Neki jednostavni
nelinearni regresijski modeli (2. dio)

Kristina Matijević, Požega
Bojan Kovačić, Zagreb

2. Model jednostavne logaritamske regresije

Standardni oblik regresijske jednadžbe modela jednostavne logaritamske regresije je:

$$\hat{y} = a \cdot \ln x + b, \tag{1}$$

gdje su $a \neq 0$ i $b \in \mathbf{R}$ realni parametri. Pritom vrijednosti varijable x nužno moraju biti strogo pozitivne jer je prirodno područje definicije bilo koje logaritamske funkcije skup $\mathbf{R}^+ := \langle 0, +\infty \rangle$.

Primijetimo da bismo za $a = 0$ dobili

$$\hat{y} = b, \tag{2}$$

tj. konstantnu funkciju. To je poseban slučaj modela jednostavne linearne regresije koji se u praksi vrlo rijetko primjenjuje, pa zbog toga ima smisla pretpostaviti da je $a \neq 0$.

U statističkoj analizi regresijskoga modela bitno je interpretirati značenja osnovnih parametara modela, kao i pokazatelje temeljem kojih se utvrđuje reprezentativnost modela, tj. "kvaliteta" opisa veze promatranih varijabli pomoću dotičnoga modela. Stoga ćemo najprije interpretirati značenje parametara a i b.

Iz (1) lako slijedi:

$$\hat{y} = b \text{ ako i samo ako je } x = 1. \tag{3}$$

Stoga parametar b interpretiramo kao očekivanu vrijednost varijable y za $x = 1$.

Interpretacija varijable a je posredna i temelji se na:

Propozicija 1. Neka je $p > -100$ proizvoljan, ali fiksiran realan broj. Ako se vrijednost varijable x promijeni za $p\%$, očekivana prosječna apsolutna promjena vrijednosti varijable y iznosi

$$\Delta\hat{y} = a \cdot \ln\left(1 + \frac{p}{100}\right). \tag{4}$$

Posebno, za $p = 171.8281828459045235360 3$ očekivana prosječna apsolutna promjena vrijednosti varijable y iznosi točno a.

Napomena 1. Uvjet $p > -100$ postavljen je zbog zahtjeva da vrijednost varijable x mora biti strogo pozitivna. Smanjenje konkretne vrijednosti varijable x za (barem) 100% povlačilo bi da je nova vrijednost te varijable nepozitivna (strogo negativna ili nula), a na takve vrijednosti nije moguće primijeniti ovaj regresijski model.

Napomena 2. Prigodom rješavanja "praktičnih" zadataka obično se uzima $p = 1$, tj. promatra se povećanje vrijednosti nezavisne varijable x za 1%.

Dokaz. Neka je x_1 početna vrijednost varijable x i neka je p postotak relativne promjene te varijable. Ako je $p > 0$, riječ je o povećanju vrijednosti x_1 za $p\%$, a ako je $p < 0$, riječ je o smanjenju vrijednosti x_1 za $p\%$. Pripadna očekivana vrijednost varijable y jednaka je:

$$\hat{y}_1 = a \cdot \ln x_1 + b. \tag{5}$$

Nakon promjene vrijednosti x_1 za $p\%$, nova vrijednost varijable x jednaka je

$$x_2 = x_1 + \frac{p}{100} \cdot x_1 = \left(1 + \frac{p}{100}\right) \cdot x_1. \tag{6}$$

Stoga je pripadna očekivana vrijednost varijable y:

$$\begin{aligned} \hat{y}_2 &= a \cdot \ln x_2 + b \\ &= a \cdot \ln\left[\left(1 + \frac{p}{100}\right) \cdot x_1\right] + b \\ &= a \cdot \ln\left(1 + \frac{p}{100}\right) + a \cdot \ln x_1 + b \\ &= a \cdot \ln\left(1 + \frac{p}{100}\right) + \hat{y}_1. \end{aligned} \tag{7}$$

Odatle slijedi da je apsolutna promjena vrijednosti varijable y jednaka

$$\hat{y}_2 - \hat{y}_1 = a \cdot \ln\left(1 + \frac{p}{100}\right),$$

što je i trebalo pokazati. ■

Napomena 3. U iskazu Propozicije 1. se navodi prilog *prosječno* jer se u izvodu formule za izračun vrijednosti parametra a pomoću emipirijskih vrijednosti varijabli x i y koriste prosjeci tih empirijskih vrijednosti, pa je i sâm parametar a pokazatelj prosječne promjene. Detalji se mogu naći u [1].

Pokažimo primjenu ovoga modela na primjeru.

Primjer 1. Na uzorku od 10 slučajno izabranih obitelji učenika Srednje ekonomske škole "Eustahije Brzić" iz Piškorevaca promatra se zavisnost prosjeka mjesečnih troškova na djecu o prosjeku ukupnih mjesečnih primanja obaju roditelja. (Svi prosjeci izračunani su za protekla tri mjeseca.) Dobiveni podatci prikazani su u donjoj tablici.

prosjek ukupnih mjesečnih primanja [000 kn]	prosjek ukupnih mjesečnih izdvajanja za djecu [000 kn]
4.6	0.82
5.2	0.91
5.7	0.95
6.5	0.97
7.3	1.01
7.9	1.08
8.4	1.12
9.1	1.15
9.8	1.19
10.1	1.22

a) Prikažite zavisnost ukupnoga mjesečnoga izdvajanja za djecu o ukupnim mjesečnim primanjima roditelja odgovarajućim grafikonom. Uz grafikon navedite sve potrebne oznake.

b) Odredite jednadžbu modela jednostavne logaritamske regresije koji najbolje opisuje promatranu zavisnost. Objasnite značenje svih parametara dobivenoga modela.

c) Procijenite reprezentativnost dobivenoga modela pomoću koeficijenta determinacije.

Na temelju rezultata b) podzadatka procijenite:

d) veličinu i smjer promjene prosjeka ukupnih izdvajanja za djecu ako se prosjek ukupnih mjesečnih primanja obitelji smanji za 10%;

e) prosjek ukupnih izdvajanja za djecu u obitelji čiji je prosjek mjesečnih primanja 7000 kn;

f) prosjek ukupnih mjesečnih primanja obitelji u kojoj je prosjek ukupnih izdvajanja za djecu 1050 kn;

g) najmanji prosjek ukupnih mjesečnih primanja roditelja potreban da se za djecu izdvaja prosječno ukupno najmanje 1000 kn.

Riješimo postavljene zadatke.

a) Zavisnost ukupnoga mjesečnoga izdvajanja za djecu o ukupnim mjesečnim primanjima roditelja grafički prikazujemo dijagramom rasipanja (slika 1). Taj dijagram konstruiramo pomoću MS Excela. Podrobniji opis izostavljamo, a detalji se mogu naći u [3].

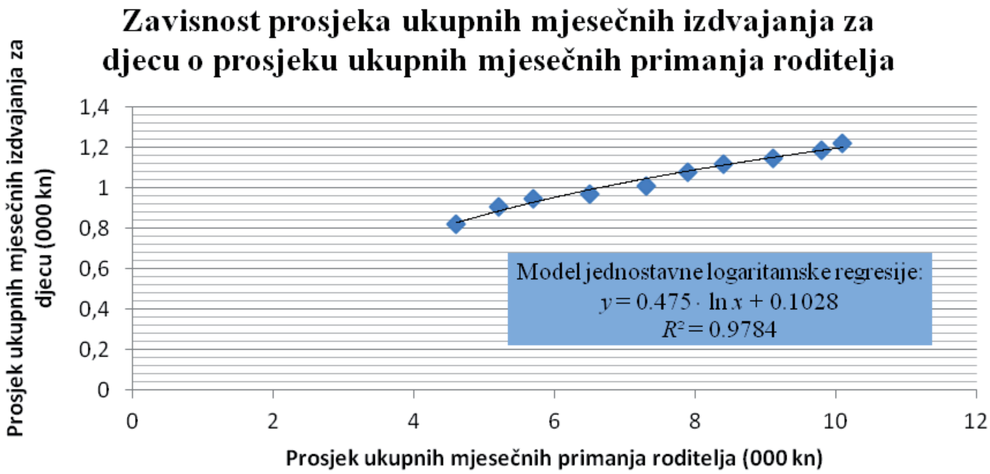
b) Jednadžba modela jednostavne logaritamske regresije, zajedno s pripadnom regresijskom krivuljom, navedena je na slici 2. Podrobniji opis dobivanja ove jednadžbe izostavljamo, a detalji se mogu naći u [3].

Parametar $a = 0.475$ interpretiramo tako da u Propoziciju 1. uvrstimo $a = 0.475$ i $p = 1$. Dakle, ako se prosjek ukupnih mjesečnih primanja roditelja poveća za 1%, onda očekivano prosječno apsolutno povećanje prosjeka ukupnih mjesečnih izdvajanja za djecu iznosi

$$\begin{aligned}\Delta y &= 0.475 \cdot \ln\left(1 + \frac{1}{100}\right) \\ &\approx 0.0047264 \text{ tisuće kn } \approx 4.73 \text{ kn.}\end{aligned} \tag{9}$$

Parametar $b = 0.1028$ znači da ako prosjek ukupnih mjesečnih primanja roditelja iznosi 1000.00 kn, onda očekivani prosjek ukupnih mjesečnih izdva-

Slika 1. Dijagram rasipanja za Primjer 1.



Slika 2. Model jednostavne logaritamske regresije za Primjer 1. i pripadna regresijska krivulja

janja za djecu iznosi 0.1028 tisuća kuna = 102.80 kn. (Čini li vam se ova interpretacija realnom?)

c) Koeficijent determinacije dobivenoga modela iznosi $R^2 = 0.9784$. Dakle, približno 97.84% zavisnosti prosjeka ukupnih mjesečnih izdvajanja za djecu o prosjeku ukupnih mjesečnih primanja roditelja objašnjeno je modelom jednostavne logaritamske regresije, pa možemo zaključiti da je model vrlo reprezentativan.

d) U rješenju ovoga zadatka primijenit ćemo Propoziciju 1. U jednakost (1) uvrstimo $a = 0.475$ i $p = -10$, pa dobivamo:

$$\begin{aligned}\Delta\hat{y} &= 0.475 \cdot \ln\left(1 - \frac{10}{100}\right)\\ &= 0.475 \cdot \ln 0.9 \approx -0.05005.\end{aligned} \tag{10}$$

Dakle, ako se prosjek ukupnih mjesečnih primanja obitelji smanji za 10%, onda će se prosjek ukupnih mjesečnih izdvajanja za djecu očekivano *smanjiti* za prosječno 50.05 kn.

e) Traženi prosjek izračunat ćemo tako da u jednadžbu dobivenoga modela jednostavne logaritamske regresije uvrstimo $x = 7$ (uz oprez s novčanim jedinicama!). Tako dobivamo:

$$\begin{aligned}\hat{y} &= 0.475 \cdot \ln 7 + 0.1028\\ &\approx 1.027107 \text{ tisuća kuna}\\ &\approx 1027.11 \text{ kn.}\end{aligned} \tag{11}$$

f) Traženi prosjek izračunat ćemo tako da u jednadžbu dobivenoga modela jednostavne logaritamske regresije uvrstimo $y = 1.05$ (uz ponovni oprez s novčanim jedinicama). Tako dobivamo logaritamsku jednadžbu:

$$1.05 = 0.475 \cdot \ln x + 0.1028. \tag{12}$$

Odatle lagano slijedi

$$\begin{aligned}x &= e^{\frac{4736}{2375}}\\ &\approx 7.34563 \text{ tisuća kuna}\\ &= 7345.63 \text{ kn.}\end{aligned} \tag{13}$$

g) Najprije ćemo riješiti nejednadžbu $y \geq 1$ po nepoznanici x. Imamo redom:

$$\begin{aligned}0.475 \cdot \ln x + 0.1028 &\geq 1,\\ 0.475 \cdot \ln x &\geq 1 - 0.1028,\\ 0.475 \cdot \ln x &\geq 0.8972 \quad / : 0.475,\\ \ln x &\geq \frac{0.8972}{0.475} = \frac{8972}{4750} = \frac{4486}{2375},\\ x &\geq e^{\frac{4486}{2375}} \approx 6.61171.\end{aligned} \tag{14}$$

Stoga je traženi najmanji iznos 6611.71 kn.

Napomena 4. Zbog "mjerne jedinice" za varijablu x, vrijednosti potencija broja e u prethodnim zadatcima zaokruživali smo na 5 decimalnih mjesta. Naime, iako neki vole što precizniji izračun, navođenje 6., 7., 8., ... znamenke iza decimalne točke u tim zadatcima nema nikakvo praktično značenje. Npr. vrijednost $0.475 \cdot \ln 0.9$ izračunana na 10 decimala iznosi 0.0500462449. Prigodom preračunavanja mogli smo napisati "Traženi iznos je 50.0462449 kn.", ali se onda postavlja pitanje: kako *objasniti značenje* znamenke 6, znamenke 2 itd. kad tisućiti (ili još manji) dio jedne kune ne postoji kao novčana jedinica. Stoga prigodom rješavanja ovakvih zadataka treba biti oprezan i s točnošću izračuna, pa zapravo "uvjetno maksimizirati" broj izračunanih znamenaka zavisno o mjernoj jedinici u kojoj su iskazani empirijski podatci.

LITERATURA

1/ I. Šošić, V. Serdar: *Uvod u statistiku*, Školska knjiga, Zagreb, 2000.

2/ I. Šošić: *Primijenjena statistika*, Školska knjiga, Zagreb, 2004.

3/ M. Papić: *Primijenjena statistika u MS Excelu*, Naklada ZORO, Zagreb, 2012.

4/ M. Vukičević, M. Papić: *Matematičko–statistički priručnik za poduzetnike*, Golden-marketing, Zagreb, 2003.