

12. Višestruka (multipla) linearna regresijska analiza

TEME

- | | |
|----|---|
| 1. | Preliminarna analiza podataka
– provjera pretpostavki |
| 2. | Procedura provedbe višestruke
linearne regresijske analize |
| 3. | Interpretacija dobivenog regresijskog modela |

12.1. Preliminarna analiza podataka – provjera pretpostavki

1) Metodološke pretpostavke:

- a) Kriterij je normalno distribuirana kvantitativna varijabla.
- b) Prediktori su normalno distribuirane kvantitativne varijable (iznimka: prediktor može biti i dihotomna varijabla).
 - Umjereno odstupanje od normalnosti se tolerira
 - U praksi umjesto kvantitativnih ponekad koristimo i unimodalno distribuirane ordinalne varijable

2) Statističke pretpostavke:

- a) Linearnost (odnos između svakog prediktora i kriterija je linearan + odnos između svih prediktora i kriterija je linearan).
- b) Varijable u modelu nemaju ekstremne vrijednosti (*outliers*).
- c) Pogreške (reziduali) su nezavisne/nekorelirane.
- d) Pogreške (reziduali) imaju jednake/homogene varijance (tzv. homoskedastičnost).
- e) Pogreške (reziduali) normalno su distribuirane.
- f) Prediktori su međusobno nezavisni/nekorelirani (nema multikolinarnosti).

Napomene:

- Više zadovoljenih pretpostavki implicira robusniju regresijsku analizu i veću "sigurnost" u interpretaciji ključnih parametara analize (parcijalni regresijski koeficijent, koeficijent višestruke determinacije)
- U stvarnosti se nerijetko događa da jedna ili više pretpostavki nije zadovoljena. Zbog kršenja nekih ne možemo provesti regresijsku analizu (npr. 1.a. ili 1.b.), a za druge postoje mogućnosti/tehnike korekcije (npr. za uklanjanje ekstremnih vrijednosti)
- Slijedeći dobru praksu, valja biti transparentan glede pretpostavki analize i temeljem toga nijansirati skromnost/opreznost u interpretaciji i donošenju zaključaka.

Osim multikolinearnosti, sve navedene pretpostavke vrijede i za jednostavnu linearnu regresijsku analizu.

Kako testiramo statističke pretpostavke za provedbu višestruke linearne regresije?

- a) Linearnost (odnos između svakog prediktora i kriterija je linearan + odnos između svih prediktora i kriterija je linearan) – *dijagrami raspršenja te QQ plotovi (Q-Q plot standardized residuals)*
- b) Varijable u modelu nemaju ekstremne vrijednosti (*outliers*) – *dijagrami raspršenja*
- c) Pogreške (reziduali) su nezavisne/nekorelirane – *Durbin-Watson test (idealno: statistik = 2)*
- d) Pogreške (reziduali) imaju jednake/homogene varijance (tzv. homoskedastičnost) – *Residuals vs. predicted plots*
- e) Pogreške (reziduali) normalno su distribuirane – *Q-Q plot standardized residual*
- f) Prediktori su međusobno nezavisni/nekorelirani (nema multikolinearnosti) – *idealno je VIF = 1, a tolerira se do VIF = 4*

Pretpostavke pod a), c) i d) provjerit ćemo prije same provedbe regresijske analize, a one pod b), e) i f) provjerit ćemo u sklopu provedbe procedure višestruke linearne regresijske analize.

12.1.1.
Provjera pretpostavki
o linearnosti veze i
normalnosti distribucije
varijabli

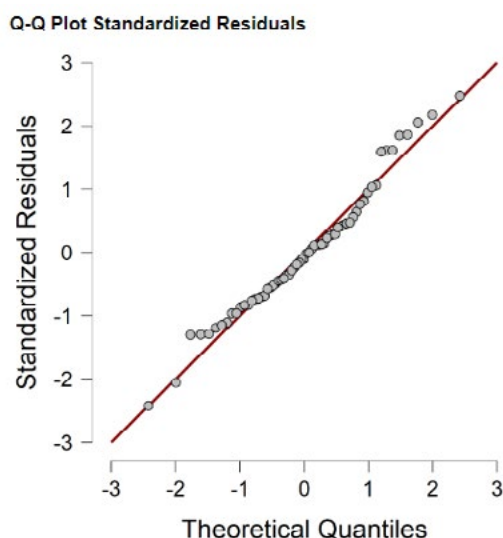
Radimo na primjerima u datoteci pod nazivom:
12_visestruka_linearna_regresija.sav

Provjera pretpostavki o **linearnosti veze i normalnosti distribucije** varijabli

Pod *Plots* označiti *Q-Q plot standardized residuals*.

Q-Q (skraćeno od engl. *quantile-quantile*) grafikon jest dijagram raspršenja s teoretskim kvantilima normalne distribucije na osi X i istim kvantilima reziduala uzorka na osi Y. Služi za usporedbu oblika distribucije varijable na uzorku s normalnom teorijskom raspodjelom. Kada su podaci normalno distribuirani, tada se točke nalaze na ili vrlo blizu dijagonalne referentne linije. Odstupanja od normalne distribucije izgledna su ako uočimo sljedeće pomake točaka od dijagonalne referentne linije:

- 4) Ako samo nekoliko točaka ne leži na pravcu, to ukazuje na mogućnost ekstremnih vrijednosti (engl. *outlier*).
- 5) Ako su točke pomaknute iznad ili ispod linije, to ukazuje na odstupanje u zakrivljenosti distribucije u odnosu na normalnu.
- 6) Ako se točke raspršuju "zmijoliko" oko linije, to ukazuje na asimetričnost distribucije.



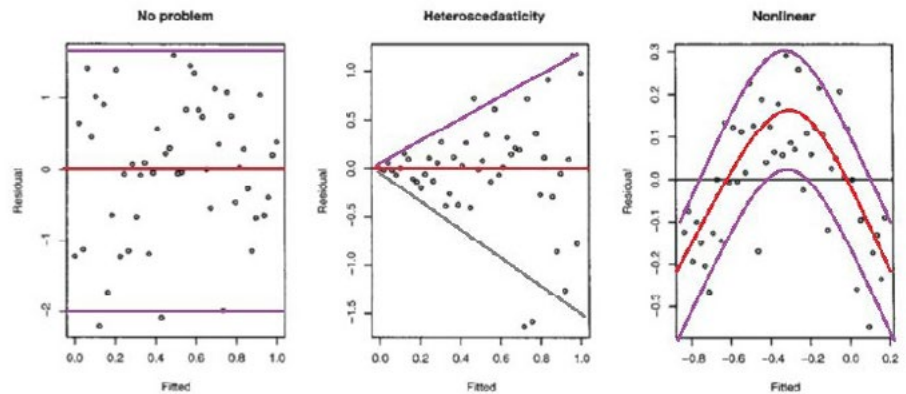
Q-Q plot pokazuje da standardizirani reziduali većinom leže duž dijagonalne linije, što ukazuje na to da ni jedna od ove dvije pretpostavke, o normalnosti raspodjele i linearnosti odnosa, nije prekršena.

12.1.2. Provjera pretpostavke o homoskedastičnosti

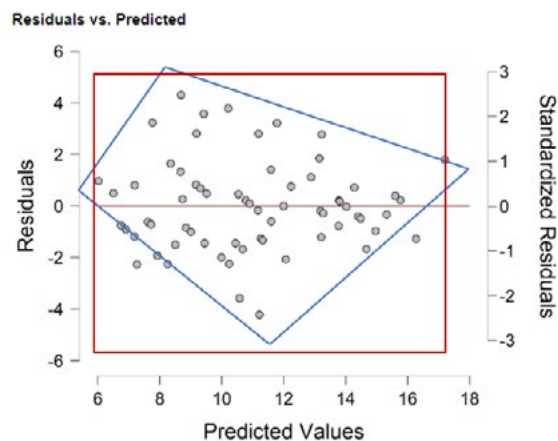
Provjera pretpostavke o **homoskedastičnosti** (hipoteza o homogenosti varijanci reziduala, tj. pretpostavka da je variranje podataka oko regresijskog pravca jednako za sve podatke prediktora).

Pod *Plots* označiti *Residuals vs. predicted*.

Mogući obrasci distribucije podataka (prema Goss-Sampson, 2018:22):



U našem primjeru oblik dijagrama raspršenja reziduala od prediciranih vrijednosti ukazuje na problem heteroskedastičnosti:



Kako na ovom grafičkom prikazu uočimo problem heteroskedastičnosti?

1. Povučemo li zamišljene linije uz rubne točke grafikona, očekujemo kvadratni oblik. U našem slučaju točke donekle čine kvadratni oblik.
2. Dobiveni bi kvadratni oblik trebao biti u što većoj mjeri horizontalan, odnosno paralelan s osi x (crvene linije). U našem je slučaju taj zamišljeni kvadratni oblik zarotiran za cca 45° , što ukazuje na problem heteroskedastičnosti.

12.1.3. Provjera pretpostavke o nezavisnosti/ nepovezanosti pogrešaka (reziduala)

Koje su posljedice heteroskedastičnosti?

- Heteroskedastičnost utječe na procjenu standardnih pogrešaka (*Standard Error*), a standardne pogreške utječu na procjenu p-vrijednosti. Prema tome, heteroskedastičnost može dovesti do lažno pozitivnih nalaza (pogreška tipa I).

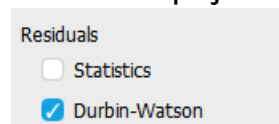
Što učiniti u slučaju heteroskedastičnosti?

- Najčešće se koristi izračun p-vrijednosti pomoću heteroskedastično dosljednih pogrešaka. U tom se slučaju mijenjaju p-vrijednosti i standardizirani regresijski koeficijenti, dok nestandardizirani regresijski koeficijenti ostaju identični. JASP trenutno još nema tu opciju.

U literaturi se te pogreške još nazivaju robusne standardne pogreške, Huber-Whiteove standardne pogreške ili sendvič-procjene. 🍷

Provjera pretpostavke o **nezavisnosti/nepovezanosti pogrešaka (reziduala)**

Odabirom opcije Durbin-Watson...



...u tablici Model Summary dobivamo Durbin-Watsonov statistik koji poprima vrijednost u rasponu od 0 do 4. Pogreške su nepovezane kada je Durbin-Watsonov statistik jednak vrijednosti 2. Ako je statistik veći od 2, imamo pozitivno povezane pogreške, a ako je statistik manji od 2, imamo negativno povezane pogreške.

Model Summary - ZADOVPOS ▼

Model	R	R ²	Adjusted R ²	RMSE	Durbin-Watson		
					Autocorrelation	Statistic	p
1	0.000	0.000	0.000	3.264	-0.088	2.139	0.572
2	0.724	0.524	0.516	2.270	-0.104	2.034	0.891
3	0.826	0.682	0.671	1.871	-0.023	1.928	0.930
4	0.846	0.716	0.702	1.782	-0.173	2.266	0.282

U našem je slučaju dobivena je vrijednost 2,266, što smatramo zanemarivom pozitivnom povezanošću pogrešaka (reziduala), tj. ne smatramo je problematičnom.

12.2. Procedura provedbe višestruke linearne regresijske analize

**Koristimo proceduru
'Linear regression'**

Višestruka linearna regresijska analiza statistička je tehnika koja se u sociologiji najčešće koristi da bi se utvrdilo u kojoj mjeri se varijacije neke varijable (zavisna varijabla, u višestrukoj regresiji naziva se **kriterijem**) mogu objasniti njezinom povezanošću s drugim varijablama (nezavisne varijable, u višestrukoj regresiji nazivaju se **prediktorima**).

Takav model, u kojem se **kriterij iskazuje kao linearna kombinacija prediktora**, izražava se jednadžbom višestruke regresije:

$$\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_kX_k$$

gdje su $X_{1,2,3,\dots,k}$ prediktorske varijable, $b_{1,2,3,\dots,k}$ parcijalni su regresijski koeficijenti (engl. *slope*) pojedinih prediktora, koji iskazuju predviđeni iznos promjene vrijednosti kriterija za jedinični porast prediktora, dok je a (engl. *intercept*) procijenjena vrijednost kriterija u slučaju da je vrijednost svih prediktora jednaka nuli.

Cilj regresijske analize jest, dakle, da se utvrdi u kojoj mjeri takva linearna kombinacija prediktora tumači varijacije kriterija te koliki je pritom doprinos (važnost) pojedinih prediktora.

Radimo na primjerima u datoteci pod nazivom:
12_visestruka_linearna_regresija.sav

Provedbu višestruke regresijske analize u JASP-u demonstrirat ćemo primjerom koji uključuje sljedeće varijable:

- ZADOVPOS – zadovoljstvo ispitanika poslom koji obavljaju;
- PLACA – iznos plaće;
- DOB – dob ispitanika;
- RUTINA – procjena rutinskog karaktera posla;
- SAMOSTAL – procjena samostalnosti pri obavljanju radnih zadaća.

Kriterijska varijabla, ona koju nastojimo protumačiti ostalima, bit će ZADOVPOS, dok će ostale varijable biti prediktori.

U slučaju kada bi se svi prediktori pokazali statistički značajnima, regresijska jednadžba glasila bi:

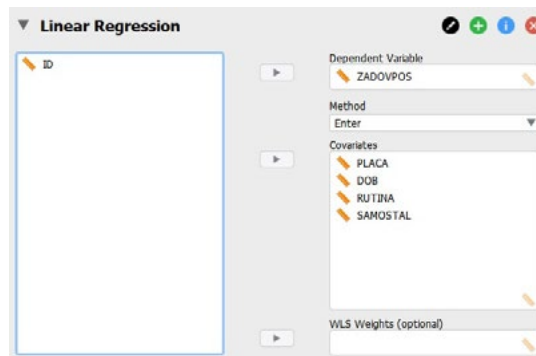
$$\text{ZADOVPOS} = a + b_1(\text{PLACA}) + b_2(\text{DOB}) + b_3(\text{RUTINA}) + b_4(\text{SAMOSTAL}).$$

Višestrukom regresijskom analizom utvrđujemo:

- 1) Koji su prediktori statistički značajni?
- 2) U kojoj mjeri svaki od prediktora doprinosi tumačenju varijance kriterija?
- 3) Koja je proporcija varijance kriterija protumačena linearnom kombinacijom svih prediktora?
- 4) Koliki je učinak pojedinog prediktora na kriterij (dok su svi ostali prediktori konstantni)?
- 5) Koliko procjenjujemo da će u prosjeku iznositi vrijednost kriterija za određenu kombinaciju vrijednosti prediktora?

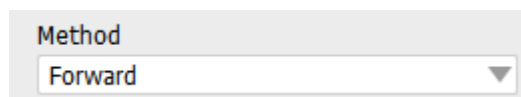
U modulu *Regression* odabiremo proceduru (*Classical*) *Linear Regression*.

Kriterijsku varijablu ZADOVPOS premjestimo u okvir *Dependent Variable*, a prediktore u okvir *Covariates*.



Potom je potrebno odabrati odgovarajuću metodu izgradnje regresijskog modela pod *Method*. Za potrebe ovog primjera odabrat ćemo *Forward* metodu.

Metode izgradnje regresijskog modela



- 1) **Forward.** Varijable ulaze u regresijski model jedna po jedna. Varijabla koja se prva razmatra za ulazak u regresijsku jednadžbu jest ona s najvećim Pearsonovim koeficijentom korelacije s kriterijem. Ulazni kriterij (F-test regresijskog koeficijenta prediktora) moguće je konfigurirati u *Method Specification*, 'Stepping Method Criteria'. Uobičajeno je da se koristi tvornička postavka, odnosno p-vrijednost F-testa. Nakon ulaska prve varijable u regresijsku jednadžbu, razmatra se onaj od preostalih prediktora koji ima najveću parcijalnu korelaciju s kriterijem (uz isključenje utjecaja onih prediktora koji su već u regresijskoj jednadžbi). Ako ta varijabla udovoljava ulaznom kriteriju, ulazi u regresijsku jednadžbu i procedura se nastavlja na isti način sa sljedećom varijablom. Izgradnja regresijskog modela zaustavlja se onda kada nijedan od preostalih prediktora ne udovoljava ulaznom kriteriju.
- 2) **Backward.** Dok *Forward* metoda započinje bez ijednog prediktora u regresijskoj jednadžbi, pa ih potom redom unosi u nju prema određenom kriteriju, postupak kod *Backward* metode upravo je obratan: u prvom koraku analize svi prediktori ulaze u regresijsku jednadžbu, da bi se potom iz nje isključivali ako se ne pokažu statistički značajnima.
- 3) **Stepwise.** Metoda koja kombinira *Forward* i *Backward* metode. Varijable redom ulaze u regresijsku jednadžbu kao kod *Forward* metode, da bi se potom u svakom koraku analize provjerilo udovoljavaju li prediktori koji su u regresijskom modelu uvjetima za ostanak, kao kod *Backward* metode.
- 4) **Enter.** Metoda u kojoj istraživač odlučuje o tome koje će varijable biti statistički testirane za ulazak u model. Na taj način moguće je

testirati određeni teorijski model. Neznačajne prediktore potom "ručno" uklanjamo iz regresijskog modela, odnosno oni ne ulaze u jednadžbu regresijskog pravca.

Prve tri opisane metode vode se u izgradnji regresijskog modela statističkim, a ne teorijskim kriterijima.

Regresijski model bez prediktora naziva se *null-model* (ili *intercept-only* model). Model s prediktorima naziva se puni model (*full-model*).

Odabir statističkih pokazatelja za definiranje i interpretaciju regresijskog modela

- 1) Model summary tablica uvijek se pojavljuje kada pokrenemo proceduru 'Linear regression' te uključuje:
 - a. R: koeficijente višestruke korelacije kriterija s prediktorima u regresijskoj jednadžbi (u pojedinim koracima analize)
 - b. R²: koeficijente višestruke determinacije, odnosno pokazuje proporciju varijance kriterija protumačenu prediktorima u regresijskoj jednadžbi (veličina učinka za cijeli regresijski model)
 - c. Adjusted R²: korigirani koeficijent višestruke determinacije konzervativnija je procjena proporcije objašnjene varijance kriterija, koja uzima u obzir broj ispitanika i broj varijabli u modelu
 - d. RMSE (*root-mean-square error*): Svojevrсна mjera preciznosti, odnosno agregirana mjera pogreške u predikciji vrijednosti kriterija. Nije standardizirana (ovisi o skali mjerenja).

Model Summary - ZADOVPOS

Model	R	R ²	Adjusted R ²	RMSE	R ² Change	F Change	df1	df2	p
1	0.000	0.000	0.000	3.264	0.000		0	64	
2	0.724	0.524	0.516	2.270	0.524	69.357	1	63	< .001
3	0.826	0.682	0.671	1.871	0.158	30.714	1	62	< .001
4	0.846	0.716	0.702	1.782	0.034	7.364	1	61	0.009

- 2) Estimates. Aktivira tablicu 'Coefficients' s (parcijalnim) regresijskim koeficijentima (*slopes*) i odsječcima (*intercepts*). Također pokazuje koji su prediktori u kojem koraku ulazili/izlazi u regresijski model.
 - a. Parcijalni regresijski koeficijenti prediktora (*Unstandardized; b*) iskazuju za koliko će se promijeniti kriterij ako se prediktor promijeni za jediničnu vrijednost, pri čemu je parcijaliziran (kontroliran) utjecaj ostalih prediktora koji su u regresijskoj jednadžbi. → Za zapis regresijske jednadžbe koristimo b-koeficijente finalnog regresijskog modela.
 - b. Parcijalni regresijski koeficijenti ne mogu se interpretirati kao pokazatelji relativne važnosti prediktora, jer njihova veličina ovisi o metrici varijabli. Ako se varijable razlikuju po jedinicama mjerenja, veličina njihovih parcijalnih regresijskih koeficijenata ne otkriva njihovu relativnu važnost. Za uspoređivanje

- b -koeficijenata koristimo parcijalne regresijske koeficijente standardiziranih prediktorskih varijabli (*Standardized*; β).
- c. Intercept je odsječak (**a**). → Za zapis regresijske jednadžbe koristimo vrijednost odsječka finalnog regresijskog modela.
- d. T-testom je testiran b -koeficijent svakog prediktora prema hipotezi: $H_0: b = 0$.
- e. **Važno!** Kada interpretirate dijelove regresijske jednadžbe (npr. ako se X poveća za jednu jedinicu mjerenja, Y se smanji za 0,665.), potrebno je naglasiti da to vrijedi ako sve ostale prediktore držimo konstantnima.

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	Collinearity Statistics	
							Tolerance	VIF
1	(Intercept)	10.969	0.405		27.093	< .001		
2	(Intercept)	2.815	1.019		2.763	0.007		
	SAMOSTAL	0.859	0.103	0.724	8.328	< .001	1.000	1.000
3	(Intercept)	-6.335	1.852		-3.420	0.001		
	SAMOSTAL	0.688	0.090	0.580	7.604	< .001	0.883	1.132
	PLACA	0.001	2.401e-4	0.422	5.542	< .001	0.003	1.132
4	(Intercept)	-1.930	2.397		-0.005	0.424		
	SAMOSTAL	0.574	0.096	0.484	5.992	< .001	0.714	1.400
	PLACA	0.001	2.413e-4	0.382	5.145	< .001	0.847	1.181
	RUTINA	-0.170	0.063	-0.219	-2.714	0.009	0.717	1.395

Note. The following covariate was considered but not included: DOB.

- 3) Model fit. Aktivira tablicu ANOVA (koristi se za testiranje statističke značajnosti regresijskog modela prema sljedećoj hipotezi: $H_0: R^2 = 0$ (koeficijent višestruke determinacije u populaciji je jednak nuli, tj. varijanca kriterija nije protumačena korelacijom s prediktorom))

ANOVA ▼

Model		Sum of Squares	df	Mean Square	F	p
2	Regression	357.346	1	357.346	69.357	< .001
	Residual	324.593	63	5.152		
	Total	681.938	64			
3	Regression	464.876	2	232.438	66.392	< .001
	Residual	217.062	62	3.501		
	Total	681.938	64			
4	Regression	488.257	3	162.752	51.259	< .001
	Residual	193.681	61	3.175		
	Total	681.938	64			

Note. The intercept model is omitted, as no meaningful information can be shown.

- 4) R squared change. Ima smisla samo za Forward, Backward i Stepwise metodu izgradnje modela. U tablici 'Model Summary' dodaje promjene R^2 (koeficijenta višestruke determinacije) nakon svakog koraka analize (ulaska/izlaska novog prediktora u model) i test statističke značajnosti te promjene. Test značajnosti promjene R^2 u svakom koraku uključuje F-omjer promjene i p-vrijednost promjene. Oprez! Ovaj F-test ne testira $H_0: R^2 = 0$. Testiranje regresijskog modela nalazimo u tablici ANOVA.

- a. R^2 Change pokazuje za koliko se povećava taj koeficijent ulaskom pojedinih prediktora u regresijsku jednadžbu. Promjena koeficijenta determinacije jest jedan od načina procjene relativne važnosti pojedinih prediktora.

Model Summary - ZADOVPOS

Model	R	R ²	Adjusted R ²	RMSE	R ² Change	F Change	df1	df2	p
1	0.000	0.000	0.000	3.264	0.000		0	64	
2	0.724	0.524	0.516	2.270	0.524	69.357	1	63	< .001
3	0.826	0.682	0.671	1.871	0.158	30.714	1	62	< .001
4	0.846	0.716	0.702	1.782	0.034	7.364	1	61	0.009

5) Descriptives (deskriptivni pokazatelji varijabli u modelu)

Descriptives

	N	Mean	SD	SE
ZADOVPOS	65	10.969	3.264	0.405
PLACA	65	7835.385	1002.876	124.391
DOB	65	39.769	12.275	1.523
RUTINA	65	13.446	4.209	0.522
SAMOSTAL	65	9.492	2.751	0.341

6) Part and partial correlations. Aktivira istoimenu tablicu u kojoj se nalaze parcijalne (*partial*) i semiparcijalne (*part*) i korelacije prediktora s kriterijem:

- Parcijalne korelacije jesu korelacije prediktora s kriterijem uz parcijaliziranje ostalih prediktora u regresijskoj jednadžbi i na prediktoru i na kriteriju. **Važno:** kvadriranjem koeficijenta parcijalne korelacije određenog prediktora dobivamo proporciju varijance kriterija koju objašnjava samo taj prediktor.
- Semiparcijalne korelacije jesu korelacije prediktora s kriterijem uz parcijaliziranje ostalih prediktora u regresijskoj jednadžbi na prediktoru. **Važno:** kvadriranjem koeficijenta semiparcijalne korelacije određenog prediktora dobivamo za koliko će se R^2 smanjiti ako uklonimo ovu varijablu iz regresijske jednadžbe.

Part And Partial Correlations ▼

Model		Partial	Part
2	SAMOSTAL	0.724	0.724
3	SAMOSTAL	0.695	0.545
	PLACA	0.576	0.397
4	SAMOSTAL	0.609	0.409
	PLACA	0.550	0.351
	RUTINA	-0.328	-0.185

Note. The intercept model is omitted, as no meaningful information can be shown.

7) Collinearity diagnostics. U tablici 'Coefficients' dodaje dvije kolone (Tolerance i Variance Inflation Factor – VIF) koje se koriste za provjeru multikolinearnosti. Dodaje i novu tablicu 'Collinearity diagnostics' koju trenutno možete zanemariti.

- Multikolinearnost u modelu znači da imamo visoke korelacije među prediktorima. Ako imamo dva izrazito povezana prediktora, moguće je da mjere isti ili gotovo isti konstrukt te stoga nam nisu potrebna oba prediktora. (npr. ako u modelu imate varijablu dobi mjerenu u godinama i ordinaliziranu varijablu dobi).
- Tolerance i VIF recipročne su mjere. Dovoljno je pogledamo jednu mjeru, npr. VIF. Ako je $VIF = 1$, nema multikolinearnosti. Ako je $VIF > 1$, postoji umjerena multikolinearnost. Ako je $VIF > 4$, postoji problematična multikolinearnost, doduše neki autori navode da se VIF može tolerirati do vrijednosti 5 ili 6. Rješenje je uklanjanje varijable (ili varijabli) iz regresijskog modela.
- Iako je riječ o pretpostavci, multikolinearnost se najčešće provjerava na punom regresijskom modelu (*full-model*), odnosno modelu sa svim prediktorima.

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	Collinearity Statistics	
							Tolerance	VIF
1	(Intercept)	10.969	0.405		27.093	< .001		
2	(Intercept)	2.815	1.019		2.763	0.007		
	SAMOSTAL	0.859	0.103	0.724	8.328	< .001	1.000	1.000
3	(Intercept)	-6.335	1.852		-3.420	0.001		
	SAMOSTAL	0.688	0.090	0.580	7.604	< .001	0.883	1.132
	PLACA	0.001	2.481e-4	0.422	5.542	< .001	0.883	1.132
4	(Intercept)	-1.930	2.397		-0.805	0.424		
	SAMOSTAL	0.574	0.096	0.484	5.992	< .001	0.714	1.400
	PLACA	0.001	2.413e-4	0.382	5.145	< .001	0.847	1.181
	RUTINA	-0.170	0.063	-0.219	-2.714	0.009	0.717	1.395

Note. The following covariate was considered but not included: DOB.

Kako višestrukom regresijskom analizom utvrđujemo:

1) Koji su prediktori statistički značajni?

- Provjerimo p-vrijednost pojedinog prediktora u punom modelu (testiranje t-omjera svakog prediktora u punom (posljednjem) modelu).

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	Collinearity Statistics	
							Tolerance	VIF
1	(Intercept)	10.969	0.405		27.093	< .001		
2	(Intercept)	2.815	1.019		2.763	0.007		
	SAMOSTAL	0.859	0.103	0.724	8.328	< .001	1.000	1.000
3	(Intercept)	-6.335	1.852		-3.420	0.001		
	SAMOSTAL	0.688	0.090	0.580	7.604	< .001	0.883	1.132
	PLACA	0.001	2.481e-4	0.422	5.542	< .001	0.883	1.132
4	(Intercept)	1.930	2.397		0.805	0.424		
	SAMOSTAL	0.574	0.096	0.484	5.992	< .001	0.714	1.400
	PLACA	0.001	2.413e-4	0.382	5.145	< .001	0.847	1.181
	RUTINA	-0.170	0.063	-0.219	-2.714	0.009	0.717	1.395

Note. The following covariate was considered but not included: DOB.

2) U kojoj mjeri svaki od prediktora doprinosi tumačenju varijance kriterija?

- Analiziramo kvadrirane vrijednosti semiparcijalne korelacije pojedinih prediktora ili analiziramo R² Change (proporcije) za svaki model nakon dodavanja/oduzimanja prediktora.

Part and Partial Correlations ▼

Model		Partial	Part
2	SAMOSTAL	0.724	0.724
3	SAMOSTAL	0.695	0.545
	PLACA	0.576	0.397
4	SAMOSTAL	0.609	0.409
	PLACA	0.550	0.351
	RUTINA	-0.328	-0.185

Note. The intercept model is omitted, as no meaningful information can be shown.

→ treba kvadrirati za odgovaranje na pitanje: Koliko će se R² smanjiti ako ovu varijablu uklonimo iz regresijske jednačbe?

Model Summary - ZADOVOS

Model	R	R ²	Adjusted R ²	RMSE	R ² Change	F Change	df1	df2	p
1	0.000	0.000	0.000	3.264	0.000		0	64	
2	0.724	0.524	0.516	2.270	0.524	69.357	1	63	< .001
3	0.826	0.682	0.671	1.871	0.158	30.714	1	62	< .001
4	0.846	0.716	0.702	1.702	0.034	7.364	1	61	0.009

Ovdje vidimo povećavanje R² u svakom koraku analize, pri tome moramo znati redoslijed prediktora koji su ulazili u model da bismo mogli interpretirati. Isti redoslijed vidimo na više mjesta, npr. iz tablice Part and Partial Correlations vidi se da je redoslijed bio sljedeći: SAMOSTAL, PLACA, RUTINA.

3) Koja proporcija varijance kriterija je protumačena linearnom kombinacijom svih prediktora?

- Provjerimo koeficijent multiple determinacije (R^2) za puni regresijski model (sugerira se koristiti korigirani, ali tada treba navesti da je korigiran).

Model Summary - ZADOVPOS

Model	R	R ²	Adjusted R ²	RMSE	R ² Change	F Change	df1	df2	p
1	0.000	0.000	0.000	3.264	0.000		0	64	
2	0.724	0.524	0.516	2.270	0.524	69.357	1	63	< .001
3	0.826	0.682	0.671	1.871	0.158	30.714	1	62	< .001
4	0.846	0.716	0.702	1.782	0.034	7.364	1	61	0.009

4) Koliki je učinak pojedinog prediktora na kriterij (dok su svi ostali prediktori konstantni)

- Provjerimo standardizirane parcijalne regresijske koeficijente (β) za svaki prediktor

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	Collinearity Statistics	
							Tolerance	VIF
1	(Intercept)	10.969	0.405		27.093	< .001		
2	(Intercept)	2.815	1.019		2.763	0.007		
2	SAMOSTAL	0.859	0.103	0.724	8.328	< .001	1.000	1.000
3	(Intercept)	-6.335	1.852		-3.420	0.001		
3	SAMOSTAL	0.688	0.090	0.580	7.604	< .001	0.883	1.132
3	PLACA	0.001	2.481e-4	0.422	5.542	< .001	0.883	1.132
4	(Intercept)	-1.930	2.397		-0.805	0.424		
4	SAMOSTAL	0.574	0.096	0.484	5.992	< .001	0.714	1.400
4	PLACA	0.001	2.413e-4	0.382	5.145	< .001	0.847	1.181
4	RUTINA	-0.170	0.063	-0.219	-2.714	0.009	0.717	1.395

Note. The following covariate was considered but not included: DOB.

5) Koliko procjenjujemo da će u prosjeku iznositi vrijednost kriterija za određenu kombinaciju vrijednosti prediktora?

- Koristimo jednadžbu regresijskog modela u nestandardiziranoj metrici, unosimo u nju željene vrijednosti pojedinih prediktora te izračunavamo očekivanu prosječnu vrijednost kriterija

Jednadžbu zapisujemo na temelju a i b koeficijenata iz punog modela:

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	Collinearity Statistics	
							Tolerance	VIF
1	(Intercept)	10.969	0.405		27.093	< .001		
2	(Intercept)	2.815	1.019		2.763	0.007		
2	SAMOSTAL	0.859	0.103	0.724	8.328	< .001	1.000	1.000
3	(Intercept)	-6.335	1.852		-3.420	0.001		
3	SAMOSTAL	0.688	0.090	0.580	7.604	< .001	0.883	1.132
3	PLACA	0.001	2.481e-4	0.422	5.542	< .001	0.883	1.132
4	(Intercept)	-1.930	2.397		-0.805	0.424		
4	SAMOSTAL	0.574	0.096	0.484	5.992	< .001	0.714	1.400
4	PLACA	0.001	2.413e-4	0.382	5.145	< .001	0.847	1.181
4	RUTINA	-0.170	0.063	-0.219	-2.714	0.009	0.717	1.395

Note. The following covariate was considered but not included: DOB.

$a = -1,930$
 $b_1 = 0,574$
 $b_2 = 0,001$
 $b_3 = -0,170$

Dakle, jednadžba ovog regresijskog modela glasi:

$$\text{ZADOVPOSL} = -1,930 + 0,574 \cdot \text{SAMOSTAL} + 0,001 \cdot \text{PLAĆA} - 0,170 \cdot \text{RUTINA}$$

Drugim riječima:

- Za jednu jedinicu porasta na skali procjene samostalnosti pri obavljanju radnih zadaća, očekuje se da će se rezultat na skali zadovoljstva poslom u prosjeku povećati za 0,574 jedinice, ako se plaća i rutina ne mijenjaju
- Za jednu kunu porasta plaće, očekuje se da će se rezultat na skali zadovoljstva poslom u prosjeku povećati za 0,001 jedinice, ako se samostalnost i rutina ne mijenjaju
- Za jednu jedinicu porasta na skali procjene rutinskog karaktera posla, očekuje se da će se rezultat na skali zadovoljstva poslom u prosjeku smanjiti za 0,170 jedinica, ako se samostalnost i plaća ne mijenjaju

Tu jednadžbu možemo iskoristiti za **predikciju rezultata kriterija uz bilo koju kombinaciju prediktora**. Npr. možemo odgovoriti na pitanje: *Koliki će biti (prosječni, najvjerojatniji) rezultat na skali zadovoljstva poslom za osobe koje na skali samostalnosti poslom imaju rezultat 10, na skali rutine rezultat 14 te imaju plaću 5 842 kn.*

$$\text{ZADOVPOSL} = -1,930 + 0,574 \cdot \text{SAMOSTAL} + 0,001 \cdot \text{PLAĆA} - 0,170 \cdot \text{RUTINA}$$

$$\text{ZADOVPOSL} = -1,930 + 0,574 \cdot 10 + 0,001 \cdot 5842 - 0,170 \cdot 14$$

$$\text{ZADOVPOSL} = -1,930 + 5,74 + 5,842 - 2,38$$

$$\text{ZADOVPOSL} = -1,930 + 5,74 + 5,842 - 2,38$$

$$\text{ZADOVPOSL} = 7,27$$

Odgovor: *Na temelju postavljenog regresijskog modela, za osobe koje na skali samostalnosti poslom imaju rezultat 10, na skali rutine rezultat 14 te imaju plaću 5 842 kn očekuje se da će u prosjeku na skali zadovoljstva poslom imati rezultat 7,27.*

Koraci za provedbu i izvještavanje o analizi – načelno:

- 1) Provjerite metodološke pretpostavke za provedbu višestruke regresijske analize
- 2) Navedite osnovne deskriptivne pokazatelje prediktora i kriterija
- 3) Navedite metodu izgradnje regresijskog modela
- 4) Testirajte značajnost regresijskog modela i navedite koliko je varijance kriterija protumačeno prediktorima
- 5) Navedite koji su prediktori značajni i koji prediktori najviše doprinose tumačenju kriterija (+ provjera multikolinearnosti)
- 6) Komentirajte koji prediktori nisu ušli u regresijski model
- 7) Navedite regresijsku jednadžbu te iz nje interpretirajte nestandardizirane parcijalne regresijske koeficijente
- 8) Interpretirajte standardizirane parcijalne regresijske koeficijente. Kontekstualizirajte nalaze.

Napomena: Ne zaboravite navesti ključne parametre (F-omjer, p-vrijednost, itd.)

Regresijska jednadžaba:

$$\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_kX_k$$

12.3. Interpretacija dobivenog regresijskog modela

Za ovaj zadatak, na temelju svih dobivenih rezultata možemo napisati **sljedeću interpretaciju**:

- 3) Višestruka regresijska analiza provedena *Forward* metodom rezultirala je s tri statistički značajna prediktora za tumačenje zadovoljstva poslom: procjena samostalnosti pri obavljanju radnih zadaća, iznos plaće i procjena rutinskog karaktera posla.
- 4) Dobiven je statistički značajan regresijski model [$F_{(3,61)} = 51,26$; $p < 0,001$] u kojem je $R^2 = 0,716$. Drugim riječima, ova tri prediktora zajedno objašnjavaju 71,6% varijance zadovoljstva poslom.
- 5) Najveći doprinos tumačenju kriterija ima samostalnost pri obavljanju radnih zadaća, potom plaća, a najmanje rutinski karakter posla (kvadrirane semiparcijalne korelacije = udjeli doprinosa pojedinog prediktora). Nije uočena problematična multikolinearnost među prediktorima (VIF-vrijednosti u rasponu od 1,181 do 1,400).
- 6) Varijabla *dob* nije ušla u regresijski model, što znači da ne pridonosi tumačenju zadovoljstva poslom.
- 7) Regresijska jednadžba glasi:
zadovposl = $-1,930 + 0,574 \cdot \text{samostal} + 0,001 \cdot \text{plaća} - 0,170 \cdot \text{rutina}$

Drugim riječima:

- Za jednu jedinicu porasta na skali procjene samostalnosti pri obavljanju radnih zadaća, očekuje se da će se rezultat na skali zadovoljstva poslom u prosjeku povećati za 0,574 jedinice, ako se plaća i rutina ne mijenjaju.
 - Za jednu kunu porasta plaće, očekuje se da će se rezultat na skali zadovoljstva poslom u prosjeku povećati za 0,001 jedinice, ako se samostalnost i rutina ne mijenjaju.
 - Za jednu jedinicu porasta na skali procjene rutinskog karaktera posla, očekuje se da će se rezultat na skali zadovoljstva poslom u prosjeku smanjiti za 0,170 jedinica, ako se samostalnost i plaća ne mijenjaju.
- 8) Najveći učinak na zadovoljstvo poslom ima samostalnost u obavljanju radnih zadataka ($\beta = 0,484$; $p < 0,001$) i to umjerene jačine i pozitivnog smjera, potom plaća (također umjerene jačine i pozitivnog smjera, $\beta = 0,382$; $p < 0,001$), a najmanji rutina posla čiji je učinak slabi i negativan ($\beta = -0,219$; $p = 0,009$), ali ipak statistički značajan.

Interval pouzdanosti za regresijski koeficijent

Koristan dodatak za interpretaciju jest i interval pouzdanosti za regresijski koeficijent, čije određivanje možemo zatražiti pod *Statistics*, opcijom '*Confidence intervals*'.

Regression Coefficients

Estimates Model fit

From 5000 bootstraps R squared change

Confidence intervals 95.0 % Descriptives

Covariance matrix Part and partial correlations

Vovk-Sellke maximum p-ratio Collinearity diagnostics

Coefficients ▼

Model		Unstandardized	Standard Error	Standardized	t	p	95% CI	
							Lower	Upper
1	(Intercept)	10.969	0.405		27.093	< .001	10.160	11.778
2	(Intercept)	2.815	1.019		2.703	0.007	0.779	4.851
	SAMOSTAL	0.059	0.103	0.724	0.320	< .001	0.653	1.065
3	(Intercept)	-6.335	1.852		-3.420	0.001	-10.038	-2.632
	SAMOSTAL	0.688	0.090	0.580	7.604	< .001	0.507	0.869
	PLACA	0.001	2.481e-4	0.422	5.542	< .001	8.791e-4	0.002
4	(Intercept)	1.930	2.397		0.805	0.424	6.723	2.064
	SAMOSTAL	0.574	0.096	0.484	5.992	< .001	0.382	0.766
	PLACA	0.001	2.413e-4	0.382	5.145	< .001	7.592e-4	0.002
	RUTINA	-0.170	0.063	-0.219	-2.714	0.009	-0.295	-0.045

Note: The following covariate was considered but not included: DOB.

Primjer interpretacije za SAMOSTAL1:

Uz 95% pouzdanosti zaključujemo da regresijski koeficijent varijable SAMOSTAL u populaciji iznosi između 0,382 i 0,766.

Literatura

- Goss-Sampson, M. A. (2018). *Statistical Analysis in JASP: a Guide for Students*. Multiple regression (str. 59-65).
URL: <https://static.jasp-stats.org/Statistical%20Analysis%20in%20JASP%20-%20A%20Students%20Guide%20v2.pdf>
- Goss-Sampson, M. A. (2019). *Statistička analiza u JASP programu: vodič za studente*. Višestruka regresija (str. 69-74).
URL: http://static.jasp-stats.org/Manuals/Statistic%cc%8cka_analiza_u_JASP_programu_v0.10.2.pdf
- Navarro, D.J., Foxcroft, D.R., i Faulkenberry, T.J. (2019). *Learning Statistics with JASP: A Tutorial for Psychology Students and Other Beginners*. Poglavlje 11. Correlation and linear regression, posebice 11.5. Multiple linear regression.
URL: <https://tomfaulkenberry.github.io/JASPbook/chapters/chapter11.pdf>