Regresija in korelacija

- Kvantitativne metode v geografiji in uvod v GIS -

dr. Gregor Kovačič, doc.

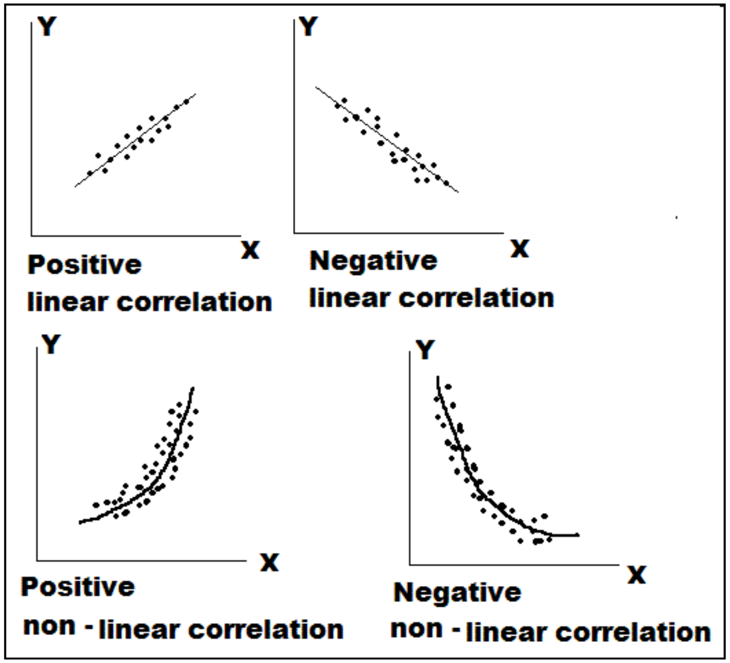
Odvisnost in povezanost

- Opazujemo primere, ko na vsaki enoti gledamo dve številski spremenljivki hkrati
- Spremenljivki imata lahko dve relaciji
  - Odvisnost
    - Vpliv vrednosti ene spremenljivke na vrednosti druge, v obratni smeri pa ni vpliva (*X* – neodvisna spremenljivka; *Y* – odvisna spremenljivka)
    - Primer: nadmorska višina in temperatura zraka
    - Namen študije odvisnosti
      - Pridobiti nova spoznanja o odvisnosti
      - Napovedovanje vrednosti odvisne spremenljivke *Y* pri izbrani vrednosti neodvisne spremenljivke *X* (nekaj podobnega, kot je napovedovanje vrednosti časovne vrste s pomočjo trenda)
  - Povezanost (soodvisnost)
    - Relacija, ko se vrednosti obeh spremenljivk spreminjajo hkrati
    - Primer: Prostornina in masa prodnikov sta povezani (soodvisni) spremenljivki → obe spremenljivki sta enakovredni (vseeno je, katero spremenljivko značimo z *X* ali *Y*)
    - Namen študije povezanosti
      - Izračunati ustrezno mero, ki vrednosti jakost povezanosti obeh spremenljivk
      - Običajni meri povezanosti sta Pearsonov koeficient korelacije in Spearmanov koeficient korelacije rangov
- Obliko in jakost odvisnosti in povezanosti med dvema spremenljivkama lahko ponazorimo:
  - Grafično (razsevni grafikon)
  - S števili, koeficienti (različno glede na tip spremenljivk)

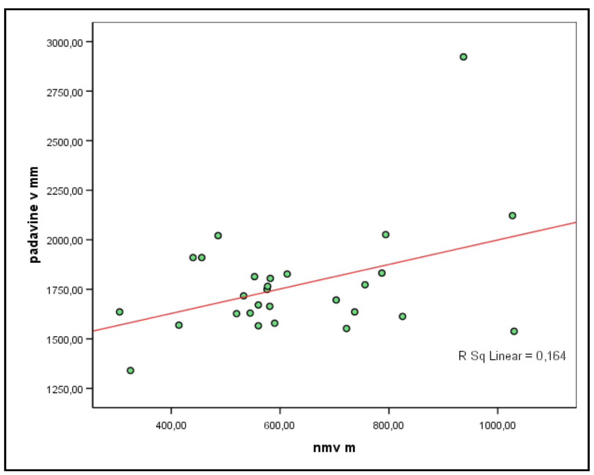
Regresija

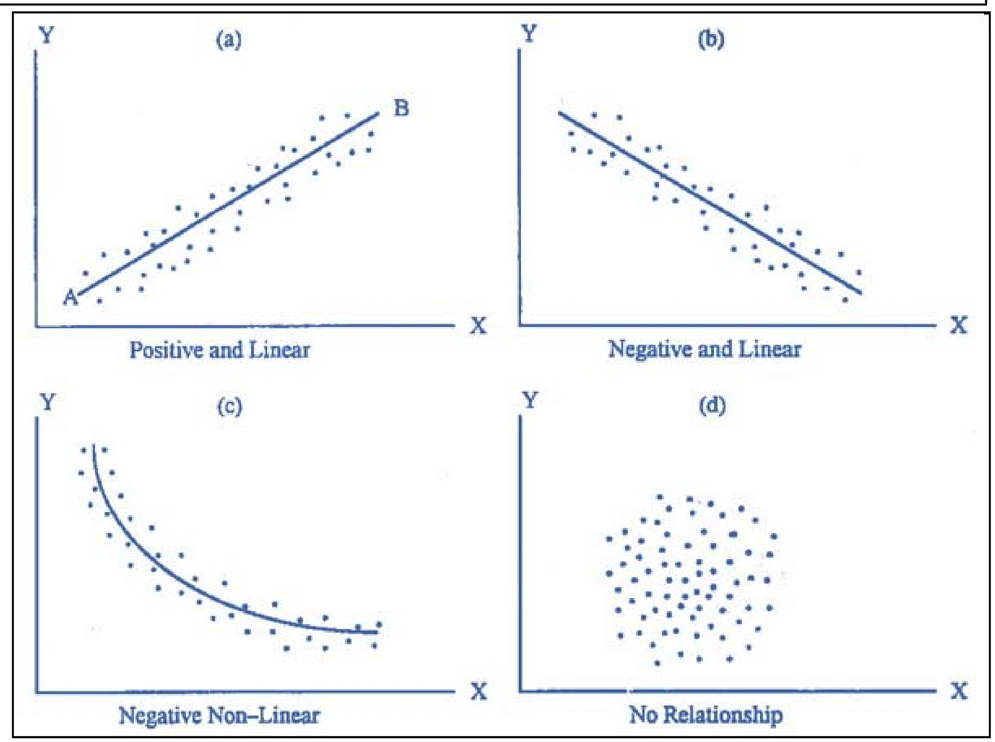
- Regresija je prilagajanje ustrezne matematične funkcije empiričnim podatkom → to funkcijo imenujemo regresijska funkcija
  - Lahko je enostavna, npr. linearna
  - Kompleksna
- Regresijska funkcija je funkcija, ki se najbolje prilega preučevanemu pojavu in kaže, kakšen bi bil vpliv spremenljivke *X* na *Y*, če razen vpliva spremenljivke *X* ne bi bilo drugih vplivov na spremenljivko *Y*.
- Metoda, s katero ob znani *X* napovemo, koliko bo (neznana) *Y*.
  - Primer: Kolikšen bo pretok reke pri vodostaju 600, če velja, da je pretok reke pri vodostaju 400 cm = 45 m^3/s.
  - Napoved vedno vsebuje napako (razen če gre za funkcijsko zvezo), ki je tem večja, čim manjša (absolutno) je korelacija med *X* in *Y*.
  - Napovedovanje vrednosti spremenljivke *Y* iz *X* zunaj okvirjev podatkov *X*, ki smo jih uporabili za postavitev linearnega modela, je zelo tvegano početje (npr. stoletnih poplav na osnovi obstoječih podatkov).
- Išče se funkcijo, pri kateri je vsota kvadratov odklonov minimalna
- Podatke najprej grafično prikažemo z razsevnim grafikonom

Razsevni grafikon ali korelacijski diagram



Razsevni diagram je grafični prikaz povezanosti med dvema spremenljivkama.
Regresijska črta kaže obliko povezanosti:
- Linearna
- Nelinearna (krivuljna)





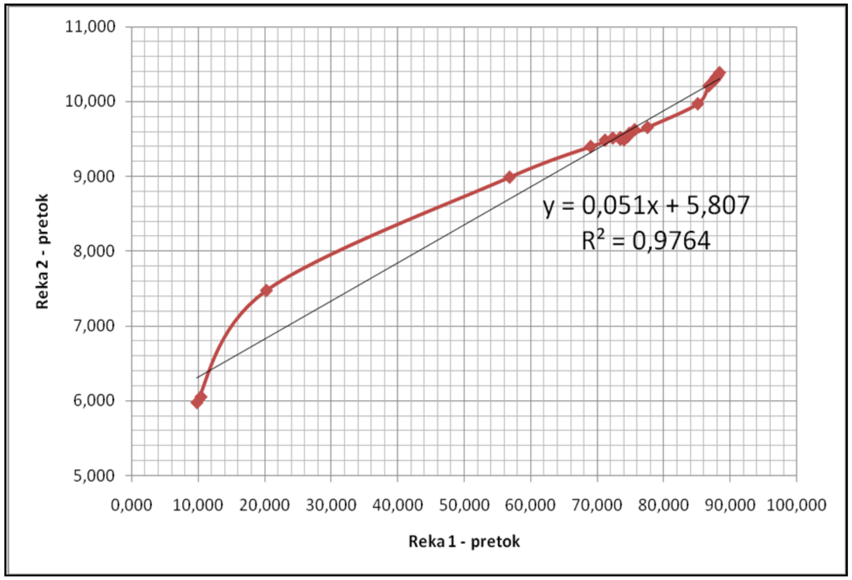
Model enostavne linearne regresije

- Grafična rešitev je premica, ki poteka skozi korelacijski oblak v razsevnem grafikonu (točke (*X,Y*)) čim bližje sredini
- Pri linearni regresiji je to premica z enačbo

$Y = \alpha + \beta \cdot X + \varepsilon$

- Kriterij, ki določa, da bo premica potekala čim bolj po sredini korelacijskega oblaka, je kriterij metode najmanjših kvadratov → vsota kvadriranih odklonov mora biti najmanjša
- Rešitev *α* in *β* dobimo iz sistema dveh linearnih enačb
  - *Y* - napovedana vrednost oz. linearna regresijska funkcija
  - *α* - presečišče z osjo *Y*; $\alpha = \mu_y - \beta \cdot \mu_x$
  - *β* - pričakovano povečanje *Y*, če se *X* poveča za 1 enoto; $\beta = C_{xy} / \sigma_x^2$
- Ocena linearnega regresijskega modela je torej premica $Y = \alpha + \beta \cdot X$

$$\sum_{t=1}^{N} (y_t - y_i^{'})^2 = \min$$



Regresijska premica

- Ob kriteriju najmanjših kvadratov odklonov

$$\sum_{t=1}^{N}(y_t - y_i^{'})^2 = \min$$

- Pri enačbi $Y = \alpha + \beta \bullet X$

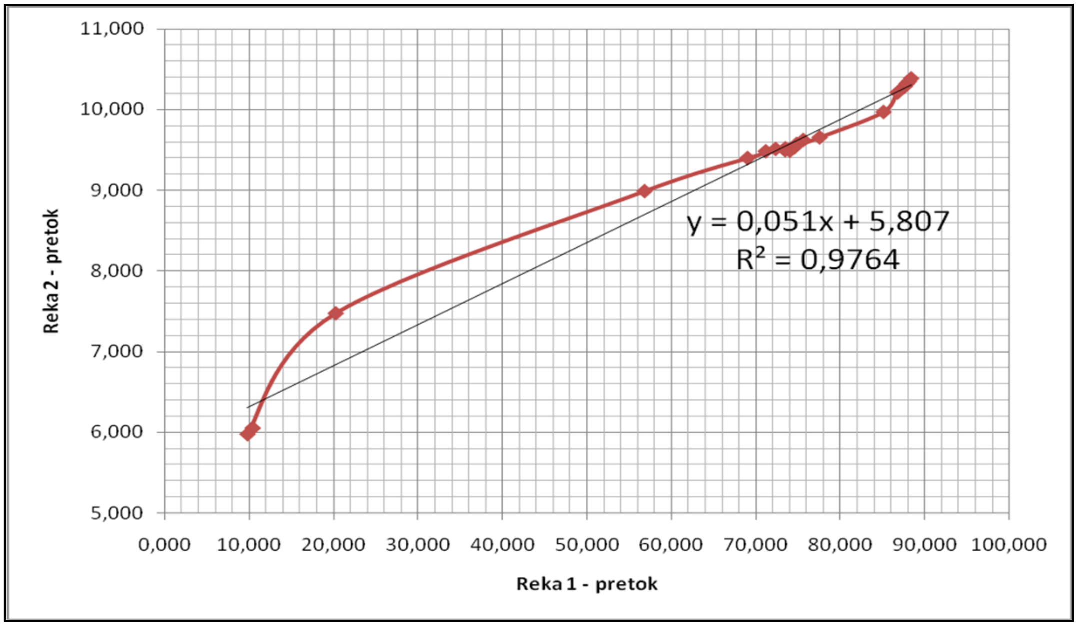
imamo rešitev

  - $\beta = C_{xy} / \sigma_x^2$
  - $\alpha = \mu_y - \beta \cdot \mu_x$

$$C_{xy} = \frac{1}{N} \cdot \sum_{i=1}^{N}(x_i - \mu_x) \cdot (y_i - \mu_y)$$

$$\sigma_x^2 = \frac{1}{N} \cdot \sum_{i=1}^{N}(x_i - \mu_x)^2$$

- MS Excel vam na ukaz sam doda linearno regresijsko premico (lahko tudi druge krivulje) in enačbo ter izračuna determinacijski koeficient (glej predavanje o trendih pri časovnih vrstah)



Regresijska premica - primer

	Reka 1 = *X*	Reka 1 = *Y*			
Datum	**Vodostaj (cm)**	**Pretok (m^3/s)**	$x_i - \mu_x$	$y_i - \mu_y$	$(x_i - \mu_x) \times (y_i - \mu_y)$
29.11.2008	57	9.753	-228.200	-61.964	14140.139
30.11.2008	89	20.198	-196.200	-51.519	10107.989
01.12.2008	184	56.797	-101.200	-14.920	1509.884
02.12.2008	219	68.998	-66.200	-2.719	179.985
03.12.2008	234	73.459	-51.200	1.742	-89.201
04.12.2008	226	71.149	-59.200	-0.568	33.614
05.12.2008	230	72.324	-55.200	0.607	-33.517
06.12.2008	242	75.595	-43.200	3.878	-167.538
07.12.2008	239	74.815	-46.200	3.098	-143.137
08.12.2008	238	74.55	-47.200	2.833	-133.727
09.12.2008	236	74.01	-49.200	2.293	-112.825
10.12.2008	234	73.459	-51.200	1.742	-89.201
11.12.2008	250	77.542	-35.200	5.825	-205.047
12.12.2008	320	85.111	34.800	13.394	466.118
13.12.2008	391	86.763	105.800	15.046	1591.888
14.12.2008	439	87.61	153.800	15.893	2444.374
15.12.2008	452	87.811	166.800	16.094	2684.513
16.12.2008	458	87.899	172.800	16.182	2796.284
17.12.2008	473	88.113	187.800	16.396	3079.206
18.12.2008	493	88.38	207.800	16.663	3462.613
AS (μ)	**285.200**	**71.717**	**0.000**	**0.000**	**41522.413**
Standardni odklon σ	**122.284**	**20.650**	**AS (*X*)**	**AS (*Y*)**	**SUM**
Kovarianca *XY* = 41522 /20	**2076.121**				
Pearson r = $C_{XY}/ \sigma_X * \sigma_Y$	**0.822**				
$\beta = C_{XY} / \sigma_X^2$	**0.139**				
$\alpha = \mu_Y - \beta \cdot \mu_X$	**32.11970567**				
$Y = \alpha + \beta \cdot X$ (320 cm)	**76.54842301**				

Regresijska enačba

$Y = \alpha + \beta \cdot X$

$Y = 0{,}139 \cdot X + 32{,}1197$

Interpretacija:

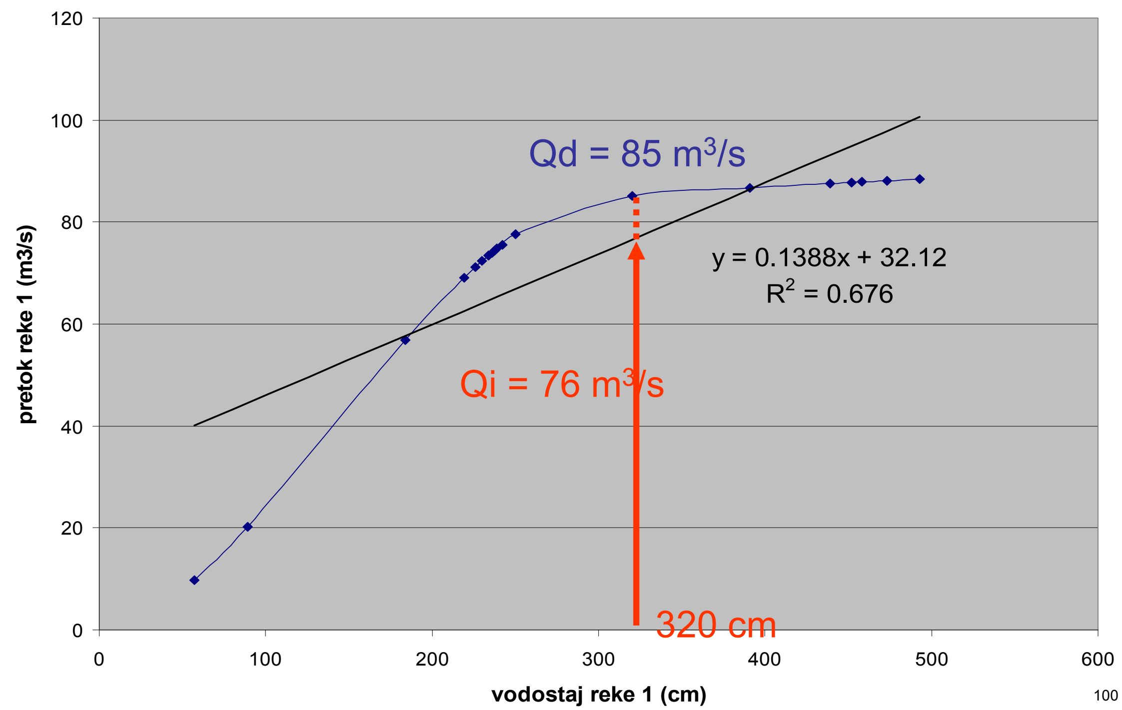
- Vrednost $\alpha = 0{,}139$ nima vsebinskega pomena
- Vrednost β pove, da se pretok reke poveča za 139 l/s, če se vodostaj poveča za 1 cm

Kolikšen je pretok reke 1 pri vodostaju 320 cm?

Ugotovitve / pojasnila:

- Izračunana vrednost je nižja od tiste v preglednici za ta pretok, razlog je v tem, da vodostaj in pretok nista linearno povezana, temveč gre za krivuljno povezanost
- Napovedovanje vrednosti spremenljivke Y iz X zunaj okvirjev podatkov X, ki smo jih uporabili za postavitev linearnega modela, je zelo tvegano početje (npr. stoletnih poplav na osnovi obstoječih podatkov).

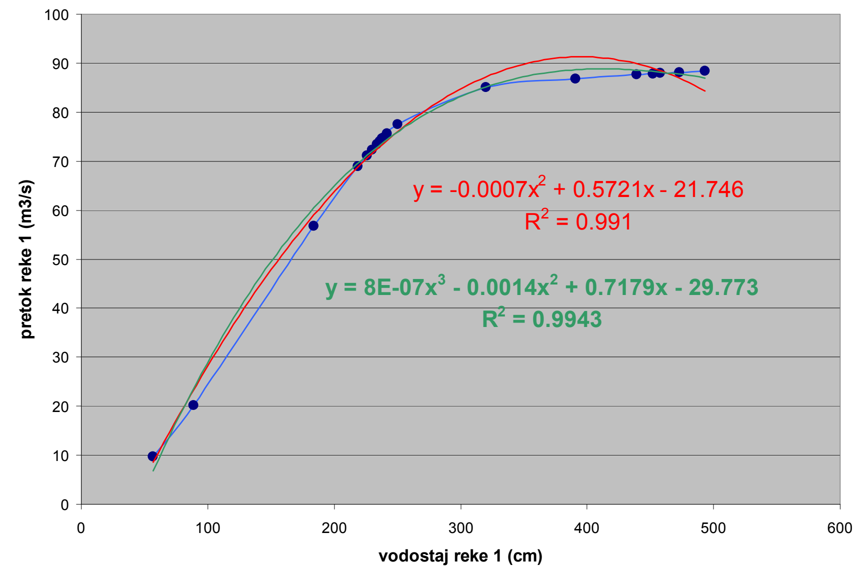
Regresijska premica in razsevni grafikon



Regresijska premica je pri vodostaju 320 cm pod krivuljo, zato je izračunani Q_i iz regresijske enačbe manjši od dejanskega Q_d.

<u>Polinomska regresijska krivulja</u>

Boljša povezanost → bolj natančne napovedi (to poznamo že iz računanja trendov)



Determinacijski koeficient (r^2)

- Regresijski model je lahko boljši ali slabši. Mera, ki vrednosti kakovost regresijskega modela je koeficient determinacije (r^2)
- Pri napovedovanju *Y* iz *X* pri vsaki od enot delamo večjo ali manjšo napako. Njeno varianco imenujemo varianca napake ali nepojasnjena varianca.

$$\sigma_{\varepsilon y}{}^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - y_i^{'})^2$$

- Njen koren je standardna napaka napovedi ali ocene. $\sigma_{\varepsilon y} = \sqrt{\sigma_{\varepsilon y}^2}$

- Pravilno napovedan del variance pa se imenuje pojasnjena varianca. Izračunamo jo enostavno kot r^2 – determinacijski koeficient, ki je določen na intervalu med 0 in 1. Ta kaže delež (lahko tudi v %) celotne variance spremenljivke *Y*, ki je pojasnjen z variiranjem spremenljivke *X*.
- V primeru linearne regresije je enak kvadratu Pearsonovega koeficienta korelacije ($\sigma_{XY}{}^2$). 1 - r^2 je nepojasnjena varianca.

$$\sigma_{xy}{}^2 = \frac{C_{xy}^2}{\sigma_y^2 \cdot \sigma_x^2}$$

Glej primer pri Pearsonovem koeficientu korelacije!

$$r_{xy} = \frac{C_{xy}}{\sigma_x \cdot \sigma_y}$$

Pearsonov koeficient korelacije - *r*

- Ugotavljanje povezanosti dveh številskih (razmernostnih) spremenljivk
- Računanje iz nestandardiziranih podatkov – kovarianca vlomljena s standardnimi odkloni spremenljivk

$$r_{xy} = \frac{C_{xy}}{\sigma_x \cdot \sigma_y}$$

- Pri tem je kovarianca enaka povprečnemu produktu odklonov od povprečja

$$C_{xy} = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - \mu_x) \cdot (y_i - \mu_y)$$

- C_{xy} - kovarianca
- σ_x – standardni odklon spremenljivke x
- σ_y - standardni odklon spremenljivke y
- N - število vseh enot (parov rangov)

$$\sigma_x^{\,2} = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - \mu_x)^2$$

$$\sigma_y^{\,2} = \frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \mu_y)^2$$

- Vrednosti na intervalu -1 do +1
  - -1 → popolna negativna povezanost
  - 0 ni povezanosti
  - +1 → popolna pozitivna povezanost

- Meri se samo linearna povezanost
- Močan vpliv osamelcev (outliers), zlasti pri majhnih vzorcih (malo podatkov)
- Korelacija spremenljivke same s seboj = 1
- Koeficient korelacije je simetričen, spremenljivki sta enakovredni, torej velja $\sigma_{xy} = \sigma_{yx}$ (varianci!)

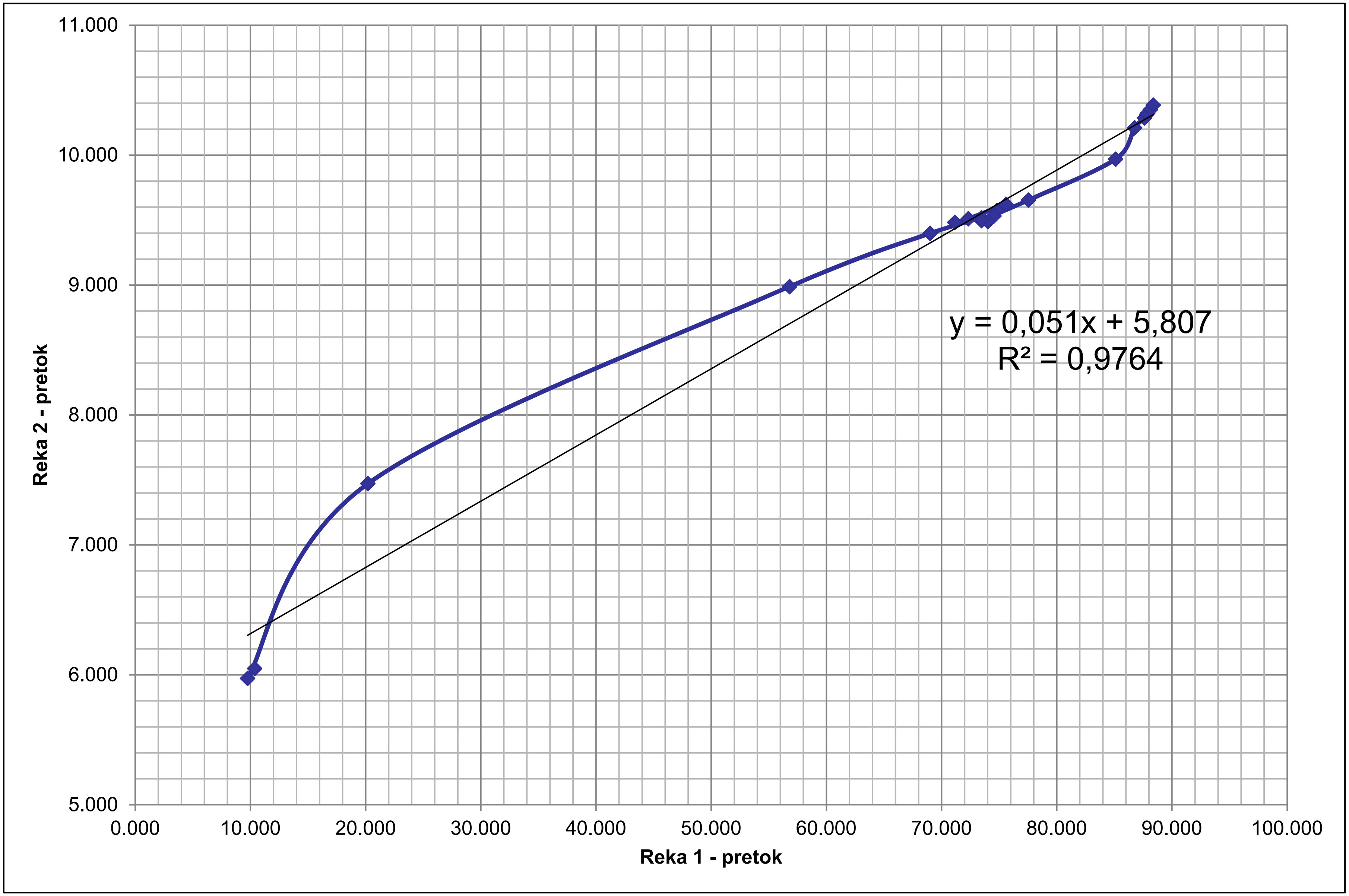
Pearsonov koeficient - primer

	Reka 1 = *x*	Reka 2 = *y*			
Datum	**pretok (m^3/s)**	**pretok (m^3/s)**	$x_i - \mu_x$	$y_i - \mu_y$	$(x_i - \mu_x) \times (y_i - \mu_y)$
28.11.2008	10,360	6,049	-58,435	-3,265	190,790
29.11.2008	9,753	5,973	-59,042	-3,341	197,259
30.11.2008	20,198	7,472	-48,597	-1,842	89,516
01.12.2008	56,797	8,988	-11,998	-0,326	3,911
02.12.2008	68,998	9,398	0,203	0,084	0,017
03.12.2008	73,459	9,518	4,664	0,204	0,951
04.12.2008	71,149	9,482	2,354	0,168	0,395
05.12.2008	72,324	9,511	3,529	0,197	0,695
06.12.2008	75,595	9,622	6,800	0,308	2,094
07.12.2008	74,815	9,574	6,020	0,260	1,565
08.12.2008	74,550	9,532	5,755	0,218	1,255
09.12.2008	74,010	9,489	5,215	0,175	0,913
10.12.2008	73,459	9,496	4,664	0,182	0,849
11.12.2008	77,542	9,654	8,747	0,340	2,974
12.12.2008	85,111	9,969	16,316	0,655	10,687
13.12.2008	86,763	10,210	17,968	0,896	16,099
14.12.2008	87,610	10,286	18,815	0,972	18,288
15.12.2008	87,811	10,313	19,016	0,999	18,997
16.12.2008	87,899	10,324	19,104	1,010	19,295
17.12.2008	88,113	10,349	19,318	1,035	19,994
18.12.2008	88,380	10,385	19,585	1,071	20,975
AS (μ)	**68,795**	**9,314**	**0,000**	**0,000**	**617,522**
Standardni odklon (σ)	**24,017**	**1,239**	**AS (*x*)**	**AS (*y*)**	**SUM**
Kovarianca = 617,522 / 21	**29,406**				
Pearson *r* = 29,406 / 24,017 x 1,239	**0,988**				

Interpretacija:

r = 0,988 → pretoka rek 1 in 2 sta zelo visoko pozitivno povezana!

Pearsonov koeficient - primer



Razlaga *r*

- Korelacijski koeficient *r* nam pove stopnjo (jakost, moč) povezave in smer povezave (+ ali -)
- Korelacijski koeficient *r* ne pokaže vzročno-posledične povezanosti (odvisnosti), ampak zgolj korelacijo
- Najbolj primerno za uporabo, če je porazdelitev *X* in *Y* normalna (ali vsaj simetrična in unimodalna; en modus)

Vrste korelacij

<u>Glede na smer povezanosti:</u>

- Pozitivna: z naraščanjem *X* narašča tudi *Y*
- Negativna: z naraščanjem *X* se vrednosti *Y* zmanjšujejo

<u>Glede na obliko povezanosti:</u>

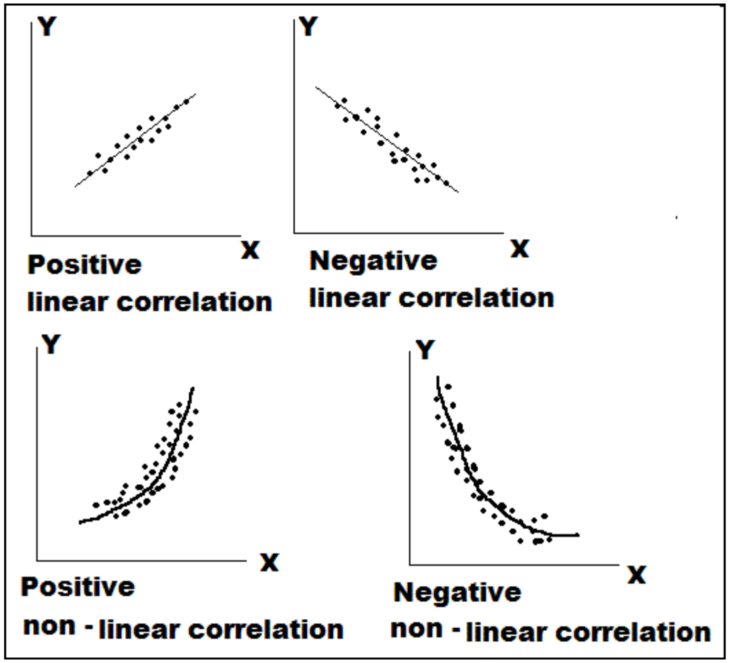
- Linearna - premočrtna
- Nelinearna – krivuljna (več vrst, krivulje 2, 3 in nadaljnjega reda; MS Excel!)

<u>Glede na jakost povezanosti:</u> ni povezanosti, neznatna, šibka, … , močna (funkcijska) → glej naprej

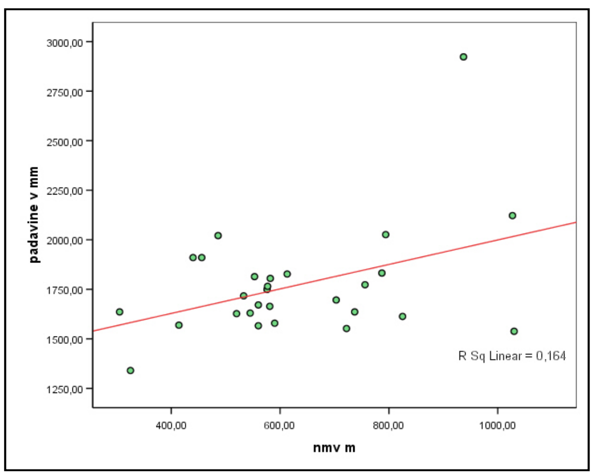
Stopnja povezanosti r

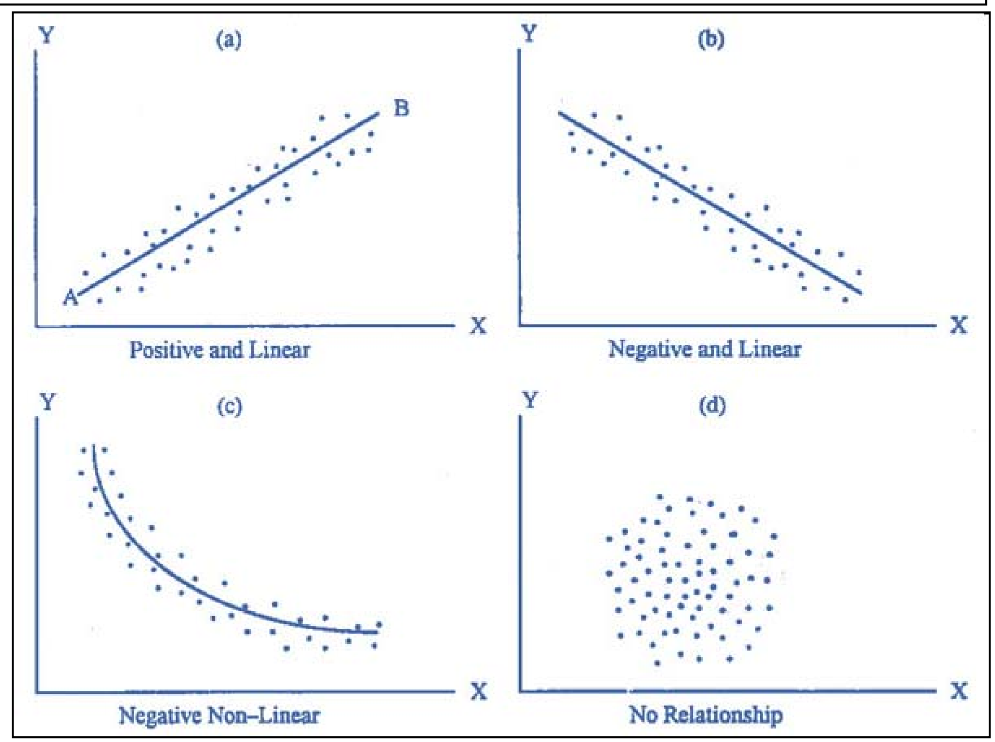
- Stopnja (jakost, moč) povezanosti (intervali *r*) se običajno poimenujejo:
  - 0 ni povezanosti
  - |0–0,2|: neznatna (pozitivna / negativna) povezanost
  - |0,2–0,4|: nizka (šibka) povezanost
  - |0,4–0,7|: srednja (zmerna) povezanost
  - |0,7–0,9|: visoka povezanost
  - |0,9–1|: zelo visoka povezanost
  - 1: popolna (funkcijska) povezanost

Razsevni grafikon ali korelacijski diagram



Razsevni diagram je grafični prikaz povezanosti med dvema spremenljivkama.
Regresijska črta kaže obliko povezanosti:
- Linearna
- Nelinearna (krivuljna)





Determinacijski koeficient - primer

- Povezanost pretokov Reke 1 in Reke 2
- Ker je r = 0,988 je r^2 = 0,976 oz. 97,6 %)
- S spreminjanem pretokov Reke 1 je pojasnjeno 98 % spreminjanja (variance) pretokov Reke 2 in obratno

Sklepanje povezanosti spremenljivk v populaciji iz vzorca

- H_0: $\sigma = 0$ - spremenljivki nista linerano povezani
- H_1: $\sigma \neq 0$ - spremenljivki sta linerano povezani
- Testna statistika za ničelno domnevo je:

$$t = \frac{r \cdot \sqrt{n}-2}{\sqrt{1}-r^2}$$

- Ničelna porazdelitev je $t(SP = n - 2)$
  - Preračunavanju se izognemo s pomočjo preglednice

V tabeli je za število enot v vzorcu **n** in za verjetnost α navedena kritična absolutna vrednost ocene Pearsonovega koeficienta korelacije r, pri kateri zavrnemo ničelno domnevo **H**$_0$: ρ = 0 pri dvostranskem preizkusu.

Primer: α = 0,05, n = 8
r = 0,7067 oziroma r = -0,7067

	α			α			α	
n	**0,05**	**0,01**	**n**	**0,05**	**0,01**	**n**	**0,05**	**0,01**
1	-	-	**31**	0,3550	0,4556	**61**	0,2521	0,3274
2	-	-	**32**	0,3494	0,4487	**62**	0,2500	0,3248
3	0,9969	0,9999	**33**	0,3440	0,4421	**63**	0,2480	0,3223
4	0,9500	0,9900	**34**	0,3388	0,4357	**64**	0,2461	0,3198
5	0,8783	0,9587	**35**	0,3338	0,4296	**65**	0,2441	0,3173
6	0,8114	0,9172	**36**	0,3291	0,4238	**66**	0,2423	0,315
7	0,7545	0,8745	**37**	0,3246	0,4182	**67**	0,2404	0,3126
8	0,7067	0,8343	**38**	0,3202	0,4128	**68**	0,2387	0,3104
9	0,6664	0,7977	**39**	0,3160	0,4076	**69**	0,2369	0,3081
10	0,6319	0,7646	**40**	0,3120	0,4026	**70**	0,2352	0,3060
11	0,6021	0,7348	**41**	0,3081	0,3978	**71**	0,2335	0,3038
12	0,5760	0,7079	**42**	0,3044	0,3932	**72**	0,2319	0,3017
13	0,5529	0,6835	**43**	0,3008	0,3887	**73**	0,2303	0,2997
14	0,5324	0,6614	**44**	0,2973	0,3843	**74**	0,2287	0,2977
15	0,5140	0,6411	**45**	0,2940	0,3801	**75**	0,2272	0,2957
16	0,4973	0,6226	**46**	0,2907	0,3761	**76**	0,2257	0,2938
17	0,4821	0,6055	**47**	0,2876	0,3721	**77**	0,2242	0,2919
18	0,4683	0,5897	**48**	0,2845	0,3683	**78**	0,2227	0,2900
19	0,4555	0,5751	**49**	0,2816	0,3646	**79**	0,2213	0,2882
20	0,4438	0,5614	**50**	0,2787	0,3610	**80**	0,2199	0,2864
21	0,4329	0,5487	**51**	0,2759	0,3575	**82**	0,2172	0,2830
22	0,4227	0,5368	**52**	0,2732	0,3542	**84**	0,2146	0,2796
23	0,4132	0,5256	**53**	0,2706	0,3509	**86**	0,2120	0,2764
24	0,4044	0,5151	**54**	0,2681	0,3477	**88**	0,2096	0,2732
25	0,3961	0,5052	**55**	0,2656	0,3445	**90**	0,2072	0,2702
26	0,3882	0,4958	**56**	0,2632	0,3415	**92**	0,2050	0,2673
27	0,3809	0,4869	**57**	0,2609	0,3385	**94**	0,2028	0,2645
28	0,3739	0,4785	**58**	0,2586	0,3357	**96**	0,2006	0,2617
29	0,3673	0,4705	**59**	0,2564	0,3328	**98**	0,1986	0,2591
30	0,3610	0,4629	**60**	0,2542	0,3301	**100**	0,1966	0,2565

(Vir: Henry R. Neave, Elementary Statistics Tables, University of Nottingham, 1979)

Sklepanje povezanosti spremenljivk v populaciji iz vzorca - primer na slidu 11

- Pri stopnji značilnosti α = 0,01 preverimo domnevo o povezanosti pretokov reke 1 in reke 2
- H_0: $\sigma = 0$ - pretoka rek nista linerano povezana
- H_1: $\sigma \neq 0$ - petoka rek sta linerano povezana

$r = \frac{29{,}406}{\sqrt{24{,}61} \cdot 1{,}24} = 0{,}988$

$N = 21$

- $r_{tab} = 0{,}5487$
- 0,988 > 0,5487 → ničelno domnevo zavržemo v korist alternativne domneve
- Pri stopnji značilnosti α = 0,01 trdimo, da obstaja pozitivna linearna povezava med pretokoma reke 1 in reke 2

V tabeli je za število enot v vzorcu **n** in za verjetnost α navedena kritična absolutna vrednost ocene Pearsonovega koeficienta korelacije r, pri kateri zavrnemo ničelno domnevo $\mathbf{H_0}$: $\rho = 0$ pri dvostranskem preizkusu.

Primer: α = 0,05, n = 8
r = 0,7067 oziroma r = -0,7067

n	α 0,05	α 0,01	n	α 0,05	α 0,01	n	α 0,05	α 0,01
1	-	-	**31**	0,3550	0,4556	**61**	0,2521	0,3274
2	-	-	**32**	0,3494	0,4487	**62**	0,2500	0,3248
3	0,9969	0,9999	**33**	0,3440	0,4421	**63**	0,2480	0,3223
4	0,9500	0,9900	**34**	0,3388	0,4357	**64**	0,2461	0,3198
5	0,8783	0,9587	**35**	0,3338	0,4296	**65**	0,2441	0,3173
6	0,8114	0,9172	**36**	0,3291	0,4238	**66**	0,2423	0,315
7	0,7545	0,8745	**37**	0,3246	0,4182	**67**	0,2404	0,3126
8	0,7067	0,8343	**38**	0,3202	0,4128	**68**	0,2387	0,3104
9	0,6664	0,7977	**39**	0,3160	0,4076	**69**	0,2369	0,3081
10	0,6319	0,7646	**40**	0,3120	0,4026	**70**	0,2352	0,3060
11	0,6021	0,7348	**41**	0,3081	0,3978	**71**	0,2335	0,3038
12	0,5760	0,7079	**42**	0,3044	0,3932	**72**	0,2319	0,3017
13	0,5529	0,6835	**43**	0,3008	0,3887	**73**	0,2303	0,2997
14	0,5324	0,6614	**44**	0,2973	0,3843	**74**	0,2287	0,2977
15	0,5140	0,6411	**45**	0,2940	0,3801	**75**	0,2272	0,2957
16	0,4973	0,6226	**46**	0,2907	0,3761	**76**	0,2257	0,2938
17	0,4821	0,6055	**47**	0,2876	0,3721	**77**	0,2242	0,2919
18	0,4683	0,5897	**48**	0,2845	0,3683	**78**	0,2227	0,2900
19	0,4555	0,5751	**49**	0,2816	0,3646	**79**	0,2213	0,2882
20	0,4438	0,5614	**50**	0,2787	0,3610	**80**	0,2199	0,2864
21	0,4329	0,5487	**51**	0,2759	0,3575	**82**	0,2172	0,2830
22	0,4227	0,5368	**52**	0,2732	0,3542	**84**	0,2146	0,2796
23	0,4132	0,5256	**53**	0,2706	0,3509	**86**	0,2120	0,2764
24	0,4044	0,5151	**54**	0,2681	0,3477	**88**	0,2096	0,2732
25	0,3961	0,5052	**55**	0,2656	0,3445	**90**	0,2072	0,2702
26	0,3882	0,4958	**56**	0,2632	0,3415	**92**	0,2050	0,2673
27	0,3809	0,4869	**57**	0,2609	0,3385	**94**	0,2028	0,2645
28	0,3739	0,4785	**58**	0,2586	0,3357	**96**	0,2006	0,2617
29	0,3673	0,4705	**59**	0,2564	0,3328	**98**	0,1986	0,2591
30	0,3610	0,4629	**60**	0,2542	0,3301	**100**	0,1966	0,2565

(Vir: Henry R. Neave, Elementary Statistics Tables, University of Nottingham, 1979)

Uporaba korelacijskega koeficienta

- Previdnost, ko povezanost ni linearna → takrat uporaba korelacijskega koeficienta kot mere povezanosti dveh številskih spremenljivk ni ustrezna
- Povezanosti dveh številskih spremenljivk so lahko nelinearne (MS Excel jih zna računati)
- Pri kvadratni porazdelitvi bi izračun Pearsonovega koeficienta korelacije pokazal vrednost okrog 0 → obrazložitev, da povezave med spremenljivkama ni, bi bila napačna
  - Povezava obstaja, vendar ni linearna
- Ko predpostavka o linearni povezanosti za dve številski spremenljivki ni utemeljena, s e uporabi druge mere povezanosti , ki ne temeljijo na linearnosti
- Npr. Koeficient korelacije rangov ali Spearmanov koeficient korelacije (r_s)

Literatura in viri

- Ferligoj, Anuška. 1995. *Osnove statistike na prosojnicah.* Ljubljana: Samozaložba Z. Batagelj.
- Statistika. 2013. »*Electronic statistic textbook, StatSoft* «. Http://www.statsoft.com.
- Rogerson, Peter A. 2006. *Statistical Methods for Geography: a student guide*. London: Sage Publications.
- Kastelec, Damijana in Katarina Košmelj, 2010. *Osnove statistike z Excelom 2007.* Ljubljana: Biotehniška fakulteta. - Dostopno tudi na medmrežju.