

Multiregressioon

Statistilise andmetöötluse keerukamate ülesannete hulka kuulub rohkem kui kahe juhusliku suuruse vahelise sõltuvuse uurimine. Olgu uuritavaid tunnuseid kolm (näiteks X , Y ja Z) ning valitsegu nende vahel lineaarne korrelatiivsõltuvus, siis võib otsida üldkogumi regressioonimudelit kujul

$$Y = \alpha + \beta \cdot x + \gamma \cdot z.$$

Üldkogumi regressiooni hindamiseks kasutatakse valimi andmetel koostatud regressioonitasapinna võrrandit

$$\hat{y} = a + b \cdot x + c \cdot z,$$

kus kordajad a , b ja c leitakse vähimruutude meetodi abil, minimiseerides hälvete ruutude summa

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (a + bx_i + cz_i - y_i)^2 = G(a, b, c).$$

Algebraliste teisenduste tulemusena saadakse parameetrite a , b ja c määramiseks lineaarvõrrandite süsteem:

$$\begin{cases} na + b\sum x_i + c\sum z_i = \sum y_i \\ a\sum x_i + b\sum x_i^2 + c\sum x_i z_i = \sum x_i y_i \\ a\sum z_i + b\sum x_i z_i + c\sum z_i^2 = \sum y_i z_i \end{cases}$$

Mitmene determinatsioonikordaja arvutatakse analoogselt lihtregressiooni juhuga ning samad valemid kehtivad ka prognoosi täpsuse ning parameetrite usalduspiiride leidmiseks.

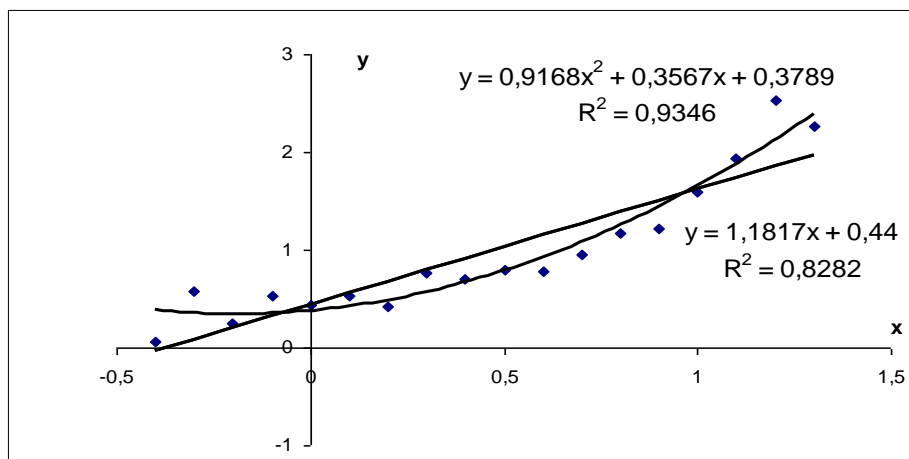
Näide. On antud ettevõtte

reklaamikulud meedias – X ,
reklaamikulud müügipunktis – Z
ja saadud müügitulud – Y .

Uurida tunnuste Y , X ja Z omavahelisi seoseid. Koostada regressioonivõrrand kujul $\hat{y} = a + b \cdot x + c \cdot z$. Kasutada saadud regressioonivõrrandit prognoosiks kohal $x = 4$, $z = 4$. Arvutada mitmene determinatsioonikordaja. Lahendus on toodud lisana Exceli failis.

Mittelineaarne regressioon

Tuleme tagasi juhu juurde, kus meil on 2 tunnust. Kahe tunnuse vaheline seos ei pruugi alati olla lineaarne, vaid võib olla kirjeldatav palju keerulisema valemiga. Lihtne ja kiire võimalus sobiva seose kuju valikuks on hajuvusdiagrammil regressioonijoone kuju muutmine. Alloleval joonisel on hajuvusdiagrammile lisatud nii lineaar- kui ka ruutfunktsiooni graafikud. Üldiselt võetakse regressioonivõrrandiks funktsioon, mille matemaatilised omadused vastavad modelleeritava nähtuse kohta eelnevalt teadaolevatele seaduspärasustele.



Paraboolne regressioon

Mittelineaarset sõltuvust võib uurida mitmel kujul, järgnevas piirdume ühe võimaluse – paraboolse regressiooni, täpsema vaatlemisega.

Paraboolse regressiooni korral valitakse regressioonikõveraks ruutparabool

$$Y = \alpha + \beta \cdot x + \gamma \cdot x^2.$$

Asendusega $z = x^2$ minnakse üle multiregressioonimudelile

$$Y = \alpha + \beta \cdot x + \gamma \cdot z.$$

Parameetrite α , β ja γ punkthinnangud a , b ja c leitakse vähimruutude meetodiga põhimõtteliselt samal viisil, nagu varasemalt kirjeldatud. Teisenduste tulemusel saadakse otsitavate a , b ja c leidmiseks lineaarne võrrandisüsteem:

$$\begin{cases} na + b\sum x_i + c\sum x_i^2 = \sum y_i \\ a\sum x_i + b\sum x_i^2 + c\sum x_i^3 = \sum x_i y_i \\ a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 = \sum x_i^2 y_i \end{cases}.$$

Üldkogumi regressiooni saame nüüd hinnata järgneva mudeli abil

$$\hat{y} = a + b \cdot x + c \cdot x^2.$$

Näide. Tabelis on toodud kuritegude arv Eestis 100 000 elaniku kohta aastatel 1992 kuni 2000.

1992	1993	1994	1995	1996	1997	1998	1999	2000
2671	2450	2369	2680	2409	2802	3154	3555	4221

Lähendada andmeid ruutparabooliga $\hat{y} = a + b \cdot x + c \cdot x^2$. Arvutada determinatsioonikordaja. Prognoosida antud andmete põhjal kuritegude arvu vaadeldud perioodile järgneval aastal. Lahendus on toodud lisana Exceli failis.