25.05.2026.

КОВАРИЈАНСА И КОРЕЛАЦИЈА

Деф. Апсолутна мера степена зависности две сл. вел. назива се КОВАРИЈАНСА

$$\operatorname{cov}(X,Y) = E(X-E(X))(Y-E(Y))$$

може се показати и:

$$\operatorname{cov}(X,Y) = E(X\cdot Y) - E(X)\cdot E(Y)$$

- X,Y независне $\Rightarrow$ $\operatorname{cov}(X,Y)=0$

али обрнуто не мора да важи,
може бити $\operatorname{cov}(X,Y)$ а да су $X\,Y$ зависне!

<u>пример</u> $X: \begin{pmatrix} -2 & -1 & 1 & 2 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$, $Y = X^2$

- X,Y види се зависност из функционалне везе

$\operatorname{cov}(X,Y) = ?$ $\qquad E(XY) - E(X)\cdot E(X)$

$E(X) = \ldots = 0$

$E(Y) = ?$

$$Y: \begin{pmatrix} 4 & 1 & 1 & 4 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

$$\downarrow$$

$$Y: \begin{pmatrix} 1 & 4 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \qquad E(Y) = \frac{1}{2} + 2 = 2.5$$

$$E(X\overset{X^2}{\overset{\downarrow}{Y}}) = E(X \cdot X^2) = E(X^3)$$

$$X: \begin{pmatrix} -2 & -1 & 1 & 2 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$$

$$X^3: \begin{pmatrix} -8 & -1 & 1 & 8 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix} \Rightarrow E(X^3) = 0$$

$$\operatorname{cov}(X,Y) = \underbrace{E(XY)}_{0} - \underbrace{E(X)}_{0}\,\underbrace{E(Y)}_{2.5} = 0$$

пример $\quad X: \begin{pmatrix} 1 & 2 \\ 1/2 & 1/2 \end{pmatrix} \quad Y = X^2 \leftarrow$ зависне су!

$\operatorname{cov}(X,Y) = ?$ $\qquad E(X) = \frac{3}{2}$, $E(Y) = ?$

$$Y: \begin{pmatrix} 1 & 4 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \qquad E(Y) = 2{,}5$$

$$E(XY) = E(X \cdot X^2) = E(X^3) = 4{,}5$$

$$X^3: \begin{pmatrix} 1 & 8 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

$$\operatorname{cov}(X,Y) = 4{,}5 - \frac{3}{2} \cdot 2{,}5 \neq 0$$

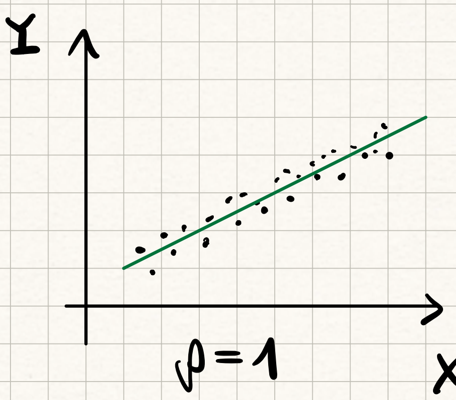
За испит СТЕПЕН ЗАВИСНОСТИ некада доби-
јамо веће вредности коваријансе ...
због тога уводимо нову меру

КОЕФ. КОРЕЛАЦИЈЕ

ДЕФ: РЕЛАТИВНА МЕРА степена зависности две сл. променљиве X, Y назива се КОЕФ. КОРЕЛАЦИЈЕ и рачуна се по формули

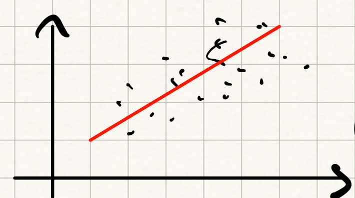
$$\rho(X,Y) = \frac{cov(X,Y)}{\underbrace{\sqrt{D(X)}}_{\sigma(X)}\ \underbrace{\sqrt{D(Y)}}_{\sigma(Y)}} \leftarrow \text{ПИРСОНОВ КОЕФ.}$$

Шта нам говори ρ?

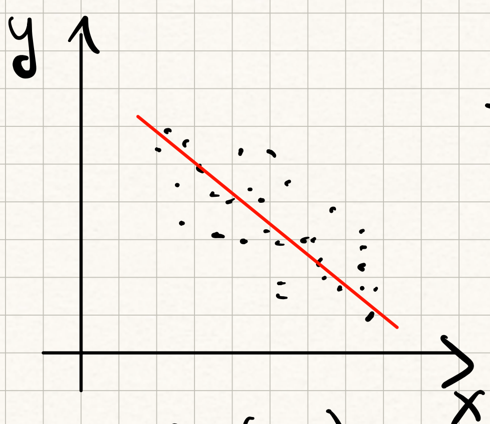


$\rho = 1$

← лин. зависност
– директно пропорционална (расте X, расте Y)

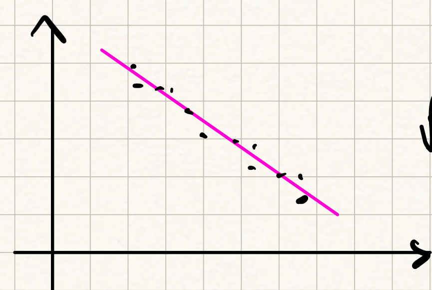


$\rho \in (0,1)$

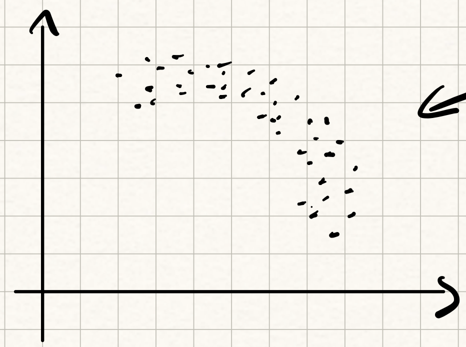


$\rho \in (-1,0)$

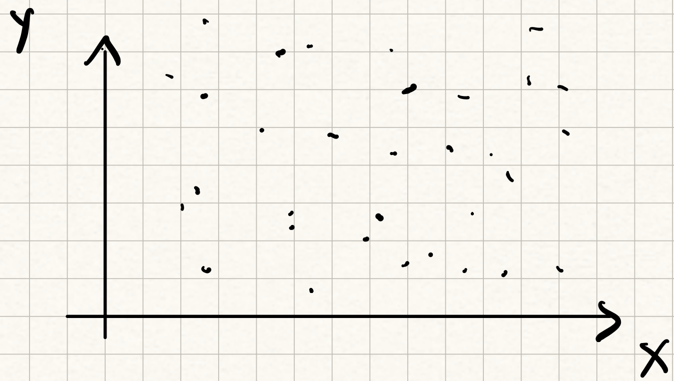
– лин. зависност
– обрнуто пропорционална (расте X, опада Y)



$\rho = -1$



постоји нелинеарна корелација али ρ (које говори само о линеарној зависности, не може ништа да нам каже о њој)



$\rho = 0$

одсуство било какве корелационе везе између појава

- $\rho > 0$ X и Y позитивно корелисане
- $\rho < 0$ X и Y негативно корелисане
- $\rho = 0$ X и Y независне сл. променљиве!

X, Y независне $\Rightarrow \operatorname{cov}(X,Y) = 0 \Rightarrow \rho(X,Y) = 0$

Али важи и: $\rho(X,Y) = 0 \Rightarrow$ X, Y независне

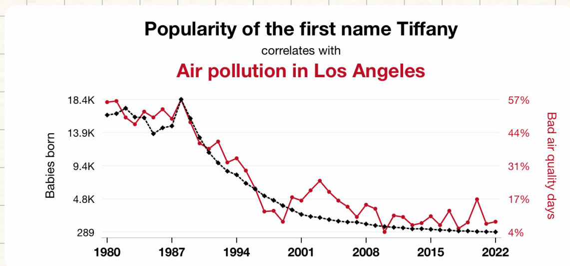
ОСОБИНЕ:

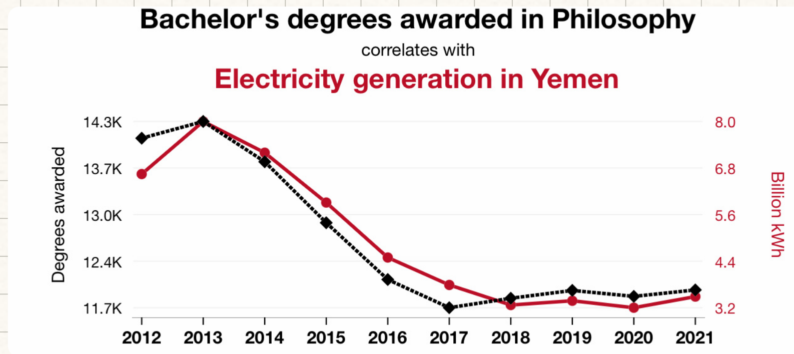
1° $-1 \le \rho \le 1$

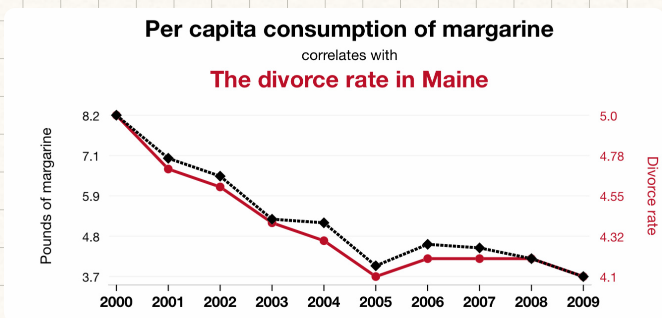
2° Указује само на ЛИНЕАРНУ ПОВЕЗАНОСТ између појава (ако је $\rho = 0$ то не значи да између појава не постоји нелинеарна корелација)

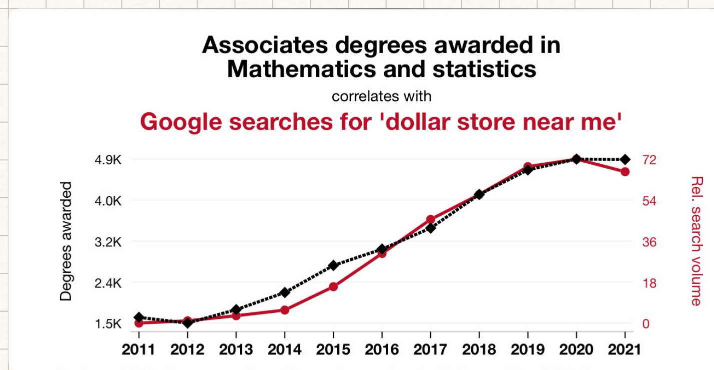
3° Корелацијом се мери повезаност између појава. Али не смемо једну прогласити за узрок, а другу за последицу БЕЗ ДОДАТНЕ АНАЛИЗЕ!

CORRELATION IS NOT CAUSATION!!!









... још примера чудних корелација
можете видети на https://www.tylervigen.com

ПРОБЛЕМ СЕ РЕШАВА УВОЂЕЊЕМ

КОЕФИЦИЈЕНТА ДЕТЕРМИНАЦИЈЕ: $\rho^2_{X,Y}$

који говори о томе колико је повезаност између података узрокована и другим факторима!

χ^2 (ХИ-КВАДРАТ) РАСПОДЕЛА

Нека су $X_1, X_2, X_3, \dots, X_n$ $\sim \mathcal{N}(0,1)$ незав. сл. променљиве са стандардном нормалном расподелом.

Збир њихових квадрата дефинише сл. пром.

$$H = X_1^2 + X_2^2 + \dots + X_n^2$$ која има χ^2 РАСПОДЕЛУ.

$$H \sim \chi^2(n)$$

n означава број сл. променљивих које се могу слободно мењати. Због тога кажемо да ова расподела има n- степени слободе.

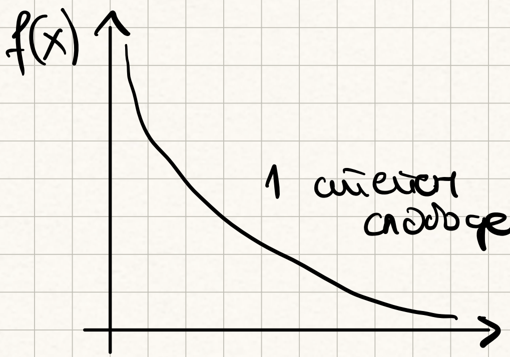
Ф-ја густине сл. променљиве са χ^2 расподелом

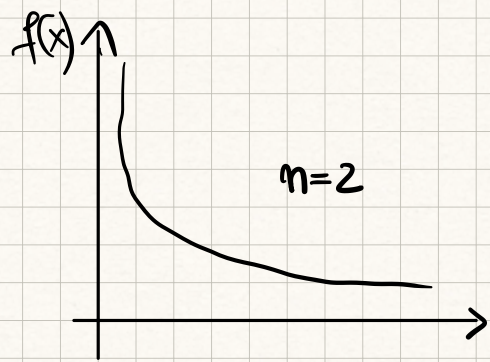
$$f(x) = \frac{1}{2^{\frac{n}{2}}\,\Gamma\left(\frac{n}{2}\right)} \cdot x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}}, \quad x>0$$

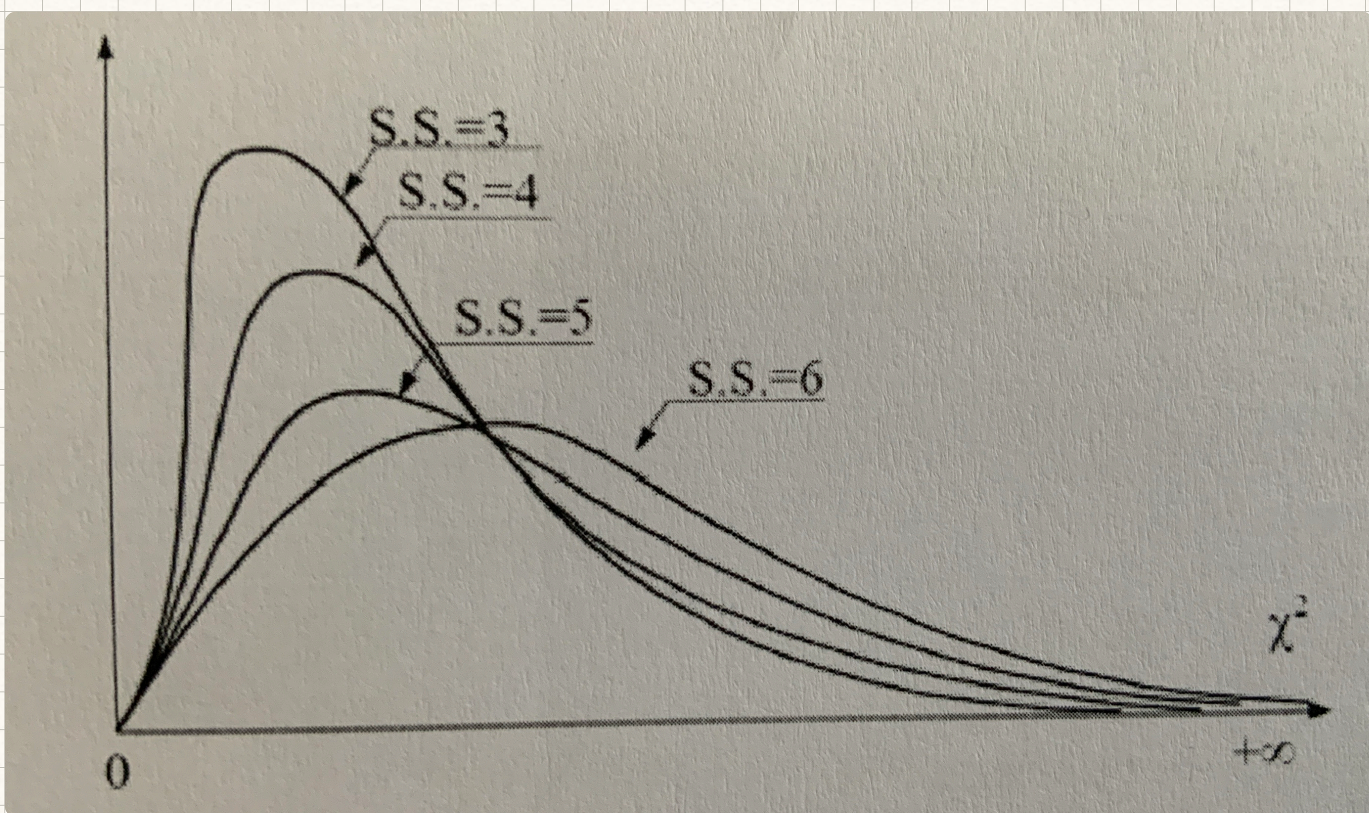
гама ф-ја

n – како се одређују степени слободе?

Степен слободе је број лин. независних сл. променљивих $X_1, X_2, \dots, X_n$ у изразу H.







ОСОБИНЕ ф-је густине

1° Дефинисана само за позитивне вредности

2° За **НЕГАТИВНЕ** вредности, узима се да је $=0$

3° Када је степен слободе **МАЛИ**, ф-ја густине је **АСИМЕТРИЧНА**, а што више степен расте, график постаје све више **СИМЕТРИЧАН**.

4° Када је n **ДОВОЉНО ВЕЛИКИ** број, расподела се асимптотски приближава **НОРМАЛНОЈ РАСПОДЕЛИ**.

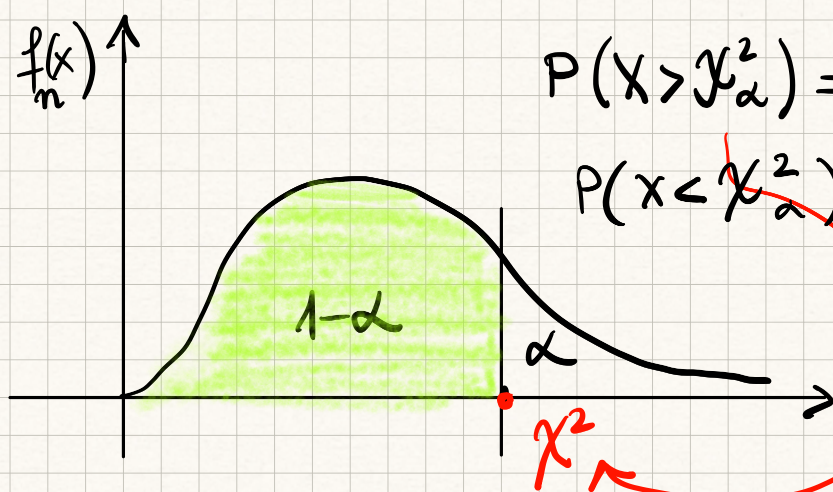
↓

За $n \geq 30$ χ-КВАДРАТ РАСПОДЕЛУ апроксимирамо нормалном $\mathcal{N}(n, 2n)$

$$E(H) = n$$

$$D(H) = 2n$$

ТАБЛИЦЕ:



$$P(X > \chi^2_\alpha) = \alpha$$

$$P(X < \chi^2_\alpha) = 1-\alpha$$

Таблице за ХИ-КВАДРАТ расподелу дају вредности χ^2_α за које важи да је вероватноћа

$$P(X > \chi^2_\alpha) = \alpha.$$

где је α задато и једнако 0,01, 0,02, ..., 0.99, а сл. променљива X има ХИ-КВАДРАТ расподелу са n- степени слободе:

n	$\alpha=0{,}99$	...	$\alpha=0{,}01$
1	0,00		
2	0,02		
⋮			
17	6.408		33,409
⋮			
30			

Нека је $X \sim \chi^2(17)$ и $\alpha=0{,}01$.

ПРАГ ЗНАЧАЈНОСТИ

Из таблице добијамо

$$P(X > 33{,}409) = 0{,}01$$

Кад n пређе 30 онда идемо на $\mathcal{N}(n, 2n)$

χ^2 расподелу у статистичким оцењивањима:

- одређујемо ИНТЕРВАЛ ПОВЕРЕЊА непознате ДИСПЕРЗИЈЕ
- за тестирање сагласности ДИСПЕРЗИЈА
- НЕПАРАМЕТАРСКЕ тестове

СТУДЕНТОВА РАСПОДЕЛА

ДЕФ: Нека су дате сл. пр. $Y \sim \mathcal{N}(0,1)$ и сл. променљива $Z \sim \chi^2(n)$, и нека су оне независне. Тада сл. променљива

$$X = \frac{Y}{\sqrt{\frac{Z}{n}}}$$

има СТУДЕНТОВУ РАСПОДЕЛУ са n-СТЕПЕНИ СЛОБОДЕ и означавамо $X \sim t(n)$.

t-расподела

Фја густине сл. променљиве са студент. расподелом зависи само од n:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n \cdot \pi} \cdot \Gamma\left(\frac{n}{2}\right)} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad , x \in \mathbb{R}$$

Може се показати

$$E(x) = 0$$

$$D(x) = \begin{cases} \infty & , n \leq 2 \\ \frac{n}{n-2} & , n > 2 \end{cases}$$

ТВРЂЕЊЕ: Ако су $Y, X_1, X_2, \ldots, X_n$ независне и све имају станд. норм. расподелу $\mathcal{N}(0,1)$ тада сл. пр.

$$X = \frac{Y}{\sqrt{\frac{X_1^2 + X_2^2 + \ldots + X_n^2}{n}}}$$

има студентову $t(n)$ расподелу.