

Osnove statistike v geografiji 1

- Metodologija geografskega raziskovanja -

dr. Gregor Kovačič, doc.

Kvantitativne raziskave

- Izdelovanje modelov, teorij, hipotez
- Razvoj merilnih naprav in merilnih metod
- Kontrola s poskusi in ravnanje s spremenljivkami
- Zbiranje empiričnih podatkov
- Modeliranje in analiza podatkov
- Evalvacija rezultatov

Kvantitativne metode

- Raziskovalne tehnike, za zbiranje kvantitativnih podatkov
 - Informacije izražene s številkami
 - Vse, kar se da izmeriti
- Za prikazovanje in opisovanje podatkov, rezultatov se uporablja:
 - Grafično prikazovanje
 - Diagrami, grafi
 - Preglednice
 - Tematske karte
 - Numerično prikazovanje
 - Različne statistike, srednje vrednosti,

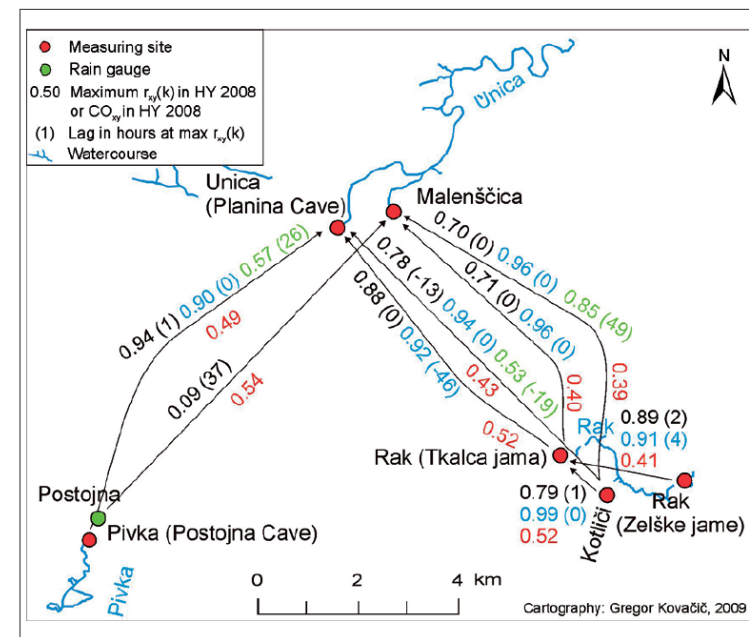
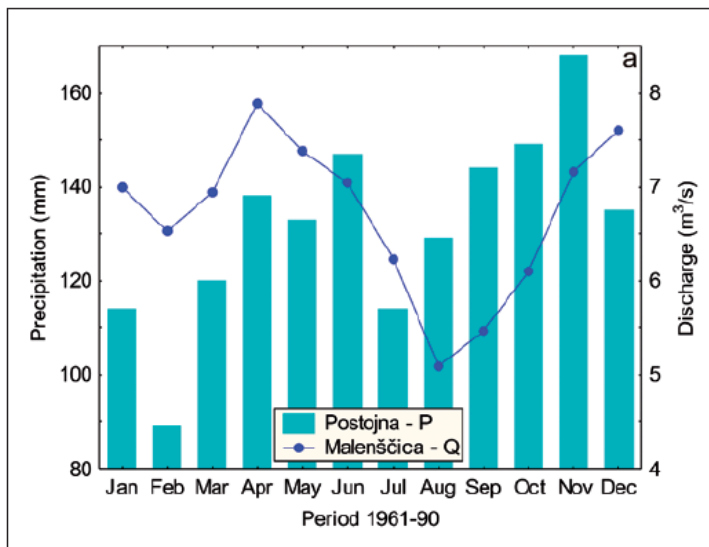


Fig. 8: Cross-correlation coefficients for the hourly discharge (black), T (blue) and EC (green) time series in the hydrological year 2008. Discharge-discharge, water level-discharge and precipitation-discharge coherence function values are indicated in red.

Tab. 2: Key characteristics of the Malenščica spring as well as other springs and watercourses included in this study in the hydrological year 2008, and characteristic lag times (hours) where the autocorrelation coefficient exceeds the 0.2 value for all measuring sites. The locations of the measuring sites are shown in Fig. 2.

Measuring site Spring or watercourse	Mean	Minimum	Maximum	Characteristic lag (hours) ($r_s = 0.2$)
Malenščica spring capture				
Discharge (m ³ /s)	5.8	1.60	9.4	557
Electric conductivity (μS/cm)	378	338	428	more than 999
Temperature (°C)	9.8	5.0	17.7	more than 999
Unica spring (Planina Cave)				
Discharge (m ³ /s)	9.1	0.16	58.03	592
Electric conductivity (μS/cm)	380	287	477	332
Temperature (°C)	9.7	4.2	14.7	2,094
Kotlički spring				
Water level (cm)	219	41	738	292
Electric conductivity (μS/cm)	370	201	447	726
Temperature (°C)	9.2	2.1	22.2	1,477
Rak spring (Zelške jame cave)				
Water level (cm)	37	0	174	635
Temperature (°C)	10.6	0.1	21.2	1,261
Rak ponor (Tkalca jama cave)				
Water level (cm)	344	5	2,259	481
Temperature (°C)	11.1	2.3	22.8	1,971
Pivka ponor (Postojna Cave)				
Discharge (m ³ /s)	3.2	0.50	31.23	542
Electric conductivity (μS/cm)	394	204	574	248
Temperature (°C)	8.3	0.1	21.9	1,803



Kvantitativne raziskave in statistika

- Statistika je zelo uporabljana veja matematike v kvantitativnih raziskavah
- Zbiranje podatkov (npr. podatki o T zraka, podatki o koncentraciji CO₂) na osnovi modelov in teorij → velikost vzorca
- Preučevanje vzročno-posledičnih povezav, odvisne in neodvisne spremenljivke
- Preučevanje povezanosti spremenljivk, ki pa ne izraža nujno tudi povzročitev; navidezne povezave

Osnovni pojmi iz statistike

- Enota
- Populacija
- Vzorec
- Spremenljivka

Enota

- Posamezni preučevani element
- Najmanjša enota preučevanja
- Statistična enota \neq enota merjenja
- Enote opazovanja so lahko:
 - Občine,
 - KS
 - Države
 - Porečje
 - Posamezna rastlina
 - Vodomerna postaja
 - Klimatološka postaja
 - Prodnik
 - Vzorci s terena
 - Pikel, pridobljen z daljinskim zaznavanjem

Populacija

- Množica vseh preučevanih elementov
- Populacija je lahko:
 - Sedimentno gradivo
 - Mreža klimatoloških postaj
 - Mreža vodomernih postaj

Vzorec

- Podmnožica populacije, na osnovi katere ponavadi sklepamo o lastnostih cele populacije
- Primer: populacija je sedimentno gradivo na kraškem polju; vzorec je v tem primeru izkopano gradivo na izbranih mestih, ki ga preučimo, in na osnovi značilnosti preučenega gradiva sklepamo o sedimentni zgradbi celotnega kraškega polja → indukcija

Spremenljivka

- Statistična spremenljivka je merljiva lastnost ali atribut posamezne statistične enote
- Spremenljivka navadno označuje lastnost, ki se skozi čas, prostor ali med različnimi statističnimi enotami spreminja
- Zvezne in nezvezne (diskretne) spremenljivke (glej slide naprej)
- Izkopano gradivo na izbranem mestu (enota) preučimo na osnovi več spremenljivk, npr.:
 - Analiza zrnatosti
 - Zaobljenost gradiva
 - Okroglost gradiva
 - Velikost delcev
 - Litološka sestava izvornega gradiva
 - Položaj v izkopanem profilu
 - Itd.

Vrste spremenljivk glede na vsebino

- Krajevne (geografske) - spreminjajo se krajevno, navezujejo se na lokacijo, npr. koeficient gozdnatosti v posameznih porečjih
- Časovne - spreminjajo se s časom, navezujejo se na čas, npr. spreminjanje pretoka na vodomerni postaji
- Stvarne ali vsebinske - sem sodijo vse ostale oblike spremenljivk, delimo jih glede na to, kako lahko izražamo njihovo vrednost

Vrste spremenljivk glede na tip izražanja vrednosti

- Opisne (atributivne)
 - Vrednosti lahko opišemo samo z besedo (npr. barva, oblika, spol)
- Številске (numerične)
 - Vrednosti lahko izrazimo s števili (npr. starost, masa, nadmorska višina)

Zvezne in nezvezne (diskretne) številске spremenljivke

- Zvezna - številška spremenljivka, ki ima lahko v določeni množici možnih vrednosti katerokoli vrednost, na primer: starost, vodostaj, čas, dolžina, teža, ...
- Nezvezna - številška spremenljivka, ki ima končno vrednost. Na primer: šolska ocena, število dreves na prostorsko enoto, število pojedениh jabolok, število skokov, ... Običajno so izražene s celimi števili (1, 4, 7, 23, ...).

Odvisne in neodvisne spremenljivke

- Neodvisna spremenljivka vpliva na odvisno spremenljivko
 - Razlog, vzrok, sila, ki povzroči spremembo odvisne spremenljivke
- Odvisna spremenljivka se spreminja zaradi odziva na spreminjanje neodvisne spremenljivke
- Na odvisno spremenljivko lahko vpliva več neodvisnih spremenljivk

Vrste spremenljivk glede na tip merjenja

- Nominalne
- Ordinalne
- Intervalne
- Razmernostne (kardinalne)

Glede na tovrstni tip statistične spremenljivke je možno različno število statističnih analiz. Najmanj jih je moč opraviti z nominalnimi spremenljivkami, največ pa z razmernostnimi.

Nominalne spremenljivke

- Vrednosti niso razdeljene hierarhično, pač pa enakovredno
- Določimo lahko le, ali se dve spremenljivki med seboj razlikujeta
- Ne moremo določiti manj ali bolj vredne izmed spremenljivk (npr. kraj bivanja, tip kamnine, tip prsti)

Ordinalne spremenljivke

- Vrednosti so razdeljene hierarhično
- **Možno je vrednostno razvrščanje spremenljivk**
- Ni možno določanje količine razlik med njimi (npr. kategorije onesnaženosti rek, šolske ocene, ...)

Intervalne spremenljivke

- Vrednosti so razdeljene hierarhično in po skali (merilni lestvici)
- Mogoče je določiti razliko med posameznimi vrednostmi
- Te spremenljivke nimajo določene absolutne ničelne vrednosti (npr. temperatura v °C, geografska dolžina, ...)

Razmernostne spremenljivke

- Hierarhično razdeljene vrednosti, označene po skali (merski lestvici)
- Imajo tudi absolutno ničlo (npr. količina padavin, globina vrtače, masa prodnika, ...)

Statistika

- Opisna (deskriptivna) statistika
 - Zbiranje, urejanje in prikazovanje podatkov
- Sklepna (inferenčna, analitična) statistika
 - Sklepanje iz podatkov, dobljenih na majhnih vzorcih

Univariantna, bivariantna ali
multivariantna analiza

Opisna statistika

- Je skupina statističnih metod, ki se ukvarjajo s kvantitativnim opisovanjem glavnih značilnosti podatkovne baze
- Metode iščejo opisne (meta) podatke o populaciji in njenih sestavnih delih, da bi ustvarile pregledni opis
- Je zbiranje, urejanje in prikazovanje podatkov

Temelji opisne statistike

- Grafični opis, katerega osnova je predstavitev s pomočjo grafov,
- Tabelarni opis, ki ustvarja pregled s pomočjo preglednic
- Statistični povzetki, ki na osnovi nekaterih računov predstavijo bistvene podatke
 1. Srednje vrednosti
 2. Mere razpršenosti

1. Srednje vrednosti

- Mediana M_e
- Modus M_o
- Aritmetična sredina μ (povprečje, mean)
- Geometrijska sredina G
- Harmonična sredina H

Čim bolj vrednosti variirajo, tem bolj se posamezne vrednosti odklanjajo od srednje vrednosti in tem slabše ta srednja vrednost predstavlja značilnosti spremenljivke.

Mediana M_e

- Tista vrednost spremenljivke, od katere ima polovica enot manjše, polovica pa večje vrednosti spremenljivke
 - Določanje, če rangirani (razvrščeni) oz. urejeni v ranžirni vrsti:
 - Če liho število: mediana enaka vrednosti srednje enote v ranžirni vrsti $N = 2m + 1$; $M_e = (m + 1)$
 - Če sodo število: mediana je povprečje srednjega para podatkov $N = 2m$;
- $$M_e = \frac{x_m + x_{m+1}}{2}$$
- Ni povezana z nobeno teoretično porazdelitvijo
 - Uporabno, ko je statistična masa porazdeljena nesimetrično
 - Primerna za ordinalne spremenljivke

Primer:

LIHO: 3,6,9,10,12,16,25; ker je $N = 7$, je mediana na 4. mestu in sicer $M_e = 10$

SODO: 3,6,9,10,12,16; ker je $N = 6$, je $M_e = 9,5$

Modus M_o

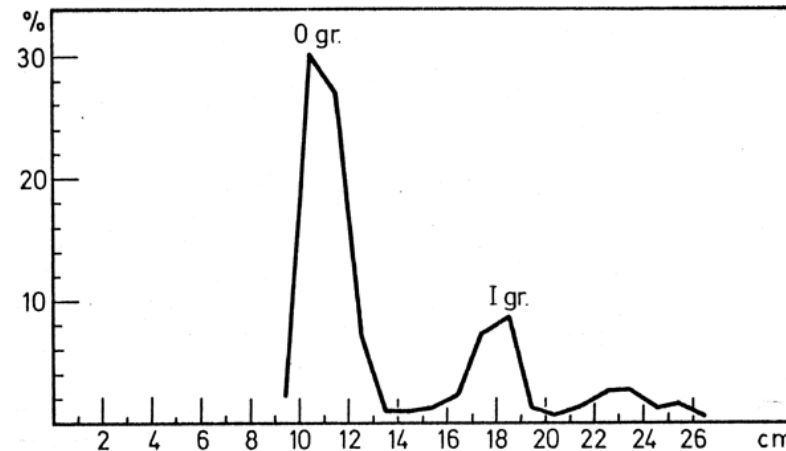
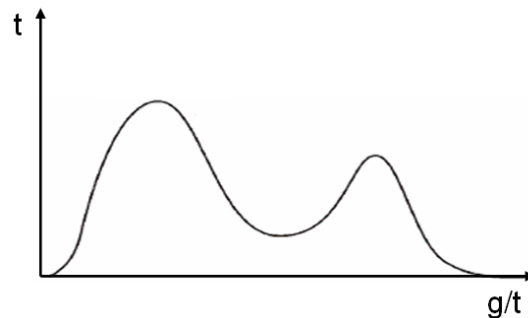
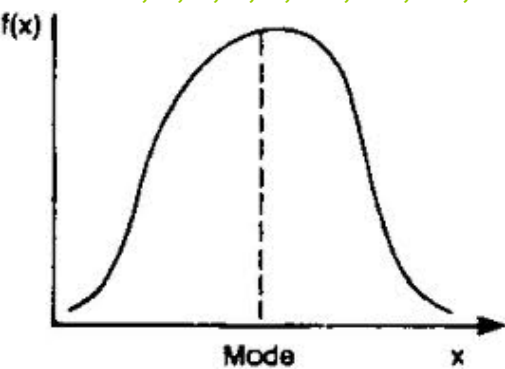
- Najpogostejša vrednost spremenljivke
- Ugotavljamo pri razmeroma veliki populaciji, pri manjših pa uporaba ni primerna
- Modusov je lahko več ali pa ni nobenega
- Primeren tudi za nominalne spremenljivke
- Opis porazdelitve populacije:
 - Unimodalna
 - Bimodalna
 - Polimodalna

Primer:

3,6,9,9,9,9,9,10,10,10,12,16,25; $Mo_1 = 9$; $Mo_2 = 10$

3,6,9,10,12,16; ni modusa

3,6,6,6,9,10,13,14; $Mo = 6$



Aritmetična sredina ali povprečje μ

- Najpogosteje uporabljena srednja vrednost
- Seštejemo vrednosti vseh enot in delimo s številom enot (μ oz. \bar{x}):
- Nanjo vplivajo posamezne vrednosti vsake statistične enote
- Vsota odklonov posameznih vrednosti od težišča navzgor je enaka tisti navzdol
- Vsota vseh odklonov od aritmetične sredine je enaka 0
- Primerna je za številske, približno normalno porazdeljene spremenljivke → povezana z normalno oz. Gaussovo porazdelitvijo

Primer:

1,2,3,4,5,6,7,8; $\mu = (1+2+3+4+5+6+7+8) / 8 = 4.5$

$$\sum_{i=1}^N (x_i - \mu) = 0$$

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

Geometrijska in harmonična sredina

- Geometrijska sredina je enaka N-temu korenu iz produkta N vrednosti številske spremenljivke, MS Excel izračuna samo do N = 30

$$\bar{X}_g = \sqrt[n]{\prod_{i=1}^n X_i}$$

Primer: 2,4,8 $G = \sqrt[3]{2 \times 4 \times 8} = \sqrt[3]{2^6} = 4$

- Harmonična sredina je enaka recipročni vrednosti aritmetične sredine, izračunane iz recipročnih vrednosti spremenljivke, MS Excel izračuna samo do N = 30

Formula: Harmonic Mean =
$$\frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} + \frac{1}{a_4} + \dots + \frac{1}{a_n}}$$

Example: For the numbers 4 and 9,
Harmonic Mean =
$$\frac{2}{\frac{1}{4} + \frac{1}{9}} = \frac{72}{13} = 5.54$$

$$\begin{aligned} \text{Arithmetic mean} &= \frac{x_1 + x_2 + \dots + x_n}{n} \\ \text{Geometric mean} &= \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} \\ \text{Harmonic mean} &= \frac{1}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} \end{aligned}$$

2. Mere razpršenosti (variacije, variabilnosti)

- Statistična razpršenost je enaka 0, če so vsi podatki enaki in narašča z različnostjo vrednosti podatkov
- Ugotavljamo, koliko posamezni podatki odstopajo od srednje vrednosti in s tem ocenimo, koliko srednja vrednost res predstavlja lastnosti preučevane populacije
- Absolutne in relativne mere razpršenosti

2.1. Absolutne mere razpršenosti

Enake enote kot spremenljivka

- Variacijski razmik ali razpon R
- Kvartilni odklon ali razmik Q
- Povprečni absolutni odklon AD
- Standardni odklon SD
- Varianca SD^2

2.1. Variacijski razmik ali razpon R

$$R = x_{\max} - x_{\min}$$

R - variacijski razmik

x_{\max} - največja vrednost

x_{\min} - najmanjša vrednost

PRIMER:

17, 5, 19, 13, 3, 5, 17, 13, 16, 23, 13, 7, 40, 10, 22, 9, 18, 6, 12, 9

R = 37

2.1. Kvartilni odklon Q

Q1 in Q3 – kvartila

- Kvartilni razmik (odklon) je interval, **znotraj** katerega je **polovica** podatkov, polovica pa **zunaj** njega: ena četrtnina je **pod njim**, ena četrtnina pa **nad njim**.

$$Q = \frac{Q_3 - Q_1}{2}$$

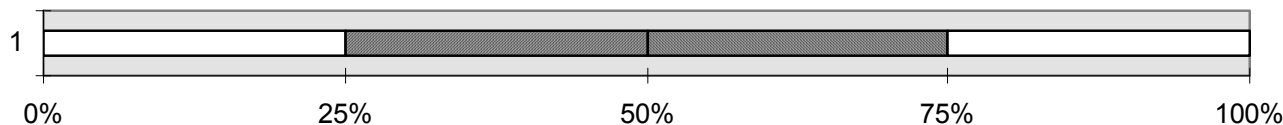
PRIMER:

Neurejeni podatki: 17, 5, 19, 13, 3, 5, 17, 13, 16, 23, 13, 7, 40, 10,
22, 9, 18, 6, 12, 9

Urejeni podatki: 3, 5, 5, 6, 7, 9, 9, 10, 12, 13, 13, 13, 16, 17, 17, 18,
19, 22, 23, 40

$$Q_1 = 7 \quad Q_3 = 17$$

$$Q = (17 - 7) / 2 = 5$$



2.1. Povprečni absolutni odklon AD

- Povprečni absolutni odklon je aritmetična sredina odklonov vrednosti posameznih podatkov od aritmetične sredine teh podatkov, vzetih ne glede na predznak (absolutno)

X_i – konkreten podatek

μ ali \bar{x} – aritmetična sredina

n – število podatkov

PRIMER:

17, 5, 19, 13, 3, 5, 17, 13, 16, 23, 13, 7

$\mu = 151/12 = 12,58$

AD = 60,68 / 12 = 5,06

$$\frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

X_i	$ X_i - \mu $
17	$ 17 - 12,58 = 4,42$
5	$ 5 - 12,58 = 7,58$
19	$ 19 - 12,58 = 6,42$
13	$ 13 - 12,58 = 0,42$
3	$ 3 - 12,58 = 9,58$
.	.
SKUPAJ	$\Sigma 60,68$

2.1. Varianca in standardni odklon SD

- Standardni odklon ali standardna deviacija (SD, sigma) je izveden iz variance
- **Varianca pa je povprečje kvadriranih odklonov posameznih podatkov od aritmetične sredine**

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

$$SD = \sqrt{\sigma^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

X_i – konkreten podatek

μ ali \bar{x} – aritmetična sredina

n – število podatkov

PRIMER:

17, 5, 19, 13, 3, 5, 17, 13, 16, 23, 13, 7

$$\mu = 146/12 = 12,58$$

$$SD = \sqrt{438,91/12} = \sqrt{36,58} = 6,05$$

X_i	$(X_i - \mu)^2$
17	$(17 - 12,58)^2 = 19,53$
5	$(5 - 12,58)^2 = 57,46$
19	$(19 - 12,58)^2 = 41,22$
13	$(13 - 12,58)^2 = 0,18$
3	$(3 - 12,58)^2 = 91,78$
.	.
SKUPAJ	$\Sigma 438,91$

2.2. Relativne mere razpršenosti

- Relativni variacijski razmik R_R
- Relativni kvartilni odklon R_Q
- Relativni povprečni absolutni odklon R_{AD}
- Relativni standardni odklon – koeficient variacije - KV

2.2. Relativne mere razpršenosti

relativni variacijski razmik R_R

$$(x_{\max} - x_{\min}) \times 2 / (x_{\max} + x_{\min})$$

PRIMER:

17, 5, 19, 13, 3, 5, 17, 13, 16, 23, 13, 7, 40, 10, 22, 9, 18, 6,
12, 9

$$R = 37 \times 2 / 43 = 1,72$$

relativni kvartilni odklon R_Q

$$R_Q = \frac{Q_3 - Q_1}{2 \cdot Me}$$

PRIMER:

3, 5, 5, 6, 7, 9, 9, 10, 12, 13, 13, 13, 16, 17, 17, 18, 19, 22,
23, 40

$$Me = 13$$

$$RQ = (17 - 7) / 2 \times 13 = 0,38$$

2.2. Relativne mere razpršenosti

relativni povprečni absolutni odklon R_{AD}

$$R_{AD} = AD / Me$$

PRIMER:

17, 5, 19, 13, 3, 5, 17, 13, 16, 23, 13, 7

$$Me = 13$$

$$R_{AD} = 5,06 / 13 = 0,39$$

X_i	$ X_i - \mu $
17	$ 17 - 12,58 = 4,42$
5	$ 5 - 12,58 = 7,58$
19	$ 19 - 12,58 = 6,42$
13	$ 13 - 12,58 = 0,42$
3	$ 3 - 12,58 = 9,58$
.	.
SKUPAJ	$\Sigma 60,68$

2.2. Relativni standardni odklon ali koeficient variacije KV

- Mera relativnega variiranja: standardni odklon utežen z aritmetično sredino, ponavadi podano kot odstotek (množeno s 100 %)
- Ko želimo primerjati variiranje različnih spremenljivk, ki so med seboj v vsebinski zvezi npr. pretoki manjšega in večjega izvira

KV - koeficient variacije

μ - aritmetična sredina (povprečje)

σ - standardni odklon

PRIMER:

17, 5, 19, 13, 3, 5, 17, 13, 16, 23, 13, 7

$\mu = 146/12 = 12,58$

SD ali $\sigma = \sqrt{438,91/12} = \sqrt{36,58} = 6,05$

$KV = 6,05 / 12,58 = 0,48$

$$KV = \frac{\sigma}{\mu}$$

Primernost uporabe posameznih mer razpršenosti glede na tip spremenljivk

Podatki	Mere razpršenosti			
	Q	R	AD	SD
Nominalni				
Ordinalni	✓			
Intervalni	✓	✓	✓	✓

Literatura in viri

- Ferligoj, A., 1995: Osnove statistike na prosojnicah. Samozaložba z Batagelj.
- Statistika, 2007: Electronic statistic textbook, StatSoft. [Dostopno na medmrežju.](#)
- Rogerson, P. A., 2000: Statistical Methods for Geography. Sage Publications Inc.
- Kastelec. D. in K. Košmelj, 2010: Osnove statistike z Excelom 2007. [Dostopno na medmrežju.](#)