

**Vežbanje za veštačke promenljive**

**Uvodne napomene**

U ovom vežbanju nastavljamo sa istraživanjem zarada zaposlenih lica u Srbiji u 2015. godini na osnovu podatka Ankete o dohotku i uslovima života dobijenih od strane RZS-a. Zarade zaposlenih lica zavise od brojnih karakteristika lica i zaposlenja kao što su obrazovanje, pol, godine radnog iskustva, sektor, zanimanje, delatnost, region, tip naselja, itd. Radi jednostavnosti posmatraćemo samo neke od promenljivih koje objašnjavaju zarade. Fokusiraćemo se na tumačenje veštačkih promenljivih, a testiranje značajnosti parametara korišćenjem t testa bi trebalo da znate sami da uradite na osnovu prethodnih vežbanja. Podsetimo se da kritična vrednost za t test za uzorke velikog obima na nivou značajnosti od 0,05 iznosi 1,96. Označićemo sa  $Y$  zaradu po času, a  $Y^* = \ln Y$ .

**Zadaci sa rešenjima**

1. Ocenjen je model zavisnosti zarada u Srbiji od tipa sektora u kojem rade zaposleni (javni ili privatni):

Tabela 1 Ocenjeni regresioni model zavisnosti zarada po času (ln) od tipa sektora (javni sektor=1/privatni sektor=0)

Zarada po času (ln)	Ocena	Standardna greška ocene	t-odnos
Konstanta	4,933	0,012	425,65
Sektor	0,378	0,018	21,28
Broj opservacija n	3.679		
R2	0,1096		
Korigovani R2	0,1094		

odnosno

$$\widehat{Y}^* = 4,933 + 0,378V$$

(0,018)

gde je  $V$  veštačka promenljiva koja se definiše prema:

$$V = \begin{cases} 1 & \text{za javni sektor} \\ 0 & \text{za privatni sektor} \end{cases}$$

- a) Da li su u proseku zarade u javnom sektoru veće nego zarade u privatnom sektoru?
- b) Koliko iznose prosečne zarade u javnom i privatnom sektoru?
- c) Ocenjen je novi model zarada u tabeli 1a, koji sadrži godine radnog iskustva kao dodatnu objašnjavajuću promenljivu (oznaka  $X_1$ ):

Tabela 1a Ocenjeni regresioni model zavisnosti zarada po času (ln) od godina radnog iskustva i tipa sektora (javni sektor=1/privatni sektor=0)

Zarada po času (ln)	Ocena	Standardna greška ocene	t-odnos
Konstanta	4,883	0,017	283,56
Godine radnog iskustva	0,003	0,001	3,89
Sektor	0,362	0,018	19,9
Broj opservacija n	3.679		
R2	0,1133		
Korigovani R2	0,1128		

$$\widehat{Y}^* = 4,883 + 0,003X_1 + 0,362V$$

$$(0,001) \quad (0,018)$$

Napisati jednačine zarada. Kako objašnjavate da je ocenjeni koeficijent za veštačku promenljivu V (tip sektora) u tabeli 1a manji nego u tabeli 1? Koji od dva modela je kvalitetniji?

### Rešenje:

- Zarade su u proseku veće za 37,8% u javnom sektoru u odnosu na privatni sektor. Razlika u zaradama između javnog i privatnog sektora se u literaturi naziva premija zarada javnog sektora.
  - Ocenjena jednačina zarada u privatnom sektoru će biti  $\widehat{Y}^* = 4,933$ , dok će u javnom sektoru biti  $\widehat{Y}^* = 4,933 + 0,378 = 5,311$ . Ovi iznosi su zapravo ocenjene prosečne vrednosti ln zarada u javnom i privatnom sektoru. Pošto je u pitanju logaritamsko linearni model, prosečna zarada po času u privatnom sektoru je iznosila  $e^{4,933} = 138,8$  RSD, dok je u javnom sektoru iznosila  $e^{5,311} = 202,6$  RSD.
  - Jednačina za privatni sektor je  $\widehat{Y}^* = 4,883 + 0,003X_1$ . Jednačina za javni sektor je  $\widehat{Y}^* = 5,245 + 0,003X_1$ . Za isti broj godina radnog iskustva (za dato  $X_1$ ) prosečne zarade u javnom sektoru su za 36,2% veće nego u privatnom sektoru. Poređenje prosečnih zarada u javnom i privatnom sektoru (tabela 1) nije „prava“ mera razlike u zaradama između dva sektora, jer ne uzima u obzir da pored sektora postoje i brojne druge determinante zarada. Preciznije, ta ocena je pristrasna ako se izostavi relevantna objašnjavajuća promenljiva, o čemu će biti više reči na narednim predavanjima. U konkretnoj analizi, uvođenjem dodatnih objašnjavajućih promenljivih razlika u zaradama između sektora može da se smanji ili poveća. S obzirom na to da imamo različit broj objašnjavajućih promenljivih, za poređenje dva modela koristimo korigovani koeficijent determinacije. Korigovani koeficijent determinacije u modelu sa jednom objašnjavajućom promenljivom (sektor) iznosi 0,1094, dok u modelu sa dve objašnjavajuće promenljive (sektor i godine radnog iskustva) iznosi 0,1128. Na osnovu vrednosti korigovanog koeficijenta determinacije možemo da zaključimo da model iz tabele 1a bolje objašnjava zarade zaposlenih lica.
2. Ocenjen je nov model zarada u tabeli 2, koji pored godina radnog iskustva uključuje informaciju o uticaju nivoa obrazovanja zaposlenih na nivo zarada. Nivo obrazovanja posmatra se kroz tri kategorije: osnovno, srednje i visoko obrazovanje. Da bismo ispitali ovaj uticaj definišemo dve veštačke promenljive (za jednu manje od tri kategorije) na sledeći način:

$$O_2 = \begin{cases} 1 & \text{za srednje obrazovanje} \\ 0 & \text{za ostalo obrazovanje} \end{cases}$$

$$O_3 = \begin{cases} 1 & \text{za visoko obrazovanje} \\ 0 & \text{za ostalo obrazovanje} \end{cases}$$

Ovakvim izborom veštačkih promenljivih kategorija osnovno obrazovanje postaje tzv. bazna ili referentna kategorija u odnosu na koju se interpretiraju dobijeni rezultati.

Tabela 2 Ocenjeni regresioni model zavisnosti zarada po času (ln) od godina radnog iskustva i obrazovnog nivoa

Zarada po času (ln)	Ocena	Standardna greška ocene	t-odnos
Konstanta	4,586	0,032	144,45
Godine radnog iskustva	0,008	0,001	10,67
Srednje obrazovanje	0,249	0,030	8,34
Visoko obrazovanje	0,705	0,032	22,15
Broj opservacija n	3.679		
R2	0,1893		
Korigovani R2	0,1886		

ili u regresionoj formi:

$$\widehat{Y}^* = 4,586 + 0,008X_1 + 0,249O_2 + 0,705O_3$$

(0,001) (0,030) (0,032)

- Protumačiti ocenjene koeficijente uz veštačke promenljive.
- Napisati ocenjene funkcije zarada za sva tri nivoa obrazovanja. Koliko iznosi prosečna zarada za osnovno, srednje i visoko obrazovanje na osnovu datih podataka?
- Koji je od dva modela predstavljenih u tabeli 1a i tabeli 2 kvalitetniji?

**Rešenje:**

- Lica sa srednjim nivoom obrazovanja imaju u proseku 24,9% veće zarade u odnosu na lica sa osnovnim obrazovanjem, dok lica sa visokim obrazovanjem imaju 70,5% veće zarade u odnosu na lica sa osnovnim obrazovanjem, pri čemu su godine radnog iskustva konstantne.
- Jednačina zarada za osnovno obrazovanje je  $\widehat{Y}^* = 4,586 + 0,008X_1$ , za srednje obrazovanje je  $\widehat{Y}^* = 4,835 + 0,008X_1$ , a za visoko obrazovanje je  $\widehat{Y}^* = 5,291 + 0,008X_1$ . Prosečna ln zarada za lica sa osnovnim obrazovanjem iznosi 4,586, za lica sa srednjim nivoom obrazovanja iznosi 4,835 (4,586+0,249), za lica sa visokim nivoom obrazovanja iznosi 5,291 (4,856+0,705). Odnosno, antilogaritmovanjem dobijemo da je prosečna zarada po času za osnovno, srednje i visoko obrazovanje  $e^{4,586} = 98$  RSD,  $e^{4,835} = 126$  RSD i  $e^{5,291} = 199$  RSD, respektivno, za isti broj godina radnog iskustva ( $X_1$  je konstantno).
- Na osnovu korigovanog koeficijenta determinacije vidimo da godine radnog iskustva i obrazovanje bolje objašnjavaju zarade zaposlenih lica u odnosu na godine radnog iskustva i sektora.

3. Potrebno je proveriti da li se prosečan nivo zarada značajno razlikuju prema regionima u Srbiji. Regioni su sledeći: Beograd, Vojvodina, Šumadija i Zapadna Srbija, i Južna i Istočna Srbija. Ako se pretpostavi da region Beograda predstavlja referentnu kategoriju, predložiti uvođenje odgovarajućih veštačkih promenljivih. Napisati funkciju zarada za svaki od regiona.

**Rešenje:**

Veštačke promenljive definišemo na sledeći način:

$$REG_2 = \begin{cases} 1 & \text{za region Vojvodine} \\ 0 & \text{za ostale regione} \end{cases}$$

$$REG_3 = \begin{cases} 1 & \text{za region Šumadije i Zapadne Srbije} \\ 0 & \text{za ostale regione} \end{cases}$$

$$REG_4 = \begin{cases} 1 & \text{za region Južne i Istočne Srbije} \\ 0 & \text{za ostale regione} \end{cases}$$

Ocenjujemo funkciju oblika:

$$Y^* = \beta_0 + \beta_1 REG_2 + \beta_2 REG_3 + \beta_3 REG_4 + \varepsilon$$

Ocenjena jednačine zarada za regione će biti:

$$\text{Beograd: } \widehat{Y}^* = b_0$$

$$\text{Vojvodina: } \widehat{Y}^* = b_0 + b_1$$

$$\text{Šumadija i Zapadna Srbija } \widehat{Y}^* = b_0 + b_2$$

$$\text{Južna i Istočna Srbija } \widehat{Y}^* = b_0 + b_3$$

### **Zaključak, komentar i tema za razmišljanje**

Ovo vežbanje ima za cilj da Vam objasni korišćenje i tumačenje veštačkih promenljivih u regresionoj analizi. Od čega zavise zarade zaposlenih lica je tema koja je puno puta razmatrana u empirijskim istraživanjima. Takođe, uvek je aktuelna tema da li je opravdano da zarade u javnom sektoru budu veće nego zarade u privatnom sektoru. U prvom primeru smo pretpostavili da jedina determinanta zarada je da li lice radi u javnom ili privatnom sektoru i dobili smo da je tzv. premija javnog sektora 37,8%. Međutim, zarade zavise i od brojnih drugih karakteristika, među kojima je sigurno jedna od bitnijih karakteristika obrazovni nivo. U javnom sektoru je učešće lica sa visokim nivoom obrazovanja oko 60%, dok je u privatnom sektoru oko 30% zaposlenih lica sa visokim nivoom obrazovanja. Na primer, u javnom sektoru rade lekari, sudije, tužioci, nastavnici koji imaju visok nivo obrazovanja, a lica sa visokim nivoom obrazovanja bi trebalo da rade za veću zaradu nego lica sa srednjim ili osnovnim nivoom obrazovanja.

Kada uzmemo u obzir sve objektivne karakteristike lica i posla (obrazovanje, radno iskustvo, pol, zanimanje, delatnost, itd.) dobijemo da je premija zarada u javnom sektoru 16,8%. Stvarna premija javnog sektora je zapravo dosta manja od premije zarada u modelu sa samo jednom objašnjavajućom promenljivom (veštačkom za javni/privatni sektor). To sugerise da se pouzdani ekonomski zaključci mogu izvesti samo na osnovu kvalitetnog ekonometrijskog modela koji obuhvata sve relevantne uticaje.