontalne linije koja predstavlja normalnu raspodelu. Trebalo bi da većina tačaka bude približno simetrično raspoređena oko te horizontalne linije i da se tačke nigde ne gomilaju.
- Na poslednjoj slici je je **pravougaoni dijagram** (engl. *boxplot*) raspodele rezultata za dve grupe. Pravougaonik predstavlja 50 procenata slučajeva, pri čemu repovi (linije koje izlaze iz pravougaonika) idu do najvećih i najmanjih vrednosti. Katkada se vide i izolovani kružići izvan tog raspona; njih je SPSS razvrstao u netipične tačke. Linija unutar pravougaonika je medijana. U narednom odeljku, u kome se govori o otkrivanju netipičnih tačaka, biće još reči o pravougaonom dijagramu.

U gornjem primeru, raspodela je bila prilično normalna. To se ne događa baš često. Na mnogim skalama i u merama koje se koriste u društvenim naukama, rezultati su raspodeljeni asimetrično, levo ili desno od srednje vrednosti (vrednost asimetrije pozitivna ili negativna). To uopšte ne mora da znači da je problem u mernoj skali, već pokazuje stvarnu prirodu konstrukta koji se meri. Primera radi, mere zadovoljstva životom često su negativno asimetrične u društvima čija je većina umereno zadovoljna svojom sudbinom. Kliničke mere anksioznosti ili potištenosti (depresije) u opštoj populaciji često su pozitivno asimetrične, jer većina ljudi ima relativno malo simptoma tih poremećaja. Neki autori u ovoj oblasti preporučuju da se asimetrično raspodeljeni rezultati statistički „transformišu“. O tome ćemo još govoriti u poglavlju 8.



Detrended Normal Q-Q Plot of ukupan subjektivno doživljen stres



Otkrivanje netipičnih tačaka

Mnoge statističke tehnike razmatrane u ovoj knjizi osjetljive su na netipične tačke (engl. *outliers*), tj. rezultate koji su mnogo iznad ili ispod većine ostalih rezultata. Tehnike opisane u prethodnom odeljku mogu se upotrebiti i za otkrivanje netipičnih tačaka.

- Najpre pogledajmo **histogram**. Bacite pogled na repove raspodele. Ima li tu usamljenih tačaka ekstremnih vrednosti? Ako ima, to su potencijalne netipične tačke. Ako prema krajevima raspodele broj tačaka opada uz približno jednak nagib, verovatno nema razloga za zabrinutost.
- Drugo, pogledajte **pravougaoni dijagram** (engl. *boxplot*). Sve vrednosti koje SPSS smatra netipičnim tačkama, tu se prikazuju kao kružići označeni brojevima. (To su ID brojevi tih slučajeva.) SPSS smatra netipičnim sve tačke koje su od ivice pravougaonika udaljene više od 1,5 njegovih dužina. Ekstremne vrednosti (označene zvezdicom *) jesu one koji su od ivice pravougaonika udaljene više od 3 njegove dužine. U gornjem primeru nije bilo ekstremnih tačaka, ali jesu dve netipične tačke, čiji su ID brojevi 24 i 157. Kada nađete takve tačke, morate razmisliti šta ćete s njima.
- Treba proveriti da li je vrednost netipične tačke zaista tolika ili je u pitanju greška. Proverite tu vrednost i da li je unutar opsega mogućih vrednosti date promenljive. Ponekad se isplati pogledati u upitniku odnosno zapisu podataka da li je podatak pogrešno unet. Ako se radi o grešci, ispravite je i ponovo nacrtajte pravougaoni dijagram. Ukoliko nema nikakve greške, morate odlučiti šta ćete s tom netipičnom tačkom. Neki pisci knjiga o statistici savetuju da se sve ekstremne netipične tačke uklone iz datoteke s podacima. Drugi savetuju da se te vrednosti zamene manje ekstremnim, kako bi i one učestvovala u analizama, ali tako da ne mogu iskriviti statističke pokazatelje. Više o tome videti u 4. poglavlju knjige autorki Tabachnik i Fidell (2007).
- Informacije u tabeli **Descriptives** pokazuju koliki problem će predstavljati slučajevi s netipičnim tačkama. Pogledajte srednju vrednost izračunatu bez gornjih i donjih 5% slučajeva, tj. **5% Trimmed Mean**. Kada se ta i prava srednja vrednost mnogo razlikuju, te tačke treba još ispitati. U ovom primeru su te dve srednje vrednosti (26,73 i 26,64) veoma bliske. Zbog toga i činjenice da se te vrednosti ne razlikuju previše od ostatka raspodele, zadržaću ih u datoteci s podacima.
- Kada zatreba da izmenite ili uklonite neke podatke iz datoteke, pređite u **Data Editor** i sortirajte podatke po opadajućem redosledu (da biste pronašli slučajeve s najvišim vrednostima) ili rastućem redosledu (za slučajeve s veoma niskim vrednostima). Na taj način će slučajevi koje treba posebno razmotriti dospeti na vrh datoteke s podacima. Pređite u kolonu s vrednostima date promenljive i izmenite ili uklonite problematičnu vrednost. Sve što izmenite u datoteci s podacima morate beležiti u svoj podsetnik.

Dodatne vežbe

Poslovno okruženje

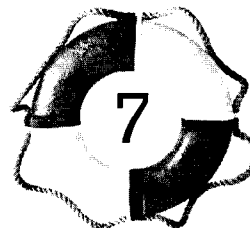
Datoteka s podacima: *staffsurvey3ED.sav*. Pojediniosti o datoteci videti u dodatku.

1. Sprovedite postupke objašnjene u ovom poglavlju da biste generisali *odgovarajuće* opisne statističke pokazatelje, pa zatim odgovorite na sledeća pitanja.
 - (a) Koliki procenat osoblja (engl. *staff*) te organizacije čine stalno zaposleni? (Upotrebite promenljivu *employstatus*.)
 - (b) Koliko je prosečno trajanje zaposlenja osoblja u toj organizaciji? (Upotrebite promenljivu *service*.)
 - (c) Koliki procenat ispitanika bi organizaciju preporučio kao dobro mesto za rad? (Upotrebite promenljivu *recommend*.)
2. Ocenite raspodelu rezultata na skali Total Staff Satisfaction (*totsatis*) za stalno zaposlene (engl. *permanent*) u odnosu na one koji su privremeno zaposleni (engl. *casual*) (promenljiva *employstatus*).
 - (a) Ima li na toj skali netipičnih tačaka zbog kojih bi trebalo da se zabrinete?
 - (b) Da li su rezultati obe grupe normalno raspodeljeni?

Zdravstvo

Datoteka s podacima: *sleep3ED.sav*. Pojediniosti o datoteci videti u dodatku.

1. Sprovedite postupke objašnjene u ovom poglavlju da biste generisali *odgovarajuće* opisne statističke pokazatelje, pa zatim odgovorite na sledeća pitanja.
 - (a) Koliki procenat ispitanika su žene (promenljiva *gender*)?
 - (b) Kolika je prosečna starost osoba u uzorku?
 - (c) Koliki procenat uzorka je naznačio da ima problema sa spavanjem (*problem*)?
 - (d) Kolika je medijana broja sati noćnog sna prosečnog radnog dana (*hourunit*)?
2. Ocenite raspodelu rezultata na skali Sleepiness and Associated Sensations (*totSAS*) za osobe koje smatraju/ne smatraju da imaju problema sa spavanjem (*problem*).
 - (a) Ima li na toj skali netipičnih tačaka zbog kojih bi trebalo da se zabrinete?
 - (b) Da li su rezultati obe grupe normalno raspodeljeni?



Upotreba dijagrama za opisivanje i analizu podataka

Numeričke vrednosti dobijene u poglavlju 6 daju korisne informacije o uzorku i promenljivama, ali se neki aspekti bolje istražuju vizuelno. SPSS ume da nacrtati više vrsta grafikona ili dijagrama (engl. *graphs, charts, plots*). U ovom poglavlju opisaću osnovne postupke za dobijanje sledećih dijagrama:

- histograma;
- stubičastih dijagrama (engl. *bar graphs*);
- linijskih dijagrama (engl. *line graphs*);
- dijagrama rasturanja (engl. *scatter plots*); i
- pravougaonih (kutijastih) dijagrama (engl. *box plots*).

U 15. verziji SPSS-a, meni **Graph** ima više opcija za crtanje dijagrama:

- Chart Builder
- Interactive
- Legacy Dialogs

U ovom poglavlju pokazaćemo kako radi opcija **Legacy Dialogs**, usaglašena s procedurama dostupnim u verzijama 12, 13 i 14 SPSS-a. Kada steknete iskustvo s dijagramima, trebalo bi da malo eksperimentišete s drugim pristupima, da se neko vreme igrate s raznim dijagramima i istražujete njihove mogućnosti. U ovom poglavlju dat je tek kratak pregled, toliko da vas uputi u osnove. Za prikazivanje raznih dijagrama upotrebljena je datoteka s podacima *survey3ED.sav*, objavljena na pratećoj Web lokaciji knjige (pojediniosti videti u str. xi i u dodatku). Ako želite da i sami na računaru pratite primere iz ovog poglavlja, pokrenite SPSS i otvorite datoteku *survey3ED.sav*.

Na kraju poglavlja data su i uputstva za doradu dijagrama da biste ih bolje prilagodili svojim potrebama. To će vam poslužiti kada zatreba da dijagram stavite u svoj istraživački rad. Detaljno je objašnjen i postupak uvoza dijagrama u Microsoftov Word.

Histogrami

Histogramima se prikazuje raspodela jedne neprekidne promenljive (npr. starosti ili subjektivno doživljenog stresa).

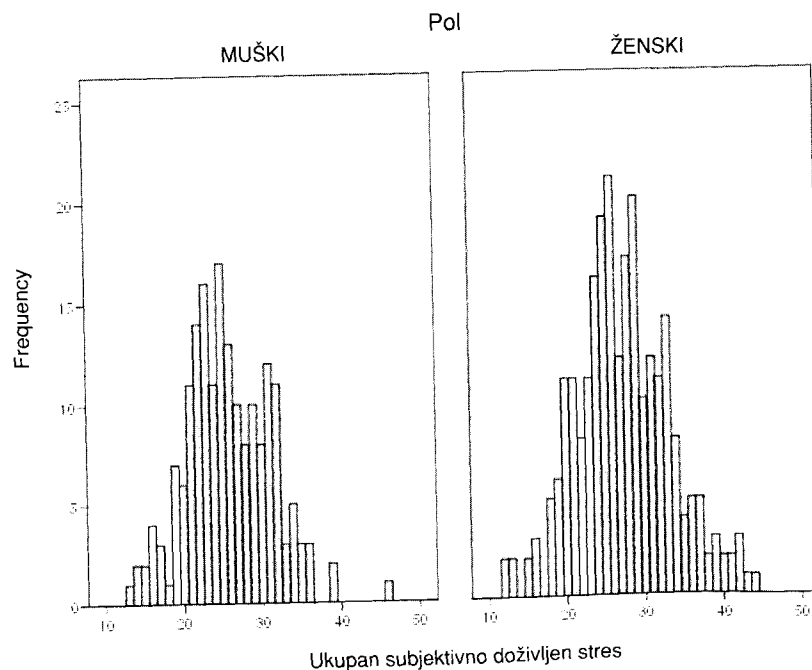
Postupak crtanja histograma

1. U glavnom meniju na vrhu prozora pritisnite **Graphs**, zatim **Legacy Dialogs**. Izaberite **Histogram**.
2. Izaberite promenljivu koja vas zanima i prebacite je u polje **Variables**. To treba da je neprekidna promenljiva (npr. ukupan subjektivno doživljen stres).
3. Ako želite zasebne histograme raznih grupa (npr. muškarci/žene), u odeljak **Panel by:** stavite odgovarajuću promenljivu (npr. pol). Izaberite **Rows** da biste te zasebne dijagrame stavili jedan iznad drugog, odnosno **Column** kada želite da ih stavite jedan pored drugog. Ja ću promenljivu pol staviti u polje **Column**.
4. Pritisnite **OK** (ili dugme **Paste** da biste komande snimili u **Syntax Editor**).

Navedeni postupak generiše ove komande:

```
GRAPH  
HISTOGRAM=usudstres  
/PANEL COLVAR=pol COLOP=CROSS
```

Rezultat je prikazan na sledećoj slici.



Tumačenje histograma

Oblik histograma pokazuje raspodelu vrednosti neprekidne promenljive. Mnoge statističke tehnike predstavljene u ovom priručniku zasnivaju se na pretpostavci da je raspodela vrednosti svih promenljivih normalna, tj. da ima oblik Gausove krive. U ovom poglavlju su vrednosti prilično normalno raspodeljene, pri čemu je najveći broj vrednosti u sredini, a prema krajevima krive sve ih je manje. Međutim, u društvenim naukama je sasvim uobičajeno da promenljive ne budu normalno raspodeljene. Raspodela može zadržati približno zvonast oblik, ali biti asimetrična ulevo ili udesno od srednje vrednosti, ili poprimiti pravougaoni oblik. O ocenjivanju normalnosti raspodele vrednosti promenljivih više se govori u poglavlju 6.

Stubičasti dijagrami

Stubičasti dijagrami mogu biti jednostavni ili veoma složeni, u zavisnosti od broja uključenih promenljivih. Stubičasti dijagram može prikazivati broj slučajeva u određenim kategorijama ili vrednost neprekidne promenljive za različite kategorije. U osnovi, potrebne su dve glavne promenljive: jedna kategorijska, a druga neprekidna. Njima se po želji može dodati još jedna kategorijska promenljiva.

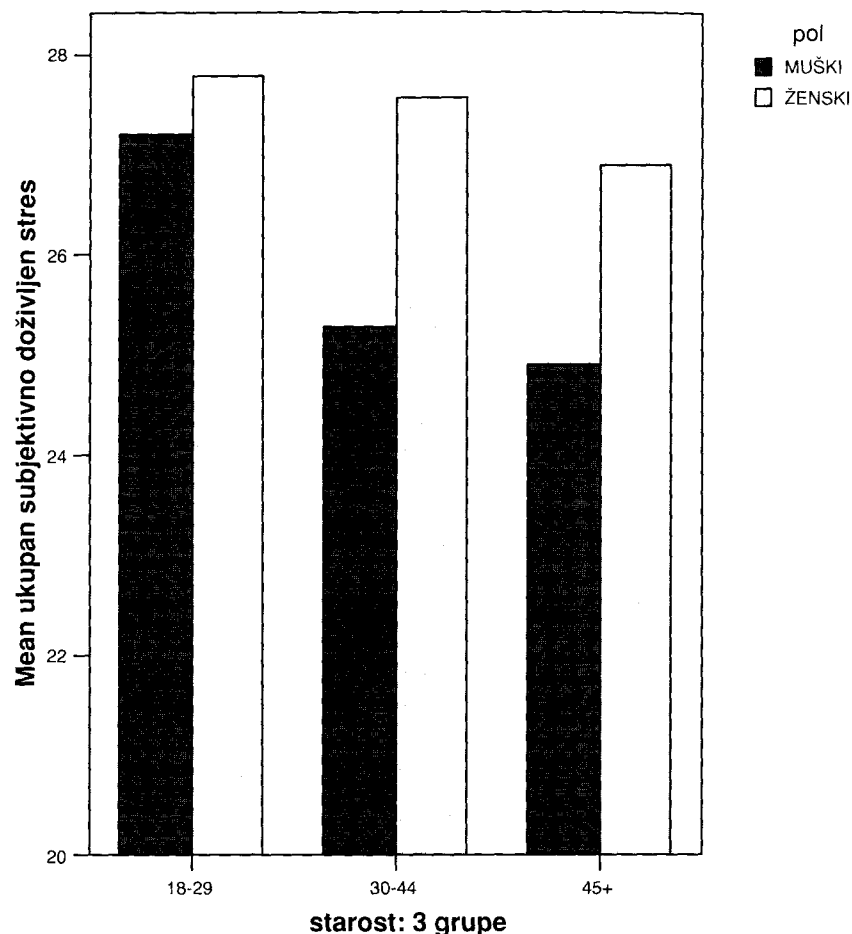
Postupak crtanja stubičastog dijagrama

1. U glavnom meniju na vrhu prozora pritisnite **Graphs**, zatim **Legacy Dialogs**. Izaberite **Bar**. Pritisnite **Clustered**.
2. U odeljku **Data in chart are** pritisnite **Summaries for groups of cases**. Pritisnite dugme **Define**.
3. U polju **Bars represent** izaberite **Other summary function**.
4. Pritisnite neprekidnu promenljivu koja vas zanima (npr. ukupan subjektivno doživljen stres). Ona bi trebalo da se pojavi u polju **Mean** (Ukupan subjektivno doživljen stres). To znači da će na skali Subjektivno doživljen stres biti prikazane srednje vrednosti raznih grupa.
5. Pritisnite prvu kategorijsku promenljivu (npr. star3grp). Pritiskom na dugme sa strelicom prebacite je u polje **Category axis**. Ta promenljiva se prikazuje na apscisi (horizontalnoj osi, X osi) stubičastog dijagrama.
6. Pritisnite drugu kategorijsku promenljivu i prebacite je u polje **Define Clusters by:**. Ona će biti navedena u legendi dijagrama.
7. Pritisnite **OK** (ili dugme **Paste** da biste komande snimili u **Syntax Editor**).

Prethodni postupak generiše ove komande:

```
GRAPH  
/BAR(GROUPED)= MEAN(usudstres) BY star3grp BY pol
```

Malo modifikovani rezultati prethodnog postupka prikazani su na slici.



Tumačenje stubičastih dijagrama

Rezultat prethodne procedure je sažet prikaz raspodele rezultata po izabranim grupama (u ovom slučaju, njih čine muškarci i žene iz različitih starosnih grupa). Gornji dijagram navodi na pomisao da žene pokazuju veći stepen subjektivno doživljenog stresa od muškaraca i da je ta razlika izraženija u starijim dobnim grupama. U grupi osoba starih između 18 i 29 godina, razlika između odgovarajućih vrednosti za muškarce i žene vrlo je mala.

Stubičaste dijagrame treba tumačiti pažljivo. Uvek pogledajte skalu Y (vertikalne) ose. Ono što vizuelno izgleda kao velika razlika, ponekad iznosi tek nekoliko jedinica merne skale i zato je verovatno malo važno. To je očigledno i na gornjem stubičastom dijagramu. Kada uzmete u obzir njegovu skalu, videćete da je razlika između grupa sasvim mala. Razlika između najmanje (muškarci starosti 45 godina i više) i najveće vrednosti (žene starosti od 18 do 29 godina) iznosi samo tri jedinice skale.

Da bi se ustanovio statistički značaj razlike između grupa vidljive na dijagramu, treba obaviti još neke statističke analize. U ovom slučaju, dvofaktorska (engl. *two-way*) analiza varijanse (uticaj dva činioca na rasturanje rezultata oko srednje vrednosti) između grupa (videti poglavlje 19) otkriva da li su te razlike statistički značajne.

Linijski dijagrami

Linijski dijagram prikazuje srednju vrednost jedne neprekidne promenljive za više različitih vrednosti neke kategorijske promenljive (npr. vreme 1, vreme 2, vreme 3). Linijski dijagrami su podesni i za grafičko prikazivanje rezultata jednofaktorske i dvofaktorske analize varijanse (tj. uticaja jednog odnosno dva činioca na rasturanje rezultata oko srednje vrednosti). Oni su neobavezan dodatak rezultatima analize varijanse (videti u poglavljima 18 i 19). U narednom postupku nacrtaćemo linijski dijagram za koji ćemo koristiti iste promenljive koje smo prethodno prikazali u stubičastom dijagramu.

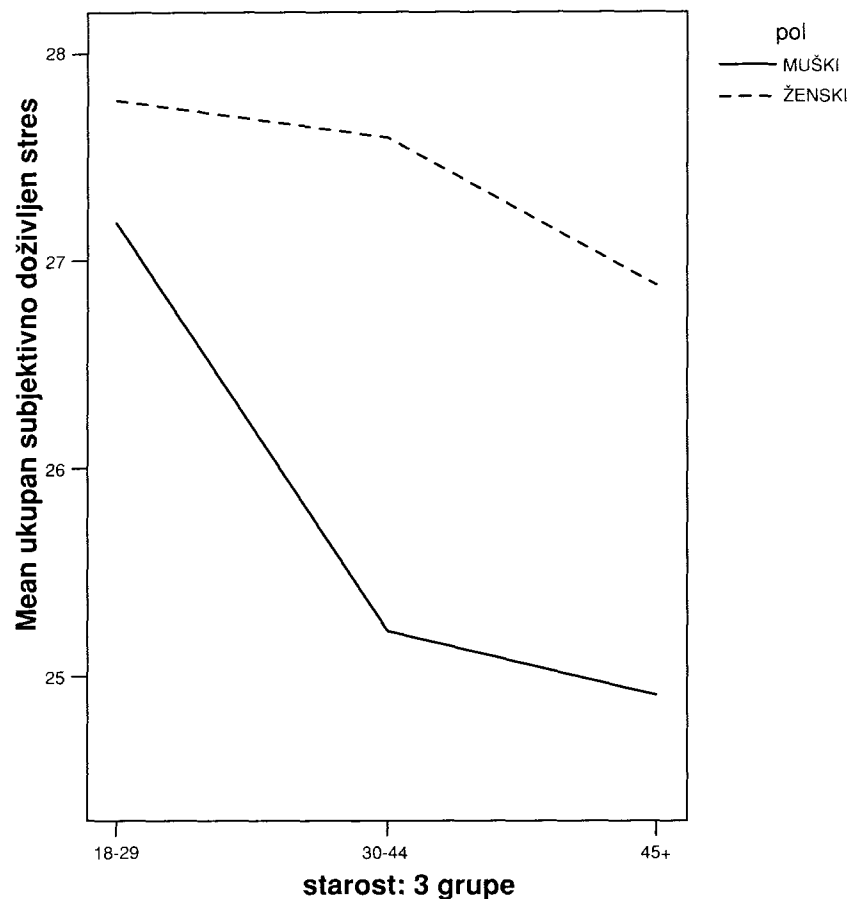
Postupak crtanja linijskog dijagrama

1. U glavnom meniju na vrhu prozora pritisnite **Graphs**, zatim **Legacy Dialogs**, pa **Line**.
2. Potvrdite polje **Multiple**. U odeljku **Data in chart are** pritisnite **Summaries for groups of cases**. Pritisnite dugme **Define**.
3. U polju **Lines represent** izaberite **Other summary function**. Pritisnite neprekidnu promenljivu koja vas zanima (npr. ukupan subjektivno doživljen stres) i zatim dugme sa strelicom. Tako ćete je prebaciti u polje **Mean** (ukupan subjektivno doživljen stres). To znači da će na skali subjektivno doživljenog stresa biti prikazane srednje vrednosti raznih grupa.
4. Pritisnite prvu kategorijsku promenljivu (npr. star3grp). Pritiskom na dugme sa strelicom prebacite je u polje **Category axis**. Ta promenljiva se prikazuje na horizontalnoj osi linijskog dijagrama.
5. Pritisnite drugu kategorijsku promenljivu i prebacite je u polje **Define Lines by:**. Ta promenljiva biće predstavljena u legendi dijagrama.
6. Ako dijagramu hoćete da dodate stubiće srazmerne odstupanjima, pritisnite dugme **Options**.
7. Pritisnite **OK** (ili dugme **Paste** da biste komande snimili u **Syntax Editor**).

Ovaj postupak će generisati sledeće komande:

```
GRAPH  
/LINE(MULTIPLE)MEAN(usudstres) BY star3grp BY pol
```

Malo modifikovani rezultati prethodnog postupka prikazani su na slici.

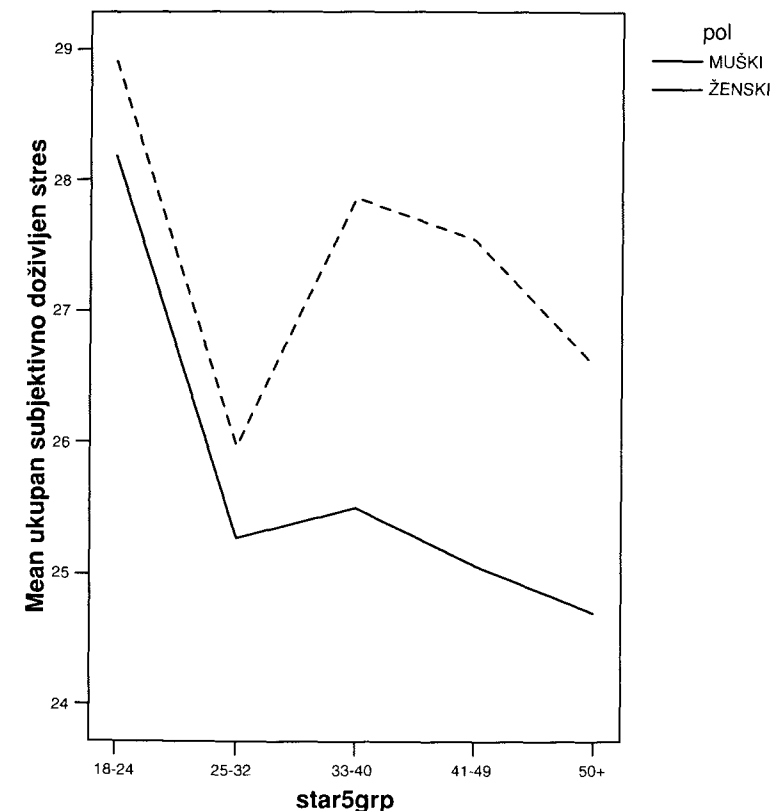


Tumačenje linijskih dijagrama

- Prvo pogledajte uticaj starosti na subjektivno doživljen stres za svaki pol. Mlađi muškarci kao da pokazuju veći stepen subjektivno doživljenog stresa od sredovečnih i od starijih muškaraca. Za žene, razlika između starosnih grupa nije toliko izražena. Starije žene su pod neznatno manjim stresom od mlađih.

- Mogli biste razmotriti i razliku između muškaraca i žena. Sveukupno, kao da muškarci subjektivno doživljavaju manje stresa od žena. Iako je razlika za mlađu grupu veoma mala, ona raste s povećanjem starosti ispitanika. Da li su te razlike statistički značajne može se utvrditi tek dvofaktorskom analizom varijanse (tj. uticaja dva činioca na rasturanje rezultata oko srednje vrednosti; videti poglavlje 19).

Prikazani rezultati navode na pomisao da uticaj starosti na subjektivno doživljen stres treba razmotriti u odnosu na pol ispitanika. U analizi varijanse, ta vrsta odnosa naziva se *uticaj interakcije* (engl. *interaction effect*). Iako linijski dijagram ne kazuje da li je taj odnos statistički značajan, on daje mnogo informacija i postavlja mnogo novih pitanja.



Pri tumačenju SPSS-ovih rezultata katkada treba razmotriti i neka druga pitanja. U ovom slučaju, rezultati navode na pomisao da bi vredelo dublje istražiti odnos starosti i subjektivno doživljenog stresa za obe grupe (muškarce i žene). Zato sam odlučila da uzorak podelim na pet starosnih grupa, umesto na tri kao na gornjem dijagramu. Time ću dobiti detaljniji uvid u uticaj starosti.

Nakon podele uzorka na pet jednakih grupa (pravljjenjem nove promenljive star5grp; uputstva za taj postupak data su u poglavlju 8), nacrtan je nov linijski dijagram. On nam daje jasniju sliku o uticaju starosti od prethodnog dijagrama sa samo tri starosne grupe.

Dijagram rasturanja

Dijagrami rasturanja se obično koriste za istraživanje odnosa između dve neprekidne promenljive (npr. starosti i samopoštovanja). Preporučljivo je dijagram rasturanja nacrtati *pre* računanja korelacije (videti poglavlje 11). Dijagram rasturanja predočava posmatraču da li je odnos promenljivih linearan ili krivolinijski. Za analizu korelacije prikladni su samo linearni odnosi.

Dijagram rasturanja pokazuje i da li su promenljive korelirane pozitivno (velikim vrednostima jedne promenljive odgovaraju veliki iznosi druge) ili negativno (velikim vrednostima jedne promenljive odgovaraju mali iznosi druge promenljive). Kod pozitivnih korelacija linija raste, tj. počinje levo s manjim vrednostima i raste udesno. Kod negativnih korelacija linija opada, tj. počinje levo s većim vrednostima i opada udesno.

Dijagram rasturanja grubo pokazuje i jačinu korelacije dve promenljive. Kada je korelacija slaba, tačke su bez reda rasute posvuda, tj. ne čine prepoznatljiv geometrijski oblik. Kada je korelacija jaka, tačke čine približno valjkast oblik, tj. vidi se gomilanje tačaka oko zamišljene prave linije.

U sledećem primeru nacrtaću dijagram rasturanja rezultata na dve skale u anketi: Ukupan subjektivno doživljen stres i Ukupna subjektivno doživljena samokontrola unutrašnjih stanja (SDSUS). ZADATA sam SPSS-u da dve grupe iz mog uzorka (muškarce i žene) na istom dijagramu rasturanja predstavi zasebno pomoću različitih oznaka (markera). Time dobijam informacije ne samo o uzorku u celini, nego i o raspodeli rezultata za muškarce i žene zasebno.

Kada želite zaseban dijagram rasturanja za svaku grupu, zadajte kategorijsku promenljivu u odeljku **Panel by:** umesto u dole prikazanom polju **Set Markers by**. Ukoliko vam treba dijagram rasturanja za ceo uzorak (ne po grupama), prosto zanemarite uputstva navedena u odeljku **Set Markers by**.

Postupak crtanja dijagrama rasturanja

1. U glavnom meniju na vrhu prozora pritisnite **Graphs**, zatim **Legacy Dialogs**, pa **Scatter/Dot**.
2. Potvrdite polje **Simple Scatter** i zatim pritisnite dugme **Define**.
3. Pritisnite prvu promenljivu koja vas zanima; to je obično ona koju smatrate zavisnom promenljivom (npr. ukupan subjektivno doživljen stres).
4. Zatim pritisnite dugme sa strelicom da biste izabranu promenljivu prebacili u polje **Y axis**. Ta promenljiva će biti prikazana na vertikalnoj osi.
5. Prebacite drugu promenljivu (npr. ukupnu SDSUS) u polje **X axis**. Ta promenljiva se prikazuje na horizontalnoj osi.

6. SPSS može da označi svaku tačku i po nekoj drugoj kategorijskoj promenljivoj. Prebacite tu promenljivu (npr. pol) u polje **Set Markers by:**. Time će tačke koje odgovaraju muškarcima odnosno ženama biti predstavljene različitim markerima (oznakama).

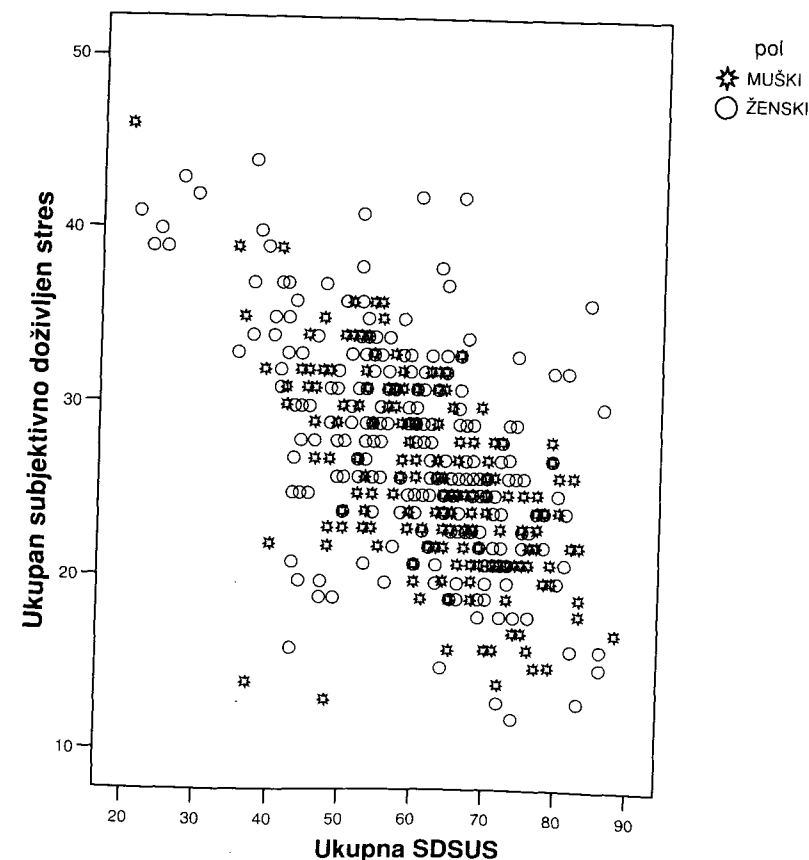
7. U polje **Label Cases by:** prebacite promenljivu ID. Tako ćete dobiti priliku da broj analiziranog slučaja očitete s dijagrama kada pronađete netipičnu tačku.

8. Pritisnite **OK** (ili dugme **Paste** da biste komande snimili u **Syntax Editor**).

Ovaj postupak će generisati sledeće komande:

```
GRAPH
/SCATTERPLOT(BIVAR)=usdsus WITH usdstres BY pol BY id
(IDENTIFY)
/MISSING=LISTWISE
```

Malo modifikovani rezultati prethodnog postupka prikazani su na slici.



Tumačenje dijagrama rasturanja

Iz prethodne slike kao da se nazire umerena negativna korelacija između dve posmatrane promenljive (subjektivno doživljenog stresa i SDSUS) u uzorku kao celini. Ispitanici s visokim nivoom subjektivno doživljene samokontrole (koja se prikazuje na horizontalnoj osi) subjektivno doživljavaju niži nivo stresa (što se prikazuje na vertikalnoj osi). S druge strane, osobe s nižim nivoom subjektivno doživljene samokontrole doživljavaju znatno viši nivo stresa. Ne nazire se nikakav krivolinijski odnos među njima, pa bi bilo umesno izračunati Pirsonovu korelaciju tih dvaju promenljivih (videti poglavlje 11).

Ne zaboravite da dijagram rasturanja ne daje konačne odgovore; nakon njega još treba izračunati odgovarajuće statističke pokazatelje (u ovom slučaju, koeficijent Pirsonove korelacije).

U gornjem primeru ispitala sam uzajamni odnos samo dve promenljive. Može se nacrtati i matrica dijagrama rasturanja za celu grupu promenljivih. To se koristi za preliminarno ispitivanje hipoteza (pretpostavki) u analizama kao sto je MANOVA.

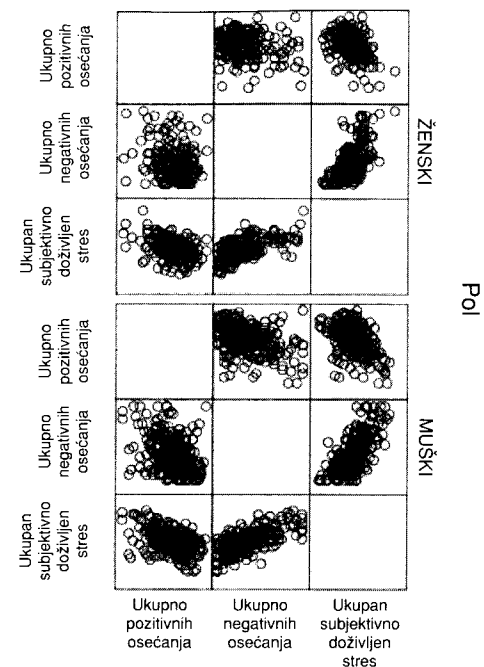
Postupak crtanja matrice dijagrama rasturanja

1. U glavnom meniju na vrhu prozora pritisnite **Graphs**, zatim **Legacy Dialogs** i najzad **Scatter/Dot**.
2. Potvrdite polje **Matrix Scatter**, pa pritisnite dugme **Define**.
3. Izaberite sve željene neprekidne promenljive (negose, pozose, sudstres) i prebacite ih u polje **Matrix Variables**.
4. Izaberite promenljivu pol i prebacite je u polje **Rows**.
5. Pritisnite dugme **Options** i potvrdite polje **Exclude cases variable by variable**.
6. Pritisnite **Continue** i zatim **OK** (ili dugme **Paste** da biste komande snimili u **Syntax Editor**).

Ovaj postupak će generisati sledeće komande:

```
GRAPH
/SCATTERPLOT(MATRIX)=negose upozose usudstres
/PANEL ROWVAR=pol ROWOP=CROSS
/MISSING=VARIABLEWISE .
```

Rezultati tog postupka prikazani su na slici.



Pravougaoni dijagram

Pravougaoni (kutijasti) dijagram se upotrebljava za poređenje raspodele vrednosti promenljivih. Može se upotrebiti za ispitivanje raspodele jedne neprekidne promenljive (npr. pozitivnih osećanja) ili se rezultati mogu razdeliti na više grupa (npr. starosnih). Za poređenje se može dodati još jedna kategorijska promenljiva (npr. muškarci i žene). U donjem primeru, istražiću raspodelu vrednosti na skali pozitivnih osećanja za muškarce i žene.

Postupak crtanja pravougaonog dijagrama

1. U glavnom meniju na vrhu prozora pritisnite **Graphs**, zatim **Legacy Dialogs** i najzad **Boxplot**.
2. Pritisnite **Simple**. U odeljku **Data in chart are** potvrdite polje **Summaries for groups of cases**. Pritisnite dugme **Define**.
3. Pritisnite neprekidnu promenljivu koja vas zanima (npr. ukupno pozitivnih osećanja). Pritiskom na dugme sa strelicom prebacite je u polje **Variable**.
4. Pritisnite kategorijsku promenljivu koja vas zanima (npr. pol). Pritiskom na dugme sa strelicom prebacite je u polje **Category axis**.

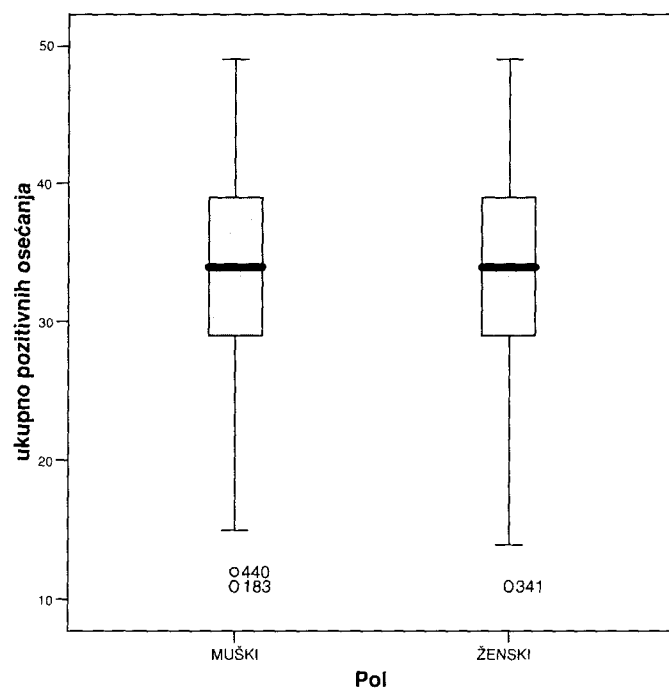
5. Pritisnite promenljivu ID prebacite je u polje **Label cases**. Tako ćete dobiti priliku da s dijagrama očitate brojeve slučajeva u kojima su se javile ekstremne vrednosti.

6. Pritisnite **OK** (ili dugme **Paste** da biste komande snimili u **Syntax Editor**).

Prethodni postupak generiše ove komande:

```
EXAMINE  
VARIABLES=upozos BY pol  
/PLOT=BOXPLOT/STATISTICS=NONE/NOTOTAL/ID=id
```

Evo rezultata tog postupka.



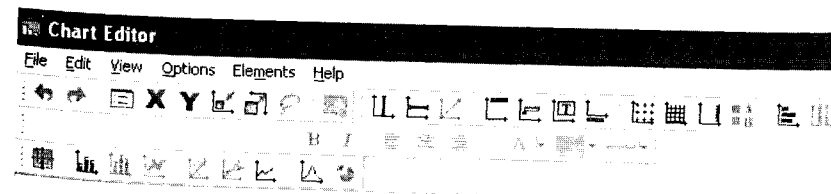
Tumačenje pravougaonog dijagrama

Pravougaoni dijagram daje mnogo informacija o raspodeli vrednosti neprekidne promenljive i mogućem uticaju druge, kategorijske promenljive (i klastera promenljive, tj. promenljive gomilanja, ako se ona upotrebi).

- Svaka raspodela vrednosti predstavljena je pravougaonikom i linijama (repovima) koje iz njega izlaze. Pravougaonik sadrži 50 procenata slučajeva, a njegova dužina je srazmerna interkvartilnom rasponu promenljive. Linija unutar pravougaonika je medijana. Repovi (linije koje izlaze iz pravougaonika) idu do najvećih i najmanjih opaženih vrednosti promenljive.
- Svi rezultati koje SPSS smatra netipičnim tačkama prikazani su kružićima pored kojih su brojevi. (To su ID brojevi tih slučajeva). Netipične tačke su slučajevi čije se vrednosti previše razlikuju od ostatka uzorka, bilo da su preveliki ili premali. SPSS smatra netipičnim sve tačke koje su od ivice pravougaonika udaljene više od 1,5 njegovih dužina. Ekstremne tačke (označene zvezdicom *) jesu one koje su od ivice pravougaonika udaljene više od 3 njegove dužine. Više o netipičnim tačkama pročitajte u poglavlju 6. U gornjem primeru prikazane su za muškarce i žene sveukupno tri netipične tačke s malim vrednostima promenljive Pozitivna osećanja.
- Osim informacija o netipičnim tačkama, pravougaoni dijagram predložava oblik raspodele rezultata raznih grupa. On pokazuje varijabilnost rezultata unutar svake grupe i omogućava vizuelni pregled razlika između grupa. U gornjem primeru, raspodele rezultata promenljive Pozitivna osećanja za muškarce i žene veoma su slične.

Dorada dijagrama

Pre nego što se dijagram štampa ili stavi u izveštaj, katkada treba doraditi njegove naslove, natpise, markere itd. Na primer, neke od dijagrama prikazanih u ovom poglavlju doradila sam da bi postali jasniji (recimo, podebljala sam linije kojima je dijagram iscrtan ili promenila popunu stubičastog dijagrama). Dijagrami se dorađuju u prozoru **Chart Editor**, koji se otvara tako što dvaput pritisnete dijagram koji treba doraditi. Prikazaće se nov prozor s tim dijagramom, dodatnim opcijama menija i ikonicama (slika 7.1). Trebalo bi da se pojavi i manji prozor **Properties**, koji služi za doradu dijagrama. Ako se ne pojavi, pritisnite dijagram desnim tasterom miša i u priručnom meniju izaberite **Properties Window**.



Slika 7.1 Traka s menijima **Chart Editor**ta.

U **Chart Editoru** mogu se sprovesti razne izmene i dorade:

- Reči u natpisima menjate tako što jednom pritisnete natpis da biste ga izabrali (oko teksta bi trebalo da se pojavi plava ivica). Pritisnite još jednom (trebalo da se pojavi crveni kursor) i tekst ćete moći da menjate. Izmenite ga i kada završite, pritisnite Enter na tastaturi.
- Položaj natpisa duž X i Y ose menjate (npr. da biste ga centrirali) tako što ga pritisnete *dvaput*. U okviru **Properties** koji će se pojaviti, pritisnite jezičak **Text**. U odeljku **Justify** izaberite željeni položaj (tačka označava centralni položaj, strelica ulevo pomera dijagram ulevo, strelica udesno ga pomera udesno).
- Da biste izmenili obeležja teksta, linije, markere, boje, popunu pozadine ili skale, *jednom* pritisnite ono svojstvo teksta koje hoćete da menjate. Zavisno od svojstva koje pritisnete, u prozoru **Properties** će se na odgovarajući način izmeniti opcije. Razne kartice u ovom okviru služe za izmenu raznih svojstava dijagrama. Želite li da menjate jednu od linija u višelinijском dijagramu (ili markere određene grupe), tu kategoriju treba da izaberete u legendi, a ne na samom dijagramu. To bi poslužilo npr. prilikom promene jedne od linija u isprekidanu, da bi bila uočljivija kada se odštampa crno-belo.

Najbolji način da naučite kako se te opcije upotrebljavaju jeste da ih koristite, tj. eksperimentišete s njima. Zato samo napred, slobodno se poigrajte!

Uvoz dijagrama u Wordove dokumente

SPSS omogućava neposredno kopiranje dijagrama u program za obradu teksta (npr. Microsoftov Word). To biste radili kada pripremate konačnu verziju izveštaja i deo rezultata želite da predstavite grafički. Ponekad dijagram predstavlja rezultate jednostavnije i jasnije nego brojevi u tabeli. Nemojte preterivati s tim; dijagrame upotrebljavajte samo kao specijalne efekte. Pre prenosa u Word, dijagram treba učiniti što je moguće jasnijim.

Napomena: Dolenavedena uputstva važe za verzije Worda koje rade pod Windowsom 95 ili novijim operativnim sistemima.

Postupak uvoza dijagrama u Wordov dokument

Windows omogućava istovremeno izvršavanje više programa. I SPSS i Word moraju biti otvoreni da bi razmena između njih bila moguća. Iz jednog programa u drugi prelazite tako što pritisnete odgovarajuću ikonicu na dnu ekrana ili pomoću menija **Window**. To je kao premeštanje papira po stolu.

1. Pokrenite Word i otvorite datoteku u koju treba staviti kopiju dijagrama. Pritisnite SPSS-ovu ikonicu na dnu ekrana da biste se vratili u SPSS.
2. Dovedite u prvi plan prozor **Viewer**.

3. Jednom pritisnite dijagram koji treba kopirati. Oko njega bi trebalo da se prikaže ivica.
4. Pritisnite **Edit** (u glavnom meniju na vrhu prozora) i zatim u tom meniju izaberite stavku **Copy Objects**. Tako ste dijagram kopirali na Clipboard (mada to ne vidite).
5. U listi minimiziranih programa na dnu ekrana pritisnite program za obradu teksta (npr. Microsoftov Word). Time ste ponovo aktivirali Word i doveli ga u prvi plan.
6. U Wordovom dokumentu dovedite kursor tamo gde želite da stavite dijagram.
7. Pritisnite **Edit** u Wordovom glavnom meniju, pa izaberite **Paste**. Ili na paleti sa alatkama pritisnite ikonicu **Paste** (liči na tablu sa štipaljkom za papir).
8. Pritisnite **File** i zatim **Save** da biste snimili Wordov dokument sa uvezenim dijagramom.
9. Ako hoćete da se vratite u SPSS i nastavite sa analizama, opet pritisnite njegovu ikonicu, koja bi trebalo da je na dnu ekrana. Kada su oba programa otvorena, možete prelaziti iz jednog u drugi i prenositi kopirane tabele, dijagrame itd. Nema potrebe da zatvarate ijedan od programa dok sve ne završite. Samo ne zaboravite da snimate datoteku dok to radite.

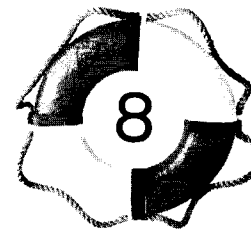
Dodatne vežbe

Poslovno okruženje

Datoteka s podacima: **staffsurvey3ED.sav**. Pojediniosti o datoteci videti u dodatku.

1. Nacrtajte pomoću SPSS-a **histogram** koji prikazuje raspodelu vrednosti na skali zadovoljstva osoblja (promenljiva *totsatis*).
2. Nacrtajte pomoću SPSS-a **stubičasti dijagram** koji prikazuje zadovoljstvo stalno zaposlenog u odnosu na osoblje zaposleno na određeno vreme manje ili jednako dve godine, 3 do 5 godina, i 6 i više godina. Upotrebite promenljive *totsatis*, *employstatus* i *servicegp3*.
3. Nacrtajte pomoću SPSS-a **dijagram rasturanja** koji prikazuje odnos broja godina zaposlenja osoblja u toj organizaciji i njihovog zadovoljstva. Najpre to pokušajte s promenljivom *service* (čija je raspodela jako asimetrična), a zatim s promenljivom pri dnu spiska promenljivih (*logservice*). Ta promenljiva je funkcijom \log_{10} transformisana prvobitna promenljiva *service*. \log_{10} bi trebalo da amortizuje veliku asimetričnost. Postupak je objašnjen u poglavlju 8.

4. Nacrtajte pomoću SPSS-a **pravougaoni (kutijasti) dijagram** koji prikazuje raspodelu vrednosti na skali zadovoljstva osoblja (*totsatis*) za razne starosne grupe (*starost*).
5. Nacrtajte pomoću SPSS-a **linijski dijagram** koji poredi zadovoljstvo osoblja raznih starosnih grupa (*agerecode*) za stalno zaposlene u odnosu na one koji su privremeno zaposleni.



Dorada podataka

Zdravstvo

Datoteka s podacima: **sleep3ED.sav**. Pojediniosti o datoteci videti u dodatku.

1. Nacrtajte pomoću SPSS-a **histogram** koji prikazuje raspodelu rezultata na Epvortovoj skali pospanosti (*ess*).
2. Nacrtajte pomoću SPSS-a **stubičasti dijagram** koji poredi rezultate na skali Sleepiness and Associated Sensations Scale (*totSAS*) za tri starosne grupe (*agegp3*) muškaraca i žena (*gender*).
3. Nacrtajte pomoću SPSS-a **dijagram rasturanja** koji prikazuje odnos rezultata na Epvortovoj skali pospanosti (Epworth Sleepiness Scale, *ess*) i na skali Sleepiness and Associated Sensations Scale (*totSAS*). Muškarce i žene (*gender*) obeležite različitim markerima.
4. Nacrtajte pomoću SPSS-a **pravougaoni (kutijasti) dijagram** koji prikazuje raspodelu rezultata na skali Sleepiness and Associated Sensations Scale (*totSAS*) za osobe koje su navele da imaju/nemaju problema sa spavanjem (*problem*).
5. Nacrtajte pomoću SPSS-a **linijski dijagram** koji poredi rezultate na skali Sleepiness and Associated Sensations Scale (*totSAS*) za razne starosne grupe (*agegp3*) muškaraca i žena (*gender*).

Nakon što unesete sve podatke i uklonite greške iz datoteke u kojoj su oni smešteni, sledeći korak je prevođenje sirovih podataka u oblik podesean za obavljanje analiza i ispitivanje hipoteza. U zavisnosti od podataka, ispitivanih obeležja (promenljivih) i vrste istraživačkih pitanja na koja tražite odgovore, taj proces bi mogao da obuhvati:

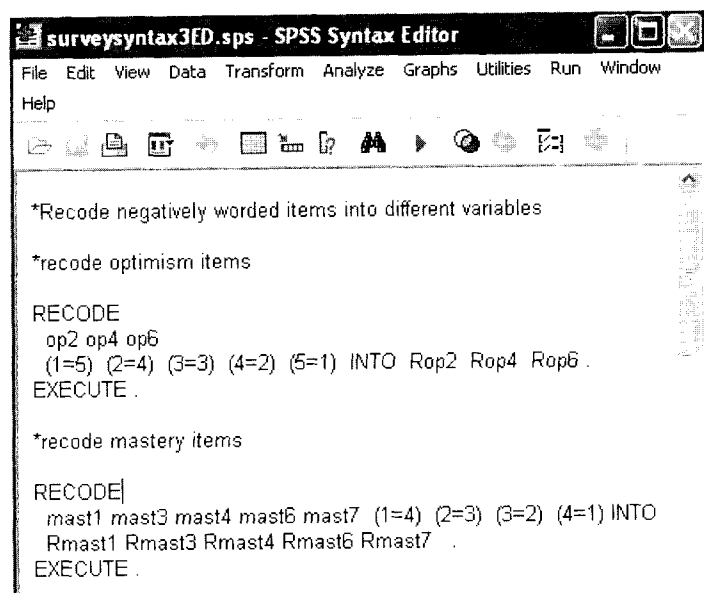
- sabiranje rezultata dobijenih za stavke (promenljive) od kojih se sastoji svaka merna skala, kako bi se dobio ukupan rezultat (engl. *total score*) na skalama na kojima se meri samopoštovanje, optimizam, subjektivno doživljen stres i sl. SPSS to radi brzo, lako i tačno; ni ne pomišljajte da to radite ručno za svakog ispitanika zasebno;
- transformisanje asimetričnih promenljivih radi analiza za koje je neophodno da rezultati budu normalno raspodeljeni;
- svođenje neprekidnih promenljivih (npr. starost) na kategorijske promenljive (npr. mladi, sredovečni i stari) radi analiza kao što je analiza varijanse; i
- smanjenje broja kategorija određene kategorijske promenljive (npr. svođenje bračnog statusa na samo dve kategorije, „jeste u vezi“/„nije u vezi“).

Kada u ovom poglavlju izmenite promenljive iz datoteke s podacima, zabeležite to u svoj podsetnik. Drugi način da pratite sve izmene u datoteci s podacima jeste korišćenje opcije Syntax, dostupne u svim SPSS procedurama. Najpre ću opisati taj proces, a zatim pokazati kako se rešifruju (engl. *recode*) i transformišu promenljive.

Beleženje SPSS procedura pomoću opcije Syntax

Kao što je već bilo objašnjeno u poglavlju 3, SPSS ima prozor **Syntax Editor** u kojem se mogu zabeležiti komande generisane pomoću menija u svakoj proceduri. Komandama pristupate prateći uputstva prikazana u narednim odeljcima u kojima su opisani postupci, ali umesto dugmeta OK na kraju pritisnite **Paste**. Otvoriće se nov prozor **Syntax Editor** sa svim komandama

koje ste prethodno zadali. Na slici 8.1 prikazan je deo komandnog prozora upotrebljenog za rešifrovanje stavki i izračunavanje ukupnih rezultata u datoteci **survey3ED.sav**. Cela komandna datoteka (**surveysyntax3ED.sps**) može se preuzeti s prateće Web lokacije ove knjige. Komande prenete u **Syntax Editor** ne izvršavaju se automatski, već treba vi da ih pokrenete. Da biste izvršili određenu komandu, izaberite je (uključujući i tačku kojom svaka komanda završava) i pritisnite **Run** u glavnom meniju ili strelicu (tj. trougao čije je teme okrenuto udesno) na paleti sa alatima. Komandnoj datoteci treba dodati komentare; to su redovi koji počinju zvezdicom (slika 8.1). Pre i posle svakog reda komentara treba ostaviti po jedan prazan red.



Slika 8.1 Prozor **Syntax Editor**.

Za svaku SPSS-ovu proceduru opisanu u nastavku, prikazaćemo i pripadajuću komandnu datoteku (sintaksu).

Izračunavanje ukupnih rezultata na skalama

Pre obavljanja statističkih analiza skupa podataka, treba izračunati ukupne rezultate na svim skalama. To se radi u dva koraka:

- **Korak 1:** obrnuti sve negativno formulisane stavke; i
- **Korak 2:** uputiti SPSS da sabere vrednosti svih stavki od kojih se podskala ili skala sastoji.

Treba razumeti skale i mere upotrebljene u istraživanju. Proverite (u uputstvu za skalu ili u članku u kome je skala bila objavljena) ima li na njoj stavki koje treba obrnuti i kako se izračunava ukupan rezultat na njoj. Neke skale se sastoje od više podskala koje treba (ili ih se ne sme) sabirati da bi se dobio ukupan rezultat. To se mora tačno uraditi, a mnogo je lakše uraditi ispravno prvi put nego kasnije ispravljati pa ponavljati analize.

Upozorenje Pre sprovođenja ovih postupaka, uvek napravite rezervnu kopiju datoteke s podacima.

Nakon što dodate nove slučajeve ili izmenite podatke, SPSS neće automatski novo izračunati rezultate. Naredne postupke sprovodite tek kada napravite konačnu verziju datoteke s podacima, kako zbog promene skupa podataka ne biste morali da ponavljate sve analize.

Korak 1: obrtanje negativno formulisanih stavki

Da bi se sprečio uticaj na ispitanike, na nekim skalama su određene stavke negativno formulisane. To se vidi na skali optimizma u anketi. Stavka 1 je pozitivno formulisana, pa velike vrednosti rezultata pokazuju *veliki* optimizam: „U nesigurnim vremenima obično očekujem da će mi se desiti ono najbolje“. Međutim, stavka 2 je negativno formulisana, što znači da velike vrednosti rezultata pokazuju *mali* optimizam: „Ako išta može da mi krene po zlu, krenuće.“ Negativno su formulisane i stavke 4 i 6. Pre izračunavanja ukupnog rezultata za ovu skalu, negativno formulisane stavke treba obrnuti tako da velike vrednosti svih stavki pokazuju veliki optimizam. Postupak obrtanja stavki 2, 4 i 6 skale optimizma prikazan je nastavku. Za skalu optimizma upotrebljena je Likert skala sa 5 podelaka; dakle, rezultati za sve stavke mogu biti u rasponu od 1 (nimalo se ne slažem) do 5 (potpuno se slažem).

Iako promenljive čije su formulacije obrnute mogu zadržati dotadašnja imena, tražićemo da SPSS od njih napravi nove promenljive, umesto da nove podatke upiše preko starih. To je mnogo bezbednije, a i ostavlja prvobitne podatke neizmenjenim.

Ukoliko želite da pratite i odmah sprovodite dolenađena uputstva, otvorite datoteku **survey3ED.sav**.

1. U glavnom meniju na vrhu prozora pritisnite **Transform**, zatim **Recode**, pa **Into Different Variables**.
2. Izaberite stavke koje treba obrnuti (op2, op4 i op6) i prebacite ih u polje **Input Variable—Output Variable**.
3. Pritisnite prvu promenljivu (op2) i zatim njeno novo ime upišite u odeljak **Output Variable** na desnoj strani prozora. U postojećoj datoteci s podacima, ja sam joj dala ime Rop2. Ako hoćete da joj date neko drugo novo ime umesto navedenoga, upišite ga (npr. revop2 ili obrop2). Ponovite to za sve ostale promenljive koje treba obrnuti (op4 i op6).

4. Pritisnite dugme **Old and new values**.

U polje **Value** u odeljku **Old value** upisite 1.

U polje **Value** u odeljku **New value** upisite 5. (Tako ste sve prvobitne vrednosti 1 pretvorili u 5.)

5. Pritisnite dugme **Add** da biste komandu $1 \leftrightarrow 5$ smestili u polje **Old > New**.

6. Ponovite isti postupak za ostale rezultate. Na primer:

U **Old value** upisite 2; u **New value** upisite 4; pritisnite **Add**.

U **Old value** upisite 3; u **New value** upisite 3; pritisnite **Add**.

U **Old value** upisite 4; u **New value** upisite 2; pritisnite **Add**.

U **Old value** upisite 5; u **New value** upisite 1; pritisnite **Add**.

7. Pritisnite **Continue** i zatim **OK** (ili dugme **Paste** da biste komandu snimili u **Syntax Editor**). Nakon prenošenja u **Syntax Editor**, izvršićete je kada je pritisnete i u meniju izaberete **Run**.

Prethodni postupak generiše sledeću sintaksu:

```
RECODE  
    op2 op4 op6  
    (1=5) (2=4) (3=3) (4=2) (5=1) INTO Rop2 Rop4 Rop6  
EXECUTE .
```

Nove promenljive sa obrnutim rezultatima trebalo bi da su na dnu datoteke s podacima. U prozoru **Data Editor** proverite da li je tako – pritisnite jezičak kartice **Variable View** i spustite se na dno spiska promenljivih. Videćete ceo niz promenljivih koje počinju slovom R; to su promenljive koje sam obrnula. Ukoliko ste sproveli opisani postupak, vaše promenljive bi trebalo da počinju sa „rev“ ili sa „obr“. Dobro je proveriti rešifrovane promenljive, videti promenu koju je rešifrovanje izazvalo. Za prvih nekoliko slučajeva u skupu podataka, treba zabeležiti rezultate dobijene s prvobitnim vrednostima promenljivih i potom pogledati da li su odgovarajuće preokrenute promenljive ispravno izmenjene.

Korak 2: izračunavanje ukupnih rezultata na skali

Nakon obrtanja negativno formulisanih stavki na skali, spremni ste za izračunavanje ukupnih rezultata na skali za svakog subjekta (ispitanika). To treba raditi tek kada imate potpunu datoteku s podacima.

Postupak izračunavanja ukupnih rezultata na skali

1. U glavnom meniju na vrhu prozora pritisnite **Transform**, zatim **Compute**.

2. U polje **Target variable** upišite novo ime promenljive „ukupan rezultat na skali“. Bilo bi dobro da svim ukupnim rezultatima date prefiks U (odnosno T ako su imena promenljivih na engleskom), jer ćete ih tako lakše pronalaziti u abecednim spiskovima promenljivih prilikom obavljanja analiza. Ja sam ukupnom rezultatu na skali optimizma dala ime „TOPTIM“, skraćeno od „total optimism“.

Važno: Ne sme vam se desiti da za novo ime upotrebite neko od već postojećih; time biste izgubili sve prvobitne podatke koji se odnose na promenljivu istog imena. Zato u šifarniku proverite da li je vaše „novo ime“ zaista novo.

3. Pritisnite dugme **Type and Label**. U polje **Label** upišite duže opisno ime skale (npr. ukupan optimizam). Pritisnite **Continue**.

4. U spisku promenljivih na levoj strani, pritisnite prvu stavku skale (op1).

5. Pritisnite dugme sa strelicom > da biste izabranu stavku prebacili u polje **Numeric Expression**.

6. Pritisnite + na kalkulatoru.

7. Ponavljajte prethodni postupak dok sve stavke skale ne prebacite u to polje. U ovom primeru najpre bismo odabrali neobrnute stavke (op3, op5), a zatim prethodnim postupkom dobijene obrnute stavke (Rop2, Rop4, Rop6), smeštene na dno spiska promenljivih.

8. Dvršen numerički izraz trebalo bi da glasi:

$op1 + op3 + op5 + Rop2 + Rop4 + Rop6$.

9. Ponovo proverite da li je sve tačno urađeno i da li su znakovi + na ispravnim mestima. Pritisnite **OK** (ili dugme **Paste** da biste komandu snimili u **Syntax Editor**). Da biste je izvršili nakon prenošenja u **Syntax Editor**, pritisnite je i u meniju izaberete **Run**.

Prethodni postupak generiše ovu sintaksu:

```
COMPUTE toptim = op1 + op3 + op5 + Rop2 + Rop4 + Rop6  
EXECUTE .
```

Na kraju skupa podataka napravili ste novu promenljivu TOPTIM. Ukupni rezultati za svakog ispitanika na ovoj skali sastoje se od zbira rezultata za svaku stavku (promenljivu) – od op1 do op6, rešifrovanih gde treba. Tamo gde pojedinoj stavki nedostaje podatak, neće biti ni ukupnog rezultata. To se vidi tako što u odgovarajućoj ćeliji datoteke s podacima imate tačku umesto rezultata. Videćete u literaturi da neki istraživači idu korak dalje tako što ukupan rezultat na skali dele ukupnim brojem stavki od kojih se skala sastoji. Tako olakšavaju tumačenje ukupnog rezultata na skali, zato

što je skala vraćena u prvobitni oblik upotrebljen za sve stavke (npr. od 1 do 5 za stavove od potpunog neslaganja do potpunog slaganja). I to se radi pomoću SPSS-ovog menija **Transform, Compute**. Novoj promenljivoj treba dati novo ime i upisati odgovarajuću formulu (npr. TOPTIM/6).

Ne zaboravite da pojediniosti o svim novim promenljivama upišete u šifarnik. Treba upisati ime nove promenljive, šta ona predstavlja i sve pojediniosti o tome kako je izračunata. Ako je neka od njenih stavki bila obrnuta, to treba navesti zajedno s detaljima o tome koje su stavke bile sabrane kako bi se dobio ukupan rezultat na skali. U šifarnik bi trebalo upisati i mogući raspon vrednosti nove promenljive (videti dodatak). Time ste sebi pripremili jasan vodič za prepoznavanje vrednosti koje su greškom izvan dozvoljenog raspona.

Kada napravite novu promenljivu, pokrenite za nju proceduru **Descriptives** kako biste proverili da li su njene vrednosti odgovarajuće (videti poglavlje 5). Tako ćete steći i osećaj za raspodelu rezultata nove promenljive.

- Proverite u upitniku koliki je mogući raspon vrednosti koje se mogu zabeležiti? Za skalu sa deset stavki i ako se odgovori označavaju brojevima od 1 do 4, najmanji mogući ukupan rezultat je 10, a najveći 40. Osoba koja na svaku stavku odgovori sa 1 imaće ukupan iznos $10 \times 1 = 10$. Osoba koja na svaku stavku odgovori sa 4 imaće ukupan iznos $10 \times 4 = 40$.
- Proverite rezultat procedure **Descriptives** da ne bi bilo vrednosti izvan dozvoljenog raspona. (Videti poglavlje 5.)
- Uporedite srednju (prosečnu) vrednost ukupnog rezultata dobijenog na skali sa odgovarajućim srednjim vrednostima navedenim u literaturi. Da li ste dobili rezultat sličan njihovom? Ako niste, zašto? Da niste pogrešili prilikom rešifrovanja promenljivih? Ili je vaš uzorak drugačiji od onoga upotrebljenog u drugim istraživanjima?

Trebalo bi da obavite i druge analize radi provere raspodele rezultata za novu promenljivu „ukupan rezultat na skali“:

- Proverite da li je raspodela rezultata asimetrična (*skewness*) te da li spoljoštenija (*kurtosis*) od normalne. (Videti poglavlje 6.)
- Napravite histogram rezultata i pogledajte njihov raspon. Da li su normalno raspodeljeni? Ako nisu, možda bi za neke analize rezultate trebalo da transformišete. (To će biti objašnjeno u nastavku.)

Transformisanje promenljivih

Često ćete (na svoju žalost!) prilikom pregleda raspodele rezultata na nekoj skali, tj. pri nekom merenju (npr. samopoštovanja, anksioznosti) primetiti da kriva raspodele nije ni lepa ni normalna. Ponekad su rezultati pozitivno asimetrični, tj. većina ispitanika je odgovorila malim vrednostima na skali

(npr. potištenosti). Katkada je raspodela negativno asimetrična; tu je većina rezultata bliža velikim vrednostima (npr. samopoštovanja). Pošto se mnogi parametarski testovi mogu obavljati samo na normalno raspodeljenim rezultatima, šta da radite sa asimetričnim raspodelama?

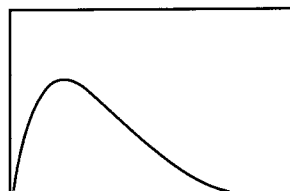
Jedno rešenje bi bilo da ne radite parametarske testove (npr. Pirsonovu korelaciju, analizu varijanse), nego njihove neparametarske alternative (Spirmanovu korelaciju ranga, Kruskal-Voliov test). U programskom paketu SPSS postoji više korisnih neparametarskih tehnika, a razmotrene su u poglavlju 16.

Druga mogućnost (kada nemate normalnu raspodelu rezultata) jeste da transformišete promenljive. To znači da raznim matematičkim funkcijama modifikujete rezultate dok raspodelu približno ne normalizujete. Ima više vrsta transformacija koje se upotrebljavaju ukoliko to prvobitni oblik raspodele nalaže. O ovom pristupu se vodi velika rasprava u literaturi. Jedni zdušno zagovaraju transformisanje promenljivih da bi se bolje zadovoljile polazne pretpostavke raznih parametarskih tehnika (najčešće normalnost raspodele), a drugi navode zašto to nije dobro. Diskusiju o pristupima transformaciji i pitanjima koje ona pokreće imate u 4. poglavlju knjige koju su napisali Tabachnick i Fidell (2007).

Na slici 8.2 predstavljene su neke od uobičajenih problematičnih raspodela i transformacije koje za njih preporučuju Tabachnick i Fidell (2007, str. 87). Uporedite svoju raspodelu s prikazanim i utvrdite kojoj od njih je najbližija. Uz svaku od predloženih transformacija dala sam i njenu formulu. Ne plašite se – to su samo formule koje SPSS upotrebljava za transformaciju. Vi ćete od SPSS-a dobiti gotovu, nadajmo se normalno raspodeljenu, novu promenljivu za analize. U narednom odeljku upoznaćete SPSS-ove procedure za transformisanje. Međutim, pre nego što sami pokušate da upotrebite neku transformaciju, treba dobro da proučite 4. poglavlje knjige autorki Tabachnick i Fidell (2007) ili odgovarajući deo nekog sličnog udžbenika.

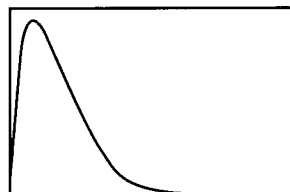
Postupak transformacije promenljivih

1. U glavnom meniju pri vrhu prozora pritisnite **Transform**, zatim **Compute Variable**.
2. **Target variable**. U ovo polje upišite novo ime promenljive. Pokušajte da njime naznačite vrstu transformacije i prvobitno ime promenljive. Na primer, da sam iz promenljive tnegaff morala da vadim kvadratni koren (engl. *square root*), nazvala bih novu promenljivu sqnegaff. Budite dosledni u primeni skraćenica kojima označavate transformacije.
3. **Functions**. U ovoj listi je mnoštvo funkcija za normalizaciju (transformaciju), od kojih treba da izaberete onu najprikladniju za svoju promenljivu. Pogledajte oblik svoje raspodele i uporedite ga sa slikom 8.2. Zapišite formulu navedenu pored slike koja je najpribližnija vašoj raspodeli. Nju ćete upotrebiti za transformaciju.



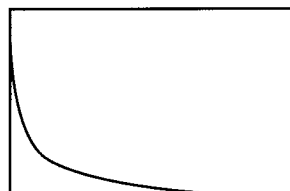
Kvadratni koren

Formula: nova promenljiva = SQRT (stara promenljiva)



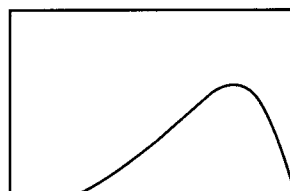
Logaritam

Formula: nova promenljiva = LG10 (stara promenljiva)



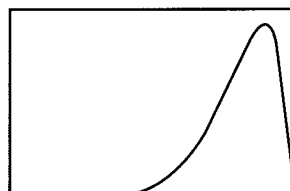
Recipročna vrednost

Formula: nova promenljiva = 1/ (stara promenljiva)



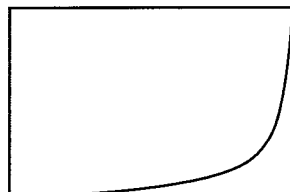
Preslikavanje i kvadratni koren

Formula: nova promenljiva = SQRT (K - stara promenljiva),
gde je K = najveća moguća vrednost + 1



Preslikavanje i logaritam

Formula: nova promenljiva = LG10 (K - stara promenljiva),
gde je K = najveća moguća vrednost + 1



Preslikavanje i recipročna vrednost

Formula: nova promenljiva = 1/ (K - stara promenljiva),
gde je K = najveća moguća vrednost + 1

Slika 8.2 Raspodela rezultata i preporučene transformacije.

4. Za transformacije u kojima se izračunava kvadratni koren ili logaritam.

U polju **Function group** pritisnite **Arithmetic**, pa premotavajte listu u donjem polju sve dok u njoj ne nađete potrebnu funkciju (npr. SQRT ili LG10). Pritisnite tu funkciju i potom strelicu naviše kako biste izabranu funkciju prebacili u polje **Numeric Expression**. Funkciji treba saopštiti koju promenljivu želite ponovo da izračunate. Pronađite je u listi promenljivih i pritisnite strelicu da biste je prebacili u polje **Numeric Expression**. Ko hoće, može sam da upiše formulu, a da uopšte ne koristi liste **Functions** i **Variables**. Jedino treba paziti da se ne pogreši prilikom pisanja formule.

5. Za transformacije u kojima se kriva raspodele preslikava.

Treba da odredite vrednost K za svoju promenljivu. To je najveća vrednost koju ta promenljiva može doseći (pogledajte u šifarniku) + 1. Upišite taj broj u polje **Numeric Expression**. Dovršite ostatak formule pomoću polja **Functions** ili ga sami dopišite.

6. Za transformacije u kojima se izračunava recipročna vrednost.

Recipročna vrednost se dobija deljenjem broja 1 sa rezultatom. Zato u polje **Numeric Expression** upisite 1, zatim kosu crtu i ime promenljive ili ostatak formule (npr. 1/ tsfest).

7. Proverite završenu formulu u polju **Numeric Expression**.

Zapišite je u šifarnik pored imena novonapravljene promenljive.

8. Pritisnite dugme **Type and Label**.

U polje **Label** upišite kratak opis nove promenljive (ili upotreblijenu formulu).

9. Proverite u polju **Target Variable** da li ste novoj promenljivoj dali *ново* ime (da nije zaostalo staro). Ako greškom upišete staro, izgubićete sve prvobitne rezultate. Zato uvek proverite dvaput.

10. Pritisnite **OK** (ili **Paste** ako tu komandu hoćete da prebacite u prozor **Syntax Editor**).

Da biste je izvršili nakon prenošenja u **Syntax Editor**, pritisnite je i u glavnom meniju izaberite **Run**. Biće napravljena nova promenljiva i dodata na kraj datoteke s podacima.

11. Pokrenite proceduru **Analyze, Frequencies** da biste uporedili asimetričnost i spljoštenost stare i nove promenljive. Da li su se poboljšali (približili nuli)?

12. Ponovo u meniju **Frequencies**, pritisnite dugme **Charts** i izaberite **Histogram** da biste pogledali raspodelu rezultata za novu promenljivu. Da li se raspodela popravila (približila normalnoj)? Ako nije, treba da nađete neku prikladniju transformaciju.

Ukoliko nijedna od ponuđenih transformacija ne normalizuje raspodelu dovoljno, razmislite o tome da za analizu podataka upotrebite neparametarske tehnike (videti poglavlje 16). Druga mogućnost za veoma asimetrične raspodele jeste da neprekidnu promenljivu podelite na više diskretnih grupa. Slede uputstva kako se to radi.

Postupak rešifrovanja kategorijske promenljive

1. U glavnom meniju na vrhu prozora pritisnite **Transform**, zatim **Recode**, pa **Into Different Variables**. (Obavezno treba da piše 'different variables', jer će samo tako prvobitna promenljiva ostati neizmenjena za druge analize.)
2. Izaberite promenljivu koju treba rešifrovati (npr. škola). U polje **Name** upišite ime nove promenljive koju treba napraviti (npr. skolaresif). U odeljak **Label** upišite njen duži opis. Pritisnite dugme **Change**.
3. Pritisnite dugme **Old and New Values**.
4. U odeljku **Old Value** nalazi se polje **Value**. Upišite prvu šifru ili vrednost tekuće promenljive (npr. 1). U odeljak **New Value** upišite novu vrednost koja će biti upotrebljavana (ili istu vrednost, ako će i dalje biti upotrebljavana). U ovom slučaju rešifrovaću na istu vrednost i zato upisati 1 u **Old Value** i u **New Value**. Pritisnite dugme **Add**.
5. Za drugu vrednost upisaću 2 u **Old Value**, ali 1 u **New Value**. Time ću sve vrednosti 1 i 2 iz prvobitnog šifrovanja rešifrovati u jednu grupu nove promenljive koja će biti napravljena, s vrednošću 1.
6. Za treću vrednost prvobitne promenljive upisaću 3 u **Old Value** i 2 u **New Value** kako bi vrednosti nove promenljive bile u neprekidnom nizu {1, 2, 3, 4, 5}. Pritisnite **Add**. Ponovite za sve preostale vrednosti. U tabeli **Old > New**, u ovom primeru bi trebalo da vidite sledeće šifre: 1→1; 2→1; 3→2; 4→3; 5→4; 6→5.
7. Pritisnite **Continue** i zatim **OK** (ili dugme **Paste** da biste komandu snimili u **Syntax Editor**). Da biste je izvršili nakon prenošenja u **Syntax Editor**, pritisnite je i u meniju izaberite **Run**.
8. Pređite u prozor **Data Editor** na karticu **Variable View**. Upišite odgovarajuće duže opise novih vrednosti (1=nezavršena srednja škola, 2=završena srednja škola, 3=dodatna obuka, 4=završene dodiplomske studije, 5=završene postdiplomske studije). Ne zaboravite da se ove šifre razlikuju od šifara prvobitne promenljive; ne smete ih pobrkati.

Prethodni postupak generiše ovu komandu:

```
RECODE
  skola
  (1=1) (2=1) (3=2) (4=3) (5=4) (6=5) INTO skolaresif .
EXECUTE.
```

Kada rešifrujete promenljivu, obavezno pokrenite proceduru **Frequencies** i za staru promenljivu (škola) i za novu (skolaresif), smeštenu na dno datoteke podataka. Proverite da li su frekvencije nove promenljive ispravne. Na primer, nova promenljiva skolaresif bi u prvoj grupi trebalo da ima $2+53=55$

slučajeva. To su dve osobe koje su označile 1 za prvobitnu promenljivu (osnovna škola) i 53 osobe koje su označile 2 (nezavršena srednja škola).

Prikazanu proceduru **Recode** (rešifrovanja) upotrebljavamo u razne svrhe. Kasnije, kada budemo radili statističke analize, doći ćete u situaciju da rešifrujete vrednosti određene promenljive. Na primer, u poglavlju 14 (Logistička regresija) rešifrovaćemo promenljive prvobitno šifrovane sa 1=da, 2=ne na nov sistem šifrovanja 1=da, 0=ne. To se radi onako kako je opisano u prethodnom odeljku. Važno je da pre nego što počnete rešifrovanje budete sasvim načisto šta predstavljaju prvobitne vrednosti, a šta treba da budu nove.

Dodatne vežbe

Poslovno okruženje

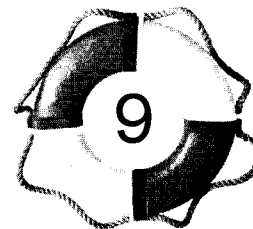
Datoteka s podacima: **staffsurvey3ED**. sav. Pojediniosti o datoteci videti u dodatku.

1. Vežbajte postupke opisane u ovom poglavlju da biste dobili ukupne rezultate na skali, koristeći stavke iz anketa o zadovoljstvu zaposlenih (**Staff Satisfaction Survey**). Saberite stavke koje mere slaganje sa svakom stavkom skale (tj. Q1a+Q2a+Q3a... do Q10a). Novu promenljivu nazovite *staffsatis*.
2. Pogledajte opisne statističke pokazatelje novog ukupnog rezultata (*staffsatis*) i uporedite ih sa istovrsnim pokazateljima za promenljivu *totsatis*, koju već imate u datoteci s podacima. To je ukupan rezultat koji sam već izračunala umesto vas.
3. Koliki su najmanji i najveći mogući rezultat za novu promenljivu? Uputstvo: treba uzeti u obzir broj stavki skale i broj kojim se šifruje odgovor za svaku stavku (videti u dodatku).
4. Pogledajte raspodelu promenljive service (staž) na histogramu. Videćete da je ta raspodela veoma asimetrična jer većina ljudi ima malo staža (manje od 2 godine), a jedan mali broj veoma mnogo (više od 30 godina). Uporedite tu raspodelu sa onima na slici 8.2 i isprobajte nekoliko transformacija. Ne zaboravite da svaki put pogledate raspodelu transformisanih promenljivih. Da li je ijedna od transformisanih promenljivih normalnije raspodeljena?
5. Podelite promenljivu koja se odnosi na broj godina radnog staža (service) u tri grupe postupkom **Visual Binning** iz menija **Transform**. Pritisnite dugme **Make Cutpoints** i izaberite **Equal Percentiles**. U odeljku **Number of Cutpoints** zadajte 2. Novu promenljivu nazovite *gp3service* kako bi se razlikovala od promenljive koju sam istim postupkom napravila u datoteci s podacima (*service3gp*). Pomoću procedure **Frequencies** prebrojte učestalosti nove promenljive da biste videli koliko je ispitivanih slučajeva u svakoj grupi.

Zdravstvo

Datoteka s podacima: *sleep3ED.sav*. Pojediniosti o datoteci videti u dodatku.

1. Vežbajte postupke opisane u ovom poglavlju tako što ćete izračunati ukupne rezultate na skali pospanosti i srodnih osećanja (Sleepiness and Associated Sensations Scale). Saberite stavke *fatigue*, *lethargy*, *tired*, *sleepy*, *energy*. Novu promenljivu nazovite *sleeptot*. Vodite računa o tome da nijednu od tih stavki ne treba obrnuti pre sabiranja.
2. Proverite opisne statističke pokazatelje za novi ukupan rezultat (*sleep-tot*) i uporedite ih sa istovrsnim pokazateljima za promenljivu *totSAS*, koju već imate u datoteci s podacima. To je ukupan rezultat koji sam već izračunala umesto vas.
3. Koliki su najmanji i najveći mogući rezultat za novu promenljivu? Uputstvo: treba uzeti u obzir broj stavki skale i broj kojim se šifruje odgovor za svaku stavku (videti u dodatku).
4. Na histogramu pogledajte raspodelu promenljive koja meri koliko cigareta dnevno popuše pušači u uzorku (*smokenum*). Videćete da je ta raspodela veoma asimetrična jer većina ljudi malo puši (manje od 10 cigareta dnevno), dok je mali broj osoba na samom kraju skale, s više od 70 cigareta dnevno. Uporedite tu raspodelu sa onima na slici 8.1 i isprobajte nekoliko transformacija. Ne zaboravite da pogledate raspodelu transformisanih promenljivih. Da li je ijedna od transformisanih promenljivih normalnije raspodeljena?
5. Podelite starosnu promenljivu (*age*) na tri grupe pomoću procedure **Visual Binning** iz menija **Transform**. Pritisnite dugme **Make Cutpoints** i izaberite **Equal Percentiles**. U odeljku **Number of Cutpoints** zadajte 2 za broj presečnih tačaka. Novu promenljivu nazovite *gp3age* kako bi se razlikovala od promenljive koju sam istim postupkom napravila u datoteci s podacima (*age3gp*). Prebrojte učestalosti nove promenljive pomoću procedure **Frequencies** da biste proverili koliko je ispitivanih slučajeva u svakoj grupi.



Provera pouzdanosti merne skale

Kada birate merne skale za svoje istraživanje, uzmite one pouzdane. Pouzdanost se može posmatrati s više aspekata, što je bilo razmotreno u poglavlju 1. Jedno od glavnih pitanja odnosi se na unutrašnju saglasnost skale, tj. stepen srodnosti stavki od kojih se skala sastoji. Da li sve one mere isti konstrukt? Među najčešće upotrebljavanim pokazateljima unutrašnje saglasnosti je Kronbahov koeficijent alfa. U idealnom slučaju bi Kronbahov koeficijent trebalo da bude veći od 0,7 (DeVellis, 2003). Međutim, vrednosti Kronbahovog koeficijenta alfa veoma su osetljive na broj stavki na skali. Kratke skale (tj. skale sa manje od deset stavki) često imaju prilično male Kronbahove koeficijente (npr. 0,5). U tom slučaju, možda je prikladnije da se izračuna srednja vrednost korelacije između stavki (engl. *mean inter-item correlation*). Briggs i Cheek (1986) preporučuju vrednosti od 0,2 do 0,4 kao optimalan raspon korelacije između stavki.

Pouzdanost skale se menja zavisno od uzorka za koji se koristi. Zato je pouzdanost svake skale neophodno proveriti na konkretnom uzorku. To se obično stavlja u odeljak Metodologija teze odnosno izveštaja (članka) o istraživanju. Ako skala sadrži negativno formulisane stavke (što je često u psihološkim istraživanjima), njih treba obrnuti *pre* provere pouzdanosti. Uputstva o tome su data u poglavlju 8.

Pre nego što nastavite, u priručniku za skalu (ili stručnom članku u kojem je skala bila opisana) obavezno pročitajte treba li obrtati stavke skale i sve što se odnosi na njene eventualne podskale. Katkada skale imaju više podskala koje treba ili ne treba kombinovati da se dobije ukupan rezultat na skali. Po potrebi ćete izračunati pouzdanost svih podskala i ukupne skale.

Ako za svoje istraživanje razvijate sopstvenu skalu, treba mnogo toga da pročitate o načelima i postupcima razvoja skala. O tome ima dobrih i čitljivih knjiga, među kojima su i one koje su napisali Streiner & Norman (2003), DeVellis (2003) i Kline (2005).

Objašnjenje primera

Pokazaću vam ovu tehniku na primeru datoteke s podacima **survey-3ED.sav**, dostupnoj na pratećoj Web lokaciji knjige. Sve pojedinosti istraživanja, upitnik i upotrebljene skale dati su u dodatku. Ukoliko želite da pratite i odmah sprovedite navedena uputstva, pokrenite SPSS i otvorite datoteku **survey3ED.sav**. U narednom postupku ispituje se unutrašnja saglasnost jedne od skala u tom upitniku. To je skala Satisfaction with Life (zadovoljstvo životom) (Pavot, Diener, Colvin & Sandvik, 1991), koja se sastoji od pet stavki. U datoteci s podacima one su nazvane **lifsat1**, **lifsat2**, **lifsat3**, **lifsat4**, **lifsat5**.

Postupak provere pouzdanosti skale

Važno: pre nego što počnete, trebalo bi da proverite da li su sve negativno formulisane stavke skale preokrenute (videti poglavlje 8). Ako to ne uradite, dobićete veoma niske (i netačne) vrednosti Kronbahovog koeficijenta alfa. U ovom konkretnom slučaju, ne treba obratiti nijednu stavku.

1. U glavnom meniju na vrhu prozora pritisnite **Analyze**, zatim **Scale** i najzad **Reliability Analysis**.
2. Pritisnite sve pojedinačne stavke od kojih se skala sastoji (npr. **lifsat1**, **lifsat2**, **lifsat3**, **lifsat4**, **lifsat5**). Prebacite ih u polje **Items**.
3. U odeljku **Model** treba da je izabrana opcija **Alpha**.
4. U polje **Scale label** upišite opisno ime skale odnosno podskale (Zadovoljstvo životom).
5. Pritisnite dugme **Statistics**. U odeljku **Descriptives for** redom pritisnite **Item**, **Scale** i **Scale if item deleted**. U odeljku **Inter-Item** pritisnite **Correlations**. U odeljku **Summaries** pritisnite **Correlations**.
6. Pritisnite **Continue** i zatim **OK** (ili **Paste** da biste komandu snimili u **Syntax Editor**).

Ovaj postupak generiše sledeću sintaksu:

RELIABILITY

```
/VARIABLES=lifsat1 lifsat2 lifsat3 lifsat4 lifsat5
/SCALE('Zadovoljstvo životom') ALL/MODEL=ALPHA
/STATISTICS=DESCRIPTIVE SCALE CORR
/SUMMARY=TOTAL CORR.
```

Scale: Life Satisfaction

Case Processing Summary

		N	%
Cases	Valid	436	99.3
	Excluded ^a	3	.7
	Total	439	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.890	.895	5

Item Statistics

	Mean	Std. Deviation	N
lifsat1	4.37	1.528	436
lifsat2	4.57	1.554	436
lifsat3	4.69	1.519	436
lifsat4	4.75	1.641	436
lifsat5	3.99	1.855	436

Inter-Item Correlation Matrix

	lifsat1	lifsat2	lifsat3	lifsat4	lifsat5
lifsat1	1.000	.763	.720	.573	.526
lifsat2	.763	1.000	.727	.606	.481
lifsat3	.720	.727	1.000	.721	.587
lifsat4	.573	.606	.721	1.000	.594
lifsat5	.526	.481	.587	.594	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.630	.481	.763	.282	1.587	.009	5

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
lifsat1	18.00	30.667	.758	.649	.861
lifsat2	17.81	30.496	.752	.654	.862
lifsat3	17.69	29.852	.824	.695	.847
lifsat4	17.63	29.954	.734	.574	.866
lifsat5	18.39	29.704	.627	.421	.896

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
22.38	45.827	6.770	5

Tumačenje rezultata provere pouzdanosti

- Proverite da li je tačan broj slučajeva (u tabeli **Case Processing Summary**) i broj stavki (u tabeli **Reliability Statistics**).
- Proverite ima li negativnih brojeva u tabeli **Inter-Item Correlation Matrix** (korelacije između stavki skale). Svi ti brojevi bi trebalo da budu pozitivni, što pokazuje da stavke mere isto obeležje. Prisustvo negativnih brojeva moglo bi da ukaže na to da neke stavke nisu bile ispravno preokrenute pre izračunavanja rezultata. Netačan rezultat bi se pokazao i u tabeli **Item-Total Statistics** tako što bi se pojavile negativne vrednosti za **Corrected-Item Total Correlation**. Sve njih treba pažljivo proveriti kada se dobije Kronbahov alfa manji od očekivanog. (Pogledajte šta su o toj skali pisali drugi istraživači.)
- Proverite koliki je **Cronbach's Alpha** (Kronbahov koeficijent alfa) naveden u tabeli **Reliability Statistics**. U ovom primeru on iznosi 0,89, što pokazuje veoma dobru pouzdanost i unutrašnju saglasnost skale za ovaj uzorak. Prihvatljive vrednosti alfe su iznad 0,7; međutim, poželjne su vrednosti preko 0,8.
- Brojevi **Corrected Item-Total Correlation** prikazani u tabeli **Item-Total Statistics** pokazuju stepen korelacije svake stavke sa ukupnim rezultatom. Mali brojevi (manji od 0,3) ovde pokazuju da stavka meri nešto drugo, a ne ono što meri cela skala. Ako cela skala ima premali Kronbahov alfa (npr. manji od 0,7), a vi ste već uklonili sve netačno izračunate rezultate, razmislite o tome da uklonite stavke s niskom korelacijom stavka – ukupan rezultat.
- U koloni **Alpha if Item Deleted** dat je uticaj uklanjanja svake stavke sa skale na iznos koeficijenta alfa. Uporedite te vrednosti s konačnom vrednošću alfe. Ako je ijedan broj u toj koloni veći od konačne vrednosti alfe, možda bi tu stavku trebalo izbaciti sa skale jer bi se time koeficijent alfa povećao. Za etabrirane, dobro validirane skale, o tome bi se mislilo samo ako je alfa mali (manji od 0,7). Međutim, uklanjanje stavki s postojeće skale znači da svoje rezultate nećete moći da uporedite sa ostalim istraživanjima u kojima je ta skala upotrebljena.
- Za skale s malim brojem stavki (npr. manjim od 10), ponekad je teško dobiti pristojnu vrednost Kronbahovog koeficijenta alfa. Tada biste u izveštaju mogli navesti srednju vrednost korelacije između stavki, navedenu u tabeli **Summary Item Statistics**. U ovom slučaju, srednja vrednost korelacije između stavki iznosi 0,63, a korelacije parova stavki od 0,48 do 0,76. To bi značilo da je korelacija (veza) među stavkama veoma snažna, što za mnogo skala ne važi.

Predstavljanje rezultata provere pouzdanosti skala

Uobičajeno je da se unutrašnja saglasnost skala upotrebljenih u istraživanju izračuna i navede u odeljku izveštaja Metodologija, pod naslovom Merila, Metrika ili Materijali. Nakon opisa skale (broj stavki, upotrebljena skala odgovora, istorija upotrebe), treba sažeto navesti šta su o njenoj pouzdanosti napisali projektant skale i drugi istraživači, a zatim u jednoj rečenici rezultate dobijene na sopstvenom uzorku.

Na primer:

Prema tvrdnji autora Pavot, Diener, Colvin i Sandvik (1991), skala *Satisfaction with Life* ima dobru unutrašnju saglasnost i Kronbahov koeficijent alfa od 0,85. U ovoj studiji je izračunati Kronbahov koeficijent alfa imao vrednost 0,89.

Dodatne vežbe

Poslovno okruženje

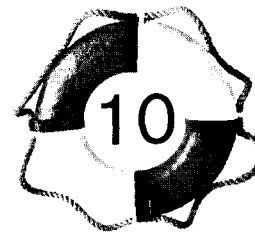
Datoteka s podacima: **staffsurvey3ED.sav**. Pojediniosti o datoteci videti u dodatku.

1. Proverite pouzdanost skale *Staff Satisfaction Survey*, koja se sastoji od stavki Q1a do Q10a za izražavanje različitog stepena slaganja. Nijednu stavku te skale nije potrebno obrnuti.

Zdravstvo

Datoteka s podacima: **sleep3ED.sav**. Pojediniosti o datoteci videti u dodatku.

1. Proverite pouzdanost skale *Sleepiness and Associated Sensations*, koja se sastoji od stavki *fatigue*, *lethargy*, *tired*, *sleepy*, *energy*. Nijednu stavku te skale nije potrebno obrnuti.



Izbor prikladnih statističkih tehnika

Za većinu studenata istraživačkih tehnika, jedan od najtežih (možda čak i zastrašujućih) delova istraživačkog procesa jeste pronalaženje (izbor) statističke tehnike prikladne za analizu datih podataka. Na većini statističkih kurseva uči se izračunavanje koeficijenta korelacije i kako se radi t-test, pa im obično ne ostaje dovoljno vremena da studente nauče da izaberu statistički pristup prikladan za pronalaženje odgovora na konkretna istraživačka pitanja. U većini istraživačkih projekata upotrebljavaju se različite statističke tehnike, u zavisnosti od pitanja na koja treba odgovoriti i prirode podataka koje treba analizirati. Zato je važno da steknete makar i elementarno poznavanje različitih statističkih tehnika, vrsta pitanja na koja one mogu odgovoriti, njihovih zahteva i pretpostavki na kojima počivaju.

Zato iskopajte svoje udžbenike statistike i pregledajte osnovne tehnike i načela na kojima one počivaju. Isto tako, trebalo bi da prelistate članke u časopisima o vašoj temi i identifikujete statističke tehnike upotrebljene u tim studijama. Raznim oblastima statistički se pristupa na različite načine, pa je važno da utvrdite kako su drugi istraživači analizirali podatke. Tražite dugačke i detaljne članke u časopisima u kojima jasno i jednostavno piše koje statističke tehnike su korišćene. Sakupite takve članke i spremite ih u zasebnu fasciklu radi lakšeg korišćenja. Dobro će vam doći i kasnije, kada budete razmatrali kako da predstavite rezultate svojih analiza.

U ovom poglavlju razmotrićemo razne dostupne statističke tehnike i korak po korak proći kroz proces izbora. Ukoliko vas već i sama reč statistika baca u paniku, smatrajte sve ovo izborom recepta po kome ćete večeras pripremiti jelo. Šta imate u frižideru, šta vam se jede (supa, pečenje, nešto prženo, kuvano jelo) i koji je postupak? Statističkim jezikom rečeno, razmotrićemo vrste istraživačkih pitanja koja postoje, promenljive (obeležja, karakteristike) koje treba analizirati i prirodu samih podataka. Prođite kroz ovaj proces korak po korak i videćete da je konačna odluka često iznenađujuće jednostavna. Kada utvrdite šta imate i šta želite da uradite, često preostaje samo jedan način da se to postigne. Najvažniji deo ovog procesa je jasno napisati šta imate i šta s tim treba da uradite.

Pregled raznih statističkih tehnika

Ovaj odeljak ima dva osnovna dela. Prvo ćemo razmotriti tehnike istraživanja *veza između promenljivih* (npr. starosti i optimizma), a potom načine istraživanja *razlika između grupa* (npr. polnih razlika u stavovima o optimizmu). Tehnike sam ovako podelila zato što je tako ustrojena većina udžbenika iz statistike, a i većinu studenata su tako učili osnove statistike. Time se pomalo veštački ističe razlika između ta dva skupa tehnika. U stvari, između raznih statističkih tehnika ima mnogo sličnosti, što se na prvi pogled ne vidi. Celovito razmatranje te teme nije predmet ove knjige. Ko o tome želi da sazna više, preporučujem da najpre pročita 17. poglavlje knjige koju su napisale Tabachnick i Fidell (2007). Tu je dat prikaz opšteg linearnog modela, pod koji se mogu svrstati mnoge statističke tehnike.

Razne statističke tehnike namerno sam opisala sažeto i jednostavno da bi ih početnici lakše razumeli. U ovom poglavlju nisu čak ni nabrojane sve dostupne tehnike, ali su date osnove koje su dovoljne da čovek počne da ih upotrebljava i tako stekne samopouzdanje.

Istraživanje veza između raznih obeležja

U anketnom istraživanju često nisu važne razlike između grupa, nego jačina veze između obeležja (promenljivih). Može se upotrebiti više tehnika.

Korelacija

Za istraživanje jačine veze između dve neprekidne promenljive upotrebljavaju se Pirsonova i Spirmanova korelacija. Korelacija pokazuje smer (pozitivan ili negativan) i jačinu veze. Pozitivna korelacija pokazuje da obe promenljive zajedno i opadaju i rastu. Negativna korelacija pokazuje da jedna promenljiva opada kada druga raste i obrnuto. To je tema 11. poglavlja.

Delimična korelacija

Delimična (parcijalna) korelacija je proširenje Pirsonove korelacije. Pomoću nje isključujemo uticaj treće, remetilačke promenljive. Delimična korelacija uklanja uticaj remetilačke promenljive (npr. svesno davanje netačnih, ali društveno poželjnih odgovora), čime je omogućeno dobijanje tačnije slike veze između dve promenljive od interesa. Delimična korelacija je tema 12. poglavlja.

Višestruka regresija

Višestruka regresija je sofisticiranije proširenje korelacije, kojim se izračunava mogućnost da se pomoću skupa nezavisnih promenljivih predvidi vrednost jednog *neprekidnog* zavisnog obeležja. Razne vrste višestruke regresije služe za poređenje prediktivne mogućnosti (predviđanja) određenih nezavisnih promenljivih i pronalaženje najboljeg skupa promenljivih za predikciju jedne zavisne promenljive. Videti poglavlje 13.

Faktorska analiza

Faktorska analiza služi za svođenje velikog skupa promenljivih ili stavki skale na manji broj dimenzija ili faktora, s kojima je lakše raditi. To se postiže sažimanjem oblika korelacije koji leže u njihovoj osnovi i pronalaženjem grupa tesno povezanih stavki. Ova tehnika se često koristi prilikom razvoja skala i merila, za identifikaciju pripadne strukture. Videti poglavlje 15.

Sažetak

Sve navedene analize obuhvataju istraživanje veza između neprekidnih promenljivih. Kada imate samo kategorijske promenljive, za ispitivanje njihove uzajamne veze može se upotrebiti hi-kvadrat test veza i nezavisnosti (npr. za odgovor na pitanje da li pol klijenata utiče na stopu njihovog odustajanja od određenog programa tretmana). U toj situaciji zanima vas broj osoba u svakoj kategoriji (broj muškaraca i žena koji odustaju od tog programa ili ga završavaju), a ne njihove prosečne vrednosti na nekoj skali posmatranog obeležja. Spomenuću još neke tehnike o kojima treba barem znati da postoje. Više o njima naći ćete u Tabachnick i Fidell (2007). To su:

- *Diskriminaciona analiza* (engl. *discriminant function analysis*) služi za ispitivanje mogućnosti da se pomoću skupa nezavisnih promenljivih predvidi vrednost jednog *kategorijskog* zavisnog obeležja, tj. da se odredi koje promenljive najbolje predviđaju pripadnost grupi. (Diskriminaciona funkcija je linearna kombinacija nezavisnih promenljivih koja najbolje razdvaja slučajeve na *a priori* definisane grupe.) U ovom slučaju, zavisna promenljiva je obično neki jasan kriterijum (položio/pao, prekinuo/nastavio tretman). Videti poglavlje 9 u knjizi Tabachnicka i Fidella (2007).
- *Kanonska korelacija* (engl. *canonical correlation*) služi za analizu uzajamnih veza dva *skupa* promenljivih. Na primer, moglo bi se istražiti kako razne demografske promenljive utiču na merila opšteg raspoloženja i sposobnosti prilagođenja. Videti poglavlje 12, Tabachnick i Fidell (2007).
- *Strukturno modelovanje* (engl. *structural equation modelling*) relativno je nova i veoma sofisticirana tehnika za ispitivanje raznih modela međuveza u skupu promenljivih. Zasnovana je na višestrukoj regresiji i tehnikama faktorske analize. Služi za izračunavanje važnosti svake nezavisne promenljive u modelu i testiranje koliko dobro ceo model odgovara podacima, kao i za poređenje alternativnih modela. Sam SPSS nema modul za strukturno modelovanje, ali podržava odgovarajući dodatni program AMOS. Videti poglavlje 14, Tabachnick i Fidell (2007).

Ispitivanje razlika između grupa

Postoji još jedna porodica statističkih tehnika za utvrđivanje statistički značajnih razlika između grupa. U nastavku ćemo prikazati parametarske verzije tih testova prikladne za podatke na intervalnim skalama s normalnom raspodelom rezultata i njihove neparametarske alternative.

T-testovi

T-testovi se upotrebljavaju kada imate *dve* grupe (recimo, muškarce i žene) ili dva skupa podataka (pre i posle), i želite da uporedite srednje vrednosti nekog neprekidnog obeležja (promenljive). Postoje dve glavne vrste t-testova. T-testove uparenih uzoraka (ili ponovljenih merenja, engl. *repeated measures*) upotrebljavate kada vas zanimaju promene vrednosti posmatranog obeležja dobijene od subjekata testiranih u Vreme 1 i zatim ponovo u Vreme 2 (obično posle neke intervencije ili događaja). Ti uzorci su povezani pošto se radi o *istim* ljudima testiranim u dva navrata. T-testovi nezavisnih uzoraka upotrebljavaju se kada imate dve *različite* (nezavisne) grupe ljudi (muškarce i žene) i želite da uporedite njihove rezultate za posmatrano obeležje. U tom slučaju informacije prikupljate samo jednom, ali od dve grupe ljudi. T-testovi su obrađeni u poglavlju 17. Njihove neparametarske alternative, Man-Vitnijev U test i Viloksonov test ranga, predstavljeni su u poglavlju 16.

Jednofaktorska analiza varijanse

Jednofaktorska analiza varijanse (engl. *one-way ANOVA*) slična je t-testu, ali se koristi kada imate *dve ili više grupa* i želite da uporedite njihove srednje vrednosti za jednu neprekidnu promenljivu (obeležje). Jednofaktorska znači da se istražuje uticaj samo jedne nezavisne promenljive na zavisnu. ANOVA kazuje da li se grupe razlikuju, ali ne kazuje gde je razlika značajna (gp1/gp3, gp2/gp3 itd.). Naknadnim poređenjem može se utvrditi koje grupe se međusobno značajno razlikuju. Umesto da se porede sve grupe, mogu se ispitati i razlike između određenih grupa; to su planirana poređenja. Slično t-testovima, postoje dve vrste jednofaktorske analize varijanse: ANOVA ponovljenih merenja (kada se isti ljudi ispituju u više od dva navrata) i ANOVA različitih grupa (ili nezavisnih uzoraka), kada se porede srednje vrednosti posmatranog obeležja u dve ili više grupa. Jednofaktorska ANOVA je obrađena u poglavlju 18, dok su njene neparametarske alternative (Kruskal-Volisonov test i Fridmanov test) predstavljene u poglavlju 16.

Dvofaktorska analiza varijanse

Dvofaktorska analiza varijanse (engl. *two-way ANOVA*) služi za ispitivanje uticaja dve nezavisne promenljive na jednu zavisnu. Prednost dvofaktorske analize varijanse je to što omogućava ispitivanje jačine interakcije, tj. uticaja druge nezavisne promenljive na dejstvo prve; na primer, kada posumnjate da se optimizam povećava s godinama, ali samo kod muškaraca. Ona meri i osnovne, zasebne uticaje, tj. celokupan uticaj svake nezavisne promenljive

(npr. pola, starosti). Postoje dve vrste dvofaktorske analize varijanse: ANOVA različitih grupa (engl. *between-groups ANOVA*), kada se ispituju grupe koje se međusobno razlikuju, i ANOVA ponovljenih merenja (engl. *repeated measures ANOVA*), kada se isti ljudi ispituju u više navrata. Neka istraživanja su projektovana tako da u istoj studiji kombinuju analize varijanse različitih grupa i ponovljenih merenja. To se onda na engleskom naziva Mixed Between-Within Designs ili Split Plot, tj. kombinovana ANOVA. Dvofaktorska ANOVA obrađena je u poglavlju 19, a kombinovana ANOVA u poglavlju 20.

Multivarijaciona analiza varijanse

Multivarijaciona analiza varijanse (engl. *multivariate analysis of variance*, MANOVA) služi za poređenje srednje vrednosti posmatranog obeležja grupa u više različitih, ali *povezanih*, zavisnih promenljivih; na primer, poredite uticaj različitih tretmana na razne merljive ishode (npr. anksioznost, depresiju, fizičke simptome). Multivarijaciona ANOVA može biti urađena uz jednofaktorske, dvofaktorske ili višefaktorske analize varijanse sa jednom, dve ili više nezavisnih promenljivih. MANOVA je obrađena u poglavlju 21.

Analiza kovarijanse

Analiza kovarijanse (ANCOVA) služi za statističku kontrolu mogućih uticaja dodatne, remetilačke (engl. *confounding*) promenljive (engl. *covariate*). Ovo je korisno kada posumnjate da se vaše grupe razlikuju po nekom obeležju koje utiče na dejstvo nezavisnih promenljivih na zavisnu. Kako biste bili sigurni da uticaj potiče od nezavisne promenljive, ANCOVA statistički uklanja dejstvo remetilačke promenljive. Analiza kovarijanse može se obaviti kao deo jednofaktorske, dvofaktorske ili multivarijacione analize varijanse. ANCOVA je obrađena u poglavlju 22.

Proces donošenja odluka

Pošto ste videli šta vam stoji na raspolaganju, vreme je da izaberete tehnike koje odgovaraju vašim potrebama. Prilikom izbora odgovarajuće statističke analize, treba uzeti u obzir više činilaca. To su vrsta pitanja na koja tražite odgovore, vrsta stavki i merne skale u vašem upitniku, priroda podataka dostupnih za svaku promenljivu i pretpostavke koje moraju biti zadovoljene za svaku statističku tehniku. Proći ćemo korak po korak kroz proces odlučivanja.

Korak 1: na koja pitanja tražite odgovore?

Napišite spisak sa svim pitanjima na koja bi istraživanje trebalo da odgovori. Videćete da se neka pitanja mogu postaviti na različite načine. U svakoj oblasti od interesa, pitanje pokušajte da postavite na više načina. Te alternative ćete upotrebiti kada budete razmatrali razne statističke pristupe koje biste mogli primeniti. Na primer, zanima vas uticaj starosti na optimizam. To pitanje se može postaviti na više načina:

- Postoji li veza između starosti i nivoa optimizma?
- Da li su starije osobe optimističnije od mlađih?

Ova dva pitanja se razlikuju i za dobijanje odgovora na njih potrebne su različite statističke tehnike. Od prirode prikupljenih podataka zavisi koje pitanje ćemo proglasiti za prikladnije. Zato za svaku oblast od interesa postavite više pitanja.

Korak 2: pronađite stavke i skale koje ćete upotrebiti za traženje odgovora na ta pitanja

Vrsta stavki i skala u upitniku i studiji igra veliku ulogu pri izboru statističkih tehnika koje su prikladne za traženje odgovora na istraživačka pitanja. Zato je prilikom projektovanja istraživanja toliko važno imati u vidu predviđene analize. Na primer, način prikupljanja informacija o starosti ispitanika (videti primer u 1. koraku) odrediće koje su statističke analize dostupne. Ako od ispitanika zatražite da izaberu jednu od dve opcije (ispod 35 godina/preko 35 godina), izbor analiza biće vrlo ograničen, zato što promenljiva starost može imati samo dve vrednosti. S druge strane, ukoliko od ispitanika zatražite da svoju starost navedu u godinama, izbor će biti širi zato što promenljiva može poprimiti vrednosti u širokom opsegu od 18 do 80 i više. U toj situaciji, mogli biste za neke analize (kao što je ANOVA) svesti raspon vrednosti na manji broj kategorija, a za druge analize (npr. korelaciju) zadržati ceo opseg vrednosti.

Ako ste za svoje istraživanje razdelili upitnik ili anketu, vratite se na konkretne stavke upitnika i šifarnika i pronađite svako pojedinačno pitanje (npr. starost) i ukupne vrednosti posmatranih obeležja na skalama (npr. optimizma) koje ćete upotrebiti u svojim analizama. Identifikujte svaku promenljivu, kako je bila merena, koliko je bilo mogućnosti za odgovor i mogući raspon vrednosti (brojeva, šifara) u koje su odgovori pretvoreni.

Ukoliko je studija obuhvatala eksperiment, proverite kako je bila merena svaka zavisna i nezavisna promenljiva. Da li se vrednosti promenljive sastoje od broja tačnih odgovora, opservatorove ocene konkretnog ponašanja ili dužine vremena koje je subjekat proveo baveći se određenom aktivnošću? Bez obzira na prirodu istraživanja, treba da vam je jasno kako je svaka promenljiva bila merena.

Korak 3: identifikujte prirodu svake promenljive

Sledeći korak je identifikacija prirode svake promenljive u studiji, tj. za svaku promenljivu treba utvrditi da li je nezavisna ili zavisna. Te informacije ne potiču od samih podataka, nego od vašeg shvatanja oblasti i teme studije, relevantnih teorija i prethodnih istraživanja. Mora vam biti jasno u glavi (i u pitanjima postavljenim u istraživanju) kakva je veza između vaših promenljivih – koje (nezavisne) utiču na druge, a koje (zavisne) trpe uticaj drugih. Ima nekih analiza (npr. korelacija) gde nije neophodno utvrditi koje su pro-

menljive nezavisne a koje zavisne. Za druge analize, kao što je ANOVA, to vam mora biti jasno. Korisno je nacrtati model uzajamnog odnosa promenljivih kako ga sami vidite (pogledajte korak 4 u nastavku).

Za svaku promenljivu treba znati i njen nivo merenja. Zavisno od toga da li su promenljive kategorijske ili neprekidne, upotrebljavaju se različite statističke analize, pa morate znati s čim radite. Da li su vaše promenljive:

- kategorijske (nominalni podaci, npr. pol: muški/ženski);
- ordinalne (rangirani podaci: prvi, drugi, treći); ili
- neprekidne (intervalni podaci, npr. starost u godinama ili vrednosti na skalama optimizma)?

U nekim prilikama treba promeniti nivo merenja određenih promenljivih. Odgovori za neprekidne promenljive mogu se svesti na manji broj kategorija (videti 8. poglavlje). Na primer, starost se može podeliti na različite kategorije (npr. ispod 35 godina/preko 35 godina). To bi bilo podesno za analizu varijanse (proceduru ANOVA), a i u slučaju da neprekidna promenljiva ne zadovoljava neku od polaznih pretpostavki određenih analiza (npr. ima veoma asimetričnu raspodelu). Međutim, sažimanje podataka ima očigledan nedostatak jer se njime gube informacije. 'Sabijanjem' ljudi u istu grupu kada se gube važne razlike između njih. Zato dobre i loše strane treba pažljivo odvagati.

Dodatne informacije potrebne za neprekidne i kategorijske promenljive

Za *neprekidne* promenljive trebalo bi da prikupite informacije o raspodeli rezultata (npr. da li im je raspodela normalna ili jako asimetrična?). Koji je raspon njihovih mogućih vrednosti? (Kako se to radi objašnjeno je u poglavlju 6.) Kada promenljiva obuhvata *kategorije* (npr. grupa 1/grupa 2, muškarci/žene), utvrdite koliko osoba spada u svaku od kategorija i da li su te grupe približno jednake ili veoma različite po broju članova?). Da li je neka od mogućih kategorija prazna? (Videti poglavlje 6.) Sve informacije koje ovde prikupite o promenljivama kasnije će se koristiti za sužavanje izbora dostupnih statističkih analiza.

Korak 4: nacrtajte dijagram za svako istraživačko pitanje

Moji studenti često ostaju bez teksta kada treba da objasne šta istražuju. Ponekad je lakše, a i jasnije, sažeti ključne tačke pomoću dijagrama. Ideja je da deo informacija prikupljenih u koracima 1 i 2 objedinimo u jednostavnom formatu koji će pomoći pri izboru prikladne statističke tehnike ili izabrati jednu od više opcija.

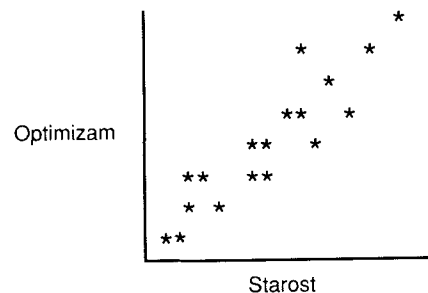
Trebalo bi da razmislite o jednom od ključnih pitanja: da li me zanima *odnos/veza* dve promenljive ili *poređenje* dve grupe subjekata? Možda će vam biti lakše da odgovorite kada za svako pitanje sažete prikupljene informacije i nacrtate dijagram. Ilustrovaću to navođenjem informacija i crtanjem dijagrama za više istraživačkih pitanja.

Pitanje 1: Postoji li veza između starosti i nivoa optimizma?

Promenljive:

- Starost – neprekidna: starost u godinama od 18 do 80; i
- Optimizam – neprekidna: vrednosti na skali optimizma, u rasponu od 6 do 30.

Iz literature ste izvukli hipotezu da optimizam raste sa životnim dobom. Ta veza između dve neprekidne promenljive može se ilustrovati ovako:



Kada očekujete da vrednost na skali optimizma raste sa životnim dobom, tačke crtate počev od donjeg levog ugla dijagrama prema gornjem desnom uglu. Ukoliko prognozirate da vrednost na skali optimizma opada sa životnim dobom, tačke crtate počev od gornjeg levog ugla dijagrama prema donjem desnom uglu.

Pitanje 2: Da li su muškarci skloniji optimizmu od žena?

Promenljive:

- Pol – nezavisna, kategorijska (dve grupe): muškarci i žene; i
- Optimizam – zavisna, neprekidna: vrednosti na skali optimizma, u rasponu od 6 do 30.

Rezultati dobijeni kao odgovor na ovo pitanje, s jednom kategorijskom promenljivom (sa samo dve grupe) i jednom neprekidnom promenljivom, mogu se sažeti ovako:

	Muškarci	Žene
Srednja vrednost na skali optimizma		

Pitanje 3: Da li se starost različito utiče na optimizam muškaraca i žena?

Kada biste istraživali zajednički uticaj starosti i pola na vrednost na skali optimizma, mogli biste podeliti svoj uzorak na tri starosne grupe (ispod 30, 31–49 godina i 50 i više).

Promenljive:

- Pol – nezavisna, kategorijska: muškarci/žene;
- Starost – nezavisna, kategorijska: subjekti podeljeni na tri jednake grupe; i
- Optimizam – zavisna, kategorijska: vrednosti na skali optimizma, raspon od 6 do 30.

Dijagram bi mogao izgledati ovako:

		Starost		
		Ispod 30	31–49	50 i više
Srednja vrednost na skali optimizma	Muškarci			
	Žene			

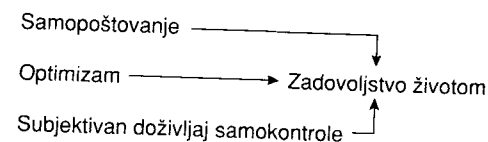
Pitanje 4: Koliki se deo varijanse u zadovoljstvu životom može objasniti pomoću datog skupa osobina ličnosti (samopoštovanje, optimizam, subjektivan doživljaj samokontrole)?

Možda treba da uporedite prediktivnu mogućnost više nezavisnih za jednu zavisnu promenljivu. Takođe vas zanima koliki deo varijanse zavisne promenljive potiče od varijanse tog skupa nezavisnih promenljivih, tj. objašnjen je njom.

Promenljive:

- Samopoštovanje – nezavisna, neprekidna;
- Optimizam – nezavisna, neprekidna;
- Subjektivan doživljaj samokontrole – nezavisna, neprekidna; i Zadovoljstvo životom – zavisna, neprekidna.

Vaš dijagram bi mogao da izgleda ovako:



Korak 5: Zaključite da li je prikladna parametarska ili neparametarska statistička tehnika

Samo da bi studentima bilo teže, mnoštvo dostupnih statističkih tehnika podeljeno je u dve glavne grupe: parametarske i neparametarske. Parametarska statistika je moćnija, ali počiva na više pretpostavki, tj. njene pretpostavke o podacima su strože. Na primer, sve parametarske metode počivaju na pretpostavci da je raspodela rezultata analize u populaciji iz koje je izvučen uzorak normalna.

Svaka parametarska tehnika (kao što su t-testovi, ANOVA, Pirsonova korelacija) ima i svoje dodatne pretpostavke. Da li su one zadovoljene ili ne,

treba proveriti *pre* sprovođenja analiza. Za svaku tehniku obrađenu u preostalim poglavljima, biće navedene konkretne pretpostavke na kojima počiva. *Šta ako pretpostavke na kojima počiva statistička tehnika koju želite da upotrebite nisu zadovoljene?* Nažalost, to se često događa u istraživanjima iz oblasti društvenih nauka. Mnogi od atributa koje hoćemo da izmerimo nisu normalno raspodeljeni. Neki su jako asimetrični, pri čemu većina rezultata ima malu vrednost (npr. depresija); drugi su asimetrični tako da većina rezultata ima veliku vrednost na skali (npr. samopoštovanje).

Kada pretpostavke na kojima počiva statistička tehnika koju želite da upotrebite nisu zadovoljene, na raspolaganju vam je više mogućnosti koje ćemo sada detaljno opisati.

1. mogućnost

Mogli biste ipak upotrebiti tu parametarsku tehniku i nadati se da time niste ozbiljno narušili valjanost svojih nalaza. Neki autori tvrde da je većina statističkih postupaka prilično robusna, tj. da one dobro podnose manja odstupanja od pretpostavki, naročito kada je uzorak pristojne veličine. Ako odlučite da ipak uradite neku parametarsku analizu, to ćete morati nekako da opravdate u izveštaju, pa prikupite podesne citate autora statističkih knjiga, prethodnih istraživača itd. koji podržavaju takvu odluku. Proverite šta kažu časopisi o oblasti koju istražujete, pogotovo oni članci koji opisuju upotrebu istih skala. Pominju li slične probleme? Ukoliko ih pominju, šta su ti autori preduzeli? Jednostavan i čitljiv prikaz robusnosti raznih statističkih testova pročitajte u knjizi Conea i Fostera (1993).

2. mogućnost

Mogli biste modifikovati podatke tako da zadovolje pretpostavke na kojima počiva statistički test (npr. normalnost raspodele). Neki autori predlažu transformaciju promenljivih čija raspodela nije normalna (videti poglavlje 8). Mišljenja o tome su podeljena, pa ćete morati mnogo toga da pročitate kako biste uspešno opravdali svoj postupak (videti Tabachnick & Fidell, 2007).

3. mogućnost

Kada podaci ne zadovoljavaju pretpostavke parametarskih tehnika, možete umesto njih upotrebiti neku neparametarsku tehniku. Mnoge često korišćene parametarske tehnike imaju svoje neparametarske alternative. I one počivaju na nekim pretpostavkama, ali manje strogim. Te neparametarske alternative (npr. Kruskal-Volison test, Man-Vitnjev U test, hi-kvadrat) najčešće su manje moćne, tj. manje osetljive prilikom otkrivanja veza i razlika između grupa. U poglavlju 16 obrađene su neke od uobičajenih neparametarskih tehnika.

Korak 6: donošenje konačne odluke

Nakon što prikupite informacije koje se odnose na istraživačka pitanja, nivo merenja svih promenljivih i karakteristike dostupnih podataka, konačno ste

u položaju da razmotrite sve mogućnosti. U narednom tekstu sažela sam ključne elemente osnovnih statističkih analiza s kojima ćete se sretati. Prođite duž tog spiska, nađite primer vrste istraživačkog pitanja na koje treba da odgovorite i proverite imate li sve potrebne sastojke. Takođe razmislite ima li i drugih načina na koje biste mogli postaviti isto pitanje i stoga primeniti drugačiji statistički pristup. Na kraj poglavlja stavila sam sažetu tabelu koja će vam pomoći u procesu odlučivanja.

Potražite dodatne informacije o tehnikama za koje ste se odlučili i poslušajte se da dobro shvatite načela i pretpostavke na kojima počivaju. Za to je dobro upotrebiti više različitih izvora: razni autori imaju različita mišljenja. Treba dobro da shvatite sporna pitanja – možda ćete čak morati da opravdate korišćenje određene statističke tehnike u svojoj situaciji – zato obavezno mnogo toga pročitajte.

Osnovne osobine glavnih statističkih tehnika

Ovaj odeljak je podeljen na dva pododeljka:

1. tehnike za istraživanje veza (odnosa) između promenljivih (obrađene u četvrtom delu knjige); i
2. tehnike za istraživanje razlika između grupa (obrađene u petom delu knjige).

Istraživanje veza (odnosa) između promenljivih

Hi-kvadrat za nezavisnost

Primer istraživačkog pitanja: Kakav je odnos između pola osobe i stope odustajanja od terapije?

Treba vam:

- jedna kategorijska nezavisna promenljiva (npr. pol: muškarci/žene); i
- jedna kategorijska zavisna promenljiva (npr. odustajanje: Da/Ne).

Zanima vas broj osoba u svakoj kategoriji (ne vrednosti na nekoj skali).

Dijagram:

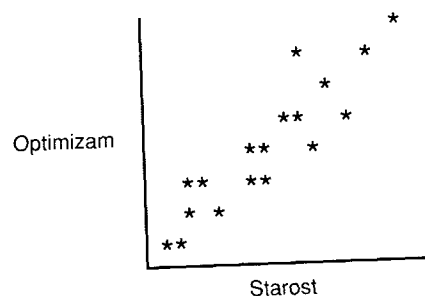
		Muškarci	Žene
Odustajanje	Da		
	Ne		

Korelacija

Primer istraživačkog pitanja: Postoji li veza između starosti i vrednosti na skali optimizma? Raste li optimizam s povećanjem životne dobi?

Treba vam: dve neprekidne promenljive (npr. starost, vrednosti na skali optimizma)

Dijagram:



Neparametarska alternativa: Spirmanova korelacija ranga

Delimična korelacija

Primer istraživačkog pitanja: Nakon uklanjanja uticaja svesno netačnih, ali društveno poželjnih odgovora, postoji li još uvek veza između optimizma i zadovoljstva životom?

Treba vam: Tri neprekidne promenljive (npr. optimizam, zadovoljstvo životom, svesno davanje netačnih, ali društveno poželjnih odgovora)

Neparametarska alternativa: Ne postoji.

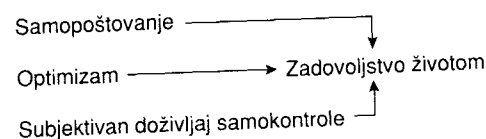
Višestruka regresija

Primer istraživačkog pitanja: Koliki deo varijanse pri ispitivanju zadovoljstva životom potiče od varijanse (tj. može biti pripisan varijansi) sledećeg skupa promenljivih: samopoštovanje, optimizam, subjektivan doživljaj samokontrole? Koja od ovih promenljivih bolje predviđa zadovoljstvo životom?

Treba vam:

- jedna neprekidna zavisna promenljiva (npr. zadovoljstvo životom); i
- dve ili više neprekidnih nezavisnih promenljivih (npr. samopoštovanje, optimizam, subjektivan doživljaj samokontrole).

Dijagram:



Neparametarska alternativa: Ne postoji.

Istraživanje razlika između grupa

T-test nezavisnih uzoraka

Primer istraživačkog pitanja: Da li su muškarci skloniji optimizmu od žena?

Treba vam:

- jedna kategorijska nezavisna promenljiva sa samo *dve* grupe (npr. pol: muškarci/žene);
- jedna neprekidna zavisna promenljiva (npr. vrednost na skali optimizma).

Subjekti mogu pripadati samo *jednoj* grupi.

Dijagram:

	Muškarci	Žene
Srednja vrednost na skali optimizma		

Neparametarska alternativa: Man-Vitnijev test

T-test uparenih uzoraka (ponovljenih merenja)

Primer istraživačkog pitanja: Smanjuje li 10-nedeljna obuka u meditaciji anksioznost učesnika? Da li se nivo anksioznosti menja od vremena 1 (pre intervencije) do vremena 2 (posle intervencije)?

Treba vam:

- jedna kategorijska nezavisna promenljiva (npr. vreme 1 / vreme 2); i
- jedna neprekidna zavisna promenljiva (npr. vrednost na skali anksioznosti).

Isti subjekti ispitani u *dva* zasebna navrata: vreme 1 (pre intervencije) i vreme 2 (posle intervencije).

Dijagram:

	Vreme 1	Vreme 2
Srednja vrednost na skali anksioznosti		

Neparametarska alternativa: Vilkoksonov test ranga

Jednofaktorska analiza varijanse različitih grupa

Primer istraživačkog pitanja: Postoji li razlika u vrednostima na skali optimizma kod osoba mlađih od 30, između 31–49, i starih 50 i više godina?

Treba vam:

- jedna kategorijska nezavisna promenljiva sa dve ili više grupa (npr. starost: ispod 30/31–49/50 i više); i
- jedna neprekidna zavisna promenljiva (npr. vrednost na skali optimizma).

Dijagram:

	Starost		
	Ispod 30	34–49	50 i više
Srednja vrednost na skali optimizma			

Neparametarska alternativa: Kruskal-Willisonov test

Dvofaktorska analiza varijanse različitih grupa

Primer istraživačkog pitanja: Koliko starost utiče na rezultate na skali optimizma za muškarce i za žene?

Šta vam treba:

- dve kategorijske nezavisne promenljive (npr. pol: muškarci/žene; starosna grupa: ispod 30/31–49/50 i više); i
- jedna neprekidna zavisna promenljiva (npr. vrednost na skali optimizma).

Dijagram:

		Starost		
		Ispod 30	34–49	50 i više
Srednja vrednost na skali optimizma	Muškarci			
	Žene			

Neparametarska alternativa: Ne postoji.

Napomena: analiza varijanse se može proširiti tako da obuhvati tri ili više nezavisnih promenljivih. (To se najčešće naziva faktorskom analizom varijanse).

Kombinovana analiza varijanse različitih grupa i ponovljenih merenja

Primer istraživačkog pitanja: Koja intervencija (povećanje matematičkog znanja/izgradnja samopouzdanja) delotvornije smanjuje strah učesnika od statistike, meren u tri navrata (pre intervencije, odmah posle intervencije, tri meseca posle intervencije)?

Treba vam:

- jedna nezavisna promenljiva različitih grupa (npr. vrsta intervencije);
- jedna nezavisna promenljiva ponovljenih merenja istih grupa (npr. vreme 1, vreme 2, vreme 3); i
- jedna neprekidna zavisna promenljiva (npr. vrednosti na testu kojim se ispituje strah od statistike).

Dijagram:

		Vreme		
		Vreme 1	Vreme 2	Vreme 2
Srednja vrednost na testu kojim se ispituje strah od statistike	Intervencija povećanjem matematičkog znanja			
	Intervencija izgradnjom samopouzdanja			

Neparametarska alternativa: Ne postoji.

Multivarijaciona analiza varijanse

Primer istraživačkog pitanja: Da li su muškarci bolje prilagođeni od žena po opštem telesnom i duševnom zdravlju (meri se nivo anksioznosti, depresije i subjektivno doživljenog stresa)?

Treba vam:

- jedna kategorijska nezavisna promenljiva (npr. pol: muškarci/žene); i
- dve ili više zavisnih promenljivih (npr. anksioznost, depresija, subjektivno doživljen stres).

Dijagram:

	Muškarci	Žene
Anksioznost		
Depresija		
Subjektivno doživljen stres		

Neparametarska alternativa: Ne postoji.

Napomena: multivarijaciona analiza varijanse može se koristiti uz jednofaktorsku (jedna nezavisna promenljiva), dvofaktorsku (dve nezavisne promenljive) i višefaktorsku analizu varijanse. Može se uzeti u obzir i dejstvo drugih promenljivih (kovarijansi).

Analiza kovarijanse

Primer istraživačkog pitanja: Postoji li značajna razlika u rezultatima ispitivanja straha od statistike između članova grupe koja povećava matematičko znanje i članova grupe koja gradi samopouzdanje, kada se oduzme uticaj njihovih prethodnih rezultata na tom testu?

Treba vam:

- jedna kategorijska nezavisna promenljiva (npr. vrsta intervencije);
- jedna neprekidna zavisna promenljiva (npr. vrednosti na skali straha od statistike u vreme 2); i
- jedna ili više neprekidnih remetilackih promenljivih (npr. vrednosti na skali straha od statistike u vreme 1).

Neparametarska alternativa: Ne postoji.

Napomena: analiza kovarijanse se može obaviti kao deo jednofaktorske (jedna nezavisna promenljiva), dvofaktorske (dve nezavisne promenljive) ili višefaktorske analize varijanse (dve ili više zavisnih promenljivih).

Zbirna tabela svojstva osnovnih statističkih tehnika

Namena	Primer pitanja	Parametarska tehnika	Neparametarska alternativa	Nezavisna promenljiva	Zavisna promenljiva	Glavna svojstva
Istraživanje veza/odnosa	Koja je veza između pola i stope odustajanja od terapije?	Ne postoji	Hi-kvadrat Poglavlje 16	jedna kategorijska promenljiva <i>Pol: M/Ž</i>	jedna kategorijska promenljiva <i>Odustaje/završava terapiju: Da/Ne</i>	Gleda se broj slučajeva u svakoj kategoriji ne rezultati na skali
	Postoji li veza između starosti i vrednosti na skali optimizma?	Pirsonov koeficijent (r) linearne korelacije Poglavlje 11	Spirmanov koeficijent (ρ) korelacije ranga Poglavlje 11	dve neprekidne promenljive <i>Starost, Vrednosti na skali optimizma</i>		Jedan uzorak sa rezultatima dva različita merenja ili isto merenje u vreme 1 i vreme 2
	Nakon oduzimanja uticaja svesno datih netačnih, ali društveno poželjnih odgovora, da li još uvek postoji veza između optimizma i zadovoljstva životom?	Delimična korelacija Poglavlje 12	Ne postoji	dve neprekidne promenljive i jedna neprekidna promenljiva za koju treba oduzeti uticaj neregularnih odgovora <i>Optimizam, zadovoljstvo životom, rezultati na skali društveno poželjnih odgovora</i>		Jedan uzorak sa rezultatima dva različita merenja ili isto merenje u vreme 1 i vreme 2
	Koliki deo varijanse u vrednostima na skali zadovoljstva životom može biti objašnjen varijansom samopoštovanja, subjektivno doživljene samokontrole i optimizma? Koja od tih promenljivih je najbolji prediktor?	Višestruka regresija Poglavlje 13	Ne postoji	skup dve i više neprekidnih nezavisnih promenljivih <i>Samopoštovanje, subjektivno doživljena samokontrola, optimizam</i>	jedna neprekidna zavisna promenljiva <i>Zadovoljstvo životom</i>	Jedan uzorak sa rezultatima svih merenja
	Koja je pripadna struktura stavki koje sačinjavaju skalu pozitivnih i negativnih osećanja? Koliko faktora postoji?	Faktorska analiza Poglavlje 15	Ne postoji	skup povezanih neprekidnih promenljivih <i>Stavke na skali pozitivnih i negativnih osećanja</i>		Jedan uzorak, više rila
Poređenje grupa	Da li su muškarci skloniji od žena da odustanu od terapije?	Ne postoji	Hi-kvadrat Poglavlje 16	jedna kategorijska nezavisna promenljiva <i>Pol</i>	jedna kategorijska zavisna promenljiva <i>Odustaje/završava terapiju</i>	Zanima vas broj ljudi u svakoj kategoriji vrednosti na nekoj skali
	Menjaju li se rezultati ispitanika na skali anksioznosti od vremena 1 do vremena 2?	T-test uparenih uzoraka Poglavlje 17	Vilkoksonov test ranga Poglavlje 16	jedna kategorijska nezavisna promenljiva (dva nivoa) <i>vreme 1/vreme 2</i>	jedna neprekidna zavisna promenljiva <i>Vrednosti na skali anksioznosti</i>	Isti ljudi u dva navrata

Namena	Primer pitanja	Parametarska tehnika	Neparametarska alternativa	Nezavisna promenljiva	Zavisna promenljiva	Glavna svojstva
	Postoji li razlika u vrednostima na skali optimizma između ljudi mlađih od 35 godina, osoba starih 36-49 i onih od 50 i više godina?	Jednofaktorska ANOVA različitih grupa Poglavlje 18	Kruskal-Volisonov test Poglavlje 16	jedna kategorijska nezavisna promenljiva (tri i više nivoa) <i>Starosna grupa</i>	jedna neprekidna zavisna promenljiva <i>Vrednosti na skali optimizma</i>	Tri ili više grupa; različiti ljudi u svakoj grupi
	Menjaju li se rezultati ispitanika na skali anksioznosti od vremena 1 do vremena 2 i vremena 3?	Jednofaktorska ANOVA ponovljenih merenja Poglavlje 18	Fridmanov test Poglavlje 16	jedna kategorijska nezavisna promenljiva (tri i više nivoa) <i>vreme 1/ vreme 2/ vreme 3</i>	jedna neprekidna zavisna promenljiva <i>Vrednosti na skali anksioznosti</i>	Tri ili više grupa; isti ljudi u dva navrata
	Postoji li razlika u vrednostima na skali optimizma između muškaraca i žena mlađih od 35 godina, osoba starih 36-49, i onih od 50 i više godina?	Dvofaktorska ANOVA različitih grupa Poglavlje 19	Ne postoji	dve kategorijske nezavisne promenljive (dva i više nivoa) <i>Starosna grupa, pol</i>	jedna neprekidna zavisna promenljiva <i>Vrednosti na skali optimizma</i>	Dve ili više grupa za svaku nezavisnu promenljivu: različiti ljudi u svakoj grupi
	Koja intervencija (povećanje matematičkog znanja/izgradnja samopouzdanja) delotvornije smanjuje strah učesnika od statistike, meren u tri navrata?	Kombinovana ANOVA različitih grupa i ponovljenih merenja Poglavlje 20	Ne postoji	jedna nezavisna promenljiva različitih grupa (dva i više nivoa) jedna nezavisna promenljiva iste grupe (dva i više nivoa) <i>Vrsta intervencije, vreme</i>	jedna neprekidna zavisna promenljiva <i>Vrednosti na skali straha od statistike</i>	Dve ili više grupa sa različitim ljudima u svakoj grupi, od koje se svaka meri u dva ili više navrata
	Postoji li razlika između muškaraca i žena, podeljenih u tri starosne grupe, u pogledu raznih merila prilagođenosti (anksioznosti, depresije i subjektivno doživljenog stresa)?	Multivarijaciona ANOVA (MANOVA) Poglavlje 21	Ne postoji	jedna ili više kategorijskih nezavisnih promenljivih (dva i više nivoa) <i>Starosna grupa, pol</i>	dve ili više povezanih neprekidnih zavisnih promenljivih <i>Vrednosti na skali anksioznosti, depresije i subjektivno doživljenog stresa</i>	
	Postoji li značajna razlika u rezultatima ispitivanja straha od statistike između pripadnika grupe koja povećava matematičko znanje i grupe koja gradi samopouzdanje, kada se ukloni uticaj njihovih rezultata na tom testu u vreme 1?	Analiza kovarijance (ANCOVA) Poglavlje 22	Ne postoji	jedna ili više kategorijskih nezavisnih promenljivih (dva i više nivoa) jedna neprekidna, kovarijantna promenljiva <i>Vrsta intervencije, vrednosti na skali straha od statistike u vreme 1</i>	jedna neprekidna zavisna promenljiva <i>Vrednosti na skali straha od statistike u vreme 2</i>	

Literatura za dalje usavršavanje

Statističke tehnike obrađene u ovom poglavlju samo su mali deo dostupnih načina analize podataka. Morate biti svesni postojanja i mogućih načina upotrebe velikog broja tehnika kako biste mogli da izaberete onu najprikladniju za vašu situaciju. Čitajte što više možete.

Osnovne tehnike (t-test, analizu varijanse, korelaciju) učite iz svog udžbenika statistike ili iz knjiga koje su napisali Cooper i Schindler (2003); Gravetter i Wallnau (2004); Peat, J. (2001); Runyon, Coleman i Pittenger (2000); Norman i Streiner (2000). Podrobnije informacije, naročito o multivarijacionoj statistici, videti u knjigama Hair, Black, Babin, Anderson i Tatham (2006) ili Tabachnick i Fidell (2007).

DEO IV

Statističke tehnike za istraživanje veza između promenljivih

U poglavljima iz ovog dela knjige, razmotrićemo neke SPSS-ove tehnike za istraživanje veza između promenljivih. Usredsređićemo se na otkrivanje i opisivanje tih veza. Sve ovde objašnjene tehnike zasnovane su na korelaciji. Korelacione tehnike često koriste istraživači angažovani u neeksperimentalnim projektima. Za razliku od eksperimentalnih projekata, ovde se promenljive namerno ne modifikuju niti kontrolišu, već se opisuju u svom prirodnom stanju. Tim tehnikama se može:

- istražiti veza između parova promenljivih (korelacija);
- predvideti vrednosti jedne promenljive na osnovu vrednosti druge (bivarijantna regresija);
- predvideti vrednosti zavisne promenljive na osnovu vrednosti više nezavisnih promenljivih (višestruka regresija); i
- identifikovati strukturu grupe povezanih promenljivih (faktorska analiza).

Ovom porodicom tehnika testiramo modele i teorije, predviđamo ishode, i ocenjujemo pouzdanost i valjanost skala.

Tehnike obrađene u četvrtom delu knjige

SPSS ima ceo niz tehnika za istraživanje veza. One se razlikuju po vrsti istraživačkog pitanja na koje treba odgovoriti i vrsti dostupnih podataka. U ovoj knjizi su obrađene samo one koje se najčešće upotrebljavaju.

Korelacija (poglavljje 11) upotrebljava se za opisivanje jačine i smera veze između dve (obično neprekidne) promenljive. Može se upotrebiti i kada je jedna od tih promenljivih dihotomna, tj. može imati samo dve vrednosti

(npr. pol: muškarci/žene). Statistički pokazatelj koji se dobija nazvan je Pirsonov koeficijent linearne korelacije (r). Izračunava se i statistička značajnost pokazatelja (veze između dve promenljive) r . Parcijalna ili delimična korelacija (poglavlje 12) služi za istraživanje veze između dve promenljive uz statističku kontrolu uticaja treće. To je prikladno za situacije kada sumnjate da na vezu dve promenljive možda utiče treća. Delimična korelacija statistički uklanja uticaj treće promenljive, čime dobijate jasniju sliku veze dvaju promenljivih od interesa.

Višestruka regresija (poglavlje 13) služi za predikciju (predviđanje) vrednosti jedne zavisne neprekidne promenljive pomoću grupe nezavisnih promenljivih. Njome se može ispitati prediktivna moć skupa promenljivih i oceniti relativan doprinos svake promenljive pojedinačno.

Logistička regresija (poglavlje 14) koristi se umesto višestruke regresije kada je zavisna promenljiva kategorijska. Njome se može ispitati prediktivna moć skupa promenljivih i oceniti relativan doprinos svake promenljive pojedinačno.

Faktorska analiza (poglavlje 15) služi za istraživanje strukture velikog broja povezanih promenljivih (npr. stavki od kojih se sastoji skala). Koristi se za svođenje velikog broja povezanih promenljivih na manji broj dimenzija ili komponenata, s kojim je lakše raditi. U ostatku uvoda u četvrti deo knjige, daću pregled osnovnih načela korelacije zajedničkih za sve tehnike obrađene u ovom delu knjige. To treba proučiti pre nego što pokušate da primenite bilo koji od postupaka obrađenih u ovom delu.

Pregled osnovnih načela

Koeficijent korelacije (npr. Pirsonove linearne korelacije) pokazuje smer i jačinu linearne veze između dve promenljive. Pirsonov koeficijent korelacije (r) ima vrednost u opsegu od -1 do +1. Predznak pokazuje da li je korelacija pozitivna (obe promenljive zajedno i opadaju i rastu) ili negativna (jedna promenljiva opada kada druga raste i obrnuto). Apsolutna vrednost tog koeficijenta (kada zanemarimo njegov predznak) pokazuje jačinu veze. Savršena korelacija, tj. koeficijent 1 ili -1 pokazuje da se vrednost jedne promenljive može tačno utvrditi kada znamo vrednost druge. S druge strane, korelacija 0 pokazuje da između te dve promenljive ne postoji nikakva veza. Poznavanje vrednosti jedne promenljive nimalo ne pomaže u predviđanju vrednosti druge.

Vezu dve promenljive treba ispitati i vizuelno, crtanjem dijagrama rasturanja. Na njemu su svi parovi rezultatnih vrednosti promenljivih dobijeni od subjekata u uzorku. Iznosi prve promenljive se crtaju duž X (horizontalne) ose, a odgovarajući rezultati druge duž Y (vertikalne) ose. Pregledom dijagrama rasturanja vidi se i smer veze (pozitivan ili negativan) i njena jačina (što je detaljnije objašnjeno u poglavlju 11). Dijagram rasturanja savršene korelacije ($r=1$ ili -1) jeste prava linija. Međutim, kada je $r=0$, dijagram rasturanja izgleda kao oblak tačaka koje ne čine nikakav geometrijski oblik.

Činioci koje treba uzeti u obzir prilikom tumačenja koeficijenta korelacije

Kada tumačite rezultate korelacione analize ili drugih tehnika zasnovanih na korelaciji, morate voditi računa o nekoliko stvari. Ovde su navedene ključne teme, ali vam savetujem da u svojim udžbenicima iz statistike pronađete odgovarajuća objašnjenja (na primer, pogledajte Gravetter & Wallnau, 2004, str. 520–576).

Nelinearna veza

Koeficijent korelacije (npr. Pirsonov r) pokazuje linearnu (pravolinijsku) vezu između promenljivih. Kada su promenljive povezane nelinearno (npr. krivolinijski), Pirsonov r će pokazivati da je veza mnogo slabija nego što jeste. Zato uvek pogledajte dijagram rasturanja, naročito kada dobijete malu vrednost r .

Netipične tačke

Netipične tačke (čije su vrednosti znatno manje ili veće od vrednosti ostalih tačaka u skupu podataka) mnogo kvare koeficijent korelacije, pogotovo u malim uzorcima. U nekim okolnostima, netipične tačke čine vrednost r mnogo većom nego što bi ona trebalo da bude, a u drugim podbacuju u oceni jačine veze. Netipične tačke se lako uočavaju na dijagramu rasturanja; to su usamljene tačke, izuzeci. Možda su posledica greške prilikom unošenja podataka (upisano 11 umesto 1), netačnog odgovora ispitanika ili je u pitanju stvarni odgovor neobične osobe! Kada pronađete netipičnu tačku, proverite da li se radi o grešci i ispravite je ako treba. Inače, razmislite o uklanjanju ili rešifrovanju te neobične vrednosti da bi se smanjio njen uticaj na koeficijent r . (Videti raspravu o netipičnim tačkama u poglavlju 6.)

Ograničen opseg rezultata

Koeficijente korelacije morate veoma pažljivo tumačiti kada potiču od malog podopsega stvarno mogućeg raspona rezultata (npr. kada koeficijent inteligencije, IQ, proučavate na uzorku univerzitetskih studenata). Koeficijenti korelacije dobijeni proučavanjem ograničenog podopsega rezultata često se razlikuju od onih kada je uzorkom obuhvaćen pun opseg rezultata. Da bi se dobio tačan i pouzdan pokazatelj jačine veze između dve promenljive, svaka od njih bi trebalo da ima najširi mogući opseg rezultata. Ukoliko proučavate ekstremne grupe (npr. klijente s visokim nivoom anksioznosti), ne uopštavajte korelaciju na slučajeve izvan opsega podataka upotrebljenih u uzorku.

Korelacija u odnosu na kauzalnost

Korelacija je pokazatelj da postoji veza između dve promenljive; međutim, ona ne pokazuje da jedna promenljiva *prouzrokuje* onu drugu. Korelacija između dve promenljive (A i B) može biti posledica činjenice da A prouzrokuje

B, da B prouzrokuje A, ili da (kako bi stvari bile još komplikovanije) treća promenljiva (C) prouzrokuje i A i B. Mogućnost da treća promenljiva prouzrokuje obe opservirane promenljive nikad ne treba gubiti iz vida. To ilustruje čuvena priča o jakoj korelaciji koju je neki istraživač otkrio između potrošnje sladoleda i broja prijavljenih ubistava u Njujorku. Da li konzumiranje sladoleda prouzrokuje nasilno ponašanje ljudi? Ne. Na obe promenljive (potrošnju sladoleda i stopu kriminala) utiču vremenske prilike. Tokom letnjih vrućina rastu i potrošnja sladoleda i stopa kriminala. Uprkos dobijenoj pozitivnoj korelaciji, time nije dokazano da lizanje sladoleda prouzrokuje ubilačko ponašanje. Što je odlično, pošto bi proizvođači sladoleda inače brzo zatvorili svoje fabrike!

Upozorenje je jasno – pazite se mogućeg uticaja treće, remetilacke promenljive kada projektujete svoje istraživanje. Ako sumnjate da bi neke druge promenljive mogle uticati na rezultat, pogledajte možete li ih izmeriti u isto vreme. Delimičnom korelacijom (opisanom u poglavlju 12) može se statistički kontrolisati uticaj dodatnih promenljivih i tako steći jasniji i manje kontaminiran pokazatelj veze dve promenljive od interesa.

Statistička značajnost u odnosu na praktičnu

Nemojte se previše uzbuđivati kada dobijete statistički značajne koeficijente korelacije. Na velikim uzorcima, statističku značajnost mogu doseći čak i sasvim mali koeficijenti korelacije. Praktičan značaj korelacije 0,2 veoma je ograničen, makar njen statistički značaj bio dokazan. Usredsredite se na stvarnu veličinu Pirsonovog koeficijenta r i iznos zajedničke varijanse dve promenljive. Prilikom tumačenja jačine korelacije morate uzeti u obzir druga istraživanja u istoj oblasti. Ako su drugi istraživači u toj oblasti uspeali da predvide samo 9 procenata varijanse (jer su dobili koeficijent korelacije 0,3) određenog ishoda (npr. anksioznosti), onda bi vaša studija koja objašnjava 25 procenata u poređenju s tim bila impresivna. U nekoj drugoj oblasti, 25 procenata objašnjene varijanse može izgledati kao mali i nevažan rezultat.

Pretpostavke

Sve tehnike obrađene u četvrtom delu knjige imaju nekoliko zajedničkih pretpostavki, koje ćemo sada razmotriti. Kada budete radili analize obrađene u poglavljima 11, 12, 13, 14 i 15, trebalo bi da se vraćate na ova objašnjenja.

Nivo merenja

Skala za merenje promenljivih u većini tehnika obrađenih u četvrtom delu knjige trebalo bi da bude intervalna (neprekidna). Izuzetak od toga bi bile jedna dihotomna nezavisna promenljiva (sa samo dve vrednosti: npr. pol) i jedna neprekidna zavisna promenljiva. Međutim, u svakoj kategoriji dihotomne promenljive trebalo bi da imate približno jednak broj osoba ili analiziranih slučajeva.

Spirmanov koeficijent r_o , što je koeficijent korelacije prikladan za ordinalne ili rangirane podatke, obrađuje se u poglavlju 11 zajedno sa svojom parametarskom alternativom, Pirsonovim koeficijentom korelacije r . R_o se često upotrebljava u zdravstvenoj i medicinskoj literaturi, a sve više i u psihološkim istraživanjima, zato što su istraživači postali svesniji mogućih problema koje ume da prouzrokuje pretpostavka da su numeričke udaljenosti između rangova (npr. Likert skale) približno jednake ili srazmerne razlikama u intenzitetu posmatranih obeležja, što je svojstvo intervalnih skala.

Povezani parovi

Svaki subjekat mora dati svoju ocenu i promenljive X i promenljive Y (povezani parovi). Oba podatka moraju poticati od istog subjekta.

Nezavisnost opservacija

Opservacije od kojih se podaci sastoje moraju biti uzajamno nezavisne, tj. na bilo koju opservaciju ili merenje ne sme uticati nijedna druga opservacija ili merenje. Stevens (1996, str. 238) tvrdi da je kršenje ove pretpostavke veoma ozbiljna stvar. Tu pretpostavljenu nezavisnost narušava nekoliko situacija, sto je dokazano u sledećim istraživanjima (videti Stevens, 1996, str. 239; i Gravetter & Wallnau, 2004, str. 251):

- Istraživanje učinka studenata koji rade u parovima ili malim grupama. Ponašanje svakog člana grupe utiče na sve ostale članove, čime se krši pretpostavka o nezavisnosti.
- Istraživanje navika i preferencija dece u vezi s gledanjem TV-a, kada su deca iz iste porodice. Ponašanje jednog deteta u porodici (koje, recimo, gleda program A) najčešće utiče na svu ostalu decu iz te porodice; zato opservacije nisu nezavisne.
- Istraživanje metoda poučavanja u učionici i njihovog uticaja na ponašanje i performanse studenata. Prisustvo malog broja problematičnih studenata može uticati na sve ostale; zato merenja ponašanja i performansi pojedinaca nisu nezavisna.

Treba sumnjati na svaku situaciju gde se opservacije ili merenja prikupljaju u grupnom okruženju ili su subjekti podvrgnuti nekom obliku međusobne interakcije. Kada budete projektovali istraživanje, pokušajte da obezbedite nezavisnost svih opservacija. Ako sumnjate da je ova pretpostavka narušena, Stevens (1996, str. 241) preporučuje da zadate strožu vrednost alfa (npr. $p < 0,01$).

Normalnost

Rezultati dobijeni za sve promenljive trebalo bi da su normalno raspodeljeni. To se može proveriti pregledom histograma rezultata za svaku promenljivu (uputstva videti u poglavlju 6).

Linearnost

Veza između dve promenljive trebalo bi da je linearna. To znači da bi na dijagramu rasturanja trebalo da vidite (približno) pravu liniju, a ne krivu.

Homogenost varijanse

Promenljivost rezultata dobijenih za promenljivu X trebalo bi da je slična za sve vrednosti promenljive Y. Proverite da li je tako na dijagramu rasturanja (uputstva videti u poglavlju 6). Trebalo bi da po celoj dužini izgleda kao približno jednako debela cigara.

Nedostajući podaci

U istraživanju, naročito onom koje obuhvata ljude, veoma retko se u svakoj opservaciji (od svake osobe) dobiju svi podaci. Zato je važno da ustanovite šta nedostaje u datoteci s podacima. Pokrenite proceduru **Descriptives** i utvrdite procenat nedostajućih podataka u svakoj promenljivoj. Kada pronađete promenljivu za koju nedostaje neočekivano mnogo podataka, zapitajte se zašto je tako. Trebalo bi da razmislite i o tome da li su nedostajući podaci raspodeljeni slučajno ili tu postoji neka pravilnost (npr. mnoge žene nisu odgovorile na pitanje o svojoj starosti). SPSS ima proceduru **Missing Value Analysis** koja olakšava pronalaženje pravilnosti u nedostajućim podacima (videti poslednju opciju u meniju **Analyze**).

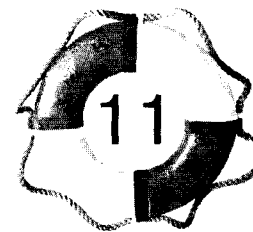
Trebalo bi da razmislite i o tome šta ćete s nedostajućim podacima kada dođe vreme da uradite statističke analize. U mnogim SPSS-ovim statističkim procedurama, pomoću dugmeta **Options** birate načine na koji će SPSS tretirati nedostajuće podatke. Birajte pažljivo, pošto time znatno utičete na statističke rezultate. To je posebno važno kada prilažete spisak promenljivih i za sve njih ponavljate iste analize (npr. korelacije između grupe promenljivih, t-testovi niza zavisnih promenljivih).

- Opcija **Exclude cases listwise** znači da će se analizirati samo slučajevi u kojima za *sve promenljive* navedene u polju **Variables** postoje *svi* podaci. Svaki ispitivan slučaj kome nedostaje makar i delić podataka uopšte neće biti analiziran. Tako biste mogli znatno i nepotrebno da ograničite veličinu uzorka.
- Opcija **Exclude cases pairwise** znači da će slučaj (osoba) biti isključen samo iz onih analiza za koje nedostaje neki od neophodnih podataka. Dakle, i takvi slučajevi će se analizirati kad god je to moguće, tj. kad god sadrže podatke potrebne za pojedinu analizu.
- Opcija **Replace with mean**, dostupna u nekim SPSS-ovim statističkim postupcima (npr. u višestrukoj regresiji), znači da će biti izračunata srednja vrednost svih promenljivih i da će njome biti zamenjeni nedostajući podaci. Ovu opciju *nikada* ne bi trebalo da koristite, pošto može znatno da iskrivi rezultate analize, naročito onda kada nedostaje mnogo podataka.

Kad god sprovedite neki statistički postupak, pritisnite dugme **Options** i proverite koja od navedenih opcija je potvrđena (jer se podrazumevana opcija menja u zavisnosti od postupka). Ako nemate jak razlog da postupite drugačije, preporučila bih vam da slučajeve isključite samo iz onih analiza za koje im nedostaju podaci (*pairwise*). Jedina situacija kada bi vam moglo trebati da analize ograničite samo na slučajeve koji imaju podatke za sve promenljive (*listwise*) jeste ona kada treba razmotriti samo podskup slučajeva koji daje potpun skup rezultata.

Brojevi koji čudno izgledaju

Rezultat koji dobijate od SPSS-a, katkada bi mogao sadržati čudne brojeve oblika 1,24E-02. To su mali brojevi koje je SPSS prikazao u tzv. naučnoj notaciji. Da vam se to ne bi dešavalo, u glavnom meniju izaberite **Edit**, zatim **Options**, pa potvrdite opciju **No scientific notation for small numbers in tables** na kartici **General**.



Korelacija

Korelacija opisuje jačinu i smer linearne veze između dve promenljive. Pomoću SPSS-a se može izračunati više statističkih pokazatelja, koje korisnik bira u zavisnosti od nivoa merenja promenljivih i prirode podataka. U ovom poglavlju predstaviti postupke dobijanja i tumačenja koeficijenta r Pirsonove linearne korelacije, kao i koeficijenta r_o Spirmanove korelacije ranga. Koeficijent Pirsonove linearne korelacije prikladan je za intervalne (neprekidne) promenljive. Može se upotrebiti i kada imate jednu neprekidnu promenljivu (npr. rezultate merenja samopoštovanja) i jednu dihotomnu promenljivu (npr. pol: M/Ž). Spirmanova korelacija ranga je prikladna za ordinalne veličine (koje se mogu rangirati) i naročito se koristi kada podaci ne zadovoljavaju kriterijume za Pirsonovu korelaciju.

SPSS izračunava obe vrste korelacije. Prvo ćete dobiti jednostavnu bivarijantnu korelaciju (tj. korelaciju između dve promenljive), poznatu i kao korelacija nultog reda (engl. *zero-order correlation*). U SPSS-u takođe možete istraživati odnos između dve promenljive, uz kontrolu treće promenljive. Ta tehnika je poznata kao parcijalna korelacija. U ovom poglavlju, predstavljena je procedura dobijanja koeficijenta Pirsonove korelacije proizvod/momenat i neparametarskog Spirmanovog koeficijenta r_o . Parcijalna korelacija je obrađena u poglavlju 12.

Koeficijenti Pirsonove korelacije (r) mogu poprimiti samo vrednosti od -1 do +1. Predznak pokazuje da li je korelacija pozitivna (obe promenljive zajedno i opadaju i rastu) ili negativna (jedna promenljiva opada kada druga raste i obrnuto). Apsolutna vrednost tog koeficijenta (kada zanemarimo njegov predznak) pokazuje jačinu veze. Savršena korelacija, koja iznosi 1 ili -1, pokazuje da se vrednost jedne promenljive može tačno utvrditi kada znamo vrednost druge. Dijagram rasturanja koji ilustruje takvu vezu prava je linija. S druge strane, korelacija jednaka nuli pokazuje da između te dve promenljive ne postoji nikakva veza. Poznavanje vrednosti jedne promenljive tada nimalo ne pomaže u predviđanju vrednosti druge. Dijagram rasturanja tada izgleda kao oblak nasumično raspoređenih tačaka.

Za tumačenje korelacije potrebno je znati odgovore na više pitanja. To su uticaj nelinearnih veza, netipičnih tačaka, ograničenosti opsega uzorka, odnos korelacije i kauzalnosti, i odnos statističke i praktične značajnosti izračunatih koeficijenata korelacije. Te teme su razmotrene u uvodu u četvrti deo knjige. Pre nego što nastavite čitanje ovog poglavlja, trebalo bi da pročitate taj uvod.

Objašnjenje primera

Upotrebu korelacije pokazaću na primeru istraživanja uzajamnih veza nekih promenljivih sadržanih u datoteci s podacima **survey3ED.sav**, s prateće Web lokacije ove knjige. Ta anketa je projektovana da istraži činioce koji utiču na psihološku prilagođenost i opšte raspoloženje ispitanika (celovit opis studije videti u dodatku). U ovom primeru, hoću da ocenim korelaciju između osećanja samokontrole ispitanika i njihovog nivoa subjektivno doživljenog stresa. Dve upotrebene promenljive detaljno su opisane u sledećoj tabeli. Ukoliko želite da pratite i odmah sprovedite navedena uputstva, pokrenite SPSS i otvorite datoteku **survey3ED.sav**.

Kratak pregled korelacije

Primer istraživačkog pitanja: Postoji li veza između nivoa samokontrole ljudi nad njihovim unutrašnjim stanjima i njihovog nivoa subjektivno doživljenog stresa? Da li osobe sa visokim nivoom samokontrole doživljavaju manje stresa?

Šta vam treba: Dve promenljive, obe neprekidne; ili jedna neprekidna, a druga dihotomna (može imati samo dve vrednosti).

Šta se postiže: Korelacija opisuje vezu između dve neprekidne promenljive, i to i jačinu te veze i njen smer.

Pretpostavke: Pretpostavke na kojima počiva korelacija detaljno su objašnjene u uvodu u četvrti deo knjige.

Neparametarska alternativa: Spirmanova korelacija ranga (ro).

Preliminarne analize za korelaciju

Pre izračunavanja korelacije treba nacrtati dijagram rasturanja. Na njemu proverite da li su zadovoljene pretpostavke o linearnosti i homogenosti varijanse (videti uvod u četvrti deo knjige). Pregledom dijagrama rasturanja steći ćete i bolju predstavu o prirodi veze između ispitivanih promenljivih.

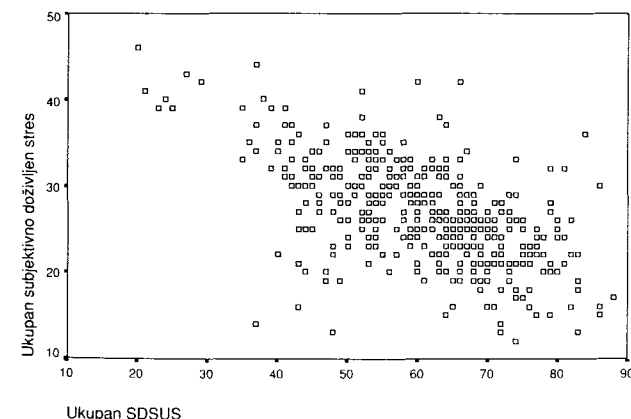
Postupak crtanja dijagrama rasturanja

1. U meniju na vrhu prozora pritisnite **Graphs**, zatim **Legacy Dialogs**.
2. Pritisnite **Simple Scatter** i potom **Define**.
3. Pritisnite prvu promenljivu (npr. Ukupan subjektivno doživljen stres, usudstres) i prebacite je u polje **Y-axis**. (Ona će biti prikazana duž vertikalne ose.) Po konvenciji, zavisna promenljiva obično se prikazuje duž Y-ose.
4. Pritisnite drugu promenljivu (npr. Ukupan SDSUS) i prebacite je u polje **X-axis**. (Ona će biti prikazana duž horizontalne ose.) Tu se najčešće prikazuje nezavisna promenljiva.
5. U polje **Label Cases by:** stavite promenljivu ID da bi uz svaku netipičnu tačku bio njen ID broj.
6. Pritisnite **Continue** i zatim **OK** (ili dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Ovaj postupak generiše sledeću komandu:

```
GRAPH  
/SCATTERPLOT(BIVAR)=usdsus WITH usudstres BY id (IDENTIFY)  
/MISSING=LISTWISE.
```

Rezultat opisanog postupka prikazan je na slici.



Tumačenje dijagrama rasturanja

Na dijagramu se može sagledati više svojstava raspodele dve ispitivane promenljive.

Korak 1: provera netipičnih tačaka

Pogledajte ima li na dijagramu rasturanja netipičnih tačaka, tj. tačaka izdvojenih od ostalih, izuzetaka, eventualno mnogo većih ili manjih od ostalih. Kada netipične tačke imaju ekstremne vrednosti, treba proveriti da li su ti podaci uopšte bili ispravno uneti? Da to nisu greške? Netipične tačke znatno utiču na neke analize, pa to вреди ispitati. Pojedini udžbenici statistike preporučuju da se ekstremne netipične tačke izbace iz skupa podataka. Drugi predlažu da im se promene šifre (kodovi) tako da im vrednost ne bude toliko ekstremna (videti poglavlje 6).

Kada prepoznate netipičnu tačku i poželite da je identifikujete ID brojem datog slučaja (opservacije), upotrebite ikonicu **Data Label Mode** u prozoru **Chart Editor**. Dvaput pritisnite dijagram da biste aktivirali prozor **Chart Editor**. Pritisnite ikonicu koja pomalo liči na centar streljačke mete (ili izaberite **Data Label Mode** u meniju **Elements**) i dovedite kursor na tačku dijagrama koju želite da identifikujete. Pritisnite tačku jednom i prikazaće se njen ID broj (ako ste u koraku 5 prethodnog postupka izabrali ID) odnosno SPSS-ov broj slučaja (engl. *case number*). Prikazivanje brojeva isključićete ako ponovo pritisnete istu ikonicu.

Korak 2: pregled raspodele vrednosti

Raspodela vrednosti kazuje više stvari:

- Da li su tačke raspršene posvuda? To bi značilo da je korelacija veoma mala.
- Da li su tačke lepo raspoređene u obliku tanke cigare? To bi značilo da je korelacija veoma velika.
- Može li se povući prava linija kroz glavnu gomilu tačaka ili bi tačke bolje predstavila neka kriva linija? Ako je kriva linija očigledna (što sugerise krivolinijsku vezu između promenljivih), Pirsonovu korelaciju uopšte ne treba izračunavati, pošto ona počiva na pretpostavci da je veza između promenljivih linearna (pravolinijska).
- Koji je oblik glavne gomile tačaka? Da li je jednake širine od jednog kraja do drugog? Ili je na početku uska pa postaje sve šira? U tom slučaju, podaci možda ne zadovoljavaju pretpostavku o homogenosti varijanse (engl. *homoscedasticity*, *homogeneity of variance*).

Korak 3: određivanje smera veze između promenljivih

Dijagram rasturanja kazuje da li je veza između dve promenljive pozitivna ili negativna. Kada bi se kroz tačke nacrtala linija, koji smer bi pokazivala – sleva nadesno, da li bi pokazivala naviše ili naniže? Trend naviše prikazuje pozitivnu vezu, tj. da velikim vrednostima X odgovaraju velike vrednosti Y. Linija naniže kazuje da se radi o negativnoj korelaciji, tj. da malim vrednostima X odgovaraju velike vrednosti Y. U ovom primeru, izgleda da je korelacija negativna i umereno jaka.

Pošto ste istražili raspodelu rezultata na dijagramu rasturanja i utvrdili da je veza između promenljivih približno linearna i da su rezultati ravnomerno raspoređeni u obliku cigare, možete nastaviti izračunavanje koeficijenta Pirsonove ili Spirmanove korelacije.

Pre nego što počnete sledeći postupak, u glavnom meniju izaberite **Edit**, zatim **Options**, i proverite da li je potvrđena opcija **No scientific notation for small numbers in tables**.

Postupak izračunavanja Pirsonovog koeficijenta r ili Spirmanovog ro

1. U meniju na vrhu prozora pritisnite **Analyze**, zatim **Correlate**, pa **Bivariate**.
2. Izaberite svoje dve promenljive i prebacite ih u polje **Variables** (npr. ukupan subjektivno doživljen stres, ukupan SDSUS). Ovde se može zadati cela lista promenljivih, ne samo dve. U rezultujućoj matrici (tabeli) biće navedene korelacije svih mogućih parova promenljivih. Tabela će biti velika ako u listi zadate više od nekoliko promenljivih.
3. U odeljku **Correlation Coefficients**, podrazumevana opcija je **Pearson** (što znači da se podrazumevano izračunava ova korelacija). Ako hoćete da izračunate i Spirmanovu korelaciju ranga, potvrdite polje **Spearman rho**.
4. Pritisnite dugme **Options**. Za nedostajuće vrednosti (odeljak **Missing Values**), potvrdite polje **Exclude cases pairwise**. Na kartici **Options** može se zadati i izračunavanje srednje vrednosti i standardnog odstupanja.
5. Pritisnite **Continue** i zatim **OK** (ili dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Opisani postupak generiše sledeće komande:

```
CORRELATIONS
/VARIABLES=usudstres usdsus
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.
NONPAR CORR
/VARIABLES=usudstres usdsus
/PRINT=SPEARMAN TWOTAIL NOSIG
/MISSING=PAIRWISE.
```

Rezultat prethodnog postupka (sa obe izračunate korelacije, i Pirsonovom i Spirmanovom) prikazan je u tabelama.

Correlations

		Total perceived stress	Total PCOISS
Total perceived stress	Pearson Correlation	1	-.581**
	Sig. (2-tailed)		.000
	N	433	426
Total PCOISS	Pearson Correlation	-.581**	1
	Sig. (2-tailed)	.000	
	N	426	430

** Correlation is significant at the 0,01 level (2-tailed).

Nonparametric correlations

Correlations

			Total perceived stress	Total PCOISS
Spearman's rho	Total perceived stress	Correlation Coefficient	1,000	-.556**
		Sig. (2-tailed)		.000
		N	433	426
	Total PCOISS	Correlation Coefficient	-.556**	1,000
		Sig. (2-tailed)	.000	
		N	426	430

** Correlation is significant at the 0,01 level (2-tailed).

Tumačenje rezultata korelacije

Rezultate obe korelacije, i Pirsonove i Spirmanove, SPSS prikazuje u tabeli koeficijenata korelacije između svakog para navedenih promenljivih, uz nivo značajnosti i broj analiziranih slučajeva (opservacija) u uzorku. Rezultati Pirsonove korelacije dati su u odeljku **Correlation**. Kada zatražite da se izračuna Spirmanov koeficijent ρ (*rho*), ti rezultati se prikazuju u odeljku **Nonparametric Correlations**. Rezultati obe vrste korelacije, i parametarske i neparametarske, tumače se jednako.

Korak 1: provera informacija o uzorku

U tabeli **Correlations** najpre treba pogledati N, tj. broj slučajeva (opservacija) u uzorku. Da li je taj broj tačan? Ako mnogo podataka nedostaje, treba da utvrdite zašto je tako. Da niste zaboravili da potvrdite polje **Exclude cases pairwise** u odeljku **Missing Data** na kartici **Options**? Druga opcija (**Exclude cases listwise**) znači da su iz svih analiza isključeni svi slučajevi kojima nedostaje vrednost bilo koje promenljive. Time se N često preoštro (i nepotrebno) smanjuje. U gornjem primeru imamo 426 slučajeva sa vrednostima na obe skale upotrebljene u ovoj analizi. Naravno, iz nje su morali da se izbace svi slučajevi kojima nedostaje podatak za bilo koju *od te dve* promenljive.

Korak 2: utvrđivanje smera veze

Drugo što treba pogledati jeste smer veze između promenljivih. Da li je ispred vrednosti koeficijenta korelacije negativan predznak? Ako jeste, to znači da između te dve promenljive postoji negativna korelacija (tj. da velike vrednosti na jednoj skali prate male vrednosti na drugoj). Tumačenje ovoga se menja u zavisnosti od načina dodeljivanja vrednosti promenljivama. Uvek proverite šta piše u upitniku i ne zaboravite da mnoge skale imaju negativno formulisane stavke koje treba obrnuti pre dodeljivanja vrednosti. Šta veliki brojevi (vrednosti) zapravo znače? To često zbunjuje studente, pa vam mora biti savršeno jasno pre nego što se usudite da tumačite rezultate.

U ovom primeru, koeficijent Pirsonove korelacije (-0,58) i Spirmanov koeficijent ρ (-0,56) negativni su, što pokazuje negativnu korelaciju između subjektivno doživljene samokontrole i stresa. Što *više* samokontrole osoba oseća da ima, to *manje* stresa misli da doživljava.

Korak 3: utvrđivanje jačine veze

Treće što treba pogledati u rezultatima jeste veličina koeficijenta korelacije. On može imati vrednost od -1,00 do 1,00. Taj broj pokazuje jačinu veze između te dve promenljive. Vrednost 0 pokazuje da nikakva veza (korelacija) ne postoji; vrednost 1,0 pokazuje da je korelacija potpuna i pozitivna, dok vrednost -1,0 pokazuje da je korelacija potpuna i negativna.

Kako tumačimo korelacije čije su vrednosti između 0 i 1? Razni autori daju različita tumačenja; međutim, Cohen (1988, str. 79–81) daje sledeće smernice za veličinu korelacije:

mala $r=0,10$ do $0,29$
srednja $r=0,30$ do $0,49$
velika $r=0,50$ do $1,0$

Navedene smernice važe bez obzira na to da li ispred koeficijenta r stoji negativan predznak. Ne zaboravite, negativan predznak pokazuje smer veze, a ne njenu jačinu. *Jačina* korelacije je jednaka kada su koeficijenti $r=0,5$ odnosno $r=-0,5$. Razlikuje se samo njen *smer*.

U gornjem primeru, između dve promenljive postoji jaka korelacija (iznad 0,5), što govori u prilog tvrdnje da je veza između subjektivno doživljene samokontrole i stresa jaka.

Korak 4: izračunavanje koeficijenta determinacije

Koeficijent determinacije pokazuje koliki je deo varijanse dve promenljive zajednički; još se kaže „koliki je deo varijanse jedne promenljive objašnjen (ili prouzrokovao) varijansom druge“. Zvuči impresivno, a izračunava se jednostavno; samo kvadrirajte vrednost r (pomnožite r sa samim sobom). Da biste to pretvorili u 'procenat varijanse', dobijeni proizvod pomnožite sa 100

(pomerite decimalni zarez za dva mesta udesno). Na primer, dve promenljive s koeficijentom korelacije $r=0,2$ imaju samo $0,2 \times 0,2 = 0,04 = 4$ procenta zajedničke varijanse. Te dve promenljive nemaju mnogo veze. Međutim, korelacija $r=0,5$ znači već 25 procenata zajedničke varijanse ($0,5 \times 0,5 = 0,25$).

U našem primeru, Pirsonova korelacija iznosi 0,581; kada se digne na kvadrat, dobija se 33,76 procenata zajedničke varijanse. Subjektivno doživljena samokontrola objašnjava gotovo 34 procenta varijanse u odgovorima ispitanika na skali subjektivno doživljenog stresa. To je sasvim pristojan deo objašnjene varijanse, kada se uporedi s mnogim istraživanjima sprovedenim u društvenim naukama.

Korak 5: ocena nivoa značajnosti

Sledeće što treba pogledati je nivo značajnosti (označen sa Sig. 2 tailed). Ovaj podatak se često pogrešno tumači, pa obratite pažnju. Nivo statističke značajnosti ne pokazuje jačinu veze između dve promenljive (nju pokazuje r odnosno r^2), nego s koliko poverenja treba posmatrati dobijene rezultate. Na značajnost dobijenog iznosa koeficijenta r odnosno r^2 jako utiče veličina uzorka. Iz malih uzoraka (npr. $n=30$) mogu se izračunati umerene korelacije koje nisu statistički značajne na uobičajenom nivou $p<0,05$. Međutim, već i vrlo male korelacije izračunate iz velikih uzoraka ($N=100+$) mogu biti statistički značajne. Naravno da statističku značajnost koeficijenta korelacije treba navesti u izveštaju o istraživanju, ali se usredsredite na jačinu veze i veličinu zajedničke varijanse (videti korak 4).

Predstavljanje rezultata korelacije

Rezultati Pirsonove korelacije u gornjem primeru mogli bi se ovako predstaviti u izveštaju o istraživanju. Ukoliko treba da predstavite rezultate Spirmanove korelacije, samo zamenite r vrednošću r_s datom u drugoj tabeli.

Veza između subjektivno doživljene samokontrole unutrašnjih stanja (kako je meri skala SDSUS) i subjektivno doživljenog stresa (merenog na istoimenoj skali) istražena je pomoću koeficijenta Pirsonove linearne korelacije. Obavljene su preliminarne analize da bi se dokazalo zadovoljenje pretpostavki o normalnosti, linearnosti i homogenosti varijanse. Izračunata je jaka negativna korelacija između te dve promenljive, $r = -0,58$, $n = 426$, $p < ,0005$, pri čemu visoke nivoje subjektivno doživljene samokontrole prate niski nivoi subjektivno doživljenog stresa.

Korelacijom se često istražuju veze unutar cele grupe promenljivih, a ne samo veza između dve promenljive kao u gornjem primeru. U takvim slučajevima, bilo bi nezgrapno u jednom pasusu navoditi sve izračunate koeficijente korelacije; bolje je prikazati ih u tabeli. Evo načina na koji se to može uraditi:

Tabela 1: Koeficijenti Pirsonove linearne korelacije između izmerenih vrednosti subjektivno doživljene samokontrole i opšteg raspoloženja

Skala					
1. Ukupna SDSUS	–	–,581 **	–,484 **	,456 **	,373 **
2. Uk. subj. doživ. stres		–	,674 **	–,442 **	–,494 **
3. Ukupna neg. osećanja			–	–,294 **	–,316 **
4. Ukupna poz. osećanja				–	,415 **
5. Uk. zadovolj. životom					–

** $p < ,001$ (2-tailed).

Računanje korelacije između grupa promenljivih

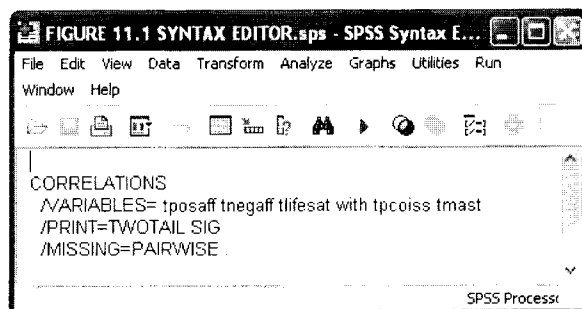
U prethodnom odeljku sa opisom postupka pokazala sam kako se dobijaju koeficijenti korelacije između dve neprekidne promenljive. Kada imate grupu promenljivih i želite da istražite uzajamne veze svih njih, zadajte SPSS-u da sve to uradi u jednoj proceduri – samo prebacite sve te promenljive u polje **Variables**. Međutim, tako bi se mogla dobiti ogromna matrica korelacije, koja se teško čita i tumači.

Ponekad je od interesa samo određeni podskup svih mogućih veza. Na primer, zanima vas odnos između vrednosti kontrole (Mastery, PCOISS) i raznih vrednosti prilagođenja (pozitivnih osećanja, engl. *positive affect*, negativnih osećanja, engl. *negative affect*, zadovoljstva životom, engl. *life satisfaction*). Ne želite celu matricu korelacije zato što bi ona sadržala koeficijente korelacije svih promenljivih, dakle i svakog para mera prilagođenja. Postoji način da to ograničite, tj. da odredite koji će se koeficijenti korelacije prikazati. To se radi u **Syntax Editoru** (opisanom u poglavlju 3). Sledi postupak pomoću kojeg se u **Syntax Editoru** određuje koje će koeficijente korelacije SPSS izračunati.

Postupak izračunavanja koeficijenata korelacije između dve grupe promenljivih

1. Otvorite meni **Analyze** na vrhu prozora, izaberite u njemu **Correlation**, pa pritisnite **Bivariate**.
2. Promenljive koje vas zanimaju prebacite u polje **Variables**, najpre celu prvu grupu promenljivih (npr. *tposaff*, *tnegaff*, *tlifesat*), a zatim drugu (npr. *tpcoiss*, *tmast*).
3. U ispisu rezultata, prva grupa promenljivih biće nanizana po vertikalni tabele u njenim redovima, a druga grupa po horizontali u kolonama.
4. Dužu listu stavite na prvo mesto; time sprečavate da tabela zbog prevelike širine ne može da stane na stranicu.

5. Pritisnite **Paste**. Otvoriće se **Syntax Editor** (videti primer na slici 11.1).
6. Postavite kursor između prve grupe promenljivih (npr. tposaff, tnegaff, tlifesat) i ostalih promenljivih (npr. tpcoiss i tmast). Upišite reč „with“ (tposaff tnegaff tlifesat **with** tpcoiss tmast). Time ste od SPSS-a zatražili da izračuna koeficijente korelacije između promenljivih tmast i tpcoiss s jedne strane i svih drugih navedenih promenljivih s druge strane. Trebalo bi da dobijete komandu kao na slici 11.1.
7. Da biste izvršili tu novu komandu, izaberite sve – od **CORRELATIONS** do tačke na kraju (uključujući i tačku). *Veoma je važno da i ta tačka bude izabrana.*
8. Kada je taj tekst izabran, u glavnom meniju pritisnite dugme (ikonu) sa troglom koji pokazuje udesno (>) ili otvorite meni **Run** i u njemu izaberite **Selection**. Time ste SPSS-u naredili da izvrši datu komandu.



Slika 11.1 Prozor **Syntax Editor**.

Sledi rezultat koji ćete dobiti u SPSS-u kada završite opisani postupak.

Correlations		Total PCOISS	Total Mastery
Total Positive Affect	Pearson Correlation	,456**	,432**
	Sig. (2-tailed)	,000	,000
	N	429	436
Total Negative Affect	Pearson Correlation	-,484**	-,464**
	Sig. (2-tailed)	,000	,000
	N	428	435
Total Life Satisfaction	Pearson Correlation	,373**	,444**
	Sig. (2-tailed)	,000	,000
	N	429	436

** Correlation is significant at the 0,01 level (2-tailed).

Kada se rezultati predstave na ovaj način, lako se poredi relativna jačina korelacije moje dve kontrolne skale (Total PCOISS, Total Mastery) sa svakom merom prilagođenja.

Poređenje koeficijenata korelacije dve grupe

U istraživanju korelacije ponekad treba uporediti jačinu koeficijenata korelacije dve zasebne grupe. Na primer, želite da pogledate vezu između optimizma i negativnih osećanja zasebno za muškarce i žene. Sledi opis jednog načina da se to uradi.

Postupak poređenja koeficijenata korelacije za dve grupe subjekata

Korak 1: podelite uzorak

1. Prozor **Data Editor** treba da je otvoren i u prednjem planu. (Ako je trenutno otvoren prozor **Viewer**, otvorite meni **Window** i u njemu izaberite **SPSS Data Editor**.)
2. U glavnom meniju na vrhu prozora pritisnite **Data**, zatim **Split File**.
3. Pritisnite **Compare Groups**.
4. Promenljivu grupisanja (npr. pol) prebacite u polje **Groups based on**. Pritisnite dugme **OK**.
5. Tako ste uzorak podelili po polu, pa će sve ubuduće zahtevane analize biti ponovljene za svaku od te dve grupe zasebno.

Korak 2: korelacija

1. Prema postupku opisanom u jednom od prethodnih odeljaka, zadajte izračunavanje korelacije između dve promenljive od interesa (npr. Total optimism, Total negative affect). Ona će biti izračunata i prikazana zasebno za svaku od dve grupe.

Važno: Kada završite sa zasebnim pregledom statističkih pokazatelja za muškarce i žene, ne zaboravite da isključite opciju **Split File**. Ona (podela uzorka na grupe) ostaje da važi sve dok je ne isključite. To se radi u prozoru **Data Editor** koji treba da je u prednjem planu. Pritisnite **Data**, **Split File** i zatim prvo dugme: **Analyze all cases, do not create groups**. Pritisnite **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Ovaj postupak generiše sledeće komande:

```

SORT CASES BY pol.
SPLIT FILE
  LAYERED BY pol.
CORRELATIONS
/VARIABLES=toptim tnegaff
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.
    
```


Postupak daje ove rezultate:

Correlations

			Total Optimism	Total Negative Affect
SEX	MALES	Total Optimism	Pearson	
			Correlation	1,000
			Sig. (2-tailed)	-,220**
			N	,003
				184
				184
		Total Negative Affect	Pearson	
			Correlation	-,220**
			Sig. (2-tailed)	1,000
			N	,003
				184
				185
FEMALES		Total Optimism	Pearson	
			Correlation	1,000
			Sig. (2-tailed)	-,394**
			N	,000
				251
				250
		Total Negative Affect	Pearson	
			Correlation	-,394**
			Sig. (2-tailed)	1,000
			N	,000
				250
				250

** Correlation is significant at the 0,01 level (2-tailed).

Tumačenje rezultata korelacije izračunate za dve grupe

U prethodnoj tabeli rezultata, korelacija između ukupnog optimizma (Total optimism) i ukupnih negativnih osećanja (Total negative affect) za muškarce bila je $r = -0,22$, dok je za žene bila neznatno veća, $r = -0,39$. Iako se ta dva broja razlikuju, da li je ta razlika dovoljno velika da se može smatrati značajnom, a ne slučajnom? U narednom odeljku detaljno je opisan jedan način provere statističke značajnosti razlike između ta dva koeficijenta korelacije. Važno je imati u vidu da se taj proces razlikuje od provere statističke značajnosti koeficijenata korelacije, čiji su rezultati dati u gornjoj tabeli. Goreprikazani nivoi značajnosti (za muškarce: Sig. = ,003; za žene: Sig. = ,000) daju verovatnoću nulte hipoteze da je koeficijent korelacije u populaciji jednak 0. Međutim, doleopisani test značajnosti ocenjuje verovatnoću da je razlika opaženih korelacija za dve grupe (muškaraca i žena) posledica greške uzorka, dok u stvari nema razlike u jačini ispitivane veze za muškarce i za žene.

Provera statističke značajnosti razlike između koeficijenata korelacije

U ovom odeljku, opisaćemo postupak utvrđivanja da li se korelacije dveju grupa značajno razlikuju, tj. da li je ta razlika slučajna ili ne. Nažalost, to SPSS još ne ume da uradi, pa ćete morati sami pomoću kalkulatora. I ako vam sledeći odeljak bude izgledao nerazumljiv, sledite dati postupak korak po korak. Prvo ćemo vrednosti koeficijenata korelacije r pretvoriti u vrednosti z , zatim pomoću jednačine izračunati opaženu vrednost z , tj. z_{ops} . Taj broj ćemo oceniti na osnovu utvrđenih pravila odlučivanja i tako izračunati verovatnoću da je razlika u korelaciji opažena između dve grupe bila posledica slučajnosti.

Pretpostavke

Kao i uvek, najpre ćemo proveriti pretpostavke. Pretpostavlja se da su vrednosti korelacije r za dve grupe dobijeni iz slučajnih uzoraka i da su grupe slučajeva (opservacija) nezavisne (da isti subjekti nisu dvaput testirani). Pretpostavlja se da je raspodela rezultata za obe grupe normalna (videti histograme za te grupe). Takođe je neophodno imati najmanje 20 slučajeva (opservacija) u svakoj grupi.

Korak 1: pretvoriti svaku vrednost r u odgovarajuću vrednost z

Prvi korak u procesu poređenja jeste pretvoriti dve dobijene vrednosti r u standardan oblik, koji se naziva vrednost z . To se radi iz više matematičkih razloga, prvenstveno zato da bi raspodele uzoraka bile približno normalne.

U prethodno navedenoj tabeli rezultata očitajte iznose r i N za Grupu 1 (muškarci) i Grupu 2 (žene).

$$\begin{array}{ll} \text{Mušk.: } r_1 = -0,22 & \text{Žene: } r_2 = -0,394 \\ N_1 = 184 & N_2 = 250 \end{array}$$

U tabeli 11.1 pronađite vrednost z koja odgovara svakoj od tih vrednosti r .

$$\text{Mušk.: } z_1 = -0,224 \quad \text{Žene: } z_2 = -0,418$$

Tabela 11.1: Transformacija r u z

r	z _r	r	z _r	r	z _r	r	z _r	r	z _r
,000	,000	,200	,203	,400	,424	,600	,693	,800	1,099
,005	,005	,205	,208	,405	,430	,605	,701	,805	1,113
,010	,010	,210	,213	,410	,436	,610	,709	,810	1,127
,015	,015	,215	,218	,415	,442	,615	,717	,815	1,142
,020	,020	,220	,224	,420	,448	,620	,725	,820	1,157
,025	,025	,225	,229	,425	,454	,625	,733	,825	1,172
,030	,030	,230	,234	,430	,460	,630	,741	,830	1,188
,035	,035	,235	,239	,435	,466	,636	,750	,835	1,204
,040	,040	,240	,245	,440	,472	,640	,758	,840	1,221
,045	,045	,245	,250	,445	,478	,645	,767	,845	1,238
,050	,050	,250	,255	,450	,485	,650	,775	,850	1,256
,055	,055	,255	,261	,455	,491	,655	,784	,855	1,274
,060	,060	,260	,266	,460	,497	,660	,793	,860	1,293
,065	,065	,265	,271	,465	,504	,665	,802	,865	1,313
,070	,070	,270	,277	,470	,510	,670	,811	,870	1,333
,075	,075	,275	,282	,475	,517	,675	,820	,875	1,354
,080	,080	,280	,288	,480	,523	,680	,829	,880	1,376
,085	,085	,285	,293	,485	,530	,685	,838	,885	1,398
,090	,090	,290	,299	,490	,536	,690	,848	,890	1,422
,095	,095	,295	,304	,495	,543	,695	,858	,895	1,447
,100	,100	,300	,310	,500	,549	,700	,867	,900	1,472
,105	,105	,305	,315	,505	,556	,705	,877	,905	1,499
,110	,110	,310	,321	,510	,563	,710	,887	,910	1,528
,115	,116	,315	,326	,515	,570	,715	,897	,915	1,557
,120	,121	,320	,332	,520	,576	,720	,908	,920	1,589
,125	,126	,325	,337	,525	,583	,725	,918	,925	1,623
,130	,131	,330	,343	,530	,590	,730	,929	,930	1,658
,135	,136	,335	,348	,535	,597	,735	,940	,935	1,697
,140	,141	,340	,354	,540	,604	,740	,950	,940	1,738
,145	,146	,345	,360	,545	,611	,745	,962	,945	1,783
,150	,151	,350	,365	,550	,618	,750	,973	,950	1,832
,155	,156	,355	,371	,555	,626	,755	,984	,955	1,886
,160	,161	,360	,377	,560	,633	,760	,996	,960	1,946
,165	,167	,365	,383	,565	,640	,765	1,008	,965	2,014
,170	,172	,370	,388	,570	,648	,770	1,020	,970	2,092
,175	,177	,375	,394	,575	,655	,775	1,033	,975	2,185
,180	,182	,380	,400	,580	,662	,780	1,045	,980	2,298
,185	,187	,385	,406	,585	,670	,785	1,058	,985	2,443
,190	,192	,390	,412	,590	,678	,790	1,071	,990	2,647
,195	,198	,395	,418	,595	,685	,795	1,085	,995	2,994

Izvor: McCall (1990); prvobitni izvor Edwards, A.L. (1967). *Statistical methods* (2nd edition). Holt, Rinehart & Winston.

Korak 2: ubacite te brojeve u jednačinu za izračunavanje z_{ops}
 Evo jednačine koja nam je potrebna. Samo ubacite brojeve na odgovarajuća mesta u njoj i izračunajte vrednost z pomoću kalkulatora. (Biće vam lakše ako ste brojeve pripremili i ispisali čitko i pregledno.)

$$z_{obs} = \frac{z_1 - z_2}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}}$$

$$z_{obs} = \frac{,224 - ,418}{\sqrt{\frac{1}{184 - 3} + \frac{1}{250 - 3}}}$$

$$z_{obs} = \frac{-,194}{\sqrt{\frac{1}{181} + \frac{1}{247}}}$$

$$z_{obs} = \frac{-,194}{\sqrt{,0055 + ,004}}$$

$$z_{obs} = \frac{-,194}{\sqrt{,0095}}$$

$$z_{obs} = \frac{-,194}{,0975} = -1,99$$

Korak 3: utvrdite da li je vrednost z_{ops} statistički značajna
 Kada dobijete vrednost z_{ops} između -1,96 i +1,96, *ne možete* tvrditi da između dva koeficijenta korelacije postoji statistički značajna razlika, tj. da ona nije slučajna.

Rečeno jezikom statistike, nultu hipotezu (da nema razlike između grupa) smete da odbacite *jedino* kada je dobijena vrednost z izvan datih granica. Dakle, pravilo odlučivanja je sledeće:

Kada je $-1,96 < z_{ops} < 1,96$: koeficijenti korelacije se ne razlikuju statistički značajno.

Kada je z_{ops} manje od ili jednako -1,96 ili kada je z_{ops} jednako ili veće od 1,96: koeficijenti se razlikuju statistički značajno, tj. razlika nije slučajna.

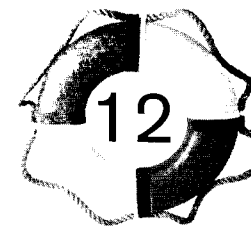
U navedenom primeru izračunatom za muškarce i žene, dobili smo z_{ops} vrednost -1,99. Ona je izvan datih granica, pa smemo da zaključimo da postoji statistički značajna razlika u jačini korelacije između optimizma i negativnih osećanja za muškarce i žene. Optimizam objašnjava značajno veći deo varijanse u negativnim osećanjima žena nego muškaraca; ta razlika nije slučajna.

Dodatne vežbe

Zdravstvo

Datoteka s podacima: **sleep3ED.sav**. Pojednosti o datoteci videti u dodatku.

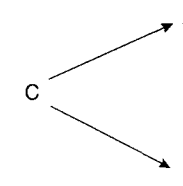
1. Proverite jačinu korelacije između rezultata na skalama Sleepiness and Associated Sensations (*totSAS*) i Epworth Sleepiness (*ess*).
2. Pomoću komandi u Syntax Editoru ocenite korelaciju između skale Epworth Sleepiness (*ess*) i svake stavke sa skale Sleepiness and Associated Sensations (*fatigue, lethargy, tired, sleepy, energy*).



Delimična korelacija

Delimična korelacija je slična Pirsonovoj linearnoj korelaciji (opisanoj u poglavlju 11), osim što se može sprečiti uticaj jedne dodatne promenljive. To je obično promenljiva na koju se sumnja da utiče na dve promenljive od interesa. Kada se statistički ukloni uticaj te remetilačke promenljive, dobija se jasnija i tačnija slika veze između dve ispitivane promenljive.

Uticaj remetilačke promenljive bio je razmotren u uvodu u četvrti deo knjige (videti odeljak o odnosu korelacije i kauzalnosti). Remećenjem se smatra uticaj, makar i mali, treće promenljive (C) na ispitivanu vezu između dve promenljive (A i B). Ako se taj uticaj (kontaminacija) ne uračuna, dobiće se nerealno velika vrednost koeficijenta korelacije dve promenljive. Taj odnos se ovako može predstaviti grafički:



U ovom slučaju izgleda kao da su A i B povezani, ali je njihova prividna veza u velikoj meri posledica uticaja C. Kada statistički uračunate (uklonite) uticaj promenljive C, obično se smanji korelacija između A i B, pa se dobija i manji koeficijent te korelacije.

Objašnjenje primera

Kao ilustraciju delimične korelacije upotrebiću primer iz poglavlja 11, ali ću analizu proširiti tako da obuhvati uticaj još jedne promenljive. Istražiću vezu između rezultata na skali subjektivno doživljene samokontrole unutrašnjih stanja (engl. *Perceived Control of Internal States scale*, PCOISS) i rezultata na skali subjektivno doživljenog stresa (engl. *Perceived Stress scale*), ne gubeći iz vida mogućnost da ispitanici ponekad svesno daju netačne, ali društveno poželjne odgovore (engl. *socially desirable responding bias*). Time

uračunavam uticaj ljudske sklonosti da prilikom popunjavanja upitnika sebe predstave na pozitivan ili društveno poželjan način (da se pretvaraju da su dobri). Ta tendencija se meri Marlou–Kraunovom skalom društvene poželjnosti (Crowne & Marlowe, 1960). Skraćena verzija te skale (Strahan & Gerbasi, 1972) bila je stavljena u upitnik pomoću kojeg su izmerene druge dve promenljive. Njeno ime u datoteci s podacima je tmarlow.

Ukoliko želite da pratite primer i navedena uputstva, pokrenite SPSS i otvorite datoteku survey3ED.sav s prateće Web lokacije knjige.

Kratak pregled delimične korelacije

Primer istraživačkog pitanja: Nakon uračunavanja sklonosti subjekata da se predstave u pozitivnom svetlu na skalama samoocenjivanja, da li i dalje postoji značajna veza između subjektivno doživljene samokontrole unutrašnjih stanja (SDSUS) i nivoa subjektivno doživljenog stresa?

Šta vam treba: Tri neprekidne promenljive:

- dve promenljive čiju vezu istražujete (npr. ukupna SDSUS, ukupan subjektivno doživljen stres); i
- jedna promenljiva čiji uticaj treba ukloniti (npr. ukupna društvena poželjnost: tmarlow).

Šta se postiže: Delimična korelacija služi za istraživanje veze između dve promenljive, pri čemu se statistički kontroliše (uklanja) uticaj treće promenljive koja bi mogla da kontaminira vezu između ispitivane dve promenljive ili da utiče na nju.

Pretpostavke: Pretpostavke na kojima se zasniva računanje korelacije navedene su u uvodu u četvrti deo knjige.

Pre nego što počnete sledeći postupak, otvorite meni **Edit** i izaberite u njemu stavku **Options**; treba da je potvrđeno polje **No scientific notation for small numbers in tables**.

Postupak za delimičnu korelaciju

- Otvorite meni **Analyze**, izaberite u njemu stavku **Correlate**, pa pritisnite **Partial**.
- Pritisnite dve neprekidne promenljive čiju korelaciju hoćete da izračunate (npr. ukupna SDSUS, ukupan subjektivno doživljen stres). Kada pritisnete strelicu, prebacićete ih u polje **Variables**.
- Pritisnite promenljivu čiji uticaj želite da uklonite (npr. tmarlow, koja meri društvenu poželjnost). Prebacite je u polje **Controlling for**.
- Pritisnite dugme **Options**.
 - U odeljku **Missing Values** potvrdite polje **Exclude cases pairwise**.
 - U odeljku **Statistics** potvrdite polje **Zero order correlations**.

5. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Opisani postupak generiše sledeću komandu:

```
PARTIAL CORR
/VARIABLES= usdsus usudstres BY tmarlow
/SIGNIFICANCE=TWOTAIL
/STATISTICS=CORR
/MISSING=ANALYSIS.
```

Rezultati postupka dati su u tabeli:

Correlations					
Kontrola promenljivih			ukupno SDSUS	ukupan subjektivno doživljen stres	ukupna društvena poželjnost
-none-a	ukupna SDSUS	Correlation	1,000	-,581	,295
		Significance (2-tailed)	.	,000	000
		df	0	424	425
	ukupan subjektivno doživljen stres	Correlation	-,581	1,000	-,228
		Significance (2-tailed)	,000	.	,000
		df	424	0	426
	ukupna društvena poželjnost	Correlation	,295	-,228	1,000
		Significance (2-tailed)	,000	,000	.
		df	425	426	0
ukupna društvena poželjnost	ukupna SDSUS	Correlation	1,000	-,552	
		Significance (2-tailed)	.	,000	
		df	0	423	
	ukupan subjektivno doživljen stres	Correlation	-,552	1,000	
		Significance (2-tailed)	,000	.	
		df	423	0	

^a. Cells contain zero-order (Pearson) correlations.

Tumačenje rezultata delimične korelacije

SPSS-ova tabela rezultata ima dva dela:

1. U gornjoj polovini tabele je uobičajena matrica Pirsonove linearne korelacije dve promenljive od interesa (npr. subjektivno doživljene samokontrole i subjektivno doživljenog stresa) kada *nije* uklonjen uticaj treće promenljive. U tom slučaju, korelacija iznosi -0,581. Reč *none* u levoj koloni kazuje da u tom delu tabele nije bilo promenljive čiji bi se uticaj kontrolisao, tj. uklonio.
2. U donjoj polovini tabele ponovljene su iste analize korelacije, ali uz uklanjanje uticaja remetilačke promenljive (npr. društvene poželjnosti). U tom slučaju, nova delimična korelacija iznosi -0,552. Uvek bi trebalo da uporedite ta dva skupa koeficijenata korelacije i utvrdite da li je uklanjanje uticaja dodatne promenljive imalo posledice po vezu između dve ispitivane promenljive. U ovom primeru, jačina korelacije se malo smanjila (od -0,581 na -0,552). Time je dokazano da zapažena veza između subjektivno doživljene samokontrole i subjektivno doživljenog stresa nije samo posledica davanja društveno poželjnih odgovora.

Predstavljavanje rezultata delimične korelacije

Rezultati ove analize mogli bi se predstaviti ovako:

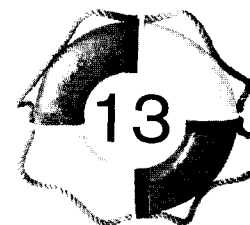
Za istraživanje veze između subjektivno doživljene samokontrole unutrašnjih stanja (merene na skali SDSUS) i subjektivno doživljenog stresa (merenog na skali USUDSTRES) upotrebljena je delimična korelacija, pri čemu je uklonjen uticaj rezultata na Marlou–Kraunovoj skali društveno poželjnih odgovora. Preliminarnim analizama je dokazano da pretpostavke normalnosti, linearnosti i homogenosti varijanse nisu narušene. Nakon uklanjanja uticaja društveno poželjnih odgovora, utvrđena je jaka negativna delimična korelacija između subjektivno doživljene samokontrole unutrašnjih stanja i subjektivno doživljenog stresa, $r = -0,55$, $n = 425$, $p < ,0005$, pri čemu visoke nivoe subjektivno doživljene samokontrole prate niski nivoi subjektivno doživljenog stresa. Poređenje sa izračunatom korelacijom nultog reda ($r = -0,58$) pokazuje da uklanjanje uticaja društveno poželjnih odgovora vrlo malo utiče na jačinu veze između te dve promenljive.

Dodatna vežba

Zdravstvo

Datoteka s podacima: sleep3ED.sav. Pojediniosti o datoteci videti u dodatku.

1. Izračunati jačinu korelacije između rezultata na skali Sleepiness and Associated Sensations (*totSAS*) i uticaja problema sa spavanjem na opšte raspoloženje (*impact6*), uz uklanjanje remetilačkog uticaja *starosti*. Uporedite (Pirsonovu) korelaciju nultog reda i koeficijent delimične korelacije. Da li uklanjanje uticaja *starosti* ima znatnije posledice?



Višestruka regresija

U ovom poglavlju ukratko ću izložiti kako se pomoću SPSS-a za Windows rade analize višestruke regresije. Pošto će izlaganje biti *veoma* pojednostavljeno, trebalo bi da više pročitate o višestrukoj regresiji pre nego što je upotrebite u istraživanju. Dobar izvor informacija je poglavlje 5 u knjizi autorki Tabachnick i Fidell (2007), koje obuhvata teoriju, različite vrste analiza višestruke regresije i pretpostavke na kojima ona počiva.

Višestruka regresija nije samo jedna tehnika već porodica tehnika pomoću kojih se može istražiti veza jedne neprekidne zavisne promenljive i više nezavisnih promenljivih ili prediktora (obično neprekidnih). Višestruka regresija se zasniva na korelaciji (obrađenoj u poglavlju 11), ali omogućuje sofisticiranije istraživanje međusobnih veza skupa promenljivih. Zato je idealna za istraživanje složenijih pitanja iz stvarnog života, pre nego onih laboratorijskih. Međutim, nemojte se nadati da će se odgovori nekim čudom pojaviti čim u višestruku regresiju ubacite promenljive. Za ovu analizu trebalo bi da imate čvrst teorijski ili konceptualni razlog, posebno za redosled promenljivih koje ulaze u jednačinu. Ne upotrebljavajte višestruku regresiju tek tako.

Višestruka regresija može odgovoriti na razna istraživačka pitanja. Ona kazuje koliko dobro određeni skup promenljivih predviđa konkretan ishod. Na primer, recimo da istražujete koliko se dobro pomoću skupa podskala na testu inteligencije može predvideti efikasnost obavljanja određenog zadatka. Višestruka regresija će dati ocenu modela kao celine (svih podskala) i relativan doprinos svih promenljivih (podskala pojedinačno) od kojih se model sastoji. Uz to, višestrukom regresijom možete ispitati da li dodavanje određene promenljive (npr. motivacije) doprinosi prediktivnoj sposobnosti modela, tj. da li poboljšava model koji već sadrži druge promenljive. Višestruka regresija se može upotrebiti i za statističko uklanjanje uticaja jedne ili više promenljivih na prediktivnu sposobnost modela. Među glavnim vrstama istraživačkih pitanja na koja višestruka regresija može odgovoriti jesu i ova:

- koliko dobro *skup promenljivih* može da predvidi određeni ishod;
- *koja promenljiva* u skupu je najbolji prediktor određenog ishoda; i
- da li određena prediktorska promenljiva i dalje dobro predviđa određeni ishod nakon uklanjanja uticaja druge promenljive (npr. sklonosti davanju društveno poželjnih odgovora).

Glavne vrste višestruke regresije

U zavisnosti od prirode pitanja na koje tražite odgovor, možete upotrebiti više vrsta višestrukih regresija. Tri glavne vrste višestrukih regresija su:

- standardna ili istovremena (simultana);
- hijerarhijska ili sekvencijalna; i
- postepena (engl. *stepwise*).

U statističkoj literaturi autori često različitim terminima nazivaju te glavne vrste višestruke regresije, što jako zbunjuje čak i iskusne istraživače, a o početnicima da i ne govorimo!

Standardna višestruka regresija

U standardnoj višestrukoj regresiji, sve nezavisne (ili prediktorske) promenljive se istovremeno unose u jednačinu. Ocenjuje se prediktivna moć svake nezavisne promenljive, tj. meri se koliko bi ona poboljšala model koji se sastoji od drugih nezavisnih promenljivih. Ova vrsta višestruke regresije se najčešće koristi. Nju biste upotreбили da imate skup promenljivih (npr. raznih skala za merenje svojstava ličnosti) i da istražujete koliki deo varijanse zavisne promenljive (npr. anksioznosti) one mogu da objasne kao grupa ili blok. Saznali biste i koliki deo jedinstvene varijanse zavisne promenljive objašnjava svaka od nezavisnih promenljivih pojedinačno.

Hijerarhijska višestruka regresija

U hijerarhijskoj (ili sekvencijalnoj) regresiji, nezavisne promenljive se unose u jednačinu redosledom koji na osnovu teorijskih razloga zadaje istraživač. Promenljive ili skupovi promenljivih unose se u koracima (ili blokovima), pri čemu se ocenjuje doprinos svake nezavisne promenljive predikciji zavisne promenljive, ali tako da se ukloni uticaj svih prethodnih promenljivih. Primera radi, nakon što uklonite uticaj starosti, istražujete koliko dobro optimizam predviđa zadovoljstvo životom; starost biste uneli u bloku 1, a optimizam u bloku 2. Kada su unete sve promenljive, ocenjuje se sposobnost celog modela da predvidi zavisno obeležje i relativan doprinos svakog bloka promenljivih.

Postepena višestruka regresija

U postepenoj regresiji, istraživač daje SPSS-u listu nezavisnih promenljivih i pušta program da, na osnovu skupa statističkih kriterijuma, izabere promenljive koje će uneti u jednačinu i redosled po kojem će ih unositi. Postoje tri verzije ovog pristupa: izbor unapred (engl. *forward selection*), brisanje unazad (engl. *backward deletion*) i postepena regresija (engl. *stepwise regression*). Ima nekoliko problema sa ovim tehnikama, te i sporenja u literaturi o tome šta je ispravna upotreba (a šta zloupotreba) svake od njih. Pre

nego sto ih primenite, preporučujem da pročitate šta je o njima napisano (videti Tabachnick & Fidell, 2007, str. 138). Neophodno je da shvatite o čemu se radi, kako izabrati odgovarajuće promenljive i kako tumačiti rezultate.

Pretpostavke na kojima se zasniva višestruka regresija

Višestruka regresija spada u pipavije statističke tehnike. Zasnovana je na više pretpostavki o podacima i ne prашta ako su narušene. Ne može se upotrebljavati na malim uzorcima, niti kada je raspodela rezultata veoma asimetrična! Sledeći sažetak glavnih pretpostavki preuzet je 5. poglavlja knjige koju su napisale Tabachnick i Fidell (2007). Ne bi bilo loše da pročitate to poglavlje pre nego što nastavite analizu. Trebalo bi da pročitate i uvod u četvrti deo ove knjige, koji obrađuje osnove korelacije, i pogledate spisak preporučenih referenci na kraju knjige. SPSS procedure za ispitivanje tih pretpostavki podrobnije su razmotrene u primerima u nastavku.

Veličina uzorka

Ovde se radi o sposobnosti uopštavanja, tj. rezultat dobijen na malom uzorku ne može se uopštiti (ponovo dobiti) na drugim uzorcima. Koliko vredi rezultat koji važi za samo jedan uzorak? Stoga treba znati koliko slučajeva ili subjekata je potrebno da bi statistički pokazatelji izračunati za uzorak važili za celu populaciju iz koje je uzorak uzet. Razni autori daju različite smernice za broj slučajeva (opservacija) potrebnih za višestruku regresiju. Stevens (1996, str. 72) preporučuje da je 'za istraživanja u društvenim naukama potrebno oko 15 subjekata po prediktoru za pouzdanu jednačinu'. Tabachnick i Fidell (2007, str. 123) daju formulu za izračunavanje veličine uzorka, koja uzima u obzir broj nezavisnih promenljivih: $N > 50 + 8m$ (gde je m = broj nezavisnih promenljivih). Za pet nezavisnih promenljivih bilo bi potrebno analizirati 90 slučajeva. Kada je zavisna promenljiva asimetrična, potrebno je još više slučajeva. Za postepenu regresiju potrebno je po 40 analiziranih slučajeva za svaku nezavisnu promenljivu.

Multikolinearnost i singularnost

Ovde se radi o odnosima između nezavisnih promenljivih. Multikolinearnost postoji kada su nezavisne promenljive jako korelirane ($r=0,9$ i više). Singularnost nastaje kada je jedna „nezavisna“ promenljiva zapravo kombinacija drugih nezavisnih promenljivih, tj. kada uopšte nije nezavisna! To bi se desilo npr. kada bi model sadržao i vrednosti na podskalama i ukupnu vrednost na skali. Višestruka regresija ne voli ni multikolinearnost ni singularnost, a ta svojstva sigurno ne poboljšavaju regresioni model, pa uvek proverite da li promenljive imaju takva svojstva pre nego što počnete analizu.

Netipične tačke

Višestruka regresija je veoma osetljiva na netipične tačke (veoma velike ili veoma male vrednosti rezultata). Pronalaženje i obrada netipičnih tačaka trebalo bi da budu deo početnog procesa čišćenja podataka (videti poglavlje 6). To treba uraditi za sve promenljive, i zavisne i nezavisne, koje ćete upotrebljavati u regresionoj analizi. Netipične tačke treba ukloniti iz skupa podataka ili im dodeliti vrednost koja jeste velika, ali se ne razlikuje previše od ostalih rezultata. Program za višestruku regresiju ima i posebne procedure za otkrivanje netipičnih tačaka. Netipične tačke zavisne promenljive mogu se prepoznati na standardizovanom dijagramu reziduala, koji SPSS ume da nacrtati. (To je opisano u primeru u nastavku.). Tabachnick i Fidell (2007, str. 128) definišu netipične tačke kao one čije su standardizovane rezidualne vrednosti iznad približno 3,3 (odnosno manje od -3,3).

Normalnost, linearnost, homogenost varijanse, nezavisnost reziduala

Svi ovi pojmovi se odnose na razne aspekte raspodele rezultata i prirodu pripadne veze između promenljivih. Te pretpostavke se mogu proveriti na dijagramu rasturanja reziduala, koji SPSS generiše u sklopu procedure višestruke regresije. Reziduali su razlike između dobijene i predviđene vrednosti zavisne promenljive (ZP). Na dijagramu rasturanja reziduala može se proveriti:

- *normalnost*: reziduali treba da su normalno raspodeljeni oko predviđenih vrednosti ZP;
- *linearnost*: reziduali treba da imaju linearnu vezu s predviđenim vrednostima ZP, tj. njihov dijagram treba da je približno prava linija;
- *homogenost varijanse*: varijansa reziduala oko predviđenih vrednosti ZP treba da je približno jednaka za sve predviđene vrednosti.

Reziduali na dijagramima rasturanja koje crta SPSS protumačeni su u nastavku poglavlja; detaljniju raspravu ove prilično složene teme pročitajte u 5. poglavlju knjige koju su napisale Tabachnick i Fidell (2007). Na kraju te knjige data je i preporučena literatura za dalje ovladavanje višestrukom regresijom.

Objašnjenje primera

Kao ilustraciju upotrebe višestruke regresije koristiću niz primera uzetih iz datoteke s podacima `survey3ED.sav`, sa prateće Web lokaciji knjige (videti str. ix). Ta je anketa bila projektovana za istraživanje faktora koji utiču na psihološko prilagođenje i opšte raspoloženje ispitanika. (Sve pojedinosti te studije date su u dodatku.) U izloženom primeru višestruke regresije istražiću uticaj subjektivno doživljenog stepena upravljanja ispitanika njihovim životima, na njihov subjektivno doživljen nivo stresa. Literatura u toj oblasti kazuje da stres ređe trpe ljudi koji smatraju da upravljaju svojim životima. U upitniku su date dve različite mere upravljanja (reference za te skale videti u dodatku). To su skala Mastery, koja meri stepen do kojeg ljudi smatraju da

upravljaju događajima u svom životu; i skala Perceived Control of Internal States Scale (PCOISS), koja meri stepen do kojeg ljudi smatraju da upravljaju svojim unutrašnjim stanjima (osećanja, misli i fizičke reakcije).

U ovom primeru istražujem koliko dobro skale Mastery i PCOISS predviđaju rezultate na skali subjektivno doživljenog stresa. Sledi opis promenljivih upotrebljenih u primerima iz ovog poglavlja. Bilo bi dobro da prorate te primere na računaru uz korišćenje navedene datoteke s podacima. Učenje uz rad je uvek bolje od pukog čitanja knjige. Slobodno se poigravajte s datotekom s podacima – promenljive upotrebljene u primerima zamenite drugim promenljivama. Pogledajte koje ćete rezultate dobiti i pokušajte da ih protumačite.

Ime datoteke: `survey3ED.sav`

Promenljive:

- Total Perceived Stress (tpstress): ukupna vrednost na skali Perceived Stress (subjektivno doživljen stres). Velike vrednosti pokazuju visok nivo stresa.
- Total Perceived Control of Internal States (tpcoiss): ukupna vrednost na skali Perceived Control of Internal States (subjektivno doživljena samokontrola unutrašnjih stanja). Velike vrednosti pokazuju veću samokontrolu unutrašnjih stanja.
- Total Mastery (tmast): ukupna vrednost na skali Mastery (ovladavanje). Velike vrednosti pokazuju veći stepen do kojeg ljudi smatraju da upravljaju događajima i okolnostima u svom životu.
- Total Social Desirability (tmarlow): ukupna vrednost na Marlow-Kraunovoj skali Social Desirability (društvena poželjnost), koja meri stepen do kojeg ljudi pokušavaju da se predstave u pozitivnom svetlu.
- Age: starost u godinama.

Najbolje se uči radeći, naročito kada treba uvežbati ovako složene tehnike. Praksom ćete izgraditi samopouzdanje za budućnost, kada ćete sami analizirati sopstvene podatke. Naredni primeri obrađuju samo standardnu višestruku regresiju (Standard Multiple Regression) i hijerarhijsku regresiju (Hierarchical Regression). Zbog kritika upućenih tehnikama postepene višestruke regresije (Stepwise Multiple Regression), one ovde neće biti ilustrovane. Ukoliko insistirate na upotrebi tih tehnika (uprkos upozorenjima!), predlažem da pročitate odgovarajući deo knjige autorki Tabachnick i Fidell (2007) ili nekih drugih naprednijih knjiga o multivarijacionoj analizi.

Kratak pregled višestruke regresije

Primer istraživačkih pitanja:

1. Koliko dobro dve date mere nivoa upravljanja životom (mastery, PCOISS) predviđaju subjektivno doživljen stres? Koliko se deo varijanse u vrednostima subjektivno doživljenog stresa može objasniti rezultatima na te dve skale?

2. Koje svojstvo bolje predviđa subjektivno doživljen stres: upravljanje spoljnim događajima (skala Mastery) ili upravljanje unutrašnjim stanjima (PCOISS)?
3. Ako uklonimo mogući uticaj starosti i davanja društveno poželjnih odgovora, može li ovaj skup promenljivih još uvek da predvidi značajan deo varijanse subjektivno doživljenog stresa?

Šta vam treba:

- jedna neprekidna zavisna promenljiva (ukupan subjektivno doživljen stres); i
- dve ili više neprekidnih nezavisnih promenljivih (mastery, PCOISS). (Mogu se upotrebiti i dihotomne nezavisne promenljive, npr. muškarci=1, žene=2.)

Šta se postiže: Višestruka regresija kazuje koliki je deo varijanse zavisne promenljive objašnjen varijansom nezavisnih promenljivih. Daje i pokazatelj relativnog doprinosa svake nezavisne promenljive. Testovi omogućuju utvrđivanje statističke značajnosti rezultata – kako celog modela tako i nezavisnih promenljivih pojedinačno.

Pretpostavke: Osnovne pretpostavke na kojima počiva višestruka regresija opisane su u prethodnom delu ovog poglavlja. Neke od tih pretpostavki mogu se proveriti u sklopu analiza višestruke regresije. (One će biti ilustrovane u narednim primerima.)

Standardna višestruka regresija

U ovom primeru odgovorićemo na dva pitanja:

Pitanje 1: Koliko dobro dve mere nivoa upravljanja životom (mastery, PCOISS) predviđaju subjektivno doživljen stres? Koliki se deo varijanse u rezultatima merenja subjektivno doživljenog stresa može objasniti vrednostima rezultata na te dve skale?

Pitanje 2: Koje svojstvo bolje predviđa subjektivno doživljen stres: upravljanje spoljnim događajima (skala Mastery) ili upravljanje unutrašnjim stanjima (PCOISS)?

Za istraživanje ovih pitanja upotrebiću standardnu višestruku regresiju. To znači da se sve nezavisne promenljive istovremeno unose u jednačinu. Rezultati će pokazati sposobnost ovog skupa promenljivih da predvidi nivo stresa; i kazaće nam takođe koliki deo jedinstvene varijanse zavisne promenljive objašnjava svaka od nezavisnih promenljivih (mastery, PCOISS) pojedinačno, *pored doprinosa svih* ostalih nezavisnih promenljivih iz skupa. Za svaki postupak date su i odgovarajuće SPSS komande. Više informacija o upotrebi Syntax Editora za pamćenje i snimanje SPSS komandi videti u poglavlju 3.

Pre nego što započnete sledeći postupak, otvorite meni **Edit** i izaberite u njemu stavku **Options**; treba da je potvrđeno polje **No scientific notation for small numbers in tables**.

Postupak za standardnu višestruku regresiju

1. Otvorite meni **Analyze**, izaberite stavku **Regression**, pa pritisnite **Linear**.
2. Pritisnite neprekidnu zavisnu promenljivu (npr. total perceived stress: tpstress) i prebacite je u polje **Dependent**.
3. Pritisnite nezavisne promenljive (total mastery: tmast; total PCOISS: tpcoiss) i prebacite ih u polje **Independent**.
4. U odeljku **Method** treba da je izabrana opcija **Enter**. (Time se bira standardna višestruka regresija.)
5. Pritisnite dugme **Statistics**.
 - Potvrdite polja **Estimates, Confidence Intervals, Model fit, Descriptives, Part and partial correlations** i **Collinearity diagnostics**.
 - U odeljku **Residuals** potvrdite polja **Casewise diagnostics** i **Outliers outside 3 standard deviations**. Pritisnite dugme **Continue**.
6. Pritisnite dugme **Options**. U odeljku **Missing Values** pritisnite **Exclude cases pairwise**. Pritisnite dugme **Continue**.
7. Pritisnite dugme **Plots**.
 - Pritisnite ***ZRESID** i dugme sa strelicom da biste je prebacili u polje **Y**.
 - Pritisnite ***ZPRED** i dugme sa strelicom da biste je prebacili u polje **X**.
 - U odeljku **Standardized Residual Plots**, potvrdite polje **Normal probability plot**. Pritisnite dugme **Continue**.
8. Pritisnite dugme **Save**.
 - U odeljku **Distances**, potvrdite polja **Mahalanobis** i **Cook's**.
 - Pritisnite **Continue** i zatim **OK** (ili dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Prethodni postupak generiše ovu komandu:

```
REGRESSION
/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING PAIRWISE
/STATISTICS COEFF OUTS R ANOVA COLLIN TOL ZPP
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT tpstress
/METHOD=ENTER tmast tpcoiss
/SCATTERPLOT=(*ZRESID *ZPRED)
/RESIDUALS NORM(ZRESID)
/CASEWISE PLOT(ZRESID) OUTLIERS(3)
/SAVE MAHAL COOK.
```

Evo rezultata opisanog postupka.

Correlations

		Total perceived stress	Total Mastery	Total PCOISS
Pearson Correlation	Total perceived stress	1,000	-,612	-,581
	Total Mastery	-,612	1,000	,521
	Total PCOISS	-,581	,521	1,000
Sig. (1-tailed)	Total perceived stress		,000	,000
	Total Mastery	,000		,000
	Total PCOISS	,000	,000	
N	Total perceived stress	433	433	426
	Total Mastery	433	436	429
	Total PCOISS	426	429	431

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,684 ^a	,468	,466	4,27

a. Predictors: (Constant), Total PCOISS, Total Mastery

b. Dependent Variable: Total perceived stress

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6806,728	2	3403,364	186,341	,000 ^a
	Residual	7725,756	423	18,264		
	Total	14532,484	425			

a. Predictors: (Constant), Total PCOISS, Total Mastery

b. Dependent Variable: Total perceived stress

Coefficients^a

		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero order	Partial	Part	Tolerance	VIF
1	(Constant)	50,971	1,273		40,035	,000	48,469	53,474					
	tmast												
	total	-,625	,061	-,424	-10,222	,000	-,745	-,505	-,612	-,445	-,362	,729	1,372
	mastery												
	tpcoiss												
	total	-,175	,020	-,360	-8,660	,000	-,215	-,136	-,581	-,388	-,307	,729	1,372
	PCOISS												

a. Dependent Variable: tpstress total perceived stress

Casewise Diagnostics^a

Case Number	Std. Residual	Total perceived stress	Predicted Value	Residual
152	-3,473	14	28,84	-14,84

a. Dependent Variable: Total perceived stress

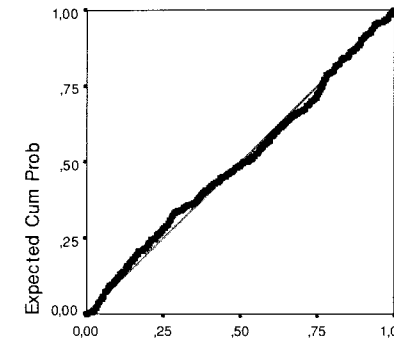
Residuals Statistics^a

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	18,03	41,31	26,74	4,001	429
Std. Predicted Value	-2,174	3,644	,002	1,000	429
Standard Error of Predicted Value	,207	,800	,341	,111	429
Adjusted Predicted Value	18,04	41,39	26,75	4,009	426
Residual	-14,849	12,612	-,002	4,268	426
Std. Residual	-3,475	2,951	,000	,999	426
Stud. Residual	-3,514	2,969	,000	1,003	426
Deleted Residual	-15,190	12,765	-,001	4,306	426
Stud. Deleted Residual	-3,562	2,997	-,001	1,006	426
Mahal. Distance	,004	13,897	1,993	2,234	429
Cook's Distance	,000	,094	,003	,008	426
Centered Leverage Value	,000	,033	,005	,005	429

a. Dependent Variable: Total perceived stress

Normal P-P Plot of Regression Standardized Residual

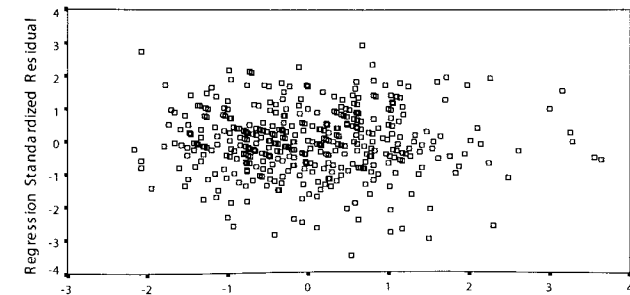
Dependent Variable : Total perceived stress



Observed Cum Prob

Scatterplot

Dependent Variable: Total perceived stress



Regression Standardized Predicted Value

Tumačenje rezultata standardne višestruke regresije

Kao i većina ostalih SPSS procedura, i regresija generiše gomilu prilično nerazumljivih brojeva. Da biste shvatili njihov smisao i značenje, vodim vas na turu kroz deo rezultata dobijenih kao odgovor na Pitanje 1.

Korak 1: provera važenja pretpostavki

Multikolinearnost

Korelacije između promenljivih u modelu date su u tabeli **Correlations**. Proverite da li nezavisne promenljive pokazuju makar i slabu vezu sa zavisnom promenljivom (bilo bi poželjno da te korelacije budu iznad 0,3). U ovom slučaju, obe skale (Total Mastery i Total PCOISS) znatno koreliraju s promenljivom Total Perceived Stress (-0,61 odnosno -0,58). Takođe proverite da nije prevelika neka od korelacija između nezavisnih promenljivih. U istu analizu verovatno ne bi trebalo da uključujete promenljive čija linearna korelacija iznosi 0,7 ili više. Kada se nađete u toj situaciji, ili izostavite jednu od tih promenljivih ili od vrednosti dve jako korelirane promenljive napravite jednu zajedničku promenljivu. U ovom primeru korelacija iznosi 0,52, što je manje od 0,7; zato ćemo zadržati sve promenljive.

Kao deo procedure višestruke regresije, SPSS obavlja i 'dijagnostiku kolinearnosti' promenljivih. Ona često ukazuje na probleme s multikolinearnošću koji se možda ne vide u matrici korelacija. Rezultati te dijagnostike dati su u tabeli **Coefficients**; to su vrednosti **Tolerance** i **VIF**. Tolerance pokazuje koliki deo varijanse date nezavisne promenljive nije objašnjen varijansama drugih nezavisnih promenljivih u modelu i računa se po formuli $1-R^2$ za svaku promenljivu. Kada je ta vrednost jako mala (manja od 0,10), to ukazuje na ogromnu korelaciju s drugim promenljivama, dakle na multikolinearnost. Ona druga vrednost, VIF (*Variance inflation factor*, faktor povećanja varijanse), recipročna je vrednost veličine Tolerance (1 podeljeno s Tolerance). Ovde bi bile problematične vrednosti VIF iznad 10 zato što ukazuju na multikolinearnost.

Navela sam uobičajene presečne tačke za utvrđivanje prisustva multikolinearnosti (vrednost Tolerance ispod 0,10 ili vrednost VIF iznad 10). Međutim, pri tim vrednostima još uvek su moguće veoma velike korelacije između nezavisnih promenljivih (iznad 0,9), pa bi njih trebalo da shvatite samo kao upozorenje i razlog da bolje pogledate matricu korelacija. U ovom primeru, Tolerance za svaku nezavisnu promenljivu iznosi 0,729, što nije manje od 0,10; dakle, pretpostavka o nepostojanju multikolinearnosti nije narušena. Taj zaključak podržava i vrednost VIF koja iznosi 1,372, što je mnogo ispod

presečne tačke 10. Ti nas rezultati ne iznenađuju, pošto je koeficijent Pirsoneve korelacije te dve nezavisne promenljive samo 0,52 (videti tabelu **Correlations**). Kada u svojim rezultatima uočite prekoračenje navedenih graničnih vrednosti, trebalo bi da iz modela izbacite jednu od jako koreliranih nezavisnih promenljivih.

Netipične tačke, normalnost, linearnost, homogenost varijanse, nezavisnost reziduala

Jedan od načina da proverite nabrojane pretpostavke jeste pregled dijagrama **Normal Probability Plot (P-P) of the Regression Standardised Residual** i **Scatterplot**, čije crtanje smo zahtevali kao deo analize. Oni su dati na kraju SPSS-ovih rezultata. Na dijagramu **Normal P-P Plot** nadamo se da tačke leže u približno pravougaonoj dijagonalnoj liniji od donjeg levog do gornjeg desnog ugla dijagrama. Ona ukazuje na to da nema velikih odstupanja od normalnosti. Na dijagramu rasturanja standardizovanih reziduala **Scatterplot** (drugi po redu), nadamo se da su reziduali približno pravougaono raspoređeni i da je većina rezultata nagomilana u centru (oko tačke 0). Ne bi nam se svideo jasno sistematski ili geometrijski oblik raspodele reziduala (npr. da reziduali čine neku jasnu krivu liniju ili da su viši na jednoj strani nego na drugoj). Odstupanja od oblika centralnog pravougaonika kazuju da je narušena neka od polaznih pretpostavki. Kada to utvrdite za svoje podatke, predlažem da u knjizi koju su napisale Tabachnick i Fidell (2007, str. 125) i pročitate ceo opis tumačenja dijagrama reziduala i kako se ocenjuju posledice narušavanja polaznih pretpostavki po analizu.

Na dijagramu rasturanja standardizovanih reziduala **Scatterplot** može se utvrditi i ima li netipičnih tačaka. Tabachnick i Fidell (2007) definišu netipične tačke kao slučajeve koji imaju standardizovan rezidual (prikazan na dijagramu rasturanja) veći od 3,3 ili manji od -3,3. U velikim uzorcima često se nađe mnogo netipičnih reziduala. Kada nađete tek nekoliko njih, verovatno ne treba da preduzimate ništa u vezi s tim.

Netipične tačke se mogu pronaći i pregledom Mahalanobisovih udaljenosti reziduala, koje izračunava program za višestruku regresiju. One nisu ispisane na ekranu računara, već su stavljene na kraj datoteke s podacima kao dodatna promenljiva (Mah_1). Da biste prepoznali netipične tačke, najpre izračunajte kritičnu vrednost hi-kvadrat, pri čemu je broj stepeni slobode jednak broju nezavisnih promenljivih. Potpunu tabelu tih vrednosti sadrži svaki udžbenik iz statistike (videti Tabachnick & Fidell, 2007, tabela C.4). Tabachnick i Fidell predlažu da se upotrebljava nivo alfa od 0,001. U skladu sa smernicama koje daju Tabachnick i Fidell (2007), navodim nekoliko kritičnih vrednosti u tabeli 13.1.

Tabela 13.1: Kritične vrednosti za izračunavanje Mahalanobisove udaljenosti

Broj nezav. promenljivih	Kritična vrednost	Broj nezav. promenljivih	Kritična vrednost	Broj nezav. promenljivih	Kritična vrednost
2	13,82	4	18,47	6	22,46
3	16,27	5	20,52	7	24,32

Izvor: preuzeto i adaptirano iz tabele u knjizi Tabacknik i Fidell (1996), prvobitno objavljene u knjizi Pearson, E.S. i Hartley, H.O. (Eds) (1958). *Biometrika tables for statisticians* (1. tom, 2. izdanje). Njujork: Cambridge University Press.

Da biste mogli koristiti ovu tabelu, trebalo bi da:

- utvrdite koliko će nezavisnih promenljivih biti uključeno u višestruku regresiju;
- pronađete taj broj u jednoj od kolona tabele sa zaglavljem Broj nezav. promenljivih; i
- u susednoj ćeliji desno pročitate odgovarajuću kritičnu vrednost.

U ovom primeru imam dve nezavisne promenljive, pa kritična vrednost iznosi 13,82. Kako biste utvrdili da li i jedan ispitivani slučaj ima Mahalanobisovu udaljenost veću od tog broja, pronađite tabelu **Residuals Statistics**, pogledajte u njoj ćeliju na preseku reda **Mahal. Distance** i kolone **Maximum**. U mojoj datoteci s podacima, najveća vrednost te udaljenosti je 13,89, što neznatno premašuje kritičnu vrednost. Treba da pronađemo slučaj koji sadrži tu vrednost.

Da bismo je pronašli, vraćamo se u prozor **Data Editor**, gde otvaramo meni **Data** i u njemu biramo stavku **Sort Cases**. Sortiranje treba obaviti po novoj promenljivoj smeštenoj na dno tabele (Mahalanobis Distance, MAH-1) u opadajućem (**Descending**) redosledu. U prozoru **Data View**, slučaj s najvećom vrednošću promenljive Mahalanobis Distance biće nakon sortiranja na vrhu datoteke s podacima (ID=66). Za ovako velik uzorak nije neobično da se pojavi nekoliko netipičnih tačaka, pa se neću previše uzbuđivati zbog ovog slučaja za koji je Mahalanobisova udaljenost tek neznatno premašila kritičnu vrednost. Kada u podacima pronađete slučajeve sa Mahalanobisovim udaljenostima mnogo većim od kritične vrednosti, trebalo bi da ih uklonite iz analiza.

Još neke informacije o netipičnim tačkama navedene su u tabeli **Casewise Diagnostics** SPSS-ovih rezultata. To su slučajevi sa standardizovanim vrednostima reziduala iznad 3,0 odnosno ispod -3,0. U normalno raspodeljenom uzorku, samo 1 procenat rezultata pada izvan tog opsega. U ovom uzorku smo našli jedan slučaj (broj 165) sa standardizovanim rezidualom -3,48. U tabeli **Casewise Diagnostics** vidimo da je ta osoba navela 14 kao vrednost svog ukupnog subjektivno doživljenog stresa, dok je naš model predvideo vrednost 28,85. Jasno je da naš model nije dobro predvideo rezultat za tu osobu; ona oseća mnogo manji stres nego što smo predvideli.

Kako bismo se uverili da ovaj čudan slučaj ne utiče previše na rezultate celog modela, pogledaćemo vrednost udaljenosti **Cook's Distance** dat pri dnu tabele **Residuals Statistics**. Tabachnick i Fidell (2007, str. 75) tvrde da problem mogu biti slučajevi za koje ta udaljenost premašuje vrednost 1. U našem primeru, najveća (**Maximum**) vrednost udaljenosti **Cook's Distance** iznosi 0,094, pa ne očekujemo nikakve probleme. Kada u svojim podacima dobijete najveću vrednost te udaljenosti preko 1, vratite se u datoteku s podacima i sortirajte slučajeve po novoj promenljivoj koju je SPSS napravio na kraju datoteke (**Cook's Distance COO_1**). Proverite sve slučajeve čija udaljenost premašuje 1 i ako treba, uklonite ih iz analiza.

Korak 2: vrednovanje modela

U SPSS-ovim rezultatima pronađite tabelu **Model Summary** i u njoj broj ispod zaglavlja **R Square**. Kao što se sećate, r^2 je koeficijent determinacije koji kazuje koliki deo varijanse zavisne promenljive (subjektivno doživljenog stresa) objašnjava model (koji obuhvata promenljive **Total Mastery** i **Total PCOISS**). U ovom slučaju, r^2 je jednako 0,468. Izraženo u procentima (pomnoženo sa 100 pomeranjem decimalne tačke za dva mesta udesno), to znači da naš model (koji obuhvata **Mastery** i **PCOISS**) objašnjava 46,8 procenata varijanse subjektivno doživljenog stresa. To je sasvim pristojan rezultat (naravno kada se uporedi s nekim rezultatima objavljenim u časopisima!).

Možda ste primetili da se u SPSS-ovim rezultatima navodi i veličina **Adjusted R Square**, tj. korigovano r^2 . Problem je to što je r^2 previše optimistička procena stvarne vrednosti koeficijenta determinacije u populaciji kada se izračuna na malom uzorku (Tabachnick & Fidell, 2007). Pokazatelj **Adjusted R square** koriguje tu vrednost i daje bolju procenu stvarne vrednosti koeficijenta determinacije u populaciji. Kada imate mali uzorak, izračunajte i navedite tu vrednost, a ne običan r^2 (**R Square**). Da biste ocenili statističku značajnost tog pokazatelja, pronađite tabelu **ANOVA**. Tu su rezultati testova nulte hipoteze da je r^2 u populaciji jednako 0. Model u ovom primeru dostiže statističku značajnost (**Sig.** = ,000; to zapravo znači da je $p < 0,0005$).

Korak 3: vrednovanje svake nezavisne promenljive

Sledeće što želimo da znamo jeste koliko je koja promenljiva u modelu doprinela predikciji zavisne promenljive. Taj podatak ćemo naći u tabeli **Coefficients**. Pogledajte kolonu **Beta** u odeljku **Standardised Coefficients**. Da biste mogli da poredite razne promenljive, gledajte standardizovane koeficijente, ne one nestandardizovane. Standardizovan znači da su vrednosti promenljivih konvertovane na istu skalu kako bi se mogle porediti. Za sastavljanje regresione jednačine upotrebili biste nestandardizovane koeficijente u koloni **B**.

Postupak za hijerarhijsku višestruku regresiju

1. U glavnom meniju na vrhu prozora pritisnite **Analyze**, zatim **Regression**, pa **Linear**.
2. Izaberite neprekidnu zavisnu promenljivu (npr. total perceived stress) i prebacite je u polje **Dependent**.
3. Prebacite promenljive čiji uticaj želite da uklonite u polje **Independent** (npr. age, total social desirability). One će biti unete u analizu u prvom bloku promenljivih (Block 1 of 1).
4. Pritisnite dugme **Next**. Prikazaće se drugo polje za nezavisne promenljive, u koje treba uneti drugi blok promenljivih (trebalo bi da piše Block 2 of 2).
5. Izaberite sledeći blok nezavisnih promenljivih (npr. total mastery, Total PCOISS).
6. U polju **Method** treba da je potvrđena podrazumevana opcija (**Enter**).
7. Pritisnite dugme **Statistics**. Potvrdite polja **Estimates**, **Model fit**, **R squared change**, **Descriptives**, **Part and partial correlations** i **Collinearity diagnostics**. Pritisnite dugme **Continue**.
8. Pritisnite dugme **Options**. U odeljku **Missing Values** pritisnite **Exclude cases pairwise**. Pritisnite dugme **Continue**.
9. Pritisnite dugme **Save**. Pritisnite opcije **Mahalanobis** i **Cook's**. Pritisnite **Continue** i zatim **OK** (ili dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Prethodni postupak generiše sledeću komandu:

```
REGRESSION
/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING PAIRWISE
/STATISTICS COEFF OUTS R ANOVA COLLIN TOL CHANGE ZPP
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT tpstress
/METHOD=ENTER tmarlow age /METHOD=ENTER tmast tpcoiss
/SCATTERPLOT=(*ZRESID *ZPRED )
/RESIDUALS NORM(ZRESID)
/SAVE MAHAL COOK.
```

Ovo je deo rezultata prethodne procedure.

Model Summary^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.238 ^a	.057	.052	5.69	.057	12.711	2	423	.000
2	.688 ^b	.474	.469	4.26	.417	166.873	2	421	.000

a. Predictors: (Constant), AGE, Total social desirability

b. Predictors: (Constant), AGE, Total social desirability, Total Mastery, Total PCOISS

c. Dependent Variable: Total perceived stress

ANOVA^c

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	823,865	2	411,932	12,711	.000 ^a
	Residual	13708,620	423	32,408		
	Total	14532,484	425			
2	Regression	6885,760	4	1721,440	94,776	.000 ^b
	Residual	7646,724	421	18,163		
	Total	14532,484	425			

a. Predictors: (Constant), AGE, Total social desirability

b. Predictors: (Constant), AGE, Total social desirability, Total Mastery, Total PCOISS

c. Dependent Variable: Total perceived stress

Coefficients^a

		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	31,076	,983		31,605	,000					
	total social desirability	-,599	,140	-,209	-4,271	,000	-,228	-,203	-,202	,928	1,08
	age	-,031	,022	-,070	-1,438	,151	-,127	-,070	-,068	,928	1,08
2	(Constant)	51,922	1,366		38,008	,000					
	total social desirability	-,149	,108	-,052	-1,373	,171	-,228	-,067	-,049	,871	1,15
	age	-,021	,017	-,047	-1,239	,216	-,127	-,060	-,044	,860	1,16
	total mastery	-,641	,062	-,435	-10,286	,000	-,612	-,448	-,364	,699	1,43
	total PCOISS	-,160	,022	-,327	-7,373	,000	-,581	-,338	-,261	,635	1,57

a. Dependent Variable: total perceived stress

Tumačenje rezultata hijerarhijske višestruke regresije

Rezultati ove analize liče na rezultate prethodne, ali sadrže i dodatne podatke. U polju **Model Summary** navedena su dva modela. **Model 1** označava prvi blok unetih promenljivih (Total social desirability i age), dok **Model 2** obuhvata sve promenljive unete u oba bloka (Total social desirability, age, Total mastery, Total PCOISS).

Korak 1: vrednovanje modela

Pogledajte kolike su vrednosti **R Square** u prvoj tabeli **Model summary**. Nakon unošenja promenljivih u Bloku 1 (age, social desirability), ceo model objašnjava 5,7 procenata varijanse ($0,057 \times 100$). Nakon unošenja promenljivih u Bloku 2 (Total Mastery, Total PCOISS), model *kao celina* objašnjava 47,4 procenata ($0,474 \times 100$). Treba uočiti da ta druga vrednost r^2 (R square) obuhvata sve promenljive iz oba bloka, a ne samo one unete u drugom koraku.

Da bismo utvrdili koliki deo te ukupne varijanse objašnjavaju promenljive koje nas zanimaju (Mastery, PCOISS) nakon što se ukloni uticaj promenljivih age i socially desirable responding (tj. starosti i davanja društveno poželjnih odgovora), pogledaćemo kolonu sa zaglavljem **R Square change**. U prethodno datim rezultatima, u redu za **Model 2**, vidimo da vrednost **R square change** iznosi 0,42. To znači da Mastery i PCOISS objašnjavaju dodatna 42 procenata ($0,42 \times 100$) varijanse subjektivno doživljenog stresa, čak i kada se statistički ukloni uticaj starosti i davanja društveno poželjnih odgovora. To je statistički značajan doprinos, što pokazuje iznos **Sig. F change** u tom redu (0,000). Tabela **ANOVA** pokazuje da model kao celina (sa oba bloka promenljivih) jeste značajan [$F(4, 421) = 94,78, p < 0,0005$].

Korak 2: vrednovanje svake nezavisne promenljive

Da bismo utvrdili koliko svaka promenljiva doprinosi konačnoj jednačini, pogledaćemo red za **Model 2** u tabeli **Coefficients**. Tu su zbirno prikazani rezultati dobijeni sa svim promenljivama unetim u jednačinu. U koloni **Sig.** vidimo da samo dve promenljive daju statistički značajan doprinos (jer je samo za njih vrednost u ćeliji manja od 0,05). Po redu važnosti, to su: Mastery (beta = -0,44) i Total PCOISS (beta = -0,33). Ni promenljiva age niti social desirability ne daju jedinstven doprinos. Ne zaboravite da koeficijenti beta predstavljaju jedinstven doprinos svake promenljive, nakon što se statistički uklone uticaji preklapanja sa svim drugim promenljivama. U drugim jednačinama, sa drugačijim skupom nezavisnih promenljivih ili drugačijim uzorkom, te vrednosti bi se promenile.

Predstavljanje rezultata višestruke regresije

U zavisnosti od vrste sprovedenih analiza i prirode istraživačkog pitanja, rezultati višestruke regresije mogu se predstaviti na više načina. U priručniku *Publication Manual* Američkog udruženja psihologa, date su smernice za predstavljanje rezultata višestruke regresije i odgovarajući primer (*Publication Manual*, 2001, str. 160, 163). U najmanju ruku, trebalo bi da na vrstu obavljene analize (standardna ili hijerarhijska), standardizovane koeficijente ako je istraživanje bilo teorijsko, odnosno nestandardizovane koeficijente (i njihove standardne greške) ako je istraživanje bilo primenjeno. Za hijerarhijsku višestruku regresiju, trebalo bi da navedete i vrstu promene r^2 (R square change) u svakom koraku i odgovarajuće verovatnoće.

Sledi primer načina predstavljanja rezultata analiza sprovedenih u poglavlju. U istraživanjima s mnogo promenljivih, rezultate je prikladno predstaviti tabelarno (videti American Psychological Association, 2001, str. 163; ili Tabachnick & Fidell, 2007, str. 177). Da se ovi rezultati predstavljaju u tezi, trebalo bi u prilogu navesti još nešto (npr. tabelu korelacija) - verite to u razgovoru sa svojim supervizorom).

Bilo bi dobro da primere predstavljanja raznih statističkih analiza tražite u časopisima relevantnim za konkretnu oblast. Zbog strogo ograničenog prostora, razni časopisi postavljaju različite zahteve i očekivanja autore.

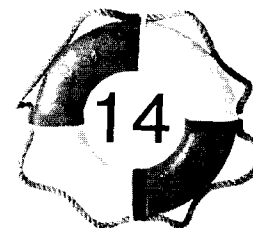
Za ocenu mogućnosti dva merila upravljanja sopstvenim životom (skala upravljanja: Mastery, skala subjektivno doživljene samokontrole unutrašnjih stanja: SDSUS) da predvide nivo stresa (skala subjektivno doživljenog stresa), izvršena je hijerarhijska višestruka regresija nakon što je uklonjen uticaj starosti i davanja društveno poželjnih odgovora. Preliminarnim analizama je dokazano da pretpostavke normalnosti, linearnosti, multikolinearnosti i homogenosti varijanse nisu bile narušene. U prvom koraku bile su unete promenljive starost i davanje društveno poželjnih odgovora, što je objasnilo 6% varijanse subjektivno doživljenog stresa. Nakon unošenja skale Mastery i SDSUS u drugom koraku, modelom kao celinom objašnjeno je 47,4% ukupne varijanse, $F(4, 421) = 94,78, p < 0,001$. Dva spomenuta merila (obeležja) objasnila su dodatnih 42% varijanse u stresu, nakon što je uklonjen uticaj starosti i davanja društveno poželjnih odgovora; r^2 se promenilo za = 0,42, F se promenilo za $(2, 421) = 166,87, p < 0,001$. U konačnom modelu, statistički su bila značajna samo dva merila upravljanja životom, pri čemu je skala Mastery imala veći koeficijent beta (beta = -0,44, $p < 0,001$) od skale SDSUS (beta = -0,33, $p < 0,001$).

Dodatne vežbe

Zdravstvo

Datoteka s podacima: *sleep3ED.sav*. Pojediniosti o datoteci videti u dodatku.

1. Standardnom višestrukom regresijom istražite faktore koji utiču na stepen pospanosti ljudi po danu. Za zavisnu promenljivu uzmite ukupan iznos na skali Sleepiness and Associated Sensations (*totSAS*). Za nezavisne promenljive uzmite pol (*sex*), starost (*age*), ocenu fizičke kondicije (*fitrate*) i stepen depresije (*depress*). Izračunajte koliki deo varijanse u rezultatima merenja ukupne pospanosti objašnjava taj skup promenljivih (pogledajte koliki je r na kvadrat). Koje promenljive daju jedinstven značajan doprinos (pogledajte vrednosti koeficijenata β)?
2. Ponovite gornju analizu, ali ovoga puta upotrebite hijerarhijsku višestruku regresiju. U prvom bloku promenljivih unesite pol i starost, a u drugom ocenu fizičke kondicije i stepen depresije. Kada uklonite uticaj demografskih promenljivih pol i starost, da li ostale dve prediktorske promenljive značajno doprinose objašnjavanju varijanse u rezultatima merenja pospanosti? Koliki dodatni deo varijanse u pospanosti objašnjavaju fizička kondicija i depresija, nakon što se ukloni uticaj pola i starosti?



Logistička regresija

U poglavlju 13 o višestrukoj regresiji, opisali smo tehniku za ocenu uticaja skupa prediktora na zavisnu promenljivu (subjektivno doživljen stres). U tom slučaju, zavisnu promenljivu smo merili kao neprekidnu (s vrednostima u opsegu od 10 do 50). Međutim, ima mnogo situacija kada je zavisna promenljiva koja nas zanima kategorijska (npr. pobediti/izgubiti; položiti/pasti; mrtav/živ). Nažalost, višestruka regresija nije prikladna za kategorijske zavisne promenljive. U višestrukoj regresiji, zavisna promenljiva (ono što pokušavate da objasnite ili predvidite) mora biti neprekidna, a njene vrednosti prilično normalno raspodeljene.

Logistička regresija omogućava ispitivanje modela za predikciju kategorijskih ishoda sa dve ili više kategorija. Prediktorske (nezavisne) promenljive mogu biti kategorijske, neprekidne ili mešavina obe vrste u istom modelu. SPSS ima celu porodicu tehnika za logističku regresiju, koje služe za istraživanje prediktivne moći skupova ili blokova promenljivih i omogućuju zadanje načina unosa promenljivih. Svrha ovog primera je da pokaže osnove logističke regresije. Zato ću upotrebiti metodu prisilnog unosa (Forced Entry Method) – podrazumevanu proceduru dostupnu u SPSS-u. To je pristup u kojem se sve prediktorske promenljive ispituju u istom bloku da bi se ocenila njihova prediktivna moć, dok se statistički uklanja uticaj ostalih prediktora u modelu. Druge tehnike regresije – recimo, one postepene (dodavanje unapred i uklanjanje unazad) – omogućuju zadanje velike grupe mogućih prediktora, odakle SPSS sam bira podskup koji ima najveću moć predviđanja. Postepene tehnike regresije su predmet kritike i u logističkoj i u višestrukoj regresiji, zato što na njih može jako uticati slučajna promenljivost podataka, kada se promenljive uključuju ili isključuju iz modela zbog čisto statističkih razloga (videti raspravu u Tabachnick & Fidell, 2007, str. 456).

U ovom poglavlju pokazaću kako se pomoću SPSS-a obavlja logistička regresija za dihotomnu zavisnu promenljivu (tj. onu koja može imati samo dve vrednosti ili kategorije). Upotrebićemo SPSS-ovu proceduru Binary Logistic. Kada zavisna promenljiva ima više od dve kategorije, upotrebljava se skup procedura Multinomial Logistic (koje nisam prikazala u ovoj knjizi, ali

su dostupne u SPSS-u – videti meni Help). Logistička regresija je složena tehnika, pa smatram neophodnim da o njoj više pročitate ukoliko nameravate da je koristite (videti Hosmer & Lemeshow, 2000; Peat, 2001; Tabachnick & Fidell, 2007; Wright, 1995).

Pretpostavke

Veličina uzorka

Kao i u većini ostalih statističkih tehnika, morate razmotriti veličinu i prirodu uzorka ukoliko nameravate da upotrebite logističku regresiju. Jedno od pitanja je broj slučajeva u uzorku u odnosu na broj prediktora (nezavisnih promenljivih) koje želite da uključite u model. Analiza postaje problematična (čak toliko da numerički divergira) kada je uzorak mali, a želite da uključite mnogo prediktora. To je problem naročito za kategorijske prediktore sa ograničenim brojem slučajeva u svakoj kategoriji. Uvek za svaki prediktor pokrenite proceduru Descriptive Statistics, pa ako je broj slučajeva (opservacija) u određenoj kategoriji premali, preraspodelite ih u druge kategorije ili ih uklonite iz analize.

Multikolinearnost

Kao što je bilo rečeno u poglavlju 13, uvek treba proveriti da li postoje jake međukorelacije prediktorskih (nezavisnih) promenljivih. Bilo bi idealno da su prediktorske promenljive jako povezane sa zavisnom promenljivom, ali ne i međusobno. Nažalost, u proceduri logističke regresije SPSS formalno nema test multikolinearnosti, pa vam preostaje postupak opisan u poglavlju 13, kada izaberete **collinearity diagnostics** nakon što pritisnete dugme **Statistics**. Ostale rezultate zanemarite, samo pogledajte tabelu **Coefficients** i njene kolone sa zaglavljem **Collinearity Statistics**. Veoma male vrednosti pokazatelja **Tolerance** (manje od 0,1) ukazuju na to da promenljiva ima visoke korelacije sa ostalim promenljivama u modelu. Tada treba da promenite skup promenljivih od kojih ste mislili da sastavite model i uklonite jednu od jako međukoreliranih promenljivih.

Netipične tačke

Neophodno je da proverite postoje li netipične tačke, to jest slučajevi koje model ne objašnjava dobro. Rečeno jezikom logističke regresije, model možda odlično predviđa određen slučaj u jednoj kategoriji, ali je u stvarnosti on u nekoj drugoj kategoriji. Netipični slučajevi se mogu prepoznati na dijagramu reziduala, što je naročito važan korak kada model slabo predviđa opservacije.

Objašnjenje primera

Logističku regresiju ću ilustrovati na primeru datoteke sa stvarnim podacima (**sleep3ED.sav**), s prateće Web lokacije ove knjige (videti Web adresu na str. xi). Ti podaci su dobijeni u anketi sprovedenoj na uzorku univerzitetskog osoblja da bi se utvrdila veličina i uticaj problema u vezi sa snom (videti **data**). U anketi su ispitanici odgovarali na pitanje smatraju li da imaju problema sa spavanjem (da/ne). To će biti zavisna promenljiva u ovoj analizi. Skup prediktora (nezavisnih promenljivih) obuhvata pol, starost, broj sati noćnog sna, teško padanje u san i lako buđenje.

Prvobitnim vrednostima svake od tih promenljivih dodeljene su nove šifre (kodovi) da bi se obezbedila njihova prikladnost za ovu analizu. I kategorijskim promenljivama je promenjen kôd, i to tako da je 0=ne i 1=da.

Priprema podataka: šifrovanje odgovora

Da biste razumeli rezultate logističke regresije, moraćete pažljivo da pripremite šifrovanje odgovora za svaku promenljivu. Za dihotomne zavisne promenljive, odgovore treba šifrovati sa 0 i 1 (ili rešifrovati postojeće vrednosti SPSS-ovom procedurom **Recode** – videti poglavlje 8). Vrednost 0 treba dodeliti svakom odgovoru koji ne sadrži obeležje od interesa. U ovom primeru, nulom se šifrjuje odgovor Ne na pitanje 'Imate li problema sa spavanjem?' Broj 1 ukazuje na odgovor Da. Slično su šifrovane nezavisne promenljive. Tu smo odgovor Da šifrovali sa 1, za odgovore na pitanja ima li osoba problema da zaspi i ima li problema da ostane u snu. Za neprekidne nezavisne promenljive (broj sati noćnog sna), veći brojevi treba da pokazuju jače ispoljavanje ispitivanog obeležja.

Ime datoteke: **sleep3ED.sav**

Promenljive:

- Problemi sa spavanjem (engl. *problem with sleep*), rešifrovana (**prob-sleeprec**): vrednost rešifrovana na 0=ne, 1=da
- Pol (engl. *sex*): 0=ženski, 1=muški
- Starost (engl. *age*): u godinama
- Broj sati noćnog sna (engl. *hours sleep/weeknight*) (**hourwnit**): u satima
- Teško padanje u san (engl. *problem getting to sleep*), rešifrovana (**getsleprec**): vrednost rešifrovana na: 0=ne, 1=da
- Lako buđenje (engl. *problem staying asleep*), rešifrovana (**stayslprec**): vrednost rešifrovana na: 0=ne, 1=da

Kratak pregled logističke regresije

Primer istraživačkog pitanja: Koji faktori predviđaju verovatnoću da će ispitanici odgovoriti da imaju problema sa spavanjem?

Šta vam treba:

- Jedna kategorijska (dihotomna) zavisna promenljiva (problemi sa spavanjem: Ne/Da, šifrovano 0/1); i
- Dve ili više neprekidnih ili kategorijskih prediktorskih (nezavisnih) promenljivih. Dihotomne promenljive šifrovati sa 0 i 1 (npr. pol, teško padanje u san, lako buđenje). Neprekidne promenljive meriti tako da veći brojevi pokazuju jače ispoljavanje obeležja od interesa (npr. starost, broj sati noćnog sna).

Šta se postiže: Logistička regresija služi za ocenu koliko dobro skup prediktorskih promenljivih predviđa ili objašnjava kategorijsku zavisnu promenljivu. Dobija se pokazatelj adekvatnosti modela (skupa prediktorskih promenljivih) tj. ocena kvaliteta predviđanja rezultata. Dobija se pokazatelj relativne važnosti svake prediktorske promenljive ili interakcije između njih. Dobija se zbir pokazatelja tačnosti klasifikacije slučajeva na osnovu modela, što omogućava izračunavanje osetljivosti i određenosti (specifičnosti) modela i njegovih pozitivnih i negativnih prediktivnih vrednosti.

Pretpostavke: Logistička regresija ne počiva na pretpostavkama o raspodeli rezultata merenja prediktorskih promenljivih; međutim, ona je osetljiva na visoke korelacije između prediktorskih promenljivih (multikolinearnost). Na rezultate logističke regresije jako utiču i netipične tačke.

Pre nego što počnete sa sledećim postupkom, otvorite meni **Edit** i izaberite u njemu stavku **Options**; treba da je potvrđeno polje **No scientific notation for small numbers in tables**.

Postupak za logističku regresiju

1. Otvorite meni **Analyze**, izaberite u njemu stavku **Regression**, pa pritisnite **Binary Logistic**.
2. Izaberite kategorijsku zavisnu promenljivu (npr. probsleeprec) i prebacite je u polje **Dependent**.
 - Pritisnite prediktorske promenljive (sex, age, getsleprec, stayslprec, hourwnit) i prebacite ih u polje **Covariates**.
 - U odeljku **Method** treba da piše **Enter**.
3. Ako imate kategorijskih prediktora, pritisnite dugme **Categorical** na dnu okvira za dijalog. Izaberite sve kategorijske promenljive (sex, getsleprec, stayslprec) i prebacite ih u polje **Categorical covariates**.
 - Redom izaberite svaku kategorijsku promenljivu, pa pritisnite dugme **First** u odeljku **Change contrast**. Kada pritisnete dugme **Change**, izabrane promenljive prikazaće se reč first. Tako ste zadali da grupa koja će biti upotrebljena kao referentna bude prikazana prva. Ponovite to za sve kategorijske promenljive. Pritisnite dugme **Continue**.

4. Pritisnite dugme **Options**. Potvrdite polja **Classification plots**, **Hosmer-Lemeshow goodness of fit**, **Casewise listing of residuals** i **CI for Exp(B)**.
5. Pritisnite **Continue** i zatim **OK** (ili dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Prethodni postupak generiše ovu komandu:

```
LOGISTIC REGRESSION VARIABLES probsleeprec
/METHOD = ENTER sex age getsleprec stayslprec hourwnit
/CONTRAST (sex)=Indicator(1) /CONTRAST (getsleprec)=Indicator(1)
/CONTRAST (stayslprec)=Indicator(1)
/CLASSPLOT /CASEWISE OUTLIER(2)
/PRINT = GOODFIT CI(95)
/CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .
```

Sledi deo rezultata prethodnog postupka.

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	241	88,9
	Missing Cases	30	11,1
	Total	271	100,0
Unselected Cases		0	,0
Total		271	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
no	0
yes	1

Categorical Variables Codings

		Frequency	Parameter (1)
prob stay asleep	no	138	,000
rec	yes	103	1,000
prob fall asleep	no	151	,000
rec	yes	90	1,000
sex	female	140	,000
	male	101	1,000

Block 0: Beginning Block

Iteration History^{a,b,c}

		-2 Log likelihood	Coefficients Constant
Step	1	328,996	-,290
0	2	328,996	-,293
	3	328,996	-,293

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 328,996

c. Estimation terminated at iteration number 3 because parameter estimates changed by less than ,001.

Classification Table^{a,b}

Observed		Predicted		Percentage Correct
		prob sleep recode 01 no	yes	
Step 0	prob sleep recode no	138	0	100,0
	01 yes	103	0	,0
Overall Percentage				57,3

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-,293	,130	5,047	1	,025	,746

Variables not in the Equation

Step	Variables	Score	df	Sig.
0	sex(1)	1,209	1	,272
	age	,795	1	,373
	getsleprec(1)	19,812	1	,000
	stayslprec(1)	58,183	1	,000
	hourwnit	17,709	1	,000
Overall Statistics		70,017	5	,000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	76,020	5	,000
	Block	76,020	5	,000
	Model	76,020	5	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	252,976 ^a	,271	,363

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	10,019	8	,264

Classification Table^a

Observed		Predicted		Percentage Correct
		prob sleep recode 01 no	yes	
Step 1	prob sleep recode no	110	28	79,7
	01 yes	32	71	68,9
Overall Percentage				75,1

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B)	
Step 1	sex(1)	-,108	,315	,118	1	,731	,897	,484	1,663
	age	-,006	,014	,193	1	,660	,994	,968	1,021
	getsleprec(1)	,716	,339	4,464	1	,035	2,046	1,053	3,976
	stayslprec(1)	1,984	,325	37,311	1	,000	7,274	3,848	13,748
	hourwnit	-,448	,165	7,366	1	,007	,639	,462	,883
	Constant	1,953	1,451	1,812	1	,178	7,053		

a. Variable(s) entered on step 1: sex, age, getsleprec, stayslprec, hourwnit.

Casewise List^b

Case	Selected Status ^a	Observed	Predicted	Predicted Group	Temporary Variable	
		prob sleep recode 01			Resid	ZResid
42	S	n**	,870	y	-,870	-2,583
224	S	y**	,126	n	,874	2,633
227	S	y**	,133	n	,867	2,554
235	S	y**	,119	n	,881	2,721
265	S	y**	,121	n	,879	2,697

a. S = Selected, U = Unselected cases, and ** = Misclassified cases.

b. Cases with studentized residuals greater than 2,000 are listed.

Tumačenje rezultata logističke regresije

Kao i većina SPSS-ovih analiza, logistička regresija daje obilje rezultata, pa ću pomenuti samo ključne aspekte.

Najpre treba pogledati pojedinosti o veličini uzorka, date u tabeli **Case Processing Summary**. Broj analiziranih slučajeva mora biti jednak onome koji ste očekivali. Sledeća tabela, **Dependent Variable Encoding**, kazuje kako je SPSS postupio sa šiframa zavisne promenljive (u ovom slučaju, sa odgovorima na pitanje imaju li ljudi problema sa spavanjem). SPSS zahteva da te promenljive budu šifrovane sa 0 i 1, ali će ih i sam rešifrovati ako to prethodno sami niste uradili (npr. ako su vaše šifre 1 i 2). Htela sam da moja promenljiva 'problem with sleep' bude šifrovana sa 0=nema problema i 1=problem, pa sam napravila novu promenljivu (procedurom Recode – videti poglavlje 8) i rešifrovala prvobitne odgovore 1=da, 2=ne na format poželjan u SPSS-u, tj. na 1=da, 0=ne. Takvo šifrovanje (kada se postojanje problema označava sa 1, a nepostojanje sa 0) olakšava tumačenje rezultata.

U sledećoj tabeli **Categorical Variables Codings** pogledajte šifrovanje nezavisnih (prediktorskih) promenljivih. Takođe, u koloni sa zaglavljem **Frequency** proverite broj slučajeva u svakoj kategoriji. Nemojte raditi sa grupama koje imaju malo slučajeva.

U sledećem odeljku, **Block 0**, dati su rezultati analize bez ijedne nezavisne promenljive od kojih se model sastoji. S njima ćemo kasnije uporediti rezultate modela koji obuhvata prediktorske promenljive. U tabeli **Classification** stoji da je ispravno klasifikovanih slučajeva ukupno bilo 57,3 procenata. U ovom primeru, SPSS je klasifikovao (nagađao) da nijedna osoba neće imati problema sa spavanjem (samo zato što je veći procenat ljudi na pitanje odgovorio sa Ne). Nadamo se da ćemo kasnije, kada unesemo naš skup prediktorskih promenljivih, poboljšati tačnost tih predviđanja.

Predite u sledeći odeljak, **Block 1**. Tu su rezultati ispitivanja modela (skupa prediktorskih promenljivih). U tabeli **Omnibus Tests of Model Coefficients** dati su zbirni pokazatelji performansi modela, dakle razlika u odnosu na rezultate dobijene za Block 0, kada u model nije bila uneta nijedna prediktorska promenljiva. Taj test se zove *goodness of fit*, tj. koliko dobro model predviđa rezultate. Za taj skup rezultata želimo da budu veoma značajni (veličina **Sig.** bi trebalo da bude manja od 0,05). U ovom slučaju, značajnost iznosi 0,000 (što zapravo znači $p < 0,0005$). Stoga proglašavamo da model (s našim skupom promenljivih kao prediktorima) predviđa bolje od SPSS-ovog prvobitnog nagađanja prikazanog u odeljku Block 0, dobijenog uz pretpostavku da niko neće odgovoriti kako ima problema sa spavanjem. Pokazatelj hi-kvadrat iznosi 76,02 uz 5 stepeni slobode i to treba navesti u izveštaju.

I rezultati prikazani u tabeli **Hosmer and Lemeshow Test** podržavaju tvrdnju da je model dobar. Tvrdi se da je to SPSS-ov najpouzdaniji test kvaliteta predikcije modela. On se tumači sasvim drugačije od prethodno

razmotrenog omnibus testa. Kada je u pitanju **Hosmer-Lemeshow Goodness of Fit Test**, indikator slabog predviđanja je značajnost manja od 0,05; dakle, model je podržan ako je veća od 0,05. U našem primeru, pokazatelj hi-kvadrat za Hosmer-Lemeshow test iznosi 10,019 uz značajnost 0,264. To je više od 0,05, što znači da podržava model.

Tabela **Model Summary** kazuje još nešto o upotrebljivosti modela. Vrednosti **Cox & Snell R Square** i **Nagelkerke R Square** pokazuju koliki deo varijanse zavisne promenljive model objašnjava (minimum je 0, a maksimum približno 1). To su takozvani pseudopokazatelji vrednosti r^2 (r na kvadrat), a ne njeni stvarni iznosi, koji se mogu videti u rezultatima višestruke regresije. U ovom primeru, ta dva pokazatelja su 0,271 i 0,363, što znači da dati skup promenljivih (model) objašnjava između 27,1 i 36,3 procenata varijanse.

Prelazimo na tabelu **Classification**. U njoj su pokazatelji koliko tačno model predviđa kategoriju (ima problema sa spavanjem/nema problema sa spavanjem) za svaki ispitivani slučaj. Možemo ih uporediti sa istoimenom tabelom u Block 0 i videti poboljšanje postignuto uključanjem prediktorskih promenljivih u model. Taj model ispravno klasifikuje 75,1 procenata svih slučajeva (engl. *percentage accuracy in classification*, PAC), što je bolje od 57,3 procenata navedenih u Block 0. Rezultati prikazani u ovoj tabeli služe i za izračunavanje drugih statističkih pokazatelja, koji se često navode u medicinskoj literaturi.

Osetljivost (engl. *sensitivity*) modela je procentualni udeo grupe sa ispitivanim obeležjem (npr. problem sa spavanjem) koji je model tačno prepoznao (stvarno pozitivnih). U ovom primeru, model je tačno klasifikovao 68,9 procenata osoba koje imaju problema sa spavanjem. Određenost (engl. *specificity*) modela je procentualni udeo grupe koja nema ispitivano obeležje (nema problema sa spavanjem) koji je model tačno prepoznao (stvarno negativnih). U ovom primeru, određenost iznosi 79,7 procenata (osoba koje nemaju problema sa spavanjem, za koje model tačno predviđa da nemaju tih problema).

Pozitivna prediktivna vrednost je procentualni udeo slučajeva koje model klasifikuje kao da imaju ispitivano obeležje, a da se ono zaista opaža u toj grupi. Da biste to izračunali u tekućem primeru, podelite broj slučajeva u ćeliji predicted=yes, observed=yes (71) ukupnim brojem ćelija predicted=yes ($28 + 71 = 99$) i pomnožite sa 100 da dobijete procentualni iznos. Dobili bismo 71 podeljeno sa $99 \times 100 = 71,7$ procenata. Dakle, pozitivna prediktivna vrednost iznosi 71,7 procenata, što pokazuje da je naš model tačno izabrao 71,7 procenata osoba za koje je prognozirano da imaju problema sa spavanjem. Negativna prediktivna vrednost je procentualni udeo slučajeva koje model klasifikuje kao da nemaju obeležje, a da se ono zaista ne opaža u toj grupi. U tekućem primeru, odgovarajući brojevi iz klasifikacione tabele jesu: 110 podeljeno sa $(110 + 32) \times 100 = 77,5$ procenata. Više o upotrebi klasifikacionih tabela potražite u knjigama Wright (1995, str. 229) ili Peat (2001, str. 237), gde je dat i jedan jednostavan primer.

Tabela **Variables in the Equation** daje podatke o doprinosu ili važnosti svake prediktorske promenljive. Taj test je izumio Wald, pa su vrednosti pokazatelja za svaki prediktor navedeni u koloni sa zaglavljem **Wald**. Pogledajte niz kolonu **Sig.** ima li brojeva manjih od 0,05. To su promenljive koje značajno doprinose prediktivnim mogućnostima modela. Utvrdila sam da su tri promenljive značajne (stayslprec $p = 0,000$, getsleprec $p = 0,035$, hourwtnit $p = 0,007$). U ovom primeru, glavni faktori koji utiču na to da li će osoba odgovoriti da ima problema sa spavanjem jesu: težina padanja u san, lako buđenje i broj sati noćnog sna. Pol i starost ispitanika nisu značajno doprileli modelu.

Koeficijenti **B** u drugoj koloni ekvivalentni su koeficijentima **B** dobijenim u analizi višestruke regresije. Ti brojevi se stavljaju u jednačinu za izračunavanje verovatnoće da analizirani slučaj spada u određenu kategoriju. Pogledajte da li su koeficijenti **B** pozitivni ili negativni. To će vam ukazati na smer veze (koji faktori povećavaju verovatnoću odgovora Da, a koji je smanjuju). Ako ste ispravno šifrovali sve zavisne i nezavisne kategorijske promenljive (sa 0=ne ili nepostojanje obeležja; 1=da ili postojanje obeležja), negativne vrednosti koeficijenata **B** pokazuju da povećanje vrednosti nezavisne promenljive ima za posledicu smanjenje verovatnoće da će ta osoba odgovoriti sa 1 u zavisnoj promenljivoj (što u ovom slučaju pokazuje postojanje problema sa spavanjem). U ovom primeru, promenljiva koja meri broj sati noćnog sna ima negativan koeficijent **B** (-0,448). To kazuje da što više sati osoba spava noću, to je manje verovatno da će odgovoriti kako ima problema sa spavanjem. Za ostale dve kategorijske promenljive (težina padanja u san, lako buđenje), koeficijenti **B** su pozitivni. To znači da osobe koje smatraju da teško padaju u san odnosno da se lako bude, češće odgovaraju Da na pitanje smatraju li da imaju problema sa spavanjem.

Još jedan koristan podatak sadrži kolona **Exp(B)** tabele **Variables in the Equation**. Ti brojevi su količnici verovatnoće (engl. *odds ratios*) za svaku nezavisnu promenljivu. Kako pišu Tabachnick i Fidell (2007), količnik verovatnoće (KV) je 'promena verovatnoće pripadanja jednoj kategoriji ishoda kada se vrednost određenog prediktora poveća za jednu mernu jedinicu' (str. 461). U našem primeru, verovatnoća da će osoba odgovoriti sa Da na pitanje ima li problema sa spavanjem jeste 7,27 puta veća za onoga ko odgovori da se lako budi u odnosu na osobu koja se ne budi lako, kada su svi ostali faktori jednaki.

Značajan prediktor je i broj sati noćnog sna, sudeći po pripadnoj veličini **Sig.** ($p=0,007$). Međutim, količnik verovatnoće za ovu promenljivu iznosi 0,639, dakle manje od 1. To znači sledeće: što više sati osoba spava noću, manje je verovatno da će odgovoriti kako ima problema sa spavanjem. Za svaki dodatni sat noćnog sna, verovatnoća da će osoba odgovoriti da ima problema sa spavanjem opada za 0,639, kada su svi ostali faktori jednaki.

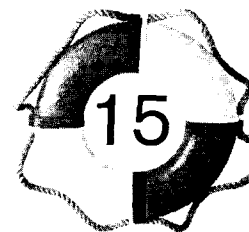
Imajte u vidu da je naš prediktor neprekidna promenljiva; zato govorimo o povećanju (ili smanjenju, ako je KV manji od 1) verovatnoće za svaku jedinicu povećanja prediktorske promenljive (u ovom slučaju, godišnjeg). Kada je prediktorska promenljiva kategorijska, poredimo verovatnoće za te dve kategorije. Za kategorijske promenljive sa više od dve kategorije, svaka kategorija se poredi s referentnom grupom (najčešće grupom čija je šifra najmanji broj, ako ste zadali **First** u odeljku **Contrast** u okviru za dijalog **Define Categorical Variables**).

Za količnike verovatnoće manje od 1, u izveštaju bismo mogli navesti njihovu recipročnu vrednost (1 podeljen s tim brojem) da bismo olakšali tumačenje. Primera radi, u ovom slučaju 1 podeljeno sa 0,639 jednako je 1,56. To kazuje sledeće: za svaki sat noćnog sna manje, verovatnoća da će ispitanik odgovoriti da ima problema sa spavanjem raste za 1,56. Ukoliko odlučite da navedete recipročne vrednosti količnika verovatnoće, trebalo bi da navedete i recipročne vrednosti intervala poverenja (videti sledeći pasus).

Za svaki količnik verovatnoće u koloni **Exp(B)** tabele **Variables in the Equation** prikazan je pripadni 95-procentni interval poverenja (95,0% **CI for EXP(B)**), njegova donja i gornja granica. Njega treba navesti u rezultatima. Jednostavno rečeno, to je opseg za koji sa 95-procentnom sigurnošću tvrdimo da obuhvata stvarnu vrednost količnika verovatnoće. Ne zaboravite, broj naveden kao količnik verovatnoće samo je procena ili nagađanje stvarne vrednosti na osnovu podataka iz uzorka. Poverenje koje imamo u tvrdnju da je to tačna vrednost (za celu populaciju) menja se u zavisnosti od veličine uzorka. Mali uzorak znači da će interval poverenja oko procenjenog količnika verovatnoće biti veoma širok. Interval se naglo smanjuje s povećanjem uzorka. U ovom primeru, interval poverenja za promenljivu lako se budi (za stayslprec, KV = 7,27) ide od 3,85 do 13,75. Dakle, iako navodimo da je izračunati KV = 7,27, možemo biti 95 procenata sigurni da stvarna vrednost KV cele populacije leži između 3,85 i 13,75, što je veoma širok opseg brojeva. Interval poverenja u tom slučaju ne sadrži broj 1; stoga je rezultat statistički značajan uz $p < 0,05$. Kada interval poverenja sadrži broj 1, količnik verovatnoće nije statistički značajan jer ne možemo isključiti mogućnost da je stvarni količnik verovatnoće 1, što ukazuje na jednaku verovatnoću oba odgovora (da/ne).

Poslednja tabela rezultata, **Casewise List**, govori o slučajevima u uzorku koje model ne predviđa dobro. Prikazani su slučajevi s vrednostima pokazatelja **ZResid** većim od 2 (u ovom primeru, to su slučajevi broj 42, 224, 227, 235, 265). Slučajeve čiji su pokazatelji veći od 2,5 (ili manji od -2,5) treba pažljivije ispitati, pošto su svakako netipični (99 procenata slučajeva ima taj pokazatelj između -2,5 i +2,5). Iz ostalih informacija u tom spisku vidi se da je za jedan slučaj (broj 42) bilo predviđeno da će biti u kategoriji Da (imam problema sa spavanjem), a zapravo je (u koloni **Observed**) na pitanje odgovoreno sa Ne. Za sve ostale netipične slučajeve bilo je predviđeno da će odgovoriti Ne, a te osobe su odgovorile Da. Za sve te slučajeve (a pogotovo

one čiji je pokazatelj ZResid preko 2,5) treba proveriti unete informacije i uopšte saznati više o njima. Možda ćete naići na grupe slučajeva koje model ne predviđa dobro (npr. ljude koji rade u smenama). Razmislite o tome da iz datoteke s podacima uklonite slučajeve s velikim vrednostima pokazatelja ZResid i da ponovite analize.



Faktorska analiza

Predstavljanje rezultata logističke regresije

Rezultati ove procedure mogu se predstaviti ovako:

Direktna logistička regresija bila je sprovedena kako bi se ocenio uticaj više faktora na verovatnoću da će ispitanici odgovoriti da imaju problema sa spavanjem. Model sadrži pet nezavisnih promenljivih (pol, starost, teško padanje u san, lako buđenje i broj sati noćnog sna). Ceo model (sa svim prediktorima) bio je statistički značajan, $\chi^2(5, N=241) = 76,02, p < 0,001$, što pokazuje da model razlikuje ispitanike koji jesu i one koji nisu odgovorili da imaju problema sa spavanjem. Model u celini objašnjava između 27,1% (r na kvadrat Koksa i Snela) i 36,3% (r na kvadrat Nagelkerkea) varijanse u spavačkom statusu, i tačno klasifikuje 75,1% slučajeva. Kao što je prikazano u tabeli 1, samo su tri nezavisne promenljive dale jedinstven statistički značajan doprinos modelu (broj sati noćnog sna, teško padanje u san i lako buđenje). Najjači prediktor odgovora da osoba ima problema sa spavanjem bilo je lako buđenje, čiji je količnik verovatnoće 7,27. To pokazuje da ispitanici koji se lako bude, preko 7 puta češće odgovaraju da imaju problema sa spavanjem od onih koji se ne bude lako, uz sve ostale faktore u modelu jednake. Količnik verovatnoće za broj sati noćnog sna iznosi 0,64 i manji je od 1, što pokazuje da za svaki dodatni sat sna, ispitanici 0,64 puta ređe odgovaraju da imaju problema sa spavanjem, uz sve ostale faktore u modelu jednake.

Tabela 14.1: Predviđanje verovatnoće odgovora da osoba ima problema sa spavanjem

	B	Stand. greška	Wald	Stepeni slobode	p	Količnik verovatnoće	Interval 95-postotnog poverenja za količnik verovatnoće	
							Donja granica	Gornja granica
Pol	-0,11	0,31	0,12	1	0,73	0,90	0,48	1,66
Starost	-0,01	0,01	0,19	1	0,66	0,99	0,97	1,02
Padanje u san	0,72	0,34	4,46	1	0,03	2,05	1,05	3,98
Lako buđenje	1,98	0,32	37,31	1	0,00	7,27	3,85	13,75
Broj sati noćnog sna	-0,45	0,17	7,37	1	0,01	0,64	0,46	0,88
Konstanta	1,95	1,45	1,81		0,18	7,05		

Faktorska analiza se razlikuje od mnogih drugih tehnika predstavljenih u ovoj knjizi. Ona nije namenjena za testiranje hipoteza, niti može reći da li se jedna grupa značajno razlikuje od druge. SPSS je smatra tehnikom za 'smanjenje količine podataka'. Ona prima veliki skup promenljivih i traži način da te podatke sažme pomoću manjeg broja faktora ili komponenata. To se radi pronalaženjem 'gomila' ili grupa u međukorelacijama skupa promenljivih. To je gotovo nemoguće uraditi od oka osim kada ima malo promenljivih.

Porodica tehnika faktorske analize ima brojne primene. Mnogo je upotrebljavaju istraživači koji se bave razvojem i vrednovanjem testova i skala. Projektant skale počinje s velikim brojem pitanja i pojedinačnih stavki skale i, tehnikama faktorske analize, prečišćava i sažima te stavke da bi dobio manji broj usaglašenih podskala. Mnoštvo povezanih promenljivih može se faktorskom analizom svesti na manji broj podesan za druge analize, kao što su višestruka regresija ili multivarijaciona analiza varijanse (više od dve promenljive).

U literaturi su opisana dva glavna pristupa faktorskoj analizi – istraživački (engl. *exploratory*) i potvrđujući (engl. *confirmatory*). U ranim fazama istraživanja često se istraživačkom faktorskom analizom prikupljaju podaci o međusobnim vezama skupa promenljivih. S druge strane, potvrđujuća faktorska analiza je složeniji i sofisticiraniji skup tehnika, koje se u kasnijoj fazi istraživanja upotrebljavaju za ispitivanje (potvrdu) konkretnih hipoteza i teorija o zajedničkoj podstrukturi skupa promenljivih.

Termin faktorska analiza obuhvata više različitih, mada srodnih tehnika. Po načinu sprovođenja one se dele na analizu glavnih komponenata (engl. *principal components analysis*, PCA) i analizu zajedničkih faktora (engl. *common factor analysis*, CFA). Ta dva skupa tehnika slična su po mnogo čemu, pa ih istraživači često upotrebljavaju jedan umesto drugog. Oba dovode do manjeg broja linearnih kombinacija prvobitnih promenljivih na način koji zadržava (ili objašnjava) glavninu varijanse (promenljivosti) u strukturi korelacija. Međutim, oni se i razlikuju po više osnova. U analizi glavnih komponenata, prvobitne promenljive se transformišu u manji skup linearnih

kombinacija, uz korišćenje svih varijansi u promenljivama. S druge strane, u analizi zajedničkih faktora oni se procenjuju pomoću matematičkog modela, pri čemu se analizira samo zajednička varijansa. (Više o tome videti u 13. poglavlju knjige Tabachnick & Fidell, 2007).

Premda oba pristupa (PCA i CFA) često daju slične rezultate, knjige o toj temi često se razlikuju po tome koju od njih preporučuju. Stevens (1996, str. 362–3) priznaje sklonost ka analizi glavnih komponenata i za to daje više razloga. On tvrdi da je PCA matematički jednostavnija i prikladna za psihometriju, te da se njome izbegavaju neki od mogućih problema sa 'neodređenošću faktora' koji se pripisuju CFA (Stevens, 1996, str. 363). Tabachnick i Fidell (2007) u svom prikazu PCA i CFA zaključuju: 'Kada tražite teorijsko rešenje nekontaminirano jedinstvenom varijabilnošću i varijabilnošću greška ... CFA je alatka za vas. S druge strane, PCA je bolja za uobičajeno empirijsko sažimanje skupa podataka' (str. 635).

Odlučila sam da u ovom poglavlju prikazem analizu glavnih komponenata. Za dalje istraživanje ostalih pristupa, videti knjigu koju su napisale Tabachnick i Fidell (2007).

Napomena: iako PCA strogo uzev daje komponente, mnogi autori terminom *faktor* opisuju rezultate i PCA i CFA. Stoga nemojte pretpostavljati da je autor upotrebio CFA kada u časopisu pročitate termin *faktor*. Faktorska analiza je opšti termin kojim se opisuje cela ta porodica tehnika.

Još jedna oblast moguće zabune posledica je zloupotrebe reči *faktor*, koja ima različita značenja i upotrebu u raznim vrstama statističkih analiza. U faktorskoj analizi, ona se odnosi na grupu povezanih promenljivih; u tehnikama analize varijanse, ona se odnosi na nezavisnu promenljivu. To su potpuno različite stvari uprkos istom imenu. Nastojte da vam razlika između njih bude jasna dok radite razne analize.

Koraci od kojih se sastoji faktorska analiza

Faktorska analiza se sprovodi u tri glavna koraka. (Termin upotrebljavam u opštem smislu za bilo koju tehniku koja pripada ovoj porodici; dakle, i za analizu glavnih komponenata.)

Korak 1: ocena prikladnosti podataka za faktorsku analizu

Kada se utvrđuje prikladnost određenog skupa podataka za faktorsku analizu, treba razmotriti dva glavna pitanja: veličinu uzorka i jačinu veze između promenljivih (ili stavki). Iako se preporuke autora o veličini uzorka razlikuju, svi se slažu u jednom: što veći, to bolji. U malim uzorcima, koeficijenti korelacije između promenljivih manje su pouzdani i često se menjaju od jednog uzorka do drugog. Zato faktori dobijeni iz malih skupova podataka, kada se uopšte koriste, daju lošije rezultate od onih dobijenih iz velikih uzoraka. Tabachnick i Fidell (2007) razmatraju to pitanje i kažu: 'Istraživač

je mirniji kada za faktorsku analizu ima najmanje 300 slučajeva' (str. 613). Međutim, i one priznaju da bi manji uzorak (npr. 150 slučajeva) trebalo da bude dovoljan kada rešenja imaju nekoliko promenljivih velike težine (iznad 0,80). Stevens (1996, str. 372) navodi da se zahtevana veličina uzorka sve više smanjuje kako se s godinama ta tema sve više istražuje. On daje brojne preporuke za pouzdanost faktorskih struktura i zahtevanu veličinu uzorka (videti Stevens, 1996, 11. poglavlje).

Neki autori tvrde da nije toliko važna veličina uzorka već količnik broja subjekata i broja stavki (promenljivih). Nunnally (1978) preporučuje količnik 10 prema 1, tj. deset slučajeva (opservacija) za svaku stavku koju treba faktorski analizirati. Drugi tvrde da je pet opservacija po stavki dovoljno u većini slučajeva (videti raspravu u knjizi Tabachnick & Fidell, 2007). Preporučujem vam da više pročitate o tome, naročito ako imate mali uzorak (manji od 150) ili mnogo promenljivih.

Drugo pitanje koje treba razmotriti jeste jačina korelacije među promenljivama (stavkama). Tabachnick i Fidell preporučuju da se u matrici korelacija potraže koeficijenti veći od 0,3. Ako ih je malo, ti podaci su neprikladni za faktorsku analizu. SPSS ima dva testa opravdanosti primene faktorske analize: Bartletov test sferičnosti (Bartlett 1954) i Kajzer-Mejer-Olkinov (KMO) pokazatelj adekvatnosti uzorka (Kaiser, 1970, 1974). Bartletov test sferičnosti treba da je značajan ($p < 0,05$) da bi primena faktorske analize bila opravdana. KMO pokazatelj poprima vrednosti između 0 i 1, pa se 0,6 preporučuje kao najmanji iznos prihvatljiv za dobru faktorsku analizu (Tabachnick & Fidell, 2007).

Korak 2: izdvajanje faktora

Izdvajanje (ekstrakcija) faktora obuhvata određivanje najmanjeg broja faktora koji dobro predstavljaju međuveze u skupu promenljivih. Postoje razni pristupi za identifikaciju (izdvajanje) broja pripadnih zajedničkih faktora ili dimenzija. Među najčešće dostupnim tehnikama izdvajanja su i:

- analiza glavnih komponenata;
- analiza glavnih faktora;
- faktorizacija slike;
- faktorizacija metodom maksimalne verodostojnosti;
- alfa faktorizacija;
- neponderisana metoda najmanjih kvadrata; i
- opšta metoda najmanjih kvadrata.

Kao što je već rečeno, najčešće se radi analiza glavnih komponenata i nju ćemo pokazati na primeru u nastavku. Istraživač treba sam da odredi broj faktora koji po njegovom/njenom mišljenju najbolje opisuju odnose između promenljivih. To znači pronaći ravnotežu između dva suprotna zahteva: naći

jednostavno rešenje s najmanjim mogućim brojem faktora; i potrebe da se objasni što veći deo varijanse originalnog skupa podataka. Tabachnick i Fidell (2007) preporučuju istraživački pristup, tj. da se eksperimentiše s raznim brojevima faktora dok se ne otkrije zadovoljavajuće rešenje.

Za određivanje broja faktora koje treba zadržati postoji više tehnika:

- Kajzerov kriterijum (kriterijum karakterističnih vrednosti);
- dijagram prevoja (engl. *scree test*); i
- paralelna analiza.

Kajzerov kriterijum

Među najčešće upotrebljavanim tehnikama je Kajzerov kriterijum ili kriterijum karakterističnih vrednosti. Karakteristična vrednost faktora je ukupna varijansa svih promenljivih objašnjena tim faktorom. Po ovom pravilu, za dalje istraživanje zadržavaju se samo oni faktori čije su karakteristične vrednosti veće od 1,0, zato što već i originalne promenljive zbog standardizacije imaju varijansu 1. (Ovo će vam postati jasnije kada prorađimo primer u nastavku poglavlja.) Međutim, Kajzerov kriterijum kritikuju da u nekim situacijama zadržava previše faktora.

Kriterijum dijagrama prevoja

Broj faktora se može odrediti i Katelovim kriterijumom ili kriterijumom dijagrama prevoja (Catell, 1966). Treba nacrtati karakteristične vrednosti (engl. *eigenvalues*) svih faktora (SPSS to ume da uradi) i na dijagramu naći tačku u kojoj se oblik krive menja i ona prelazi u horizontalu. Katel preporučuje da se zadrže svi faktori iznad lakta, tj. prevoja tog dijagrama, pošto oni najviše doprinose objašnjavanju varijanse u skupu podataka.

Paralelna analiza

Sve popularnija tehnika, naročito u literaturi iz oblasti društvenih nauka (npr. Choi, Fuqua & Griffin, 2001; Stober, 1998), jeste Hornova paralelna analiza (Horn, 1965). Paralelna analiza znači uporediti iznose karakterističnih vrednosti sa onima dobijenim na jednako velikom skupu slučajno generisanih podataka. Zadržavaju se samo oni faktori čije su karakteristične vrednosti veće od odgovarajućih vrednosti dobijenih na nasumično generisanim podacima. Dokazano je da je ovaj način određivanja broja potrebnih komponenata najtačniji, zato što i u Kajzerovom i Katelovom kriterijumu postoji sklonost ka precenjivanju broja komponenata (Hubbard & Allen, 1987; Zwick & Velicer, 1986). Ako svoje rezultate nameravate da objavite u časopisu posvećenom psihologiji ili obrazovanju, moraćete da izračunate i predstavite rezultate paralelne analize. Mnogi časopisi (npr. *Educational and Psychological Measurement*, *Journal of Personality Assessment*) to postavljaju kao preduslov da rukopis uopšte razmotre za objavljivanje. U nastavku poglavlja demonstriraćemo sve tri tehnike u sklopu urađenog primera.

Korak 3: rotacija i tumačenje faktora

Kada je broj faktora određen, sledeći korak je da pokušate da ih protumačite. Da bi se ovaj proces olakšao, pribegava se 'rotaciji' faktora. Time se samo rešenje ne menja, ali se struktura faktorskih težina (engl. *factor loadings*) – tj. koeficijenata korelacije između promenljivih i faktora – predstavlja tako da se rezultati lakše tumače, jer se rotiranjem mera menja perspektiva posmatranja podataka. Nažalost, SPSS ne ume da tumači faktore niti da im dodeli opise. On samo pokazuje koje su promenljive 'zgomilane zajedno'. Na osnovu sopstvenog razumevanja sadržaja promenljivih (te pripadne teorije i prethodnih istraživanja), treba sami da predložite moguća tumačenja.

Rotirani faktori mogu biti ortogonalni (nekorelirani) ili kosi (korelirani). Tabachnick i Fidell (2007) kažu da ortogonalna rotacija daje rešenja koja se lakše tumače i predstavljaju; međutim, njima se istraživaču (obično netačno) nameće da su i pripadni konstrukti nezavisni (nekorelirani). U kosim rotacijama dozvoljeni su korelirani faktori, ali je njih teže protumačiti, opisati i predstaviti (Tabachnick & Fidell, 2007, str. 638). U praksi, ta dva pristupa (ortogonalni i kosi) često daju vrlo slična rešenja, naročito onda kada je struktura korelacija između stavki jasna (Tabachnick & Fidell, 2007). Mnogi istraživači urade i ortogonalnu i kosu rotaciju, a u izveštaju predstave onu koja je jasnija i koju je lakše protumačiti. Nadaite se da ćete dobiti ono što Thurstone (1947) naziva 'jednostavnom strukturom'. To znači da je svaka promenljiva jako korelirana sa samo jednom komponentom (tj. da samo tom faktoru daje veliku težinu), a svaku komponentu predstavlja pomoću više s njom jako koreliranih promenljivih (tj. da svakom faktoru više promenljivih da veliku težinu).

U te dve široke kategorije rotacija, spada više različitih tehnika SPSS-a (ortogonalne: Varimax, Quartimax, Equamax; kose: Direct Oblimin, Promax). Od ortogonalnih rotacija najčešće se upotrebljava metoda Varimax, koja pokušava da minimizira broj promenljivih s visokim apsolutnim vrednostima faktorskih težina. Najčešće upotrebljavana kosa rotacija je Direct Oblimin. Poređenje navedenih vrsta rotacija dato je u knjizi autorki Tabachnick i Fidell (2007, str. 639). U primeru predstavljenom u ovom poglavlju, prikazaću oblimin rotaciju. Uvek bi trebalo da počnete sa oblimin rotacijom, pošto ona daje i stepen korelacije između faktora.

Dodatna literatura

U ovom poglavlju dat je tek veoma kratak prikaz faktorske analize. Pokušala sam da je pojednostavim koliko god je to moguće, ali je faktorska analiza zapravo sofisticirana i složena porodica tehnika. Ako faktorsku analizu nameravate da upotrebite na sopstvenim podacima, preporučujem da pročitate više o njoj. Kao temeljno i lako čitljivo štivo o tome, preporučujem knjigu koju su napisali Pett, Lackey i Sullivan (2003). Za prikaz složenijih tema, videti knjigu autorki Tabachnick i Fidell (2007).

Objašnjenje primera

Da bih prikazala upotrebu faktorske analize, istražiću strukturu jedne od skala u datoteci s podacima **survey3ED.sav**, dostupnoj na pratećoj Web lokaciji. Anketa je bila osmišljena za istraživanje faktora koji utiču na psihološko prilagođenje i opšte raspoloženje ispitanika. (Celovit opis studije videti u dodatku.) Jedna od upotrebljenih skala bila je skala pozitivnih i negativnih osećanja (engl. *Positive and Negative Affect scale*, PANAS: Watson, Clark & Tellegen, 1988), slika 15.1. Ta skala se sastoji od dvadeset prideva koji opisuju različita psihička stanja/raspoloženja, deset pozitivnih (npr. ponosno, aktivno, odlučno) i deset negativnih (npr. nervozno, razdražljivo, uzrujano). Autori skale tvrde da se PANAS sastoji od dve dimenzije (ili faktora): pozitivnog osećanja i negativnog osećanja. U istraživanju ove strukture na uzorku postojeće zajednice, stavke skale biće podvrgnute analizi glavnih komponenata (PCA). Za razvoj i vrednovanje skala obično se upotrebljava upravo ta vrsta faktorske analize.

Ukoliko želite da pratite primer i odmah sprovedite navedena uputstva, pokrenite SPSS i otvorite datoteku **survey3ED.sav**, dostupnu na pratećoj Web lokaciji knjige. Promenljive korišćene u analizi nose oznake od pn1 do pn20. Na slici 15.1 predstavljena je skala korišćena u anketi. Kada budete pokušavali da protumačite dobijene faktore, pozivaćete se na te pojedinačne stavke. Sve pojedinosti i reference za skalu date su u dodatku.

Ova skala se sastoji od reči koje opisuju razna stanja i osećanja. Za svaku stavku naznačite do koje mere ste se tako osećali tokom nekoliko poslednjih sedmica. Na liniju pored svake stavke napišite broj od 1 do 5.				
vrlo malo ili nimalo	malo	umereno	prilično	izuzetno mnogo
1	2	3	4	5
1.–zainteresovano		8.–ucveljeno		15.–uzbuđeno
2.–uzrujano		9.–snažno		16.–kao krivac
3.–prestravljeno		10.–odbojno		17.–oduševljeno
4.–ponosno		11.–razdražljivo		18.–čilo
5.–posramljeno		12.–nadahuto		19.–nervozno
6.–odlučno		13.–pažljivo		20.–usplahireno
7.–aktivno		14.–uplašeno		

Slika 15.1 Skala pozitivnih i negativnih osećanja (PANAS)

Kratak pregled faktorske analize

Primer istraživačkog pitanja: Koju faktorsku strukturu ima skala pozitivnih i negativnih osećanja (Positive and Negative Affect scale, PANAS)? Prošla istraživanja ukazuju na strukturu od dva faktora (pozitivno osećanje/negativno osećanje). Da li je struktura skale u ovoj studiji, upotrebljenoj na uzorku postojeće zajednice, saglasna s prethodnim istraživanjima?

Šta vam treba: Skup koreliranih neprekidnih promenljivih.

Šta se postiže: Faktorskom analizom pokušavamo da prepoznamo (identifikujemo) mali skup faktora koji predstavlja unutrašnje veze u grupi povezanih promenljivih.

Pretpostavke:

1. *Veličina uzorka.* Idealno bi bilo da uzorak ima preko 150 slučajeva i da bude najmanje pet slučajeva po svakoj promenljivoj (videti raspravu u koraku 1, u prethodnom delu poglavlja).
2. *Podobnost korelacione matrice za faktorizaciju.* Da bi bila podobna za faktorsku analizu, korelaciona matrica bi trebalo da ima barem deo korelacija većih od $r = 0,3$. Bartletov pokazatelj bi trebalo da bude statistički značajan uz $p < 0,05$, a Kajzer-Mejer-Olkinov pokazatelj jednak ili veći od 0,6. Te brojeve treba navesti u rezultatima faktorske analize.
3. *Linearnost.* Pošto je faktorska analiza zasnovana na korelaciji, pretpostavlja se da su promenljive u linearnoj vezi. Svakako ne bi bilo moguće nacrtati i proveriti dijagrame rasturanja svih parova promenljivih. Tabachnick i Fidell (2007) predlažu da se 'nasumično izabere i proveriti' neka kombinacija promenljivih. Ukoliko se ne pojavi jasan dokaz nelinearne (krivolinijske) veze, najčešće je bezbedno nastaviti, ako su dovoljni veličina uzorka i broj slučajeva po svakoj promenljivoj (videti 1. pretpostavku).
4. *Netipične tačke među analiziranim slučajevima.* Faktorska analiza ume da bude osetljiva na netipične tačke, pa bi u sklopu procesa početnog čišćenja podataka (videti poglavlje 6) trebalo da proverite da li ih ima; ukoliko naidete na netipične tačke, uklonite ih ili rešifrujte na manje ekstremnu vrednost.

Postupak faktorske analize

Pre nego što započnete sledeći postupak, otvorite meni **Edit** i u njemu izaberite stavku **Options**; treba da je potvrđeno polje **No scientific notation for small numbers in tables**.

Postupak (1. deo)

1. U glavnom meniju na vrhu prozora otvorite meni **Analyze**, u njemu izaberite stavku **Data Reduction**, pa pritisnite **Factor**.
2. Izaberite sve potrebne promenljive (ili stavke na skali). U ovom slučaju, ja bih izabrala stavke od kojih se sastoji skala PANAS (od pn1 do pn20). Prebacite ih u polje **Variables**.
3. Pritisnite dugme **Descriptives**.
U odeljku **Statistics** treba da je potvrđeno polje **Initial Solution**.
U odeljku **Correlation Matrix**, izaberite opcije **Coefficients** i **KMO and Bartlett's test of sphericity**. Pritisnite **Continue**.

4. Pritisnite dugme **Extraction**.

U odeljku **Method** treba da piše **Principal components** (ili izaberite neku drugu tehniku izdvajanja faktora, recimo **Maximum likelihood** tj. metodu maksimalne verodostojnosti).

U odeljku **Analyze** treba da je potvrđeno polje **Correlation matrix**.

U odeljku **Display**, pritisnite **Screeplot**; treba da je potvrđeno i polje **Unrotated factor solution**.

U odeljku **Extract** treba da je izabrano radio-dugme **Eigenvalues over 1**, ili, kada želite da zadate broj faktora, pritisnite **Number of factors** i upišite taj broj. Pritisnite **Continue**.

5. Pritisnite dugme **Rotation**. Izaberite **Direct Oblimin** i pritisnite **Continue**.

6. Pritisnite dugme **Options**.

U odeljku **Missing Values**, pritisnite **Exclude cases pairwise**.

U odeljku **Coefficient Display Format**, pritisnite **Sorted by size** i **Suppress absolute values less than** _____. U to polje upišite 0,3. To znači da će biti prikazane samo faktorske težine veće od 0,3, da bi se rezultati lakše tumačili.

7. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Ovaj postupak generiše sledeću komandu:

```
FACTOR
/VARIABLES pn1 pn2 pn3 pn4 pn5 pn6 pn7 pn8 pn9 pn10 pn11 pn12
pn13 pn14 pn15 pn16 pn17 pn18 pn19 pn20
/MISSING PAIRWISE /ANALYSIS pn1 pn2 pn3 pn4 pn5 pn6 pn7 pn8
pn9 pn10 pn11 pn12 pn13 pn14 pn15 pn16 pn17 pn18 pn19 pn20
/PRINT INITIAL CORRELATION KMO EXTRACTION ROTATION
/FORMAT SORT BLANK(,3)
/PLOT EIGEN
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PC
/CRITERIA ITERATE(25) DELTA(0)
/ROTATION OBLIMIN
/METHOD=CORRELATION.
```

Evo kako izgleda odabrani deo rezultata prethodnog postupka.

Correlation Matrix

	PN1	PN2	PN3	PN4	PN5	PN6	PN7	PN8	PN9	PN10
Correl PN1	1,000	-.139	-.152	.346	-.071	.352	.407	-.250	.416	-.122
PN2	-.139	1,000	.462	-.141	.271	-.127	-.197	.645	-.188	.411
PN3	-.152	.462	1,000	-.102	.247	-.097	-.255	.494	-.200	.234
PN4	.346	-.141	-.102	1,000	-.156	.295	.331	-.152	.396	-.056
PN5	-.071	.271	.247	-.156	1,000	-.067	-.248	.278	-.201	.258
PN6	.352	-.127	-.097	.295	-.067	1,000	.329	-.048	.426	.077
PN7	.407	-.197	-.255	.331	-.248	.329	1,000	-.232	.481	-.093
PN8	-.250	.645	.494	-.152	.278	-.048	-.232	1,000	-.181	.380
PN9	.416	-.188	-.200	.396	-.201	.426	.481	-.181	1,000	-.070
PN10	-.122	.411	.234	-.056	.258	.077	-.093	.380	-.070	1,000
PN11	-.210	.500	.333	-.179	.266	-.043	-.214	.464	-.210	.583
PN12	.482	-.204	-.135	.315	-.063	.401	.400	-.175	.407	-.074
PN13	.491	-.171	-.165	.329	-.137	.336	.391	-.199	.427	-.114
PN14	-.151	.406	.810	-.107	.302	-.090	-.271	.459	-.198	.263
PN15	.413	-.136	-.085	.317	-.062	.276	.329	-.098	.362	-.067
PN16	-.177	.314	.330	-.121	.539	-.099	-.221	.378	-.164	.314
PN17	.562	-.208	-.190	.368	-.156	.396	.484	-.218	.465	-.134
PN18	.466	-.196	-.181	.338	-.189	.451	.458	-.234	.462	-.066
PN19	-.148	.459	.560	-.124	.285	-.050	-.234	.480	-.198	.339
PN20	-.176	.425	.424	-.171	.245	-.025	-.204	.431	-.219	.367

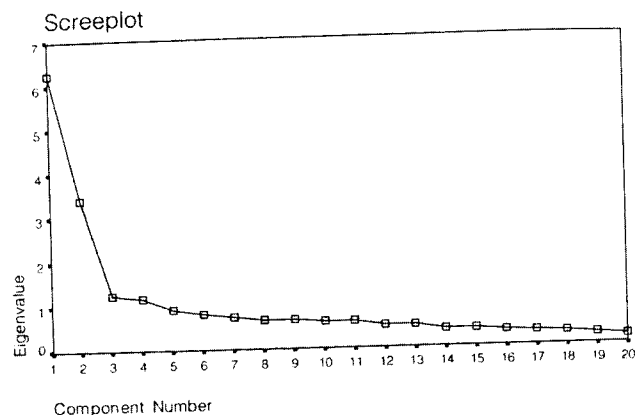
KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.874
Bartlett's Test of Sphericity	Approx. Chi-Square	3966.539
	df	190
	Sig.	.000

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6,250	31,249	31,249	6,250	31,249	31,249
2	3,396	16,979	48,228	3,396	16,979	48,228
3	1,223	6,113	54,341	1,223	6,113	54,341
4	1,158	5,788	60,130	1,158	5,788	60,130
5	.898	4,490	64,619			
6	.785	3,926	68,546			
7	.731	3,655	72,201			
8	.655	3,275	75,476			
9	.650	3,248	78,724			
10	.601	3,004	81,728			
11	.586	2,928	84,656			
12	.499	2,495	87,151			
13	.491	2,456	89,607			
14	.393	1,964	91,571			
15	.375	1,875	93,446			
16	.331	1,653	95,100			
17	.299	1,496	96,595			
18	.283	1,414	98,010			
19	.223	1,117	99,126			
20	.175	.874	100,000			

Extraction Method: Principal Component Analysis.

Component Matrix^a

	Component			
	1	2	3	4
PN17	.679	.474		
PN18	.639	.404		
PN7	.621			
PN8	-.614	.420		
PN9	.609	.323		
PN13	.607	.413		
PN1	.600	.381		
PN2	-.591	.408		
PN3	-.584	.449	-.457	
PN14	-.583	.456	-.451	
PN12	.582	.497		
PN19	-.569	.545		
PN11	-.554	.366	.462	
PN20	-.545	.459		
PN4	.474			
PN15	.477	.483		
PN6	.432	.437		
PN10	-.416	.426	.563	
PN5	-.429			.649
PN16	-.474	.357		.566

^aExtraction Method: Principal Component Analysis.Pattern Matrix^a

	Component			
	1	2	3	4
pn17	.836			
pn12	.795			
pn13	.735			
pn18	.732			
pn15	.721			
pn1	.708			
pn9	.638			
pn6	.604			
pn7	.578			
pn4	.531			
pn3		.909		
pn14		.888		
pn19		.799		
pn20		.677		
pn8		.477	.413	
pn10			.808	
pn11			.707	
pn2		.434	.473	
pn5				.830
pn16				.773

Extraction Method: Principal Component Analysis.

Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 8 iterations.

Tumačenje rezultata

Kao većina ostalih SPSS procedura, i ova daje mnogo rezultata. U ovom odeljku, provešću vas kroz ključni deo potrebnih informacija.

Korak 1

Kako biste proverili da li je skup podataka prikladan za faktorsku analizu, pogledajte da li je vrednost pokazatelja **Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO)** jednaka ili veća od 0,6 i da li je vrednost pokazatelja **Barlett's Test of Sphericity** značajna (tj. da je vrednost **Sig.** 0,05 ili manje). U ovom primeru, pokazatelj KMO iznosi 0,874, a i Bartletov pokazatelj je značajan ($p = 0,000$), te je faktorska analiza opravdana. U tabeli

Correlation Matrix (koja ovde nije prikazana zbog ograničenog prostora), potražite koeficijente korelacije jednake ili veće od 0,3 (videti 2. pretpostavku). Ako ih u vašoj matrici nema baš mnogo, razmotrite da li je faktorska analiza takvih podataka uopšte opravdana.

Korak 2

Da bismo odredili koliko komponenata (faktora) da izdvojimo, razmotrićemo deo rezultata. Po Kajzerovom kriterijumu, zanimaju nas samo komponente čija je karakteristična vrednost 1 ili više. Da bismo odredili koliko komponenata zadovoljava taj kriterijum, pogledaćemo tabelu **Total Variance Explained**. Pročitajte brojeve u prvih nekoliko kolona, pod zaglavljem **Initial Eigenvalues**. Navedene su karakteristične vrednosti svih komponenata. U ovom primeru, samo prve četiri komponente imaju karakteristične vrednosti iznad 1 (6,25, 3,396, 1,223, 1,158). Te četiri komponente objašnjavaju ukupno 60,13 procenata varijanse (videti kolonu **Cumulative %**).

Korak 3

Videćete da je broj komponenata koje zadovoljavaju Kajzerov kriterijum često prevelik, pa obavezno treba pogledati i dijagram prevoja (**Screeplot**) koji je SPSS nacrtao. Potražite na njemu prevojnu tačku. Zadržavaju se samo komponente iznad te tačke. U ovom primeru, sasvim je jasan lom dijagrama na spoju druge i treće komponente. Komponente 1 i 2 objašnjavaju mnogo veći deo varijanse od preostalih komponenata. Na osnovu ovog dijagrama, preporučila bih da se zadrže samo dve komponente (tj. izdvoje samo dva faktora). Postoji još jedna tačka loma, posle četvrte komponente. U zavisnosti od konteksta istraživanja, i njoj bi možda trebalo posvetiti pažnju. Ne zaboravite, faktorska analiza je tehnika istraživanja podataka, pa su tumačenje rezultata i način na koji ćete ih upotrebiti prepušteni vašem sudu, a ne bilo kakvim čvrstim i strogim statističkim pravilima.

Korak 4

Treći način određivanja broja faktora koje treba zadržati jeste paralelna analiza (takođe obrađena u prethodnom delu poglavlja). U tom postupku koristi se spisak karakterističnih vrednosti dat u tabeli **Total Variance Explained** i još neki brojevi koje morate pribaviti iz jednog drugog statističkog programa (koji je napisao Marley Watkins, 2000), dostupnog na pratećoj Web lokaciji knjige. Sledite hiperveze do Web lokacije **Additional Material** i tamošnju komprimovanu datoteku **parallel analysis.zip** preuzmite na svoj računar. Raspakujte je na čvrsti disk i pritisnite datoteku **MonteCarloPA.exe**.

Pokrenuće se program Monte Carlo PCA for Parallel Analysis. Od vas će zahtevati tri podatka: broj promenljivih koje analizirate (u ovom slučaju, 20); broj subjekata u uzorku (u ovom slučaju, 435); i broj replika (zadajte 100). Pritisnite **Calculate**. Program će u pozadini generisati 100 skupova slučajnih brojeva, koji su iste veličine kao datoteka sa stvarnim podacima (20 promenljivih \times 435 slučajeva). Izračunaće i prikazati srednje vrednosti karakterističnih vrednosti tih 100 uzoraka slučajnih brojeva (videti ispis u tabeli 15.1).

Vaš zadatak je da redom uporedite prvu, drugu itd. karakterističnu vrednost dobijenu u SPSS-u sa odgovarajućom prvom, drugom itd. vrednošću dobijenom paralelnom analizom slučajnih brojeva. Ako je SPSS-ov faktor veći od vrednosti praga dobijene paralelnom analizom, zadržite taj faktor; ako je manji, odbacite ga. U tabeli 15.2 dati su zbirni rezultati tog poređenja u ovom primeru. Rezultati paralelne analize podržavaju naš zaključak donesen na osnovu dijagrama prevoja, da za dalje istraživanje zadržimo samo dva faktora.

Tabela 15.1: Rezultati paralelne analize

7/03/2004	11:58:37 AM		
Number of variables:	20		
Number of subjects:	435		
Number of replications:	100		
Eigenvalue	Random Eigenvalue	Standard Dev	
1	1.3984	.0422	
2	1.3277	.0282	
3	1.2733	.0262	
4	1.2233	.0236	
5	1.1832	.0191	
6	1.1433	.0206	
7	1.1057	.0192	
8	1.0679	.0193	
9	1.0389	.0186	
10	1.0033	.0153	
11	0.9712	.0180	
12	0.9380	.0175	
13	0.9051	.0187	
14	0.8733	.0179	
15	0.8435	.0187	
16	0.8107	.0185	
17	0.7804	.0190	
18	0.7449	.0194	
19	0.7090	.0224	
20	0.6587	.0242	
7/03/2004	11:58:50 AM		
MonteCarlo PCA for Parallel Analysis			
Watkins, M. W. (2000). MonteCarlo PCA for parallel analysis [computer software]. State College, PA: Ed & Psych Associates.			

Tabela 15.2: Poređenje karakterističnih vrednosti dobijenih u PCA i vrednosti praga dobijenih paralelnom analizom

Redni broj komponente	Stvama karakteristična vrednost iz PCA	Vrednost dobijena paralelnom analizom	Odluka
1	6.250	1.3984	prihvatiti
2	3.396	1.3277	prihvatiti
3	1.223	1.2733	odbaciti
4	1.158	1.2233	odbaciti
5	.898	1.1832	odbaciti

Korak 5

Vraćamo se na rezultate iz SPSS-a; poslednja tabela koju treba pogledati je **Component Matrix**. U njoj su nerotirane faktorske težine svake od stavki (promenljivih) za te četiri komponente (faktora). SPSS podrazumevano upotrebljava Kajzerov kriterijum (zadržati sve komponente čije su karakteristične vrednosti iznad 1). Videćete u tabeli da većina stavki ima pristojno velike faktorske težine (iznad 0,4) za prve dve komponente. Vrlo malo stavki uopšte ima faktorske težine za komponente 3 i 4. To ukazuje da bi rešenje sa samo dva faktora bilo primerenije.

Korak 6

Pre nego što donesete konačnu odluku o broju faktora, trebalo bi da pogledate rešenje sa četiri rotirana faktora prikazano u tabeli **Pattern Matrix**. Prikazane su faktorske težine (veće od 0,3) za ta četiri faktora; komponenta 1 ima deset faktorskih težina iznad 0,3, komponenta 2 pet, komponenta 3 četiri, a komponenta 4 samo dve. Idealno bi bilo da svaka komponenta ima tri ili više faktorskih težina, pa ovo rešenje nije optimalno, što još jednom podržava našu odluku da zadržimo samo dva faktora.

Uz SPSS-ove podrazumevane opcije dobili smo rešenje sa četiri faktora. Moramo se vratiti i nametnuti rešenje sa dva faktora.

Postupak (2. deo)

Ponovite sve korake iz 1. dela postupka, ali nakon što pritisnete dugme **Extraction**, pritisnite **Number of factors** i upišite broj 2. Pritisnite **Continue**.

Evo dela tih rezultata.

Communalities

	Initial	Extraction
pn1	1,000	,505
pn2	1,000	,516
pn3	1,000	,543
pn4	1,000	,308
pn5	1,000	,258
pn6	1,000	,377
pn7	1,000	,445
pn8	1,000	,553
pn9	1,000	,475
pn10	1,000	,355
pn11	1,000	,440
pn12	1,000	,586
pn13	1,000	,538
pn14	1,000	,548
pn15	1,000	,462
pn16	1,000	,352
pn17	1,000	,686
pn18	1,000	,572
pn19	1,000	,620
pn20	1,000	,507

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	
1	6,250	31,249	31,249	6,250	31,249	31,249	5,277
2	3,396	16,979	48,228	3,396	16,979	48,228	5,157
3	1,223	6,113	54,341				
4	1,158	5,788	60,130				
5	,898	4,490	64,619				
6	,785	3,926	68,546				
7	,731	3,655	72,201				
8	,655	3,275	75,476				
9	,650	3,248	78,724				
10	,601	3,004	81,728				
11	,586	2,928	84,656				
12	,499	2,495	87,151				
13	,491	2,456	89,607				
14	,393	1,964	91,571				
15	,375	1,875	93,446				
16	,331	1,653	95,100				
17	,299	1,496	96,595				
18	,283	1,414	98,010				
19	,223	1,117	99,126				
20	,175	,874	100,000				

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance

Component Matrix^a

	Component	
	1	2
pn17	,679	,474
pn18	,639	,404
pn7	,621	
pn8	-,614	,420
pn9	,609	,323
pn13	,607	,413
pn1	,600	,381
pn2	-,591	,408
pn3	-,584	,449
pn14	-,583	,456
pn12	,582	,497
pn19	-,569	,545
pn11	-,554	,366
pn20	-,545	,459
pn4	,474	
pn16	-,474	,357
pn5	-,429	
pn15	,477	,483
pn6	,432	,437
pn10	-,416	,426

Extraction Method: Principal Component Analysis.
a. 2 components extracted.

Pattern Matrix^a

	Component	
	1	2
pn17	,825	
pn12	,781	
pn18	,742	
pn13	,728	
pn15	,703	
pn1	,698	
pn9	,656	
pn6	,635	
pn7	,599	
pn4	,540	
pn19		,806
pn14		,739
pn3		,734
pn8		,728
pn20		,718
pn2		,704
pn11		,645
pn10		,613
pn16		,589
pn5		,490

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.
a. Rotation converged in 6 iterations.

Structure Matrix

	Component	
	1	2
pn17	,828	
pn12	,763	
pn18	,755	
pn13	,733	
pn1	,710	
pn9	,683	
pn15	,670	
pn7	,646	-,338
pn6	,605	
pn4	,553	
pn19		,784
pn8		,742
pn14		,740
pn3		,737
pn2		,717
pn20		,712
pn11		,661
pn16		,593
pn10		,590
pn5		,505

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Component Correlation Matrix

Component	1	2
1	1,000	-,277
2	-,277	1,000

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Tumačenje rezultata (2. deo: oblimin rotacija dvofaktorskog rešenja)

Prvo treba proveriti procentualni udeo varijanse objašnjen ovim dvofaktorskim rešenjem, prikazan u tabeli **Total Variance Explained**. Dvofaktorsko rešenje objašnjava samo 48,2 procenata varijanse, dok četvorofaktorsko rešenje objašnjava preko 60 procenata.

Nakon rotacije dvofaktorskog rešenja, treba pogledati tri nove tabele na kraju rezultata. Prvo pogledajte tabelu **Component Correlation Matrix** (na samom kraju rezultata). Tu je data jačina korelacije između ta dva faktora (u ovom slučaju, korelacija je sasvim mala, -0,277). Na osnovu tog podatka treba zaključiti da li je bila opravdana pretpostavka da te dve komponente nisu uzajamno zavisne (što je pretpostavka za ispravnost Varimax rotacije) ili treba primeniti (ovde prikazano) rešenje oblimin rotacijom i njega predstaviti u izveštaju.

U ovom slučaju je korelacija između te dve komponente sasvim mala, pa očekujemo da Varimax i oblimin rotacije daju vrlo slična rešenja. Međutim, kada su komponente jače korelirane (npr. iznad 0,3), videćete da će se rezultati tih rotacija razlikovati. Tada treba uraditi oblimin rotaciju i u rezultati-ma navesti njeno rešenje.

Oblimin rotacija daje dve tabele faktorskih težina. U tabeli **Pattern Matrix** prikazane su faktorske težine svih promenljivih. Da biste prepoznali značenje komponente i dali joj prikladno ime, potražite stavke s najvećim faktorskim težinama za nju. U ovom primeru, glavne faktorske težine komponente 1 daju stavke 17, 12, 18 i 13. Pročitajte (na slici 15.1) nazive tih stavki i videćete da se radi o pozitivnim osećanjima (oduševljeno, nadahnu-to, čilo, pažljivo). Glavne stavke za komponentu 2 (19, 14, 3, 8) jesu nega-tivna osećanja (nervozno, uplašeno, prestravljeno, usplahireno). U ovom slučaju je prepoznavanje značenja i imenovanje komponentata bilo lako, ali nije uvek tako.

U tabeli **Structure Matrix**, koju u sklopu rezultata daje samo oblimin rota-cija, date su korelacije između promenljivih i faktora. Kada se predstavlja oblimin rotirano rešenje u izveštaju, daju se obe tabele.

U prethodnom delu rezultata data je tabela **Communalities**. U njoj su brojevi koji predstavljaju zajedničkim faktorima objašnjen deo varijanse za sva-ku promenljivu (stavku). Mali brojevi (npr. manji od 0,3) možda ukazuju na to da se ta stavka ne uklapa dobro u svoju komponentu sa ostalim stavkama. Primera radi, u ovom dvofaktorskom rešenju, stavka pn5 ima, u odnosu na ostale promenljive, najmanji zajedničkim faktorima objašnjen deo varijanse, tj. zajednički varijabilitet (engl. *communality*) od 0,258, a i najmanju faktor-sku težinu (0,49) za komponentu 2 (videti **Pattern Matrix**). Da se bavite po-boljšavanjem ili prečišćavanjem merne skale, na osnovu takvih informacija

stavku biste izbacili sa skale. Uklanjanje stavki s malim delom varijanse koji je objašnjen zajedničkim faktorima, obično poveća ukupnu objašnjenu vari-jansu. Vrednosti zajedničkog varijabiliteta znatno se menjaju u zavisnosti od broja zadržanih faktora, pa ih je obično bolje interpretirati nakon određi-vanja (na osnovu dijagrama prevoja i paralelne analize) broja faktora koji će biti zadržani.

Upozorenje: U ovom primeru smo dobili vrlo 'čist' rezultat. Svaka pro-menljiva je samo po jednoj komponenti dala veliku faktorsku težinu i svakoj komponenti su brojne promenljive dale velike faktorske težine. (To je pri-mer 'jednostavne strukture'.) Raspravu o tome videti u knjizi koju su napi-sale Tabachnick i Fidell (2007, str. 647). Nažalost, tako jednostavan rezultat se ne dobija za sve podatke. Promenljive često daju umereno velike faktorske težine raznim komponentama, a nekim komponentama faktorske težine daju samo jedna ili dve promenljive. U takvim slučajevima, možda bi trebalo rotirati neki drugi broj komponentata (npr. jednu više i jednu manje) da se vidi može li se otkriti bolje rešenje. Kada utvrdite da neke promenljive naprosto ne daju težine dobijenim komponentama, razmislite o tome da ih uklonite i zatim ponovite analizu. Da biste umeli da donesete takve odluke, pročitajte što više o tome. Razumljiv priručnik za početnike je knjiga koju su napisali Pett, Lackey & Sullivan (2003).

Predstavljanje rezultata faktorske analize

Vrsta i količina informacija koje treba navesti u odeljku rezultata menja se u zavisnosti od oblasti istraživanja, vrste izveštaja koji se priprema i mesta na kojem će biti objavljen. Oni koji objavljuju radove iz oblasti psihologije i obrazovanja, moraju da zadovolje stroge zahteve u pogledu obaveznog sadržaja članka o istraživanju koje obuhvata faktorsku analizu. Treba nave-sti pojedinosti o upotrebljenoj metodi izdvajanja faktora, kriterijumima po-moću kojih je određen broj faktora (među njima obavezno i paralelnoj analizi), upotrebljenoj vrsti rotacije faktora (npr. Varimax, Oblimin), ukup-noj objašnjenoj varijansi, početnim iznosima karakterističnih vrednosti i iz-nosima nakon rotacije.

Tabelarno treba prikazati sve faktorske težine (ne samo one iznad 0,3). Kada je rešenje dobijeno Varimax rotacijom, tabelu treba nazvati 'faktorske težine / koeficijenti korelacije promenljivih i faktora'. Ukoliko je rešenje do-bijeno oblimin rotacijom, treba zasebno navesti i faktorske težine i koefici-jente korelacije promenljivih i faktora (može i u istoj tabeli, kao dole), te koeficijente međusobne korelacije faktora.

Rezultati faktorske analize u gornjem primeru mogli bi se predstaviti ovako:

20 stavki skale pozitivnih i negativnih osećanja (Positive and Negative Affect scale, PANAS) bilo je podvrgnuto analizi glavnih komponentata (PCA) u 15. verziji SPSS-a. Pre sprovođenja PCA, bila je ocenjena prikladnost podataka za faktorsku analizu. Pregledom korelacione matrice otkriveno je mnogo koeficijentata vrednosti 0,3 i više. Vrednost Kajzer-Mejer-Oklinovog pokazatelja bio je 0,87, što premašuje preporučenu vrednost 0,6 (Kaiser, 1970, 1974). I Bartletov test sferičnosti (Bartlett, 1954) dostigao je statističku značajnost, što sve ukazuje na faktorabilnost korelacione matrice.

Analiza glavnih komponentata otkrila je prisustvo četiri komponente s karakterističnim vrednostima preko 1, koje objašnjavaju 31,2%, 17%, 6,1% i 5,8% varijanse. Pregledom dijagrama prevoja utvrđeno je postojanje jasne tačke loma iza druge komponente. Na osnovu Katelovog kriterijuma (1966) odlučeno je da se za dalje istraživanje zadrže dve komponente. To su podržali i rezultati paralelne analize, sa samo dve komponente čije karakteristične vrednosti premašuju odgovarajuće vrednosti praga dobijene pomoću jednako velike matrice slučajnih brojeva (20 promenljivih \times 435 ispitanika).

To dvokomponentno rešenje objasnilo je ukupno 48,2% varijanse, pri čemu je doprinos 1. komponente 31,25%, a 2. komponente 17,0%. Da bi se lakše protumačile te dve komponente, sprovedena je oblimin rotacija. Rotirano rešenje je otkrilo postojanje jednostavne strukture (Thurstone, 1947), pri čemu obe komponente imaju mnogo velikih faktorskih težina i sve promenljive daju znatne težine samo po jednoj od komponentata. Tumačenje te dve komponente bilo je saglasno s prethodnim istraživanjima skale PANAS, pri čemu su stavke pozitivnih osećanja dale velike težine komponenti 1, a stavke negativnih osećanja velike težine komponenti 2. Između ta dva faktora postoji slaba negativna korelacija ($r = -0,28$). Rezultati ove analize podržavaju korišćenje stavki pozitivnih osećanja i stavki negativnih osećanja u obliku zasebnih skala, kao što su predložili autori skale (Watson, Clark & Tellegen, 1988).

U izveštaju treba navesti i matricu svih faktorskih težina (ne samo onih većih od 0,3) i matricu korelacije promenljivih i faktora. Ukoliko je rešenje dobijeno oblimin rotacijom, treba zasebno navesti i faktorske težine i koeficijente korelacije promenljivih i faktora (može i u istoj tabeli, kao dole), te koeficijente međusobne korelacije faktora. Da biste dobili sve faktorske težine, morate ponoviti analizu za odabrano konačno rešenje (npr. u ovom slučaju za dvofaktorsku oblimin rotaciju), ali ovoga puta uz isključenu opciju da se prikažu samo koeficijenti iznad 0,3 (videti odeljak sa opisom postupka). Pritisnite **Options** i u odeljku **Coefficient Display Format** uklonite kvačicu iz drugog polja, **Suppress absolute values less than ,3**.

Tabela 1: Matrica faktorskih težina i korelacija promenljivih i faktora za PCA sa oblimin rotacijom dvofaktorskog rešenja za stavke skale PANAS

Stavka	Faktorske težine		Koeficijenti korelacije promenljivih i faktora		Deo varijanse objašnjen zajedničkim faktorima
	Komponenta 1	Komponenta 2	Komponenta 1	Komponenta 2	
17. oduševljeno	0,825	-0,012	0,828	-0,241	0,686
12. nadahnuto	0,781	0,067	0,763	-0,149	0,586
18. čilo	0,742	-0,047	0,755	-0,253	0,572
13. pažljivo	0,728	-0,020	0,733	-0,221	0,538
15. uzbuđeno	0,703	0,119	0,710	-0,236	0,462
1. zainteresovano	0,698	-0,043	0,683	-0,278	0,505
9. snažno	0,656	-0,097	0,670	-0,076	0,475
6. odlučno	0,635	0,107	0,646	-0,338	0,377
7. aktivno	0,599	-0,172	0,605	-0,069	0,445
4. ponosno	0,540	-0,045	0,553	-0,195	0,308
19. nervozno	0,079	0,806	-0,144	0,784	0,620
14. uplašeno	-0,003	0,739	-0,253	0,742	0,548
3. prestravljeno	-0,010	0,734	-0,207	0,740	0,543
8. ucveljeno	-0,052	0,728	-0,213	0,737	0,553
20. usplahireno	0,024	0,718	-0,242	0,717	0,507
2. uzrujano	-0,047	0,704	-0,175	0,712	0,516
11. razdražljivo	-0,057	0,645	-0,236	0,661	0,440
10. odbojno	0,080	0,613	-0,176	0,593	0,355
16. kao krivac	-0,013	0,589	-0,090	0,590	0,352
5. posramljeno	-0,055	0,490	-0,191	0,505	0,258

Napomena. Podebljane su glavne faktorske težine / koeficijenti korelacije za svaku stavku.

Ako rezultate ove analize predstavljate u tezi (za razliku od časopisa), trebalo bi da u dodatak uključite i dijagram prevoja i tabelu nerotiranih faktorskih težina (iz tabele **Component Matrix**). Tako bi čitalac teze dobio priliku da proveriti slaže li se s vašom odlukom da zadržite samo dve komponente.

Predstavljanje rezultata u časopisu obično je mnogo kraće, zbog ograničenog prostora. Ukoliko želite da vidite objavljen članak o faktorskoj analizi, potražite ga na adresi <http://www.hqlo.com/content/3/1/82> ili sledite hiperveze na pratećoj Web lokaciji ove knjige.

Dodatne vežbe

Poslovno okruženje

Datoteka s podacima: **staffsurvey3ED.sav**. Pojediniosti o datoteci videti u dodatku.

1. U skladu sa uputstvima datim u celom poglavlju, sprovedite analizu glavnih komponenata sa obličin rotacijom za deset stavki slaganja (od Q1a do Q10a) koje sačinjavaju anketu Staff Satisfaction Survey. Videćete da dijagram prevoja ukazuje na to da treba zadržati samo jednu komponentu, iako dva faktora imaju karakteristične vrednosti veće od 1. Pokrenite paralelnu analizu za 523 slučaja i 10 stavki. Njeni rezultati pokazuju da samo jedna komponenta ima karakterističnu vrednost koja premašuje ekvivalentnu veličinu dobijenu iz skupa slučajnih brojeva, što ukazuje na to da stavke skale Staff Satisfaction ocenjuju samo jednu zajedničku dimenziju (faktor).

Zdravstvo

Datoteka s podacima: **sleep3ED.sav**. Pojediniosti o datoteci videti u dodatku.

1. Pomoću postupaka prikazanih u poglavlju 15 istražite zajedničku osnovu niza pitanja osmišljenih za ocenu uticaja problema sa spavanjem na razne aspekte života. Te stavke su nazvane od *impact1* do *impact7*. Pokrenite paralelnu analizu (za 121 slučaj i 7 stavki) da biste proverili koliko faktora treba zadržati.

DEO V

Statističke tehnike za poređenje grupa

U petom delu ove knjige, istražujemo neke SPSS-ove tehnike za ocenjivanje razlike između grupa ili okolnosti. Te tehnike su vrlo složene i oslanjaju se na obimnu teoriju i statistička načela. Pre nego što započnete analizu u SPSS-u, neophodno je da steknete barem elementarno znanje o statističkim tehnikama koje nameravate da upotrebite. U tome vam može pomoći mnogo dostupnih udžbenika iz statistike. (Na kraju ove knjige, u odeljku Preporučena literatura, dat je spisak nekih prikladnih naslova.) Bilo bi dobro da tu građu pregledate odmah. Tako biste lakše shvatili šta SPSS izračunava, šta to znači i kako protumačiti složene nizove brojeva u njegovim rezultatima. U narednim poglavljima pretpostavljam da već imate osnovna znanja o statistici i da poznajete njenu terminologiju.

Tehnike obrađene u petom delu knjige

Za ispitivanje značajnih razlika između grupa postoji ceo niz tehnika. Programski paket SPSS sadrži mnogo statističkih tehnika, a ovde ću obraditi samo one glavne, i to i parametarske i neparametarske. Parametarske tehnike počivaju na više pretpostavki o populaciji iz koje je izvučen uzorak (npr. da su rezultati normalno raspodeljeni) i prirodi tih podataka (da su mereni na intervalnim skalama). Neparametarske tehnike nemaju tako stroge pretpostavke i često su prikladnije za male uzorke ili kada su prikupljeni podaci izmereni samo na ordinalnim skalama (čiji se iznosi mogu rangirati). U narednoj tabeli dat je spisak svih tehnika obrađenih u ovom poglavlju.

Lista parametarskih tehnika i njihovih neparametarskih ekvivalenata obrađenih u petom delu

Parametarska tehnika	Neparametarska tehnika
Ne postoji	hi-kvadrat za kvalitet aproksimacije (fitovanja)
Ne postoji	hi-kvadrat za nezavisnost
T-test nezavisnih uzoraka	Man-Vitnijev U test
T-test uparenih uzoraka	Vilkoksonov test ranga
Jednofaktorska ANOVA različitih grupa	Kruskal-Volisov test
Jednofaktorska ANOVA ponovljenih merenja	Fridmanov test
Dvofaktorska analiza varijanse (različitih grupa)	Ne postoji
Kombinovana ANOVA	Ne postoji
Multivarijaciona analiza varijanse (MANOVA)	Ne postoji
Analiza kovarijanse	Ne postoji

U poglavlju 10, provela sam vas kroz proces odlučivanja o tome koja je statistička tehnika prikladna za konkretno istraživačko pitanje. Odgovor se menja u zavisnosti od prirode istraživačkog pitanja, vrste podataka na raspolaganju i broja promenljivih i grupa. (Ako dosad niste pročitali to poglavlje, trebalo bi da to uradite pre nego što nastavite.) Evo ključnih tačaka koje treba imati u vidu kada se traži prikladna statistička tehnika:

- T-testovi se upotrebljavaju kada imate samo dve grupe (npr. muškarci/žene) ili dve tačke u vremenu (npr. pre intervencije, posle intervencije).
- Tehnike analize varijanse se upotrebljavaju kada imate *dve ili više* grupa ili tačaka u vremenu.
- Tehnike uparenih uzoraka ili ponovljenih merenja upotrebljavaju se za testiranje *istih ljudi u više navrata*, ili kada imate uparene uzorke.
- Tehnike analize različitih grupa i nezavisnih uzoraka upotrebljavaju se kada su subjekti u svim grupama *različiti ljudi* (ili nezavisni).
- Jednofaktorska analiza varijanse se upotrebljava kada imate samo jednu nezavisnu promenljivu (npr. pol).
- Dvofaktorska analiza varijanse se upotrebljava kada imate dve *nezavisne* promenljive (pol, starosna grupa).
- Analiza varijanse više zavisnih promenljivih (multivarijaciona analiza) upotrebljava se kada imate više *zavisnih* promenljivih (anksioznost, depresija).
- Analiza kovarijanse (ANCOVA) upotrebljava se kada treba statistički kontrolisati (ukloniti) uticaj dodatne, remetilačke promenljive koja utiče na vezu između nezavisne i zavisne promenljive.

Pre nego što se upustimo u istraživanje nekih dostupnih tehnika, treba razmotriti više zajedničkih pitanja. Te teme će biti relevantne za mnoga poglavlja iz ovog dela knjige, pa ćete se vraćati ovde kako budete prolazili kroz ostatak knjige.

Pretpostavke

Na nekim opštim pretpostavkama počivaju sve ovde razmotrene parametarske tehnike (npr. t-testovi, analiza varijanse), a pojedine tehnike se zasnivaju i na dodatnim pretpostavkama. Opšte pretpostavke predstavljamo u ovom poglavlju, a specifičnije pretpostavke u narednim poglavljima, po potrebi. Trebalo bi da se vraćate na ovaj uvod kada budete primenjivali neku od tehnika predstavljenih u petom delu knjige. Razmatranje postupaka provere zadovoljenja pretpostavki videti u Tabachnick i Fidell (2007, 4. poglavlje). Dalje razmatranje posledica narušavanja tih pretpostavki videti u knjigama Stevens (1996, 6. poglavlje) i Glass, Peckham i Sanders (1972).

Nivo merenja

Za sve parametarske pristupe pretpostavka je da se zavisna promenljiva meri na intervalnoj skali (pa rastojanja između brojčanih vrednosti odgovaraju rastojanjima između obeležja koja se mere). Dakle, upotrebljava se neprekidna skala, a ne diskretne kategorije. Kad god je to moguće prilikom projektovanja istraživanja, pokušajte da upotrebite neprekidne mere zavisne promenljive, a ne kategorijske. Tako stičete mogućnost upotrebe većeg broja tehnika analize podataka.

Slučajnost uzorkovanja

Parametarske tehnike obrađene u petom delu knjige zasnivaju se na pretpostavci da su rezultati dobijeni iz slučajnog uzorka populacije. U stvarnim istraživanjima ta pretpostavka često nije zadovoljena.

Nezavisnost opservacija

Opservacije od kojih se sastoje podaci moraju biti uzajamno nezavisne, tj. ni na jednu opservaciju ili merenje ne sme uticati nijedna druga opservacija ili merenje. Kršenje ove pretpostavke ima vrlo ozbiljne posledice; raspravu videti u knjizi Stevensa (1996, str. 238). U brojnim istraživačkim situacijama, krši se pretpostavka o nezavisnosti. U nastavku dajem primere takvih studija (preuzete iz knjiga koje su napisali Stevens, 1996, str. 239; i Gravetter & Wallnau, 2004, str. 251):

- Istraživanje učinka studenata koji rade u parovima ili malim grupama. Ponašanje svakog člana grupe utiče na sve ostale članove i tako se krši pretpostavka o nezavisnosti.

- Istraživanje navika i preferencija dece u vezi s gledanjem TV-a, kada su deca iz iste porodice. Ponašanje jednog deteta u porodici (koje, recimo, gleda program A) najčešće utiče na svu ostalu decu iz te porodice; zato opservacije nisu nezavisne.
- Istraživanje metoda poučavanja u učionici i njihovog uticaja na ponašanje i učinak studenata. Prisustvo malog broja problematičnih studenata može uticati na sve ostale; zato merenja ponašanja i učinka pojedinaca nisu nezavisna.

Za svaku situaciju u kojoj se opservacije ili merenja prikupljaju u grupnom okruženju ili su subjekti podvrgnuti nekom obliku interakcije, mogu biti potrebne specifičnije tehnike, kao što je modelovanje u više nivoa (hijerarhijsko). Taj pristup je sada uobičajen u istraživanjima koja obuhvataju decu u učionici, u školama, u gradovima; ili studije s pacijentima, različitim medicinskim specijalistima, u ordinaciji/kancelariji, u gradu ili zemlji. Više o tome videti u 15. poglavlju knjige Tabachnicka i Fidella (2007).

Normalnost raspodele

Parametarske tehnike daju tačne rezultate samo kada su populacije iz kojih su uzorci uzeti normalno raspodeljene. U mnogim istraživanjima (naročito u društvenim naukama), vrednosti zavisne promenljive nisu normalno raspodeljene. Srećom, većina tehnika je prilično robusna, tj. narušavanje ove pretpostavke prouzrokuje malu netačnost rezultata. Kada su uzorci dovoljno postavke prouzrokuje malu netačnost rezultata. Kada su uzorci dovoljno veliki (npr. preko 30 opservacija), kršenje ove pretpostavke ne bi trebalo da prouzrokuje veće probleme. Raspodela rezultata za svaku grupu može se proveriti pregledom histograma koji se crtaju pomoću SPSS-ovog menija **Graphs** (videti 7. poglavlje). Podrobniji opis tog procesa videti u knjizi Tabachnicka i Fidella (2007, 4. poglavlje).

Homogenost varijanse

Parametarske tehnike iz ovog dela knjige temelje se na pretpostavci da su uzorci dobijeni iz populacija jednakih varijansi. To znači da je promenljivost rezultata u svim grupama jednaka. Za ispitivanje te homogenosti u sklopu t-testova i analiza varijansi, SPSS obavlja Leveneov test jednakosti varijansi. Njegovi rezultati su deo SPSS-ovog izlaza za te tehnike. Budite pažljivi u tumačenju rezultata tog testa: nadate se da pokazatelj *nije* značajan (tj. da mu je značajnost *veća* od 0,05). Kada dobijete značajnost manju od 0,05, to znači da varijanse dveju grupa nisu jednake i da ne važi pretpostavka o homogenosti varijanse. Ne paničite kada vam se to dogodi. Analiza varijanse je prilično neosetljiva na narušavanje te pretpostavke ukoliko su veličine grupa približno slične (npr. najveća/najmanja = 1,5; Stevens, 1996, str. 249). U t-testovima dobijate dva niza rezultata, jedan za situacije u kojima pretpostavka nije narušena, a drugi za one kada jeste. U tom slučaju, samo upotrebite onaj skup rezultata koji odgovara podacima.

Greška prve vrste, greška druge vrste i moć testa

T-testovi i analize varijansi služe za ispitivanje hipoteza. U toj vrsti analiza uvek je moguće doneti pogrešan zaključak. Mogu se napraviti dve različite greške. Kada odbacimo nultu hipotezu koja je u stvari tačna, radi se o grešci prve vrste. To se dešava kada zaključimo da između grupa postoji razlika, a ona zapravo ne postoji. Verovatnoću te greške minimiziramo tako što izaberemo malu vrednost alfa; najčešće se koriste vrednosti 0,05 ili 0,01, tj. svesno rizikujemo da jedanput u 20 odnosno 100 slučajeva odbacimo nultu hipotezu (o nepostojanju razlike) kada je tačna.

U ispitivanju (verifikaciji) hipoteza može se napraviti i druga vrsta greške, kada ne odbacimo pogrešnu nultu hipotezu (tj. kada zaključimo da između grupa ne postoji razlika, a ona zapravo postoji). Nažalost, te dve greške su obrnuto srazmerne. Što više smanjujemo verovatnoću greške I vrste, povećavamo verovatnoću da ćemo napraviti grešku II vrste.

Idealno bi bilo da se pomoću upotrebljenih testova tačno utvrdi postoji li zaista razlika između grupa. To je tzv. *moć* testa, tj. verovatnoća da se otkrije postojeća razlika, odnosno da se ne napravi greška II vrste. Testovi se razlikuju po svojoj moći; recimo, kada su njihove osnovne pretpostavke zadovoljene, parametarski testovi (kao što su t-testovi, analiza varijanse itd.) potencijalno su moćniji od neparametarskih testova. Međutim, na moć testa u datoj situaciji utiču i drugi činioci:

- veličina uzorka;
- veličina uticaja istraživane razlike između grupa, tj. uticaja nezavisne promenljive; i
- alfa nivo (rizik greške I vrste) koji je zadao istraživač (npr. 0,05/0,01).

Moć testa se jako menja u zavisnosti od veličine uzorka upotrebljenog u studiji. Stevens (1996) tvrdi da 'moć ne predstavlja problem' (str. 6) kada je uzorak velik (npr. 100 ili više subjekata). Međutim, u istraživanjima na malim uzorcima (npr. $n=20$) morate biti svesni mogućnosti da neznačajan rezultat može biti posledica nedovoljne moći testa. Stevens (1996) predlaže da se za male grupe po potrebi poveća alfa (rizik greške I vrste), npr. na 0,10 ili 0,15 umesto uobičajenih 0,05.

Postoje i tabele (videti Cohen, 1988) u kojima se može očitati veličina uzorka potrebna za određenu (dovoljnu) moć testa, za datu veličinu uticaja razlike koju želite da otkrijete. Sve više je i softverskih programa koji to umeju da izračunaju (npr. G*Power, dostupan na adresi <http://www.psych.uni-duesseldorf.de/aap/projects/gpower/>).

I neki SPSS programi izračunavaju moć obavljenog testa, vodeći računa o veličini uticaja razlike i veličini uzorka. Idealno bi bilo da imate 80 procenata šanse da otkrijete postojanje veze. Kada dobijete neznačajan rezultat i

imate veoma mali uzorak, trebalo bi da pogledate moć upotrebljenog testa. Uz moć testa manju od 0,80 (80 procenata šanse da otkrijete postojeću razliku), treba pažljivo protumačiti razlog neznačajnosti rezultata. Ona može biti posledica nedovoljne moći testa, a ne nepostojanja razlike između grupa. Analiza moći pokazuje koliko poverenja treba imati u rezultate kada se ne odbaci nulta hipoteza o jednakosti grupa. Što je veća moć testa, to više tada treba biti uveren da stvarna razlika između grupa ne postoji.

Planirana poređenja/naknadne analize

Analizom varijanse utvrđujete da li postoje značajne razlike između raznih grupa ili okolnosti. Katkada će vas zanimati da li se grupe kao celina razlikuju (da li nezavisna promenljiva na neki način utiče na vrednosti zavisne promenljive). U drugim istraživačkim kontekstima usredsredićete se više na ispitivanje razlika između pojedinih, za razliku od svih mogućih, grupa. Vodite računa o toj razlici, pošto se za svaku od tih namena koristi drugačija analiza.

Planirana (ili *a priori*) poređenja služe za ispitivanje konkretnih hipoteza (obično izvučenih iz teorije ili prethodnih istraživanja) u vezi s razlikama unutar određenog podskupa grupa (npr. da li se grupe 1 i 3 značajno razlikuju?). Ta poređenja treba specificirati (isplanirati) pre nego što analizirate podatke, umesto da pecate po rezultatima u nadi da ćete naći nešto zanimljivo!

Ukoliko nameravate da specificirate mnogo različitih, ali istovremenih poređenja, moraćete postupati pažljivo. Planirana poređenja ne uklanjaju povećani rizik od grešaka I vrste, koji je posledica velikog broja paralelnih hipoteza koje se ispituju. Greška I vrste znači odbaciti nultu hipotezu (npr. da nema razlike između grupa) koja je u stvari tačna. Drugim rečima, povećan je rizik da ćete misliti da ste otkrili značajan rezultat (razliku), a on je zapravo sasvim slučajan. Kada istražujete veliki broj razlika, bezbedniji je drugačiji pristup tj. naknadna (post-hoc ili *a posteriori*) poređenja, koja štite od grešaka I vrste.

Treća mogućnost je da na alfa nivo (rizik greške I vrste) koji ćete upotrebiti za procenu statističke značajnosti, primenite tzv. Bonferonijevo prilagođenje. To znači zadati stroži alfa nivo za svako poređenje, da bi alfa u svim testovima zajedno ostao na razumnom nivou. To se postiže deljenjem alfa nivoa (najčešće 0,05) brojem poređenja koje nameravate da obavite; zatim se ta nova vrednost koristi kao zahtevani alfa nivo. Primera radi, za tri nameravana poređenja novi alfa nivo bio bi 0,05 podeljeno s 3, što je jednako 0,017. Raspravu o tome videti u Tabachnick i Fidell (2007, str. 52).

Naknadna (post-hoc ili *a posteriori*) poređenja upotrebljavaju se kada želite da obavite ceo niz poređenja, tj. istražite razlike između svih mogućih grupa ili uslova u studiji. Ukoliko odaberete taj pristup, analiza treba da se sastoji od dva koraka. Prvo se izračuna ukupan F pokazatelj koji kazuje ima li značajnih razlika između grupa u projektu. Ako je ukupan F pokazatelj

značajan (što ukazuje da postoji razlika između grupa), možete nastaviti i obaviti dodatne testove za identifikaciju tih razlika (npr. da li se Grupa 1 razlikuje od Grupe 2 ili Grupe 3, da li se razlikuju Grupa 2 i Grupa 3).

Naknadna poređenja štite od moguće greške I vrste kao posledice velikog broja različitih poređenja. To se postiže zadavanjem strožih kriterijuma za značajnost, koju je utoliko teže postići. Kod malih uzoraka to ume da bude problem, zato što je ponekad vrlo teško dobiti značajan rezultat, čak i kada je vidljiva razlika u rezultatima između grupa veoma velika.

Postoje i brojni naknadni testovi koji se razlikuju po svojoj prirodi i strogoći. Razlikuju se i pretpostavke na kojima se zasnivaju. U nekima se pretpostavlja da su varijanse dve grupe jednake (npr. Tukey); u drugima se ne pretpostavlja jednakost varijansi (npr. Dunnettov C test). Među najčešće upotrebljavanim naknadnim testovima su Tukejev test „zaista značajne različitosti“ (engl. *Honestly Significant Different*, HSD) i Šefeov (Scheffe) test. Od ta dva, Šefeov test je bezbednija metoda za smanjenje rizika od greške I vrste. Međutim, to se plaća u moći. Tim testom je teže (manje verovatno) otkriti stvarno postojeću razliku između grupa.

Veličina uticaja

Sve tehnike razmotrene u ovom delu knjige pokazuju da li je razlika između grupa statistički značajna (tj. neslučajna). Za većinu istraživača i studenata nastupi trenutak uzbuđenja kada utvrde da su njihovi rezultati statistički značajni! Međutim, istraživanje znači više od dobijanja statističke značajnosti. Verovatnoća ne pokazuje stepen povezanosti promenljivih (jačinu veze). Kada su izračunate za velike uzorke, čak i vrlo male razlike između grupa postaju statistički značajne. To ne znači da je ta razlika dovoljno velika da bi imala ikakvu praktičnu ili teorijsku važnost.

Jedan od načina da ocenite važnost svojih rezultata jeste da izračunate veličinu uticaja (engl. *effect size*), tj. jačinu veze između promenljivih. To je skup pokazatelja koji pokazuje relativnu veličinu razlika između srednjih vrednosti ili iznos ukupne varijanse u zavisnoj promenljivoj koji se može predvideti na osnovu poznavanja vrednosti nezavisne promenljive (Tabachnick & Fidell, 2007, str. 54).

Ima više pokazatelja veličine uticaja. Za poređenje grupa najčešće se upotrebljavaju pokazatelji *parcijalni eta kvadrat* i *Koenov d*. SPSS izračunava parcijalni eta kvadrat u sklopu analize varijanse, ali ne i u sklopu t-testova; no, to je lako izračunati na osnovu drugih njegovih rezultata.

Pokazatelj veličine uticaja *parcijalni eta kvadrat* srazmeran je delu varijanse zavisne promenljive koji je objašnjen nezavisnom promenljivom. Može imati vrednosti u opsegu od 0 do 1. S druge strane, *Koenov d* predstavlja razliku između grupa izraženu brojem standardnih odstupanja. Pazite da ne pobrkate te pokazatelje kada budete tumačili jačinu veze. Adresa Web

lokacije na kojoj se brzo i lako mogu izračunati oba spomenuta pokazatelja veličine uticaja glasi: <http://web.uccs.edu/lbecker/Psy590/escalc3.htm>.

Cohen (1988, str. 22) predložio je sledeće smernice za tumačenje veličine uticaja (kada se ocenjuje istraživanje koje obuhvata poređenje grupa). [Koenova preporuka se odnosi na pokazatelj eta kvadrat, ali se može primeniti i na tumačenje pokazatelja parcijalni eta kvadrat. Formula za parcijalni ima neznatno drugačiji imenilac. Više o tome pročitajte u knjizi Tabachnicka i Fidella, 2007, str. 55.]

Veličina uticaja	Eta kvadrat (% objašnjene varijanse)	Koenov d (jedinica standardnog odstupanja)
Mali	0,01 ili 1%	0,2
Srednji	0,06 ili 6%	0,5
Veliki	0,138 ili 13,8%	0,8

Vodite računa o tome da Koen daje drugačije smernice za korelacione projekte (obrađene u četvrtom delu knjige). Gornje vrednosti važe za poređenja grupa.

Nedostajući podaci

Kada radite istraživanje, naročito ono sa ljudima, vrlo retko ćete u svakoj opservaciji tj. od svakog ispitanika dobiti kompletne podatke. Zato pregledajte šta nedostaje u datoteci s podacima. Pokrenite proceduru **Descriptives** i utvrdite procentualni udeo nedostajućih podataka za svaku promenljivu. Kada određenoj promenljivoj neočekivano nedostaje mnogo podataka, zapitajte se zašto je tako? Trebalo bi da razmislite i o tome da li su nedostajući podaci raspodeljeni nasumično po slučajevima ili tu postoji neka pravilnost (npr. mnoge žene nisu odgovorile na pitanje o svojoj starosti). SPSS ima proceduru **Missing Value Analysis** koja olakšava pronalaženje pravilnosti u nedostajućim podacima (videti poslednju opciju u meniju **Analyze**). Više o tome pročitajte u 4. poglavlju knjige Tabachnicka i Fidella, 2007.

Takođe razmotrite šta ćete s nedostajućim podacima kada dođe vreme da uradite statističke analize. U mnogim SPSS-ovim statističkim procedurama, preko dugmeta **Options** birate način na koji će SPSS tretirati nedostajuće podatke. Birajte pažljivo, pošto time znatno utičete na statističke rezultate. To je posebno važno kada prilažete spisak promenljivih i za sve njih ponavljate iste analize (npr. korelacije između grupe promenljivih, t-testovi niza zavisnih promenljivih).

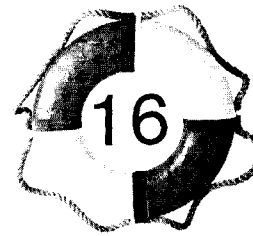
- Opcija **Exclude cases listwise** znači da će biti analizirani samo slučajevi u kojima za *sve* promenljive navedene u polju **Variables** postoje *svi* podaci. Svaki slučaj kome nedostaje makar i delić podataka uopšte neće biti analiziran. Time biste znatno i nepotrebno ograničili veličinu uzorka.

- Opcija **Exclude cases pairwise** (katkada piše *Exclude cases analysis by analysis*) znači da će slučaj (osoba) biti isključen samo iz onih analiza za koje mu nedostaje neki od neophodnih podataka. Dakle, i takvi slučajevi će biti analizirani kad god je to moguće, tj. kad god imaju podatke potrebne za određenu analizu.
- Opcija **Replace with mean**, dostupna u nekim SPSS-ovim statističkim postupcima (npr. višestrukoj regresiji), znači da će biti izračunata srednja vrednost svih promenljivih i da će njome biti zamenjeni nedostajući podaci. Ovu opciju *nikada* ne bi trebalo da koristite, pošto može znatno da iskrivi rezultate analize, naročito onda kada nedostaje mnogo podataka.

Kad god radite neki statistički postupak, pritisnite dugme **Options** i proverite koja je od navedenih opcija potvrđena (jer se podrazumevana opcija menja u zavisnosti od postupka). Ako nemate jak razlog da postupite drugačije, preporučila bih vam da slučajeve isključite samo iz onih analiza za koje im nedostaju podaci (**pairwise**). Jedina situacija kada bi vam moglo zatrebati da analize ograničite samo na slučajeve koji imaju podatke za sve promenljive (**listwise**) jeste ona kada treba razmotriti samo podskup slučajeva koji daje potpun skup rezultata.

Brojevi koji čudno izgledaju

Izlazni rezultat iz SPSS-a ponekad sadrži čudne brojeve oblika 1,24E-02. To su mali brojevi koje je SPSS prikazao u tzv. naučnoj notaciji. Da vam se to ne bi dešavalo, u glavnom meniju izaberite **Edit**, zatim **Options**, i potvrdite opciju **No scientific notation for small numbers in tables** na kartici **General**.



Neparametarske tehnike

U knjigama o statistici često se govori o dve vrste statističkih tehnika: parametarskim i neparametarskim. Po čemu se razlikuju ta dva skupa tehnika? Zašto je ta razlika važna? Reč 'parametarski' potiče od 'parametar', tj. obeležje populacije. U parametarskim testovima (npr. t-testovi, analiza varijanse), pretpostavljaju se neka svojstva populacije iz koje je uzorak uzet. To su najčešće pretpostavke o obliku raspodele populacije (npr. da je normalno raspodeljena). S druge strane, zahtevi za neparametarske tehnike nisu tako strogi i nista se ne pretpostavlja o pripadnoj raspodeli populacije. (Zato se katkada nazivaju testovima bez raspodele, engl. *distribution-free tests*.)

Uprkos tome sto su manje pipave, neparametarske statističke tehnike imaju i svojih loših strana. Manje su osetljive od odgovarajućih parametarskih rođaka i zato rede otkrivaju postojeće razlike između grupa. Kada imate pravu vrstu podataka, uvek je bolje primeniti parametarsku tehniku. Dakle, u kojim okolnostima biste izabrali da upotrebite neparametarsku statističku tehniku?

Neparametarske tehnike su idealne za podatke merene na nominalnim (kategorijskim) ili ordinalnim skalama (čiji se iznosi mogu rangirati). Korisne su i kada imate veoma male uzorke ili kada podaci ne zadovoljavaju stroge pretpostavke parametarskih tehnika. U SPSS-u se mogu koristiti najrazličitije neparametarske tehnike, ali ćemo u ovom poglavlju razmotriti samo one glavne. Evo koje će teme biti obrađene i koje su njihove parametarske alternative (obrađene u kasnijim poglavljima).

Kratak pregled tehnika obrađenih u ovom poglavlju

Neparametarska tehnika	Parametarska alternativa
Hi-kvadrat test kvaliteta podudaranja	Ne postoji
Hi-kvadrat test nezavisnosti	Ne postoji
Kapa test (mera slaganja)	Ne postoji
Man-Vitnjev U test	T-test nezavisnih uzoraka (17. poglavlje)
Vilkoksonov test ranga	T-test uparenih uzoraka (17. poglavlje)
Kruskal-Volisov test	Jednofaktorska ANOVA različitih grupa (18. poglavlje)
Fridmanov test	Jednofaktorska ANOVA ponovljenih merenja (18. poglavlje)

Tehnike u ovom poglavlju namenjene su prvenstveno za poređenje grupa. Neparametarska alternativa korelacije (Spirmanova korelacija ranga, koeficijent ro) predstavljena je u poglavlju 11.

Pretpostavke neparametarskih tehnika

Iako su opšte pretpostavke neparametarskih tehnika blaže i zato češće zadovoljene, uvek bi trebalo proveriti da li je tako u svakom konkretnom slučaju.

- *Slučajnost uzoraka.*
- *Nezavisnost opservacija.* Svaka osoba ili opservacija može se brojati samo jedanput, ne smeju se pojavljivati u više kategorija ili grupa, i podaci jednog subjekta ne smeju uticati na podatke drugih. U ovom pogledu izuzetak su tehnike ponovljenih merenja (Vilkoksonov test ranga, Fridmanov test), kada se isti subjekti ponovno testiraju u više navrata ili u različitim okolnostima.

Neke tehnike razmotrene u ovom poglavlju imaju i dodatne pretpostavke koje treba proveriti. Te specifične pretpostavke su objašnjene u odgovarajućim odeljcima.

U ovom poglavlju prikazaćemo razne neparametarske tehnike na primerima iz nekoliko datoteka s podacima, koje se mogu preuzeti s prateće Web lokacije ove knjige. Sve pojedinosti o tim datotekama date su u dodatku. Ukoliko želite da pratite primere i odmah sprovedite navedena uputstva, pokrenite SPSS i otvorite odgovarajuću datoteku s podacima. Neparametarske tehnike opisaćemo samo ukratko. Za dalje usavršavanje, pročitajte knjige koju su napisali Daniel (1990), Gravetter i Wallnau (2004), Siegel i Castellan (1988) i Peat (2001).

Hi-kvadrat

Postoje dve vrste testa hi-kvadrat, obe za kategorijske podatke:

1. Metoda hi-kvadrat za ispitivanje kvaliteta podudaranja (engl. *goodness of fit*); istražuje proporciju slučajeva koji spadaju u razne kategorije *jedne* promenljive i poredi ih sa hipotetičkim (teorijskim, pretpostavljenim) vrednostima tih proporcija.
2. Metoda hi-kvadrat za testiranje nezavisnosti; određuje da li su povezane *dve* kategorijske promenljive. Ona poredi učestalosti (frekvencije) slučajeva u raznim kategorijama *jedne* promenljive sa raznim kategorijama druge promenljive. Na primer: da li je proporcija pušača i nepušača jednaka za muškarce i žene? Ili, drugačije rečeno: da li su muškarci češće pušači od žena?

Opširniji prikaz testa hi-kvadrat dat je u 17. poglavlju knjige Gravettera i Wallnaua (2004).

Hi-kvadrat test za ispitivanje kvaliteta podudaranja

Ovaj test, koji se naziva i hi-kvadrat test jednog uzorka, često se upotrebljava za poređenje proporcije slučajeva iz određenog uzorka s hipotetičkim vrednostima ili onima prethodno dobijenim u nekoj poredbenoj populaciji. Dovoljno je da datoteka s podacima sadrži samo jednu kategorijsku promenljivu i određenu proporciju s kojom želite da uporedite dobijene rezultate. Može se ispitati hipoteza da nema razlike u proporciji kategorija (50%/50%) ili određena proporcija dobijena u nekoj prethodnoj studiji.

Primer istraživačkog pitanja: Istražiću da li je procentualni udeo pušača u populaciji opisanoj u datoteci survey3ED.sav ekvivalentan onome (20%) navedenom u literaturi prethodno objavljene velike australijske studije.

Šta vam treba:

- Jedna kategorijska promenljiva sa dve ili više kategorija: pušač (Da/Ne).
- Jedna hipotetička proporcija (20% pušača; 80% nepušača ili 0,2/0,8).

Postupak za obavljanje hi-kvadrat testa kvaliteta podudaranja

Da biste pratili ovaj primer, otvorite datoteku **survey3ED.sav**.

1. Otvorite meni **Analyze**, pritisnite njegovu stavku **Non-parametric Tests** i zatim **Chi-Square**.
2. Pritisnite kategorijsku promenljivu **smoke** (tj. pušač) i prebacite je u polje **Test Variable List**.

3. U odeljku **Expected Values** pritisnite opciju **Values**. U polje **Values** treba da upišete dva broja.

- Prvi broj (0,2) odgovara očekivanoj proporciji prve šifrovane vrednosti te promenljive (1=da, pušač). Pritisnite dugme **Add**.
- Upišite drugi broj (0,8), što je očekivana proporcija druge šifrovane vrednosti te promenljive (2=ne, nepušač). Pritisnite dugme **Add**.
- Kada promenljiva ima više od dve moguće vrednosti, za svaku od njih treba upisati odgovarajuću proporciju.

4. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Tim postupkom se generiše sledeća komanda:

```
NPART TEST
/CHISQUARE=smoke
/EXPECTED=.2, .8
/MISSING ANALYSIS.
```

Evo kako izgledaju rezultati tog postupka.

smoker

	Observed N	Expected N	Residual
1 YES	85	87.2	-2.2
2 NO	351	348.8	2.2
Total	436		

Test Statistics

	smoker
Chi-Square ^a	.069
df	1
Asymp. Sig.	.792

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 87.2.

Tumačenje rezultata hi-kvadrat testa podudaranja

U prvoj tabeli date su opažene učestalosti (frekvencije) iz tekuće datoteke s podacima; vidimo da su pušači njih 85 od ukupno 436 (ili 19,5%). U koloni Expected N date su učestalosti očekivane na osnovu prethodno zadate proporcije (20%). U ovom slučaju, očekivano je da bude 87 pušača, a opaženo 85.

U tabeli Test Statistics dati su rezultati SPSS-ovog testa Chi-Square, koji poredi očekivane i opažene vrednosti. U ovom slučaju, razlika između njih je vrlo mala, a nije ni statistički značajna (Sig.=0,79).

Predstavljanje rezultata

Treba navesti vrednost hi-kvadrat, broj stepeni slobode (engl. *degrees of freedom*, *df*) i iznos *p*.

Hi-kvadrat test podudaranja pokazuje da se proporcija pušača u tekućem uzorku (19,5%), ne razlikuje mnogo od vrednosti 20% dobijenoj u prethodnoj australijskoj studiji, $\chi^2 (1, n = 436) = 0,07, p < 0,79$.

Hi-kvadrat test nezavisnosti

Ovim testom se istražuje veza između dve kategorijske promenljive. Svaka od njih može imati dve ili više kategorija. Test poredi učestalosti ili proporcije slučajeva opažene u svakoj od kategorija, s vrednostima koje bi bile očekivane da između dve merene promenljive nema nikakve veze. Zasniva se na unakrsnoj tabeli, tj. tabeli u kojoj su kategorije jedne promenljive ukrštene s kategorijama druge (npr. muškarci/žene; pušač/nepušač); svaka ćelija unakrsne tabele sadrži po jednu kombinaciju kategorija posmatranih promenljivih.

Kada SPSS naide na tabelu 2 sa 2 (po dve kategorije u svakoj promenljivoj), rezultat hi-kvadrat testa obuhvata i jednu dodatnu „korekciju neprekidnosti prema Jeitsu“ (*Yates' Correction for Continuity*). Ona bi trebalo da kompenzuje (po mišljenju nekih autora) precenjenu vrednost hi-kvadrat koja se dobija u tabeli 2 sa 2.

U sledećem postupku, na primeru datoteke *survey3ED.sav* pokazaću kako se upotrebljava hi-kvadrat test za projekat 2 sa 2. Kada studija obuhvata promenljive s više od dve kategorije (npr. 2 sa 3, 4 sa 4), videćete da se izlaz iz SPSS-a neznatno razlikuje.

Kratak pregled hi-kvadrat testa nezavisnosti

Primer istraživačkog pitanja: Pitanje se može formulisati na više načina: postoji li veza između pola i pušačkog ponašanja? Da li su muškarci češće pušači od žena? Da li je proporcija muškaraca pušača jednaka toj proporciji kod žena?

Šta vam treba: Dve kategorijske promenljive sa po dve ili više kategorija:

- Pol (Muškarac/Žena)
- Pušač (Da/Ne)

Dodatne pretpostavke: Najmanja očekivana učestalost u svim ćelijama trebalo bi da bude 5 ili više. Neki autori predlažu blaže kriterijume: najmanje 80 procenata ćelija trebalo bi da imaju očekivane učestalosti 5 ili više. Za tabele 1 sa 2 ili 2 sa 2, preporučuje se da očekivana učestalost bude najmanje 10. Kada tabela 2 sa 2 ne zadovoljava ovu pretpostavku, umesto vrednosti hi-kvadrat treba upotrebiti Fišerov „tačan pokazatelj verovatnoće“ (engl. *Fisher's Exact Probability Test*), koji se navodi u sklopu rezultata hi-kvadrat testa.

Postupak za obavljanje hi-kvadrat testa nezavisnosti

Da biste pratili ovaj primer, otvorite datoteku **survey3ED.sav**.

1. Otvorite meni **Analyze**, pritisnite stavku **Descriptive Statistics**, zatim **Crosstabs**.
2. Pritisnite onu promenljivu (npr. sex/pol) čije će kategorije zauzimati redove tabele; pritisnite strelicu da biste izabranu promenljivu prebacili u polje **Row(s)**.
3. Pritisnite onu promenljivu (npr. smoke/pušač) čije će kategorije zauzimati kolone tabele, pa pritisnite strelicu da biste izabranu promenljivu prebacili u polje **Column(s)**.
4. Pritisnite dugme **Statistics**. Potvrdite polja **Chi-square** i **Phi and Cramer's V**. Pritisnite **Continue**.
5. Pritisnite dugme **Cells**.
 - U polju **Counts** pritisnite **Observed**.
 - U odeljku **Percentage**, potvrdite polja **Row**, **Column** i **Total**.
6. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

CROSSTABS

```
/TABLES=sex BY smoke
/FORMAT=AVALUE TABLES
/STATISTIC=CHISQ PHI
/CELLS= COUNT ROW COLUMN TOTAL
/COUNT ROUND CELL .
```

Sledi deo rezultata opisanog postupka za tabelu 2 sa 2. Izlaz se malo razlikuje za promenljive sa više od dve kategorije, ali u njemu i dalje treba tražiti iste ključne podatke.

			SMOKE		Total
			YES	NO	
SEX	MALES	Count	33	151	184
		Expected Count	35,9	148,1	184,0
		% within SEX	17,9%	82,1%	100,0%
		% within SMOKE	38,8%	43,0%	42,2%
		% of Total	7,6%	34,6%	42,2%
	FEMALES	Count	52	200	252
		Expected Count	49,1	202,9	252,0
		% within SEX	20,6%	79,4%	100,0%
		% within SMOKE	61,2%	57,0%	57,8%
		% of Total	11,9%	45,9%	57,8%
Total		Count	85	351	436
		Expected Count	85,0	351,0	436,0
		% within SEX	19,5%	80,5%	100,0%
		% within SMOKE	100,0%	100,0%	100,0%
		% of Total	19,5%	80,5%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.494 ^b	1	.482		
Continuity Correction ^a	.337	1	.562		
Likelihood Ratio	.497	1	.481		
Fisher's Exact Test				.541	.282
Linear-by-Linear Association	.493	1	.483		
N of Valid Cases	436				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 35,87.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	-.034	.482
	Cramer's V	.034	.482
N of Valid Cases		436	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Tumačenje rezultata hi-kvadrat testa nezavisnosti

Pretpostavke

Najpre treba proveriti da li je prekršena jedna od pretpostavki testa hi-kvadrat u pogledu najmanje očekivane ćelijske učestalosti, koja bi trebalo da bude 5 ili više (ili da najmanje 80 procenata ćelija ima očekivane učestalosti 5 ili više). Ta informacija je data u fusnoti ispod tabele **Chi-Square Tests**. U fusnoti b u primeru kaže se '0 cells (,0%) have expected count less than 5. The minimum expected count is 35,87'. To znači da nismo prekršili pretpostavku, pošto su očekivane učestalosti u svim ćelijama veće od 5 (u našem primeru, veće od 35,87).

Hi-kvadrat testovi

U rezultatima nas najviše zanima vrednost **Pearson Chi-Square**, data u tabeli **Chi-Square Tests**. Međutim, kada se ima tabela 2 sa 2 (tj. kada svaka promenljiva ima samo dve kategorije), treba upotrebiti vrednost u drugom redu tabele (**Continuity Correction**). To je tzv. Jejtsova korekcija (Yates' Correction for Continuity); ona kompenzuje precenjenu vrednost hi-kvadrat koja je posledica malog broja dimenzija tabele 2 sa 2. U gornjem primeru, korišćena vrednost iznosi 0,337, uz značajnost od 0,56 (datu u koloni **Asymp. Sig. (2-sided)**). Da bi rezultat bio značajan, veličina **Sig.** treba da je 0,05 ili manja. U ovom primeru, vrednost 0,56 je (11 puta) *veća* od alfa vrednosti 0,05, pa možemo zaključiti da naš rezultat *nije* značajan. To znači da se proporcija muškaraca koji puše ne razlikuje značajno od proporcije žena koje puše. Nema nikakve veze između pušačkog statusa i pola.

Unakrsno tabeliranje

Procenat pušača svakog pola može se očitati u zbirnim informacijama datim u tabeli **SEX*SMOKE Crosstabulation**. S toliko informacija naguranih u svaku ćeliju, ova tabela vas može zbuniti. Da biste saznali koji procenat muškaraca su pušači, pročitajte prvi red tabele koji se odnosi na muškarce. U ovom slučaju, gledamo šta piše u nastavku **% within sex**. U ovom primeru, 17,9 procenata muškaraca su pušači i 82,1 procenata nepušači; 20,6 procenata žena je pušača i 79,4 procenata nepušača.

Da nas zanima procenat celog uzorka koji spada u pušače, prešli bismo naniže u poslednji red, koji sabira rezultate za oba pola. U ovom slučaju, pogledali bismo šta piše u redu **% of Total**. Prema tim rezultatima, 19,5 procenata uzorka puši, a 80,5 procenata ne puši.

Veličina uticaja

U proceduri **Crosstabs** izračunava se više pokazatelja veličine uticaja (tj. jačine veze između promenljivih). Za tabele 2 sa 2, najčešće se koristi koeficijent **fi** (**phi coefficient**), što je koeficijent korelacije u opsegu od 0 do 1, pri čemu veći broj pokazuje jaču vezu između dve promenljive. U ovom primeru,

koeficijent **fi** je -0,034, što se smatra vrlo malim uticajem po Koenovom (1988) kriterijumu od 0,10 za mali, 0,30 za srednji i 0,50 za veliki uticaj.

Za tabele veće od 2 sa 2, u izveštaj treba staviti **Kramerov** pokazatelj **V** (**Cramer's V**), koji uzima u obzir broj stepeni slobode. Za ocenu veličine uticaja kod većih tabela koriste se malo drugačiji kriterijumi. Da biste utvrdili koji kriterijum da upotrebite, najpre oduzmite 1 od broja kategorija u rednoj promenljivoj (**R-1**), a zatim oduzmite 1 od broja kategorija u kolonskoj promenljivoj (**K-1**). Od ta dva broja, zadržite onaj manji.

Za **R-1** ili **K-1** jednako 1 (dve kategorije): mali=0,01, srednji=0,30, veliki=0,50

Za bilo **R-1** bilo **K-1** jednako 2 (tri kategorije): mali=0,07, srednji=0,21, veliki=0,35

Za bilo **R-1** bilo **K-1** jednako 3 (četiri kategorije): mali=0,06, srednji=0,17, veliki=0,29

Više o pokazateljima koeficijent **fi** i **Kramerov V** pročitajte u knjizi Gravettera i Wallnaua (2004, str. 605).

Rezultate ove analize mogli biste predstaviti ovako:

Hi-kvadrat test nezavisnosti (uz korekciju neprekidnosti prema Jejtisu) nije pokazao značajnu vezu između pola i pušačkog statusa, $c^2(1, n = 436) = 0,34$, $p = 0,56$, $fi = -0,03$.

Mera slaganja kapa

Još jedan od statističkih pokazatelja za kategorijske promenljive u SPSS-ovoj proceduri **Crosstabs** jeste mera slaganja kapa (**Kappa**). Ona je uobičajena u medicinskoj literaturi za ocenu slaganja dva lekara (npr. dijagnoze dva kliničara) ili saglasnosti dva različita dijagnostička testa (novorazvijen test u odnosu na zlatni standard).

Kratak pregled kapa testa

Primer istraživačkog pitanja: Koliko su saglasne dijagnostičke klasifikacije mernih skala **Edinburgh Postnatal Depression Scale (EPDS)** i **Depression, Anxiety and Stress Scale (DASS)**?

Šta vam treba: Dve kategorijske promenljive s jednakim brojem kategorija (npr. dijagnostička klasifikacija 1. lekara ili 1. testa: 0=nije deprimirana, 1=deprimirana; i dijagnostička klasifikacija iste osobe koju je dao 2. lekar ili 2. test).

Pretpostavke: Jednak broj kategorija za 1. i za 2. lekara.

Parametarske alternative: Ne postoji.

U narednom primeru istražićemo stepen slaganja između dve mere depresije u uzorku žena posle porođaja. U datoteci **depress3ED.sav**, rezultati za svaku ženu na skalama **Edinburgh Postnatal Depression Scale (EPDS: Cox, Holden**

& Sagovsky, 1987) i Depression, Anxiety and Stress Scale (DASS-Dep: Lovibond & Lovibond, 1995) klasifikovani su u skladu s preporučenim graničnim vrednostima svake skale. Dobijene su dve promenljive s vrednostima 0 (nije deprimirana) i 1 (deprimirana). Cilj istraživanja je utvrditi da li žene prepoznate kao deprimirane na skali EPDS, isto tako klasifikuje i skala DASS Depression (DASS-Dep).

Postupak za dobijanje koeficijenta kapa

Da biste pratili ovaj primer, otvorite datoteku s podacima **depress3ED.sav**.

1. Otvorite meni **Analyze**, izaberite stavku **Descriptive Statistics**, pa pritisnite **Crosstabs**.
2. Pritisnite onu promenljivu (npr. DASSdep2) koja će biti redna u unakrsnoj tabeli i strelicu da je prebacite u polje **Row(s)**.
3. Pritisnite onu promenljivu (npr. EPDSgp2) koja će biti kolonska u unakrsnoj tabeli i strelicu da je prebacite u polje **Column(s)**.
4. Pritisnite dugme **Statistics**. Pritisnite **Kappa**. Pritisnite dugme **Continue**.
5. Pritisnite dugme **Cells**.
6. U polju **Counts** pritisnite dugme **Observed**.
7. U odeljku **Percentage** pritisnite dugme **Row**. Pritisnite dugme **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Taj postupak generiše komandu:

```
CROSSTABS
  /TABLES=DASSdep2 BY EPDSgp2
  /FORMAT=AVALUE TABLES
  /STATISTIC=KAPPA
  /CELLS= COUNT ROW COLUMN TOTAL
  /COUNT ROUND CELL .
```

Ovako izgledaju rezultati:

DASSdep2 * EPDSgp2 Crosstabulation

			EPDSgp2		Total
			0 not depressed	1 likely depressed	
DASSdep2 0 not depressed	Count		225	34	259
	% within DASSdep2		86,9%	13,1%	100,0%
	% within EPDSgp2		94,1%	42,5%	81,2%
	% of Total		70,5%	10,7%	81,2%
1 mild to severe depression	Count		14	46	60
	% within DASSdep2		23,3%	76,7%	100,0%
	% within EPDSgp2		5,9%	57,5%	18,8%
	% of Total		4,4%	14,4%	18,8%
Total	Count		239	80	319
	% within DASSdep2		74,9%	25,1%	100,0%
	% within EPDSgp2		100,0%	100,0%	100,0%
	% of Total		74,9%	25,1%	100,0%

Symmetric Measures

		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Measure of Agreement	Kappa	.563	.055	10.231	.000
N of Valid Cases		319			

a. Not assuming the null hypothesis

b. Using the asymptotic standard error assuming the null hypothesis.

Tumačenje koeficijenta kapa

Glavni rezultat za nas je u tabeli **Symmetric Measures**, gde očitavamo da mera slaganja (engl. *Measure of Agreement*) Kappa iznosi 0,56, uz značajnost $p < 0,0005$. Kako navodi Peat (2001, str. 228), vrednost 0,5 pokazatelja kapa predstavlja umereno slaganje, iznad 0,7 predstavlja dobro slaganje, a iznad 0,8 vrlo dobro slaganje. Dakle, u ovom primeru je stepen slaganja između klasifikacije slučajeva kao deprimiranih pomoću skale EPDS i skale DASS-Dep umeren.

Osetljivost i određenost

Pomoću učestalosti (frekvencija) i procenata u tabeli **Crosstabulation** mogu se izračunati i **osetljivost** i **određenost** mere ili testa. U medicinskoj literaturi je uobičajeno da se pomoću njih ocenjuje tačnost dijagnostičkog testa u otkrivanju postojanja ili nepostojanja bolesti, odnosno tačnost novog testa u odnosu na neki postojeći zlatni standard. **Osetljivost** odražava proporciju slučajeva postojanja bolesti (stanja) koji su tačno dijagnostifikovani, dok **određenost** predstavlja proporciju slučajeva bez date bolesti (stanja) koji su tačno dijagnostifikovani. U ovom primeru, ispitujemo saglasnost klasifikacije na skali DASS-Dep u odnosu na najčešće upotrebljavani test za klasifikaciju, skalu EPDS.

Osetljivost skale DASS-Dep može se izračunati ako se pročitaju vrednosti iz druge kolone tabele. Od 80 slučajeva koje je skala EPDS (naš zlatni standard) prepoznala kao deprimirane, 46 je tako klasifikovala i skala DASS-Dep. To daje osetljivost od 57,5 procenata (46/80).

Određenost se izračunava tako što se pročitaju vrednosti iz prve kolone, gde su osobe koje su pomoću skale EPDS bile klasifikovane kao *nedeprimirane*. Skala DASS-Dep je 'tačno' klasifikovala 225 od ukupno 239 takvih, što predstavlja određenost od 94,1 procenata. Obe te skale su veoma usklađene u pogledu slučajeva klasifikovanih kao nedeprimirani; međutim, postoji određena nesaglasnost između slučajeva koje skale prepoznaju kao deprimirane. Više o pojedinostima ovog istraživanja videti u dodatku, a članak objavljen o toj temi potražite na adresi: <http://www.biomedcentral.com/1471-244X/6/12>.

Man-Vitnijev U test

Ova tehnika se upotrebljava za ispitivanje razlika između dve nezavisne grupe na neprekidnoj skali. Primera radi, da li se muškarci i žene razlikuju po nivou samopoštovanja? Ovaj test je neparametarska alternativa t-testu nezavisnih uzoraka. Umesto da poredi srednje vrednosti dveju grupa, kao sto radi t-test, Man-Vitnijev U test poredi njihove medijane. Dobijene vrednosti neprekidne promenljive pretvara u rangove za obe grupe i potom izračunava da li se rangovi tih grupa značajno razlikuju. Pošto se rezultati pretvaraju u rangove, stvarna raspodela rezultata nije važna.

Kratak pregled Man-Vitnijevog U testa

Primer istraživačkog pitanja: Da li se muškarci i žene razlikuju po nivou samopoštovanja? Osećaju li muškarci veće samopoštovanje od žena?

Sta vam treba: Dve promenljive:

- jedna kategorijska promenljiva sa dve grupe (npr. pol/sex);
- jedna neprekidna promenljiva (npr. ukupno samopoštovanje/total self-esteem).

Pretpostavke: Videti opšte pretpostavke za neparametarske tehnike date na početku poglavlja.

Parametarska alternativa: t-test nezavisnih uzoraka.

Postupak za obavljanje Man-Vitnijevog U testa

Da biste pratili ovaj primer, otvorite datoteku s podacima **survey3ED.sav**.

1. U glavnom meniju na vrhu prozora otvorite meni **Analyze**, izaberite u njemu stavku **Nonparametric Tests**, pa pritisnite **2 Independent Samples**.
2. Pritisnite odabranu neprekidnu (zavisnu) promenljivu (npr. ukupno samopoštovanje tj. tsfest) i prebacite je u polje **Test Variable List**.
3. Pritisnite odabranu kategorijsku (nezavisnu) promenljivu (npr. pol tj. sex) i prebacite je u polje **Grouping Variable**.
4. Pritisnite dugme **Define Groups**. Upišite šifre za Grupu 1 (npr. 1) i za Grupu 2 (npr. 2). Njima ste šifrovali vrednosti promenljive pomoću koje se sada definišu grupe – pogledajte u svoj šifarnik. Ako ne možete da se setite šifara za svaku grupu, desnim tasterom miša pritisnite polje **Grouping Variable** i izaberite **Variable Information**. Prikazaće se okvir sa opisima i brojevima (šiframa) koje treba upisati za tu promenljivu. Pritisnite dugme **Continue**.
5. U odeljku **Test Type** treba da je potvrđeno polje **Mann-Whitney U**.
6. Pritisnite dugme **Options** i izaberite **Descriptives**.
7. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

```
NPART TESTS  
/M-W= tsfest BY sex(1 2)  
/MISSING ANALYSIS.
```

Prikazan je odabran deo rezultata ovog postupka.

Test Statistics^a

	Total self esteem
Mann-Whitney U	21594.000
Wilcoxon W	53472.000
Z	-1.227
Asymp. Sig. (2-tailed)	.220

a. Grouping Variable: SEX

Ranks

	sex	N	Mean Rank	Sum of Ranks
Total Self esteem	1 MALES	184	227.14	41794.00
	2 FEMALES	252	212.19	53472.00
	Total	436		

Tumačenje rezultata Man-Vitnijevog U testa

Glavne vrednosti koje treba gledati u izlazu jesu **Z** i nivo značajnosti, dat u redu **Asymp. Sig (2-tailed)**. Kada uzorak ima više od 30 slučajeva, SPSS izračunava i iznos z-aproksimacije, koja obuhvata i korekciju zbog međusobnih veza između podataka. U gornjem primeru, vrednost **Z** je -1,23 (zaokruženo), uz nivo značajnosti od $p=0,22$. Iznos verovatnoće (p) nije manji od niti jednak 0,05, pa taj rezultat (**Z**) nije značajan. Nema statistički značajne razlike između nivoa samopoštovanja muškaraca i žena.

Kada otkrijete statistički značajnu razliku između grupa, trebalo bi da opišete i smer te razlike, tj. u kojoj grupi je neprekidna promenljiva prosečno veća. U tabeli **Ranks** to se vidi u koloni **Mean Rank**. Međutim, bilo bi bolje da u izveštaju navedete medijanu svake grupe. Nažalost, njih ova procedura ne izračunava, pa je potreban još jedan korak.

Postupak izračunavanja medijana svake grupe

1. Otvorite meni **Analyze**, pa pritisnite **Compare means** i izaberite **Means**.
2. Pritisnite odabranu neprekidnu promenljivu (npr. ukupno samopoštovanje tj. tsfest) i prebacite je u polje **Dependent List**.
3. Pritisnite odabranu kategorijsku promenljivu (npr. pol tj. sex) i prebacite je u polje **Independent List**.
4. Pritisnite dugme **Options**. Pritisnite dugme **Median** u odeljku **Statistics** i pređite u polje **Cell Statistics**. Pritisnite dugmad **Mean** i **Standard Deviation**, i uklonite kvačicu iz polja **Cell Statistics**.
5. Pritisnite dugme **Continue**.
6. Pritisnite **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

```
MEANS
  TABLES=tslfest BY sex
  /CELLS COUNT MEDIAN .
```

Evo i tabele rezultata sa izračunatim medijanima:

Report

Total Self esteem

sex	N	Median
1 MALES	184	35,00
2 FEMALES	252	34,50
Total	436	35,00

Veličina uticaja

SPSS ne izračunava statističke pokazatelje veličine uticaja, ali se pomoću vrednosti z navedene u rezultatima može izračunati približna vrednost r .

$r = z / \text{kvadratni koren od } N$ gde je N = ukupan broj slučajeva (opervacija).

U ovom primeru, $z = -1,23$ i $N = 436$; stoga r iznosi 0,06. To bi se smatralo vrlo malim uticajem prema Koenovom (1988) kriterijumu (0,1=mali uticaj, 0,3=srednji uticaj, 0,5=veliki uticaj).

Rezultate ove analize mogli biste predstaviti ovako:

Man-Vitnjev U test nije otkrio značajnu razliku u nivou samopoštovanja muškaraca ($Md = 35$, $n = 184$) i žena ($Md = 34,5$, $n = 252$), $U = 21594$, $z = -1,23$, $p = 0,22$, $r = 0,06$.

Vilkoksonov test ranga

Vilkoksonov test ranga – još se naziva Vilkoksonov test ekvivalentnih parova ili test rangova sa znakom – namenjen je za ponovljena merenja, tj. kada se subjekti mere u dva navrata ili pod dva različita uslova. To je neparametarska alternativa t-testu ponovljenih merenja, ali umesto poređenja srednjih vrednosti, Vilkoksonov test pretvara rezultate u rangove i njih poredi u trenutku 1 i trenutku 2. Vilkoksonov test se može upotrebiti i kada su subjekti upareni po specifičnim kriterijumima.

Kao ilustraciju ove tehnike upotrebila sam podatke iz datoteke **experim3ED.sav** s prateće Web lokacije knjige (videti str. xi, a pojedinosti studije u dodatku). U ovom primeru, upoređiću rezultate na testu kojim se ispituje strah od statistike (engl. *Fear of Statistics Test*), popunjavanom pre i posle intervencije koja je trebalo da pomogne studentima da ovladaju statistikom.

Kratak pregled Vilkoksonovog testa ranga

Primer istraživačkog pitanja: Ima li promene rezultata ispitivanja straha od statistike od vremena 1 do vremena 2?

Šta vam treba: Jedna grupa subjekata merena na istoj neprekidnoj skali ili po istom kriterijumu u dva navrata. Promenljive su rezultati u trenutku 1 ili u okolnostima 1, i rezultati u trenutku 2 ili u okolnostima 2.

Pretpostavke: Videti opšte pretpostavke za neparametarske tehnike, navedene na početku poglavlja.

Parametarska alternativa: t-test uparenih uzoraka.

Postupak obavljanja Vilkoksonovog testa ranga

Da biste pratili ovaj primer, otvorite datoteku **experim3ED.sav**.

1. Na vrhu prozora otvorite meni **Analyze**, izaberite u njemu stavku **Nonparametric Tests**, pa pritisnite **2 Related Samples**.
2. Pritisnite promenljive koje predstavljaju rezultate u trenutku 1 i trenutku 2 (npr. fost1, fost2). Prebacite ih u kućicu **Test Pairs List**.
3. U odeljku **Test Type** treba da je potvrđeno polje **Wilcoxon**. Pritisnite **OK**.
4. Pritisnite dugme **Options**. Izaberite **Quartiles**; time ste zadali izračunavanje medijana za svaki od trenutaka u kojima su obavljana ispitivanja.
5. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

```
NPAR TEST
  /WILCOXON=fost1 WITH fost2 (PAIRED)
  /STATISTICS QUANTILES
  /MISSING ANALYSIS.
```

Prikazaćemo samo deo rezultata:

Descriptive Statistics

	N	Percentiles		
		25th	50th (Median)	75th
fear of stats time1	30	37.00	40.00	44.00
fear of stats time2	30	34.50	38.00	40.00

Test Statistics^a

	fear of stats time2 - fear of stats time1
Z	-4,180 ^a
Asymp. Sig. (2-tailed)	,000

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Tumačenje rezultata Vilkoksonovog testa ranga

Glavne vrednosti koje treba gledati u izlaznom rezultatu jesu **Z** i njen nivo značajnosti dat u redu **Asymp. Sig (2-tailed)**. Kada je nivo značajnosti jednak ili manji od 0,05 (npr. 0,04, 0,01, 0,001), može se zaključiti da je razlika između rezultata značajna. U ovom primeru, verovatnoća **Sig.** iznosi 0,000 (što zapravo znači manje od 0,0005). Stoga treba zaključiti da je razlika između te dve grupe rezultata značajna (neslučajna).

Veličina uticaja

Veličina uticaja u ovom testu može se izračunati na način opisan za Man-Vitnjev U test, tj. deljenjem veličine z kvadratnim korenom iz N . Međutim, u ovoj situaciji N = broj opservacija u dva vremenska trenutka, a ne broj slučajeva. U ovom primeru, $Z = 4,18$, $N = 60$ (slučajeva $\times 2$); zato je $r = 0,54$, što ukazuje na veliki uticaj prema Koenovom (1988) kriterijumu: 0,1 = mali uticaj, 0,3 = srednji uticaj, 0,5 = veliki uticaj.

Rezultati ove analize mogu se predstaviti ovako:

Vilkoksonov test ranga otkrio je statistički značajno smanjenje straha od statistike nakon učestvovanja u programu obuke, $z = -4,18$, $p < 0,001$, uz veliku razliku ($r = 0,54$). Medijana rezultata na skali straha od statistike (Fear of Statistics Scale) opala je od ($Md = 40$) pre programa do ($Md = 38$) posle programa.

Kruskal-Volisov test

Kruskal-Volisov test (naziva se i Kruskal-Volisaov H test) neparametarska je alternativa jednofaktorskoj analizi varijanse različitih grupa. Služi za poređenje rezultata neke neprekidne promenljive za *tri ili više grupa*. Po prirodi je jednak Man-Vitnjevom U testu prikazanom u prethodnom delu poglavlja, ali omogućava poređenje više od dve grupe. Rezultati se pretvaraju u rangove, pa se porede srednji rangovi svake grupe. To je analiza *različitih grupa*, pa u svakoj grupi moraju biti drugi ljudi.

Kratak pregled Kruskal-Volisovog testa

Primer istraživačkog pitanja: Ima li razlike u nivoima optimizma osoba koje pripadaju trima starosnim grupama?

Šta vam treba: Dve promenljive:

- jedna kategorijska nezavisna promenljiva s tri ili više kategorija (npr. agegp3: 18–29, 30–44, 45+); i
- jedna neprekidna zavisna promenljiva (npr. ukupan optimizam).

Pretpostavke: Videti opšte pretpostavke za neparametarske tehnike, navedene na početku poglavlja.

Parametarska alternativa: Jednofaktorska analiza varijanse različitih grupa.

Postupak obavljanja Kruskal-Volisovog testa

Da biste pratili ovaj primer, otvorite datoteku **survey3ED.sav**.

1. Na vrhu prozora otvorite meni **Analyze**, izaberite u njemu stavku **Nonparametric Tests**, pa pritisnite **K Independent Samples**.
2. Pritisnite odabranu neprekidnu zavisnu promenljivu (npr. ukupan optimizam) i prebacite je u polje **Test Variable List**.
3. Pritisnite odabranu kategorijsku nezavisnu promenljivu (npr. agegp3) i prebacite je u polje **Grouping Variable**.
4. Pritisnite dugme **Define Range**. U polje **Minimum** upišite početnu vrednost kategorijske promenljive (npr. 1). U polje **Maximum** upišite najveću vrednost kategorijske promenljive (npr. 3). Pritisnite dugme **Continue**.
5. U odeljku **Test Type** treba da je potvrđeno polje **Kruskal-Wallis H**.
6. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

NPAR TESTS

```
/K-W=toptim BY agegp3(1 3)
/MISSING ANALYSIS.
```

Prikazaćemo samo deo rezultata ovog testa.

Ranks			
	AGEGP	N	Mean Rank
Total	18-29	147	198,18
Optimism	30-44	153	216,06
	45+	135	241,80
Total		435	

Test Statistics ^{a,b}	
	Total Optimism
Chi-Square	8,573
df	2
Asymp. Sig.	.014

a. Kruskal Wallis Test

b. Grouping Variable : AGEGP3

Trebalo bi da izračunate i medijane vrednosti optimizma za svaku starosnu grupu. Sledite postupak za dobijanje medijana opisan u odeljku posvećenom Man-Vitnijevom U testu.

```
MEANS
TABLES=toptim BY agegp3
/CELLS COUNT MEDIAN .
```

Ovo je tabela rezultata sa izračunatim medijanama:

Report		
Total Optimism		
age in 3 groups	N	Median
1 <= 29	147	22,00
2 30 - 44	153	22,00
3 45+	135	23,00
Total	435	22,00

Tumačenje rezultata Kruskal-Volisovog testa

Glavne vrednosti koje treba gledati u izlaznom rezultatu jesu: vrednost pokazatelja hi-kvadrat (Chi-Square), broj stepeni slobode (df) i nivo značajnosti (prikazan kao Asymp. Sig.). Kada je nivo značajnosti manji od 0,05 (npr. 0,04, 0,01, 0,001), može se zaključiti da je razlika u dobijenim vrednostima neprekidne promenljive između te tri grupe značajna. Zatim treba pogledati srednju vrednost ranga (Mean Rank) za te tri grupe prikazan u prvoj tabeli rezultata. Tu vidite koja od tri grupe ima najveći rang, što odgovara najvećoj vrednosti neprekidne promenljive.

U navedenim rezultatima, nivo značajnosti je 0,01 (zaokruženo). To je manje od alfa nivoa od 0,05, pa rezultat govori da postoji razlika u nivoima optimizma različitih starosnih grupa. Pregledom srednjih (prosečnih) vrednosti rangova grupa vidimo da je optimizam na najvišem nivou u najstarijoj grupi (45 i više godina), a najmanji u najmlađoj.

Rezultati ove analize mogu se predstaviti ovako:

Kruskal-Volisov test je otkrio statistički značajnu razliku nivoa optimizma tri različite starosne grupe (Gp1, n = 147: 18–29 god, Gp2, n = 153: 30–44 god, Gp3, n = 135: 45 i više god), $\chi^2(2, n = 435) = 8,57, p = 0,014$. Najstarija starosna grupa (45 i više godina) ima veću medijanu rezultata (Md = 23) od ostale dve starosne grupe, čija medijana iznosi 22.

Naknadni testovi i veličina uticaja

Kada dobijete statistički značajan rezultat Kruskal-Volisovog testa, još uvek ne znate koje grupe se statistički značajno međusobno razlikuju. Da biste to utvrdili, treba da obavite još nekoliko naknadnih Man-Vitnijevih U testova između parova grupa (npr. grupe najstarijih i grupe najmlađih). Ako nameravate da poredite sve parove grupa (1 sa 2, 1 sa 3 i 2 sa 3), morate primeniti Bonferonijevu korekciju alfa vrednosti (videti uvod u peti deo knjige), kako biste izbegli greške prve vrste.

Bonferonijevo prilagođenje znači podeliti alfa vrednost 0,05 brojem testova koje nameravate da obavite i potom upotrebljavati tako revidiran alfa nivo kao kriterijum za određivanje značajnosti. Ovde bi to značilo stroži alfa nivo od $0,05/3 = 0,017$. Međutim, umesto poređenja svih mogućih parova grupa, bolje je izabrati samo nekoliko ključnih grupa i uporediti njih, čime se alfa nivo zadržava na upotrebljivoj visini (ne premaloj). Za svako grupno poređenje može se izračunati pokazatelj veličine uticaja (tj. jačine veze između promenljivih), po postupku opisanom u odeljku posvećenom Man-Vitnijevom U testu.

Fridmanov test

Fridmanov test je neparametarska alternativa jednofaktorskoj analizi varijanse ponovljenih merenja (videti poglavlje 18). Upotrebljava se kada uzmete isti uzorak subjekata ili slučajeva i izmerite ih u tri ili više navrata ili pod tri različita uslova.

Kratak pregled Fridmanovog testa

Primer istraživačkog pitanja: Ima li promene rezultata ispitivanja straha od statistike obavljenim u tri vremenska perioda (pre intervencije, neposredno posle intervencije i tri meseca nakon intervencije)?

Sta vam treba: Jedan uzorak subjekata, merenih na istoj skali ili u tri vremenski različita navrata ili pod tri različita uslova.

Pretpostavke: Videti opšte pretpostavke za neparametarske tehnike, navedene na početku poglavlja.

Parametarska alternativa: Jednofaktorska analiza varijanse ponovljenih merenja (istih subjekata).

Postupak za obavljanje Fridmanovog testa

Da biste mogli pratiti ovaj primer, otvorite datoteku **experim3ED.sav**.

1. Otvorite meni **Analyze**, izaberite u njemu stavku **Nonparametric Tests**, pa pritisnite **K Related Samples**.
2. Pritisnite promenljive koje predstavljaju ta tri merenja (npr. fost1, fost2, fost3).
3. U odeljku **Test Type** treba da je potvrđeno polje **Friedman**.
4. Pritisnite dugme **Statistics**. Potvrdite polje **Quartiles**. Pritisnite dugme **Continue**.
5. Pritisnite **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Tim postupkom generišete ovu komandu:

```
NPARTESTS
/FRIEDMAN = fost1 fost2 fost3
/STATISTICS QUARTILES
/MISSING LISTWISE.
```

Evo i rezultata:

Descriptive Statistics

	N	Percentiles		
		25th	50th (Median)	75th
fost1	30	37,00	40,00	44,00
fost2	30	34,50	38,00	40,00
fost3	30	32,00	35,50	40,00

Ranks

	Mean Rank
fear of stats time1	2,78
fear of stats time2	2,03
fear of stats time3	1,18

Test Statistics^a

N	30
Chi-Square	41,568
df	2
Asymp. Sig.	,000

^a Friedman Test

Tumačenje rezultata Fridmanovog testa

Rezultati ovog testa pokazuju da postoje značajne razlike u rezultatima na skali straha od statistike dobijenim u tri vremenska razdoblja. To pokazuje vrednost **Sig.** od 0,000 (što zapravo znači manje od 0,0005). Poredeći srednje vrednosti rangova (**Mean Ranks**) za tri skupa rezultata, izgleda kao da strah od statistike opada s vremenom.

Rezultate ove analize mogli biste predstaviti ovako:

Rezultati Fridmanovog testa pokazuju da postoji statistički značajna razlika rezultata na skali straha od statistike dobijenih u tri vremenska razdoblja (pre intervencije, neposredno posle intervencije, 3 meseca posle intervencije), χ^2 (2, n = 30) = 41,57, $p < 0,005$). Pregled medijana je pokazao opadanje straha od statistike u vremenskom razdoblju od pre intervencije ($Md = 40$) do posle intervencije ($Md = 38$), kao i dalje opadanje 3 meseca posle intervencije ($Md = 35,5$).

Naknadni testovi i veličina uticaja

Kada se utvrdi da postoji statistički značajna razlika koja nastaje negde između tri tačke u vremenu, sledeći korak bi bili naknadni testovi poređenja vremenskih tačaka od interesa (videti prethodno razmatranje naknadnih testova u vezi s Kruskal-Volisovim testom). U ovom slučaju, naknadno ispitivanje bi obuhvatilo pojedinačne Vilkoksove testove ranga, uz Bonferoni korekciju alfa vrednosti kako bi se izbegla greška I vrste. Ja bih se verovatno odlučila za poređenje rezultata u trenutku 1 (pre intervencije) s rezultatima u trenutku 2 (posle intervencije), i zatim onih u trenutku 2 (posle intervencije) sa onima dobijenim u trenutku 3 (3 meseca posle intervencije). To su dva testa; zato bi moj revidirani alfa nivo za utvrđivanje statističke značajnosti bio $0,05 / 2 = 0,025$. Pokazatelje veličine uticaja (jačine veze) trebalo bi izračunati za svako konkretno poređenje obavljeno Vilkoksovim testom ranga.

Dodatne vežbe

Poslovno okruženje

Datoteka s podacima: **staffsurvey3ED.sav**. Pojediniosti o datoteci videti u dodatku.

1. Pomoću hi-kvadrat testa nezavisnosti uporedite proporciju stalnih i privremeno zaposlenih (*employstatus*) koji bi preporučili organizaciju kao dobro mesto za rad (*recommend*).
2. Pomoću Man-Vitnijevog U testa uporedite dobijene vrednosti zadovoljstva osoblja (*totsatis*) za stalno i za privremeno zaposlene (*employstatus*).
3. Pomoću Kruskal-Volisovog testa uporedite dobijene vrednosti zadovoljstva osoblja (*totsatis*) za sve kategorije dužine radnog staža (upotrebite promenljivu *servicegp3*).

Zdravstvo

Datoteka s podacima: *sleep3ED.sav*. Pojediniosti o datoteci videti u dodatku.

1. Pomoću hi-kvadrat testa nezavisnosti uporedite proporciju muškaraca i žena (*gender*) koji odgovaraju da imaju problema sa spavanjem (*problem*).
2. Pomoću Man-Vitnijevog U testa uporedite srednje vrednosti pospanosti (ukupne vrednosti na skali Sleepiness and Associated Sensations Scale: *totSAS*) za muškarce i žene (*gender*).
3. Pomoću Kruskal-Volisovog testa uporedite srednje vrednosti merenja pospanosti (ukupne vrednosti na skali Sleepiness and Associated Sensations Scale: *totSAS*) za tri starosne grupe definisane promenljivom *agegp3* (≤ 37 , 38–50, 51+).



T-testovi

SPSS ima više vrsta t-testova. Mi ćemo ovde razmotriti dve:

- *t-test nezavisnih uzoraka*, koji se upotrebljava za poređenje srednje vrednosti obeležja merenog u dve različite grupe ljudi ili u različitim okolnostima; i
- *t-test uparenih uzoraka*, koji se upotrebljava za poređenje srednje vrednosti obeležja *iste* grupe ljudi merenog u dva navrata ili kada su subjekti upareni (engl. *matched pairs*).

U oba slučaja, porede se vrednosti određene neprekidne promenljive merene u *dve* grupe ili u *dva* navrata. Kada grupa odnosno okolnosti ima više od dve, umesto t-testa upotrebljava se analiza varijanse.

Obe vrste t-testova razmotrene u ovom poglavlju počivaju na nekoliko pretpostavki koje treba proveriti pre obavljanja analiza. Opšte pretpostavke za obe vrste t-testova navedene su u uvodu u peti deo knjige. T-test uparenih uzoraka ima i jednu jedinstvenu dodatnu pretpostavku, koju takođe treba proveriti. Ona je detaljno objašnjena u kratkom pregledu t-testova uparenih uzoraka. Pre nego što nastavite da čitate ovo poglavlje, trebalo bi da pročitate uvod u peti deo knjige.

T-test nezavisnih uzoraka

T-test nezavisnih uzoraka (engl. *independent-samples t-test*) upotrebljavamo za poređenje srednje vrednosti neke neprekidne promenljive u dve različite grupe subjekata.

Objašnjenje primera

Da bismo ilustrovali upotrebu ove tehnike, koristićemo datoteku s podacima *survey3ED.sav*. U primeru se porede razlike između polova u izmerenom nivou samopoštovanja. Imamo dve promenljive, pol (*sex*), gde su muškarci šifrovani sa 1, a žene sa 2, i *tslfest* (skraćeno od total self-esteem), što je ukupan rezultat dobijen na skali samopoštovanja od deset stavki. (Videti u

dotatku pojedinosti o istraživanju, promenljivama i upitniku upotrebljenom za prikupljanje podataka.) Ukoliko želite da pratite primer i odmah sprovedite navedena uputstva, pokrenite SPSS i otvorite datoteku **survey3ED.sav**.

Kratak pregled t-testa nezavisnih uzoraka

Primer istraživačkog pitanja: Da li postoji značajna razlika između srednjih vrednosti samopoštovanja kod muškaraca i žena?

Šta vam treba: Dve promenljive:

- jedna kategorijska, nezavisna promenljiva (npr. muškarci/žene); i
- jedna neprekidna, zavisna promenljiva (npr. vrednost samopoštovanja).

Šta se postiže: T-test nezavisnih uzoraka kazuje da li postoji statistički značajna razlika u prosečnom rezultatu merenja nekog obeležja u dve grupe (tj. da li se muškarci i žene značajno razlikuju po nivou samopoštovanja). Rečeno jezikom statistike, ispitujemo verovatnoću da dva skupa rezultata (dobijenih od muškaraca i od žena) potiču iz iste populacije.

Pretpostavke: Pretpostavke za ovaj test obrađene su u uvodu u peti deo knjige. Trebalo bi da pročitate taj deo teksta pre nego što nastavite.

Neparametarska alternativa: Man-Vitnijev U test (videti 16. poglavlje).

Postupak za obavljanje t-testa nezavisnih uzoraka

1. Otvorite meni **Analyze**, izaberite u njemu stavku **Compare means**, pa pritisnite **Independent Samples T test**.
2. Pritisnite odabranu zavisnu (neprekidnu) promenljivu (npr. ukupno samopoštovanje, **tslfest**) i prebacite je u polje **Test variable**.
3. Pritisnite odabranu nezavisnu (kategorijsku) promenljivu (npr. pol, **sex**) i prebacite je u polje **Grouping variable**.
4. Pritisnite dugme **Define groups** i upišite brojeve pomoću kojih su grupe šifrovane u skupu podataka. U tekućoj datoteci s podacima, 1=muškarci, 2=žene; zato u polje **Group 1** upišite 1; a u polje **Group 2** upišite 2. Ako ne možete da se setite šifara za svaku grupu, desnim tasterom miša pritisnite ime promenljive i u kontekstnom meniju izaberite **Variable Information**. Prikaže se okvir sa opisima i brojevima (šiframa) koje treba upisati za tu promenljivu. Pritisnite dugme **Continue**.
5. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

```
T-TEST
GROUPS = sex(1 2)
/MISSING = ANALYSIS
/VARIABLES = tslfest
/CRITERIA = CI(.95) .
```

Evo odabranog dela rezultata ovog postupka.

	SEX	N	Mean	Std. Deviation	Std. Error Mean
Total self-esteem	MALES	184	34.02	4.91	.36
	FEMALES	252	33.17	5.71	.36

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
Total self-esteem	Equal variances assumed	3.506	.062	1.622	434	.105	.85	.52	-.18	1.87
	Equal variances not assumed			1.661	422.349	.098	.85	.51	-.16	1.85

Tumačenje rezultata t-testa nezavisnih uzoraka

Korak 1: provera informacija o grupama

U tabeli **Group Statistics** SPSS je ispisao srednju vrednost (engl. *mean*) i standardno odstupanje (engl. *standard deviation*) obeležja od te vrednosti za svaku grupu (u ovom slučaju: muškarci/žene). Naveden je i broj osoba u svakoj grupi (N). Uvek najpre proverite te brojeve. Da li su tačni iznosi N za muškarce i žene? Nedostaje mnogo podataka? Ako je tako, otkrijte razlog. Možda ste upisali pogrešne šifre za muškarce odnosno žene (0 i 1, umesto 1 i 2). Pročitajte šta piše u šifarniku.

Korak 2: provera pretpostavki

U prvom delu tabele **Independent Samples Test** dati su rezultati Leveneovog testa jednakosti varijansi. Ispituje se da li je jednaka varijansa (promenljivost) rezultata u dve grupe (muškaraca i žena). Ishod ovog testa određuje koju t-vrednost (od dve koje je SPSS izračunao) treba smatrati tačnom i upotrebiti.

- Kada je veličina **Sig.** (u odeljku **Levene's Test for Equality of Variances**) veća od 0,05 (npr. 0,07, 0,10), treba upotrebiti prvi red tabele, izračunat za slučaj jednakih varijansi (**Equal variances assumed**).
- Ukoliko nivo značajnosti Leveneovog testa iznosi $p=0,05$ ili manje (npr. 0,01, 0,001), to znači da varijanse dve grupe (muškarci/žene) *nisu* jednake. Dakle, podaci ne zadovoljavaju pretpostavku o jednakosti varijansi. Ne paničite; SPSS je ljubazno izračunao alternativnu vrednost t, koja kompenzuje činjenicu da varijanse nisu jednake. Treba upotrebiti drugu vrstu tabele rezultata t-testa, izračunatu bez pretpostavljanja jednakosti varijansi (**Equal variances not assumed**).

U gornjem primeru, nivo značajnosti Leveneovog testa je 0,062. To je više od graničnih 0,05; znači da pretpostavka o jednakosti varijansi nije bila narušena. Zato u izveštaju treba navesti vrednost t iz prvog reda tabele.

Korak 3: procena razlike između grupa

Da biste utvrdili postoji li značajna razlika između dve grupe, pogledajte kolonu Sig. (2-tailed) u odeljku t-test for Equality of Means. Date su dve vrednosti – prva za slučaj jednakih varijansi, druga za nejednake varijanse. Odaberite vrednost za koju vam rezultat Leveneovog testa kazuje da je tačan (videti korak 2).

- Kada je broj u koloni Sig. (2-tailed) jednak ili manji od 0,05 (npr. 0,03, 0,01, 0,001), tada postoji značajna razlika između srednjih vrednosti zavisne promenljive u svakoj od dve grupe.
- Kada je broj iznad 0,05 (npr. 0,06, 0,10), razlika između dve grupe nije značajna, već slučajna.

U gornjem primeru, broj u koloni Sig. (2-tailed) iznosi 0,105. Pošto je ta veličina iznad zahtevane granične vrednosti 0,05, treba zaključiti da ne postoji statistički značajna razlika između srednjih vrednosti samopoštovanja muškaraca i žena. Tabela sadrži i srednju vrednost te razlike između dve grupe (Mean Difference), koja iznosi 0,85; a u odeljku 95% Confidence Interval of the Difference, donju (Lower) i gornju (Upper) granicu intervala koji sa verovatnoćom od 95% sadrži stvarnu veličinu te razlike.

Izračunavanje veličine uticaja u t-testu nezavisnih uzoraka

Veličina uticaja (engl. *effect size*) koju svim ovim testovima pokušavamo da otkrijemo, razmotrena je u uvodu u peti deo knjige. Pokazatelji veličine uticaja ukazuju na veličinu razlike između grupa, a ne samo da li je ta razlika slučajna ili ne. Veličina uticaja se procenjuje pomoću više pokazatelja, od kojih se najčešće upotrebljavaju eta kvadrat i Koenov d. Eta kvadrat može imati vrednost u opsegu od 0 do 1 i predstavlja proporciju varijanse u zavisnoj promenljivoj objašnjenju nezavisnom promenljivom (grupisanja). SPSS ne izračunava pokazatelj eta kvadrat u t-testovima, ali ga možemo sami izračunati pomoću ostalih datih rezultata.

Formula za eta kvadrat glasi:

$$\text{Eta kvadrat} = \frac{t^2}{t^2 + (N1 + N2 - 2)}$$

Kada u formulu uvrstimo odgovarajuće brojeve iz gornjeg primera:

$$\text{Eta kvadrat} = \frac{1,62^2}{1,62^2 + (184 + 252 - 2)}$$

$$\text{Eta kvadrat} = ,006$$

Smernice (Cohen, 1988, str. 284–7) za tumačenje ove veličine glase:

0,01=mali uticaj,

0,06=umeren utical,

0,14=veliki uticaj.

Što se tiče našeg primera, veličina uticaja iznosi samo 0,006, dakle vrlo je mali. Procentualno izraženo (pomnožite izračunati eta kvadrat sa 100), polna razlika objašnjava samo 0,6 procenata varijanse samopoštovanja.

U literaturi iz nekih oblasti (npr. medicine), veličina uticaja često se izražava Koenovim pokazateljem d. Brz i kratak način izračunavanja pokazatelja veličine uticaja, među kojima i Koenovog pokazatelja d, potražite na Web adresi <http://web.uccs.edu/lbecker/Psy590/escalc3.htm>.

Predstavljanje rezultata t-testa nezavisnih uzoraka

Rezultate ove analize mogli biste predstaviti ovako:

T-testom nezavisnih uzoraka upoređeni su rezultati ispitivanja samopoštovanja muškaraca i žena. Nije bilo značajne razlike rezultata kod muškaraca ($M = 34,02$, $SD = 4,91$) odnosno žena, $M = 33,17$, $SD = 5,71$; $t(434) = 1,62$, $p = 0,11$ (obostrano). Razlika između srednjih vrednosti obeležja po grupama (prosečna razlika = 0,85, 95% CI: -1,80 do 1,87) bila je vrlo mala (eta kvadrat = 0,006).

T-test uparenih uzoraka

T-test uparenih uzoraka (ili ponovljenih merenja) upotrebljavamo kada imamo samo jednu grupu ljudi (ili preduzeća, ili mašina itd.), a podatke od njih prikupljamo u dva navrata ili pod dva različita uslova. Primer takve situacije su eksperimenti u kojima merenje obavljam pre i posle neke intervencije. Svaku osobu ocenimo na nekoj neprekidnoj skali u trenutku 1, i zatim ponovo u trenutku 2, nakon što smo ih podvrgli nekoj eksperimentalnoj obradi ili intervenciji. Isti način se upotrebljava za uparene subjekte (tj. svaka osoba je uparena s nekom drugom na osnovu određenih kriterijuma, kao što su starost, pol i sl.). Jedan deo para biva podvrgnut intervenciji 1, a drugi intervenciji 2. Zatim se za svaki par porede rezultati na nekoj neprekidnoj skali.

T-testovi uparenih uzoraka služe i za merenje odgovora iste osobe na dva različita pitanja, recimo da oceni važnost dva aspekta života (npr. zdravlja, finansijske obezbeđenosti) u pogledu njihovog uticaja na zadovoljstvo životom. U ovom slučaju, oba aspekta treba oceniti na istoj skali (npr. od 1=nimalo važno do 10=veoma važno).

Objašnjenje primera

Kao ilustraciju t-testa uparenih uzoraka upotrebila sam podatke iz datoteke **experim3ED.sav** (dostupne na pratećoj Web lokaciji knjige). To je veštačka datoteka s podacima, napravljena i modifikovana tako da ilustruje razne statističke tehnike. Sve pojedinosti projekta istraživanja, upotrebljenih mera itd. date su u dodatku.

U donjem primeru istražiću uticaj intervencije koja je trebalo da poveća samopouzdanje studenata u pogledu njihove sposobnosti da ovladaju statistikom. Od studenata je bilo traženo da popune test za ispitivanje straha od statistike (engl. *Fear of Statistics Test*, FOST) i pre (trenutak 1) i posle intervencije (trenutak 2). Upotrebiću ove dve promenljive iz datoteke s podacima: FOST1 (rezultati ispitivanja straha od statistike u trenutku 1) i FOST2 (rezultati ispitivanja straha od statistike u trenutku 2). Da biste pratili ovaj primer, pokrenite SPSS i otvorite datoteku s podacima **experim3ED.sav**.

Kratak pregled t-testa uparenih uzoraka

Primer istraživačkog pitanja: Da li se značajno razlikuju rezultati ispitivanja straha od statistike, nakon što su ispitanici učestvovali u intervenciji koja je trebalo da poveća njihovu veru u sposobnost da uspešno ovladaju statistikom? Utiče li ta intervencija na rezultate učesnika ispitivanja straha od statistike?

Šta vam treba: Jedan skup subjekata (ili uparenih subjekata). Svaka osoba (odnosno par) treba da da oba skupa odgovora. Dve promenljive:

- jedna kategorijska nezavisna promenljiva (u ovom slučaju, to je vreme; sa dva različita nivoa: trenutak 1, trenutak 2);
- jedna neprekidna, zavisna promenljiva (npr. rezultati ispitivanja straha od statistike) merena u dva navrata ili pod različitim okolnostima.

Šta se postiže: T-test uparenih uzoraka kazuje da li postoji statistički značajna razlika u srednjim vrednostima rezultata dobijenih u trenutku 1 i trenutku 2.

Pretpostavke: Osnovne pretpostavke t-testova obrađene su u uvodu u peti deo knjige. Trebalo bi da pročitate taj deo teksta pre nego što nastavite. **Dodatna pretpostavka:** razlika između dva rezultata dobijena za svakog ispitanika treba da je normalno raspodeljena. Kršenje te pretpostavke ne prouzrokuje ozbiljne probleme kada uzorci imaju više od 30 subjekata.

Neparametarska alternativa: Vilkoksonov test ranga (videti 16. poglavlje).

Postupak obavljanja t-testa uparenih uzoraka

1. Otvorite meni **Analyze**, izaberite u njemu stavku **Compare Means**, pa pritisnite **Paired Samples T test**.
2. Pritisnite dve promenljive koje želite da uporedite za svakog ispitanika (npr. **fost1**: strah od statistike u trenutku 1, **fost2**: strah od statistike u trenutku 2).

3. Kada su obe promenljive izabrane, prebacite ih u polje **Paired Variables** tako što ćete pritisnuti dugme sa strelicom. Pritisnite **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

```
T-TEST
  PAIRS = fost1 WITH fost2 (PAIRED)
  /CRITERIA = CI(.95)
  /MISSING = ANALYSIS.
```

Evo rezultata prethodnog postupka:

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	fear of stats time1	40,17	30	5,16	,94
	fear of stats time2	37,50	30	5,15	,94

Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	fear of stats time1 - fear of stats time2	2,67	2,71	,49	1,66	3,68	5,394	29	,00

Tumačenje rezultata t-testa uparenih uzoraka

Tumačenje rezultata ove analize sastoji se od dva koraka.

Korak 1: određivanje ukupne značajnosti

U tabeli **Paired Samples Test** pogledajte poslednju kolonu, sa zaglavljem **Sig. (2-tailed)**; to je tražena verovatnoća donošenja pogrešnog zaključka. Kada je ta vrednost manja od 0,05 (npr. 0,04, 0,01, 0,001), treba zaključiti da postoji značajna razlika između dva rezultata. U gornjem primeru, verovatnoća je 0,000. Taj broj je zapravo zaokružen na tri decimalna mesta, što znači da je stvarna verovatnoća manja od 0,0005. Ta vrednost je znatno (sto puta) manja od zadatog alfa nivoa od 0,05. Zato treba zaključiti da postoji značajna razlika u rezultatima ispitivanja straha od statistike dobijenim u trenutku 1 odnosno trenutku 2. Zabeležite vrednost **t** (u ovom slučaju, 5,39) i broj stepeni slobode (**df=29**), pošto njih treba navesti u izveštaju o istraživanju.

Trebalo bi zabeležiti i da je prosečno smanjenje straha od statistike (**Mean**) bilo 2,67, te da se interval 95-procentne pouzdanosti proteže od donjih (**Lower**) 1,66 do gornjih (**Upper**) 3,68.

Korak 2: poređenje srednjih vrednosti

Pošto smo ustanovili da postoji značajna razlika, u sledećem koraku treba utvrditi koji skup rezultata sadrži veće srednje vrednosti (skup dobijen u trenutku 1 ili onaj u trenutku 2). To ćemo uraditi uvidom u prvu tabelu **Paired Samples Statistics**. U njoj su srednje vrednosti (**Mean**) izračunate za oba skupa rezultata pojedinačno. U našem slučaju, strah od statistike u trenutku 1 prosečno iznosi 40,17, dok u trenutku 2 iznosi 37,50. Dakle, zaključujemo da su se rezultati ispitivanja straha od statistike značajno smanjili od trenutka 1 (pre intervencije) do trenutka 2 (posle intervencije).

Upozorenje: Iako smo dobili značajnu razliku u rezultatima pre i posle intervencije, ne možemo reći da je ta intervencija prouzrokovala smanjenje vrednosti rezultata ispitivanja straha od statistike. Nažalost, istraživanja nisu tako jednostavna! Na smanjenje dobijenih rezultata ispitivanja straha od statistike moglo je uticati i mnogo drugih činilaca. Tome je moglo doprineti i samo proticanje vremena, bez ikakve intervencije. Tokom tog razdoblja mogli su se desiti i drugi događaji koji su uticali na stavove studenata prema statistici. Možda su na učesnike uticali stariji studenti statistike, koji su ih obavestili da je predavač „dobar“ zato što je ispit lako položiti. Možda su svi dobili ilegalnu kopiju ispitnih pitanja (s tačnim odgovorima)! Ima još mnogo drugih mogućih kontaminirajućih činilaca. Kad god je moguće, istraživač bi trebalo da predvidi te remetilačke činioce i statistički ukloni njihov uticaj ili ih uključi u eksperiment. U opisanom slučaju, istraživanje bi poboljšala kontrolna grupa koja ne bi bila izložena intervenciji, ali bi po svemu ostalom bila slična učesnicima. Time bi se isključio uticaj proticanja vremena, drugih događaja, podvala, ilegalnih kopija itd. na rezultate istraživanja.

Izračunavanje veličine uticaja u t-testu uparenih uzoraka

Iako se iz navedenih rezultata vidi da je vrlo mala verovatnoća da je dobijena razlika između dobijena dva skupa vrednosti bila slučajna, oni nam ne kažu mnogo o veličini uticaja date intervencije, tj. kolika je ta razlika. Jedan od načina da se to utvrdi bilo bi izračunavanje nekog pokazatelja te veličine (više o tome videti u uvodu u peti deo knjige). Pokazaćemo kako se izračunava i tumači eta kvadrat, jedan od najčešće upotrebljivanih pokazatelja veličine uticaja.

Eta kvadrat se može dobiti pomoću ove formule:

$$\text{Eta kvadrat} = \frac{(5,39)^2}{(5,39)^2 + 30 - 1} = \frac{29,05}{29,05 + 30 - 1} = ,50$$

Kada u formulu uvrstimo odgovarajuće brojeve iz gornjeg primera:

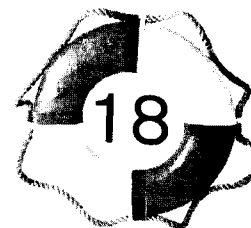
$$\text{Eta kvadrat} = \frac{(5,39)^2}{(5,39)^2 + 30 - 1} = \frac{29,05}{29,05 + 30 - 1} = ,50$$

Smernice (Cohen, 1988, str. 284–7) za tumačenje ove veličine glase: 0,01=mali uticaj, 0,06=umeren uticaj, 0,14=veliki uticaj. Pošto smo dobili eta kvadrat jednak 0,50, zaključujemo da je veoma velika razlika između rezultata ispitivanja straha od statistike dobijenih pre i posle intervencije.

Predstavljanje rezultata t-testa uparenih uzoraka

Osnovne pojedinosti koje treba navesti su ime testa, svrha testa, vrednost pokazatelja t, broj stepeni slobode (df), verovatnoća (alfa) dobijanja slučajnog rezultata i srednje vrednosti obeležja i standardna odstupanja za svaku grupu odnosno merenje. Pri objavljivanju rada u većini današnjih časopisa, potrebno je da se navede i pokazatelj veličine uticaja (npr. eta kvadrat). Rezultate opisane analize mogli biste predstaviti ovako:

T-testom uparenih uzoraka procenjen je uticaj intervencije na rezultate ispitivanja straha studenata od statistike (FOST). Utvrđeno je statistički značajno smanjenje vrednosti FOST od trenutka 1 ($M = 40,17$, $SD = 5,16$) do trenutka 2 ($M = 37,5$, $SD = 5,15$), $t(29) = 5,39$, $p < 0,0005$ (obostrano). Prosečno smanjenje vrednosti FOST bilo je 2,27, dok se interval 95-procentnog poverenja proteže od 1,66 do 3,68. Vrednost eta kvadrat (0,50) pokazuje da je uticaj intervencije velik.



Jednofaktorska analiza varijanse

Dodatne vežbe

Poslovno okruženje

Datoteka s podacima: *staffsurvey3ED.sav*. Pojediniosti o datoteci videti u dodatku.

1. Postupkom objašnjenim u odeljku o t-testu nezavisnih uzoraka uporedite srednje vrednosti zadovoljstva osoblja (*totsatis*) za stalno odnosno privremeno zaposlene (*employstatus*). Da li je razlika između srednjih vrednosti zadovoljstva značajna?

Zdravstvo

Datoteka s podacima: *sleep3ED.sav*. Pojediniosti o datoteci videti u dodatku.

1. Postupkom objašnjenim u odeljku o t-testu nezavisnih uzoraka uporedite srednje vrednosti pospanosti, dobijene na skali pospanosti i pridruženih osećanja (total score of Sleepiness and Associated Sensations Scale: *totSAS*) za muškarce odnosno žene (*gender*). Da li je razlika između srednjih vrednosti pospanosti značajna?

U prethodnom poglavlju, pomoću t-testova smo poredili rezultate ispitivanja dobijene u dve različite grupe odnosno u istoj grupi ali pod različitim uslovima. Međutim, u mnogim istraživanjima zanima nas poređenje prosečnih rezultata u više od dve grupe. Tada se upotrebljava analiza varijanse (ANOVA). Jednofaktorska analiza varijanse znači da postoji samo jedna nezavisna promenljiva (faktor), podeljena na više nivoa ili grupa, odnosno uslova. Primera radi, kada biste na osnovu ocena učenika iz matematike poredili delotvornost tri različita stila podučavanja, imali biste jedan faktor (stil podučavanja) s tri nivoa (npr. ceo razred, aktivnosti u malim grupama, samostalne aktivnosti uz korišćenje računara). Zavisna promenljiva je neprekidna (u ovom slučaju, to su ocene na testu iz matematike, koje se u SAD izražavaju procentualno, tj. u opsegu od 0 do 100).

Analiza varijanse je tako nazvana zato što poredi varijansu (promenljivost rezultata) između raznih grupa (za koju se veruje da je prouzrokuje nezavisna promenljiva) s tom promenljivošću unutar svake grupe (za koju se smatra da je posledica slučajnosti). Izračuna se pokazatelj (količnik) F, koji predstavlja varijansu između grupa podeljenu varijansom unutar grupa. Velika vrednost količnika F pokazuje da je veća promenljivost između grupa (koju prouzrokuje nezavisna promenljiva) nego unutar svake grupe (što su odstupanja ili reziduali, engl. *error term*, rezultata od srednje vrednosti).

Statistički značajan pokazatelj F kazuje da treba odbaciti nultu hipotezu, tj. tvrdnju da su prosečne vrednosti obeležja u populaciji jednake. Međutim, on ne kazuje koje se grupe razlikuju; to tek treba utvrditi naknadnim testovima. Alternativa sprovođenju naknadnih testova nakon dobijanja značajnog iznosa omnibus F testa, jeste da studiju isplanirate tako da obavite samo određena poređenja (tzv. planirana poređenja). Poređenje naknadnih i planiranih poređenja razmotreno je u uvodu u peti deo knjige; tamo je navedena i literatura za dublje proučavanje. Oba pristupa imaju svoje dobre i loše strane; pažljivo razmotrite svoj izbor pre nego što otpočnete analize. Naknadni testovi smanjuju verovatnoću grešaka I vrste, ali je taj pristup stroži, što otežava dobijanje statistički značajnih razlika. Ukoliko nemate jasne konceptualne razloge za poređenje samo određenih grupa, bolje je obaviti naknadnu analizu svih grupa.

U ovom poglavlju, razmotrićemo dve vrste jednofaktorske analize varijanse (ANOVA):

- ANOVA različitih grupa, koja se upotrebljava kada u svakoj grupi imate različite subjekte ili slučajeve (to je „projekat s nezavisnim grupama“); i
- ANOVA ponovljenih merenja, koja se upotrebljava kada iste subjekte merite pod različitim okolnostima (ili u različitim vremenskim trenucima); to je „projekat sa istim subjektima“.

U narednom odeljku posvećenom ANOVA analizi različitih grupa, ilustrovaćemo upotrebu i naknadnih testova i planiranih poređenja.

Jednofaktorska ANOVA različitih grupa s naknadnim testovima

Jednofaktorska ANOVA različitih grupa upotrebljava se kada imate jednu nezavisnu promenljivu (grupisanja) s tri ili više nivoa (grupa) i jednu zavisnu neprekidnu promenljivu. Reč jednofaktorska u nazivu kazuje da postoji samo jedna nezavisna promenljiva, a sintagma različitih grupa znači da u svakoj grupi imate različite subjekte ili slučajeve.

Kratak pregled jednofaktorske analize varijanse različitih grupa s naknadnim testovima

Primer istraživačkog pitanja: Da li se razlikuju rezultati merenja optimizma mladih, sredovečnih i starih subjekata?

Šta vam treba: Dve promenljive:

- jedna kategorijska nezavisna promenljiva s tri ili više kategorija. To može biti i neprekidna promenljiva rešifrovana tako da se dobiju tri grupe sa istim brojem članova (npr. starosne grupe – subjekti podeljeni u tri starosne kategorije: 29 godina ili mlađi, između 30 i 44, te 45 ili više godina).
- jedna neprekidna zavisna promenljiva (npr. optimizam).

Šta se postiže: Jednofaktorska ANOVA kazuje postoje li značajne razlike između srednjih vrednosti zavisne promenljive u tri grupe. Potom se naknadnim testovima može utvrditi koje grupe se razlikuju.

Pretpostavke: Opšte pretpostavke na kojima počiva ANOVA, razmotrene su u uvodu u peti deo knjige.

Neparametarska alternativa: Kruskal-Volison test (videti poglavlje 16).

Objašnjenje primera

Pokazaću ovu tehniku na primeru datoteke s podacima `survey3ED.sav`, dostupne na pratećoj Web lokaciji knjige (videti str. xi). Podaci potiču iz ankete sprovedene radi istraživanja faktora koji utiču na psihološko prilagođenje,

zdravlje i opšte raspoloženje ispitanika. Podaci su iz stvarnog istraživanja koje je obavila grupa mojih postdiplomskih studenata. Sve pojedinosti projekta istraživanja, upitnika, upotrebljenih skala itd. date su u dodatku. Ukoliko želite da pratite primer i odmah sprovedite navedena uputstva, pokrenite SPSS i otvorite datoteku `survey3ED.sav`.

Evo pojedinosti o promenljivama upotrebljenim u analizi.

Ime datoteke: `survey3ED.sav`

Promenljive:

- Ukupan optimizam (Total optimism, `Toptim`): Ukupan rezultat dobijen na skali optimizma. Brojevi od 6 do 30, gde veći brojevi pokazuju veći optimizam.
- Starost podeljena na tri gupe (Age 3 group, `Agegp3`): Ova promenljiva je dobijena rešifrovanjem, deljenjem starosti na tri grupe iste veličine (videti uputstva u poglavlju 8): grupa 1: 18–29 godina = 1, grupa 2: 30–44 godina = 2, grupa 3: 45+ godina = 3

Postupak obavljanja jednofaktorske analize varijanse različitih grupa s naknadnim testovima

1. Otvorite meni **Analyze**, izaberite u njemu stavku **Compare Means**, pa pritisnite **One-way ANOVA**.
2. Pritisnite odabranu zavisnu (neprekidnu) promenljivu (npr. Total optimism). Prebacite je u polje **Dependent List** tako što ćete pritisnuti dugme sa strelicom.
3. Pritisnite odabranu nezavisnu kategorijsku promenljivu (npr. `agegp3`) i prebacite je u polje **Factor**.
4. Pritisnite dugme **Options** i potvrdite polja **Descriptive**, **Homogeneity of variance test**, **Brown-Forsythe**, **Welsh** i **Means Plot**.
5. U odeljku **Missing values** (nedostajuće vrednosti) treba da je izabrana opcija **Exclude cases analysis by analysis**. Pritisnite dugme **Continue**.
6. Pritisnite dugme **Post Hoc**. Pritisnite dugme **Tukey**.
7. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

```
ONEWAY  
toptim BY agegp3  
/STATISTICS DESCRIPTIVES HOMOGENEITY BROWN-FORSYTHE  
WELCH  
/PLOT MEANS  
/MISSING ANALYSIS  
/POSTHOC = TUKEY ALPHA(.05).
```

Evo rezultata prethodnog postupka.

Oneway

Descriptives

Total Optimism

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1 18 - 29	147	21,36	4,551	,375	20,62	22,10	7	30
2 30 - 44	153	22,10	4,147	,335	21,44	22,77	10	30
3 45+	135	22,96	4,485	,386	22,19	23,72	8	30
Total	435	22,12	4,429	,212	21,70	22,53	7	30

Test of Homogeneity of Variances

Total Optimism

Levene Statistic	df1	df2	Sig.
,746	2	432	,475

ANOVA

Total Optimism

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	179,069	2	89,535	4,641	,010
Within Groups	8333,951	432	19,292		
Total	8513,021	434			

Robust Tests of Equality of Means

Total Optimism

	Statistic ^a	df1	df2	Sig.
Welch	4,380	2	284,508	,013
Brown-Forsythe	4,623	2	423,601	,010

a. Asymptotically F distributed

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Total Optimism
Tukey HSD

(I) age 3 groups	(J) age 3 groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1 18 - 29	2 30 - 44	-,744	,507	,308	-1,94	,45
	3 45+	-1,595*	,524	,007	-2,83	-,36
2 30 - 44	1 18 - 29	,744	,507	,308	-,45	1,94
	3 45+	-,851	,519	,230	-2,07	,37
3 45+	1 18 - 29	1,595*	,524	,007	,36	2,83
	2 30 - 44	,851	,519	,230	-,37	2,07

*. The mean difference is significant at the ,05 level.

Tumačenje rezultata jednofaktorske analize varijanse različitih grupa s naknadnim testovima

Descriptives

U ovoj tabeli dati su podaci o svakoj grupi (broj ispitanika u grupi, srednje vrednosti, standardno odstupanje (engl. *standard deviation*), minimum i maksimum, itd.). Uvek najpre pogledajte ovu tabelu. Da li su tačni brojevi ispitanika u svakoj grupi?

Test homogenosti varijansi

Opcija „Homogeneity of variance test“ (koju ste uključili u 4. koraku) daje Leveneov test homogenosti varijanse, pomoću kojeg se ispituje jednakost varijansi u rezultatima u svakoj od tri grupe. Pročitajte veličinu značajnosti (Sig.) za Leveneov test. Kada je taj broj *veći* od 0,05 (npr. 0,08, 0,12, 0,28), *niste* prekršili pretpostavku o homogenosti varijanse. U ovom primeru, veličina Sig. iznosi 0,475. Pošto je to više od 0,05, nismo prekršili pretpostavku o homogenosti varijanse. Da smo utvrdili narušavanje te pretpostavke, potražili bismo u rezultatima tabelu **Robust Tests of Equality of Means**. Tamo su prikazani rezultati dva testa (Welsh i Brown-Forsythe) otporna na kršenje pretpostavke o homogenosti varijanse.

ANOVA

U ovoj tabeli dati su zbrojevi kvadrata (engl. *sum of squares*) odstupanja (reziduala) rezultata od njihove srednje vrednosti, broj stepeni slobode itd. za analizu različitih grupa (engl. *between-groups*) i analizu istih subjekata (engl. *within-groups*). Trebalo bi da ih prepoznate jer ste ih vidali u udžbenicima za statistiku. Najviše nas zanima kolona Sig. Kada je vrednost Sig. manja od ili jednaka 0,05 (npr. 0,03, 0,01, 0,001), postoji statistički značajna razlika između srednjih vrednosti zavisne promenljive u tri grupe. To još ne znači da znate koja se grupa razlikuje od kojih drugih grupa. U tabeli **Multiple Comparisons** rezultata naknadnih testova (engl. *post-hoc tests*), date su i statističke značajnosti razlika između svakog para grupa. Srednje vrednosti za svaku grupu date su u tabeli **Descriptives**. U ovom primeru, sveukupna značajnost (kolona Sig. u tabeli ANOVA) iznosi 0,01, što je manje od 0,05, tj. pokazuje da je rezultat neke od grupa statistički značajan. Pošto je dobijena razlika statistički značajna, sada treba pogledati rezultate naknadnih testova.

Multiple comparisons

Ovu tabelu treba gledati samo kada je razlika u tabeli ANOVA statistički značajna, tj. kada je vrednost Sig. jednaka ili manja od 0,05. Naknadni testovi iz ove tabele kazuju tačno gde su razlike između grupa. Pogledajte niz kolonu **Mean Difference**. Potražite zvezdice (*) pored ispisanih brojeva.

Zvezdica znači da se dve upoređene grupe međusobno značajno razlikuju na nivou $p < 0,05$. U koloni Sig. dati su tačni iznosi značajnosti. Prema gornjim rezultatima, samo grupa 1 i grupa 3 se statistički značajno međusobno razlikuju. Drugim rečima, starosna grupa 18–29 godina i grupa 45+ godina značajno se razlikuju po nivou optimizma ispitanika.

Dijagrami srednjih vrednosti

Pomoću ovog dijagrama lako je uporediti srednje vrednosti rezultata dobijenih za razne grupe. Na njemu se vidi da je starosna grupa 18–29 godina zabeležila najmanji optimizam, a starosna grupa 45+ godina najveći.

Upozorenje: ovi dijagrami umeju da zavedu. U zavisnosti od skale upotrebljene na Y osi (ovde su to rezultati merenja optimizma), čak i male razlike mogu izgledati vrlo velike. U gornjem primeru, stvarna razlika srednjih vrednosti vrlo je mala (21,36, 22,10, 22,96), a na dijagramu izgleda znatna. Pouka: ne uzbuđujte se zbog dijagrama dok ne uporedite srednje vrednosti (date u tabeli Descriptives) i skalu dijagrama.

Izračunavanje veličine uticaja

U ovoj analizi, SPSS ne daje veličinu uticaja, ali ćemo je lako izračunati iz njegovih rezultata. (Veličina uticaja opisana je u uvodu u peti deo knjige.) Iz rezultata navedenih u tabeli ANOVA izračunaćemo eta kvadrat, jedan od najčešće upotrebljavanih pokazatelja veličine uticaja. Formula glasi (dobro bi vam došao kalkulator):

$$\text{Eta kvadrat} = \frac{\text{Zbir kvadrata odstupanja različitih grupa}}{\text{Ukupan zbir kvadrata}}$$

U ovom primeru, treba samo da podelite zbir kvadrata odstupanja različitih grupa (179,07) ukupnim zbirom kvadrata odstupanja (8513,02). Dobija se eta kvadrat jednak 0,02, što po Koenovom kriterijumu (Cohen, 1988, str. 284–7) kazuje da je uticaj razlike mali. Koen klasifikuje 0,01 kao mali uticaj, 0,06 kao srednji uticaj i 0,14 kao veliki uticaj.

Upozorenje: U ovom primeru smo dobili statistički značajan rezultat, ali je stvarni uticaj razlike srednjih vrednosti grupa vrlo mali (21,36, 22,10, 22,96). To je očigledno iz malog pokazatelja uticaja razlike (eta kvadrat = 0,02). Kada je uzorak dovoljno velik (ovde, $N = 435$), sasvim male razlike postaju statistički značajne, čak i kada je razlika između grupa praktički nevažna. Dakle, uvek pažljivo tumačite rezultate i uzmite u obzir sve dostupne informacije. Nemojte pridavati preveliku važnost statističkoj značajnosti, pošto treba povesti računa i o mnogim drugim činiocima.

Predstavljanje rezultata jednofaktorske analize varijanse različitih grupa s naknadnim testovima

Jednofaktorskom analizom varijanse istražen je uticaj starosti na nivoe optimizma, merene testom Life Orientation Test (LOT). Subjekti su po starosti podeljeni u tri grupe (grupa 1: 29 ili manje godina; grupa 2: 30 do 44 godine; grupa 3: 45 i više godina). Utvrđena je statistički značajna razlika na nivou $p < 0,05$ u LOT rezultatima tri starosne grupe: $F(2, 432) = 4,6$, $p = 0,01$. Uprkos statističkoj značajnosti, stvarna razlika između srednjih vrednosti grupa vrlo je mala. Veličina te razlike, izražena pomoću pokazatelja eta kvadrat, iznosi 0,02. Naknadna poređenja pomoću Tukeyevog HSD testa kazuju da se srednja vrednost grupe 1 ($M = 21,36$, $SD = 4,55$) značajno razlikuje od srednje vrednosti grupe 3 ($M = 22,96$, $SD = 4,49$). Grupa 2 ($M = 22,10$, $SD = 4,15$) se ne razlikuje značajno ni od grupe 1 ni od grupe 3.

Jednofaktorska ANOVA različitih grupa s planiranim poređenjima

U gornjem primeru, poredili smo rezultate merenja optimizma u svakoj od tri grupe. Međutim, ima situacija kada istraživače zanima samo poređenje određenih grupa. Primera radi, u eksperimentalnoj studiji s pet različitih intervencija može nas zanimati da li je intervencija 1 bolja od svih ostalih. Dakle, tada ne bismo poredili sve moguće kombinacije grupa. Kada nas zanima samo određeni podskup svih mogućih poređenja, treba sprovesti planirana poređenja umesto da radimo naknadne testove, zato što bi njihova moć bila premala. (Moć testa je verovatnoća da se u njemu ne napravi greška druge vrste, tj. da se ne odbaci pogrešna hipoteza; videti razmatranje moći testova u uvodu u peti deo knjige). Planirana poređenja su osjetljivija u otkrivanju razlika. S druge strane, u naknadnim testovima zadaju se stroži nivoi značajnosti da bi se smanjio rizik od greške I vrste, zbog većeg broja testova koji se sprovode. Odluku da li sprovesti planirana poređenja ili naknadne testove treba doneti pre nego što počnete analizu. Nije primereno uraditi oboje i zatim birati rezultate koji vam se više sviđaju! Kao ilustraciju planiranih poređenja upotrebiću iste podatke kao u prethodnom primeru, iako nije uobičajeno da se isti podaci analiziraju na oba načina. U ovom slučaju, postavimo malo drugačije pitanje: da li su subjekti u najstarijoj starosnoj grupi (45+ godina) optimističniji od onih u dve mlađe starosne grupe (18–29 godina; 30–44 godina)?

Zadavanje vrednosti koeficijenata

U narednom postupku, SPSS će tražiti da odaberete grupe koje želite da poredite. Da biste to mogli da uradite, morate zadati veličine koeficijenata. Mnogi studenti to početka ne znaju da urade, pa ćemo to sada objasniti.

Prvo treba da identifikujete grupe na osnovu različitih vrednosti nezavisne promenljive (agegp3):

- Grupa 1 (šifrovana sa 1): starost 18–29
- Grupa 2 (šifrovana sa 2): starost 30–44
- Grupa 3 (šifrovana sa 3): starost 45+

Zatim treba da odlučite koje grupe ćete porediti, a koje zanemariti. Daću za to nekoliko primera.

Primer 1

Za poređenje grupe 3 sa ostale dve grupe, koeficijenti bi trebalo da glase:

- Grupa 1: -1
- Grupa 2: -1
- Grupa 3: 2

Zbir koeficijenata uvek mora biti 0. Porede se grupe čiji se koeficijenti razlikuju. Zanemaruju se (ne učestvuju u poređenju) grupe čiji je koeficijent 0. Kada nekoj grupi date koeficijent 0, ostale koeficijente izmenite tako da njihov zbir bude 0.

Primer 2

Za poređenje grupe 3 s grupom 1 (a da grupa 2 uopšte ne učestvuje u poređenju), koeficijenti bi trebalo da budu:

- Grupa 1: -1
- Grupa 2: 0
- Grupa 3: 1

Vrednosti koeficijenata svake grupe unosite u odeljku **Contrasts** sledećeg SPSS postupka.

Postupak obavljanja jednofaktorske analize varijanse različitih grupa s planiranim poređenjima

1. Otvorite meni **Analyze**, izaberite u njemu stavku **Compare Means**, pa pritisnite **One-way ANOVA**.
2. Pritisnite odabranu zavisnu (neprekidnu) promenljivu (npr. total optimism). Pritiskom na strelicu prebacite tu promenljivu u polje **Dependent List**.
3. Pritisnite odabranu nezavisnu, kategorijsku promenljivu (npr. agegp3). Prebacite je u polje **Factor**.
4. Pritisnite dugme **Options**, pa potvrdite polja **Descriptive**, **Homogeneity-of-Variance** i **Means Plot**.

5. U odeljku **Missing Values** treba da je potvrđeno polje **Exclude cases analysis by analysis**. Pritisnite dugme **Continue**.

6. Pritisnite dugme **Contrasts**.

- U polje **Coefficients** upišite koeficijent prve grupe. (U gornjem primeru 1, to bi bio koeficijent -1.) Pritisnite dugme **Add**.
- Upišite koeficijent druge grupe (-1). Pritisnite dugme **Add**.
- Upišite koeficijent treće grupe (2). Pritisnite dugme **Add**.
- Ako ste ispravno uneli koeficijente, u polju **Coefficient Total** na dnu tabele pišaće 0.

7. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

```
ONEWAY
  toptim BY agegp3
  /CONTRAST= -1 -1 2
  /STATISTICS DESCRIPTIVES HOMOGENEITY BROWN-FORSYTHE
  WELCH
  /PLOT MEANS
  /MISSING ANALYSIS
  /POSTHOC = TUKEY ALPHA(.05).
```

Evo kako izgleda odabrani deo rezultata gornjeg postupka.

Contrast Coefficients

Contrast	AGEGP3		
	18-29	30-44	45+
1	-1	-1	2

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Total	Assume equal variances	1	2.45	.91	2.687	432	.007
Optimism	Does not assume equal	1	2.45	.92	2.654	251.323	.008

Tumačenje rezultata jednofaktorske analize varijanse različitih grupa s planiranim poređenjima

Tabele **Descriptives** i **Test of homogeneity of variances** sastavni su deo rezultata ovog postupka, ali je njihov sadržaj jednak kao u prethodnom primeru jednofaktorske analize varijanse (ANOVA) s naknadnim testovima. Zato ćemo ovde razmotriti samo onaj deo rezultata koji je relevantan za planirana poređenja.

Korak 1

U tabeli **Contrast Coefficients** navedeni su koeficijenti koje ste zadali za svaku grupu. Proverite da li je to ono što ste hteli.

Korak 2

Zanimaju nas glavni rezultati, dati u tabeli **Contrast Tests**. Računamo da su varijanse jednake pošto je Leveneov test (komentarisan u prethodnom primeru analize varijanse s naknadnim testovima) pokazao da je razlika varijansi statistički neznčajna, dakle slučajna; zato koristimo prvi red tabele (**Assume equal variances**). Značajnost zadatog poređenja (data u koloni **Sig.**) iznosi 0,007. To je manje od 0,05, pa zaključujemo da postoji statistički značajna razlika između grupe 3 (45+ godina) i ostale dve grupe. Iako značajna, stvarna razlika između srednjih vrednosti tih grupa vrlo je mala (21,36, 22,10, 22,96). Više o tome pročitajte u raspravi o rezultatima prethodnog primera.

Verovatno ste primetili da je rezultat analize planiranih poređenja dat kao pokazatelj t , a ne kao količnik F uobičajen u analizi varijanse. Odgovarajuću vrednost količnika F dobijate kvadriranjem iznosa t . U ovom primeru t iznosi 2,687, što kvadrirano daje 7,22. U izveštaju o ovakvoj analizi treba navesti i broj stepeni slobode. Prvi stepen slobode (za sva planirana poređenja) jeste 1; drugi je dat u koloni df iza kolone t (ovde je to 432). Dakle, te rezultate bismo opisali sa $F(1, 432) = 7,22, p = 0,007$.

Jednofaktorska ANOVA ponovljenih merenja

U jednofaktorskoj analizi varijanse ponovljenih merenja, svaki subjekat se meri pod dva ili više različitih uslova, odnosno na istoj neprekidnoj skali u tri ili više navrata. Služi i za poređenje odgovora ispitanika na tri ili više različitih pitanja odnosno stavki istog pitanja, ali se odgovori moraju meriti na istoj skali (npr. od 1=nimalo se ne slažem, do 5=potpuno se slažem).

Kratak prikaz jednofaktorske analize varijanse ponovljenih merenja

Primer istraživačkog pitanja: Ima li razlike u rezultatima merenim u tri navrata?

Šta vam treba: Jedna grupa subjekata merenih na istoj skali u tri navrata ili pod tri različita uslova, *ili* odgovori (mereni na istoj skali) svake osobe na tri različita pitanja odnosno stavke istog pitanja. Tu se radi o dve promenljive:

- jedna nezavisna promenljiva (kategorijska) (npr. trenutak 1/ trenutak 2/ trenutak 3); i
- jedna zavisna promenljiva (neprekidna) (npr. rezultati na testu poverenja). Rezultati testa, dobijeni u raznim vremenskim trenucima, biće u datoteci podataka prikazani u zasebnim kolonama.

Šta se postiže: Opisanom tehnikom se otkriva postojanje značajne razlike između tri skupa rezultata.

Pretpostavke: Videti raspravu o opštim pretpostavkama za analizu varijanse datim u uvodu u peti deo knjige.

Neparametarska alternativa: Fridmanov test (videti 16. poglavlje).

Objašnjenje primera

Prikažaću ovu tehniku na podacima iz datoteke **experim3ED.sav** (dostupne na pratećoj Web lokaciji). Pojednostosti o toj datoteci s podacima pročitajte u dodatku. Grupa studenata bila je pozvana da učestvuje u intervenciji (programu obuke) koja je trebalo da poveća njihovo poverenje u sopstvene sposobnosti za savladavanje statistike. Tri nivoa poverenja (mereno na skali subjektivnih doživljaja) bila su merena pre intervencije (trenutak 1), neposredno nakon intervencije (trenutak 2) i ponovo tri meseca kasnije (trenutak 3).

Da biste pratili ovaj primer, pokrenite SPSS i otvorite datoteku **experim3ED.sav**. Sledi prikaz imena upotrebljenih promenljivih, sa opisima i pojednostostima navedenim u datoteci s podacima:

Ime datoteke: **experim3ED.sav**

Promenljive:

- Rezultati merenja poverenja u svoje sposobnosti u trenutku 1 (**Confid1**): ukupni rezultati na testu kojim se ispituje poverenje u svoje sposobnosti za savladavanje statistike (**Confidence in Coping with Statistics**), sprovedenom pre intervencije. Opseg mogućih vrednosti od 10 do 40. Veće vrednosti pokazuju viši nivo poverenja.
- Rezultati ispitivanja poverenja u trenutku 2 (**Confid2**): ukupni rezultati na testu, sprovedenom neposredno nakon intervencije.
- Rezultati ispitivanja poverenja u trenutku 3 (**Confid3**): ukupni rezultati na testu sprovedenom 3 meseca posle intervencije.

Postupak obavljanja jednofaktorske analize varijanse ponovljenih merenja

1. Otvorite meni **Analyze**, izaberite u njemu stavku **General Linear Model**, pa pritisnite **Repeated Measures**.
2. U polje **Within Subject Factor Name** upišite ime koje će predstavljati vašu nezavisnu promenljivu (npr. Vreme ili Uslov). To nije stvarno ime te promenljive, već samo opis koji ste dali nezavisnoj promenljivoj.
3. U polje **Number of Levels** upišite broj nivoa odnosno grupa (vremenskih perioda) o kojem se radi (u ovom primeru, 3).
4. Pritisnite **Add**.
5. Pritisnite dugme **Define** na desnoj strani.

6. Izaberite tri promenljive koje predstavljaju promenljivu ponovljenih merenja (npr. confid1, confid2, confid3). Pritiskom na dugme sa strelicom prebacite ih u polje **Within Subjects Variables**.
7. Pritisnite polje **Options** u donjem desnom uglu prozora.
8. Potvrdite polja **Descriptive Statistics** i **Estimates of effect size** u odeljku **Display**. Ako hoćete da zadate naknadne testove, u odeljku **Factor and Factor Interactions** izaberite ime svoje nezavisne promenljive (npr. Vreme) i prebacite je u polje **Display Means for**. Potvrdite polje **Compare main effects**. U odeljku **Confidence interval adjustment**, pritisnite strelicu nadole i izaberite drugu opciju **Bonferroni**.
9. Pritisnite **Continue** i zatim **OK** (ili dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

GLM

```
confid1 confid2 confid3
/WSFACTOR = time 3 Polynomial
/METHOD = SSTYPE(3)
/PLOT = PROFILE( time )
/EMMEANS = TABLES(time) COMPARE ADJ(BONFERRONI)
/PRINT = DESCRIPTIVE ETASQ
/CRITERIA = ALPHA(.05)
/WSDESIGN = time.
```

Evo rezultata prethodnog postupka.

Descriptive Statistics

	Mean	Std. Deviation	N
confidence time1	19,00	5,37	30
confidence time2	21,87	5,59	30
confidence time3	25,03	5,20	30

Multivariate Tests^a

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
time	Pillai's Trace	,749	41,711 ^a	2,000	28,000	,000	,749
	Wilks' Lambda	,251	41,711 ^a	2,000	28,000	,000	,749
	Hotelling's Trace	2,979	41,711 ^a	2,000	28,000	,000	,749
	Roy's Largest Root	2,979	41,711 ^a	2,000	28,000	,000	,749

a. Exact statistic

b.
Design: Intercept
Within Subjects Design: time

Mauchly's Test of Sphericity^b

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^a		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TIME	,348	28,517	2	,000	,605	,640	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept+GROUP
Within Subjects Design: TIME

Pairwise Comparisons

Measure: MEASURE_1

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference ^a	
					Lower Bound	Upper Bound
1	2	-2,867*	,868	,008	-5,072	-,661
	3	-6,033*	,833	,000	-8,149	-3,918
2	1	2,867*	,868	,008	-,661	5,072
	3	-3,167*	,447	,000	-4,304	-2,030
3	1	6,033*	,833	,000	3,918	8,149
	2	3,167*	,447	,000	2,030	4,304

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Tumačenje rezultata jednofaktorske analize varijanse ponovljenih merenja

Primitili ste da su rezultati ove procedure brojni i naizgled složeni. Među njima su testovi pretpostavljene sferičnosti, te rezultati analiza varijanse za jednu (engl. *univariate*) odnosno više promenljivih (engl. *multivariate*); ova poslednja se naziva i multivarijaciona analiza varijanse. Detaljno razmatranje razlika između rezultata univarijacione i multivarijacione analize varijanse nije tema ove knjige; u ovom poglavlju razmotrićemo samo rezultate multivarijacione analize (više o tome videti u knjizi Stevensa, 1996, str. 466–9). Tumačimo samo rezultate SPSS-ove multivarijacione analize varijanse, zato što se univarijaciona temelji na sferičnosti podataka. Sferičnost znači da je varijansa rezultata merenja različitosti populacije za bilo koja dva uslova, jednaka toj varijansi za bilo koja druga dva uslova (što najčešće nije tačno). SPSS meri sferičnost Moklijevim (Mauchly) testom.

S druge strane, za multivarijacionu analizu varijanse nije neophodna sferičnost. Videćete u našem primeru da pretpostavka o sferičnosti nije zadovoljena, što pokazuje iznos Sig. od 0,000 u tabeli **Mauchly's Test of Sphericity**. Mada ima načina da se narušavanje te pretpostavke kompenzuje, bezbednije je pogledati rezultate multivarijacione analize varijanse datim u SPSS-ovom izlazu.

Pogledajmo ključne veličine koje treba razmotriti u rezultatima.

Tabela Descriptive Statistics

U prvoj tabeli rezultata, dati su opisni statistički pokazatelji (Mean, Standard deviation, N) tri skupa rezultata. Dobro je proveriti imaju li te veličine smisla. Da li je tačan broj osoba u svakoj grupi? Imaju li smisla srednje vrednosti merenog obeležja, uzimajući u obzir upotrebljenu skalu? U gornjem primeru, najniži rezultat merenja poverenja zabeležen je u trenutku 1 (pre intervencije), a najviši u trenutku 3 (tri meseca nakon završenog kursa iz statistike).

Tabela Multivariate Tests

U ovoj tabeli, zanimljiva je vrednost Wilks' Lambda i njoj pridružena verovatnoća u koloni Sig. Svi testovi multivarijacione analize varijanse dali su isti rezultat, ali se najčešće navodi Vilksov pokazatelj lambda. To je količnik zbira kvadrata odstupanja (rezultata od srednje vrednosti) unutar grupe i ukupnog zbira odstupanja; lambda poprima vrednosti između 0 i 1, pri čemu iznosi blizu 1 kazuju da se srednje vrednosti posmatrane nezavisne promenljive za grupu ne razlikuju mnogo, dok vrednosti blizu 0 kazuju da se one razlikuju. U ovom primeru, Vilksov lambda iznosi 0,25 uz verovatnoću 0,000 (što zapravo znači $p < 0,0005$). Pošto je verovatnoća p manja od 0,05, treba zaključiti da se grupne srednje vrednosti razlikuju, tj. da postoji statistički značajan uticaj vremena. To navodi na pomisao da se tokom ta tri vremenska razdoblja promenio nivo poverenja ispitanika u njihove sposobnosti za savladavanje statistike.

Veličina uticaja

Premda smo utvrdili statistički značajnu razliku između tri skupa rezultata, treba da procenimo i veličinu te razlike, tj. uticaja intervencije (videti raspravu o veličinama uticaja u uvodu u peti deo knjige). Zanima nas pokazatelj parcijalno eta kvadrat tj. Partial Eta Squared, naveden u poslednjoj koloni tabele Multivariate Tests. Eta kvadrat je količnik dela varijanse zavisne promenljive objašnjenog različitim kategorijama nezavisne promenljive, tj. varijansom između grupa, i ukupne varijanse, jednake zbiru prethodne i varijanse unutar grupa. Eta kvadrat poprima vrednosti između 0 i 1. Kada je jednaka 0, to kazuje da su srednje vrednosti posmatrane zavisne promenljive za sve kategorije jednake, tj. da nezavisna promenljiva ne utiče na zavisnu. Suprotno tome, kada je eta kvadrat jednak 1, znači da se posmatrano obeležje ne menja unutar kategorija, već samo između različitih kategorija. U ovoj studiji, dobijeni eta kvadrat iznosi 0,749. Na osnovu smernica koje je predložio Koen (Cohen, 1988, str. 284–7) (0,01=mali uticaj, 0,06=umeren uticaj, 0,14=veliki uticaj), reklo bi se da je uticaj intervencije (kursa iz statistike) vrlo velik.

Tabela Pairwise Comparisons

Kada dobijete statistički značajan rezultat iz goreopisanih analiza, to samo znači da među ispitivanim grupama postoji neka razlika. Još uvek ne znate koje se grupe odnosno skupovi rezultata međusobno razlikuju. (U ovom slučaju, to su grupe rezultata izmerenih u trenucima 1, 2, 3 tj. skupovi vreme 1, vreme 2, vreme 3.) To ćemo pročitati u tabeli Pairwise Comparisons, gde se porede svi parovi vremenskih tačaka i pripadne verovatnoće (u koloni Sig.) i pokazuje da li su razlike između parova značajne. U ovom primeru, značajne su sve razlike, pošto su sve vrednosti Sig. manje od 0,05.

Predstavljanje rezultata jednofaktorske analize varijanse ponovljenih merenja

Rezultate jednofaktorske analize varijanse ponovljenih merenja mogli biste predstaviti ovako:

Jednofaktorskom analizom varijanse ponovljenih merenja upoređeni su iznosi na testu poverenja u sopstvene sposobnosti za savladavanje statistike, dobijenih u Vreme 1 (pre intervencije), Vreme 2 (posle intervencije) i Vreme 3 (tri meseca posle intervencije). U tabeli 1 date su njihove srednje vrednosti i standardna odstupanja. Utvrđen je značajan uticaj vremena, Vilksov lambda = 0,25, $F(2, 28) = 41,17$, $p < 0,0005$, multivarijaciono parcijalno eta kvadrat = 0,75.

Tabela 1

Opisni statistički pokazatelji rezultata ispitivanja poverenja u sopstvene sposobnosti za savladavanje statistike, dobijenih u Vreme 1, Vreme 2 i Vreme 3.

Vremensko razdoblje	N	Srednja vrednost	Standardno odstupanje
Vreme 1 (pre intervencije)	30	19,00	5,37
Vreme 2 (neposredno posle intervencije)	30	21,87	5,59
Vreme 3 (3 meseca posle intervencije)	30	25,03	5,20

Dodatne vežbe

Poslovno okruženje

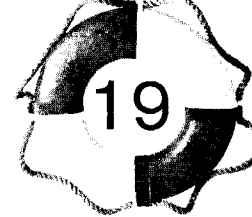
Datoteka s podacima: staffsurvey3ED.sav. Pojediniosti o datoteci videti u dodatku.

1. Sprovedite jednofaktorsku analizu varijanse s naknadnim testovima (gde je prikladno) radi poređenja ukupnih nivoa zadovoljstva osoblja (totsatis) za različite dužine zaposlenja (upotrebite promenljivu servicegp3).

Zdravstvo

Datoteka s podacima: sleep3ED.sav. Pojediniosti o datoteci videti u dodatku.

1. Sprovedite jednofaktorsku analizu varijanse s naknadnim testovima (gde je prikladno) radi poređenja srednjih vrednosti pospanosti na skali pospanosti i pridruženih osećanja (Sleepiness and Associated Sensations Scale total score: *totSAS*), za tri starosne grupe definisane promenljivom *agegp3* (≤ 37 , 38–50, 51+).



Dvofaktorska analiza varijanse različitih grupa

U ovom poglavlju upoznaćemo dvofaktorsku analizu varijanse različitih grupa (engl. *two-way between-groups analysis of variance*). Dvofaktorska znači da postoje dve nezavisne promenljive, a *različitih grupa* da su različiti ljudi u svakoj grupi. Ta tehnika omogućava istraživanje i pojedinačnog i zajedničkog uticaja dve nezavisne promenljive na jednu zavisnu. U poglavlju 18, pomoću jednofaktorske analize varijanse različitih grupa poredili smo nivo optimizma u tri starosne grupe (18–29, 30–44, 45+). Utvrdili smo značajnu razliku između grupa, dok su naknadni testovi pokazali da je najveća razlika između grupe najmlađih i grupe najstarijih. Stariji ljudi su pokazali viši nivo optimizma.

Sledeće pitanje koje se može postaviti glasi: da li to važi i za muškarce i za žene? Jednofaktorska ANOVA ne može da odgovori na takvo pitanje; njom je analiziran uzorak kao celina koja obuhvata i muškarce i žene. U ovom poglavlju, produbiću to istraživanje i razmotriti uticaj pola osoba na spomenuti nalaz. Zato ću imati dve nezavisne promenljive (starosna grupa i pol) i jednu zavisnu (optimizam).

Prednost dvofaktorske analize varijanse je to što se može ispitati osnovni uticaj svake nezavisne promenljive i mogući uticaj njihove interakcije. Uticaj interakcije postoji kada se uticaj jedne nezavisne promenljive na zavisnu menja u zavisnosti od vrednosti druge nezavisne promenljive. Primera radi, možda ćemo u ovom slučaju otkriti da se uticaj starosti na optimizam razlikuje kod muškaraca i kod žena. Možda se optimizam muškaraca povećava s godinama, dok se kod žena smanjuje. U tom slučaju bismo rekli da postoji uticaj interakcije (faktora, nezavisnih promenljivih) i da se uticaj starosti može opisati tek kada se saopšti na koju se grupu (muškarce/žene) tvrđenje odnosi.

Ako vam nije jasna razlika između zasebnih uticaja i uticaja interakcije, predlažem da tu građu proučite u nekom dobrom udžbeniku za statistiku (videti Gravetter & Wallnau, 2004; Harris, 1994; Runyon, Coleman & Pittenger, 2000; Tabachnick & Fidell, 2007). Pre nego što nastavite sa ovim poglavljem, preporučila bih vam da pročitate i uvod u peti deo ove knjige, gde sam razmotrila nekoliko tema relevantnih za tehnike analize varijanse.

Objašnjenje primera

Kao ilustraciju razmatrane tehnike upotrebiću datoteku s podacima `survey3ED.sav`, dostupnu na pratećoj Web lokaciji knjige (videti str. xi). Ti podaci potiču iz ankete sprovedene kako bi se istražili činioci koji utiču na psihološko prilagođenje, zdravlje i opšte raspoloženje ljudi. Radi se o stvarnim podacima iz istraživanja koje je obavila jedna grupa mojih postdiplomaca. Sve pojedinosti te studije, upitnika i upotrebljenih skala date su u dodatku. Ako želite da i sami pratite postupak opisan u ovom poglavlju, pokrenite SPSS i otvorite datoteku `survey3ED.sav`. U analizi su korišćene ove promenljive:

Ime datoteke: `survey3ED.sav`

- Ukupan optimizam (Toptim): ukupana vrednost na skali optimizma. Rezultati mogu imati vrednosti u opsegu od 6 do 30, pri čemu veći brojevi pokazuju viši nivo optimizma.
- Starosna grupa (Agegp3): rešifrovana promenljiva, dobijena deljenjem promenljive age na tri jednake grupe: grupa 1: 18–29 = 1; grupa 2: 30–44 = 2; grupa 3: 45+ = 3 (uputstva za rešifrovanje videti u poglavlju 8).
- Pol (promenljiva sex): muškarci (Males)=1, žene (Females)=2.

Kratak pregled dvofaktorske analize varijanse

Primer istraživačkog pitanja: Kako starost i pol osobe utiču na njen optimizam? Da li pol osobe menja uticaj njene starosti na optimizam?

Šta vam treba: Tri promenljive:

- dve kategorijske nezavisne promenljive (npr. pol: muškarci/žene; starosna grupa: mladi, sredovečni, stari); i
- jedna neprekidna zavisna promenljiva (npr. ukupan optimizam).

Šta se postiže: Dvofaktorska ANOVA služi za istovremeno ispitivanje uticaja svake nezavisne promenljive na zavisnu promenljivu, pri čemu se identifikuje i eventualni uticaj njihove interakcije. Primera radi, omogućava ispitivanje (a) polnih razlika u nivou optimizma, (b) razlike u optimizmu mladih, sredovečnih i starih osoba, i (c) interakcije te dve promenljive – da li starost različito utiče na optimizam muškaraca i žena?

Pretpostavke: Pretpostavke na kojima se zasniva ANOVA navedene su u uvodu u peti deo knjige.

Neparameterska alternativa: Ne postoji.

Postupak obavljanja dvofaktorske analize varijanse

1. Otvorite meni **Analyze**, izaberite u njemu stavku **General Linear Model**, i pritisnite **Univariate**.
2. Pritisnite odabranu zavisnu promenljivu (npr. toptim) i prebacite je u polje **Dependent variable**.
3. Pritisnite dve odabrane nezavisne kategorijske promenljive (sex, agegp3) i prebacite ih u polje **Fixed Factors**.
4. Pritisnite dugme **Options**.
 - Potvrdite polja **Descriptive Statistics**, **Estimates of effect size** i **Homogeneity tests**.
 - Pritisnite **Continue**.
5. Pritisnite dugme **Post Hoc**.
 - U spisku **Factors** na levoj strani, izaberite jednu ili više nezavisnih promenljivih od interesa (koje imaju tri ili više nivoa odnosno grupa), npr. agegp3.
 - Pritiskom na dugme sa strelicom prebacite je u odeljak **Post Hoc Tests for**.
 - Izaberite naknadni test koji želite da sprovedete (u ovom slučaju, **Tukey**).
 - Pritisnite **Continue**.
6. Pritisnite dugme **Plots**.
 - U polje **Horizontal** stavite nezavisnu promenljivu koja ima više grupa (npr. agegp3).
 - U polje **Separate Lines** stavite drugu nezavisnu promenljivu (npr. sex).
 - Pritisnite **Add**.
 - U odeljku **Plots** trebalo bi da pišu imena dve odabrane promenljive (npr. agegp3*sex).
7. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

UNIANOVA

```
toptim BY agegp3 sex
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/POSTHOC = agegp3 ( TUKEY )
/PLOT = PROFILE( agegp3*sex )
/PRINT = DESCRIPTIVE ETASQ HOMOGENEITY
/CRITERIA = ALPHA(.05)
/DESIGN = agegp3 sex agegp3*sex.
```

Evo kako izgledaju rezultati prethodnog postupka.

Descriptive Statistics

Dependent Variable: Total Optimism

AGEGP	SEX	Mean	Std. Deviation	N
18-29	MALES	21,38	4,33	60
	FEMALES	21,34	4,72	87
	Total	21,36	4,55	147
30-44	MALES	22,38	3,55	68
	FEMALES	21,88	4,58	85
	Total	22,10	4,15	153
45+	MALES	22,23	4,09	56
	FEMALES	23,47	4,70	79
	Total	22,96	4,49	135
Total	MALES	22,01	3,98	184
	FEMALES	22,20	4,73	251
	Total	22,12	4,43	435

Levene's Test of Equality of Error Variances^a

Dependent Variable: Total Optimism

F	df1	df2	Sig.
1,083	5	429	,369

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

^a. Design: Intercept +AGEGP3 +SEX+AGEGP3 * SEX

Tests of Between-Subjects Effects

Dependent Variable: total optimism

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	238,647 ^a	5	47,729	2,475	,032	,028
Intercept	206790,069	1	206790,069	10721,408	,000	,962
agegp3	150,863	2	75,431	3,911	,021	,018
sex	5,717	1	5,717	,296	,586	,001
agegp3 * sex	55,709	2	27,855	1,444	,237	,007
Error	8274,374	429	19,288			
Total	221303,000	435				
Corrected Total	8513,021	434				

a. R Squared = ,028 (Adjusted R Squared = ,017)

Multiple Comparisons

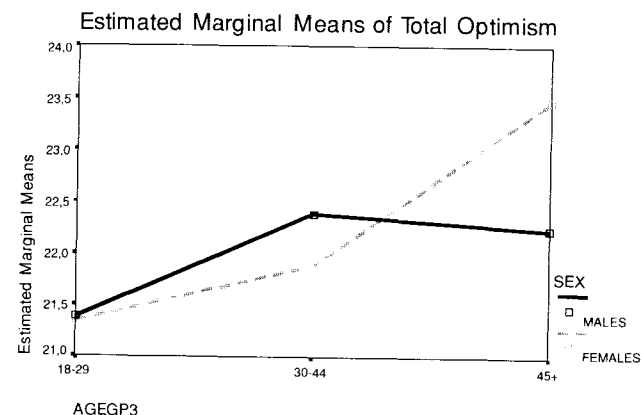
Dependent Variable: Total Optimism

Tukey HSD

(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
18-29	30-44	-,74	,51	,307	-1,93	,44
	45+	-1,60*	,52	,007	-2,82	-,37
30-44	18-29	,74	,51	,307	-,44	1,93
	45+	-,85	,52	,228	-2,07	,36
45+	18-29	1,60*	,52	,007	,37	2,82
	30-44	,85	,52	,228	-,36	2,07

Based on observed means.

* The mean difference is significant at the ,05 level.



Tumačenje rezultata dvofaktorske analize varijanse

- **Descriptive Statistics.** U ovoj tabeli su srednje vrednosti (Mean), standardna odstupanja (Std deviations) i N svake podgrupe. Proverite da su ti brojevi tačni. Pregledom strukture tih brojeva steći ćete uvid u uticaj odabranih nezavisnih promenljivih.
- **Levene's Test of Equality of Error Variances.** Ovde su rezultati test jedne od temeljnih pretpostavki analize varijanse. Najviše nas zanima značajnost tog rezultata, navedena u koloni Sig.. Odgovaralo bi nam da taj broj bude *veći* od 0,05 i stoga *neznačajan*. Značajan rezultat (Sig. manji od 0,05) kazuje da varijansa zavisne promenljive nije jednaka u svim grupama. Kada to utvrdite u svojoj studiji, preporučuje se

da zadate stroži nivo značajnosti (npr. 0,01) za vrednovanje rezultata dvofaktorske ANOVA analize, tj. da zaseban uticaj i uticaj interakcije smatrate značajnim samo ako je **Sig.** veće od 0,01. U gornjem primeru, **Sig.** iznosi 0,369. Pošto je to više od 0,05, treba zaključiti da homogenost varijansi nije narušena, tj. da je varijansa zavisne promenljive jednaka u svim grupama.

Dvofaktorska ANOVA glavni deo rezultata daje u tabeli **Tests of Between-Subjects Effects**. Te informacije nisu uvek poredane redom kojim ih treba analizirati.

Uticaji interakcije

Najpre treba proveriti da li postoji uticaj interakcije (npr. da se uticaj starosti na nivo optimizma menja u zavisnosti od toga da li je osoba muškarac ili žena). Kada utvrdite da je uticaj interakcije značajan, tumačenje osnovnih (zasebnih) uticaja nezavisnih promenljivih više nije lako ni jednostavno. Razlog je sledeći: da biste opisali uticaj jedne nezavisne promenljive, morate specificirati odgovarajuću vrednost druge nezavisne promenljive. U prikazanim rezultatima iz SPSS-a, treba pogledati red **agegp3*sex**. Značajnost interakcije je data u ćeliji kolone **Sig.** u tom redu. Kada je njena vrednost manja od ili jednaka 0,05 (npr. 0,03, 0,01, 0,001), uticaj interakcije je značajan. U našem primeru, uticaj interakcije nije značajan (**agegp3*sex: sig.** = 0,237). To kazuje da nema značajne razlike u uticaju starosti na optimizam muškaraca odnosno žena. *Upozorenje:* kada proveravate značajnost ovih rezultata, vodite računa o tome da čitate brojeve u koloni **Sig.**, pošto mnogo studenata misli da se značajnost čita u koloni **Partial Eta Squared**, što donosi katastrofalne posledice!

Zasebni uticaji

Utvrdili smo da uticaj interakcije nezavisnih promenljivih nije značajan; stoga je bezbedno i lako tumačiti njihove zasebne uticaje. Radi se jednostavno o uticaju jedne nezavisne promenljive (npr. uticaj pola na sve starosne grupe zajedno). U prvoj koloni sleva pronađite promenljivu od interesa (npr. **agegp3**). Da biste utvrdili postoji li njen zaseban uticaj, pročitajte odgovarajući broj u koloni **Sig.** u produžetku tog reda. Kada je očitana vrednost manja od ili jednaka 0,05 (npr. 0,03, 0,01, 0,001), zaseban uticaj te nezavisne promenljive je značajan. U gornjem primeru, postoji značajan zaseban uticaj starosne grupe (**agegp3: Sig.**=0,021), ali ne i značajan zaseban uticaj pola (**sex: Sig.**=0,586). To znači da se muškarci i žene ne razlikuju po svom optimizmu, ali da postoji razlika u nivoima optimizma mladih, sredovečnih i starijih osoba.

Veličina uticaja

Veličina uticaja promenljive **agegp3** data je u koloni **Partial Eta Squared** (0,018). Prema Koenovom kriterijumu (Cohen, 1988), taj uticaj se može klasifikovati kao mali (videti uvod u peti deo knjige). Dakle, mada statistički značajna tj. neslučajna, stvarna razlika između srednjih vrednosti vrlo je mala. U tabeli **Descriptives** videćete da srednje vrednosti za tri starosne grupe (za oba pola zajedno) iznose 21,36, 22,10, 22,96. Razlika između tih grupa nema praktičnog značaja.

Naknadni testovi

Iako znamo da se naše starosne grupe razlikuju, ne znamo konkretno koje se razlikuju: da li se **gp1** razlikuje od **gp2**, da li se **gp2** razlikuje od **gp3**, da li se **gp1** razlikuje od **gp3**? Odgovor na ta pitanja daće nam naknadni (engl. *post-hoc*) testovi (videti njihov opis u uvodu u peti deo knjige). Naknadni testovi su relevantni samo kada nezavisna promenljiva ima više od dva nivoa (grupe). Ti testovi redom porede sve parove grupa i kazuju da li se njihove srednje vrednosti značajno razlikuju. Rezultate naknadnih testova SPSS daje u sklopu ANOVA postupka. Međutim, ne bi trebalo da ih gledate dok zajedničkom (omnibus) analizom varijanse ne otkrijete neki značajan zaseban uticaj ili uticaj interakcije. U ovom primeru, dobili smo značajan zaseban (osnovni) uticaj promenljive **agegp3** (za razliku od promenljive **sex**, koja ga nema); zato je opravdano da naknadnim testovima ispitujemo uticaj **agegp3**.

Tabela Multiple Comparisons

Rezultati naknadnih testova dati su u tabeli **Multiple Comparisons**. Zadali smo da se izračuna Tukeyev (Tukey) test „zaista značajne razlike“ (Honestly Significant Difference, HSD), jedan od češće upotrebljivanih. Potražite u koloni **Sig.** brojeve manje od 0,05; to su značajni rezultati, označeni i zvezdicom u koloni **Mean Difference**. U gornjem primeru, značajno se međusobno razlikuju samo grupa 1 (18–29) i grupa 3 (45+).

Dijagrami

Na kraju rezultata iz SPSS-a nacrtan je dijagram nivoa optimizma muškaraca i žena za sve tri starosne grupe. Taj dijagram je veoma koristan jer omogućava vizuelni uvid u odnose između promenljivih, što je obično lakše od dešifrovanja velike tabele brojeva. Iako su na kraju rezultata, te dijagrame bi trebalo pogledati prve da biste bolje shvatili uticaj dve odabrane nezavisne promenljive. *Upozorenje:* kada tumačite dijagrame, imajte u vidu skalu upotrebljenu za crtanje zavisne promenljive. Ponekad se ono što na dijagramu izgleda kao ogromna razlika, zapravo svodi na tek nekoliko jedinica mere. Videćete to u tekućem primeru. Na prvom dijagramu izgleda kao da postoji velika razlika u rezultatima muškaraca i žena starije starosne grupe (45+). Međutim, kada pogled vratite na vertikalnu skalu na levoj strani dijagrama, videćete da je ta razlika zapravo mala (22,2 prema 23,5).

Predstavljanje rezultata dvofaktorske analize varijanse

Rezultati opisane analize mogu se predstaviti ovako:

Dvofaktorskom analizom varijanse različitih grupa istražen je uticaj pola i starosti na nivo optimizma, meren na skali Life Orientation Test (LOT). Subjekti su bili podeljeni u tri starosne grupe (grupa 1: 18–29 godina; grupa 2: 30–44 godina; grupa 3: 45 i više godina). Uticaj interakcije između pola i starosne grupe nije bio statistički značajan, $F(2, 429) = 1,44$, $p = 0,24$. Utvrđen je statistički značajan zaseban uticaj starosti, $F(2, 429) = 3,91$, $p = 0,02$; međutim, uticaj je mali (parcijalni eta kvadrat = 0,02). Naknadna poređenja pomoću Tukeyevog HSD testa pokazuju da se srednja vrednost rezultata u starosnoj grupi 18–29 godina ($M = 21,36$, $SD = 4,55$) značajno razlikuje od one u grupi 45+ godina ($M = 22,96$, $SD = 4,49$). Starosna grupa 30–44 godina ($M = 22,10$, $SD = 4,15$) ne razlikuje se značajno od ostalih grupa. Zaseban uticaj pola, $F(1, 429) = 0,30$, $p = 0,59$, nije dosegao statističku značajnost.

Dodatne analize kada se dobije značajan uticaj interakcije

Kada dobijete značajan uticaj interakcije, trebalo bi da odnose dodatno istražite naknadnim testovima. (To važi samo ako jedna od odabranih promenljivih ima tri ili više nivoa.) Jedan od načina da se to uradi jeste analiza jednostavnih uticaja. To znači da ćete zasebno razmotriti rezultate svih podgrupa. Uzorak treba podeliti na grupe jedne od nezavisnih promenljivih i za njih zasebno sprovesti jednofaktorske analize varijanse kojima se istražuje uticaj one druge promenljive. Da smo u gornjem primeru utvrdili značajan uticaj interakcije, mogli smo uzorak podeliti po polu i istražiti uticaj starosti na optimizam zasebno za muškarce i za žene. Za podelu uzorka i ponavljanje analiza za svaku grupu, upotrebite SPSS-ovu opciju **Split File**. Ta opcija služi za podelu uzorka na grupe jedne kategorijske promenljive i zasebno ponavljanje analiza za svaku grupu.

SPSS: Split File (Analyze > Split File > Split by groups)

Postupak za podelu uzorka

1. Otvorite meni **Data** i u njemu izaberite stavku **Split File**.
2. Pritisnite **Organize output by groups**.
3. Prebacite promenljivu grupisanja (sex) u polje **Groups based on**.
4. Tako ste uzorak podelili po polu (muškarci/žene) i zadali zasebno ponavljanje svih budućih analiza za te dve grupe.
5. Pritisnite dugme **OK**.

Nakon podele uzorka, istraživač sprovodi jednofaktorsku analizu varijanse (videti 18. poglavlje) i poredi nivoe optimizma u tri starosne grupe. Kada je procedura **Split File** uključena, dobićete zasebne rezultate za muškarce i žene.

Dodatne vežbe

Poslovno okruženje

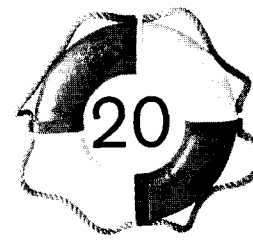
Datoteka s podacima: **staffsurvey3ED.sav**. Pojediniosti o datoteci videti u dodatku.

1. Sprovedite dvofaktorsku analizu varijanse s naknadnim testovima (ako treba) za poređenje nivoa zadovoljstva osoblja (*totsatis*) za različito trajanje zaposlenja (upotrebite promenljivu *servicegp3*), i to i za stalno i za privremeno zaposlene (*employstatus*).

Zdravstvo

Datoteka s podacima: **sleep3ED.sav**. Pojediniosti o datoteci videti u dodatku.

1. Sprovedite dvofaktorsku analizu varijanse s naknadnim testovima (ako treba) za poređenje srednjih vrednosti pospanosti na skali pospanosti i pridruženih osećanja (Sleepiness and Associated Sensations Scale total score: *totSAS*) muškaraca i žena (*gender*) podeljenih u tri starosne grupe definisane promenljivom *agegp3* (≤ 37 , 38–50, 51+).



Kombinovana analiza varijanse

U prethodnim poglavljima posvećenim analizi varijanse, upoznali smo analize različitih grupa subjekata (poređenje dve ili više grupa) i analize istih subjekata ili ponovljenih merenja (ista grupa subjekata ispitivanih pod dva ili više uslova). Dosad smo te pristupe razmatrali zasebno. Međutim, ima situacija kada u istoj studiji treba kombinovati oba ta pristupa, tako da jedna promenljiva bude „različitih subjekata“, a druga „istih subjekata“. Primera radi, želite da istražite uticaj određene intervencije na nivoe anksioznosti klijenata (ispitivanjem pre i posle intervencije), ali i da saznate da li se taj uticaj razlikuje za muškarce i za žene. U tom slučaju, imate dve nezavisne promenljive: jedna je „različitih subjekata“ (pol: muškarci/žene); druga je „istih subjekata“ (vreme). U tom slučaju, grupu muškaraca i žena biste podvrgnuli intervenciji i izmerili njihove nivoe anksioznosti u trenutku 1 (pre intervencije) i ponovo u trenutku 2 (posle intervencije).

SPSS omogućava kombinovanje promenljivih „različitih subjekata“ i „istih subjekata“ u istoj analizi. Takva analiza se u nekim udžbenicima naziva „split-plot ANOVA“ (SPANOVA). Ja sam se opredelila za termin koji koriste Tabachnick i Fidell (2007) – kombinovana analiza varijanse (engl. *mixed between-within subjects ANOVA*), pošto smatram da on najbolje opisuje o čemu se radi. Ta tehnika je proširenje analize varijanse ponovljenih merenja, prethodno razmotrene u poglavlju 18. Bilo bi dobro da pročitate to poglavlje pre nego što nastavite sa ovim.

U ovom poglavlju daću vrlo kratak prikaz kombinovane analize varijanse. Ukoliko tu tehniku nameravate da upotrebite u svom istraživanju, pročitate više o njoj (npr. Keppel & Zedeck, 2004; Harris, 1994; Stevens, 1996; Tabachnick & Fidell, 2007).

Objašnjenje primera

Kao ilustraciju kombinovane analize varijanse, upotrebiću datoteku s podacima **experim3ED.sav** dostupnu na pratećoj Web lokaciji knjige (videti str. xi). To su podaci iz fiktivne studije koja obuhvata ispitivanje uticaja dve različite intervencije preduzete kako bi se pomoglo studentima da se izbore sa svojom anksioznošću povodom predstojećeg kursa iz statistike (videti u dodatku sve pojedinosti o toj studiji). Studenti su bili podeljeni u dve jednake grupe, koje su popunile test za ispitivanje straha od statistike. Potom je jedna grupa pohađala kurs matematičkih veština, a druga učestvovala u programu izgradnje samopouzdanja. Grupe su opet popunile isti test neposredno posle te intervencije, kao i tri meseca kasnije. Ta datoteka sa izmišljenim podacima dostupna je na pratećoj Web lokaciji knjige. Ako želite da pratite doleopisane postupke, pokrenite SPSS i otvorite datoteku **experim3ED.sav**.

U ovom primeru, upoređiću uticaj kursa matematičkih veština (grupa 1) i kursa izgradnje samopouzdanja (grupa 2) na rezultate učesnika na testu za ispitivanje straha od statistike, popunjavanom u tri navrata. Slede imena upotrebljenih promenljivih i njihovi opisi u datoteci s podacima.

Ime datoteke: **experim3ED.sav**

Promenljive:

- Vrsta obuke (grupa): 1=Matematičke veštine 2= Izgradnja samopouzdanja
- Rezultati na testu za ispitivanje straha od statistike, popunjavanom u trenutku 1 (Fost1): ukupni rezultati na testu za ispitivanje straha od statistike popunjavanom pre intervencije. Rezultati poprimaju vrednosti u opsegu od 20 do 60. Veće vrednosti pokazuju veći strah od statistike.
- Rezultati na testu za ispitivanje straha od statistike, popunjavanom u trenutku 2 (Fost2): ukupni rezultati na testu za ispitivanje straha od statistike popunjavanom neposredno posle intervencije.
- Rezultati na testu za ispitivanje straha od statistike, popunjavanom u trenutku 3 (Fost3): ukupni rezultati na testu za ispitivanje straha od statistike, popunjavanom 3 meseca posle intervencije.

Kratak pregled kombinovane analize varijanse

Primer istraživačkog pitanja: Koja intervencija delotvornije smanjuje rezultate učesnika na testu kojim se ispituje strah od statistike, popunjavanom u tri navrata (pre intervencije, neposredno posle intervencije, 3 meseca kasnije)? Postoji li promena u rezultatima učesnika na tom testu, popunjavanom u tri navrata (pre intervencije, posle intervencije, 3 meseca kasnije)?

Šta vam treba: Najmanje tri promenljive:

- jedna kategorijska nezavisna promenljiva različitih subjekata sa dva ili više nivoa (grupa1/ grupa2);
- jedna kategorijska nezavisna promenljiva istih subjekata sa dva ili više nivoa (trenutak1/ trenutak2/ trenutak3);
- jedna neprekidna zavisna promenljiva (rezultati učesnika pri ispitivanju straha od statistike, mereni u svakom vremenskom periodu).

Šta se postiže: Pomoću ove analize se ispituje postoje li osnovni (zasebni) uticaji svake nezavisne promenljive i da li je interakcija između dve promenljive značajna. U ovom primeru, kazaće nam da li su se promenili rezultati testa za ispitivanje straha od statistike, popunjavanom u tri navrata (zasebni uticaj vremena). Upoređiću delotvornost te dve intervencije (matematičke veštine / izgradnja samopouzdanja) za smanjivanje straha od statistike (zaseban uticaj grupe). Najzad, kazaće nam da li se promena rezultata u vremenu razlikuje u te dve grupe (uticaj interakcije).

Pretpostavke: Videti uvod u peti deo knjige, gde su razmotrene opšte pretpostavke na kojima se temelji ANOVA.

Dodatna pretpostavka: Homogenost međukorelacija. Struktura korelacija među nivoima promenljive istih subjekata trebalo bi da bude jednaka za svaki nivo promenljive različitih subjekata. Zadovoljenost ove pretpostavke se ispituje u sklopu analize Boksovim (Box) pokazateljem M. Zbog velike osetljivosti tog pokazatelja, treba koristiti konzervativniji nivo alfa, čija je vrednost 0,001. Pretpostavka je zadovoljena kada pokazatelj nije značajan (tj. kada je pripadna verovatnoća veća od 0,001).

Neparametarska alternativa: Ne postoji.

Postupak za sprovođenje kombinovane analize varijanse

1. Otvorite meni **Analyze**, izaberite u njemu stavku **General Linear Model**, pa pritisnite **Repeated measures**.
2. U polje **Within-Subject Factor Name** upišite opisno ime faktora različitih subjekata (npr. vremena). To nije stvarno ime promenljive, već opisno po vašem izboru.
3. U polje **Number of Levels** upišite broj nivoa tog faktora (u ovom slučaju imamo tri vremenska perioda; zato upišite 3).
4. Pritisnite dugme **Add**. Pritisnite dugme **Define**.
5. Pritisnite promenljive koje predstavljaju faktor istih subjekata (npr. rezultate na testu za ispitivanje straha od statistike popunjenom u trenutku 1, trenutku 2 i trenutku 3).
6. Pritiskom na strelicu prebacite te promenljive u polje **Within-Subjects Variables**. Trebalo bi da se u polju ispišu imena odabranih promenljivih (samo kratka imena: fost1, fost2, fost3).

7. Pritisnite odabranu promenljivu različitih subjekata (npr. vrsta kursa). Pritisnite na strelicu prebacite tu promenljivu u polje **Between-Subjects Factors**.
8. Pritisnite dugme **Options**.
 - U odeljku **Display** potvrdite polja **Descriptive statistics, Estimates of effect size, Homogeneity tests**.
 - Pritisnite **Continue**.
9. Pritisnite dugme **Plots**.
 - Pritisnite faktor istih grupa (npr. vreme) i prebacite ga u polje **Horizontal Axis**.
 - U polju **Separate Lines** pritisnite promenljivu različitih grupa (npr. grupa).
10. Pritisnite **Add**. Trebalo bi da se u polju ispišu imena odabranih promenljivih (npr. vreme*grupa).
11. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

```
GLM
fost1 fost2 fost3 BY grupa
/WSFACTOR = vreme 3 Polynomial
/METHOD = SSTYPE(3)
/PLOT = PROFILE( vreme*grupa )
/PRINT = DESCRIPTIVE ETASQ HOMOGENEITY
/CRITERIA = ALPHA(.05)
/WSDESIGN = vreme
/DESIGN = grupa .
```

Prikažaćemo deo rezultata tog postupka.

Descriptive Statistics

	Type of class	Mean	Std. Deviation	N
fear of stats time1	maths skills	39,87	4,60	15
	confidence building	40,47	5,82	15
	Total	40,17	5,16	30
fear of stats time2	maths skills	37,67	4,51	15
	confidence building	37,33	5,88	15
	Total	37,50	5,15	30
fear of stats time3	maths skills	36,07	5,43	15
	confidence building	34,40	6,63	15
	Total	35,23	6,02	30

Box's Test of Equality of Covariance Matrices^a

Box's M	1,520
F	,224
df1	6
df2	5680
Sig.	,969

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept +GROUP
Within Subjects Design: TIME

Multivariate Tests^b

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
time	Pillai's Trace	,663	26,593 ^a	2,000	27,000	,000	,663
	Wilks' Lambda	,337	26,593 ^a	2,000	27,000	,000	,663
	Hotelling's Trace	1,970	26,593 ^a	2,000	27,000	,000	,663
	Roy's Largest Root	1,970	26,593 ^a	2,000	27,000	,000	,663
time * grupa	Pillai's Trace	,131	2,034 ^a	2,000	27,000	,150	,131
	Wilks' Lambda	,869	2,034 ^a	2,000	27,000	,150	,131
	Hotelling's Trace	,151	2,034 ^a	2,000	27,000	,150	,131
	Roy's Largest Root	,151	2,034 ^a	2,000	27,000	,150	,131

a. Exact statistic

b. Design: Intercept+group
Within Subjects Design: time

Mauchly's Test of Sphericity^b

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^a		
					Greenhouse-Geisser	Huynh-Feldt	Lower bound
TIME	,348	28,517	2	,000	,605	,640	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept + GROUP
Within Subjects Design: TIME

Levene's Test of Equality of Error Variances^a

	F	df1	df2	Sig.
fear of stats time1	,893	1	28	,353
fear of stats time2	,767	1	28	,389
fear of stats time3	,770	1	28	,388

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+GROUP
Within Subjects Design: TIME

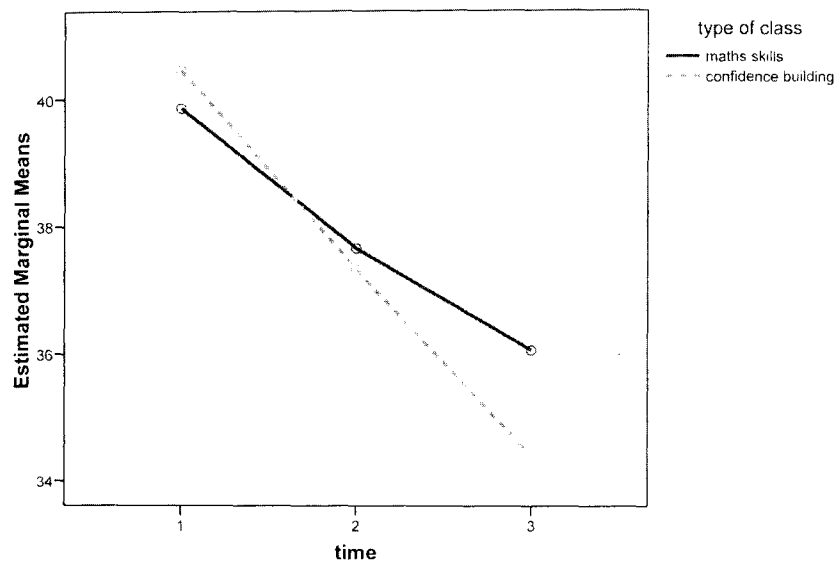
Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	127464,100	1	127464,100	1531,757	,000	,982
group	4,900	1	4,900	,059	,810	,002
Error	2330,000	28	83,214			

Estimated Marginal Means of MEASURE_1



Tumačenje rezultata kombinovane analize varijanse

Primitili ste (još jednom) da ova SPSS-ova tehnika generiše prilično mnogo složenih rezultata. Ako ste proučili prethodna poglavlja, prepoznate deo rezultata iz drugih postupaka analize varijanse. Tu su rezultati testova pretpostavljene sferičnosti, te rezultati univarijacione i multivarijacione analize varijanse (za jednu odnosno više promenljivih). Celovitoj raspravi o razlici između rezultata univarijacione i multivarijacione analize varijanse nije mesto u ovoj knjizi; u ovom poglavlju, razmotrićemo samo rezultate multivarijacione analize (više o tome videti u knjizi koju je napisao Stevens, 1996,

str. 466–9). Tumačimo samo rezultate SPSS-ove multivarijacione analize varijanse, zato što se univarijaciona zaniva na sferičnosti podataka. Sferičnost znači da je varijansa rezultata populacije za bilo koja dva uslova, jednaka toj varijansi za bilo koja druga dva uslova (što najčešće nije zadovoljeno). SPSS meri sferičnost Moklijevim testom, ali multivarijaciona analiza varijanse ne podrazumeva sferičnost. Videćete u našem primeru da pretpostavka o sferičnosti nije zadovoljena, što pokazuje vrednost **Sig.** od 0,000 u tabeli **Mauchly's Test of Sphericity**. Mada ima načina da se narušavanje te pretpostavke kompenzuje, sigurnije je pogledati rezultate multivarijacione analize varijanse date u SPSS-ovom izlazu.

Pogledajmo ključne veličine koje treba razmotriti u rezultatima.

Tabela Descriptive Statistics

U prvoj tabeli rezultata date su srednje vrednosti, standardna odstupanja i broj subjekata (**Mean**, **Std Deviation**, **N**) tri skupa rezultata. Dobro je proveriti imaju li te veličine smisla. Da li je tačan broj osoba u svakoj grupi? Imaju li smisla srednje vrednosti merenog obeležja, uzimajući u obzir upotrebljenu skalu? U gornjem primeru, videćete da su najviše srednje vrednosti straha (39,87 i 40,47) zabeležene u trenutku 1, da su niže u trenutku 2 (37,67 i 37,33) i još niže u trenutku 3 (36,07 i 34,40). Međutim, ne znamo da li su te razlike dovoljno velike da budu statistički značajne.

Pretpostavke

Pogledajte tabelu **Levene's Test of Equality of Error Variances** i proverite da li je narušena pretpostavka o homogenosti varijansi. Želeli bismo da vrednost **Sig.** bude neznačajna (veća od 0,05). U ovom slučaju, ta verovatnoća je veća od 0,05 za svaku promenljivu (0,35, 0,39, 0,39); dakle, možemo bezbedno dalje.

Zatim treba proveriti pokazatelj **Box's Test of Equality of Covariance Matrices**. Želeli bismo vrednost **Sig.** veću od 0,001. U ovom primeru, ta verovatnoća iznosi 0,97; dakle, pretpostavka je zadovoljena.

Uticaoj interakcije

Pre nego što pogledamo veličine zasebnih uticaja, prvo treba da ocenimo uticaj interakcije. Da li je vremenska promena rezultata jednaka za dve grupe (matematičke veštine/izgradnja samopouzdanja)? To nam kazuje drugi niz redova (**time*group**) tabele **Multivariate Tests**. Zanima nas pokazatelj **Wilks' Lambda** i njegova verovatnoća navedena u koloni **Sig.** Svi testovi multivarijacione analize varijanse dali su isti rezultat, ali se najčešće navodi Vilksov pokazatelj **lambda**. U ovom slučaju, uticaj interakcije nije statistički značajan (verovatnoća **Sig.** Vilksovog pokazatelja **lambda** iznosi 0,15, što je više od našeg nivoa alfa od 0,05).

Zasebni uticaji

Pošto smo pokazali da uticaj interakcije *nije* značajan, smemo da krenemo dalje i ocenimo zaseban (osnovni) uticaj svake nezavisne promenljive. Da je uticaj interakcije bio značajan, zasebne uticaje bismo tumačili veoma oprežno zato što značajna interakcija znači da se uticaj jedne promenljive menja u zavisnosti od vrednosti druge, te da uopšteni zaključci (o njihovim zasebnim uticajima) najčešće nisu tačni. Kada dobijete značajnu interakciju, tumačenje uvek zasnivajte na onom što vam kazuje dijagram.

U ovom primeru, pokazatelj Wilks' Lambda za promenljivu vreme iznosi 0,337 uz verovatnoću Sig. od 0,000 (što zapravo znači $p < 0,0005$). Pošto je dobijeni iznos verovatnoće p manji od 0,05, zaključujemo da vreme ima statistički značajan uticaj. To bi značilo da su se rezultati na testu za ispitivanje straha od statistike promenili u tri posmatrana vremenska perioda. Zaseban uticaj vremena je značajan.

Premda smo utvrdili da postoji statistički značajna razlika između tri vremenska perioda, treba da procenimo i veličinu te razlike (videti raspravu o veličinama uticaja u uvodu u peti deo knjige). Zanima nas pokazatelj parcijalno eta kvadrat tj. **Partial Eta Squared**, naveden u poslednjoj koloni tabele **Multivariate Tests**. U ovoj studiji, dobijeni eta kvadrat za vreme iznosi 0,663. Na osnovu smernica koje je predložio Koen (Cohen, 1988, str. 284–7) i koje se najčešće koriste (0,01=mali uticaj, 0,06=umeren uticaj, 0,14=velik uticaj), reklo bi se da je uticaj intervencija veoma velik.

Pošto smo istražili uticaje promenljive istih subjekata, razmotrićemo zaseban uticaj promenljive različitih subjekata (vrsta kursa: matematičke veštine/ izgradnja samopouzdanja).

Uticaj promenljive različitih subjekata

Zanimaju nas rezultati u tabeli **Tests of Between-Subjects Effects**. Pogledajte red **grupa** (to je skraćeno ime SPSS-ove promenljive za vrstu kursa). Verovatnoća Sig. iznosi 0,81. To nije manje od našeg alfa nivoa od 0,05, pa zaključujemo da zaseban uticaj grupe nije značajan. Nije bilo značajne razlike u rezultatima ispitivanja straha od statistike dveju grupa (onih koji su vežbali matematičke veštine i onih koji su bili podvrgnuti intervenciji izgradnje samopouzdanja).

U tabeli **Tests of Between-Subject Effects** data je i veličina uticaja promenljive različitih subjekata. U ovom slučaju, **Partial Eta Squared** za grupu iznosi 0,002. To je vrlo malo. Zato ne iznenađuje da promenljiva različitih subjekata nije dosegla statističku značajnost.

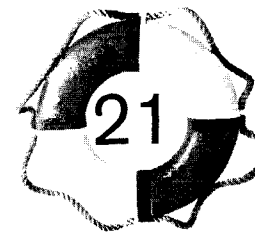
Predstavljanje rezultata kombinovane analize varijanse

Rezultati ove tehnike predstavljaju se kombinacijom načina predstavljanja analize varijanse različitih grupa (videti 19. poglavlje) i one ponovljenih merenja (videti 18. poglavlje). Uvek treba najpre navesti koliki je uticaj interakcije, zato što od toga zavisi mogućnost tumačenja zasebnih (osnovnih) uticaja nezavisnih promenljivih.

Kombinovanom analizom varijanse ocenjen je uticaj dva različita kursa (Matematičkih veština, Izgradnje samopouzdanja) na rezultate učenika ispitivanja straha od statistike, dobijene u tri vremenska perioda (pre intervencije, neposredno posle intervencije i tri meseca kasnije). Nije bilo značajne interakcije između vrste intervencije i vremena, Vilksov lambda = 0,87, $F(2, 27) = 2,03$, $p = 0,15$, parcijalno eta kvadrat = 0,13. Utvrđen je znatan zaseban (osnovni) uticaj vremena, Vilksov lambda = 0,34, $F(2, 27) = 26,59$, $p < 0,0005$, parcijalno eta kvadrat = 0,66, pri čemu je u obe grupe zabeleženo smanjenje rezultata učenika na testu dobijenih u tri navrata (videti tabelu 1). Zaseban uticaj dve vrste intervencije nije bio značajan, $F(1, 28) = 0,059$, $p = 0,81$, parcijalno eta kvadrat = 0,002, što bi značilo da su oba kursa podjednako delotvorna.

Tabela 1: Rezultati ispitivanja straha od statistike polaznika kursa matematičkih veština odnosno izgradnje samopouzdanja, mereni u tri vremenska perioda

	Matematičke veštine			Izgradnja samopouzdanja		
	n	M	SD	n	M	SD
Vremenski period						
Pre intervencije	15	39,87	4,60	15	40,47	5,82
Posle intervencije	15	37,67	4,51	15	37,33	5,88
3 meseca kasnije	15	36,07	5,43	15	34,40	6,63



Multivarijaciona analiza varijanse

U prethodnim poglavljima, analizom varijanse smo poredili grupe po *jednom obeležju* (jednoj zavisnoj promenljivoj). Međutim, ima mnogo istraživačkih situacija kada treba porediti grupe po više obeležja. To je sasvim uobičajeno u kliničkim istraživanjima, gde je težište na vrednovanju uticaja određene intervencije na razna merila ishoda (npr. anksioznost, depresiju, fizičke simptome).

Multivarijaciona analiza varijanse (MANOVA) proširenje je analize varijanse, koje se upotrebljava kada ima više od jedne zavisne promenljive. Trebalo bi da su te zavisne promenljive na neki način povezane ili da postoji neki konceptualni razlog zašto se razmatraju zajedno. MANOVA poredi te grupe i kazuje da li je verovatno da su srednje razlike u uticajima grupa na tu kombinaciju zavisnih promenljivih nastale slučajno. Da bi to uradila, MANOVA pravi novu zbirnu zavisnu promenljivu, dobijenu linearnom kombinacijom svih prvobitnih zavisnih promenljivih. Zatim analizira varijansu te nove, kombinovane zavisne promenljive. MANOVA kazuje da li postoji značajna razlika između grupa u pogledu uticaja na tu složenu zavisnu promenljivu; dobija se i univarijaciona analiza svake od zavisnih promenljivih zasebno.

Možda je neko od čitalaca pomislio: a što ne bismo prosto sproveli niz ANOVA za svaku zavisnu promenljivu zasebno? Mnogi istraživači rade upravo to. Nažalost, sprovođenjem niza ANOVA povećavate verovatnoću (rizik) da napravite grešku I vrste. (Greške prve i druge vrste razmotrene su u uvodu u peti deo knjige.) Jednostavno rečeno, to znači sledeće: što više analiza sprovedete, verovatnije je da ćete dobiti značajan rezultat, čak i kada razlika između grupa u stvari nema. MANOVA je bolja zato što uzima u obzir taj povećani rizik od greške I vrste. Međutim, to ima svoju cenu. MANOVA je mnogo složeniji skup procedura, a ima i više dodatnih pretpostavki koje moraju biti zadovoljene.

Ukoliko smatrate da je MANOVA još malo preteška za vas, imate još jednu mogućnost. Sprovedite niz ANOVA za svaku zavisnu promenljivu zasebno, ali smanjite rizik greške I vrste tako što ćete zadati strožu vrednost alfa. Jedan od načina kontrole grešaka I vrste zbog više testova jeste Bonferronijevo prilagođenje. To znači da uobičajenu vrednost alfa (najčešće 0,05) delite brojem testova koje nameravate da sprovedete. Ako imate tri nezavisne promenljive, podelili biste 0,05 sa 3 (što iznosi 0,017 nakon zaokruživanja) i tu novu verovatnoću biste koristili kao graničnu. Razlike između grupa potom smatrate statistički značajnim tek ako je dobijena verovatnoća (Sig.) manja od 0,017.

MANOVA se može upotrebljavati za jednofaktorske, dvofaktorske i višefaktorske analize varijanse (kada ima više nezavisnih promenljivih, tj. faktora), kao i u analizi kovarijanse (kada se statistički uklanja uticaj remetilačke promenljive). U ovom poglavlju, daću primer koji ilustruje jednostavnu, jednofaktorsku MANOVA analizu. Složenije analize se ne razmatraju u ovoj knjizi. Ako nameravate da koristite MANOVA analizu, zaista vam preporučujem da to ne radite dok više o tome ne pročitate i sve dobro ne proučite. Predlažem vam sledeću literaturu: Tabachnick i Fidell (2007), Hair i ostali (2006) i Stevens (1996).

Objašnjenje primera

Za prikazivanje MANOVA analize koristila sam datoteku s podacima *survey3ED.sav*, dostupnu na pratećoj Web lokaciji knjige. Studija je detaljno opisana u dodatku. U ovom primeru, istražićemo razliku između muškaraca i žena u pogledu nekoliko merila raspoloženja. To su: negativno raspoloženje (mereno na skali negativnih osećanja, Negative Affect), pozitivno raspoloženje (mereno na skali pozitivnih osećanja, Positive Affect) i subjektivno doživljen stres (mereno na skali Total Perceived Stress). Ako želite da pratite dolepisane postupke, pokrenite SPSS i otvorite datoteku *survey3ED.sav*.

Kratak pregled jednofaktorske MANOVA analize

Primer istraživačkog pitanja: Da li se muškarci i žene razlikuju po opštem raspoloženju? Da li su muškarci bolje prilagođeni od žena po svojim pozitivnim i negativnim raspoloženjima i nivou subjektivno doživljenog stresa?

Šta vam treba: Za jednofaktorsku MANOVA analizu:

- jedna kategorijska, nezavisna promenljiva (npr. pol); i
- dve ili više neprekidnih, zavisnih promenljivih (npr. negativno osećanje, pozitivno osećanje, subjektivno doživljen stres).

MANOVA se može proširiti i na dvofaktorska i višefaktorska istraživanja, kada ima više kategorijskih nezavisnih promenljivih.

Šta se postiže: Porede se dve ili više grupa po srednjim vrednostima određene kombinacije (grupe) obeležja (zavisnih promenljivih). Ispituje se nulta hipoteza da se srednje vrednosti određenog skupa zavisnih promenljivih (obeležja populacije) ne menjaju u zavisnosti od nivoa faktorske promenljive (grupisanja).

Pretpostavke: MANOVA počiva na nekoliko pretpostavki. One su detaljnije razmotrene u sledećem odeljku. Trebalo bi da pročitate i raspravu o pretpostavkama u uvodu u peti deo knjige.

1. Veličina uzorka
2. Normalnost raspodele
3. Netipične tačke
4. Linearnost
5. Homogenost regresije
6. Multikolinearnost i singularnost
7. Homogenost matrica varijanse i kovarijanse

Neparametarska alternativa: Ne postoji.

Ispitivanje pretpostavki

Pre sprovođenja MANOVA analize, proverićemo da li naši podaci zadovoljavaju pretpostavke popisane u kratkom pregledu. Neki od tih testova nisu, strogo uzev, neophodni za ovako velike uzorke, ali ću ih prikazati da biste videli potreban postupak.

1. Veličina uzorka

U svakoj ćeliji mora biti više slučajeva (opservacija) nego što ima zavisnih promenljivih. To je apsolutni minimum; bilo bi dobro da ih je još više. Kada je uzorak velik, manje su i posledice kršenja nekih drugih pretpostavki (npr. normalnosti). U ovom primeru, u svakoj ćeliji treba da ima najmanje tri slučaja (toliko ima zavisnih promenljivih). Ukupno imamo šest ćelija (dva nivoa nezavisne promenljive: muškarci/žene, i po tri zavisne promenljive za svaki nivo). U rezultatima MANOVA analize navodi se i broj slučajeva u svakoj ćeliji. U ovom primeru, u ćelijama imamo mnogo više od neophodno potrebnog broja slučajeva (videti tabelu **Descriptive Statistics** u rezultatima SPSS-ove procedure).

2. Normalnost raspodele

Ispitivanje značajnosti rezultata MANOVA analize zasnovano je na multivarijacionoj normalnoj raspodeli, ali je u praksi ova analiza prilično otporna na umereno narušavanje normalnosti (osim kada je ono posledica postojanja netipičnih tačaka). Tabachnick i Fidell (2007, str. 251) tvrde da već i 20 slučajeva po ćeliji obezbeđuje robusnost. Treba proveriti i univarijacionu

normalnost (videti 6. poglavlje) i multivarijacionu normalnost (pomoću tzv. Mahalanobisovih udaljenosti). Postupci provere normalnosti pomažu i u otkrivanju netipičnih tačaka (videti pretpostavku 3).

3. Netipične tačke

MANOVA je veoma osetljiva na netipične tačke (tj. rezultate koji se mnogo razlikuju od većine ostalih). Treba proveriti postojanje univarijacionih netipičnih tačaka (zasebno za svaku zavisnu promenljivu), a i multivarijacionih netipičnih tačaka. Multivarijacione netipične tačke su subjekti s čudnom kombinacijom dobijenih vrednosti raznih zavisnih promenljivih (npr. vrlo velike vrednosti jedne promenljive i vrlo mali druge). Univarijacione netipične tačke otkriva procedura **Explore** (videti 6. poglavlje). Postupak provere postojanja multivarijacionih netipičnih tačaka i multivarijacione normalnosti prikazaću u nastavku.

Provera multivarijacione normalnosti

Multivarijacionu normalnost ćemo proveriti kada od SPSS-a – preko menija **Regression** – zatražimo da izračuna Mahalanobisove udaljenosti. Mahalanobisova udaljenost je udaljenost određenog slučaja od centroida ostalih slučajeva; centroid je tačka koju tvore srednje vrednosti svih promenljivih (Tabachnick & Fidell, 2007). Ova analiza otkriva sve slučajeve čudnih kombinacija vrednosti *sve tri zavisne promenljive*.

Doleopisanim postupkom napravićemo novu promenljivu (nazvanu mah_2) u datoteci s podacima. Svaka osoba (subjekat) dobija određenu vrednost te promenljive, srazmernu stepenu različitosti svoje kombinacije rezultata od ostatka uzorka. Upoređićemo te Mahalanobisove udaljenosti sa određenom kritičnom vrednošću, očitano u hi-kvadrat tabeli kritičnih vrednosti. Kada rezultat dobijen za promenljivu mah_2 određenog subjekta premaši tu vrednost, smatramo ga netipičnom tačkom. MANOVA može da istrpi nekoliko netipičnih tačaka, naročito ako njihove vrednosti nisu previše ekstremne a uzorak je razumne veličine. Kada je netipičnih tačaka previše ili kada one imaju previše ekstremne vrednosti, uklonite te slučajeve iz uzorka ili transformišite te promenljive (videti Tabachnick & Fidell, 2007, str. 72).

Postupak izračunavanja Mahalanobisovih udaljenosti

1. Otvorite meni **Analyze**, izaberite u njemu stavku **Regression**, pa pritisnite **Linear**.
2. Pritisnite onu promenljivu u datoteci s podacima koja jedinstveno identifikuje svaki od ispitivanih slučajeva. (U ovom primeru, to je promenljiva ID.) Prebacite tu promenljivu u polje **Dependent**.

3. U polje **Independent** stavite neprekidne zavisne promenljive koje ste odabrali za MANOVA analizu (npr. ukupna negativna osećanja / tnegaff, ukupna pozitivna osećanja / tposaff, ukupan subjektivno doživljen stres / tpstress).
4. Pritisnite dugme **Save**. U odeljku **Distances** pritisnite **Mahalanobis**.
5. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

```
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT id
/METHOD=ENTER tposaff tnegaff tpstress
/SAVE MAHAL.
```

Evo tabele rezultata prethodnog postupka.

Residuals Statistics^a

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	239,04	317,73	276,25	13,576	432
Std. Predicted Value	-2,741	3,056	,000	1,000	432
Standard Error of Predicted Value	8,344	34,955	15,220	4,870	432
Adjusted Predicted Value	233,16	316,81	276,23	13,654	432
Residual	-297,100	448,062	,000	165,473	432
Std. Residual	-1,789	2,698	,000	,997	432
Stud. Residual	-1,805	2,711	,000	1,001	432
Deleted Residual	-302,419	452,382	,012	167,023	432
Stud. Deleted Residual	-1,810	2,732	,000	1,002	432
Mahal. Distance	,090	18,101	2,993	2,793	432
Cook's Distance	,000	,025	,002	,003	432
Centered Leverage Value	,000	,042	,007	,006	432

a. Dependent Variable: id

Pri dnu ove tabele je red **Mahal. Distance**. Pogledajte ćeliju u koloni **Maximum**. Zabeležite tu vrednost (u ovom primeru, to je 18,1). Upoređićemo taj broj sa određenom kritičnom vrednošću.

Tu kritičnu vrednost očitavate iz hi-kvadrat tabele, pri čemu je broj stepeni slobode jednak broju zavisnih promenljivih. Uzmite da je koeficijent alfa jednak 0,001. Da bih vam olakšala ceo postupak, navela sam ključne vrednosti u tabeli 21.1, za istraživanja sa do deset zavisnih promenljivih.

Pronađite kolonu s brojem zavisnih promenljivih koji imate u svojoj studiji (u ovom primeru, to je broj 3). U susednoj ćeliji desno očitajte kritičnu vrednost (u ovom slučaju, ona iznosi 16,27).

Tabela 21.1: Kritične vrednosti za vrednovanje izračunatih Mahalanobisovih udaljenosti

Broj zavisnih promenljivih	Kritična vrednost	Broj zavisnih promenljivih	Kritična vrednost	Broj zavisnih promenljivih	Kritična vrednost
2	13,82	5	20,52	8	26,13
3	16,27	6	22,46	9	27,88
4	18,47	7	24,32	10	29,59

Izvor: izdvojeno i prilagođeno iz tabele u knjizi autorki Tabachnik & Fidell (1996); prvobitni izvor: Pearson, E. S. and Harley, H. O. (Eds) (1958). Biometrika tables for statisticians (vol. 1, 2nd edn). New York: Cambridge University Press.

Uporedite maksimalnu vrednost Mahalanobisovih udaljenosti dobijenu u rezultatima (npr. 18,1) sa očitanom kritičnom vrednošću. Kada je maksimalna vrednost veća od kritične, u datoteci s podacima imate multivarijacione netipične tačke. U mom primeru, Mahalanobisova udaljenost barem jednog slučaja premašuje kritičnu vrednost od 16,27, dakle multivarijacione netipične tačke postoje. Koliko ih je i koliko se razlikuju od ostalih slučajeva, to tek treba utvrditi. Da je maksimalna vrednost Mahalanobisovih udaljenosti bila manja od kritične vrednosti, mogla sam bezbedno zaključiti da nema ekstremnih multivarijacionih netipičnih tačaka i nastaviti proveru ostalih pretpostavki.

Više o netipičnim tačkama najlakše ćemo saznati sledeći uputstva iz 5. poglavlja da sortiramo slučajeve (u opadajućem redosledu) po novoj promenljivoj (MAH_2) dodatoj na dno datoteke s podacima. Nazvana je MAH_2 zato što sam već imala MAH_1, napravljenu u poglavlju posvećenom višestrukoj regresiji. U prozoru **Data View**, slučajeve će biti poredani od najvećeg do najmanjeg iznosa MAH_2, a najveći 18,1 biće na vrhu kolone MAH_2.

U našem primeru, rezultat samo jedne osobe je premašio kritičnu vrednost. To je osoba čiji je ID=415, a rezultat 18,10. Pošto se radi o samo jednoj osobi a ni njen rezultat nije prevelik, ostaviću tu osobu u datoteci s podacima. Da je netipičnih tačaka bilo mnogo, trebalo bi transformisati tu grupu promenljivih (videti 8. poglavlje) ili ukloniti te slučajeve iz datoteke s podacima. Kada se nađete u toj situaciji, obavezno pročitajte više o navedenim mogućnostima pre nego što se odlučite za jednu od njih (Tabachnick & Fidell, 2007, str. 72).

4. Linearost

Ova pretpostavka znači da između svih parova zavisnih promenljivih treba da postoji pravolinijski odnos. To se može oceniti na više načina, od kojih je najjednostavniji nacrtati matricu dijagrama rasturanja za sve parove zavisnih promenljivih, zasebno za grupe (muškarce i žene) nezavisne promenljive.

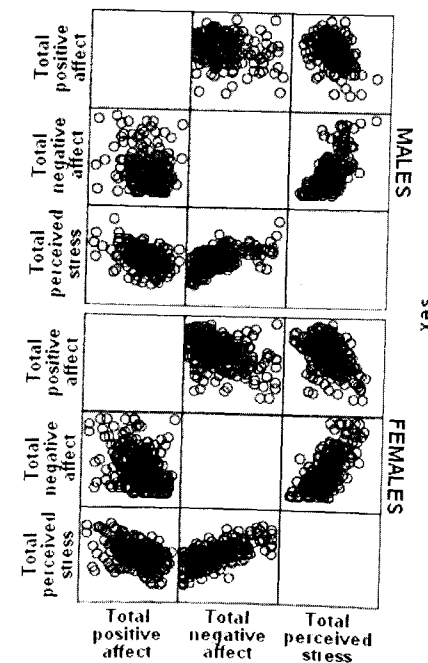
Postupak za generisanje matrice dijagrama rasturanja

1. Otvorite meni **Graphs**, izaberite u njemu stavku **Legacy Dialogs**, pa pritisnite **Scatter/Dot**.
2. Pritisnite **Matrix Scatter**. Pritisnite dugme **Define**.
3. Izaberite sve zavisne promenljive (negaff, posaff, pstress) i prebacite ih u polje **Matrix Variables**.
4. Izaberite promenljivu sex i prebacite je u polje **Rows**.
5. Pritisnite dugme **Options** i potvrdite polje **Exclude cases variable by variable**.
6. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

```
GRAPH
/SCATTERPLOT(MATRIX)=tposaff tnegaff tpstress
/PANEL ROWVAR=sex ROWOP=CROSS
/MISSING=VARIABLEWISE.
```

Evo i dobijenih dijagrama rasturanja.



Na ovim dijagramima ne primećujem znakove očigledne nelinearnosti, te ću pretpostavku o linearnosti smatrati zadovoljenom.

5. Homogenost regresije

Ova pretpostavka je važna samo ako nameravate da radite postepenu analizu (engl. *stepdown analysis*). To se radi kada iz nekog teorijskog ili konceptualnog razloga rangirate zavisne promenljive. Postupak je veoma složen i ne obrađuje se u ovoj knjizi. Koga zanima da sazna više, neka pogleda knjigu koji su napisale Tabachnick i Fidell (2007, str. 252).

6. Multikolinearnost i singularnost

MANOVA najbolje radi kada su zavisne promenljive umereno korelirane. U slučaju manjih korelacija, trebalo bi da razmislite o sprovođenju zasebne univarijacione analize varijanse za svaku zavisnu promenljivu. Kada su zavisne promenljive jako korelirane, to se naziva multikolinearnost. Do toga može doći kada je jedna od promenljivih kombinacija drugih (npr. ukupni rezultati na skali sastavljenoj od podskala, koje su takođe uključene kao zavisne promenljive). To bi bila singularnost, koju možete može izbeći tako što ćete znati kakve su vam promenljive i kako se dobijaju rezultati.

Iako za pronalaženje multikolinearnosti ima vrlo sofisticiranih načina, najjednostavnije je da pokrenete proceduru **Correlation** i proverite jačinu korelacije zavisnih promenljivih (videti 11. poglavlje). Razlog za brigu su korelacije oko 0,8 ili 0,9. Kada pronađete neku toliku korelaciju, trebalo bi da uklonite jedan od jako koreliranih parova zavisnih promenljivih ili da ih objedinite u zajedničko merilo.

7. Homogenost matrica varijanse i kovarijanse

Srećom, rezultate ispitivanja ove pretpostavke MANOVA daje u sklopu svog izlaza. Taj test se naziva **Box's M Test of Equality of Covariance Matrices** i detaljnije je razmotren u nastavku, u sklopu tumačenja rezultata.

Sprovođenje MANOVA analize

Sledi opis postupka sprovođenja jednofaktorske multivarijacione analize varijanse radi istraživanja polnih razlika u našem skupu zavisnih promenljivih (Total Negative Affect, Total Positive Affect, Total Perceived Stress). Dodavanjem još nezavisnih promenljivih, ova tehnika se može proširiti na više-faktorsku analizu. Mogla bi se proširiti i tako da obuhvati remetićke promenljive (videti analizu kovarijanse razmotrenu u 22. poglavlju).

Postupak obavljanja MANOVA analize

1. Otvorite meni **Analyze**, izaberite u njemu stavku **General Linear Model**, pa pritisnite **Multivariate**.
2. U polje **Dependent Variables** unesite sve zavisne promenljive (npr. Total negative affect, Total positive affect, Total perceived stress).
3. U polje **Fixed Factors** unesite nezavisnu promenljivu (npr. sex).
4. Pritisnite dugme **Model**. U polju **Specify Model** treba da je izabrano radio-dugme **Full factorial**.
5. Pri dnu polja **Sum of squares** treba da piše **Type III**. To je podrazumevana metoda izračunavanja zbirova kvadrata. Pritisnite **Continue**.
6. Pritisnite dugme **Options**.
 - U odeljku **Factor and Factor interactions** pritisnite odabranu nezavisnu promenljivu (npr. sex). Prebacite je u polje **Display Means for**.
 - U odeljku **Display** istog prozora, potvrdite polja **Descriptive Statistics**, **Estimates of effect size** i **Homogeneity tests**.
7. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

```
GLM
tposaff tnegaff tpstress BY sex
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/EMMEANS = TABLES(sex)
/PRINT = DESCRIPTIVE ETASQ HOMOGENEITY
/CRITERIA = ALPHA(.05)
/DESIGN = sex.
```

Prikazaćemo samo deo rezultata prethodnog postupka.

Descriptive Statistics

	SEX	Mean	Std. Deviation	N
Total Positive Affect	MALES	33,63	6,99	184
	FEMALES	33,69	7,44	248
	Total	33,66	7,24	432
Total Negative Affect	MALES	18,71	6,90	184
	FEMALES	19,98	7,18	248
	Total	19,44	7,08	432
Total perceived stress	MALES	25,79	5,41	184
	FEMALES	27,42	6,08	248
	Total	26,72	5,85	432

Box's Test of Equality of Covariance Matrices^a

Box's M	6,942
F	1,148
df1	6
df2	1074772
Sig.	,331

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

^a. Design: Intercept+SEX

Multivariate Tests^b

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillar's Trace	,987	10841,625 ^a	3,000	428,000	,000	,987
	Wilks' Lambda	,013	10841,625 ^a	3,000	428,000	,000	,987
	Hotelling's Trace	75,993	10841,625 ^a	3,000	428,000	,000	,987
	Roy's Largest Root	75,993	10841,625 ^a	3,000	428,000	,000	,987
sex	Pillar's Trace	,024	3,569 ^a	3,000	428,000	,014	,024
	Wilks' Lambda	,976	3,569 ^a	3,000	428,000	,014	,024
	Hotelling's Trace	,025	3,569 ^a	3,000	428,000	,014	,024
	Roy's Largest Root	,025	3,569 ^a	3,000	428,000	,014	,024

^a. Exact statistic

^b. Design: Intercept+sex

Levene's Test of Equality of Error Variances^a

	F	df1	df2	Sig.
Total Positive Affect	1,065	1	430	,303
Total Negative Affect	1,251	1	430	,264
Total perceived stress	2,074	1	430	,151

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

^a. Design: Intercept+SEX

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	total positive affect	,440 ^a	1	,440	,008	,927	,000
	total negative affect	172,348 ^b	1	172,348	3,456	,064	,008
	total perceived stress	281,099 ^c	1	281,099	8,342	,004	,019
Intercept	total positive affect	478633,634	1	478633,634	9108,270	,000	,955
	total negative affect	158121,903	1	158121,903	3170,979	,000	,881
	total perceived stress	299040,358	1	299040,358	8874,752	,000	,954
sex	total positive affect	,440	1	,440	,008	,927	,000
	total negative affect	172,348	1	172,348	3,456	,064	,008
	total perceived stress	281,099	1	281,099	8,342	,004	,019
Error	total positive affect	22596,218	430	52,549			
	total negative affect	21442,088	430	49,865			
	total perceived stress	14489,121	430	33,696			
Total	total positive affect	512110,000	432				
	total negative affect	184870,000	432				
	total perceived stress	323305,000	432				
Corrected Total	total positive affect	22596,657	431				
	total negative affect	21614,435	431				
	total perceived stress	14770,220	431				

^a. R Squared = ,000 (Adjusted R Squared = -,002)

^b. R Squared = ,008 (Adjusted R Squared = ,006)

SEX

Dependent Variable	SEX	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Total Positive Affect	MALES	33,625	,534	32,575	34,675
	FEMALES	33,690	,460	32,785	34,594
Total Negative Affect	MALES	18,707	,521	17,683	19,730
	FEMALES	19,984	,448	19,103	20,865
Total perceived stress	MALES	25,788	,428	24,947	26,629
	FEMALES	27,419	,369	26,695	28,144

Tumačenje rezultata MANOVA analize

Predstavljamo ključne aspekte rezultata MANOVA analize.

Tabela Descriptive Statistics

Proverite da li su tačni brojevi u ovoj tabeli, naročito da li brojevi N odgovaraju onome što znate o svom uzorku. Ti brojevi su 'broj slučajeva po ćeliji' (videti 1. pretpostavku). U svakoj ćeliji, broj slučajeva mora biti veći od broja zavisnih promenljivih. Kada u svakoj ćeliji imate preko 30 slučajeva, eventualno narušena normalnost ili jednakost varijansi neće mnogo nauditi.

Tabela Box's Test

Izlazna tabela Box's Test of Equality of Covariance Matrices kazuje da li podaci narušavaju pretpostavku homogenosti matrica varijanse i kovarijanse. Kada je vrednost **Sig. veća** od 0,001, pretpostavka *nije* narušena. Tabachnick i Fidell (2007, str. 281) upozoravaju da je Boksov pokazatelj M često prestrog kada je uzorak velik. Srećom, u našem primeru vrednost **Sig. Boksovog pokazatelja M** iznosi 0,33; dakle, tu pretpostavku nismo narušili.

Tabela Levene's Test

Pogledajte sada tabelu Levene's Test of Equality of Error Variances. U njenoj koloni **Sig.** potražite brojeve *manje* od 0,05. Svaki takav kazuje da je narušena pretpostavka o jednakosti varijanse njegove promenljive. U tekućem primeru, nijedna promenljiva nema značajnu vrednost Leveneovog pokazatelja; zato smemo da pretpostavimo da su varijanse jednake. Ako utvrdite kršenje pretpostavke o jednakosti varijansi, zadajte stroži nivo alfa za izračunavanje značajnosti univarijacionog F-testa te promenljive. Tabachnick i Fidell (2007) predlažu alfa nivo od 0,025 ili 0,01, umesto uobičajenog nivoa 0,05.

Tabela Multivariate Tests

Ovaj skup multivarijacionih testova značajnosti kazuje da li se grupe statistički značajno razlikuju po linearnoj kombinaciji zavisno promenljivih. Na raspolaganju vam je nekoliko pokazatelja (Wilks' Lambda, Hotelling's Trace, Pillai's Trace). Jedan od najčešće navodenih pokazatelja je **Wilks' Lambda**. Tabachnick i Fidell (2007) za opštu upotrebu preporučuju upravo Vilksov lambda. Međutim, kada su vam podaci loši (mali uzorak, nejednak broj slučajeva po ćelijama, narušene pretpostavke), robusniji je pokazatelj Pillai's Trace (poređenje navedenih pokazatelja videti u literaturi: Tabachnick & Fidell, 2007, str. 252). Kada imate samo dve grupe, F-testovi pokazatelja Wilks' Lambda, Hotelling's Trace i Pillai's Trace identični su.

Tabela Wilks' Lambda

Traženi pokazatelj ćete naći u *drugom* odeljku tabele **Multivariate Tests**, u redu označenim imenom nezavisne promenljive (grupisanja, u ovom slučaju SEX). Bilo bi pogrešno upotrebiti brojeve iz prvog odeljka, koji se odnosi na presečnu tačku. Pročitajte vrednost pokazatelja **Wilks' Lambda** i njegov nivo značajnosti (Sig.). Kada je nivo značajnosti *manji* od 0,05, onda treba zaključiti da postoji razlika između grupa. Ovde smo dobili iznos Vilksovog lambda od 0,976, uz značajnost 0,014. To je manje od 0,05; dakle, postoji statistički značajna razlika između muškaraca i žena u pogledu njihovog opšteg raspoloženja.

Tabela Between-subjects effects

Kada dobijete značajan rezultat prethodnog multivarijacionog testa značajnosti, tada vam je dozvoljeno da dalje istražujete svaku od svojih zavisnih promenljivih. Da li se muškarci i žene razlikuju po svim zavisnim merilima ili samo po nekim? Te informacije daće izlazna tabela **Tests of Between-Subjects Effects**. Pošto je ovde bilo sprovedeno više zasebnih analiza, preporučuje se da zadate niži nivo alfa da biste smanjili verovatnoću (rizik) da napravite grešku I vrste (tj. da rezultat proglasite značajnim kada on to nije). To se najčešće radi primenom Bonferonijevog prilagođenja. U svom najjednostavnijem obliku, to znači podeliti prvobitni nivo alfa (obično 0,05) brojem analiza koje nameravate da sprovedete (sophisticiranije verzije te formule videti u Tabachnick & Fidell, 2007, str. 270). U ovom slučaju, istražujemo tri zavisne promenljive; zato bismo 0,05 podelili sa 3, što daje novu vrednost alfa od 0,017. Svoje rezultate ćemo smatrati značajnim samo kada je pripadna verovatnoća greške prve vrste (Sig.) manja od 0,017.

U tabeli **Tests of Between-Subjects Effects**, siđite do trećeg skupa vrednosti u redu sa imenom nezavisne promenljive (u ovom slučaju, SEX). Tu su navedene sve zavisne promenljive i pripadajući univarijacioni pokazatelji F,

broj stepeni slobode df i verovatnoća greške prve vrste Sig.. Njih tumačite jednako kao što biste tumačili običnu jednofaktorsku analizu varijanse. U koloni Sig. potražite brojeve manje od 0,017 (to je naš novi, prilagođen nivo alfa). U ovom primeru, samo je za jednu zavisnu promenljivu (Total perceived stress) značajnost manja od naše granične (Sig. od 0,004). U ovoj studiji, jedina značajna razlika između muškaraca i žena bila je u subjektivno doživljenom stresu.

Veličina uticaja

Važnost (veličina) uticaja pola na subjektivno doživljen stres može se izračunati pomoću pokazatelja u poslednjoj koloni **Partial Eta Squared**. Parcijalno eta kvadrat predstavlja proporciju varijanse u zavisnoj promenljivoj (tj. u vrednostima subjektivno doživljenog stresa) koju objašnjava nezavisna promenljiva (pol/sex). U ovom primeru, ta vrednost je 0,019, što se, prema opšteprihvaćenim kriterijumima (Cohen, 1988, str. 284–7), smatra veoma malim uticajem (veličina uticaja razmotrena je u uvodu u peti deo knjige). To znači da pol objašnjava samo 1,9 procenata varijanse u rezultatima merenja subjektivno doživljenog stresa.

Poređenje srednjih vrednosti obeležja po grupama

Iako znamo da se muškarci i žene razlikuju po subjektivno doživljenom stresu, ne znamo ko je imao veće vrednosti tog obeležja. To ćemo saznati u poslednjoj izlaznoj tabeli, **SEX**. Srednja vrednost subjektivno doživljenog stresa za muškarce bila je 25,79, a za žene 27,42. Mada je ta razlika statistički značajna, ona je zapravo vrlo mala, manja od 2 jedinice na skali.

Naknadne analize

U gornjem primeru, nezavisna promenljiva je imala samo dva nivoa (dakle dve grupe: muškarci, žene). Kada nezavisna promenljiva ima tri ili više nivoa, neophodno je naknadnim univarijacionim analizama utvrditi koje su razlike značajne (da li se grupa 1 razlikuje od grupe 2, da li se grupa 2 razlikuje od grupe 3 itd.). Jedan od načina da se to uradi bila bi jednofaktorska ANOVA zavisnih promenljivih koje su bile značajne u analizi MANOVA (npr. subjektivno doživljen stres). Unutar jednofaktorske ANOVA procedure (videti 18. poglavlje), treba zahtevati naknadne testove promenljive s tri ili više nivoa. Kad god sprovedite više analiza, ne zaboravite da smanjite nivo alfa Bonferonijevim prilagođenjem.

Predstavljanje rezultata MANOVA analize

Rezultati multivarijacione analize varijanse mogu se predstaviti ovako:

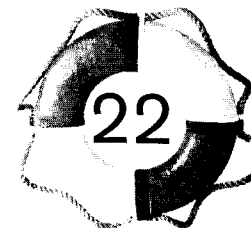
Jednofaktorskom multivarijacionom analizom varijanse istražene su polne razlike u psihološkom raspoloženju. Upotrebene su tri zavisne promenljive: pozitivna osećanja, negativna osećanja i subjektivno doživljen stres. Nezavisna promenljiva bio je pol. Preliminarnim ispitivanjem proverene su pretpostavke o normalnosti, linearnosti, univarijacionim i multivarijacionim netipičnim tačkama, homogenosti matrica varijanse-kovarijanse, i multikolinearnosti; ozbiljnije narušavanje pretpostavki nije primećeno. Utvrđena je statistički značajna razlika između muškaraca i žena u pogledu kombinacije zavisnih promenljivih, $F(3, 428) = 3,57$, $p = 0,014$; Vilksov $\lambda = 0,98$; parcijalni eta kvadrat $= 0,02$. Kada su rezultati zavisnih promenljivih razmotreni zasebno, jedina razlika koja je dosegla statističku značajnost (uz po Bonferoniju prilagođen nivo alfa od 0,017) bio je subjektivno doživljen stres, $F(1, 430) = 8,34$, $p = 0,004$, parcijalni eta kvadrat $= 0,02$. Pregledom prosečnih vrednosti rezultata utvrđeno je da su kod žena zabeležen neznatno veći nivoi subjektivno doživljenog stresa ($M = 27,42$, $SD = 6,08$) nego kod muškaraca ($M = 25,79$, $SD = 5,41$).

Dodatna vežba

Zdravstvo

Datoteka s podacima: **sleep3ED.sav**. Pojediniosti o datoteci videti u dodatku.

1. Sprovedite jednofaktorsku MANOVA analizu kako biste videli da li postoje polne razlike za svaku stavku sa skale pospanosti i pridruženih osećanja (Sleepiness and Associated Sensations Scale). Kao zavisne promenljive, trebaće vam *fatigue* (zamor), *lethargy* (bezvoljnost), *tired* (umor), *sleepy* (pospanost), *energy* (energija).



Analiza kovarijanse

Analiza kovarijanse (ANCOVA) jeste proširenje analize varijanse (razmotrene u 18. poglavlju), koje služi za istraživanja razlika između grupa uz statističko uklanjanje (kontrolu) uticaja jedne dodatne (neprekidne) remetilacke promenljive. Na tu dodatnu promenljivu (kovarijat) sumnjate da možda utiče na vrednosti zavisne promenljive. SPSS regresionim postupcima uklanja varijansu u zavisnoj promenljivoj koju prouzrokuju kovarijati, i potom na korigovane rezultate primenjuje uobičajene tehnike analize varijanse. Uklanjanjem uticaja dodatnih promenljivih, ANCOVA povećava moć ili osetljivost F-testa, tj. povećava verovatnoću da ćete otkriti razliku između grupa, ako ona postoji.

Analiza kovarijanse može biti upotrebljena u sklopu jednofaktorskih, dvofaktorskih i multivarijacionih ANOVA tehnika. U ovom poglavlju, razmotreni su postupci analize kovarijanse pridruženi sledećim analizama:

- jednofaktorska ANOVA različitih grupa (jedna nezavisna promenljiva, jedna zavisna promenljiva); i
- dvofaktorska ANOVA različitih grupa (dve nezavisne promenljive, jedna zavisna promenljiva).

U narednim odeljcima ukratko ću vas upoznati s pozadinom ove tehnike i njenim osnovnim pretpostavkama koje treba ispitati; zatim sledi nekoliko urađenih primera. Pre nego što pokrenete te analize, predlažem da pročitate uvodne odeljke u ovom poglavlju i ponovo prelistate odgovarajuća poglavlja u kojima je opisana ANOVA (18. i 19.). Multivarijaciona ANCOVA ovde neće biti ilustrovana. Koga to zanima, predlažem mu knjigu čije su autorka Tabachnick i Fidell (2007).

Čemu služi ANCOVA

ANCOVA se može upotrebiti za poređenje dve grupe koje se testiraju pre i posle (npr. poređenje uticaja dve različite intervencije, mereno u svakoj grupi pre i posle intervencija). Rezultati na testu pre intervencije tretiraju se kao kovarijat za kontrolu, tj. statističko uklanjanje prethodno postojećih razlika između grupa. Zato je ANCOVA vrlo korisna kada su uzorci mali ili kada su

uticaji razlika mali do umereni (videti raspravu o veličinama uticaja u uvodu u peti deo knjige). U takvim okolnostima, veoma uobičajenim u društvenim naukama, Stevens (1996) preporučuje da se upotrebe dva ili tri pažljivo odabrana kovarijata (obeležja), kako bi se smanjila varijansa greške i povećala verovatnoća otkrivanja značajne razlike između grupa, ukoliko postoji.

ANCOVA je podesna i kada subjekti nisu slučajno dodeljeni grupama, već su upotrebljene postojeće grupe, npr. razredi učenika. Pošto se te grupe mogu razlikovati po više atributa (a ne samo po onom koji istražujete), ANCOVA se može upotrebiti da bi se smanjile neke od tih razlika. Dobro odabrani kovarijati smanjuju remetilački uticaj razlika između grupa. To svakako nije idealna situacija, pošto nije moguće ukloniti uticaj svih mogućih razlika; međutim, njime se ta sistematska greška smanjuje. Sprovođenje analize kovarijanse netaknutih ili postojećih grupa predmet je sporenja autora u ovoj oblasti. Kada se nađete u toj situaciji, bilo bi dobro da pročitate više o tome. Neka od tih pitanja sumarno je obradio Stevens (1996, str. 324–7).

Izbor prikladnih kovarijata

ANCOVA se može upotrebiti za istovremenu kontrolu više kovarijata. Međutim, njih treba pažljivo odabrati (videti Stevens, 1996, str. 320; i Tabachnick & Fidell, 2007, str. 203, 211). Prilikom identifikacije mogućih kovarijata, morate dobro poznavati teoriju i prethodna istraživanja sprovedena u toj oblasti. Promenljive odabrane kao kovarijati moraju biti neprekidne, pouzdano izmerene (videti 9. poglavlje) i jako korelirane sa odabranom zavisnom promenljivom. Idealno bi bilo da izaberete mali skup kovarijata koji su međusobno umereno korelirani, tako da svaki od njih jedinstveno doprinosi objašnjenju varijanse. Kovarijat mora biti izmeren *pre* tretmana odnosno eksperimentalne obrade kako bi se sprečilo da tretman utiče i na vrednosti kovarijata.

Alternative analizi kovarijanse

ANCOVA ima više ograničenja i temeljnih pretpostavki zbog kojih projekat istraživanja ili podaci mogu biti neprikladni za analizu kovarijanse. Tabachnick & Fidell (2007, str. 221) i Stevens (1996, str. 327) navode više alternative analizi kovarijanse. Bilo bi dobro da istražite te alternative i u konkretnim okolnostima procenite koji je put najbolji.

Osnovne pretpostavke analize kovarijanse

Pored pretpostavki uobičajenih za svaku analizu varijanse i razmotrenih u uvodu u peti deo knjige, ANCOVA ima više svojih posebnih pretpostavki i problematičnih pitanja. U ovom poglavlju ćemo razmatrati samo ključne pretpostavke na kojima se zasniva ANCOVA. Tabachnick i Fidell (2007, str.

200) detaljno i dobro obrađuju ovu temu, uključujući i nejednake uzorkovne netipične tačke, multikolinearnost, normalnost raspodele, homogenost varijanse, linearnost, homogenost regresije i pouzdanost kovarijata. (Već je sam ovaj spisak statističkog žargona dovoljan da čoveka preplaši!).

Uticaj tretmana na merenje kovarijata

Istraživanje treba projektovati tako da se kovarijat meri *pre* tretmana odnosno eksperimentalne obrade kako bi se izbegao uticaj tretmana na kovarijat. Ukoliko tretman utiče na kovarijat, ta promena će biti korelirana promenom u zavisnoj promenljivoj. Kada ANCOVA ukloni uticaj kovarijata, ukloniće i deo uticaja tretmana i tako smanjiti verovatnoću dobijanja značajnog rezultata.

Pouzdanost kovarijata

ANCOVA pretpostavlja da su kovarijati izmereni bez greške, što je potpuno nerealna pretpostavka u dobrom delu društvenih istraživanja. Samo deo obeležja čiji uticaj treba ukloniti, recimo starost ispitanika, mogu se izmeriti prilično pouzdano; ostala obeležja (ona koja se mere pomoću skala) ne zadovoljavaju tu pretpostavku. Pouzdanost alatki za merenje može se poboljšati na nekoliko načina:

- Potražite dobre i dobro validirane skale i upitnike. Proverite da li mere ono što vi mislite da mere (nije dovoljno pročitati samo naziv skale – pročitajte i prateći priručnik, pregledajte sve stavke) i da li su prikladni za vaš uzorak.
- Proverite unutrašnju saglasnost (to je jedna vrsta pouzdanosti) skale tako što ćete izračunati njen Kronsbaheov pokazatelj alfa (videti 9. poglavlje). Alfa treba da je iznad 0,7 (bilo bi poželjno iznad 0,8) da bi se skala mogla smatrati pouzdanom.
- Ako ste sami morali da pišete pitanja, ona treba da su jasna, prikladna i jednoznačna. Pitanja i skale odgovora treba da su prikladni za sve grupe. Pre sprovođenja glavnog istraživanja, obavezno ispitajte svoja pitanja pomoću probnog (pilot) testa.
- Ukoliko upotrebljavate bilo koju vrstu opreme ili merne instrumentacije, obavezno proverite radi li ispravno i da li je propisno kalibrisana. Proverite da li je osoba koja rukuje tom opremom ovlašćena i obučena da je upotrebljava.
- Ako u istraživanju učestvuju i drugi ljudi kao opservatori ili ocenjivači ponašanja, proverite da li su obučeni i da li svi primenjuju iste kriterijume. Međusobnu usaglašenost ocenjivača proverite probnim (pilot) testiranjem pre nego što pređete na glavno istraživanje.

Međukorelacije kovarijata

Obeležja odabrana za kovarijate ne bi trebalo da su jako međusobno korelirana. Idealno bi bilo da jako koreliraju sa zavisnom promenljivom, ali ne i međusobno. Prilikom izbora prikladnih kovarijata, oslonite se na postojeću teoriju i prethodna istraživanja u toj oblasti. Preliminarnim analizama korelacije (videti 11. poglavlje) istražite jačinu veze između predloženih kovarijata. Ako utvrdite da su kovarijati koje ste nameravali da upotrebite jako međusobno korelirani (npr. $r=0,80$), uklonite jedan ili više njih (videti Stevens, 1996, str. 320). Svaki od odabranih kovarijata treba da da svoj samostalan doprinos; preklapanje među njima ne doprinosi smanjenju varijanse greške.

Linearna veza zavisne promenljive i kovarijata

ANCOVA pretpostavlja da je veza između zavisne promenljive i svakog kovarijata linearna (pravolinijska). Isto važi i za sve parove kovarijata, koliko god da ih ima. Narušavanje ove pretpostavke obično smanjuje moć (osetljivost) testa. Ne zaboravite, jedan od razloga za uključivanje kovarijata bilo je povećanje moći analize varijanse.

Linearnost se može proveravati vizuelno na dijagramima rasturanja, ali se to mora uraditi zasebno za svaku grupu (tj. svaki nivo nezavisne promenljive). Ukoliko otkrijete krivolinijsku vezu, to se može popraviti transformacijom tog kovarijata (videti 8. poglavlje) ili njegovim izbacivanjem iz analize. Često je lakše iz analize izbaciti loše kovarijate, zato što je transformisane promenljive teško protumačiti.

Homogenost nagiba regresije

Ovo impresivno zvuči, a znači samo to da bi veza između kovarijata i zavisne promenljive trebalo da bude jednaka za sve grupe. Na dijagramu regresije to se vidi kao jednakost nagiba regresione linije za svaku grupu. Nejednaki nagibi kazuju da postoji interakcija između tog kovarijata i tretmana. Ako interakcija postoji, onda rezultati analize kovarijanse zavaravaju, pa nju uopšte ne bi trebalo sprovoditi (videti Stevens, 1996, str. 323, 331; Tabachnick & Fidell, 2007, str. 202). Postupak provere ove pretpostavke dat je u primerima u nastavku poglavlja.

Jednofaktorska ANCOVA

U ovom odeljku, proći ćemo korak po korak kroz postupak jednofaktorske analize kovarijanse. Jednofaktorska ANCOVA obuhvata jednu nezavisnu, kategorijsku promenljivu (sa dva ili više nivoa ili uslova), jednu zavisnu neprekidnu promenljivu i jedan ili više neprekidnih kovarijata. Ova tehnika se često upotrebljava za vrednovanje uticaja intervencije ili eksperimentalne obrade, zato što uklanja uticaj rezultata ispitivanja dobijenih pre intervencije.

Objašnjenje primera

Za prikazivanje jednofaktorske analize kovarijanse upotrebiću datoteku s podacima `experim3ED.sav`, dostupnu na pratećoj Web lokaciji knjige (videti str. xi). To su podaci iz fiktivne studije o ispitivanju uticaja dve različite intervencije koje studentima pomažu da se izbore sa svojom anksioznošću povodom predstojećeg kursa iz statistike. (Sve pojedinosti studije videti u dodatku.) Podeljeni u dve jednake grupe, studenti su odgovorili na nekoliko pitanja, pri čemu je jedna skala merila njihov strah od statistike. Potom je jedna grupa pohađala kurs matematičkih veština, a druga je učestvovala u programu izgradnje samopouzdanja. Nakon tih intervencija, svi su ponovo odgovarali na ista pitanja koja su dobili pre programa. Spomenuta datoteka sa izmišljenim podacima dostupna je na Web lokaciji ove knjige. Ako želite da pratite dolepisane postupke, pokrenite SPSS i otvorite datoteku `experim3ED.sav`.

U ovom primeru istražujemo uticaj kursa matematičkih veština (grupa 1) i kursa izgradnje samopouzdanja (grupa 2) na rezultate učesnika u ispitivanju straha od statistike, pri čemu uklanjamo uticaj rezultata dobijenih na tom testu pre pohađanja kurseva. Slede pojedinosti o imenima upotrebljenih promenljivih i njihovi opisi u datoteci s podacima:

Ime datoteke: `experim3ED.sav`

Promenljive:

- Vrsta poduke (grupa) 1=Matematičke veštine, 2=Izgradnja samopouzdanja
- Rezultati na testu kojim se ispituje strah od statistike, popunjavanom u trenutku 1 (Fost1): ukupni rezultati na istom testu popunjavanom pre intervencije. Rezultati poprimaju vrednosti u opsegu od 20 do 60. Veće vrednosti pokazuju veći strah od statistike.
- Rezultati na testu kojim se ispituje strah od statistike, popunjavanom u trenutku 2 (Fost2): ukupni rezultati na istom testu popunjavanom neposredno posle intervencije.
- Rezultati na testu kojim se ispituje strah od statistike popunjavanom u trenutku 3 (Fost3): ukupni rezultati na istom testu, popunjavanom 3 meseca posle intervencije.

Kratak pregled jednofaktorske analize kovarijanse

Primer istraživačkog pitanja: Postoji li značajna razlika u rezultatima ispitivanja straha od statistike grupe koja je pohađala kurs matematičkih veština (grupa 1) i grupe sa kursa izgradnje samopouzdanja (grupa 2), kada se ukloni uticaj njihovih početnih odgovora na istom testu?

Šta vam treba: Najmanje tri promenljive:

- jedna kategorijska nezavisna promenljiva sa dva ili više nivoa (grupa 1/ grupa 2);
- jedna neprekidna zavisna promenljiva (rezultati učesnika pri ispitivanju straha od statistike mereni u trenutku 2);
- jedan ili više neprekidnih kovarijata (rezultati učesnika pri ispitivanju straha od statistike mereni u trenutku 1).

Šta se postiže: ANCOVA kazuje da li se rezultati ispitivanja dve grupe na strah od statistike sprovedenom u trenutku 2 značajno razlikuju nakon uklanjanja uticaja rezultata (odgovora) studenata na testu sprovedenom pre intervencija (kurseva obuke).

Pretpostavke: Sve uobičajene pretpostavke jednofaktorske analize varijanse (videti uvod u peti deo knjige). Njih treba prve proveriti. Dodatne pretpostavke analize kovarijanse (videti njihov opis dat u prethodnom delu poglavlja):

1. kovarijat se meri pre intervencije odnosno eksperimentalne obrade;
2. kovarijat se meri bez greške (ili barem onoliko pouzdano koliko je to uopšte moguće);
3. kovarijati nisu jako međusobno korelirani;
4. između zavisne promenljive i kovarijata postoji linearan odnos za sve grupe (linearnost);
5. odnos između kovarijata i zavisne promenljive jednak je za sve grupe (homogenost nagiba regresione linije).

Neparametarska alternativa: Ne postoji.

Provera pretpostavki

Pre nego što počnete proveru posebnih pretpostavki na kojima se temelji ANCOVA, morate proveriti zadovoljenost pretpostavki na kojima se zasniva obična jednofaktorska analiza varijanse (normalnost raspodele, homogenost varijanse). Pre nego što nastavite sa ovim poglavljem, trebalo bi da pročitate uvod u peti deo knjige i 18. poglavlje. Tu građu ovde neću ponavljati; razmotrićemo samo dodatne pretpostavke analize kovarijanse, navedene u kratkom pregledu.

Pretpostavka 1: merenje kovarijata

Ova pretpostavka nalaže da se kovarijat izmeri pre početka tretmana odnosno eksperimentalne obrade. Zadovoljenost ove pretpostavke ne proverava se statistički, nego tako morate projektovati istraživanje i sve istraživačke postupke. Eto zašto je važno planirati studiju tek onda kada dobro upoznate statističke tehnike koje nameravate da upotrebite.

Pretpostavka 2: pouzdanost kovarijata

I ova pretpostavka nameće odgovarajući način projektovanja istraživanja, a znači da treba izabrati najpouzdanije dostupne merne alatke. Ako u istraživanju nameravate da koristite neko psihometrijsko merilo ili skalu, proverite njenu unutrašnju saglasnost (međusobnu saglasnost stavki koje je sačinjavaju) izračunavanjem Kronbahovog pokazatelja alfa (pomoću SPSS-ove procedure **Reliability**, videti 9. poglavlje). Pošto sam podatke za ilustraciju ove tehnike izmislila, ne mogu da proverim pouzdanost ispitivanja straha od statistike. Da su to stvarni podaci, proverila bih da li je Kronbahov alfa barem 0,70 (bilo bi poželjno da je 0,80).

Pretpostavka 3: korelacije između kovarijata

Kada upotrebljavate više kovarijata, proverite da nisu prejak međusobno korelirani ($r=0,8$ i više). Tome služi SPSS-ova procedura **Correlation**, detaljno opisana u 11. poglavlju. Pošto ja imam samo jedan kovarijat, u ovom primeru to ne moram da radim.

Pretpostavka 4: linearnost

Pretpostavka o linearnom odnosu između zavisne promenljive i kovarijata, za sve grupe, može se proveriti na više načina. U sledećem postupku pokazacu vam jedan brz način.

Postupak za proveru linearnosti za sve grupe

Da biste pratili ovaj primer, otvorite datoteku s podacima **experim3ED.sav**.

Prvi korak

1. Otvorite meni **Graphs**, izaberite u njemu stavku **Legacy Dialogs**, pa pritisnite **Scatter/Dot**.
2. Pritisnite opciju **Simple Scatter**. Pritisnite dugme **Define**.
3. U polje **Y axis** stavite zavisnu promenljivu (u ovom slučaju, strah od statistike izmeren u trenutku 2, fost2).
4. U polje **X axis** stavite kovarijat (npr. strah od statistike izmeren u trenutku 1, fost1).
5. Pritisnite nezavisnu promenljivu (npr. grupa/group) i stavite je u polje **Set Markers by**.
6. Pritisnite **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

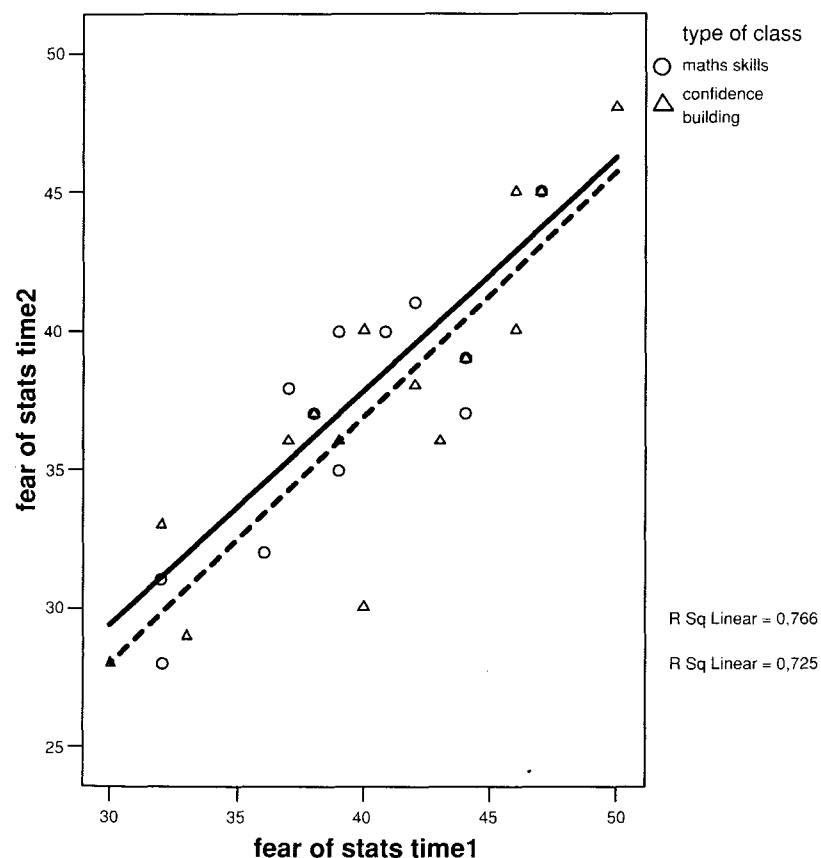
Postupak generiše ovu komandu:

```
GRAPH
/SCATTERPLOT(BIVAR)=fost1 WITH fost2 BY group
/MISSING=LISTWISE
```

Drugi korak

1. Kada je dijagram rasturanja nacrtan i prikazan u prvom planu, dvaput ga pritisnite da biste otvorili prozor **Chart Editor**.
2. Otvorite meni **Elements** i u njemu izaberite **Add Fit line at Subgroups**. Na dijagramu će se prikazati linije koje najbolje odgovaraju podacima svake od grupa. Ja sam ih na donjoj slici izmenila da bi bile jasnije (liniju jedne grupe nacrtala sam isprekidano), a izmenila sam i markere (videti uputstva za modifikovanje dijagrama u 7. poglavlju).
3. Otvorite meni **File** i u njemu izaberite **Close**.

Evo kako izgleda moj dijagram rasturanja.



Kako vam izgleda raspodela rezultata svake grupe? Da li vam odnos između zavisne promenljive i kovarijata, u svakoj grupi, liči na linearan (pravolinijski)? Najgore bi bilo da se pokaže jasan krivolinijski odnos. U gornjem primeru odnos je jasno linearan, pa pretpostavka linearnosti nije

narušena. Kada otkrijete krivolinijski odnos, uzmite neki drugi kovarijat ili pokušajte da transformišete postojeći (videti 8. poglavlje), pa ponovo nacrtajte dijagram rasturanja i proverite da li je poboljšan.

Pretpostavka 5: homogenost regresionih nagiba

Poslednja pretpostavka (homogenost regresionih nagiba) odnosi se na vezu između kovarijata i zavisne promenljive u svim grupama. Proveravate da nema interakcije između kovarijata i tretmana odnosno eksperimentalne obrade. Zadovoljenost ove pretpostavke može se proveriti na više načina.

Vizuelno, može se pogledati dijagram rasturanja zavisne promenljive u funkciji kovarijata, nacrtan prilikom ispitivanja pretpostavke 4 (videti gore). Imaju li te dve linije (koje u ovoj studiji odgovaraju dvema grupama) slične nagibe? U gornjem primeru linije su vrlo slične, pa izgleda da ova pretpostavka nije narušena. Da su linije bile primetno različito usmerene, zaključili bismo da postoji interakcija između kovarijata i tretmana (kao što se vidi po razlici među grupama). To bi značilo da je ova pretpostavka narušena.

Ova pretpostavka se može proveriti i statistički umesto grafički. To znači da bismo izračunali interakciju tretmana i kovarijata i njenu statističku značajnost. Kada je interakcija značajna na nivou alfa od 0,05, pretpostavka je narušena. Sada ćemo izračunati tu interakciju.

Postupak provere homogenosti regresionih nagiba

1. Otvorite meni **Analyze**, izaberite u njemu stavku **General Linear Model**, pa pritisnite **Univariate**.
2. U polje **Dependent Variables** prebacite zavisnu promenljivu (npr. strah od statistike izmeren u trenutku 2, skraćeno fost2).
3. U polje **Fixed Factor** prebacite nezavisnu tj. promenljivu grupisanja (npr. vrstu poduke, skraćeno group).
4. U polje **Covariate** prebacite kovarijat (npr. strah od statistike izmeren u trenutku 1, skraćeno fost1).
5. Pritisnite dugme **Model**. Pritisnite opciju **Custom**.
6. U polju **Build Terms** treba da je izabrana opcija **Interaction**.
7. Pritisnite nezavisnu promenljivu (group) i zatim je pritiskom na dugme sa strelicom prebacite u polje **Model**.
8. Pritisnite kovarijat (fost1) i zatim ga pritiskom na dugme sa strelicom prebacite u polje **Model**.
9. Ponovo pritisnite nezavisnu promenljivu (group) na levoj strani (u odeljku **Factors and Covariates**). Kada je ta promenljiva izabrana, pritisnite i zadržite taster Ctrl na tastaturi, i zatim pritisnite kovarijat (fost1). Pritiskom na dugme sa strelicom prebacite ga u polje **Model** na desnoj strani.

10. Sada bi u polju **Model** trebalo da piše ime nezavisne promenljive (group); ime kovarijata (Fost1); i jos jedan red u obliku: kovarijat * nezavisna promenljiva (group*fost1).

11. To poslednje je interakcija koju proveravamo.

12. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

```
UNIANOVA fost2 BY group WITH fost1
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE /CRITERIA = ALPHA(.05)
/DESIGN = group fost1 fost1*group .
```

Evo i rezultata tog postupka.

Tests of Between-Subjects Effects

Dependent Variable : fear of stats time2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	577,689 ^a	3	192,563	26,102	,000
Intercept	3,877	1	3,877	,525	,475
GROUP	,846	1	,846	,115	,738
FOST1	538,706	1	538,706	73,022	,000
GROUP*FOST1	,409	1	,409	,055	,816
Error	191,811	26	7,377		
Total	42957,000	30			
Corrected Total	769,500	29			

a. R Squared = ,751 (Adjusted R Squared = ,722)

U gornjoj izlaznoj tabeli, zanima nas samo značajnost interakcije, tj. red **Group*Fost1**. Ostale rezultate slobodno zanemarite. Kada je pripadni broj u koloni **Sig.** manji od ili jednak 0,05, interakcija je statistički značajna, a pretpostavka narušena. Naravno, to u ovoj situaciji *ne* želimo. Odgovarala bi nam verovatnoća *veća* od 0,05. U gornjem primeru, **Sig.** (tj. verovatnoća) iznosi 0,816, što je mnogo iznad granične vrednosti 0,05. Pretpostavka o homogenosti regresionih nagiba nije narušena. To podržava naš raniji zaključak, dobijen pregledom dijagrama rasturanja za svaku grupu.

Pošto smo time završili proveru pretpostavki, nastavljamo analizu kovarijanse u istraživanju razlika između grupa.

Postupak za jednofaktorsku analizu kovarijanse

1. Otvorite meni **Analyze**, izaberite u njemu stavku **General Linear Model**, pa pritisnite **Univariate**.
2. U polje **Dependent Variables** stavite zavisnu promenljivu (npr. strah od statistike izmeren u trenutku 2, skraćeno fost2).
3. U polje **Fixed Factor** stavite nezavisnu promenljivu tj. promenljivu grupisanja (npr. vrsta poduke, skraćeno group).
4. U polje **Covariate** stavite odabrani kovarijat (npr. strah od statistike izmeren u trenutku 1, skraćeno fost1).
5. Pritisnite dugme **Model**. U odeljku **Specify Model** pritisnite opciju **Full Factorial**. Pritisnite **Continue**.
6. Pritisnite dugme **Options**.
7. Pri vrhu, u odeljku **Estimated Marginal Means** pritisnite nezavisnu promenljivu (group).
8. Pritiskom na strelicu prebacite je u polje **Display Means for**. Tako ste zadali da se izračuna srednja vrednost zavisne promenljive u svakoj grupi, uz uklanjanje uticaja kovarijata.
9. U donjem delu okvira za dijalog **Options** izaberite opcije **Descriptive statistics, Estimates of effect size i Homogeneity tests**.
10. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

```
UNIANOVA
fost2 BY group WITH fost1
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/EMMEANS = TABLES(group) WITH(fost1=MEAN)
/PRINT = DESCRIPTIVE ETASQ HOMOGENEITY
/CRITERIA = ALPHA(.05)
/DESIGN = fost1 group .
```

U nastavku su rezultati tog postupka.

Univariate Analysis of Variance

Between-Subjects Factors

	Value Label	N
type of class	1 maths skills	15
	2 confidence building	15

Descriptive Statistics

Dependent Variable: fear of stats time2

type of class	Mean	Std. Deviation	N
1 maths skills	37,67	4,515	15
2 confidence building	37,33	5,876	15
Total	37,50	5,151	30

Levene's Test of Equality of Error Variances^a

Dependent Variable: fear of stats time2

F	df1	df2	Sig.
,141	1	28	,710

Tests the null hypothesis that the error variance of the dependent variable is equal across groups

a. Design: Intercept+fast1+group

Tests of Between-Subjects Effects

Dependent Variable: fear of stats time2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	577,279 ^a	2	288,640	40,543	,000	,750
Intercept	3,510	1	3,510	,493	,489	,018
fast1	576,446	1	576,446	80,970	,000	,750
group	5,434	1	5,434	,763	,390	,027
Error	192,221	27	7,119			
Total	42957,000	30				
Corrected Total	769,500	29				

a. R Squared = ,750 (Adjusted R Squared = ,732)

Estimated Marginal Means

type of class

Dependent Variable: fear of stats time2

type of class	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 maths skills	37,926 ^a	,690	36,512	39,341
2 confidence building	37,074 ^a	,690	35,659	38,488

a. Covariates appearing in the model are evaluated at the following values: fear of stats time1 = 40,17

Tumačenje rezultata jednofaktorske analize kovarijanse

ANCOVA daje mnogo rezultata, pa ćemo ih tumačiti korak po korak.

- Proverite da li su pojedinosti u tabeli **Descriptive Statistics** tačne.
- U tabeli **Levene's Test of Equality of Error Variances** proverite da li je narušena pretpostavka o jednakosti varijanse. Trebala bi nam verovatnoća **Sig.** veća od 0,05. Kada je ta vrednost manja od 0,05 (stoga i značajna), to znači da varijanse nisu jednake i da je pretpostavka narušena. U ovom slučaju pretpostavka nije narušena, zato što **Sig.** iznosi 0,71, što je mnogo više od graničnih 0,05.
- Glavni rezultati analize kovarijanse dati su u tabeli **Test of Between-Subjects Effects**. Istražujemo da li se grupe značajno razlikuju po vrednosti zavisne promenljive (npr. straha od statistike izmerenog u trenutku 2). Pronađite red sa imenom nezavisne promenljive (group) i pročitajte broj u odgovarajućoj ćeliji kolone **Sig.** Ako je broj u toj ćeliji *manji* od 0,05 (odnosno od zadatog nivoa verovatnoće alfa), grupe se razlikuju značajno. U ovom primeru, ta verovatnoća iznosi 0,39, što je više od 0,05; dakle, naš rezultat *nije* značajan. Kada se ukloni uticaj rezultata ispitivanja straha od statistike obavljenog pre intervencije, ne postoji značajna razlika u rezultatima osoba u grupi koja je učila matematičke veštine i osoba iz grupe za izgradnju samopouzdanja.
- Trebalo bi razmotriti i veličinu uticaja, koju pokazuje parcijalni eta kvadrat (**Partial Eta Squared**; objašnjenje veličine uticaja videti u uvodu u peti deo knjige). U ovom slučaju, ta vrednost iznosi samo 0,027 (što je, prema Koenovim smernicama iz 1988, mali uticaj). Taj broj kazuje i koji deo varijanse u zavisnoj promenljivoj objašnjava nezavisna promenljiva. Množenjem sa 100 pretvorite taj parcijalni eta kvadrat u procentualnu vrednost (pomerite decimalni zarez za dva mesta udesno). U ovom primeru, uspeali smo da objasnimo samo 2,7 procenata varijanse.
- U prikazanoj tabeli može se očitati još i uticaj kovarijata. Pronađite red sa imenom kovarijata (npr. strah od statistike izmeren u trenutku 1, tj. FOST1) i pročitajte broj u odgovarajućoj ćeliji kolone **Sig.** On kazuje da li postoji značajna veza između kovarijata i zavisne promenljive, kada se ukloni uticaj nezavisne promenljive (group). U redu Fast1 (našeg kovarijata), verovatnoća **Sig.** iznosi 0,000 (što zapravo znači manje od 0,0005). To je manje od 0,05, pa je naš kovarijat značajan. U stvari, on je objasnio 75 procenata varijanse u zavisnoj promenljivoj (parcijalni eta kvadrat od 0,75 pomnožen sa 100).
- U poslednjoj tabeli rezultata (**Estimated marginal means**) ANCOVA daje korigovane srednje vrednosti zavisne promenljive za svaku grupu. 'Korigovane' znači da je uticaj kovarijata statistički uklonjen.

Predstavljanje rezultata jednofaktorske analize kovarijanse

Rezultate jednofaktorske analize kovarijanse mogli biste predstaviti ovako:

Jednofaktorskom analizom kovarijanse upoređena je delotvornost dvaju intervencija namenjenih smanjivanju straha ispitanika od statistike. Nezavisna promenljiva je bila vrsta intervencije (poduka u matematičkim veštinama, odnosno izgradnja samopouzdanja), a zavisnu čine rezultati ispitivanja straha od statistike obavljenog nakon završetka intervencija. Kao kovarijat u analizi upotrebljeni su rezultati ispitivanja straha od statistike sprovedenog pre intervencija. Preliminarnim proverama je utvrđeno da nisu narušene pretpostavke o normalnosti, linearnosti, homogenosti varijansi, homogenosti regresionih nagiba i pouzdanosti merenja kovarijata. Nakon statističkog uklanjanja uticaja rezultata ispitivanja straha od statistike obavljenog pre intervencija, utvrđeno je da nema značajne razlike između te dve grupe u rezultatima ispitivanja straha od statistike sprovedenog posle intervencija, $F(1, 27) = 0,76$, $p = 0,39$, parcijalni eta kvadrat = 0,03. Utvrđena je jaka veza između rezultata ispitivanja straha od statistike sprovedenog pre i posle intervencija, što pokazuje parcijalni eta kvadrat od 0,75.

U prikaz rezultata ove analize trebalo bi uključiti i tabelu sa srednjim vrednostima rezultata grupa. Ako su zavisna promenljiva (trenutak 2) i kovarijat (trenutak 1) mereni na istoj skali, trebalo bi navesti srednje vrednosti u trenutku 1 i trenutku 2. Njih će izračunati procedura **Descriptives** (videti 6. poglavlje).

Ukoliko je kovarijat meren na drugačijoj skali, navedite nekorigovanu srednju vrednost (i standardno odstupanje) i korigovanu srednju vrednost (i standardno odstupanje) obe grupe. Nekorigovanu srednju vrednost pročitajte u tabeli **Descriptive Statistics**. Korigovanu srednju vrednost (iz koje je uticaj kovarijata statistički uklonjen) pročitajte u tabeli **Estimated Marginal Means**. Bilo bi dobro da navedete i broj slučajeva u svakoj grupi.

Dvofaktorska ANCOVA

U ovom odeljku, proći ćemo korak po korak kroz postupak dvofaktorske analize kovarijanse. Dvofaktorska ANCOVA obuhvata dve nezavisne, kategorijske promenljive (sa po dva ili više nivoa ili uslova), jednu zavisnu neprekidnu promenljivu, i jedan ili više neprekidnih kovarijata. Pre nego što nastavite sa ovim poglavljem, trebalo bi da dobro proučite standardnu dvofaktorsku analizu varijanse i njene pretpostavke. Pročitajte to u 19. poglavlju knjige.

Primer koji će ovde biti predstavljen proširenje je onoga datog u prethodnom odeljku o jednofaktorskoj analizi kovarijanse. Tamo sam istraživala koja intervencija (poduka u matematičkim veštinama ili izgradnja samopouzdanja) delotvornije smanjuje strah studenata od statistike. Utvrdila sam da između tih grupa nema značajne razlike.

Recimo da sam prilikom daljeg proučavanja literature u toj oblasti naišla na istraživanja iz kojih sledi da postoji razlika u reagovanju muškaraca i žena na razne intervencije. Ponekad se u literaturi ta dodatna promenljiva (npr. pol) naziva *moderatorom*, zato što moderira (modifikuje) uticaj druge nezavisne promenljive. Moderatorske promenljive su često promenljive individualne različitosti, obeležja individua koja utiču na njihov odgovor na eksperimentalnu obradu ili uslov tretmana.

Kada jednofaktorska ANCOVA da neznačajan rezultat, trebalo bi razmotriti moguće moderatorske promenljive. Istraživanja postaju veoma zanimljiva kada istraživač naleti na (ili sistematski istražuje) moderatorske promenljive koje doprinose objašnjavanju zašto neki istraživači dobijaju statistički značajne rezultate, a drugi ne. U sopstvenom istraživanju uvek uzmite u obzir faktore kao što su pol i starost, zato što oni mogu znatno da utiču na rezultate. Studije sprovedene na uzorku mladih studenata univerziteta često nije moguće uopštiti tako da važe za šire (i starije) uzorke populacije. Istraživanja sprovedena na uzorku muškaraca katkada daju sasvim drugačije rezultate kada se ponove na uzorku žena. Pouka je da prilikom projektovanja istraživanja treba razmotriti moguće moderatorske promenljive i, gde je prikladno, uključiti ih u studiju.

Objašnjenje primera

U dolenavedenom primeru koristim iste podatke kao u prethodnom odeljku, ali uz jednu dodatnu nezavisnu promenljivu (pol/gender: 1=muškarci/male, 2=žene/female). Tako dobijam priliku da proširim analizu i saznam da li pol kao moderatorska promenljiva utiče na delotvornost ta dva programa. Zanimam me da li postoji značajan uticaj interakcije između pola učesnika i vrste intervencije. Možda muškarci imaju više koristi od poduke u matematičkim veštinama, a žene od programa izgradnje samopouzdanja.

Upozorenje: uzorak upotrebljen za ilustraciju ovog primera zaista je vrlo mali, naročito kada se podeli na kategorije nezavisnih promenljivih (pol učesnika i vrsta intervencije). Ukoliko ovu tehniku upotrebljavate u sopstvenom istraživanju, uložite ozbiljan napor da dobijete sveukupno mnogo veću datoteku s podacima, sa značajnim brojem slučajeva u svakoj kategoriji svih nezavisne promenljive.

Ako želite da pratite opisane postupke, pokrenite SPSS i otvorite datoteku **experim3ED.sav**, dostupnu na pratećoj Web lokaciji knjige.

Kratak pregled dvofaktorske analize kovarijanse

Primer istraživačkog pitanja: Da li pol utiče na delotvornost dva programa namenjena smanjenju straha ispitanika od statistike? Ima li razlike u rezultatima ispitivanja straha od statistike muškaraca i žena, dobijenim posle intervencije (tj. pohađanja programa obuke), po tome kako su reagovali na program učenja matematičkih veština i program izgradnje samopouzdanja?

Šta vam treba: Najmanje četiri promenljive:

- dve kategorijske nezavisne promenljive sa po dva ili više nivoa (pol: M/Ž; grupa: matematičke veštine/izgradnja samopouzdanja);
- jedna neprekidna zavisna promenljiva (rezultati učesnika na testu kojim se ispituje strah od statistike, mereni u trenutku 2);
- jedan ili više neprekidnih kovarijata (rezultati učesnika na testu kojim se ispituje strah od statistike, mereni u trenutku 1).

Šta se postiže: ANCOVA uklanja uticaj kovarijata i zatim obavlja uobičajenu dvofaktorsku analizu varijanse kojom saznajemo sledeće:

- postoji li značajan zaseban uticaj prve nezavisne promenljive (grupa);
- postoji li značajan zaseban uticaj druge nezavisne promenljive (pol); i
- postoji li značajna interakcija između njih.

Pretpostavke: Sve uobičajene pretpostavke dvofaktorske analize varijanse (npr. normalnost raspodele, homogenost varijanse). Njih treba prve proveriti (videti uvod u peti deo knjige).

Dodatne pretpostavke analize kovarijanse: Videti njihov opis i postupke za njihovu proveru date u prethodnom delu poglavlja, posvećenom jedno-faktorskoj analizi kovarijanse.

Neparametarska alternativa: Ne postoji.

Postupak za dvofaktorsku analizu kovarijanse

1. Otvorite meni **Analyze**, izaberite u njemu stavku **General Linear Model**, pa pritisnite **Univariate**.
2. U polje **Dependent Variables** prebacite zavisnu promenljivu (npr. strah od statistike izmeren u trenutku 2, skraćeno fost2).
3. Pritisnite dve nezavisne promenljive tj. promenljive grupisanja (npr. pol/sex; grupa/group). Prebacite ih u polje **Fixed Factor**.
4. U polje **Covariate** stavite odabrane kovarijate (npr. strah od statistike izmeren u trenutku 1, skraćeno fost1).

5. Pritisnite dugme **Model**. U odeljku **Specify Model** pritisnite opciju **Full Factorial**. Pritisnite **Continue**.

6. Pritisnite dugme **Options**.

- Pri vrhu, u odeljku **Estimated Marginal Means** pritisnite dve odabrane nezavisne promenljive (npr. pol/sex, grupa/group). Pritiskom na strelicu prebacite ih u polje **Display means for**.
- Pritisnite dodatni član interakcije (npr. group*sex). Prebacite ga u to polje. Tako ste zadali da se izračuna srednja vrednost zavisne promenljive u svakoj grupi, iz koje je uklonjen uticaj tog kovarijata
- U donjem delu okvira za dijalog **Options** izaberite opcije **Descriptive statistics, Estimates of effect size i Homogeneity tests**. Pritisnite **Continue**.

7. Pritisnite dugme **Plots**.

- Izaberite prvu nezavisnu promenljivu (npr. group) i prebacite je u polje **Horizontal**. Ta promenljiva će biti prikazana duž horizontalne ose dijagrama.
- Pritisnite drugu nezavisnu promenljivu (npr. sex) i prebacite je u polje **Separate Lines**. Ta promenljiva će biti prikazana različitim linijama za svaku grupu.
- Pritisnite **Add**.

8. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Postupak generiše ovu komandu:

UNIANOVA

```
fost2 BY group sex WITH fost1
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/PLOT = PROFILE( group*sex )
/EMMEANS = TABLES(group) WITH(fost1=MEAN)
/EMMEANS = TABLES(sex) WITH(fost1=MEAN)
/EMMEANS = TABLES(sex*group) WITH(fost1=MEAN)
/PRINT = DESCRIPTIVE ETASQ HOMOGENEITY
/CRITERIA = ALPHA(.05)
/DESIGN = fost1 group sex group*sex .
```

Evo i rezultata opisanog postupka.

Univariate Analysis of Variance

Descriptive Statistics

Dependent Variable: fear of stats time2

type of class	sex	Mean	Std. Deviation	N
1 maths skills	1 male	37,25	5,497	8
	2 female	38,14	3,436	7
	Total	37,67	4,515	15
2 confidence building	1 male	40,57	5,563	7
	2 female	34,50	4,781	8
	Total	37,33	5,876	15
Total	1 male	38,80	5,596	15
	2 female	36,20	4,475	15
	Total	37,50	5,151	30

Levene's Test of Equality of Error Variances^a

Dependent Variable: fear of stats time2

F	df1	df2	Sig.
2,204	3	26	,112

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+fostr1+group+sex+group * sex

Tests of Between-Subjects Effects

Dependent Variable: fear of stats time2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	686,728 ^a	4	171,682	51,854	,000	,892
Intercept	4,137	1	4,137	1,250	,274	,048
fostr1	545,299	1	545,299	164,698	,000	,868
group	4,739	1	4,739	1,431	,243	,054
sex	4,202	1	4,202	1,269	,271	,048
group * sex	104,966	1	104,966	31,703	,000	,559
Error	82,772	25	3,311			
Total	42957,000	30				
Corrected Total	769,500	29				

a. R Squared = ,892 (Adjusted R Squared = ,875)

Estimated Marginal Means

1. type of class

Dependent Variable: fear of stats time2

type of class	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 maths skills	38,024 ^a	,472	37,053	38,996
2 confidence building	37,226 ^a	,471	36,255	38,197

a. Covariates appearing in the model are evaluated at the following values: fear of stats time1 = 40,17

2. sex

Dependent Variable: fear of stats time2

sex	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 male	38,009 ^a	,476	37,028	38,989
2 female	37,242 ^a	,476	36,261	38,222

a. Covariates appearing in the model are evaluated at the following values: fear of stats time1 = 40,17.

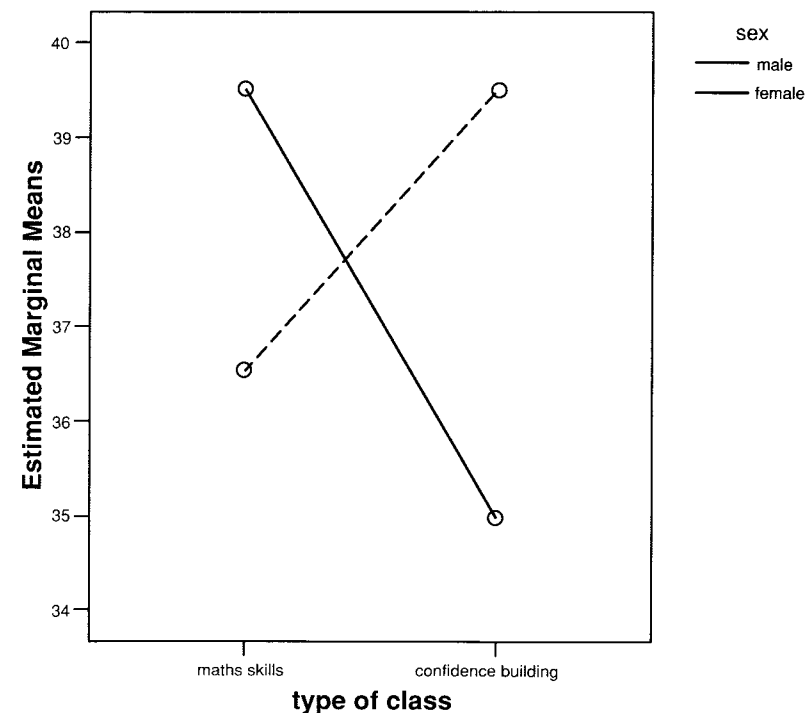
3. sex * type of class

Dependent Variable: fear of stats time2

sex	type of class	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1 male	1 maths skills	36,532 ^a	,646	35,202	37,862
	2 confidence building	39,485 ^a	,693	38,058	40,912
2 female	1 maths skills	39,517 ^a	,696	38,083	40,950
	2 confidence building	34,966 ^a	,644	33,639	36,294

a. Covariates appearing in the model are evaluated at the following values: fear of stats time1 = 40,17.

Estimated Marginal Means of fear of stats time2



Tumačenje rezultata dvofaktorske analize kovarijanse

ANCOVA daje mnogo rezultata, pa ćemo ih tumačiti korak po korak.

- Proverite da li su brojevi u tabeli **Descriptive Statistics** tačni (npr. broj slučajeva u svakoj grupi, srednje vrednosti rezultata grupa).
- U tabeli **Levene's Test of Equality of Error Variances** proverite da li je narušena pretpostavka o jednakosti varijanse. Trebala bi nam verovatnoća **Sig.** veća od 0,05. Kada je ta vrednost manja od 0,05 (stoga i značajna), to znači da varijanse nisu jednake i da je pretpostavka narušena. U ovom slučaju pretpostavka nije narušena, zato što **Sig.** iznosi 0,11, što je više od graničnih 0,05.
- Glavni rezultati analize kovarijanse dati su u tabeli **Test of Between-Subjects Effects**. Zanima nas da li postoji značajan zaseban uticaj ijedne nezavisne promenljive (grupa ili pola) i da li je interakcija između njih značajna. Od najvećeg interesa je interakcija, pa ćemo nju prvu proveriti. Kada je interakcija značajna, dva zasebna uticaja nisu važna, pošto se uticaj jedne nezavisne promenljive menja u zavisnosti od nivoa druge. Pronađite red sa imenom interakcije (u ovom slučaju, $group * sex$). i pročitajte broj u odgovarajućoj ćeliji kolone **Sig.** Ako je broj u toj ćeliji *manji* od 0,05 (odnosno od zadatog nivoa verovatnoće alfa), interakcija je značajna. U ovom primeru, ta verovatnoća iznosi 0,000 (to treba shvatiti kao manje od 0,0005), što je mnogo manje i od 0,05 i od 0,01; dakle, naš rezultat je statistički veoma značajan. Ovako značajan uticaj interakcije navodi na pomisao da muškarci drugačije reaguju na ta dva programa od žena. Zasebni uticaji obeležja **Group** i **Sex** nisu statistički značajni ($Group: p=0,24$; $Sex: p=0,27$). Ne može se reći da je jedna intervencija bolja od druge, pošto je jedna bolja za muškarce, a druga za žene. Ne može se reći da „intervencija više koristi muškarcima“, pošto to važi samo za jednu intervenciju, a za onu drugu važi suprotno.
- Treba razmotriti i veličinu uticaja, čiji je pokazatelj parcijalni eta kvadrat (objašnjenje veličine uticaja videti u uvodu u peti deo knjige). U ovom slučaju, ta veličina iznosi 0,56 (što je veliki uticaj, prema Koenovim smernicama iz 1988). Taj broj kazuje i koji deo varijanse u zavisnoj promenljivoj objašnjava nezavisna promenljiva. Množenjem sa 100 pretvorite taj parcijalni eta kvadrat u procentualnu vrednost (pomerite decimalni zarez za dva mesta udesno). U ovom primeru, uspeći smo da objasnimo čak 56 procenata varijanse.

- U toj tabeli se može očitati još i uticaj kovarijata. Pronađite red sa imenom kovarijata (npr. strah od statistike izmeren u trenutku 1, tj. FOST1) i pročitajte broj u odgovarajućoj ćeliji kolone **Sig.** On kazuje da li postoji značajna veza između kovarijata i zavisne promenljive, kada se ukloni uticaj nezavisne promenljive (group). U redu Fost1 (našeg kovarijata) piše da verovatnoća **Sig.** iznosi 0,000 (što zapravo znači manje od 0,0005). To je manje od 0,05, pa je naš kovarijat značajan. U stvari, on je objasnio čak 87 procenata varijanse u zavisnoj promenljivoj (parcijalni eta kvadrat od 0,87 pomnožen sa 100).
- U odeljku **Estimated marginal means** ANCOVA daje korigovane srednje vrednosti zavisne promenljive za svaku grupu, date za svaku nezavisnu promenljivu zasebno i potom zajedno. 'Korigovane' znači da je uticaj kovarijata statistički uklonjen. Pošto je uticaj interakcije značajan, najzanimljivija je treća (poslednja) tabela rezultata zajedničkog uticaja dve nezavisne promenljive, naslovljena **3. sex* type of class**.
- Kao neobavezan dodatak, zatražila sam da se nacrtaj dijagram korigovanih srednjih vrednosti rezultata ispitivanja straha od statistike, zasebno za muškarce i žene i za obe intervencije. (Ne brinite ukoliko se vaš dijagram razlikuje od prikazanog: naknadno sam modifikovala dijagram da bi se lakše čitao kada se odštampa. Uputstva za modifikovanje dijagrama videti u 7. poglavlju.) Na ovom dijagramu jasno se vidi da postoji interakcija između te dve nezavisne promenljive. Za muškarce, izmeren strah je bio najmanji nakon poduke iz matematičkih veština. Kod žena, strah je bio najmanji nakon izgradnje njihovog samopouzdanja. Rezultati jasno ukazuju na to da muškarci i žene različito reaguju na sprovedene intervencije i da prilikom projektovanja intervencija, npr. programa obuke, morate voditi računa o polu učesnika.

Upozorenje: ne uzbuđujte se mnogo kada dobijete značajan rezultat. Morate stalno imati u vidu cilj analize. Primera radi, ovi rezultati ne kazuju da je kurs iz matematičkih veština koristio *svim* muškarcima, niti da je izgradnja samopouzdanja koristila *svim* ženama. Ne zaboravite, poredimo prosečne rezultate grupe kao celine. Uprosečavanjem rezultata za celu grupu, neizbežno gubimo deo informacija o pojedincima. Prilikom vrednovanja programa treba razmotriti dodatne analize da bi se istražilo koliko je ljudi kojima su koristili, u odnosu na broj onih kojima su 'štetili'.

Premda se nadamo da intervencije namenjene da pomognu ljudima nijednom učesniku neće naškoditi, ponekad se to ipak događa. U oblasti upravljanja stresom, neka istraživanja su pokazala neočekivano povećanje nivoa anksioznosti učesnika nakon određenih vrsta tretmana (npr. treninga opuštanja/relaksacije). U takvoj situaciji treba utvrditi zašto se to desilo nekim osobama i prepoznati dodatne moderatorske promenljive.

Predstavljanje rezultata dvofaktorske analize kovarijanse

Rezultati ove analize mogu se predstaviti ovako:

Analizom 2 sa 2 kovarijanse različitih grupa upoređena je delotvornost dva programa obuke namenjena smanjivanju straha muškaraca i žena od statistike. Nezavisne promenljive bile su vrsta intervencije (poduka u matematičkim veštinama, odnosno izgradnja samopouzdanja) i pol učesnika. Zavisnu promenljivu čine rezultati ispitivanja straha od statistike dobijeni nakon završetka programa intervencija (u trenutku 2). Kao kovarijat u analizi, upotrebljen za uklanjanje individualnih razlika, korišćeni su rezultati ispitivanja straha od statistike dobijeni pre intervencija.

Preliminarnim proverama je utvrđeno da nisu narušene pretpostavke o normalnosti, linearnosti, homogenosti varijansi, homogenosti regresionih nagiba i pouzdanosti merenja kovarijata. Nakon statističkog uklanjanja uticaja rezultata ispitivanja straha od statistike dobijenih pre intervencija (u trenutku 1), utvrđeno je da postoji značajan uticaj interakcije, $F(1, 25) = 31,7$, $p < 0,0005$, i da je taj uticaj velik (parcijalni eta kvadrat = 0,56). Nijedan od zasebnih uticaja nije bio značajan, program: $F(1, 25) = 1,43$, $p = 0,24$; pol: $F(1, 25) = 1,27$, $p = 0,27$. Ovaj rezultat navodi na pomisao da muškarci drugačije reaguju na te dve vrste intervencija od žena. Strah muškaraca od statistike više se smanjio nakon učestvovanja u programu savladavanja matematičkih veština, dok je ženama više koristio program izgradnje samopouzdanja.

U prikaz rezultata ove analize trebalo bi uključiti i tabelu sa srednjim vrednostima grupa. Ako su zavisna promenljiva (trenutak 2) i kovarijat (trenutak 1) mereni na istoj skali, trebalo bi navesti srednje vrednosti u trenutku 1 i trenutku 2. Njih će izračunati procedura **Descriptives** (videti 6. poglavlje). Ukoliko je kovarijat meren na drugačijoj skali, navedite nekorigovanu srednju vrednost (i standardno odstupanje) i korigovanu srednju vrednost (i standardnu grešku) za obe grupe. Nekorigovanu srednju vrednost pročitajte u tabeli **Descriptive Statistics**. Korigovanu srednju vrednost (iz koje je uticaj kovarijata statistički uklonjen) pročitajte u tabeli **Estimated Marginal Means**. Bilo bi dobro da navedete i broj slučajeva u svakoj grupi.

Dodatak

Pojedinosti o datotekama s podacima

Ovaj dodatak sadrži informacije o šest datoteka s podacima, dostupnih na pratećoj Web lokaciji knjige (pojediniosti o tome videti na str. xi):

1. survey3ED.sav
2. error3ED.sav
3. experim3ED.sav
4. depress3ED.sav
5. sleep3ED.sav
6. staffsurvey3ED.sav

Prve četiri datoteke služe za praćenje postupaka opisanih u raznim poglavljima ove knjige. Poslednje dve (**sleep3ED.sav** i **staffsurvey3ED.sav**) dodatne su datoteke za vežbanje, s podacima iz raznih naučnih disciplina. **Sleep3ED.sav** sadrži podatke zdravstvene i medicinske prirode, dok **staffsurvey3ED.sav** sadrži podatke relevantne za poslovno okruženje. Na kraju nekih poglavlja date su preporučene vežbe s podacima iz tih datoteka.

Da biste mogli koristiti te datoteke s podacima, morate posetiti navedenu Web lokaciju i, prateći uputstva na ekranu, preuzeti ih na čvrsti disk, CD ili USB. Zatim bi trebalo da pokrenete SPSS i otvorite onu datoteku s podacima koju nameravate da koristite. Takve datoteke mogu da se otvore samo u SPSS-u. U ovom dodatku je za svaku datoteku dat šifarnik, sa svim pojedinostima o promenljivama i pratećim uputstvima za šifrovanje.

Survey3ED.sav

Ovo su stvarni podaci, sažeti iz jedne studije koju su sproveli moji postdiplomci psihologije obrazovanja. Studija je trebalo da istraži činioce koji utiču na psihološko prilagođenje i opšte raspoloženje ispitanika. Pripadajuća anketa je obuhvatala razne validirane skale za merenje konstrukata, za koje bi se po obimnoj literaturi o stresu i njegovom prevazilaženju moglo smatrati da utiču na subjektan doživljaj stresa. Pomoću tih skala mere se samopoštovanje, optimizam, subjektivno doživljen stepen samokontrole, tj. upravljanja sopstvenim životom, subjektivno doživljen stres, pozitivna i

negativna osećanja i zadovoljstvo životom. Bila je tu i jedna skala za merenje sklonosti ljudi da se predstave u povoljnom ili društveno poželjnom svetlu. Anketa je podeljena pripadnicima opšte populacije u Melburnu i okolnim okruzima Australije. Konačno je dobijen uzorak sa 439 osoba, od toga 42 procenta muškaraca i 58 procenata žena, starosti od 18 do 82 godine (srednja vrednost=37,4).

Error3ED.sav

Podaci u ovoj datoteci dobijeni su modifikovanjem datoteke **survey3ED.sav**, tako što su u nju namerno ubačene greške koje treba otkriti postupcima objašnjenim u 5. poglavlju. Više o njenim promenljivama itd. videti u opisu datoteke **survey3ED.sav**.

Experim3ED.sav

Ovo je skup izmišljenih podataka, napravljen da bi se dobilo nešto prikladno za ilustraciju statističkih tehnika kao što su t-test za ponovljena merenja i jednofaktorska ANOVA ponovljenih merenja. Taj skup podataka upućuje na fiktivno istraživanje uticaja dve različite intervencije koje pomažu studentima da se izbore sa sopstvenom anksioznošću u vezi s predstojećim kursom iz statistike. Podeljeni u dve jednake grupe, studenti su (u trenutku 1) popunili nekoliko testova. To su bili testovi za ispitivanje straha od statistike (Fear of Statistics), samopouzdanja u vezi s polaganjem ispita iz statistike (Confidence in Coping with Statistics) i depresije (Depression). Potom je jedna od tih grupa (grupa 1) pohađala višečasovni kurs matematičkih veština; druga grupa (grupa 2) učestvovala je u programu izgradnje samopouzdanja u polaganje ispita iz statistike. Nakon završetka oba programa (u trenutku 2) opet su popunjavali iste testove kao pre programa, i ponovo tri meseca kasnije (u trenutku 3). Bile su izmerene i njihove performanse na ispitu iz statistike.

Depress3ED.sav

Ova datoteka je priložena za prikazivanje određenih tehnika u 16. poglavlju. Obuhvata nekoliko ključnih promenljivih iz jedne stvarne studije, koju je jedan od mojih postdiplomaca sproveo istražujući činioce koji utiču na opšte raspoloženje prvorotki. Sadrži rezultate dobijene na više različitih psiholoških skala projektovanih za merenje depresije (pojediniosti videti u 16. poglavlju, o meri podudaranja kapa).

Sleep3ED.sav

Stvarni podaci, sažeti iz studije sprovedene radi istraživanja rasprostranjenosti i uticaja problema sa spavanjem na razne aspekte života. Univerzitet-sko osoblje iz Melburna u Australiji popunilo je upitnik s pitanjima o broju sati noćnog sna koje uspevaju da ostvare, problemima sa spavanjem (da li

teško zaspu, da li se lako bude) i uticaju tih problema na razne aspekte njihovih života (rad, vožnju, odnose s životnim partnerima). Uzorak je imao 271 ispitanika (55 procenata žena, 45 procenata muškaraca) starosti od 18 do 84 godine (srednja vrednost=44 god).

Staffsurvey3ED.sav

Stvarni podaci, sažeti iz studije sprovedene radi istraživanja stepena zadovoljstva osoblja jedne obrazovne ustanove sa ograncima u više mesta po Australiji. Osoblje je popunilo kratak anoniman upitnik (dat u nastavku dodatka), u kojem su dali svoja mišljenja o raznim aspektima organizacije i tretmana kojem su bili podvrgnuti kao zaposleni.

Deo A: materijali za survey3ED.sav

Pojediniosti o skalama koje obuhvata survey3ED.sav

Skale su navedene redom kojim se pojavljuju u anketi.

Skala	Referenca
Life Orientation Test Skala optimizma ili životne orijentacije (šest stavki)	Scheier, M.F. & Carver, C.S. (1985). Optimism, coping and health: An assessment and implications of generalized outcome expectancies. <i>Health Psychology</i> , 4, 219–47. Scheier, M.F., Carver, C.S. & Bridges, M.W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery and self-esteem): A re-evaluation of the Life Orientation Test. <i>Journal of Personality and Social Psychology</i> , 67, 6, 1063–78.
Mastery Scale Skala subjektivnog osećaja upravljanja sopstvenim životom (sedam stavki)	Pearlin, L. & Schooler, C. (1978). The structure of coping. <i>Journal of Health and Social Behavior</i> , 19, 2–21.
Positive and Negative Affect Scale (PANAS) Skala pozitivnih i negativnih osećanja (dvadeset stavki)	Watson, D., Clark, L.A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. <i>Journal of Personality and Social Psychology</i> , 54, 1063–70.
Satisfaction with Life Scale Skala zadovoljstva životom (pet stavki)	Diener, E., Emmons, R.A., Larson, R.J. & Griffin, S. (1985). The Satisfaction with Life scale. <i>Journal of Personality Assessment</i> , 49, 71–6.
Perceived Stress Scale Skala subjektivno doživljenog stresa (deset stavki)	Cohen, S., Kamarck, T. & Mermelstein, R. (1983). A global measure of perceived stress. <i>Journal of Health and Social Behavior</i> , 24, 385–96.
Self-esteem Scale Skala samopoštovanja (deset stavki)	Rosenberg, M. (1965). <i>Society and the adolescent selfimage</i> . Princeton, NJ: Princeton University Press.

Social Desirability Scale

Skala za merenje sklonosti ljudi da se predstave u društveno poželjnom svetlu (deset stavki)

Crowne, D.P. & Marlowe, P. (1960).

A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349–54.

Strahan, R. & Gerbasi, K. (1972). Short, homogeneous version of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 28, 191–3.

Perceived Control of Internal States Scale (PCOISS)

Skala subjektivno doživljene samokontrole unutrašnjih stanja (osamnaest stavki)

Pallant, J. (2000). Development and validation of a scale to measure perceived control of internal states. *Journal of Personality Assessment*, 75, 2, 308–37.

Šifarnik za survey3ED.sav

Puno ime promenljive	Ime promenljive u SPSS-u	Uputstvo za šifrovanje
Identifikacioni broj	id	identifikacioni broj subjekta
Pol subjekta	sex	1=muškarci, 2=žene
Starost subjekta	age	u godinama
Bračno stanje	marital	1=nevenčan(a), 2=stalan partner, 3=živi s partnerom, 4=prvi put venčan(a), 5=ponovo venčan(a), 6=živi odvojeno od supružnika, 7=razveden(a), 8=udovac/udovica
Dete	child	1=da, 2=ne
Najviša završena škola	educ	1=osnovna škola 2=deo srednje škole 3=cela srednja škola 4=viša škola ili dodatna obuka 5=dodiplomske studije na univerzitetu 6=postdiplomske studije na univerzitetu
Glavni izvor stresa	source	1= radno mesto, 2=supružnik iil partner, 3=ljubavni odnosi, 4=deca, 5=porodica, 6=zdravstveno stanje/ bolest, 7=život uopšte, 8=finansijsko stanje, 9=nedostatak slobodnog vremena
Pušač	smoke	1=da, 2=ne
Broj popušenih cigareta sedmično	smokenum	broj popušenih cigareta sedmično
Skala optimizma	op1 do op6	1=nimalo se ne slažem, 5=potpuno se slažem
Skala subjektivnog osećaja upravljanja sopstvenim životom	mast1 do mast7	1=nimalo se ne slažem, 4=potpuno se slažem

Skala pozitivnih i negativnih osećanja (PANAS)	pn1 do pn20	1=vrlo malo, 5=izuzetno mnogo
Skala zadovoljstva životom	lifsat1 do lifsat5	1=nimalo se ne slažem, 7=potpuno se slažem
Skala subjektivno doživljenog stresa	pss1 do pss10	1=nikada, 5=vrlo često
Skala samopouzdanja	sest1 do sest10	1=nimalo se ne slažem, 4=potpuno se slažem
Marlowe-Crowne skala za merenje sklonosti ljudi da se predstave u društveno poželjnom svetlu	m1 do m10	1=tačno, 2=netačno
Skala subjektivno doživljene samokontrole unutrašnjih stanja (PCOISS)	pc1 do pc18	1=nimalo se ne slažem, 5=potpuno se slažem

Ukupni rezultati na skalama koje su deo datoteke survey3ED.sav

Puno ime promenljive	Ime promenljive u SPSS-u	Uputstvo za šifrovanje
Ukupan optimizam	toptim	obrnuti stavke op2, op4, op6 sabrati vrednosti od op1 do op6 raspon od 6 do 30
Ukupan subjektivni osećaj upravljanja sopstvenim životom	tmast	obrnuti stavke mast1, mast3, mast4, mast6, mast7 sabrati vrednosti od mast1 do mast7 raspon od 7 do 28
Ukupna pozitivna osećanja	tposaff	sabrati vrednosti pn1, pn4, pn6, pn7, pn9, pn12, pn13, pn15, pn17, pn18 raspon od 10 do 50
Ukupna negativna osećanja	tnegaff	sabrati vrednosti pn2, pn3, pn5, pn8, pn10, pn11, pn14, pn16, pn19, pn20 raspon od 10 do 50
Ukupno zadovoljstvo životom	tlifesat	sabrati sve vrednosti od lifsat1 do lifsat5 raspon 5 do 35
Ukupan subjektivno doživljen stres	tpstress	obrnuti stavke psp4, pss5, pss7, pss8 sabrati sve vrednosti pss1 do pss10 raspon od 10 do 50
Ukupno samopouzdanje	tslfest	obrnuti stavke sest3, sest5, sest7, sest9, sest10 sabrati sve vrednosti od sest1 do sest10 raspon od 10 do 40
Ukupna sklonosti osobe da se predstavi u društveno poželjnom svetlu	tmarlow	obrnuti stavke m6 do m10 (rešifrovati na tačno=1, netačno=0) sabrati sve vrednosti od m1 do m10 raspon od 0 do 10
Ukupna subjektivno doživljena samokontrola sopstvenih unutrašnjih stanja	tpcoiss	obrnuti stavke pc1, pc2, pc7, pc11, pc15, pc16 sabrati sve vrednosti od pc1 do pc18 raspon od 18 do 90
Nove obrazovne kategorije	educ2	zbog malog broja slučajeva u svakoj od tih grupa, kategorije Osnovna škola i Deo srednje škole rešifrovane su u jednu grupu: 1=osnovna ili deo srednje škole 2=cela srednja škola 3=viša škola ili dodatna obuka 4=dodiplomske studije na univerzitetu 5=postdiplomske studije na univerzitetu
Starosna grupa, 3 kategorije	agegp3	1=18-29 god, 2=30-44 god, 3=45+ god.

Kopija upitnika upotrebljenog u datoteci survey3ED.sav

Na sledećim stranicama navela sam samo deo upitnika upotrebljenog za prikupljanje podataka u datoteci survey3ED.sav. Na prvoj stranici su demografska pitanja, zatim test optimizma / životne orijentacije (6 stavki) i skala pozitivnih i negativnih osećanja (20 stavki). U spisku u prethodnom delu dodatka date su sve pojedinosti o referencama iz kojih su preuzete skale korišćene u upitniku.

Upitnik

- Pol: ☐ muški
☐ ženski (označite odgovarajuće polje)
- Starost: _____ (u godinama)
- Vaš bračni status: (označite odgovarajuće polje)
☐ 1=nevenčan(a)
☐ 2=stalni partner
☐ 3=živi s partnerom
☐ 4=prvi put venčan(a)
☐ 5=ponovo venčan(a)
☐ 6=živi odvojeno od supružnika
☐ 7=razveden(a)
☐ 8=udovac/udovica
- Imate li dece koja trenutno žive s vama?
☐ da
☐ ne (označite odgovarajuće polje)
- Koju **najvišu** školu ste završili? (označite polje pored **najviše škole** koju ste završili)
☐ 1=osnovna škola
☐ 2=deo srednje škole
☐ 3=cela srednja škola
☐ 4=viša škola ili dodatna obuka
☐ 5=dodiplomske studije na univerzitetu
☐ 6=postdiplomske studije na univerzitetu
- Koji su glavni izvori stresa u vašem životu?
- Da li pušite?
☐ da
☐ na (označite odgovarajuće polje)
Ako odgovorite sa da, koliko cigareta popušite sedmično? _____

Pročitajte sledeće iskaze i razmislite koliko se slažete ili ne slažete sa svakim od njih. Pored svakog iskaza napišite broj koji, prema datoj skali, najbolje pokazuje vaše mišljenje.

nimalo se ne slažem 1 2 3 4 5 potpuno se slažem

1. ____ U nesigurnim vremenima obično očekujem da će mi se desiti ono najbolje.
2. ____ Ako išta može da mi krene po zlu, krenuće.
3. ____ Uvek sam optimista kada je u pitanju moja budućnost.
4. ____ Gotovo nikada ne očekujem da će se stvari razvijati meni u prilog.
5. ____ Sve u svemu, očekujem da će mi se desiti više dobrog nego lošeg.
6. ____ Retko očekujem da će mi se desiti nešto dobro.

Izvor: Scheier, Carver & Bridges, 1994.

Ova skala se sastoji od reči koje opisuju razna stanja i osećanja. Za svaku stavku naznačite do koje mere ste se tako osećali tokom nekoliko poslednjih sedmica. Na liniju pored svake stavke upišite broj od 1 do 5.

vrlo malo ili nimalo	malo	umereno	prilično	izuzetno mnogo
1	2	3	4	5
1.-zainteresovano		8.-ucveljeno		15.-uzbuđeno
2.-uzrujano		9.-snažno		16.-kao krivac
3.-prestravljeno		10.-odbojno		17.-oduševljeno
4.-ponosno		11.-razdražljivo		18.-čilo
5.-posramljeno		12.-nadahnuto		19.-nervozno
6.-odluno		13.-pažljivo		20.-usplahireno
7.-aktivno		14.-uplašeno		

Izvor: Watson, Clark & Tellegen, 1988.

Deo B: materijali za experim3ED.sav

Šifarnik za experim3ED.sav

Puno ime promenljive	Ime promenljive u SPSS-u	Duže opisno ime promenljive u SPSS-u	Uputstvo za šifrovanje
Identifikacioni broj	id	id	identifikacioni broj subjekta
Pol subjekta	sex	sex	1=muškarci, 2=žene
Starost subjekta	age	age	u godinama
Grupa	group	type of class	1=matematičke veštine, 2=izgradnja samopouzdanja
Test za ispitivanje straha od statistike u trenutku 1	fost1	fear of stats time1	Rezultati na testu kojim se ispituje strah od statistike, popunjavanom u trenutku 1. Opseg mogućih vrednosti od 20 do 60. Veće vrednosti pokazuju veći strah od statistike.
Vera u ovladavanje statistikom u trenutku 1	conf1	confidence time1	Rezultati ispitivanja vere u ovla- davanje statistikom, dobijeni u trenutku 1. Opseg mogućih vrednosti od 10 do 40. Veće vrednosti pokazuju veće samopouzdanje.
Depresija u trenutku 1	depress1	depression time1	Rezultati merenja na skali depresije, dobijeni u trenutku 1. Opseg mogućih vrednosti od 20 do 60. Veće vrednosti pokazuju veću depresiju.
Test za ispitivanje straha od statistike u trenutku 2	fost2	fear of stats time2	Rezultati na testu kojim se ispituje strah od statistike popunjavanom u trenutku 2. Opseg mogućih vrednosti od 20 do 60. Veće vrednosti pokazuju veći strah od statistike.
Vera u ovladavanje statistikom u trenutku 2	conf2	confidence time2	Rezultati ispitivanja vere u ovla- davanje statistikom, dobijeni u trenutku 2. Opseg mogućih vrednosti od 10 do 40. Veće vrednosti pokazuju veće samopouzdanje.
Depresija u trenutku 2	depress2	depression time2	Rezultati merenja na skali depresije, dobijeni u trenutku 2. Opseg mogućih vrednosti od 20 do 60. Veće vrednosti pokazuju

Puno ime promenljive	Ime promenljive u SPSS-u	Duže opisno ime promenljive u SPSS-u	Uputstvo za šifrovanje
Test za ispitivanje straha od statistike u trenutku 3	fost3	fear of stats time3	Rezultati na testu kojim se ispituje strah od statistike popunjavanom u trenutku 3. Opseg mogućih vrednosti od 20 do 60. Veće vrednosti pokazuju veći strah od statistike.
Vera u ovladavanje statistikom u trenutku 3	conf3	confidence time3	Rezultati ispitivanja vere u ovladavanje statistikom, dobijeni u trenutku 3. Opseg mogućih vrednosti od 10 do 40. Veće vrednosti pokazuju veće samopouzdanje.
Depresija u trenutku 3	depress3	depression time3	Rezultati merenja na skali depresije, dobijeni u trenutku 3. Opseg mogućih vrednosti od 20 do 60. Veće vrednosti pokazuju veću depresiju.
Rezultati na ispitu iz statistike	exam	exam	Rezultati na ispitu iz statistike. Opseg mogućih vrednosti 0-100.

Deo C: materijali za staffsurvey3ED.sav

Anketa za ispitivanje osoblja (odabrane stavke)

Starost: ☐ ispod 20
☐ 21 do 30
☐ 31 do 40
☐ 41 do 50
☐ preko 50 god.

Dužina zaposlenja u ovoj organizaciji (u godinama): _____

Zaposlen: ☐ stalno ☐ privremeno

Pomoću sledećih skala, za svaki od navedenih aspekata ocenite svoj stepen slaganja s njim i njegovu važnost.

Slaganje: 1=nimalo, 2=malo, 3=umereno, 4= mnogo, 5=veoma mnogo

Važnost: 1=nimalo, 2=malo važno, 3=umereno važno, 4= veoma važno, 5=izuzetno važno

		Slaganje	Važnost
1.	Da li je jasno šta se od vas očekuje na radnom mestu?	1 2 3 4 5	1 2 3 4 5
2.	Dobili ste na radnom mestu svu opremu i materijal koji su potrebni da svoj posao radite efikasno?	1 2 3 4 5	1 2 3 4 5
3.	Da li vas organizacija ažurno obaveštava o razvoju i promenama?	1 2 3 4 5	1 2 3 4 5
4.	Da li od organizacije dobijate priznanje kada dobro uradite posao?	1 2 3 4 5	1 2 3 4 5
5.	Da li vas direktor ili šef podstiču na profesionalni razvoj na radnom mestu?	1 2 3 4 5	1 2 3 4 5
6.	Smatrate li da se vaše mišljenje uvažava u organizaciji?	1 2 3 4 5	1 2 3 4 5
7.	Smatrate li da organizacija shvata koliko je vaš posao važan?	1 2 3 4 5	1 2 3 4 5
8.	Smatrate li da su vaše kolege na poslu posvećene tome da dobro rade svoj posao?	1 2 3 4 5	1 2 3 4 5
9.	Da li su vaš rad i njegovi rezultati bili ocenjivani ili razmatrani u poslednjih šest meseci?	1 2 3 4 5	1 2 3 4 5
10.	Da li ste tokom poslednje godine dana imali priliku da se usavršavate na poslu/radnom mestu?	1 2 3 4 5	1 2 3 4 5

Da li biste ovu organizaciju preporučili drugima kao dobro mesto za rad?

☐ Da ☐ Ne

Šifarnik za staffsurvey3ed.sav

Opis promenljive	Ime promenljive u SPSS-u	Uputstvo za šifrovanje
Identifikacioni broj	id	Identifikacioni broj subjekta
Mesto prebivališta zaposlenog	city	Svakom mestu boravišta dodeljen je određen broj
Starost zaposlenog	age	1=ispod 20 2=21 do 30 3=31 do 40 4=41 do 50 5=preko 50 god.
Broj godina radnog staža u organizaciji	service	Broj godina radnog staža u organizaciji (staž kraći od 1 godine pisati decimalno, npr. 6 meseci=0,5 godina)
Status zaposlenog	employstatus	1=stalno zaposlen, 2=privremeno zaposlen
P1 stepen slaganja	Q1a	1=nimalo, 2=malo, 3=umereno, 4= mnogo, 5=vrlo mnogo
P1 stepen važnosti	Q1imp	1=nimalo, 2=malo važno, 3=umereno važno, 4= veoma važno, 5=izuzetno važno
P2 stepen slaganja	Q2a	1=nimalo, 2=malo, 3=umereno, 4= mnogo, 5=vrlo mnogo
P2 stepen važnosti	Q2imp	1=nimalo, 2=malo važno, 3=umereno važno, 4= veoma važno, 5=izuzetno važno
P3 stepen slaganja	Q3a	1=nimalo, 2=malo, 3=umereno, 4= mnogo, 5=vrlo mnogo
P3 stepen važnosti	Q3imp	1=nimalo, 2=malo važno, 3=umereno važno, 4= veoma važno, 5=izuzetno važno
P4 stepen slaganja	Q4a	1=nimalo, 2=malo, 3=umereno, 4= mnogo, 5=vrlo mnogo
P4 stepen važnosti	Q4imp	1=nimalo, 2=malo važno, 3=umereno važno, 4= veoma važno, 5=izuzetno važno
P5 stepen slaganja	Q5a	1=nimalo, 2=malo, 3=umereno, 4= mnogo, 5=vrlo mnogo
P5 stepen važnosti	Q5imp	1=nimalo, 2=malo važno, 3=umereno važno, 4= veoma važno, 5=izuzetno važno

Opis promenljive	Ime promenljive u SPSS-u	Uputstvo za šifrovanje
P6 stepen slaganja	Q6a	1=nimalo, 2=malo, 3=umereno, 4= mnogo, 5=vrlo mnogo
P6 stepen važnosti	Q6imp	1=nimalo, 2=malo važno, 3=umereno važno, 4= veoma važno, 5=izuzetno važno
P7 stepen slaganja	Q7a	1=nimalo, 2=malo, 3=umereno, 4= mnogo, 5=vrlo mnogo
P7 stepen važnosti	Q7imp	1=nimalo, 2=malo važno, 3=umereno važno, 4= veoma važno, 5=izuzetno važno
P8 stepen slaganja	Q8a	1=nimalo, 2=malo, 3=umereno, 4= mnogo, 5=vrlo mnogo
P8 stepen važnosti	Q8imp	1=nimalo, 2=malo važno, 3=umereno važno, 4= veoma važno, 5=izuzetno važno
P9 stepen slaganja	Q9a	1=nimalo, 2=malo, 3=umereno, 4= mnogo, 5=vrlo mnogo
P9 stepen važnosti	Q9imp	1=nimalo, 2=malo važno, 3=umereno važno, 4= veoma važno, 5=izuzetno važno
P10 stepen slaganja	Q10a	1=nimalo, 2=malo, 3=umereno, 4= mnogo, 5=vrlo mnogo
P10 stepen važnosti	Q10imp	1=nimalo, 2=malo važno, 3=umereno važno, 4= veoma važno, 5=izuzetno važno
Preporuka	recommend	0=ne, 1=da
Promenljive koje se izračunavaju		
Ukupan rezultat na skali zadovoljstva osoblja	totsatis	Veći iznos pokazuje veće zadovoljstvo
Dužina radnog staža u organizaciji (3 grupe)	Servicegp3	1=< ili = 2 godine, 2= 3 do 5 godina, 3=6+ godina
Log transformacija dužine radnog staža	logservice	
Starost zaposlenog rešifrovana u 4 grupe	agerecode	1=18 do 30 godina, 2=31 do 40 godina, 3=41 do 50 godina, 4=50+ godina

Deo D: materijali za sleep3ED.sav

Šifarnik za sleep3ED.sav

Opis promenljive	Ime promenljive u SPSS-u	Uputstvo za šifrovanje
Identifikacioni broj	id	
Pol	sex	0=ženski, 1=muški
Starost	age	u godinama
Bračni status	marital	1=nevenčan(a), 2=venčan(a)/živi s partnerom, 3=razveden(a), 4=udovac/udovica
Najviša završena škola	edlevel	1=osnovna, 2=srednja, 3=zanat, 4=dodiplomske studije, 5=postdiplomske studije
Težina (kg)	weight	u kg
Visina (cm)	height	u cm
Ocenite svoje zdravlje	healthrate	1=vrlo slabo, 10=vrlo dobro
Ocenite svoju fizičku kondiciju	fitrate	1=vrlo slaba, 10=vrlo dobra
Ocenite svoju trenutnu težinu	weightrate	1=veoma mršav(a), 10=veoma debeo(la)
Da li pušite?	smoke	1=da, 2=ne
Koliko cigareta dnevno?	smokenum	broj cigareta dnevno
Koliko alkoholnih pića dnevno?	alcohol	broj pića na dan
Koliko kofeinskih pića dnevno?	caffeine	broj pića na dan
Sati sna/noć radnog dana	hourwnit	Prosečan broj sati sna svake noći radnog dana
Sati sna/noć neradnog dana	hourwend	Prosečan broj sati sna svake noći neradnog dana
Koliko sati sna smatrate potrebnim?	hourneed	Broj sati sna potreban da nema osećanja pospanosti
Teško zaspete?	trubslep	1=da, 2=ne
Lako se budite noću?	trubstay	1=da, 2=ne
Budite se noću?	wakenite	1=da, 2=ne
Radite noćnu smenu?	niteshift	1=da, 2=ne
Uvek se lako budite?	liteslp	1=da, 2=ne
Budite se čili radnim danom?	refreshd	1=da, 2=ne

Opis promenljive	Ime promenljive u SPSS-u	Uputstvo za šifrovanje
Zadovoljni količinom sna?	satsleep	1=veoma nezadovoljan, 10=u velikoj meri zadovoljan
Ocenite kvalitet svog sna	qualslp	1=vrlo loš, 2=loš, 3=pristojan, 4=dobar, 5=vrlo dobar, 6=izvrstan
Ocenite nivo svog stresa tokom prošlog meseca	stressmo	1=nimalo, 10=izuzetno mnogo
Pijete pilule za spavanje?	medhelp	1=da, 2=ne
Imate li problema sa snom?	problem	1=da, 2=ne
Ocenite uticaj problema sa snom na raspoloženje	impact1	1=nimalo, 10=izuzetno mnogo
Ocenite uticaj problema sa snom na nivo energije	impact2	1=nimalo, 10=izuzetno mnogo
Ocenite uticaj problema sa snom na koncentraciju	impact3	1=nimalo, 10=izuzetno mnogo
Ocenite uticaj problema sa snom na pamćenje	impact4	1=nimalo, 10=izuzetno mnogo
Ocenite uticaj problema sa snom na zadovoljstvo životom	impact5	1=nimalo, 10=izuzetno mnogo
Ocenite uticaj problema sa snom na opšte raspoloženje	impact6	1=nimalo, 10=izuzetno mnogo
Ocenite uticaj problema sa snom na ljubavne veze	impact7	1=nimalo, 10=izuzetno mnogo
Dešava vam se da u snu prestanete da dišete?	stopb	1=da, 2=ne
Nemirno spavate?	restlss	1=da, 2=ne
Jeste li ikada zaspali za volanom tokom vožnje?	drvsleep	1=da, 2=ne
Epworthova skala pospanosti	ess	Ukupan iznos na ESS skali (od 0=malo do 24=velika dnevna pospanost)
HADS skala anksioznosti	anxiety	Ukupan iznos na HADS skali anksioznosti (od 0=nema anksioznosti do 21=velika anksioznost)
HADS skala depresije	depress	Ukupan iznos na HADS skali depresije (od 0=nema depresije do 21=velika depresija)
Koliko napora ste utrošili tokom prošle sedmice?	fatigue	1=nimalo, 10=izuzetno mnogo

Opis promenljive	Ime promenljive u SPSS-u	Uputstvo za šifrovanje
Koliko ste bili letargični tokom prošle sedmice?	lethargy	1=nimalo, 10=izuzetno mnogo
Koliko ste bili umorni tokom prošle sedmice?	tired	1=nimalo, 10=izuzetno mnogo
Koliko ste bili pospani tokom prošle sedmice?	sleepy	1=nimalo, 10=izuzetno mnogo
Koliko vam je nedostajalo energije tokom prošle sedmice?	energy	1=nimalo, 10=izuzetno mnogo
Lako se budite noću, rešifrovano	stayslprec	0=ne, 1=da
Teško zaspate, rešifrovano	getsleprec	0=ne, 1=da
Kvalitet sna, rešifrovano	qualsleeprec	1=vrlo loš, loš, 2=pristojan, 3=dobar, 4=vrlo dobar, 5=izvrstan
skala Sleepiness and Associated Sensations (skala pospanosti i pridruženih osećanja)	totsas	Ukupan iznos na skali SASS (5=mala pospanost, 50=izuzetno velika pospanost)
Broj cigareta dnevno, rešifrovan u 3 grupe	cigsgp3	1=<=5, 2=6-15, 3=16+
Starost, rešifrovan u 3 grupe	agegp3	1=<=37god, 2=38-50god, 3=51+ godina
Imate li problema sa snom, rešifrovano u 0/1	probsleeprec	0=ne, 1=da

Deo E: materijali za depress3ED.sav

U 16. poglavlju pokazala sam kako se pomoću datoteke depress3ED.sav ispituje stepen podudaranja između dva merila depresije u uzorku sastavljenom od postnatalnih žena. Cilj je bio da se utvrdi da li žene identifikovane kao depresivne na skali EPDS jednako tako klasifikuje i skala DASS Depression (DASS-Dep).

Podaci u datoteci **depress3ED.sav** samo su mali deo stvarnih rezultata ispitivanja svake žene na skali Edinburgh Postnatal Depression Scale (EPDS: Cox, Holden & Sagovsky, 1987) i na skali Depression, Anxiety and Stress Scales (DASS-Dep: Lovibond & Lovibond, 1995). Rezultati su klasifikovani u skladu s preporučenim graničnim vrednostima za svaku skalu. Tako su dobijene dve promenljive (DASSdepgp2, EPDSgp2) s vrednostima 0 (nije depresivna) i 1 (depresivna).

Preporučena literatura

Ovo su neki od članaka i knjiga koji su mi bili najkorisniji u istraživanju i nastavi. Tražite nova izdanja navedenih naslova; mnogi se ažuriraju svakih nekoliko godina. Razvrstala sam ih po temama, mada je u mnogima obrađeno više tema. Naslovi koje najviše preporučujem označeni su zvezdicom.

Projektovanje istraživanja

Bowling, A. (1997). *Research methods in health: Investigating health and health services*. Buckingham: Open University Press.

*Boyce, J. (2003). *Market research in practice*. Boston: McGraw-Hill.

*Cone, J., & Foster, S. (1993). *Dissertations and theses from start to finish*. Washington: American Psychological Association.

Goodwin, C. J. (2007). *Research in psychology: Methods and design* (5th edn). New York: John Wiley.

Polgar, S., Thomas, S.A. (2000). *Introduction to research in the health sciences* (4th edn). Edinburgh: Churchill Livingstone.

Stangor, C. (2006). *Research methods for the behavioral sciences* (3rd edn). Boston: Houghton Mifflin.

Projektovanje upitnika

*De Vaus, D.A. (2002). *Surveys in social research* (5th edn). Crows Nest: Allen & Unwin.

Izbor i konstrukcija mernih skala

Dawis, R. V. (1987). Scale construction. *Journal of Counseling Psychology*, 34, 481–489.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd edn). Thousand Oaks, California: Sage.

Gable, R. K., & Wolf, M. B. (1993). *Instrument development in the affective domain: Measuring attitudes and values in corporate and school settings*. Boston: Kluwer Academic.

Kline, P. (1986). *A handbook of test construction*. New York: Methuen.

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (Eds). *Measures of personality and social psychological attitudes*. Hillsdale, NJ: Academic Press.

*Streiner, D. L., & Norman, G. R. (2003). *Health measurement scales: A practical guide to their development and use* (3rd edn). Oxford: Oxford University Press.

Osnovi statistike

Cooper, D. R., & Schindler, P. S. (2003). *Business research methods* (8th edn). Boston: McGraw-Hill.

Everitt, B. S. (1996). *Making sense of statistics in psychology: A second level course*. Oxford: Oxford University Press.

*Gravetter, F. J., & Wallnau, L. B. (2007). *Statistics for the behavioral sciences* (7th edn). Belmont, CA: Wadsworth.

Norman, G. R., & Streiner, D. L. (2000). *Biostatistics: The bare essentials* (2nd edn). Hamilton: B.C. Decker Inc.

Pagano, R. R. (1998). *Understanding statistics in the behavioral sciences* (5th edn). Pacific Grove, CA: Brooks/Cole.

*Peat, J. (2001). *Health science research: A handbook of quantitative methods*. Sydney: Allen & Unwin.

Raymondo, J. C. (1999). *Statistical analysis in the behavioral sciences*. Boston: McGraw-Hill College.

Runyon, R. P., Coleman, K. A., & Pittenger, D. J. (2000). *Fundamentals of behavioral statistics* (9th edn). Boston: McGraw-Hill.

Smithson, M. (2000). *Statistics with confidence*. London: Sage.

Napredni kurs statistike

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. & Tatham, R. L. (2006). *Multivariate data analysis* (6th edn). Upper Saddle River, NJ: Pearson Education.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd edn). Mahwah, NJ: Lawrence Erlbaum.

*Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th edn). Boston: Pearson Education.

Priprema izveštaja o istraživanju

American Psychological Association (2001). *Publication Manual* (5th edn). Washington: American Psychological Association.

McInerney, D. M. (2001). *Publishing your psychology research*. Crows Nest: Allen & Unwin.

Thomas, S. A. (2000). *How to write health sciences papers, dissertations and theses*. Edinburgh: Churchill Livingstone.

Bibliografija

Aiken, L. S. & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

American Psychological Association (2001). *Publication manual* (5th edn). Washington: American Psychological Association.

Bartlett, M. S. (1954). A note on the multiplying factors for various chi square approximations. *Journal of the Royal Statistical Society*, 16 (Series B), 296–8.

Berry, W. D. (1993). *Understanding regression assumptions*. Newbury Park, CA: Sage.

Bowling, A. (1997). *Research methods in health: Investigating health and health services*. Buckingham: Open University Press.

— (2001). *Measuring disease* (2nd edn). Buckingham: Open University Press.

— (2004). *Measuring health: A review of quality of life measurement scales*. Buckingham: Open University Press.

Boyce, J. (2003). *Market research in practice*. Boston: McGraw-Hill.

Briggs, S. R. & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, 54, 106–48.

Cattell, R. B. (1966). The scree test for number of factors. *Multivariate Behavioral Research*, 1, 245–76.

Choi, N., Fuqua, D. & N. Griffin, B. W. (2001). Exploratory analysis of the structure of scores from the multidimensional scales of perceived self efficacy. *Educational and Psychological Measurement*, 61, 475–89.

Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd edn). New York: Erlbaum.

Cohen, S., Kamarck, T. & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385–96.

Cone, J. & Foster, S. (1993). *Dissertations and theses from start to finish*. Washington: American Psychological Association.

- Cooper, D.R. & Schindler, P.S. (2003). *Business research methods* (8th edn). Boston: McGraw-Hill.
- Cox, J.L., Holden, J.M., Sagovsky, R. (1987). Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry*, 150, 782–6.
- Crowne, D.P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349–54.
- Daniel, W. (1990). *Applied nonparametric statistics* (2nd edn). Boston: PWS-Kent.
- Dawis, R.V. (1987). Scale construction. *Journal of Counseling Psychology*, 34, 481–9.
- De Vaus, D.A. (2002). *Surveys in social research* (5th edn). Crows Nest: Allen & Unwin.
- DeVellis, R.F. (2003). *Scale development: Theory and applications* (2nd edn). Thousand Oaks, California: Sage.
- Diener, E., Emmons, R.A., Larson, R.J. & Griffin, S. (1985). The Satisfaction with Life scale. *Journal of Personality Assessment*, 49, 71–6.
- Everitt, B.S. (1996). *Making sense of statistics in psychology: A second level course*. Oxford: Oxford University Press.
- Fox, J. (1991). *Regression diagnostics*. Newbury Park, CA: Sage.
- Gable, R.K. & Wolf, M.B. (1993). *Instrument development in the affective domain: Measuring attitudes and values in corporate and school settings*. Boston: Kluwer Academic.
- Glass, G.V., Peckham, P.D. & Sanders, J.R. (1972). Consequences of failure to meet the assumptions underlying the use of analysis of variance and covariance. *Review of Educational Research*, 42, 237–88.
- Goodwin, C.J. (2007). *Research in psychology: Methods and design* (5th edn). New York: John Wiley.
- Gorsuch, R.L. (1983). *Factor analysis*. Hillsdale, NJ: Erlbaum.
- Gravetter, F.J. & Wallnau, L.B. (2004). *Statistics for the behavioral sciences* (6th edn). Belmont, CA: Wadsworth.
- Greene, J. & d'Oliveira, M. (1999). *Learning to use statistical tests in psychology* (2nd edn). Buckingham: Open University Press.
- Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. & Tatham, R.L. (2006). *Multivariate data analysis* (6th edn). Upper Saddle River, NJ: Pearson Education.
- Harris, R.J. (1994). *ANOVA: An analysis of variance primer*. Itasca, Ill: Peacock.
- Hayes, N. (2000). *Doing psychological research: Gathering and analysing data*. Buckingham: Open University Press.
- Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179–85.

- Hosmer, D.W. & Lemeshow, S. (2000). *Applied logistic regression*. New York: Wiley.
- Hubbard, R. & Allen, S.J. (1987). An empirical comparison of alternative methods for principal component extraction. *Journal of Business Research*, 15, 173–90.
- Kaiser, H. (1970). A second generation Little Jiffy. *Psychometrika*, 35, 401–15.
- (1974). An index of factorial simplicity. *Psychometrika*, 39, 31–6.
- Keppel, G. & Zedeck, S. (1989). *Data analysis for research designs: Analysis of variance and multiple regression/correlation approaches*. New York: Freeman.
- (2004). *Design and analysis: a researcher's handbook* (4th edn). New York: Prentice Hall.
- Kline, P. (1986). *A handbook of test construction*. New York: Methuen.
- Kline, T.J.B. (2005). *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks: Sage.
- Lovibond, S.H. & Lovibond, P.F. (1995). *Manual for the Depression Anxiety Stress Scales* (2nd edn). Sydney: Psychology Foundation of Australia.
- Norman, G.R. & Streiner, D.L. (2000). *Biostatistics: The bare essentials* (2nd edn). Hamilton: B.C. Decker Inc.
- Nunnally, J.O. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Pagano, R.R. (1998). *Understanding statistics in the behavioral sciences* (5th edn). Pacific Grove, CA: Brooks/Cole.
- Pallant, J. (2000). Development and validation of a scale to measure perceived control of internal states. *Journal of Personality Assessment*, 75, 2, 308–37.
- Pavot, W., Diener, E., Colvin, C.R. & Sandvik, E. (1991). Further validation of the Satisfaction with Life scale: Evidence for the cross method convergence of wellbeing measures. *Journal of Personality Assessment*, 57, 149–61.
- Pearlin, L. & Schooler, C. (1978). The structure of coping. *Journal of Health and Social Behavior*, 19, 2–21.
- Peat, J. (2001). *Health science research: A handbook of quantitative methods*. Sydney: Allen & Unwin.
- Pett, M.A., Lackey, N.R. & Sullivan, J.J. (2003). *Making sense of factor analysis: the use of factor analysis for instrument development in health care research*. Thousand Oaks: Sage.
- Raymondo, J.C. (1999). *Statistical analysis in the behavioral sciences*. Boston: McGraw-Hill College.
- Robinson, J.P., Shaver, P.R. & Wrightsman, L.S. (eds). *Measures of personality and social psychological attitudes*. Hillsdale, NJ: Academic Press.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

- Runyon, R.P., Coleman, K.A. & Pittenger, D.J. (2000). *Fundamentals of Behavioral Statistics* (9th edn). Boston: McGraw-Hill.
- Scheier, M.F. & Carver, C.S. (1985). Optimism, coping and health: An assessment and implications of generalized outcome expectancies. *Health Psychology*, 4, 219–47.
- Scheier, M.F., Carver, C.S. & Bridges, M.W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A re-evaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, 67, 6, 1063–78.
- Siegel, S. & Castellan, N. (1988). *Nonparametric statistics for the behavioral sciences* (2nd edn). New York: McGraw-Hill.
- Smithson, M. (2000). *Statistics with confidence*. London: Sage.
- Stangor, C. (2006). *Research methods for the behavioral sciences* (3rd ed). Boston: Houghton Mifflin.
- Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd edn). Mahwah, NJ: Lawrence Erlbaum.
- Stober, J. (1998). The Frost multidimensional perfectionism scale revisited: more perfect with four (instead of six) dimensions. *Personality and Individual Differences*, 24, 481–91.
- Strahan, R. & Gerbasi, K. (1972). Short, homogeneous version of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 28, 191–3.
- Streiner, D.L. & Norman, G.R. (2003). *Health measurement scales: A practical guide to their development and use* (3rd edn). Oxford: Oxford University Press.
- Tabachnick, B.G. & Fidell, L.S. (2007). *Using multivariate statistics* (5th edn). Boston: Pearson Education.
- Thurstone, L.L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Watkins, M.W. (2000). *Monte Carlo PCA for parallel analysis* [computer software]. State College, PA: Ed & Psych Associates.
- Watson, D., Clark, L.A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063–70.
- Wright, R.E. (1995) Logistic regression. In L.G. Grimm & P. R. Yarnold (eds). *Reading and understanding multivariate statistics*. Washington, DC: American Psychological Association. Chapter 7.
- Zwick, W.R. & Velicer, W.F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432–42.

Spisak termina korišćenih u knjizi

ANOVA	<i>repeated</i>	oznaka tipa datoteke	<i>file extension</i>
ponovljenih merenja	<i>measures ANOVA</i>	postepeno pravougaoni dijagram, kutijasti dijagram	<i>stepwise box plot</i>
asimetrija	<i>skewness</i>	probni	<i>pilot</i>
dijagram, grafikon	<i>graph, chart, plot</i>	promenljiva	<i>variable</i>
dijagram rasturanja	<i>scatter plot</i>	proračunska tabela	<i>spreadsheet</i>
društveno poželjan odgovor	<i>socially desirable response</i>	remetilačka promenljiva	<i>confounding variable</i>
dvofaktorska analiza varijanse	<i>two-way ANOVA</i>	rešifrovanje	<i>recoding</i>
faktorska težina	<i>factor loading</i>	slučaj	<i>case</i>
grafikon, dijagram	<i>graph, chart, plot</i>	srednja vrednost	<i>mean</i>
izdvajanje faktora	<i>factor extraction</i>	srednja vrednost korelacije između stavki	<i>mean inter-item correlation</i>
izveden	<i>pivot</i>	standardno odstupanje, standardna devijacija	<i>standard deviation</i>
jednofaktorska analiza varijanse	<i>one-way ANOVA</i>	stepen slobode	<i>degree of freedom, df</i>
karakteristična vrednost	<i>eigenvalue</i>	stubičasti dijagram	<i>bar graph</i>
kombinovana analiza varijanse	<i>mixed between-within subjects ANOVA</i>	šifarnik	<i>codebook</i>
kutijasti dijagram, pravougaoni dijagram	<i>box plot</i>	t-test nezavisnih uzoraka	<i>independent-samples t-test</i>
kvadratni koren	<i>square root</i>	trakasti dijagram	<i>bar graph</i>
linijski dijagram	<i>line graph</i>	ukupan rezultat	<i>total score</i>
multivarijaciona analiza varijanse	<i>multivariate analysis of variance, MANOVA</i>	upareni subjekti	<i>matched pairs</i>
naknadni test	<i>post-hoc test</i>	uticaj interakcije	<i>interaction effect</i>
netipična tačka	<i>outlier</i>	veličina uticaja	<i>effect size</i>
određenost	<i>specificity</i>	zapisnička datoteka	<i>journal</i>
osetljivost	<i>sensitivity</i>	zbir kvadrata	<i>sum of squares</i>

Indeks

Napomena: imena SPSS procedura i okvira za dijalog ispisana su **podebljano**

analiza glavnih komponentata (PCA), 181–183
analiza kovarijanse, 107, 117, 295–316
analiza varijanse
 dvofaktorska različitih grupa, 107, 261–268
 jednofaktorska ponovljenih merenja, 254–259
 jednofaktorska različitih grupa, 115, 246–254
 kombinovana (ponovljenih merenja različitih grupa), 107, 116, 271–279
 multivarijaciona, 107, 117, 281–294
 pretpostavke, 205–207
ANCOVA. *Videti* analiza kovarijanse
ANOVA. *Videti* analiza varijanse
asimetrija raspodele, 58, 60

Bartletov test sferičnosti, 183, 187, 191
beta (standardizovani regresioni)
 koeficijenti, 161–162, 164
bibliografija, 339–342
binarna logistička regresija, 169–180
Boksov test jednakosti matrica
 kovarijanse, 273, 277, 288, 291
Bonferonijevo prilagođenje, 208, 231, 256, 282, 292

Chart Editor, prozor, 20, 79–80
Classification, tabela, 177
Compare Means, 226, 236, 240, 247, 252
Compute, 87–89
Correlate, 133, 146
Cox & Snell R Square (pseudopokazatelj veličine r na kvadrat), 177
Crosstabs, 217, 221

čišćenje podataka, 45–52

Data Editor, prozor, 18
Data Reduction, 187
datoteke s podacima
 korišćene u primerima, 317–335
 modifikovanje, 36
 otvaranje, 16
 pravljenje, 27–42
 snimanje, 16
definisane promenljivih, 31–35
delimična korelacija, 104, 114, 121, 129, 146–148
delimični eta kvadrat. *Videti* veličina uticaja
dijagrami, 67–82
 prevoja, 192
 rasturanja, 74–77
 srednjih vrednosti, 250
Direct Oblimin, 185, 188
diskriminaciona analiza, 105
diskriminaciona validnost, 7
dorada dijagrama, 79–80
dorada podataka, 83–96
duža opisna imena promenljivih, 33
dvofaktorska analiza varijanse, 106–107, 116, 261–269

eta kvadrat. *Videti* veličina uticaja
Excel datoteke, konverzija u format SPSS-a, 36–39
exp(B) količnici verovatnoće prediktora, 173, 178–179
experim3ED.sav, 318, 327–328
Explore, 60–62

Factor, 187
faktor
 izdvajanje, 183–184
 rotacija, 185

aktorska analiza
 potvrdna, 181
 istraživačka, 105, 121, 181–202
 i, koeficijent, 220–221
 format odgovora, 8–10
 frequencies, 46, 56, 91
 Fridmanov test, 231–233

General Linear Model, 104, 252–255,
 263, 273, 289, 303, 305, 310
 greška I vrste, 207–208, 245, 281–282,
 293
 greška II vrste, 207–208

Help, meni, 21, 23–24
 hi-kvadrat test kvaliteta podudaranja,
 215–217
 hi-kvadrat test nezavisnosti, 112, 217–221
 histogram, 62, 68–69
 homogenost matrica
 varijanse-kovarijanse, 283, 288,
 291
 homogenost regresije, 283, 288
 homogenost regresionih nagiba, 298,
 303–304
 homogenost varijanse, 126, 129, 152, 159,
 206, 249, 266, 277, 297
 Hosmer-Lemeshow, pokazatelj kvaliteta
 podudaranja, 176–177

imena promjenljivih 12–13, 32, 38
 ispravljanje grešaka u datoteci s podacima,
 49–51
 izbor odgovarajućih mernih skala, 5–7
 izbor prikladnih statističkih tehnika,
 103–120
 izbor promjenljivih, 22–23
 izmena SPSS-ovih opcija, 28–29
 izračunavanje ukupnih rezultata, 84–88

jednofaktorska analiza varijanse, 106, 115,
 245–260

Kajzer-Mejer-Olkinov pokazatelj
 adekvatnosti uzorka, 183, 187, 191
 Kajzerov kriterijum, 184, 192
 kanonska korelacija, 105
 kapa, mera slaganja, 221–223
 karakteristične vrednosti, 184, 188,
 192–194
 kategorijske promjenljive, 109
 kauzalnost, 123–124
 KMO. *Videti* Kajzer-Mejer-Olkinov
 pokazatelj adekvatnosti uzorka

koeficijent determinacije, 135–136
 koeficijenti kontrasta tj. poređenja grupa,
 252–254
 Koenov pokazatelj d, 209, 238
 količnik verovatnoće, 178–179
 Kolmogorov-Smirnov, test za verifikaciju
 neparametarskih hipoteza, 62
 kombinovana analiza varijanse, 107, 116,
 271–279
 konvergentna validnost, 7
 korelacija, 104, 113–114, 122–127,
 129–148
 korelacioni koeficijenti
 između grupa promjenljivih, 137–139
 poređenja dve grupe, 139–140
 korigovana korelacija stavke i ukupnog
 rezultata, 100
 korigovano R kvadrat, 161
 kosa rotacija, 185
 kovarijati, 288, 296–298
 Kramerov pokazatelj V, 221
 kriterijum dijagrama prevoja, 184
 Kronbahov koeficijent alfa, 6, 97–101
 Kruskal-Volison test, 229–231

lambda. *Videti* Vilkssov lambda
 Leveneov test, 206, 237–238, 249,
 265–266, 277, 291–292, 307, 314
 linearna korelacija, 129
 linearnost, 126, 130, 152, 159, 187, 283,
 286, 298, 300–301
 linijski dijagrami, 71–74
 logistička regresija, 121–122, 169–180

Mahalanobisova udaljenost, 159–160,
 284–286
 maksimalni rezultati, 48
 Man-Vitnijev U test, 224–226
 MANOVA. *Videti* multivarijaciona
 analiza varijanse
 medijana, 57–58
 minimalni rezultati, 48
 moć testa, 4, 207–209, 295, 298
 modifikovanje datoteke s podacima, 36
 Moklijev test sferičnosti, 257, 277
 multikolinearnost i singularnost, 151,
 158–159, 170, 172, 283, 288, 297
 Multiple Comparisons, tabela rezultata,
 249–250, 267
 multivarijaciona analiza varijanse, 107,
 117, 281–294
 multivarijaciona normalnost, 284

Nagelkerke R Square (pseudopokazatelj
 vrednosti r na kvadrat), 177
 naknadni (post-hoc) testovi, 106,
 208–209, 231, 233, 246–251, 267
 nedostajući podaci, 59, 87, 126–127, 134,
 210–211
 neparametarske statističke tehnike, 89,
 111–112, 213–234
 neprekidne promjenljive, 57, 109
 netipične tačke, 45, 53, 62, 79, 123, 132,
 152, 159–160, 170, 187, 284, 286
 nezavisna promjenljiva, 109
 nezavisni uzorci, t-test, 235–239
 nezavisnost opservacija, 125, 205–206
 Normal Q-Q, dijagram normalne
 verovatnoće, 62
 normalna raspodela, 62, 89, 206

obrtnanje negativno formulisanih stavki,
 85–86
 ograničen opseg rezultata, 123
 okviri za dijalog, 22–23
 Omnibus Tests of Model Coefficients,
 zbirni pokazatelji kvaliteta
 koeficijenata modela, 176
one-way ANOVA, 247, 252
 opcije SPSS-a, 28–31
 opisni statistički pokazatelji, 49, 55–66
 ortogonalna rotacija, 185
 osetljivost i određenost, 172, 177, 223
 otvaranje postojeće datoteke s podacima,
 16–17
 otvorena pitanja, 8–9, 13–14

paralelna analiza, 184, 192–193
 parametarske tehnike, 89, 111–112,
 203–206, 213
 Pirsonov koeficijent linearne korelacije,
 104, 121–127, 129–144
Pivot Table Editor, 20
 planirana poređenja, 208–209, 245,
 251–254
 podela datoteke s podacima na grupe, 41
 podela neprekidne promjenljive u grupe,
 92–93
 pokretanje SPSS-a, 15
 poludelimični koeficijenti korelacije, 162
 poređenje korelacionih koeficijenata dve
 grupe, 139–140
 pouzdanost, 6–7, 97–101
 pravljenje datoteke s podacima, 27–42
 pravljenje nove (prazne) datoteke za
 podatke, 17

pravougaoni dijagrami, 77–79
 premeštanje postojećih promjenljivih, 37
 priprema upitnika, 7–10
 projektovanje studije, 3–10
 pronalaženje grešaka u datoteci
 s podacima, 49–51

R kvadrat, 161–162, 166–167
Recode, 85, 94–95
 remetilne promjenljive, 5, 145
Repeated Measures, 273
 reziduali (odstupanja, greške), 152,
 159–161

sažet prikaz slučajeva, 51
Select Cases, 41–42
 skupovi, 42
 sleep3ED.sav, 318–319, 332–334
 snimanje datoteke s podacima, 16
 snimanje rezultata, 18–19
Sort Cases, 40
 sortiranje datoteke s podacima, 40
 spajanje datoteka, 39–40
 Spirmanova korelacija ranga, 125, 129,
 133
Split File, 41
 spljoštenost raspodele, 58
 SPSS-ovi prozori, 17
 srednja vrednost, 57–58
 srednja vrednost izračunata bez gornjih i
 donjih 5% slučajeva (Trimmed
 Mean), 61–62
 Staffsurvey3ED.sav, 319, 329–331
 standardizovani koeficijenti regresije, 161
 standardno odstupanje, 55–58
 strukturno modelovanje, 105
 stubičasti dijagrami, 69–71
 survey3ED.sav, 317–318, 320–326
Syntax Editor, 20–21

Šefeov test, 209
 šifarnik, 11–14, 322–323, 332–334
 šifrovanje odgovora, 13
 štampanje rezultata, 19–20

t-testovi, 106, 235–244
 nezavisnih uzoraka, 235–239
 pretpostavke, 205–207
 uparenih uzoraka, 239–243
 Tolerance, pokazatelj, 158, 170
Transform, 85, 87–89, 92
 transformacija podataka, 88–91
 transformacija promjenljivih, 83–95

- traženje grešaka u datoteci s podacima, 46–49
- Tukejev test zaista značajne različitosti (HSD), 209, 247
- uklanjanje jednog slučaja, 36
- uklanjanje promjenljive, 37
- umetanje promjenljive, 37
- umetanje slučaja, 37
- unošenje podataka, 36
- unošenje podataka u Excelu, 36–39
- unutrašnja saglasnost, 6–7, 97, 101
- upareni uzorci, t-test, 115, 239–243
- uticaji interakcije, 261, 266
- uvoz dijagrama u Wordove dokumente, 80–81
- validnost, 5, 7
- validnost konstrukta, 7
- validnost kriterijuma, 7
- validnost sadržaja, 7
- Variable View**, 31–32
- Varimax rotacija, 185, 198–199
- veličina uticaja
 - dvofaktorska analiza varijanse, 267
 - jednofaktorska analiza varijanse različitih grupa, 250
 - jednofaktorska analiza varijanse ponovljenih merenja, 259
- multivarijaciona analiza varijanse, 293
- neparametarske statističke tehnike, 220, 226, 228, 231, 295
- t-test nezavisnih uzoraka, 238–239
- t-test uparenih uzoraka, 242
- uvod, 209–210
- Viewer**, prozor, 17–18
- Vilkoksonov test ranga, 227–228
- Vilksov lambda, 258, 277
- Visual Binning, opcija, podela neprekidne promjenljive na grupe, 92–93
- višestruka regresija, 104, 114, 121, 149–168
 - hijerarhijska, 150, 163–167
 - postepena, 150–151
 - standardna, 150, 154–163
- vremenska stabilnost (pouzdanost) skale, 6
- vrste promjenljivih (atribut Type), 32
- Web lokacija, ix
- zasebni uticaji, 107, 261, 266, 273, 278, 314–315
- zatvaranje SPSS-a, 23
- značajnost, praktična u odnosu na statističku, 124