

POVEZANOST MED POJAVI:

**KOVARIANCA, REGRESIJA
IN KORELACIJA**

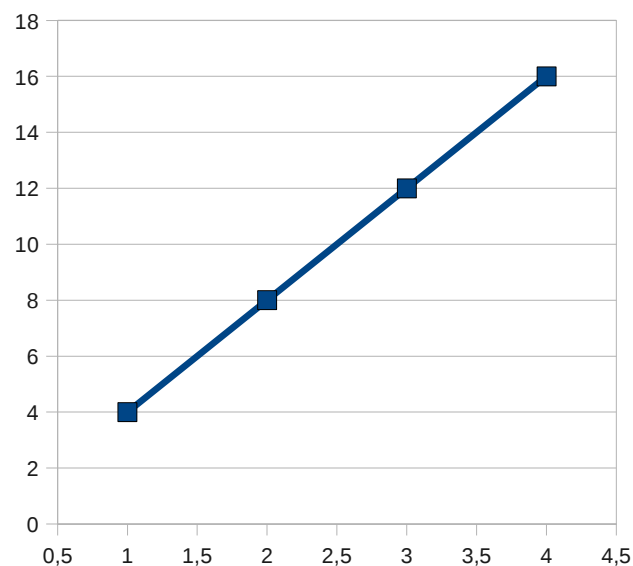
POVEZANOST MED POJAVI

- Pojavi so med seboj povezani
 - _V Sloveniji ne moremo gojiti pomaranč, ker drevesa pozebejo
 - _V gospodarski krizi je več socialnih stisk
 - _V mestih se gripa širi hitreje kot na podeželju
- „Večji/manjši, je x, večji/manjši je y“
 - _Dlje kot spiš, večja je možnost, da zamudiš na predavanja
 - _Z višjo plačo je lažje preživeti mesec
 - _Več se učiš, višjo oceno dobiš
 - _Več poješ, več tehtaš

FUNKCIJSKA IN STATISTIČNA POVEZANOST POJAVOV

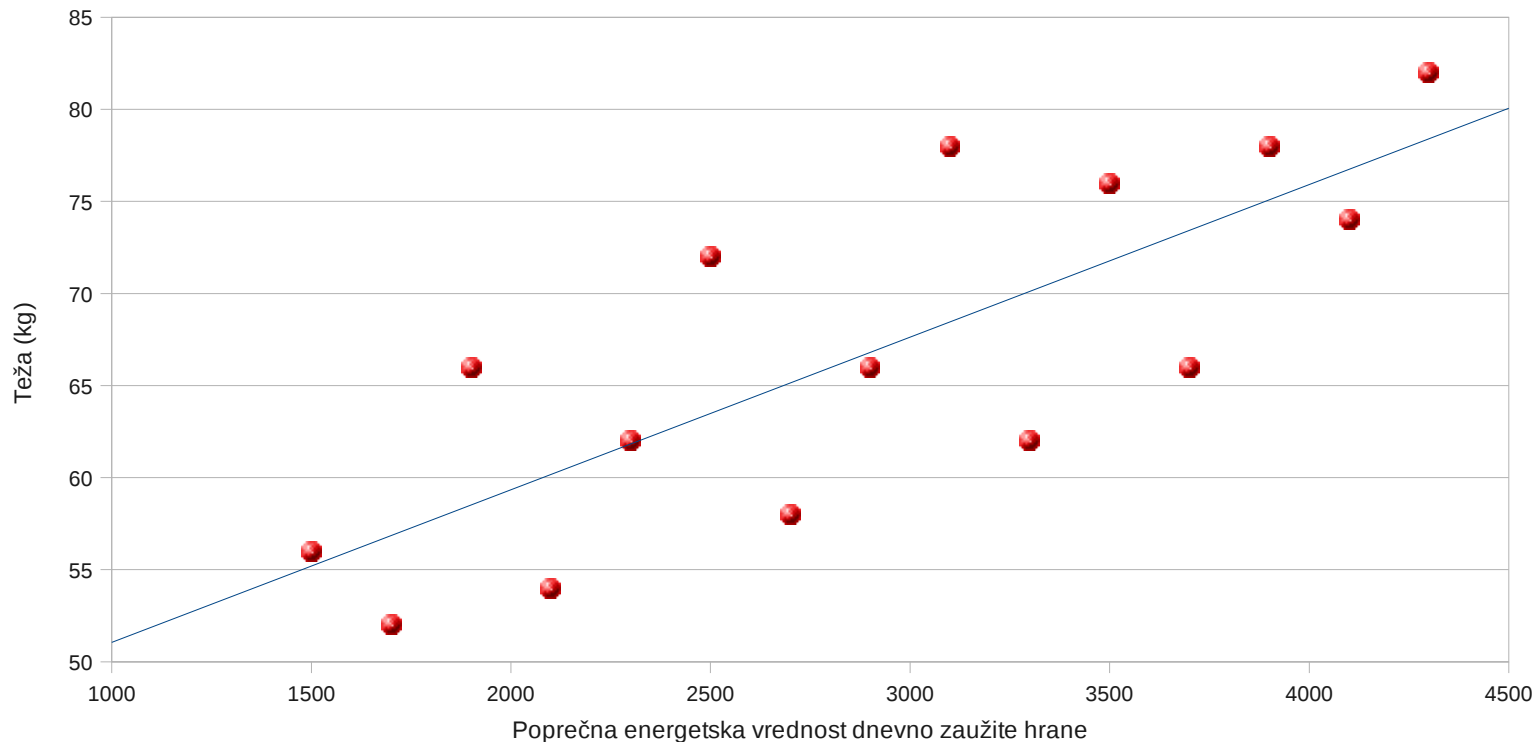
- **FUNKCIJSKA (linearna) POVEZANOST:**
- Dlje časa hodiš, dlje prideš
 - Če hodimo s hitrostjo 4 km/uro pridemo:

URE	KILOMETRI
1	4
2	8
3	12
4	16



FUNKCIJSKA IN STATISTIČNA POVEZANOST POJAVOV

- **STATISTIČNA (linearna) POVEZANOST:**
- Več ješ, več tehtaš



P O J M I

- **KOVARIANCA** (skupno variranje) - **skupne razlike**:
 - *podobnost odklonov* dveh spremenljivk od povprečja.
- **REGRESIJA** – **napovedovanje** ene spremenljivke na podlagi druge:
 - Kakšno *vrednost* odvisne spremenljivke ***pričakujemo***, če poznamo vrednost neodvisne spremenljivke.
- **KORELACIJA** – **moč povezave** med spremenljivkama:
 - Kako *podobni* sta si dve spremenljivki.

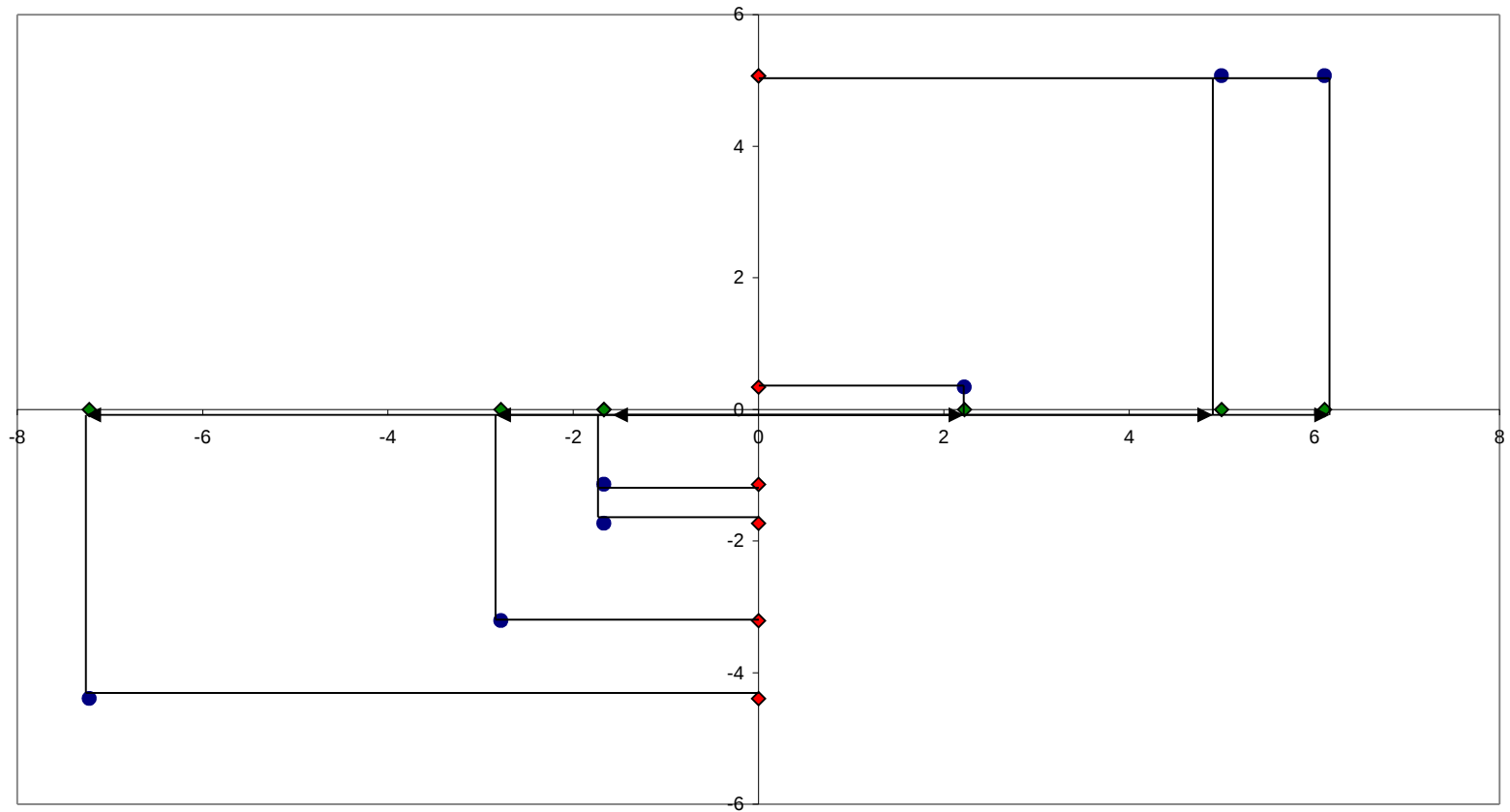
KOVARIANCA

- skupne razlike
- Ugotavljanje **podobnosti odklonov** od povprečja med dvema spremenljivkama.

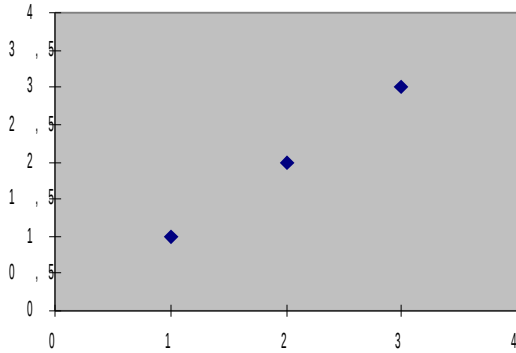
$$C_{xy} = \frac{\sum_{i=1}^n (x - \bar{x}) \times (y - \bar{y})}{n}$$

KOVARIANCA

Odkloni



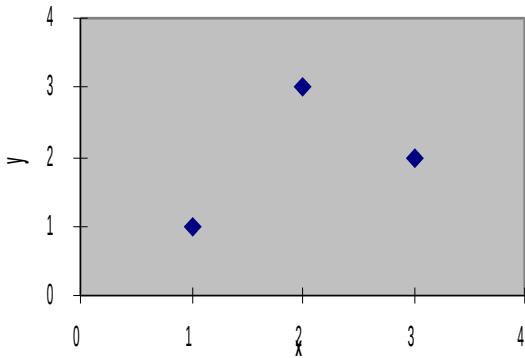
KOVARIANCA



X	Y	d_X	d_Y	$d_X \cdot d_Y$
1	1	-1	-1	1
2	2	0	0	0
3	3	1	1	1

$$C_{xy} = \frac{\sum_{i=1}^n (x - \bar{x}) \times (y - \bar{y})}{n} = \frac{2}{3} = 0,67$$

KOVARIANCA



X	Y	d _X	d _Y	d _X · d _Y
1	1	-1	-1	1
2	3	0	1	0
3	2	1	0	0

$$C_{xy} = \frac{\sum_{i=1}^n (x - \bar{x}) \times (y - \bar{y})}{n} = \frac{1}{3} = 0,33$$

KOVARIANCA

X	Y	d_X	d_Y	$d_X \cdot d_Y$
1	100	-1	-100	100
2	300	0	100	0
3	200	1	0	0

$$C_{XY} = 33,33$$

Velikost kovariance je odvisna od enote merjenja – potrebno jo je **standardizirati**

REGRESIJA

NAPOVEDOVANJE

- Kakšno vrednost **odvisne** spremenljivke lahko **pričakujemo**, če **poznamo** vrednost **neodvisne** spremenljivke.
- Odnosi, ki jih ugotavljamo z regresijsko analizo:

“Večji, ko je x , večji je y ”,

ali

“Večji, ko je x , manjši je y ”

REGRESIJA

Regresijska premica:

Premica, od katere so točke skupno najmanj oddaljene

$$y = a + b \times x + e$$

$$y' = \bar{y} + b(x - \bar{x})$$

REGRESIJA

Regresijski koeficient:

Kazalec strmine premice

$$b = \frac{C_{xy}}{\sigma_x^2} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

Regresijski koeficienti so le delno standardizirani
Povedo nam, koliko se **spreminja odvisna** spremenljivka,
če se **neodvisna** spremenljivka **spremeni za eno enoto**

KORELACIJA

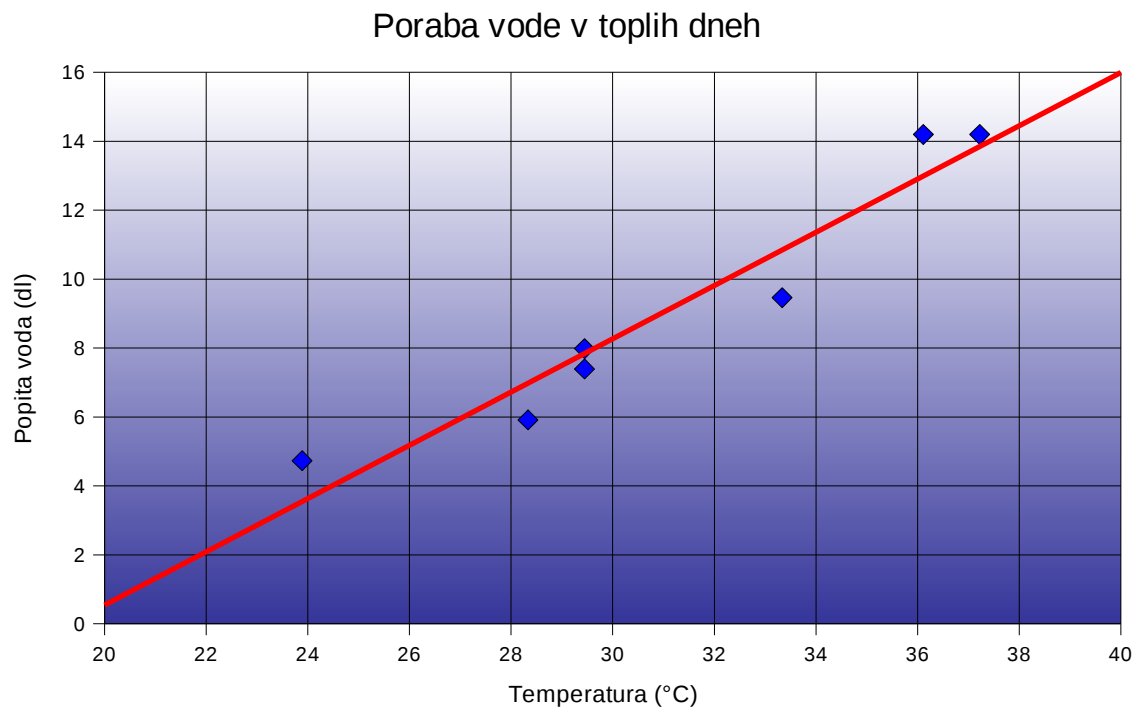
- Povezanost med dvema spremenljivkama (korelacija: „skupna usmerjenost”)
 - Kako **podobni** sta si med sabo dve spremenljivki
 - Dokončna standardizacija kovariance (od -1 do 1)
 - Kosinus kota med regresijskima premicama

$$r = \sqrt{b_x b_y} = \sqrt{\frac{C_{xy} C_{xy}}{\sigma_x^2 \sigma_y^2}} = \frac{C_{xy}}{\sigma_x \sigma_y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

REGRESIJA

PRIMER: Koliko pijače rabimo za žur?

Temperatura (°C)	Popita voda (dl)
37	14
29	8
36	14
24	5
33	9
29	7
28	6



REGRESIJA: izračun

	Temperatura t (°C)	Popita voda v (dl)	$t - \bar{t}$	$v - \bar{v}$	$(t - \bar{t})^2$	$(v - \bar{v})^2$	$(t - \bar{t})^2 \cdot (v - \bar{v})^2$
	37	14	6,11	5,07	37,35	25,70	30,98
	29	8	-1,67	-1,14	2,78	1,30	1,90
	36	14	5,00	5,07	25,00	25,70	25,35
	24	5	-7,22	-4,39	52,16	19,31	31,73
	33	9	2,22	0,34	4,94	0,11	0,75
	29	7	-1,67	-1,73	2,78	3,00	2,89
	28	6	-2,78	-3,21	7,72	10,31	8,92
Suma	218	64	0	0	133	85	103
Aritm.sred	31,11	9,13					

REGRESIJA: izračun

$$b = \frac{C_{xy}}{\sigma_x^2} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} = \frac{103}{133} = 0,77$$

$$y' = \bar{y} + b(x - \bar{x}) = 9,13 + 0,77(x - 33,11)$$