**Vežbanje za testiranje značajnosti parametara, za utvrđivanje postojanja multikolinearnosti i testiranje opravdanosti isključivanja regresora**

Na osnovu podataka Ankete o dohotku i uslovima života za Srbiju u 2015. godini ocenjena je veza logaritma zarada po času zaposlenih lica, godina starosti i godina radnog iskustva (Tabela 1), logaritma zarada po času i godina radnog iskustva (Tabela 2), logaritma zarada po času i godina starosti (Tabela 3). Zarade zaposlenih lica zavise od brojnih karakteristika lica i karakteristika zaposlenja, ali ćemo radi vežbanja pojednostaviti i za sada pretpostaviti da zavise samo od godina starosti i/ili od godina radnog iskustva. Podaci su dobijeni od RZS-a.

Tabela 1 Ocenjeni regresioni model zavisnosti zarada po času (ln) od godina radnog iskustva i godina starosti

|  |  |  |
| --- | --- | --- |
| Zarada po času (ln) | Ocena | Standardnagreška ocene |
| Konstanta | 4,927 | 0,051 |
| Godine radnog iskustva | 0,005 | 0,002 |
| Godine starosti | 0,002 | 0,002 |
| Broj opservacija n | 3.773 |  |
| $$\sum\_{}^{}e^{2}$$ | 1.180,73 |  |

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

|  |  |  |
| --- | --- | --- |
| Zarada po času (ln) | Ocena | Standardnagreška ocene |
| Konstanta | 4,971 | 0,017 |
| Godine radnog iskustva | 0,007 | 0,001 |
| Broj opservacija n | 3.773 |  |
| $$\sum\_{}^{}e^{2}$$ | 1.181,0 |  |

Tabela 3 Ocenjeni regresioni model zavisnosti zarada po času (ln) od godina starosti

|  |  |  |
| --- | --- | --- |
| Zarada po času (ln) | Ocena | Standardnagreška ocene |
| Konstanta | 4,820 | 0,037 |
| Godine starosti | 0,007 | 0,001 |
| Broj opservacija n | 3.773 |  |
| $$\sum\_{}^{}e^{2}$$ | 1.183,51 |  |

1. Prikazane rezultate zapisati u formi regresionih modela.
2. Testirati da li objašnjavajuće promenljive pojedinačno imaju statistički značajan uticaj na zavisnu promenljivu (Tabela 1).
3. Testirati da li objašnjavajuća promenljiva ima statistički značajan uticaj na zavisnu promenljivu (Tabela 2).
4. Testirati da li objašnjavajuća promenljiva ima statistički značajan uticaj na zavisnu promenljivu (Tabela 3).
5. Koji može biti razlog zašto su godine starosti značajne u regresiji u kojoj je to jedina objašnjavajuća promenljiva, a nisu značajne u regresiji gde imamo i godine radnog iskustva (Tabela 1 i Tabela 3)?
6. Izračunati faktor rasta varijanse na osnovu datog koeficijenta korelacije (Tabela 4).

Tabela 4 Koeficijent korelacije

|  |  |  |
| --- | --- | --- |
|  | Godine radnog iskustva | Godine starosti |
| Godine radnog iskustva | 1 |  |
| Godine starosti | 0,8851 | 1 |

1. Testirati F testom da godine starosti ne utiču statistički značajno na zarade (Tabela 1 i Tabela 2).
2. Protumačiti ocenu parametra za godine radnog iskustva (Tabela 2).

**Rešenja**

Označićemo zarade po času sa $Y$, $Y^{\*}=lnY$, godine radnog iskustva sa $X\_{1}$, a godine starosti sa $X\_{2}$.

1. Ocenjeni regresioni model zavisnosti zarada po času (ln) od godina radnog iskustva i godina starosti (Tabela 1).

$$\hat{Y^{\*}}=4,927+0,005X\_{1}+0,002X\_{2}$$

$$ \left(0,002\right) \left(0,002\right)$$

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

$$\hat{Y^{\*}}=4,971+0,007X\_{1}$$

$$ \left(0,001\right) $$

Ocenjeni regresioni model zavisnosti zarada po času (ln) od godina starosti (Tabela 3).

$$\hat{Y^{\*}}=4,820+0,007X\_{2}$$

$$ \left(0,001\right) $$

1. Definišemo nultu hipotezu da godine radnog iskustva ne utiču statistički značajno na zarade i da godine starosti ne utiču statistički značajno na zarade, $H\_{0}:β\_{1}=0$ i $H\_{0}:β\_{2}=0$. Alternativne hipoteze su redom $H\_{1}:β\_{1}\ne 0$ i $H\_{1}:β\_{2}\ne 0$.

$t\_{1}=\frac{b\_{1}}{s\_{b\_{1}}}=2,5$, $t\_{2}=\frac{b\_{2}}{s\_{b\_{2}}}=1$, kritična vrednost za n-k (3.773-3=3.770) i $α=0,05$ je 1,96. Možemo zaključiti da godine radnog iskustva imaju statistički značajan uticaj na zarade (odbacuje se nulta hipoteza), dok godine starosti ne utiču statistički značajno na zarade (ne odbacuje se nulta hipoteza).

1. Definišemo nultu hipotezu da godine radnog iskustva ne utiču statistički značajno na zarade $H\_{0}:β\_{1}=0$. Alternativna hipoteza je $H\_{1}:β\_{1}\ne 0$

$t\_{1}=\frac{b\_{1}}{s\_{b\_{1}}}=7$. Možemo zaključiti da godine radnog iskustva imaju statistički značajan uticaj na zarade. Ovde je odgovarajući broj stepeni slobode 3771, ali se koristi ista kritična vrednost kao u delu 2. zbog velikog obima uzorka. Preciznije, za uzorak ovako velikog obima *t* raspodela se jako dobro aproksimira standardizovanom normalnom raspodelom, za koju je 5% kritična vrednost dvostranog testa 1,96.

1. Definišemo nultu hipotezu da godine starosti ne utiču statistički značajno na zarade $H\_{0}:β\_{2}=0$. Alternativna hipoteza je $H\_{1}:β\_{2}\ne 0$.

$t\_{2}=\frac{b\_{2}}{s\_{b\_{2}}}=7$. Možemo zaključiti da godine starosti imaju statistički značajan uticaj na zarade.

1. Možemo da pretpostavimo da su godine starosti i godine radnog iskustva visoko korelisane. Starija lica u proseku imaju više godina radnog iskustva. Tabela 4 nam potvrđuje visoku korelaciju datih promenljivih.
2. $FRV=\frac{1}{1-0,8851^{2}}=4,62$. Relativno visoka vrednost faktora rasta varijanse nam ukazuje na problem multikolinearnosti.
3. Poredimo model sa i bez ograničenja na parametre, odnosno da li je opravdano isključiti godine starosti iz polaznog modela. Tabela 1 prikazuje model bez ograničenja. Tabela 2 prikazuje model sa ograničenjem, tj. model iz koga je isključena promenljiva godine starosti.

$$H\_{0}:β\_{2}=0$$

$$H\_{1}:β\_{2}\ne 0$$

$$F=\frac{{\left(\sum\_{}^{}e\_{0}^{2}-\sum\_{}^{}e\_{b}^{2}\right)}/{g}}{{\sum\_{}^{}e\_{b}^{2}}/{(n-k\_{b})}}$$

$g=1$, $n-k\_{b}=3.773-3=3.770$, $\sum\_{}^{}e\_{0}^{2}= 1.181,0$, $\sum\_{}^{}e\_{b}^{2}=1.180,73$.

$F=\frac{{\left(1.181,0-1.180,73\right)}/{1}}{{1.180,73}/{3.770}}=0,862$ < $F\_{3.770 (0,05)}^{1}=3,84$

$H\_{0}$ se ne odbacuje, odnosno zaključujemo da je opravdano isključiti godine starosti iz polaznog modela.

1. Budući da je reč o logaritamsko – linearnom modelu, dobijena ocena sugeriše da dodatna godina radnog iskustva povećava zarade po času za 0,7%.