Kvantitativne metode za analize v upravi

ANALIZA POVEZANOSTI STATISTIČNIH

SPREMENLJIVK

Prof.dr. Srečko Devjak

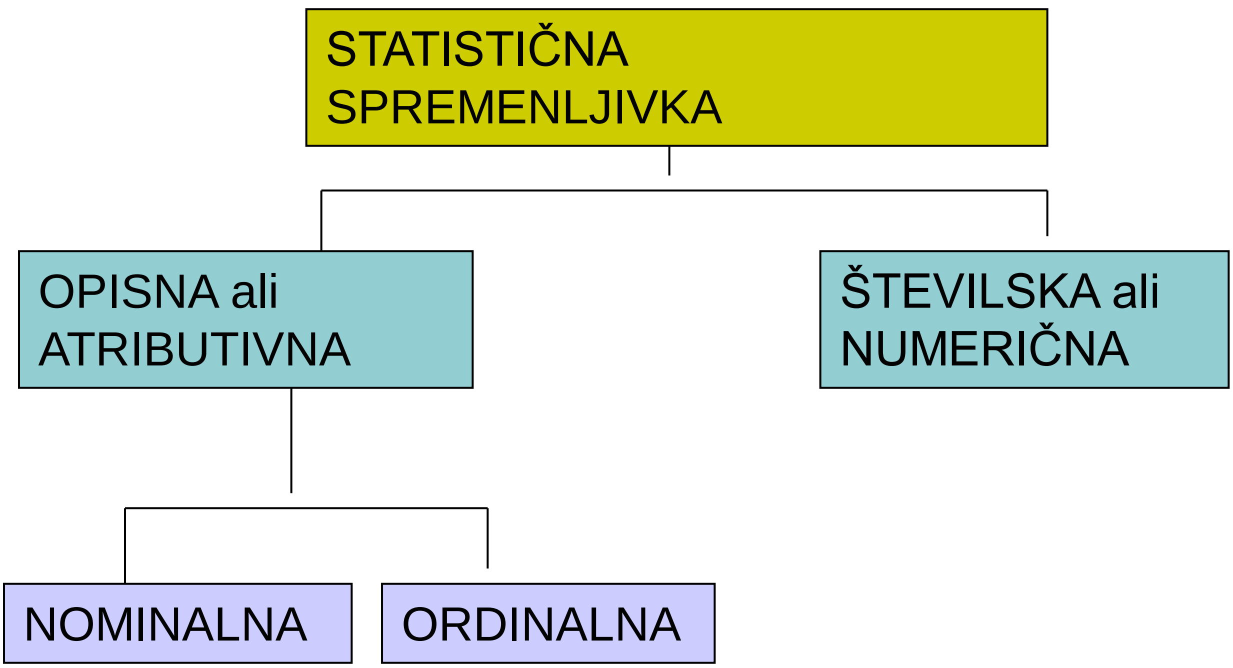
Dr. Jože Benčina

ANALIZA ODVISNOSTI MED MNOŽIČNIMI POJAVI

Proučevanje:

- Zakonitosti povezanosti statističnih pojavov,
- Opisa povezanosti: moč, smer, oblika,
- Metod analize povezanosti
- Ugotavljanja povezanosti od vrste statistične spremenljivke

Vrste statističnih spremenljivk



Povezanost statističnih spremenljivk

<u>Vrste analiz</u> **po številu spremenljivk**:

- **univariantna** – ena spremenljivka,
- **bivariantna** – dve spremenljivki,
- **multivariantni** - več kot dve spremenljivki.

<u>**Povezanost:**</u>

•**funkcijska**: y=y’(x)

•**stohastična**: y= y’(x)+ε

<u>**Vsebinski vidik**</u> **povezanosti**:

▪**regresijski** model,

▪**korelacijski** model.

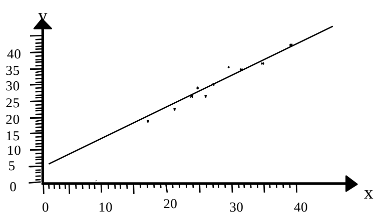
<u>**Grafično prikazovanje**</u>:

•Točke – **razsevni grafikon/diagram**,

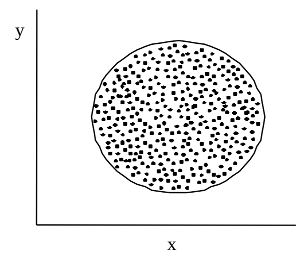
•Stolpci – redkeje.

Lastnosti povezav

MOČ

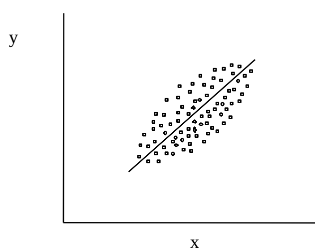


Močna povezanost

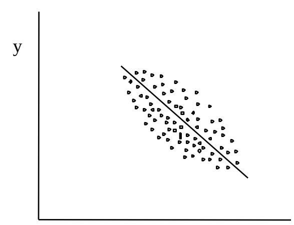


Povezanost ne obstaja

SMER

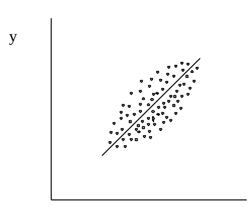


Pozitivna

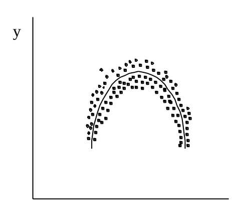


Negativna

OBLIKA

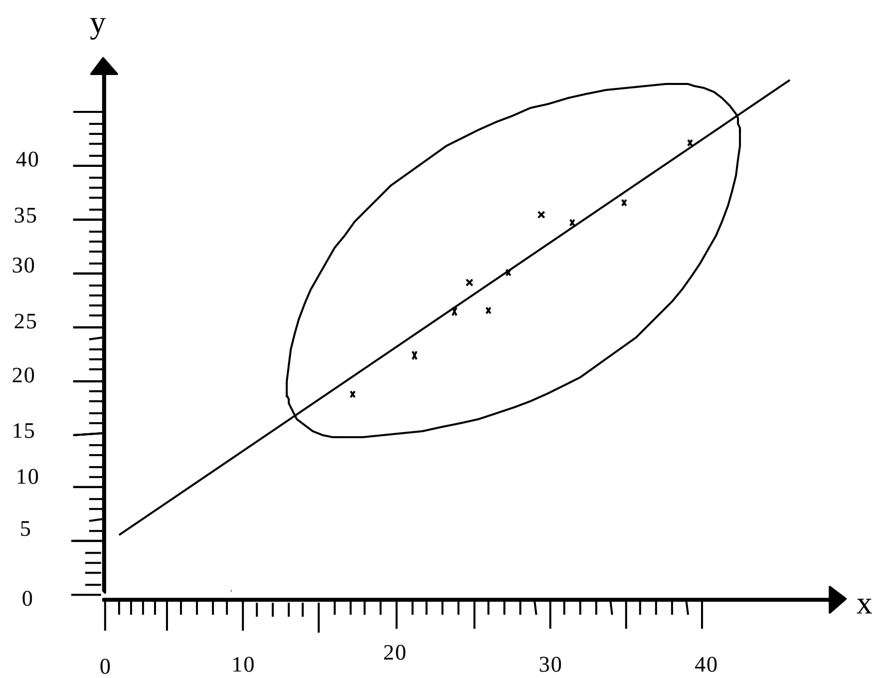


Linearna



Nelinearna

Grafični prikaz lastnosti povezanosti spremenljivk



Regresijska funkcija

Regresijska funkcija – regresijski model:

$$y_i^{'} = \alpha_1 + \beta_1 * x_i$$

Koeficienti regresijske funkcije:

$$\beta_1 = \frac{C_{xy}}{\sigma_X^2}$$

$$\alpha_1 = \overline{Y} - \beta_1 * \overline{X}$$

$$C_{xy} = C_{yx} = \frac{\sum (y_i - \overline{Y}) * (x_i - \overline{X})}{N}$$

$$\sigma_x^2 = \frac{\sum (x_i - \overline{X})^2}{N}$$

Korelacijski model:

$$y_i' = \alpha_1 + \beta_1 * x_i$$

$$x_i' = \alpha_2 + \beta_2 * y_i$$

Postopek izračuna:

Izračun varianc:

$$\sigma_y^2 = \frac{\sum (y_i - \overline{Y})^2}{N}$$

$$\sigma_x^2 = \frac{\sum (x_i - \overline{X})^2}{N}$$

Izračun kovariance:

$$C_{xy} = C_{yx} = \frac{\sum (y_i - \overline{Y}) * (x_i - \overline{X})}{N}$$

Koeficienti:

$$\beta_1 = \frac{C_{xy}}{\sigma_x^2} \qquad \beta_2 = \frac{C_{xy}}{\sigma_y^2}$$

$$\alpha_1 = \overline{Y} - \beta_1 * \overline{X} \qquad \alpha_2 = \overline{X} - \beta_2 * \overline{Y}$$

Izračun pri frekvenčnih porazdelitvah

Posamični podatki	Frekvenčna porazdelitev
$\overline{X} = M_x = \frac{1}{N}\sum_{i=1}^{N} x_i$	$\overline{X} = M_x = \frac{1}{N}\sum_{k=1}^{K} f_k x_k$
$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{X})^2$	$\sigma_x^2 == \frac{1}{N}\sum_{k=1}^{K} f_k (x_k - \overline{X})^2$
$C_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y})$	$C_{xy} = \frac{1}{N}\sum_{k=1}^{K}\sum_{j=1}^{M} f_{jk} (x_k - \overline{X})(y_j - \overline{Y})$

Moč povezanosti - vrednotenje

- **Koeficienti povezanosti**
  - Vrednosti koeficientov:

Vrednost koeficienta		Moč/jakost povezanosti
Interval		
0	+/-0,20	ni
+/-0,20	+/-0,40	šibka
+/-0,40	+/-0,70	zmerna
+/-0,70	+/-1,00	močna

Pearsonov korelacijski koeficient

$$\rho_{xy} = \frac{C_{xy}}{\sigma_y * \sigma_x}$$

Spearmanov koeficient korelacije ranga:

$$\rho_{yx} = 1 - \frac{6 * \sum (R_y - R_x)^2}{N * (N^2 - 1)}$$

Primer:

i	Občina	Odhodki občine **skupaj (v mio sit)**	Stroški uprave (v mio sit)		yj-Y	
		y	x	Ry	Rx	(Ry-Rx)2
1	HRASTNIK	1.380	202	7	7	0
2	CERKNICA	1.201	179	10	8	4
3	ROGAŠKA SLATINA	1.267	145	9	10	1
4	BLED	1.556	283	3	3	0
5	LENDAVA	1.454	338	5	2	9
6	LENART	1.427	224	6	5	1
7	LOGATEC	1.500	152	4	9	25
8	LJUTOMER	1.637	246	2	4	4
9	SEŽANA	1.940	358	1	1	0
10	IDRIJA	1.332	202	8	6	4
	Skupaj	14.695	2.330			48

$$r_{yx} = 1 - \frac{6 * \sum (R_y - R_x)^2}{N * (N^2 - 1)} = 1 - \frac{6 * 48}{10 * (100 - 1)} = 0,71$$

1.3 Merjenje jakosti korelacije pri atributivnih statističnih spremenljivkah

- Dve skupini spremenljivk:

- ***mere povezanosti ordinalnih spremenljivk in***

- ***mere povezanosti nominalnih spremenljivk***

Povezanost ordinalnih spremenljivk

- *koeficient korelacije ranga in*
- *koeficient konkordance.*

1. Koeficient korelacije ranga

Spearmanov koeficient korelacije ranga:

$$\rho_{yx} = 1 - \frac{6 * \sum (R_y - R_x)^2}{N * (N^2 - 1)}$$

2. Koeficient konkorndance

<u>*Izhodišče*</u>: dve ordinalni spremenljivki *x* in *y*

- enoti E_i pripada dvojica vrednosti (x_i, y_i)
- enoti E_j dvojico vrednosti (x_j, y_j).

<u>Varianta</u>: x_i in x_j sta v **enakem odnosu** urejenosti, kot sta vrednosti y_i in y_j:

$x_i > x_j \Rightarrow y_i > y_j$ ali $x_i < x_j \Rightarrow y_i < y_j$,

par E_i in E_j .... **konkordanten par** .

<u>Varianta</u>: x_i in x_j sta v **nasprotnem odnosu** urejenosti, kot sta vrednosti y_i in y_j:

$x_i > x_j \Rightarrow y_i < y_j$ ali $x_i < x_j \Rightarrow y_i > y_j$,

par E_i in E_j .... dis**konkordanten par** .

<u>Varianta</u>: kadar nastopi enakost vrednosti:

$x_i = x_j$ za y pa $y_i < y_j$ ali $x_i < x_j \Rightarrow y_i > y_j$,

Ali pa

$y_i = y_j$ za x pa $x_i \geq x_j$ ali $x_i < x_j$

par Ei in Ej ....**ne upoštevamo**

Koeficient konkordance

- Postopek:
  - Primerjamo vsako enoto z preostalimi – število primerjav za N enot:

$$\binom{N}{2} = \frac{N(N-1)}{2}$$

- N_{kon} – število konkordantnih parov
- N_{dis} – število diskonkordantnih parov,
- k_{kon}- koeficient konkordance

$$k_{kon} = \frac{2(N_{kon} - N_{dis})}{N(N-1)}$$

$$-1 \leq k_{kon} \leq +1$$

Primer:

- Ali je plačilni razred odvisen od stopnje izobrazbe?

Zaposleni	Stopnja izobrazbe	Plačilni razred
D1	**4**	**VI**
D2	**3**	**VII**
D3	**4**	**V**
D4	**3**	**VII**
D5	**7**	**VI**
D6	**2**	**VIII**

1.3.2 Povezanost nominalnih spremenljivk

- Mere:
  - *kontingenca in*
  - *asociacija .*

Kontingenca

Postopek:

•Kontingenčna tabela za $f(x_j, y_k)$

•Koeficient χ^2

•Koeficient kontingence C

Koeficient χ^2

$$\chi^2 = \sum_{j=1}^{J}\sum_{k=1}^{K} \frac{\left(f(x_j, y_k) - f^t(x_j, y_k)\right)^2}{f^t(x_j y_k)}$$

$f(x_j, y_k)$ frekvenca za proučevan pojav

x_j (j=1,2,...,J) , y_k (k=1,2,..,K)

$f^t(x_j, y_k)$ izračunana teoretična frekvenca enot za primer neodvisnosti med x in y

$$f^t(x_j, y_k) = \frac{f(x_j) f(y_k)}{N}$$

$f(x_j)$ in $f(y_k)$ robni frekvenci

$$N = \sum_{j=1}^{J} f(x_j) = \sum_{k=1}^{K} f(y_k)$$

Koeficient kontingence C

- Če je $\chi^2 > 0$ nadaljujemo
- Izračun koeficienta C:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

$0 \leq C \leq 1$

Popravek:

$$C_{\max} = \sqrt{\frac{m-1}{m}}$$

$m = \max(J,K)$

Popravljena vrednost:

$$C_{corr} = \frac{C}{C_{\max}}$$

Primer:

Asociacija

- Problem – *x* in *y* binarni spremenljivki

$$Q = \frac{f(x_1, y_1) f(x_2, y_2) - f(x_1, y_2) f(x_2, y_1)}{f(x_1, y_1) f(x_2, y_2) + f(x_1, y_2) f(x_2, y_1)}$$

1.3.3. Merjenje jakosti korelacije med atributivno in numerično statistično spremenljivko - korelacijsko razmerje η_{yx}

- *numerično spremenljivko označimo z y,*
- *podatke grupirajmo po spremenljivki x (numerični ali atributivni) v K skupin*
- Označimo z:
  - *N-število enot v statistični množici,*
  - *K-število grup za spremenljivko x ,*
  - *N_k-število enot v k - ti (k =1,2,...,K) grupi spremenljivke x ,*

$$\eta_{yx} = \sqrt{\frac{\sum_{k=1}^{K} Y_k^2 \Big/ N_k - Y^2/N}{\sum_{i=1}^{N} y_i^2 - Y^2/N}}$$

$$Y_k^2 = \left(\sum_{i_k=1}^{N_k} y_{i_k} \right)^2$$

$$Y^2 = \left(\sum_{i=1}^{N} y_i \right)^2$$

Primer:

Povprečne mesečne neto plače za leto 1999 po dejavnostih (Vir:SL 2000, Str.:249)

Dejavnost	Neto plača (Tisoč SIT)	Y_k^2	N_k	Y_k^2/N_k	y_i^2
Rudarstvo	125				15625
	107				11449
VSOTA 1	**232**	**53824**	**2**	**26912**	**0**
Predelovalne dejavnosti	105				11025
	71				5041
	74				5476
	81				6561
	94				8836
	95				9025
VSOTA 2	**520**	**270400**	**6**	**45066,667**	**0**
Trgovina	103				10609
	117				13689
	110				12100
	94				8836
VSOTA 3	**424**	**179776**	**4**	**44944**	**0**
		504000	**12**	**116923**	**118272**

N=	12
Y =	1.176
Y^2 =	1.382.976
Y^2/N =	115.248
Σy_i^2	118.272
$\Sigma Y_k^2/N_k$	116.923

$$\eta_{yx} = \sqrt{\frac{\sum_{k=1}^{K} Y_k^2 \big/ N_k - Y^2 \big/ N}{\sum_{i=1}^{N} y_i^2 - Y^2 \big/ N}} = 0.74$$