

Prost linearni regresioni model

Linearna regresiona analiza se zasniva na određenim pretpostavkama. Posmatrajmo regresioni model osnovnog skupa:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (4)$$

Ovaj model se zasniva na četiri pretpostavke, koje se odnose na regresioni model osnovnog skupa, a ne na regresioni model uzorka.

Pretpostavka 1. Srednja vrednost slučajne greške ε jednaka je nuli za svaku vrednost promenljive X .

Pretpostavka 2. Slučajne greške različitih opservacija su međusobno nezavisne.

Pretpostavka 3. Slučajne greške za svaku vrednost X ima konstantnu varijansu ($\sigma^2\varepsilon$).

Pretpostavka 4. Za datu vrednost X , raspodela slučajnih grešaka je normalna

13.27 Osiguravajuće društvo interesuje da li iznos životnog osiguranja zavisi od zarade pojedinaca. Odsek za istraživanja u tom društvu je prikupio informacije o šest pojedinaca. U tabeli su dati podaci o godišnjem prihodu (u hiljadama dolara) i iznosu polise životnog osiguranja (u hiljadama dolara) ovih šest osoba.

Godišnji prihod	62	78	41	53	85	34
Životno osiguranje	250	300	100	150	500	75

- Na osnovu datih podataka nacrtajte dijagram raspršenosti. Da li se na osnovu ovog dijagrama može uočiti linearna veza između godišnjeg prihoda i polise životnog osiguranja?
- Odredite liniju regresije $= b_0 + b_1x$, pri čemu je godišnji prihod objašnjavajuća promenljiva, a iznos polise životnog osiguranja zavisna promenljiva.
- Protumačite dobijene vrednosti b_0 i b_1 .
- Ucrtajte liniju regresije u dijagram raspršenosti iz dela zadatka pod a. i označite reziduale crtanjem vertikalnih linija između empirijskih tačaka i linije regresije.
- Kolika je ocenjena vrednost životnog osiguranja osobe za godišnjom zaradom od 55.000 dolara godišnje?
- Jedna od osoba iz uzorka ima godišnji prihod od 78.000 dolara i iznos životnog osiguranja od 300.000 dolara. Koliko iznosi ocenjena vrednost životnog osiguranja ove osobe? Izračunajte rezidual za ovu opservaciju.

13.39 U sledećoj tabeli su dati podaci o prosečnim zasićenim mastima (u gramima) koje se konzumiraju i nivou holesterola (u miligramima na stotinu mililitara) osam slučajno izabranih osoba muškog pola.

Unos masti	55	68	50	34	43	58	77	36
Nivo holesterola	180	215	195	165	170	204	235	150

Izračunajte sledeće: **a.** SK_{xx} SK_{yy} SP_{xy} **b.** Standardnu devijaciju grešaka **c.** SK_U , SK_N i SK_O **d.** Koeficijent determinacije

13.95 U sledećoj tabeli su dati podaci o starosti (u godinama) i broju kvarova tokom prošlog meseca u uzorku od sedam mašina jednog preduzeća.

Starost	12	7	2	8	13	9	4
Broj kvarova	10	5	1	4	12	7	2

- Ako je starost objašnjavajuća promenljiva, a broj kvarova zavisna promenljiva, kako biste formulisali hipotezu o vrednosti parametra β_1 ? (Da li očekujete pozitivnu ili negativnu vrednost β_1 ?)
- Primenom metoda najmanjih kvadrata, odredite regresionu pravu.
- Interpretirajte dobijene ocenjene vrednosti b_0 i b_1 .
- Izračunajte koeficijente r i r^2 i objasnite njihovo značenje.
- Izračunajte standardnu grešku regresije.
- Formirajte 99% interval poverenja za parametar β_1 .
- Sa nivoom značajnosti od 2,5%, testirajte da li je β_1 pozitivno.
- Sa nivoom značajnosti od 2,5%, možete li zaključiti da je koeficijent proste linearne korelacije pozitivan. Da li je vaš zaključak isti kao u delu zadatka pod g.?

13.97 Uprava jednog supermarketa želi da ispita da li postoji veza između učestalosti reklamiranja jednog proizvoda na interkom sistemu u prodavnici i broja prodatih komada tog proizvoda. U svrhu ove analize, uprava je izabrala proizvod i reklamirala ga na interkom sistemu sedam dana. U sledećoj tabeli su prikazani podaci o učestalosti reklamiranja proizvoda svakog dana i broja prodatih komada proizvoda.

Učestalost reklamiranja po danu	15	22	42	30	18	12	38
Broj prodatih komada po danu (u hiljadama)	11	22	30	26	17	15	23

- a. Ako je učestalost reklamiranja objašnjavajuća promenljiva, a broj prodatih komada zavisna promenljiva, da li očekujete pozitivnu ili negativnu vrednost β_1 u regresionom modelu $Y = \beta_0 + \beta_1 X + \varepsilon$.
- b. Primenom metoda najmanjih kvadrata, odredite regresionu pravu $= b_0 + b_1 X$. Da li je znak dobijene ocenjene vrednosti b_1 isti kao i znak β_1 koji ste očekivali u delu zadatka pod a.?
- c. Interpretirajte dobijene ocenjene vrednosti b_0 i b_1 .
- d. Izračunajte koeficijente r i r^2 i objasnite njihovo značenje.
- e. Predvidite broj prodatih komada proizvoda u danu u kojem je proizvod reklamiran 35 puta.
- f. Izračunajte standardnu grešku regresije.
- g. Formirajte 98% interval poverenja za parametar β_1 .
- h. Sa nivoom značajnosti od 1%, testirajte da li je β_1 pozitivno.
- i. Za $\alpha = 0,02$, možete li zaključiti da se koeficijent korelacije statistički značajno razlikuje od nule?

15. U sledećoj tabeli dati su podaci o temperaturi (u Farenhajtima) u 6 časova popodne i posećenosti (zaokruženo na stotine) utakmicama tima iz niže bejzbol lige tokom sedam slučajno odabranih večeri u maju mesecu.

Temperatura 61 70 50 65 48 75 53

Posećenost 10 16 12 15 8 20 18

- a. Da li smatrate da posećenost zavisi od temperature ili važi obratno?
- b. Ako je temperatura objašnjavajuća promenljiva, a posećenost zavisna promenljiva, kako biste formulisali hipotezu o parametru β_1 u regresionom modelu?
- c. Na osnovu datih podataka nacrtajte dijagram raspršenosti. Da li ovaj dijagram ukazuje na linearnu vezu između dve promenljive?
- d. Primenom metoda najmanjih kvadrata, odredite regresionu pravu. Da li je znak dobijene ocenjene vrednosti b_1 isti kao što je znak β_1 , koji ste pretpostavili u delu zadatka pod a.?
- e. Interpretirajte dobijene ocenjene vrednosti b_0 i b_1 .
- f. Izračunajte koeficijente r i r^2 i objasnite njihovo značenje.
- g. Predvidite posećenost utakmici u večernjim časovima u maju pri temperaturi od 60 stepeni.
- h. Izračunajte standardnu grešku regresije.
- i. Formirajte 99% interval poverenja za parametar β_1 .
- j. Sa nivoom značajnosti od 1%, testirajte da li je β_1 pozitivno.
- k. Formirajte 95% interval poverenja za prosečnu posećenost utakmicama u večernjim časovima u maju pri temperaturi od 60 stepeni.
- l. Formirajte 95% interval predviđanja za posećenost jednoj utakmici u večernjim časovima u maju pri temperaturi od 60 stepeni.
- m. Za $\alpha = 0,01$, možete li zaključiti da je koeficijent korelacije pozitivan?

Ispitni zadaci: Jun 2008 VI grupa, Septembar 2009 B i D grupa