BINARNA LOGISTIČKA REGRESIJA: Rešenja zadataka

Rešenja zadataka uradila koleginica Ana Perović, dok je još bila studentkinja psihologije

Zadatak 1:
*U fajlu **verov_sanse_logit.sav** u varijabli **p** nalaze se verovatnoće za jednu kategoriju dihotomne kategoričke varijable:*

a) Izračunati šanse za datu kategoriju na osnovu verovatnoća
b) Izračunati logaritam šansi (logit) za datu kategoriju

Uporediti verovatnoće, šanse i logaritme šansi.

Šanse, tj. ocena šansi koju možemo izračunati na datom uzorku, predstavljaju količnik verovatnoće da neki entitet pripadne datoj kategoriji koju posmatramo i verovatnoće da taj isti entitet pripadne drugoj kategoriji binarne kriterijumske varijable.
Logaritam šansi (logit) predstavlja nelinearnu transformaciju kriterijumske varijable, varijable koju predviđamo na osnovu kombinacije prediktorskih varijabli.

- Veza šansi da ispitanik pripada jednoj u odnosu na drugu (od dve postojeće) kategorije binarne kriterijumske varijable i linearne kombinacije skupa prediktorskih varijabli je **nelinearna** što se vidi iz obrasca:
  $Š = \exp(b_0 + b_1*x_1 + ... + b_m*x_m) \rightarrow$ šansa je eksponencijalna funkcija (označeno sa exp) linearne kombinacije skupa prediktorskih varijabli (koja je u zagradi).
- Karakteristika binarne logističke regresije je da ona pretpostavlja **linearnost** veze između logartima šansi za određenu kategoriju na binarnoj kriterijumskoj varijabli i linearne kombinacije skupa prediktorskih varijabli.

U skladu sa navedenim definicijama, šanse za datu kategoriju izračunaćemo komandom:

Transform → Compute

U prozoru za dijalog imenovan kao **Compute Variable** polja **Target Variable** nazvaćemo redom 'Šanse' i 'Logit', dok u polje **Numeric Expression** upisujemo formule po kojima se računaju tražene šanse i logit.

a) Š=p/(1-p) b) Logit=LN(p/(1-p))

Upoređivanje verovatnoća, šansi i logaritama šansi izvršićemo pomoću **Scatter - diagrama**. Komandom **Graph → Scatter** (nakon odabira opcije **Simple**, a zatim klikom na dugme **Define** u dijaloškom prozoru **Simple Scatterplot** definišemo varijable čiji prikaz želimo) u okvire za x i y osu redom ubacujemo varijable:

1) x = p ; y = Šanse
   primećujemo da je veza nelinearna
2) x = p ; y = Logit
   veza ove 2 varijable je takođe nelinearna i daje *logističku krivu*!
3) x = Šanse ; y = Logit
   odnos ove 2 varijable je, očekivano, opisan logaritamskom funkcijom

!!!VAŽNO!!!

Uvidom u rezultate na 3 gore razmatrane varijable, primećujemo da:

- do vrednosti u koloni p (verovatnoća) od oko 0.15 šanse i verovatnoće su približne veličine; objašnjenje: za vrednosti $p<0.5 \rightarrow$ vrednost p je manja od (1-p), pa se količnik p/(1-p) nalazi između 0 i 1, što čini logit, tj. vrednost ln(p/(1-p)) manjim od 0! (prema osobinama logaritamske funkcije)
- međutim, kako verovatnoća nastavlja da raste ka vrednosti od 0.5 šanse se sve više približavaju jedinici
- najzad, kada verovatnoća dostigne vrednost od 0.5 vrednost šansi je 1 (tj. 1:1); za $p = 0.5 \rightarrow$ Šanse $= 1 \rightarrow$ Logit $= 0$ (jer je ln1 = 0)
- kako verovatnoća nastavlja da raste ka 1 (svom maksimumu), šanse se povećavaju od 1 ka $+\infty$; objašnjenje: za $p>0.5 \rightarrow$ vrednost p>(1-p), pa je p/(1-p)>1, što čini logit većim od 0! (prema osobinama logaritamske funkcije)

Verovatnoća (proporcija) p	Šanse	Logaritam šansi
0.01	0.01	-4.60
0.05	0.05	-2.94
0.10	0.11	-2.20
0.15	0.18	-1.73
0.20	0.25	-1.39
0.25	0.33	-1.10
0.30	0.43	-0.85
0.35	0.54	-0.62
0.40	0.67	-0.41
0.45	0.82	-0.20
0.50	1.00	0.00
0.55	1.22	0.20
0.60	1.50	0.41
0.65	1.86	0.62
0.70	2.33	0.85
0.75	3.00	1.10
0.80	4.00	1.39
0.85	5.67	1.73
0.90	9.00	2.20
0.95	19.00	2.94
0.99	99.00	4.60

Tabela 1. Prikaz odnosa verovatnoće, šansi i logaritma šansi (logita)

Zadatak 2:

*U fajlu **imz_mladi_razg.sav** nalaze se, između ostalog, podaci o polu (varijabla **SEX**) i, u varijabli **RAZGOVOR**, odgovori mladih srednjoškolaca na pitanje «Imate li potrebu da o svojim (psihološkim) problemima razgovarate sa stručnjakom (psihologom, pedagogom, lekarom...)?» Na ovo pitanje moglo se odgovoriti samo sa DA ili NE. Pri tome odgovori NE kodirani su cifrom 1, a odgovori DA cifrom 2.*

a) Pre daljeg rada potrebno je odgovore DA kodirati cifrom 1, a odgovore NE cifrom 0.
b) Izračunati šanse odgovora DA za mladiće
c) Izračunati šanse odgovora DA za devojke
d) Izračunati količnik šansi odgovora DA za dečake u odnosu na devojčice
e) Izračunati količnik šansi odgovora DA za devojčice u odnosu na dečake
f) Izračunati recipročne vrednosti količnika šansi koji su dobijeni pod c) i d).

Šta zaključujete na osnovu vrednosti dobijenih pod f?

a) U datoj varijabli 'razgovor' odgovori 'DA' kodirani su cifrom 2, a odgovori 'NE' cifrom 1. Za dalji rad neophodno je rekodirati odgovore u oblik pogodan programu što ćemo učiniti komandom **Transform → Recode → Into Different Variables** gde čemo u polje **Numeric Variable → Output Variable** uneti varijablu 'razgovor' od koje pravimo novu varijablu koju ćemo nazvati 'razg2'. Zatim klikom na dugme **Old And New Values** odlazimo u novi prozor gde je potrebno konkretno definisati vrednosti starih (**Old**) i novih (**New**) oznaka. Stare vrednosti kodujemo na sledeći način: vrednosti '1' (NE) dodeljujemo vrednost '0', dok vrednosti '2' (DA) dodeljujemo vrednost '1'. Odmah nakon ove transformacije neophodno je podesiti parametre u polju **Values** novonastale varijable 'razg2'.

b) Za naredna izračunavanja, potrebno je formirati **TABELU KONTIGENCIJE, komandom Analyze → Descriptive Statistics → Crosstabs.** Ovom prilikom ćemo proizvoljno odabrati da nam u redovima bude prikazana varijabla 'pol', a u kolonama 'razg2'. Tabela kontigencije nam daje uvid u empirijske frekvence svake pojedinačne situacije.

SEX * RAZG2 Crosstabulation

Count

		RAZG2		
		NE	DA	Total
SEX	BOYS	200	108	308
	GIRLS	136	129	265
Total		336	237	573

Tabela 2. Prikaz ispisa tabele Crosstabulation (SEX*RAZG2) u verziji 11.5 programa SPSS for Windows

Šanse odgovora 'DA' u odnosu na odgovor 'NE' za mladiće, u oznaci Š(DA| M) dobijamo na osnovu definicije šansi navedene u prvom zadatku:

$$Š(DA|\ M) = p(DA|\ M) / (1 - p(DA|\ M)) = 0.54$$

c) Šanse odgovora 'DA' u odnosu na odgovor 'NE' za devojke, u oznaci Š(DA| Ž) dobijamo formulom:

$$Š(DA|\ Ž) = p(DA|\ Ž) / (1 - p(DA|\ Ž)) = 0.95$$

d) Količnik šansi predstavlja količnik šansi koje su dobijene za svaku od dveju kategorija neke dihotomne varijable. Količnik šansi odgovora 'DA' za dečake u odnosu na devojčice se izračunava na sledeći način:

$$KŠ\ (DA)\ M/Ž = Š(DA|\ M) / Š(DA|\ Ž) = 0.57$$

Ova šansa se tumači na sledeći način: šanse da na pitanje o potrebi za razgovorom, kojim se bavimo, dečaci odgovore 'DA' umesto 'NE' predstavljaju 57% šansi da devojčice odgovore 'DA' umesto 'NE' na isto pitanje, tj. šanse da dečaci odgovore 'DA' umesto 'NE' na postavljeno pitanje su 0.57 puta veće od šansi da devojčice odgovore 'DA' umesto 'NE' na postavljeno pitanje.

e) Količnik šansi odgovora 'DA' za devojčice u odnosu na dečake se izračunava na sledeći način:

$$KŠ\ (DA)\ Ž/M = Š(DA|\ Ž) / Š(DA|\ M) = 1.76$$

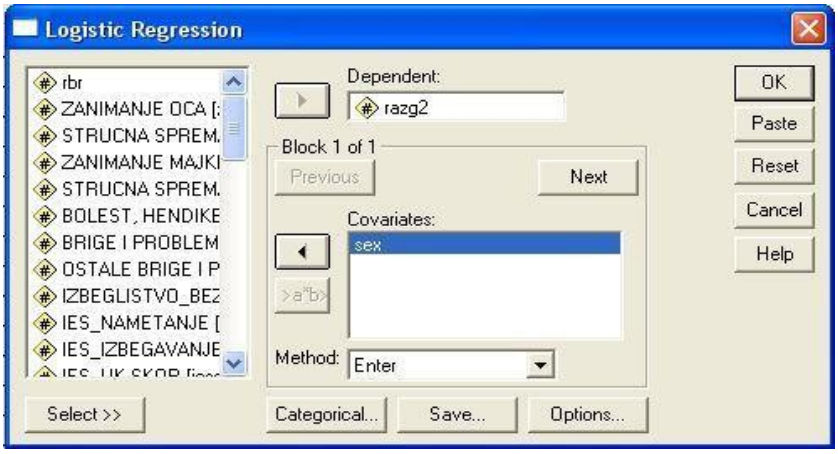
f) Na osnovu izračunatih količnika šansi može se zaključiti da je u binarnoj logističkoj regresiji KŠ u jednom smeru recipročna vrednost KŠ u drugom smeru, tj. da je:

$$KŠ\ (DA)\ M/Ž = 1 / KŠ\ (DA)\ Ž/M$$

Zadatak 3:

Fajl je isti kao u zadatku 2. Napraviti logistički regresioni model za predviđanje načina odgovaranja na ovo pitanje (pri tome od kritičnog interesa su odgovori DA) na osnovu polne pripadnosti. Uporediti ***exp(b)*** *za pol sa količnicima šansi iz zadatka 2 pod d i e. Šta zaključujete?*

Logistički regesioni model za binarnu kriterijumsku varijablu pravimo komandom **Analyze → Regression → Binary Logistic**. Zadatak nalaže da predviđanje odgovora izvršimo na osnovu polne pripadnosti, pa u polje **Dependent** unosimo kriterijumsku varijablu 'razg2', dok u polje **Covariates** unosimo sve prediktore (u našem slučaju 'pol'), pri čemu se podrazumeva da su sve prediktorske varijable kvantitativne (koje je preporučljivo standardizovati radi poređenja njihovog uticaja na kriterijumsku varijablu putem eksponenciranih logističkih koeficijenata – videti definiciju exp(b)) ili, eventualno, binarne (koje već sadrže 2 kategorije kodirane numerički sa 0 i 1).



Slika 1. Izgled dijaloga za definisanje logističke regresije sa jednim prediktorom u verziji 11.5 programa SPSS for Windows

Pri tumačenju ispisa, treba obratiti pažnju na:
- U tabeli *Case Processing Summary* sadržane su informacije o broju jedinica posmatranja korišćenih u analizi (Included in analysis) i broju jedinica posmatranja koje su izostavljene iz analize budući da barem na jednoj od korišćenih varijabli nemaju podatke (Missing cases). Naime, binarna logistička regresija je moguća samo sa onim ispitanicima koji na svim varijablama koje u njoj učestvuju imaju podatke. U protivnom, program briše ispitanike čije rezultate nemamo. Dakle, treba voditi računa da broj missing values bude sveden na minimum.
- Tabela *Dependent Variable Encoding* nam ukazuje na vrednosti koje program predlaže za kodovanje kategorija kriterijumske varijable – Internal Values. Dakle, program bi vrednosti pomenute dve kategorije kodovao u njemu pogodan oblik (values: 0 and 1) nezavisno od nas. Međutim, program to radi na sledeći način: oznaci za jednu od kategorija koja je po numeričkoj vrednosti manja, on pripisuje manju od 2 njemu pogodne vrednosti, tj. 0, a oznaci koja je numerički veća, pripisuje 1. To nama neće odgovarati u slučaju da je oznaka za odgovor 'NE' numerički veća od oznake za 'DA', jer bi onda od kritičnog interesa bili odgovori 'NE', a to nam često (kao u ovom slučaju) nije cilj. Upravo iz tog razloga mi sami kodujemo vrednosti kategorija kao u *zadatku 2. a)*.

- ***Block 0: Beginning Block*** pokazuje kako bi izgledao model predviđanja pre uvođenja prediktorskih varijabli (u našem slučaju varijable 'pol'). Model u Bloku 0 ima 3 oblika:

$\check{S}$ (DA) = p/(1-p) = exp(b_0)
ln$\check{S}$ = ln(p/(1-p)) = b_0 ; b_0 je analog intercepta u linearnoj regresionoj analizi
p (DA) = $\check{S}$/(1+$\check{S}$) = exp(b_0) / (1 + exp(b_0))

SPSS predikciju pre uvođenja prediktorskih varijabli određuje tako što svim ispitanicima pripisuje odgovor 'NE', tako da *Classification Table* možemo zanemariti. U tabeli *Variables in the Equation* očitavamo vrednosti koeficijenta exp(b_0) koji nam kazuje kakve su šanse odgovora 'DA' u odnosu na odgovor 'NE' pre uvođenja varijable 'pol' u regresiju. Ovaj koeficijent možemo dobiti deljenjem broja emprijiski dobijenih odgovora 'DA' sa brojem emprijski dobijenih odogovora 'NE' ne uzimajući u obzir varijablu 'pol'.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.349	.085	16.932	1	.000	.705

Tabela 3. Prikaz tabele Variables in the Equation iz ispisa Block 0 u verziji 11.5 programa SPSS for Windows

- ***Block 1: Method = Enter*** pokazuje promenu modela predikcije nakon uvođenja svih prediktora koji su ubačeni u prvom bloku. U našem primeru, ovim blokom je uveden prediktor 'pol' i model sada uzima sledeće forme:

$\check{S}$ (DA) = p/(1-p) = exp(b_0+b_1*POL)
ln$\check{S}$ = ln(p/(1-p)) = b_0+b_1*POL

$$p\,(DA) = Š/(1+Š) = \exp(b_0+b_1*POL) / (1 + \exp(b_0+b_1*POL))$$

Tabela *Omnibus Tests of Model Coefficients* sadrži ishode testiranja nulte hipoteze prema kojoj su populacioni logistički koeficijenti svih prediktorskih varijabli koje su u bloku 1 ubačene u model jednaki nuli: H_0: β_j=0, za svako j, tj. H_0: $\exp(\beta_j) = e^0 = 1$, za svako j. U našem slučaju ova hipoteza bi značila da je $\beta_{POL} = 0$, tj. da pol nema nikakvog uticaja na potrebu za razgovorom koju procenjujemo kao kriterijumsku varijablu. Test statistik koji služi za utvrđivanje statističke značajnosti dobijenih rezultata je **Chi square** tj. H^2 koji pod uslovom da je nulta hipoteza tačna ima hi-kvadrat distribuciju. U našem slučaju on iznosi 10.90. U koloni Significance prikazana je verovatoća da se na slučajnom uzorku dobije toliki ili veći H^2 statistik pod uslovom da je nulta hipoteza tačna. Sa tabele se očitava vrednost Sig. od 0.001 što je manje od 0.05 pa je stoga dobijeni H^2 statistički značajan i nulta hipoteza biva odbačena. Dakle, pol ima statistički značajan uticaj na kriterijumsku varijablu.

U tabeli *Model Summary* uočimo vrednost Najdželkerkeovog R-kvadrata, koji je analog koeficijentu multiple determinacije u linearnoj regresiji, i koji je zbog svog podesnog raspona mogućih vrednosti od 0 do 1, najprikladniji za praćenje dodatnog doprinosa novouvedenih prediktorskih varijabli. Očitavanjem vrednosti iz tabele vidimo da je ovaj $R^2 = 0.25$ što nije neki značajan doprinos varijable 'pol'.

Treba obratiti posebnu pažnju na tabelu pod nazivom *Variables in the Equation* koja sadrži:

- u koloni B - ocene logističkih koeficijenata za model sa prediktorima koji su uvedeni u bloku 1. Pri tome, u redu **Constant** nalazi se koeficijent b_0, a u redu koji počinje imenom određene prediktorske varijable je koeficijent b_j za prediktorsku varijablu v_j. Na osnovu logističkih koeficijenata možemo isključivo predvideti da li je uticaj prediktora na kriterijumsku varijablu pozitivan ili negativan (prostim očitavanjem znaka ovog koeficijenta), tj. da li povećanje vrednosti prediktora smanjuje ili povećava šansu kategorije koja je od kritičnog interesa u odnosu na onu drugu kategoriju binarne kriterijumske varijable. B_{POL} tj. β_{POL} ima pozitivnu vrednost, dakle prelaskom sa kategorije 0 (dečaci) na kategoriju 1 (devojčice) povećava se šansa pozitivnog odgovora u odnosu na negativni.

- Kolona **exp(B)** sadrži eksponencirane logističke koeficijente koji su veoma važni za tumačenje ishoda logističke regresije. Ako se logistički koeficijent b_j za prediktorsku varijablu v_j eksponencira, tada $\exp(b_j)$ predstavlja količnik šansi za određenu kategoriju kriterijumske varijable: šansi koje odgovaraju vrednosti jedne prediktorske varijable koja je promenjena za jednu jedinicu ('pol' sa 0 tj. muškog na 1 tj. ženski, odnosno šanse ženskog pola) i šansi koje odgovaraju vrednosti te prediktorske varijable pre no što je ona promenjena za jednu jedinicu ('pol' = 0, odnosno šanse muškog pola), pod uslovom da svi ostali prediktori u modelu ostanu nepromenjeni (ovde ih nema). U našem konkretnom slučaju, vrednost exp(b) = 1.76 što znači da žene imaju skoro 2 puta veće (ili za 76% veće) šanse odgovora 'DA' u odnosu na odgovor 'NE' od muškaraca.

Variables in the Equation

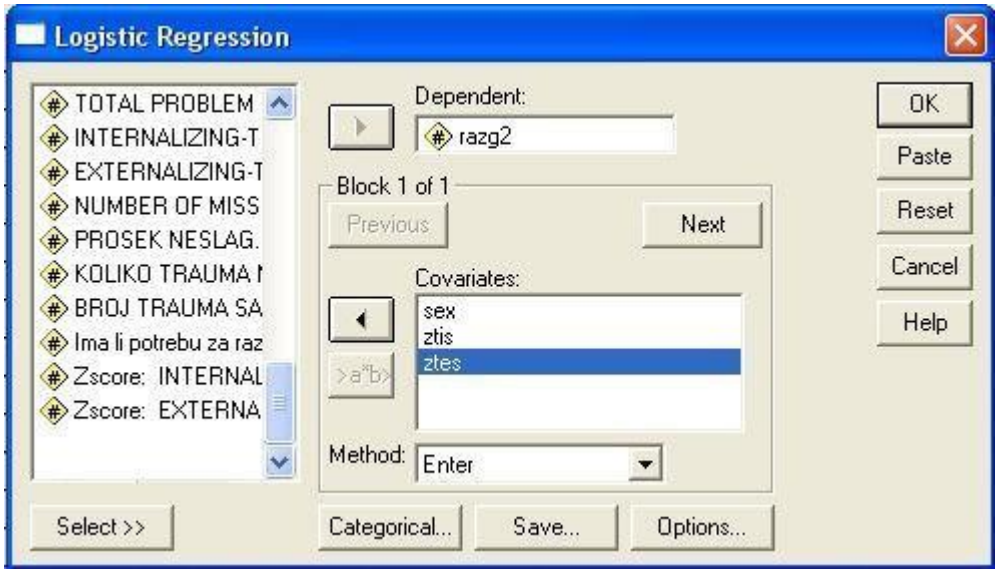
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	SEX	.563	.171	10.808	1	.001	1.757
	Constant	-1.180	.269	19.285	1	.000	.307

Tabela 4. Prikaz tabele Variables in the Equation iz ispisa Block 1 u verziji 11.5 programa SPSS for Windows

Konkretno logistički model određen samo na osnovu prediktorske varijable 'pol' izgleda ovako: ln(p(DA) / (1-p(DA))) = -1.180 + 0.563*POL

Zadatak 4:
*Fajl je isti kao u zadatku 2. Napraviti logistički regresioni model za predviđanje načina odgovaranja na ovo pitanje (pri tome od kritičnog interesa su odgovori DA) na osnovu polne pripadnosti, standardizovane izraženosti problema u ponašanju sa skala internalizacije (varijabla **TIS** sa nazivom **Internalizing total score**) i standardizovane izraženosti problema eksternalizacije (varijabla **TES** sa nazivom **Externalizing total score**). Šta zaključujete? Da li je **exp(b)** za pol jednak onom kao u modelu iz zadatka 3? Zašto?*

Na samom početku moramo standardizovati vrednosti varijabli TIS i TES. To postižemo komandom **Analyze → Descriptive Statistics → Descriptives** gde u prozoru čekiramo opciju Save standardized values as variables. Na ovaj način kreiramo 2 nove varijable ztis i ztes, na osnovu kojih ćemo vršiti dalju predikciju.
Komanda **Analyze → Regression → Binary Logistic** vodi nas u dijaloški prozor u čija polja unosimo sledeće podatke: **Dependent** (kriterijumska varijabla) je 'razg2', dok su **Covariates** (prediktori) 'sex' ('pol'), 'ztis' i 'ztes'.



Slika 2. Izgled dijaloga za definisanje logističke regresije sa tri istovremeno uvedena prediktora u verziji 11.5 programa SPSS for Windows

U ispisu treba obratiti pažnju na ***Block 1: Method = Enter*** . Polazeći od iste nulte hipoteze, H_0: $\beta_j=0$, za svako j, tabela *Omnibus Tests of Model Coefficients* pokazuje vrednost H^2 od 40, 234 i podatak da je verovatnoća da se na slučajnom uzorku dobije ova ili veća vrednost statistika (pod uslovom da je nulta hipoteza tačna) manja od 0.05 pa je stoga dobijeni test-statistik statistički značajan. Dakle, odbacujemo nultu hipotezu odnosno možemo tvrditi da je uticaj ovog skupa prediktora (stepen internalizovanja, stepen eksternalizovanja problema i pol) značajan po potrebu za razgovorom.
Da li je dodavanje još dva prediktora u model, pored polne pripadnosti, poboljšalo model za predviđanje kriterijumske varijable pokazuje poređenje Nejdželkerkeovih R kvadrata

iz tabela *Model Summary* za model u kojem je jedini prediktor bio pol i za model u kojem su pored pola u model kao prediktori uvedeni stepen internalizovanja i stepen eksternalizovanja problema. U prvom slučaju Nejdželkerkeov R-kvadrat iznosio je 0.025, a u drugom slučaju 0.091. Međutim, tumačenje Nejdželkerkeovog R-kvadrata je donekle problematično. Stoga je kao pokazatelj poboljšanja prediktivne moći modela bolje koristiti povećanje u procentu tačnih predviđanja odgovora koje možemo dobiti kao razliku procenata tačnih predviđanja (**Overal percentage**) u tabelama *Classification Table* za različite modele. Ukoliko se kao prediktor koristi samo pol, procenat tačnih predviđanja odgovora na postavljeno pitanje jednak je 58.6%, a za model u kojem su pored pola kao prediktori uvedeni stepen internalizovanja i stepen eksternalizovanja problema ovaj procenat iznosi 63.5%. Dakle, potonji model jeste nešto bolji, ali po cenu uvođenja dva nova prediktora u model.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	766.257	.019	.025

Tabela 5. Prikaz tabele Model Summary iz ispisa Block 1 zadatka 3 u verziji 11.5 programa SPSS for Windows

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	736.922	.068	.091

Tabela 6. Prikaz tabele Model Summary iz ispisa Block 1 zadatka 4 u verziji 11.5 programa SPSS for Windows

Classification Table(a,b)

Observed			Predicted		
			RAZG2		Percentage Correct
			NE	DA	
Step 0	RAZG2	NE	336	0	100.0
		DA	237	0	.0
	Overall Percentage				58.6

Classification Table(a)

Observed			Predicted		
			RAZG2		Percentage Correct
			NE	DA	
Step 1	RAZG2	NE	278	58	82.7
		DA	151	86	36.3
	Overall Percentage				63.5

Tabele 7 i 8. Prikaz tabela Classification Table iz ispisa Block 1 u zadatku 3 i Block 1 u zadatku 4 u verziji 11.5 programa SPSS for Windows

Iz kolone **B** u tabeli *Variables in the Equation* očitavamo vrednosti logističkih koeficijenata: b_0 = - 0.789 ; $b_1 = b_{POL}$ = 0.288 ; $b_2 = b_{ZTIS}$ = 0.541 ; $b_3 = b_{ZTES}$ = - 0.173. Prema tome, logistički model nakon uvođenja još 2 prediktora izgleda ovako:

$$\ln(p(DA) / (1-p(DA))) = - 0.789 + 0.288*POL + 0.541*ZTIS – 0.173*ZTES$$

!!!VAŽNO!!!
Uočimo da je u kontekstu novih prediktora logistički koeficijent za pol manji no što je bio u modelu u kojem je pol bio jedini prediktor. Isto tako, ukoliko logistički koeficijent za pol eksponenciramo, dobijamo vrednost u redu **Pol** i koloni **Exp(b)** tabele *Variables in the Equation*: $\exp(b_1) = 1.334$. Dakle, šanse odgovora DA na postavljeno pitanje (prema odgovoru NE) povećavaju se 1.33 puta kada se "vrednost" na prediktorskoj varijabli pol "poveća" za 1, a ostala dva prediktora u modelu drže konstantnim. To zapravo znači da su šanse odgovora 'DA' prema odgovoru 'NE' 1.33 puta veće za devojke nego za mladiće, pod uslovom da su devojke i mladići izjednačeni po stepenu internalizovanja i eksternalizovanja problema u ponašanju.

Variables in the Equation

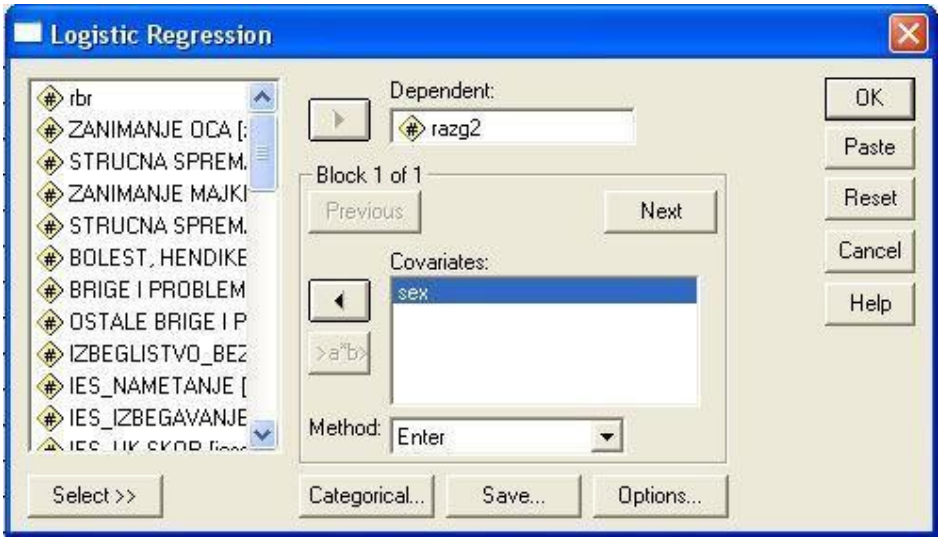
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	SEX	.288	.182	2.495	1	.114	1.334
	ZTIS	.541	.104	27.174	1	.000	1.718
	ZTES	-.173	.098	3.112	1	.078	.841
	Constant	-.789	.283	7.781	1	.005	.454

Tabela 9. Prikaz tabele Variables in the Equation iz ispisa Block 1 zadatka 4 u verziji 11.5 programa SPSS for Windows

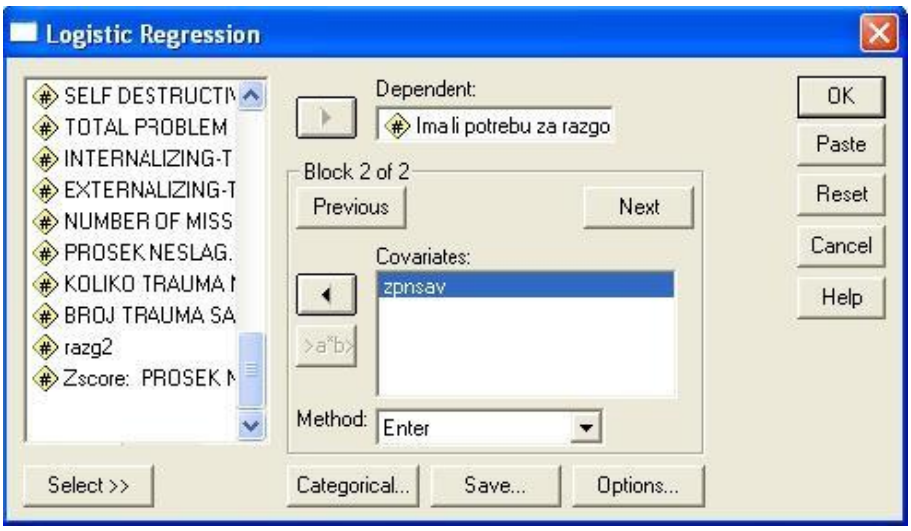
Poredeći apsolutne vrednosti logističkih koeficijenata 2 uvedena prediktora (koje se nalaze u koloni **B**) zaključujemo da je stepen internalizovanja problema važniji prediktor odgovaranja na pitanje nego stepen eksternalizovanja. Ovakvo poređenje relativne važnosti prediktorskih varijabli ima smisla samo ako su skale prediktora iste. Promena za 1 jedinicu na standardizovanim prediktorima (kakvi su naši) znači promenu od 1 standardne devijacije varijable. Prema tome, vrednosti u koloni **EXP(b)** za ova dva prediktora govore o tome da, ako se svi ostali prediktori drže konstantnim, povećanje stepena internalizovanja problema za jednu standardnu devijaciju povećava šanse odgovora 'DA' 1.72 puta (u skladu sa definicijom exp(b) iz *zadatka 3.*), dok povećanje stepena eksternalizacije za jednu standardnu devijaciju povećava šanse 'DA' 0.84 puta, pod uslovom da se svi ostali prediktori drže konstantnima. Na osnovu rezultata kolone **Sig.** očigledno je da je jedini dobar prediktor odgovora internalizacija problema.

Zadatak 5:
*Fajl je isti kao u zadatku 2. Napraviti logistički regresioni model za predviđanje načina odgovaranja na ovo pitanje (pri tome od kritičnog interesa su odgovori DA) na osnovu polne pripadnosti i izraženosti slaganja idealnog i realnog JA (varijabla **PNSAV**). Utvrditi da li uvođenje varijable **PNSAV** značajno dodatno doprinosi, pored polne pripadnosti, poboljšanju predviđanja dihotomne kriterijumske varijable (potreba za razgovorom sa stručnjakom).*

Pre no što bude uvedena u regresioni model, treba standardizovati varijablu 'Prosek neslaganja realnog i idealnog ja' ('pnsav'). Nakon toga pristupamo pomenutoj komandi za izradu logističkog regresionog modela: **Analyze → Regression → Binary Logistic.** S obzirom da nam treba dodatni doprinos ove varijable u predviđanju, iskoristićemo mogućnost da pojedine prediktorske varijable ili skupove prediktora uvodimo u logistički regresioni model postupno, a ne sve odjednom. U tom slučaju potrebno je, pri definisanju procedure za logističku regresiju, nakon ubacivanja imena prvog prediktora u okvir **Covariates** kliknuti na dugme **Next** iznad ovog okvira, ubaciti imena jednog ili više željenih prediktora u datom bloku, potom kliknuti na dugme **Next** i ubaciti sledeći prediktor ili više njih u sledećem bloku, i tako redom. U našem slučaju to znači da će prvi blok regresionog modela sadržati: 'razg2' kao **Dependent** (kriterijumsku) i varijablu 'pol' u polju **Covariates**. Drugi blok modela će sadržati istu kriterijumsku varijablu u polju **Dependent**, dok u polje **Covariates** upisujemo 'zpnsav'.



Slika 3. Izgled dijaloga za definisanje logističke regresije, uvođenje prvog prediktora u Block 1 u verziji 11.5 programa SPSS for Windows



Slika 4. Izgled dijaloga za definisanje logističke regresije, uvođenje drugog prediktora u Block 2 u verziji 11.5 programa SPSS for Windows

Ispis iz logističke regresije sa postupnim uvođenjem prediktora u osnovi sadrži iste tabele kao i ispis iz procedure u kojoj su svi prediktori uvedeni odjednom u prvom bloku, s tim što program u ispisu procedure sa postupnim uvođenjem prediktora jasno odvaja blokove modela. Ključna razlika u ispisu iz logističke regresije sa postupnim uvođenjem prediktora i logističke regresije sa istovremenim uvođenjem svih prediktora postoji u redu **Block** u tabeli *Omnibus Tests of Model Coefficients*.

U logističkoj regresiji u kojoj su sve varijable istovremeno uvedene u model, vrednosti u redu **Block** tabele *Omnibus Tests of Model Coefficients* jednake su vrednostima u redu **Model**. Razlog za to je što vrednost **Block**-a predstavlja doprinos predviđanju određenih varijabli (polje **Covariates**) koje su uvedene datim blokom, dok vrednost **Model** predstavlja test nulte hipoteze kada uzmemo u obzir sve prediktore uvedene u svim blokovima, što je u našoj regresiji apsolutno isto.

U logističkoj regresiji sa postupnim uvođenjem prediktora statistik H^2 (kolona **Chi square**), broj stepeni slobode (kolona **df**) i verovatnoća u koloni **Sig** koji su dati u redu **Block** tabele *Omnibus Tests of Model Coefficients* služe za testiranje statističke značajnosti ***dodatnog*** doprinosa objašnjenju ili predviđanju kriterijumske varijable onih prediktora koji su uvedeni u datom bloku.

U našem konkretnom slučaju dodatni doprinos koji daje varijabla 'pnsav' izražen hi-kvadratom iznosi 14.104 i statistički je značajan. Značajan je i celokupni doprinos varijabli 'pol' i 'pnsav' u predikciji odgovora na pitanje koje ispitujemo.

Zadatak 6:

*Fajl je isti kao u zadatku 2. Napraviti logistički regresioni model za predviđanje načina odgovaranja na ovo pitanje (pri tome od kritičnog interesa su odgovori DA) na osnovu polne pripadnosti, standardizovane izraženosti slaganja idealnog i realnog JA (varijabla **PNSAV**) i standardizovanog broja veoma stresnih životnih iskustava kojem je osoba bila izložena tokom života (varijabla **BRTRAUMA**, sa nazivom **Koliko trauma navodi**) Utvrditi da li uvođenje varijable **BRTRAUMA**, pored varijabli **SEX** i **PNSAV** značajno dodatno doprinosi poboljšanju predviđanja dihotomne kriterijumske varijable (potreba za razgovorom sa stručnjakom).*

Pre formiranja regresionog modela, standardizujemo varijablu 'brtrauma'. Nakon toga pristupamo pomenutoj komandi za izradu logističkog regresionog modela: **Analyze** →

Regression → Binary Logistic. Zadatak nalaže upotrebu više blokova regresije, i to na sledeći način:

- varijabla ‘razg2’ je konstantno prisutna u polju **Dependent** kao kriterijumska
- **Block 1 of 1** kao prediktore sadrži varijable ‘pol’ i ‘zpnsav’ u polju **Covariates**
- **Block 2 of 2** kao prediktore sadrži varijablu ‘zbrtrauma’ u polju **Covariates**

Praćenjem vrednosti Najdželkerkeovog koeficijenta (tabela *Model Summary*) možemo uočiti porast u procentu tačnih predviđanja uporedo sa uvođenjem novih varijabli. U bloku 1 njegova vrednost iznosi 0.65; zatim u bloku 2 ovaj procenat dostiže vrednost od 0.80.

Međutim, vrednost hi-kvadrata u redu **Block** tabele *Omnibus Tests of Model Coefficients* iznosi 5.445 i ne pokazuje statističku značajnost svog doprinosa u predviđanju dihotomne kriterijumske varijable.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	614.865	.048	.065

Tabela 10. Prikaz tabele Model Summary iz ispisa Block 1 u verziji 11.5 programa SPSS for Windows

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	609.420	.059	.080

Tabela 11. Prikaz tabele Model Summary iz ispisa Block 2 u verziji 11.5 programa SPSS for Windows

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	5,445	1	,020
	Block	5,445	1	,020
	Model	28,822	3	,000

Tabela 12. Prikaz tabele Omnibus Tests of Model Coefficients iz ispisa Block 2 u verziji 11.5 programa SPSS for Windows

Dodatni pokazatelj procenutalnog doprinosa tačnih klasifikacija varijable zbrtrauma u (odnosu na prethodni blok gde je predviđanje izvršeno samo na osnovu varijabli sex i zpnsav) možemo potražiti u tabeli *Classification Table* u redu **Overall percentage**. Primećujemo da je porast uspešnosti klasifikacije manji od 5% (tačnije iznosi 1.4%) što nikako ne može biti značajno.

Prilažem ovom prilikom tabelu *Variables in the Equation* koja je neophodna za režavanje sedmog zadatka.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	SEX	,460	,195	5,549	1	,018	1,584
	ZPNSAV	,349	,101	12,018	1	,001	1,418
	ZBRTRAUM	,227	,098	5,384	1	,020	1,255
	Constant	-1,054	,305	11,942	1	,001	,349

a Variable(s) entered on step 1: ZBRTRAUM.

Tabela 13. Prikaz tabele Variables in the Equation iz ispisa Block 2 u verziji 11.5 programa SPSS for Windows

Zadatak 7:

Fajl je isti kao u zadatku 2, a treba koristiti logistički model dobijen u zadatku 6.

*a) Kolike su <u>šanse</u> odgovora DA na pitanje «Imate li potrebu da o o svojim (psihološkim) problemima razgovara sa stručnjakom (psihologom, pedagogom, lekarom...)?» za mladu osobu ženskog pola, sa rezultatom na varijabli **PNSAV** jednakim 3 i sa 3 navedena vrlo stresna životna iskustva (varijabla **BRTRAUMA**)? A za osobu muškog pola sa rezultatom na varijabli **PNSAV** jednakim 3 i sa 3 navedena vrlo stresna životna iskustva (varijabla **BRTRAUMA**)?*

*b) Koliki je <u>logaritam šansi</u> odgovora DA na pitanje «Imate li potrebu da o o svojim (psihološkim) problemima razgovara sa stručnjakom (psihologom, pedagogom, lekarom...)?» za mladu osobu ženskog pola, sa rezultatom na varijabli **PNSAV** jednakim 3 i sa 3 navedena vrlo stresna životna iskustva (varijabla **BRTRAUMA**)? A za osobu muškog pola sa rezultatom na varijabli **PNSAV** jednakim 3 i sa 3 navedena vrlo stresna životna iskustva (varijabla **BRTRAUMA**)?*

*c) Kolika je <u>verovatnoća</u> odgovora DA na pitanje «Imate li potrebu da o o svojim (psihološkim) problemima razgovara sa stručnjakom (psihologom, pedagogom, lekarom...)?» za mladu osobu ženskog pola, sa rezultatom na varijabli **PNSAV** jednakim 3 i sa 3 navedena vrlo stresna životna iskustva (varijabla **BRTRAUMA**)? A za osobu muškog pola sa rezultatom na varijabli **PNSAV** jednakim 3 i sa 3 navedena vrlo stresna životna iskustva (varijabla **BRTRAUMA**)?*

a) Osobine ispitanika čije šanse računamo su sledeće:

- pol = 2 (ženski)
- zpnsav = 3 (kako je varijabla standardizovana, vrednost 3 znači 3 standardne devijacije iznad vrednosti aritmetičke sredine)
- zbrtrauma = 3 (i ova varijabla je standardizovana, te vrednost 3 takođe znači 3 standardne devijacije iznad vrednosti aritmetičke sredine)

Dakle, šanse ovog ispitanika su:

$$Š\ (DA) = \exp(b_0 + b_1*POL + b_2*ZPNSAV + b_3*ZBRTRAUMA)$$
$$Š\ (DA) = \exp(-1.054 + 0.460*2 + 0.349*3 + 0.227*3)$$
$$Š\ (DA) = 4.923$$

Sledi da su šanse da ispitanica ženskog pola, koja je prilično istraumirana i ima nisko samopoštovanje, želi da razgovara o svojim problemima sa stručnim licem znatno veće nego šanse za osobu muškog pola.
Šanse za osobu muškog pola i istih rezultata na preostala 2 prediktora su:

Š (DA) = exp(-1.054 + 0.460*1 + 0.349*3 + 0.227*3)
Š (DA) = 3.108

c) Za pomenutu ispitanicu, prirodni logaritam šansi odgovora 'DA' iznosi:

lnŠ (DA) = ln(exp(-1.054 + 0.460*2 + 0.349*3 + 0.227*3))
lnŠ (DA) = -1.054 + 0.460*2 + 0.349*3 + 0.227*3
lnŠ (DA) = 1.594

Za pomenutog ispitanika istih rezultata na preostala 2 prediktora kao i ispitanica, logaritam šansi odgovora 'DA' iznosi:

lnŠ (DA) = ln(exp(-1.054 + 0.460*1 + 0.349*3 + 0.227*3))
lnŠ (DA) = -1.054 + 0.460*1 + 0.349*3 + 0.227*3
lnŠ (DA) = 1.134

c) Za pomenutu ispitanicu, verovatnoća odgovora 'DA' iznosi:

$$p\,(DA) = Š/(1+Š) =$$
$$= \exp(-1.054+0.460*2+0.349*3+0.227*3) / (1+ \exp(-1.054+0.460*2+0.349*3+0.227*3))$$
$$p\,(DA) = 0.831$$

Za pomenutog ispitanika istih rezultata na preostala 2 prediktora kao i ispitanica, verovatnoća odgovora 'DA' iznosi:

$$p\,(DA) = Š/(1+Š) =$$
$$= \exp(-1.054+0.460*1+0.349*3+0.227*3) / (1+ \exp(-1.054+0.460*1+0.349*3+0.227*3))$$
$$p\,(DA) = 0.756$$