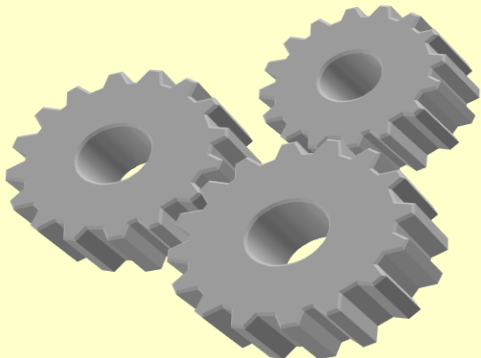




Nema u knjizi!

Regresiona i korelaciona analiza



Predavač: Dr Mirko Savić

savicmirko@ef.uns.ac.rs

www.ef.uns.ac.rs

Regresiona analiza

Regresiona analiza je skup statističkih metoda kojima se otkriva da li postoje veze između posmatranih pojava i kakve su po obliku i smeru.

***Regressio* (lat.) – vraćanje, opadanje, odstupanje**

Fransis Galton u XIX veku.

U regresionoj analizi cilj istraživanja je osnovni skup.

Najčešće se ispituje uzorak.

Prosta i višestruka regresija

Regresija je metod kojim se ispituje zavisnost između dve ili više promenljivih, odnosno pojava:

Visina zarade, radni učinak

Količina padavina, prinos kukuruza

Temperatura vazduha, broj turista

Nacionalni dohodak, javna potrošnja

Nacionalni dohodak, životni standard

Godine starosti, sportski rezultat

Nivo obrazovanja, radno mesto, visinu zarade

Količina padavina, količina đubriva, prinos pšenice

Utvrđuje se regresioni model kojem se najviše približava kvantitativno slaganje varijacija posmatranih pojava.

Prvi korak: Da li uopšte postoji veza?

Tri oblika veze:

- pravolinijski (linearni),
- krivolinijski (nelinearni),
- prostorni.

Smer (samo za linearni oblik veze):

- pozitivan,
- negativan.

Drugi korak: Određivanje zavisne (Y) i nezavisne promenljive (X_1). Na osnovu **logičke analize**.

Dve promenljive:

Visina zarade → Radni učinak

Količina padavina → Prinos kukuruza

Temperatura vazduha → Broj turista

Nacionalni dohodak → Javna potrošnja

Nacionalni dohodak → Životni standard

Godine starosti → Sportski rezultat

Tri promenljive:

Nivo obrazovanja, Radno mesto → Visinu zarade

Količina padavina, količina đubriva → Prinos pšenice

Korelaciona analiza

Korelaciona analiza je skup statističkih metoda kojima se istražuje jačina veze između posmatranih pojava.

Korelacija predstavlja međusobnu povezanost obeležja posmatranih pojava.

***Correlatia* (lat.) – stvari koje stoje u uzajamnom odnosu.**

Prvi je primenio Karl Pearson (1857-1936).

Ako nema regresije, nema ni korelacije.

Ovde nije bitno koja promenljiva je zavisna, a koja nezavisna.

Prema jačini (intenzitetu veza) između obeležja posmatranih pojava:

- funktionalna (matematička),**
- korelaciona, (stohastička, statistička).**

Korelacione veze su labavije u odnosu na funkcionalne i važe samo u proseku.

Korelacija se može podeliti prema:

- smeru (pozitivna ili negativna),**
- broju posmatranih pojava (prosta ili višestruka),**
- obliku veza (linearna ili nelinearna).**

Neposredna korelacija:

ako su pojave neposredno povezane, npr. tražnja i cena.

Posredna korelacija:

cena naftnih derivata (kerozina) → cena avio-prevoza

cena avio-prevoza → cena turističkih aranžmana.

Prosta linearna regresija

Kada se istovremeno posmatraju dve pojave između kojih postoji korelaciona veza, pri čemu se originalnim vrednostima obeležja može dobro prilagoditi linearna funkcija, odnosno prava linija.

Parametri proste linearne regresije su:

- **dijagram rasturanja (raspršenosti),**
- **jednačine,**
- **standardne greške.**

Dijagram rasturanja

Na osnovu njega može se odrediti oblik, smer i približna jačina veze između posmatranih pojava.

Ako su tačke raspršene bez ikakvog pravila, onda nema nikakve veze između pojava.

Jednačina proste linearne regresije

Za osnovni skup:

$$y = \beta_0 + \beta_1 x_1 + u$$

$$y = \alpha + \beta_1 x_1 + u$$

Gde je:

y - zavisna promenljiva,

x_1 – nezavisna promenljiva,

β_0, β_1 – nepoznati regresioni parametri u osnovnom skupu, koji se ocenjuju na osnovu uzorka,

u (nekada ε - epsilon) – slučajna greška (komponenta slučajnosti, skrivena promenljiva) koja se pokorava normalnom rasporedu, i bez nje veza između pojava bi bila funkcionalna),

Stara notacija:

Za zavisnost X_1 prema X_2 : $x_{i1.2} = A_{12} + B_{12} x_{i2} + \varepsilon_i, i = 1, 2, \dots, N.$

Za uzorak:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$$

$$\hat{y} = \hat{\alpha}_0 + \hat{\beta}_1 x_1$$

$$\hat{y} = b_0 + b_1 x_1$$

Stara notacija:

Za zavisnost X_1 prema X_2 :

$$\hat{x}_{i1.2} = a_{12} + b_{12} x_{i2} \quad , \quad i = 1, 2, \dots, n.$$

Standardne greške proste linearne korelacije

Standardna greška σ pokazuje srednju meru odstupanja originalnih vrednosti obeležja zavisne promenljive y , od linije proste linearne regresije \hat{y} mereno paralelno sa osom zavisne promenljive y . Standardna greška se izražava u jedinicama mere u kojima se izražava i zavisna promenljiva y .

REG-001 Linearna regresija i korelacija (Excel)

REG-015 Linearna regresija i korelacija (Excel)

Izbor najboljeg prostog regresionog modela

Regresioni modeli:

- linearna regresija,
- krivolinijska regresija oblika kvadratne parabole,
- krivolinijska regresija oblika stepene funkcije,
- krivolinijska regresija oblika eksponencijalne funkcije,
- krivolinijska regresija oblika hiperbolične funkcije,
- krivolinijska regresija oblika polulogaritamske funkcije.

Na osnovu najmanje standardne greške!

Pearson-ov koeficijent korelacije: r

$$-1 \leq r \leq +1.$$

$$r = \mp \sqrt{r^2}$$

Stepeni jačine:

$0 < |r| \leq 0,5$ - slaba korelaciona veza,

$0,5 < |r| \leq 0,7$ - značajna korelaciona veza,

$0,7 < |r| \leq 0,9$ - jaka korelaciona veza,

$0,9 \leq |r| < 1$ - vrlo jaka korelaciona veza.

Za obeležja koja imaju normalnu distribuciju.

Kad ne postoji normalna distribucija:

Spearman-ov koeficijent korelacije

Rangirana obeležja

Crosstabs-Statistics-Corr

Kendall-ov koeficijent korelacije (Kendall's Tau)

Rangirana obeležja

Crosstabs-Statistics-CTau

REG-022 Koeficijent korelacije

