6. Lineární regresní modely

STATISTICKÁ ZÁVISLOST

- **Korelace** popisuje vliv změny úrovně jednoho znaku na změnu úrovně jiných znaků a platí pro **kvantitativní (měřené) znaky;**

- **Kontingence** popisuje závislost **kvalitativních** (slovních, popisných) znaků, které mají více než dvě alternativy, tzv. **množných znaků** (např. druh dřeviny, národnost, apod.);

- **Asociace** popisuje závislost **kvalitativních** (slovních, popisných) znaků, které mají pouze dvě alternativy, tzv. **alternativních znaků** (např. pohlaví, odpovědi typu ano/ne, …).

KORELACE

typy podle počtu korelovaných znaků

- **Jednoduchá** popisuje vztah dvou znaků,

- **Mnohonásobná** popisuje vztahy více než dvou znaků,

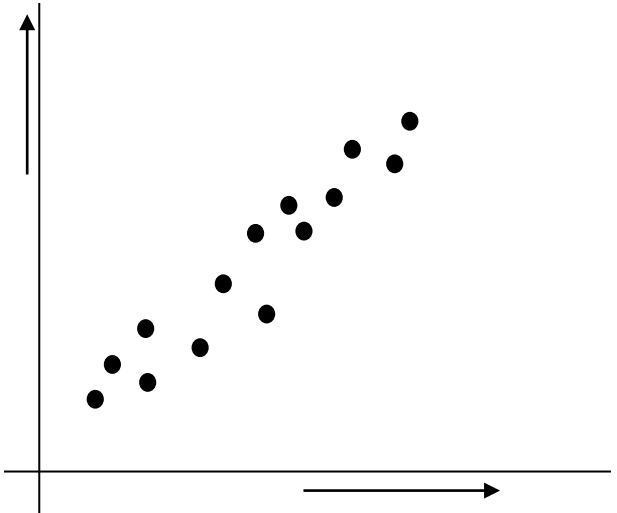
- **Parciální** popisuje závislost dvou znaků ve vícerozměrném statistickém souboru při vyloučení vlivu ostatních znaků na tuto závislost·

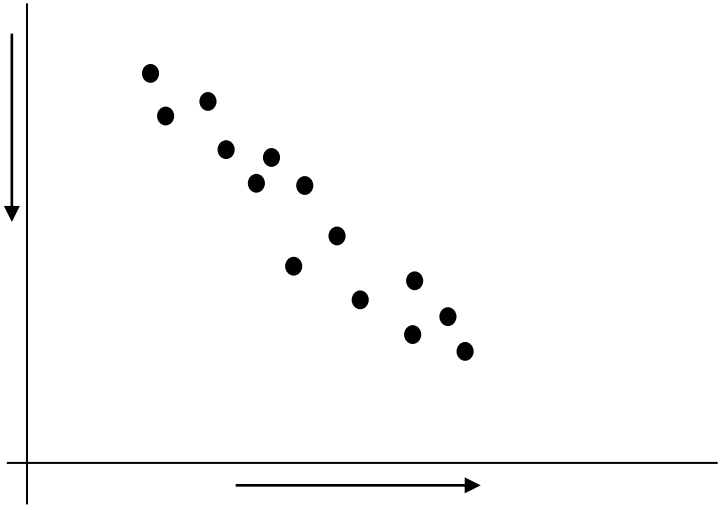
KORELACE

typy podle smyslu změny hodnot

Kladná značí, že se zvyšováním hodnot jednoho znaku se zvyšují i hodnoty druhého znaku,

Záporná značí, že se zvyšováním hodnot jednoho znaku se zmenšují hodnoty druhého znaku,





KORELACE

typy podle tvaru závislosti

- **Přímková (lineární)** značí, že grafickým obrazem závislosti je přímka (lineární trend),

- **Křivková (nelineární)** značí, že grafickým obrazem závislosti je křivka (nelineární trend).

KORELAČNÍ POČET

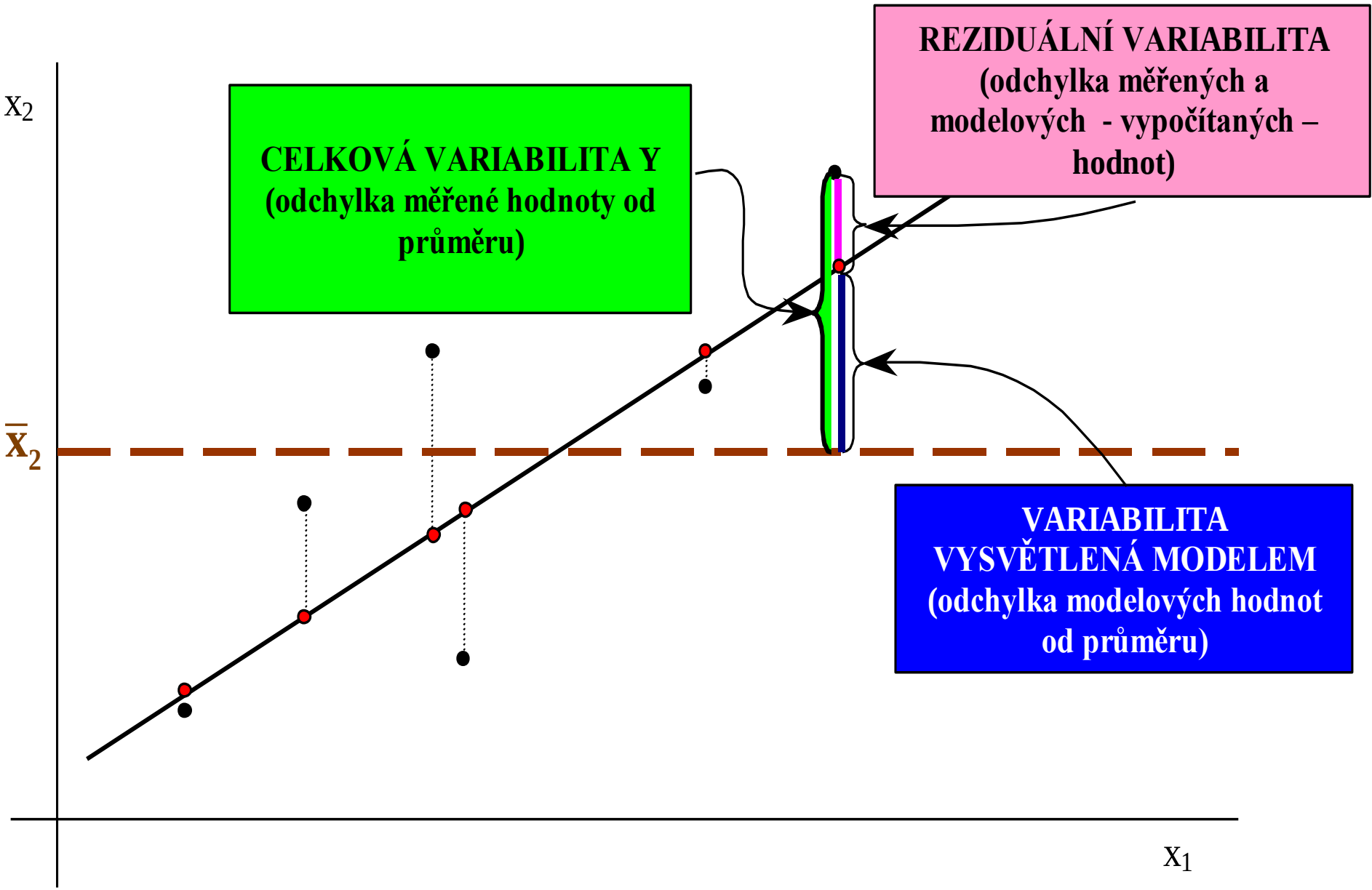
Korelační analýza

- zjišťuje ***existenci závislosti*** a její druhy,
- měří ***těsnost závislosti***,
- ověřuje ***hypotézy o statistické významnosti*** *závislosti*;

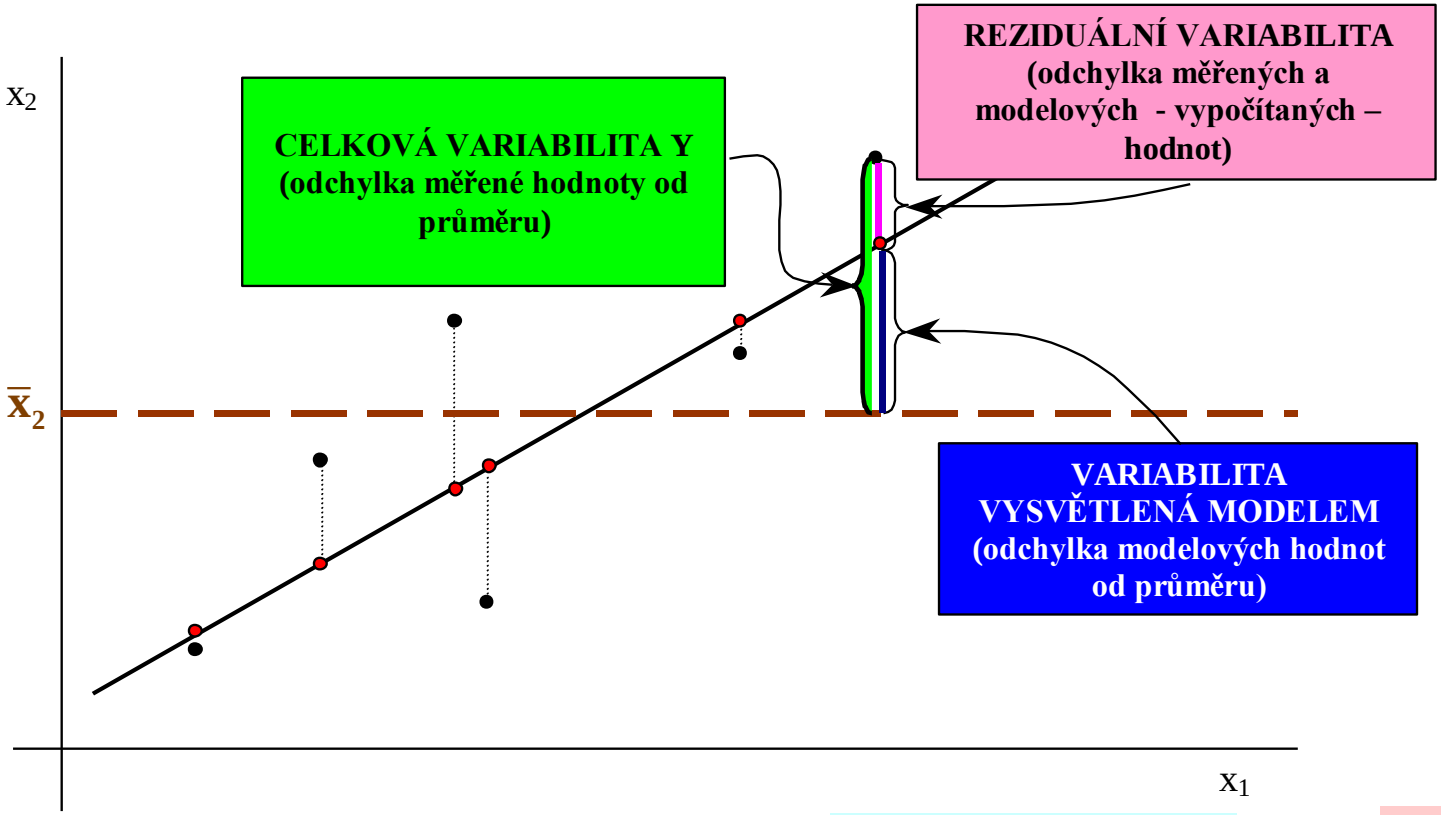
Regresní analýza

- zabývá se ***vytvořením vhodného matematického modelu*** závislosti,
- stanoví ***parametry*** tohoto *modelu*,
- ověřuje ***hypotézy o vhodnosti a důležitých vlastnostech modelu***.

MÍRA KORELAČNÍ ZÁVISLOSTI



MÍRA LINEÁRNÍ KORELAČNÍ ZÁVISLOSTI



$$\frac{\sum_{i=1}^{n}\left(x_{2i}-\bar{x}_2\right)^2}{n} = \frac{\sum_{i=1}^{n}\left(x'_{2i}-\bar{x}_2\right)^2}{n} + \frac{\sum_{i=1}^{n}\left(x_{2i}-x'_{2i}\right)^2}{n}$$

MÍRA LINEÁRNÍ KORELAČNÍ ZÁVISLOSTI

KOEFICIENT DETERMINACE

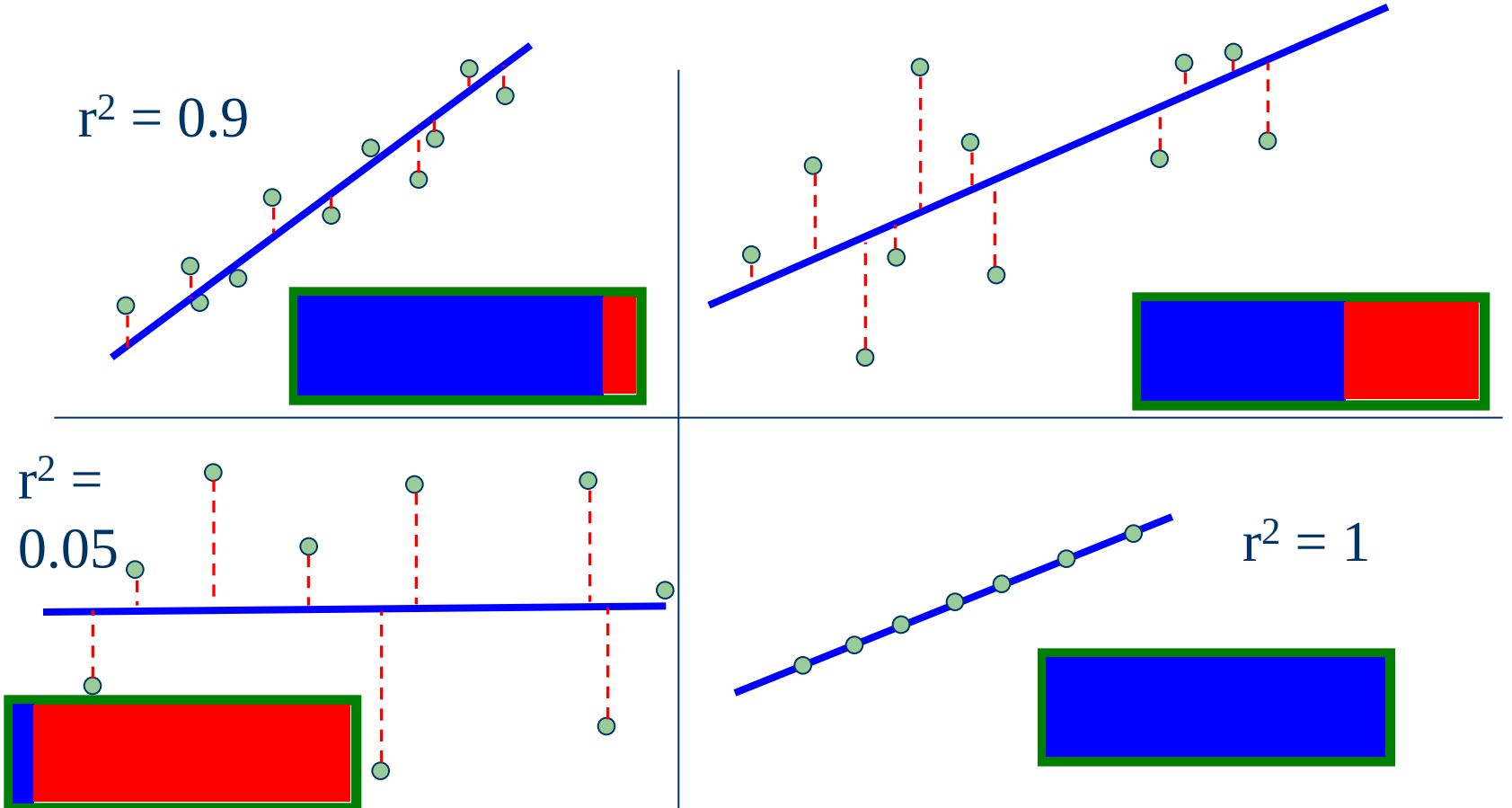
$$R^2 = \frac{S^2_{x'_2}}{S^2_{x_2}} = 1 - \frac{S^2_{x_1 x_2}}{S^2_{x_2}}$$

KOEFICIENT KORELACE

$$R = \sqrt{\frac{S^2_{x'_2}}{S^2_{x_2}}} = \sqrt{1 - \frac{S^2_{x_1 x_2}}{S^2_{x_2}}}$$

KOEFICIENT DETERMINACE

vyjadřuje, jakou část celkové variability závisle proměnné (vysvětlované proměnné) objasňuje regresní model.



KORELAČNÍ KOEFICIENT

Pro jednoduchou korelaci:

Párový představuje zvláštní případ vícenásobného korelačního koeficientu, kdy vyjadřuje míru lineární stochastické závislosti mezi náhodnými veličinami x_i a x_j**,**

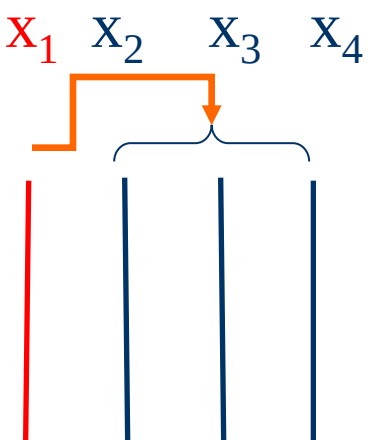
- **Pearsonův**
- **Spearmanův (korelace pořadí)**

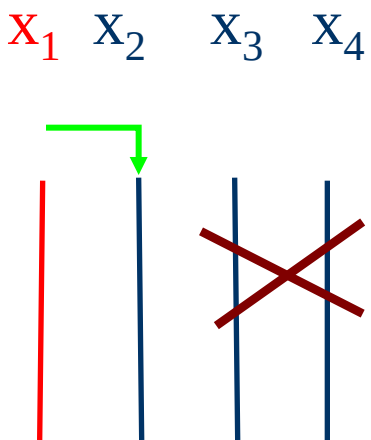
KORELAČNÍ KOEFICIENT

Pro vícenásobnou korelaci:

Vícenásobný definuje míru lineární stochastické závislosti mezi náhodnou veličinou x_1 a nejlepší lineární kombinací složek x_2, x_3, ..., x_m náhodného vektoru $\boldsymbol{x}$

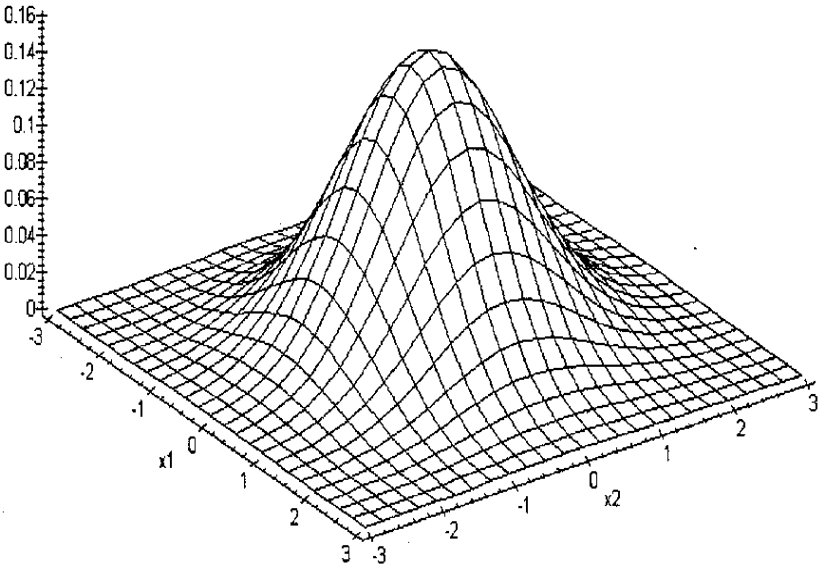
Parciální definuje míru lineární stochastické závislosti mezi náhodnými veličinami x_i a x_j při skonstantnění ostatních složek vektoru $\boldsymbol{x}$





PEARSONŮV KORELAČNÍ KOEFICIENT *r*

Podmínkou je dodržení dvourozměného normálního rozdělení



normovaná kovariance

$$r_{x_1x_2} = r_{x_2x_1} = \frac{cov_{x_1x_2}}{S_{x_1} \cdot S_{x_2}}$$

PEARSONŮV KORELAČNÍ KOEFICIENT *r*

KOVARIANCE:

- **míra intenzity vztahu** mezi složkami vícerozměrného souboru
- je mírou intenzity **lineární** závislosti
- je vždy **nezáporná**
- její **limitou je součin směrodatných odchylek**
- je **symetrickou funkcí** svých argumentů
- její **velikost je závislá na měřítku argumentů ⇒ nutnost normování**

$$\mathrm{cov}_{x_1x_2} = \frac{1}{n}\sum_{i=1}^{n}\left(x_{1i} - \bar{x}_1\right)\cdot\left(x_{2i} - \bar{x}_2\right)$$

Nevýhodou kovariance:

a) její hodnoty závisí na měřítku, ve kterém jsou vyjádřeny ξ_1 a ξ_2.

b) její velikost lze hodnotit vzhledem k součinu $\sigma_i \sigma_j$.

Standardizace podělením tímto součinem a vzniklá veličina $\varrho_{ij} = \varrho(\xi_i, \xi_j)$ se nazývá párový ***korelační koeficient***

$$\rho(\xi_i, \xi_j) = \rho_{ij} = \frac{\mathrm{cov}(\xi_i, \xi_j)}{\sigma_i \sigma_j}$$

Korelační koeficient leží v rozmezí $-1 \leq \varrho_{ij} \leq 1$.

a) $\varrho_{ij} > 0$, jde o ***pozitivně korelované*** náhodné veličiny,

b) $\varrho_{ij} < 0$, jde o ***negativně korelované*** náhodné veličiny.

PEARSONŮV KORELAČNÍ KOEFICIENT *r*

Základní vlastnosti Pearsonova korelačního koeficientu:

- je to **bezrozměrná** míra lineární korelace;
- nabývá hodnoty **0 – 1 pro kladnou korelaci, 0 – (-1) pro zápornou korelaci**;
- hodnota **0** znamená, že mezi posuzovanými veličinami **není žádný lineární vztah** (může být nelineární) nebo tento vztah zůstal na základě dat, které máme k dispozici, neprokázán;
- hodnota **1** nebo **(-1)** indikuje **funkční závislost**;
- hodnota korelačního koeficientu je stejná pro závislost x_1 na x_2 i pro opačnou závislost x_2 na x_1.

Základní vlastnosti:

a) Rovnost $|\varrho_{ij}| = 1$ ukazuje, že mezi ξ_i a ξ_j existuje přesně lineární vztah.

b) Pokud jsou náhodné veličiny ξ_i a ξ_j vzájemně nekorelované, je $\varrho_{ij} = 0$.

c) V případě, že ξ_i a ξ_j pocházejí z vícerozměrného normálního rozdělení a $\varrho_{ij} = 0$, znamená to, že jsou ***vzájemně nezávislé***.

d) Pro nelineárně závislé náhodné veličiny může být $\varrho_{ij} = 0$

e) Korelační koeficient ϱ_{ii} náhodné veličiny ξ_i samotné se sebou je roven jedné.

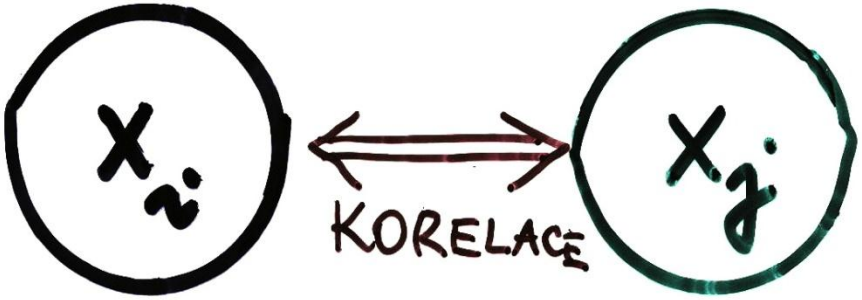
f) Korelační koeficient je invariantní vůči lineární transformaci náhodných proměnných ξ_i, ξ_j. Pro čísla a_1, a_2, b_1, b_2 platí vztah

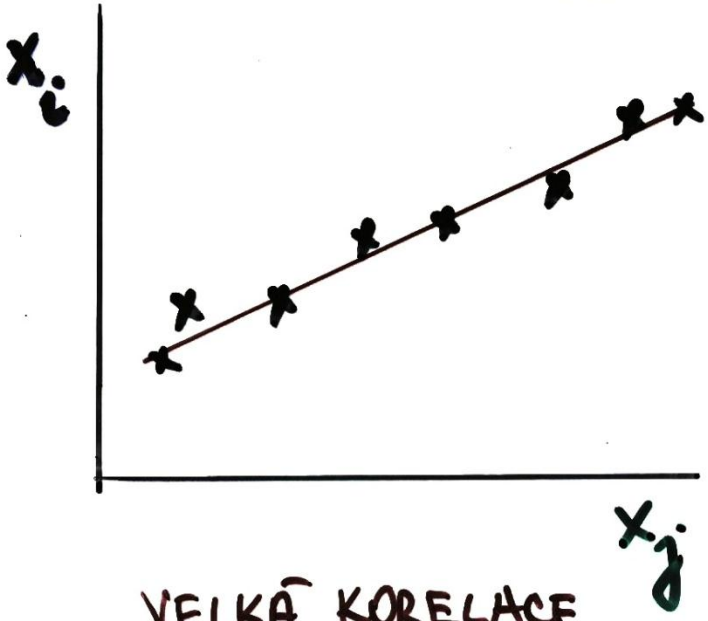
$$\rho(a_1 \xi_i + b_1, a_2 \xi_j + b_2) = \text{sign}(a_1 a_2)\, \rho(\xi_i, \xi_j)$$

kde sign(x) je znaménková funkce, pro kterou platí

$$\text{sign}(x) = \begin{cases} -1 & \text{pro } x < 0 \\ 0 & \text{pro } x = 0 \\ 1 & \text{pro } x > 1 \end{cases}$$

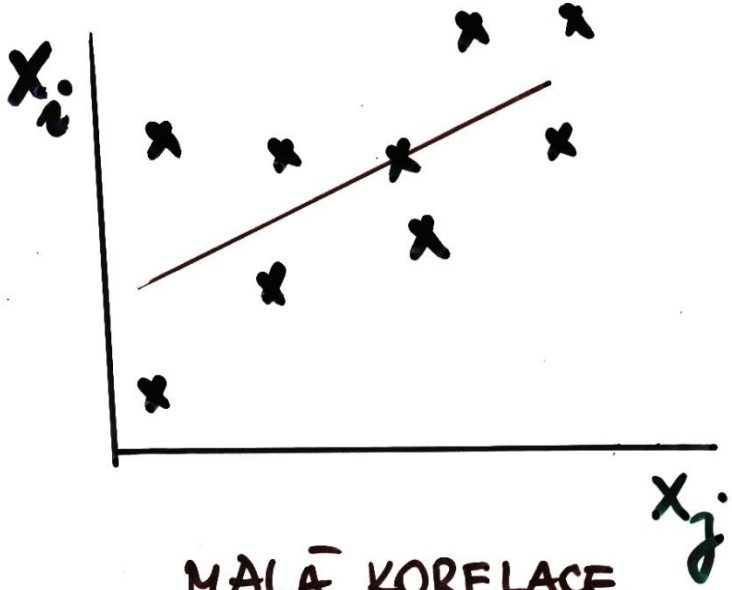
KORELAČNÍ CHARAKTERISTIKY = míry lineární závislosti mezi dvěma či více náhodnými veličinami





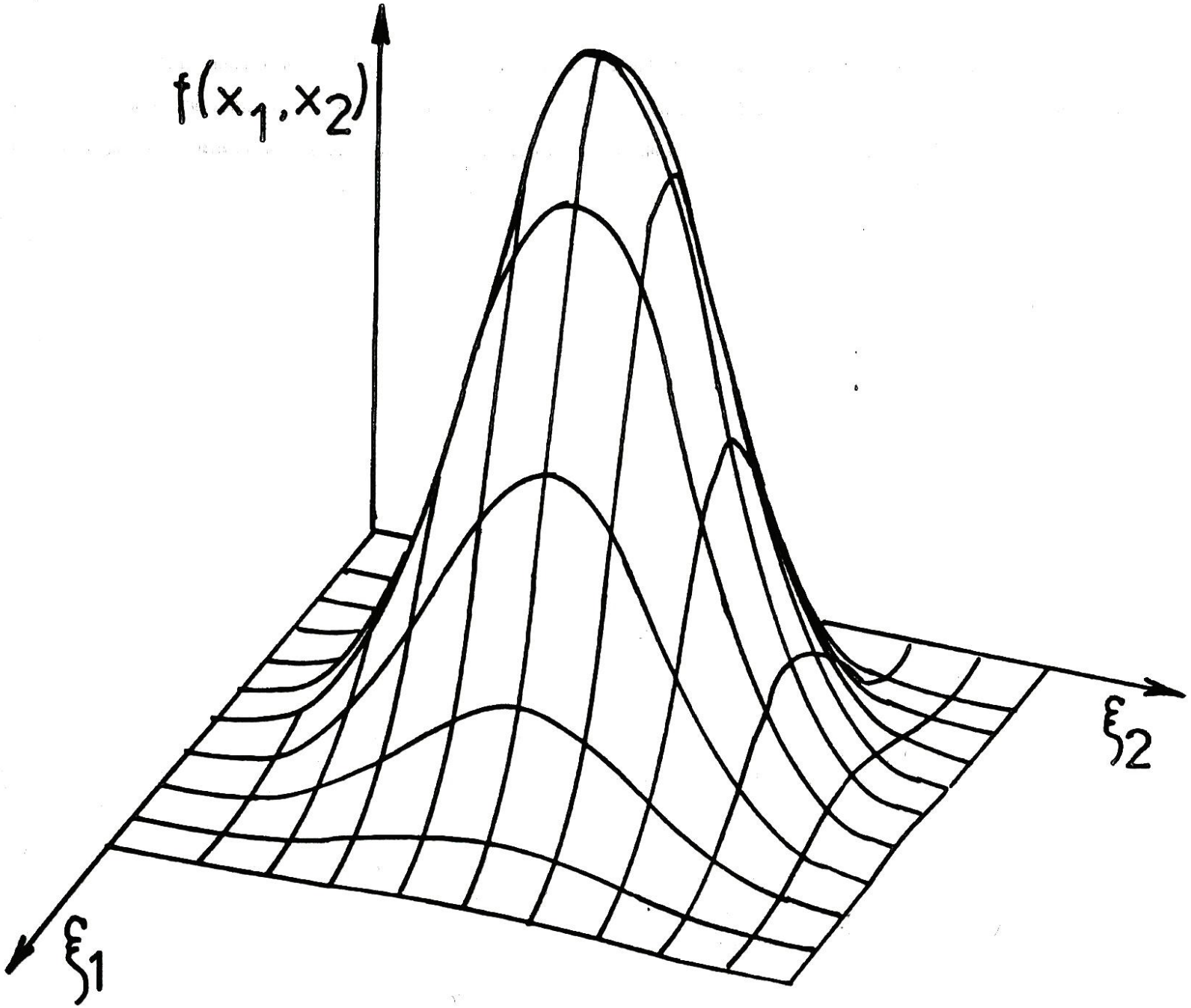
VELKÁ KORELACE

$r_{ij} \approx 1$

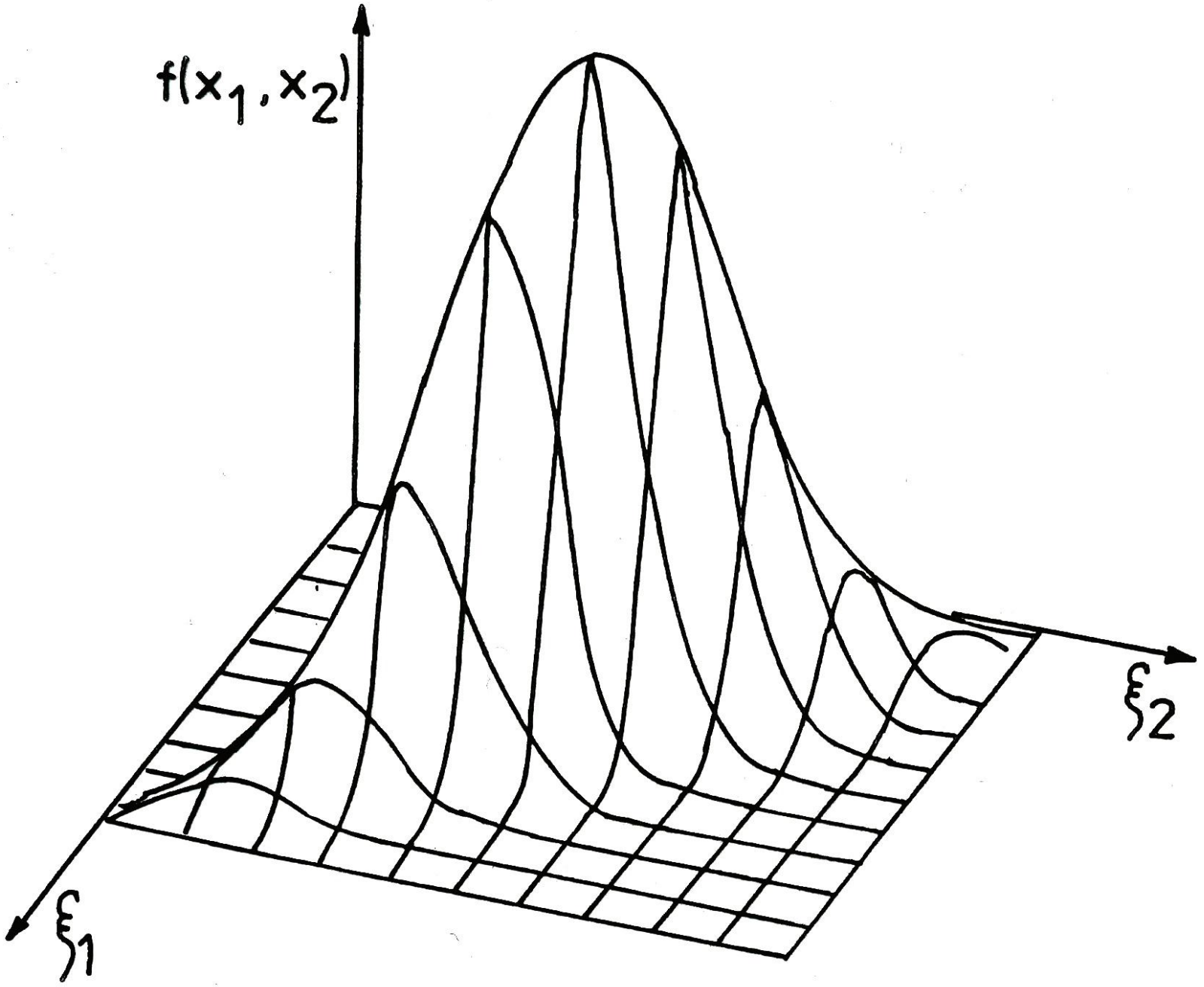


MALÁ KORELACE

$r_{ij} \approx 0.2$

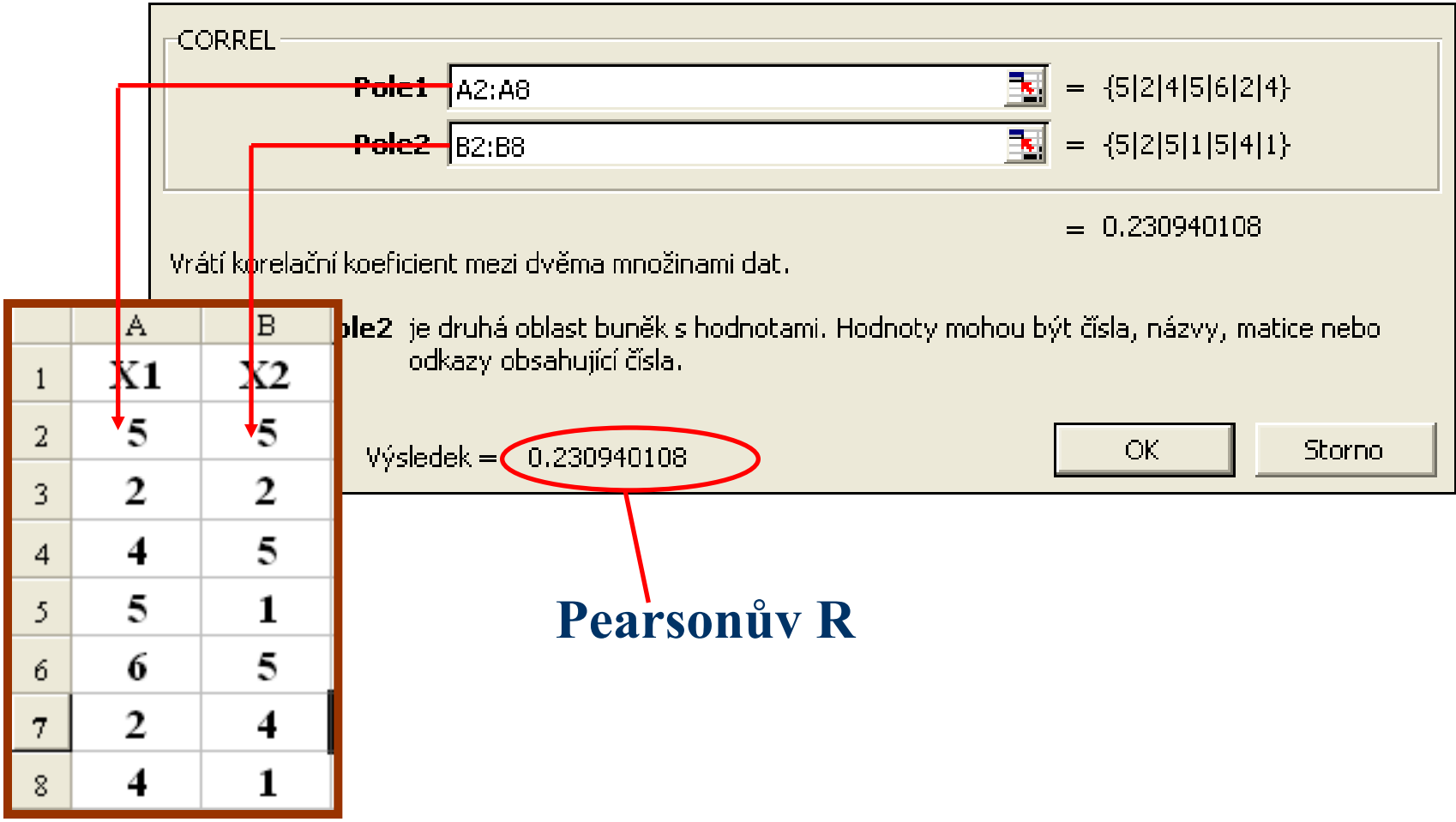


Obr. 4.3 Povrch simultánní hustoty pravděpodobnosti pro $\varrho_{12} = 0$



Obr. 4.4 Povrch simultánní hustoty pravděpodobnosti pro $\varrho_{12} = 0.9$

PEARSONŮV KORELAČNÍ KOEFICIENT *r*
výpočet v Excelu



SPEARMANŮV KORELAČNÍ KOEFICIENT

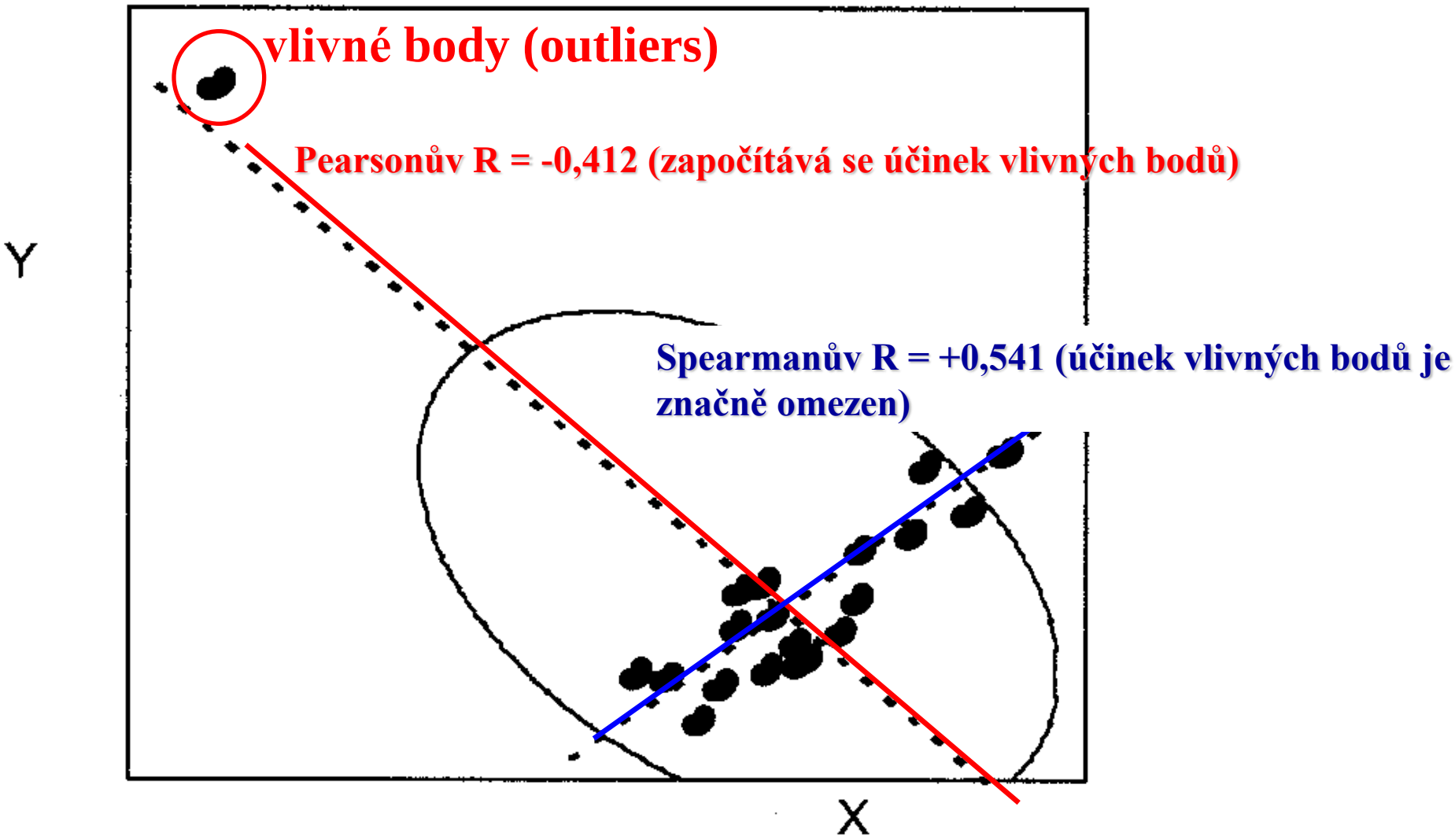
Neparametrický korelační koeficient, vycházející nikoli z hodnot, ale z jejich **pořadí.**

Používá se tehdy, nejsou-li závažným způsobem splněny předpoklady pro použití Pearsonova korelačního koeficientu.

$$r_S = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n^3 - n}$$

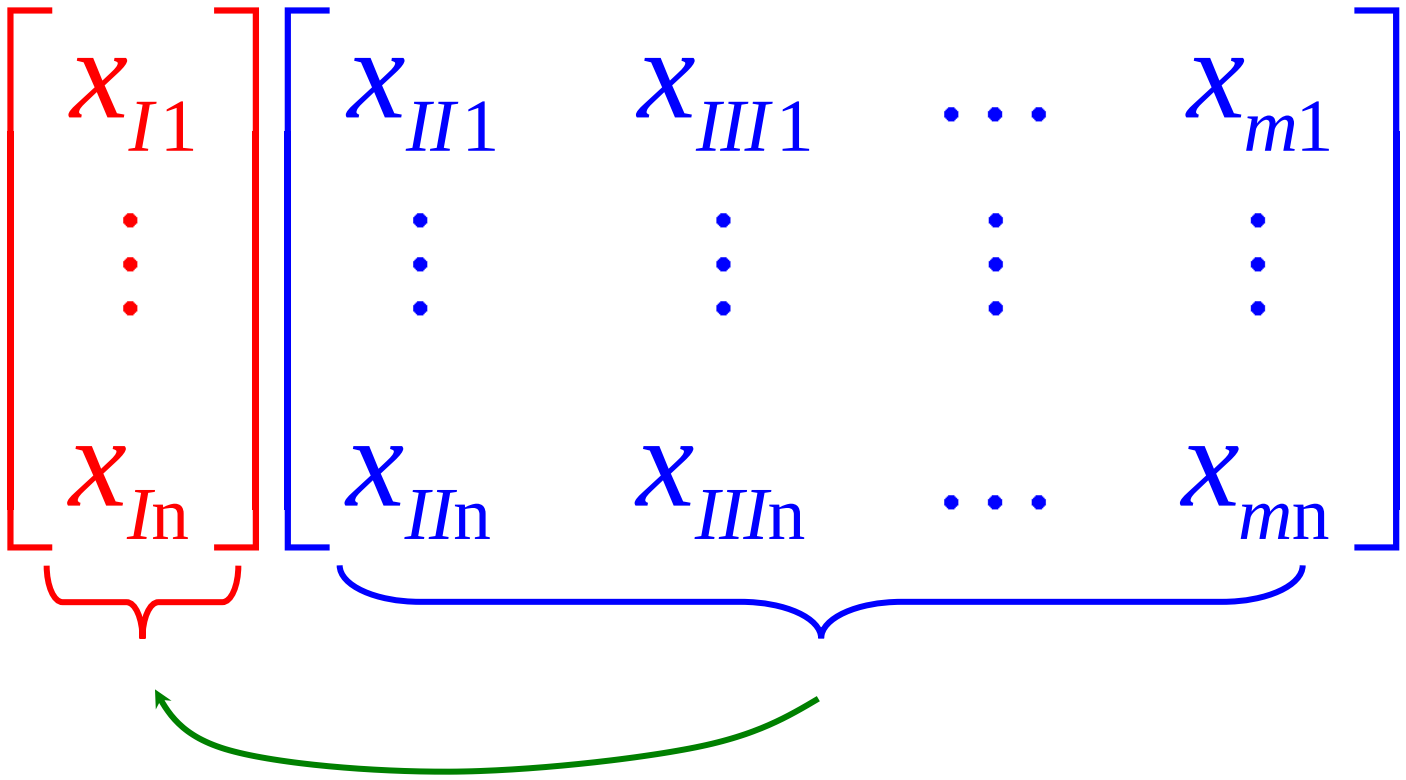
Diference mezi pořadími hodnot *x* a *y* v jednom řádku

SPEARMANŮV KORELAČNÍ KOEFICIENT



MNOHONÁSOBNÝ KORELAČNÍ KOEFICIENT

vyjadřuje sílu závislosti **jedné proměnné** na **dvou a více jiných proměnných**



MNOHONÁSOBNÝ KORELAČNÍ KOEFICIENT

Základní vlastnosti:

a) $0 \leq R \leq 1$

b) Pokud je $R = 1$, znamená to, že závisle proměnná x_1 je přesně lineární kombinací veličin $x_2, \ldots, x_m$.

c) Pokud je $R = 0$, potom jsou také všechny párové korelační koeficienty nulové.

d) S růstem počtu vysvětlujících (nezávislých) proměnných hodnota vícenásobného korelačního koeficientu neklesá, tj. platí

$$R_{1(2)} \leq R_{1(2,3)} \leq \ldots \leq R_{1(2,\ \ldots,\ m)}.$$

MNOHONÁSOBNÝ KORELAČNÍ KOEFICIENT
numerický výpočet

$$R_{1(2,3,\ldots,m)} = \sqrt{1 - \frac{\det(\mathbf{R})}{\det(\mathbf{R}_{(11)})}}$$

= **determinant korelační matice**

= **determinant korelační matice s vypuštěným sloupcem a řádkem odpovídajícím té proměnné, jejíž závislost na zbytku matice se vypočítává**

korelační koeficient 1. a 2. proměnné

$$\mathbf{R} = \begin{bmatrix} 1 & R_{12} & \cdots & R_{1i} & \cdots & R_{1m} \\ R_{21} & 1 & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & 1 & \cdots & \cdots & \cdots \\ R_{i1} & \cdots & \cdots & 1 & \cdots & R_{im} \\ \cdots & \cdots & \cdots & \cdots & 1 & \cdots \\ R_{m1} & R_{m2} & \cdots & R_{mi} & \cdots & 1 \end{bmatrix}$$

Korelační matice R

MNOHONÁSOBNÝ KORELAČNÍ KOEFICIENT

$$R_{1(2,3,\ldots,m)} = \frac{\begin{bmatrix} 1 & R_{12} & \cdots & R_{1i} & \cdots & R_{1m} \\ R_{21} & 1 & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & 1 & \cdots & \cdots & \cdots \\ R_{i1} & \cdots & \cdots & 1 & \cdots & R_{im} \\ \cdots & \cdots & \cdots & \cdots & 1 & \cdots \\ R_{m1} & R_{m2} & \cdots & R_{mi} & \cdots & 1 \end{bmatrix}}{\begin{bmatrix} 1 & R_{12} & \cdots & R_{1i} & \cdots & R_{1m} \\ R_{21} & 1 & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & 1 & \cdots & \cdots & \cdots \\ R_{i1} & \cdots & \cdots & 1 & \cdots & R_{im} \\ \cdots & \cdots & \cdots & \cdots & 1 & \cdots \\ R_{m1} & R_{m2} & \cdots & R_{mi} & \cdots & 1 \end{bmatrix}}$$

$\Rightarrow \det(\mathbf{R})$

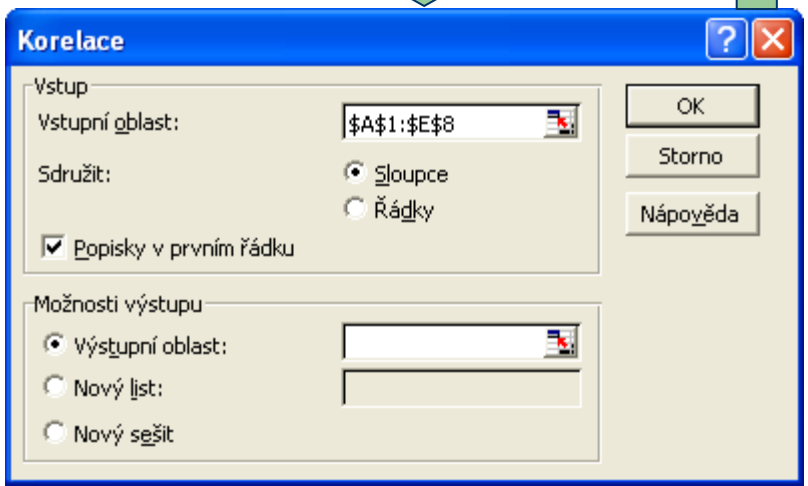
$\Rightarrow \det(\mathbf{R}_{(11)})$

$$\Rightarrow \sqrt{1 - \frac{\det(\mathbf{R})}{\det(\mathbf{R}_{(11)})}}$$

MNOHONÁSOBNÝ KORELAČNÍ KOEFICIENT
numerický výpočet v Excelu

❶

	A	B	C	D	E
1	X1	X2	X3	X4	X5
2	5	5	8	7	7
3	2	2	9	8	8
4	4	5	5	9	9
5	5	1	2	5	5
6	6	5	3	4	4
7	2	4	1	1	2
8	4	1	4	2	1



	X1	X2	X3	X4	X5
X1	1				
X2	0.23	1			
X3	-0.15	0.08	1		
X4	0.07	0.25	0.73	1	
X5	0	0.34	0.67	0.98	1

	X1	X2	X3	X4	X5
X1	1	0.23	-0.15	0.07	0
X2	0.23	1	0.08	0.25	0.34
X3	-0.15	0.08	1	0.73	0.67
X4	0.07	0.25	0.73	1	0.98
X5	0	0.34	0.67	0.98	1

MNOHONÁSOBNÝ KORELAČNÍ KOEFICIENT
numerický výpočet v Excelu

	X1	*X2*	*X3*	*X4*	*X5*
X1	1	0.23	-0.15	0.07	0
X2	0.23	1	0.08	0.25	0.34
X3	-0.15	0.08	1	0.73	0.67
X4	0.07	0.25	0.73	1	0.98
X5	0	0.34	0.67	0.98	1

$$\sqrt{1-\frac{\det(\mathbf{R})}{\det(\mathbf{R}_{(11)})}} = \sqrt{1-\frac{\mathbf{= DETERMINANT(R)}}{\mathbf{= DETERMINANT(R_{(11)})}}} =$$

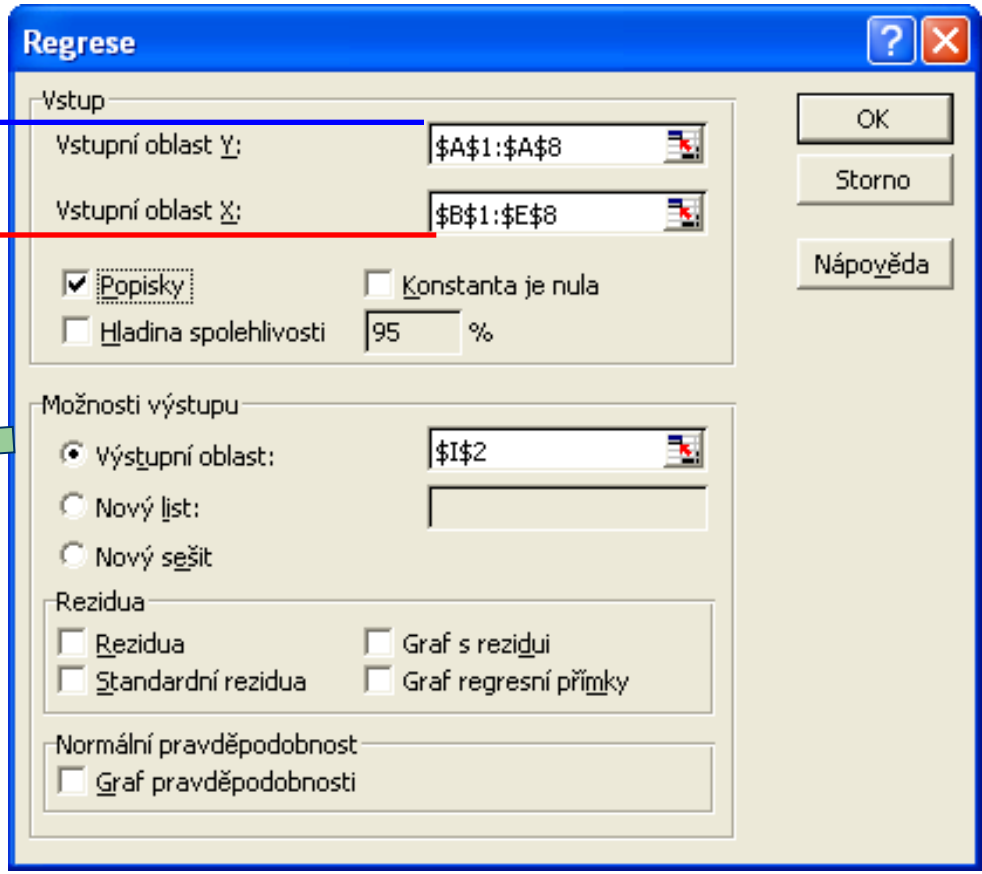
$$= \sqrt{1-\frac{0.004755585}{0.010714947}} = \mathbf{0.74577}$$

MNOHONÁSOBNÝ KORELAČNÍ KOEFICIENT
numerický výpočet v Excelu

2

	A	B	C	D	E
1	X1	X2	X3	X4	X5
2	5	5	8	7	7
3	2	2	9	8	8
4	4	5	5	9	9
5	5	1	2	5	5
6	6	5	3	4	4
7	2	4	1	1	2
8	4	1	4	2	1

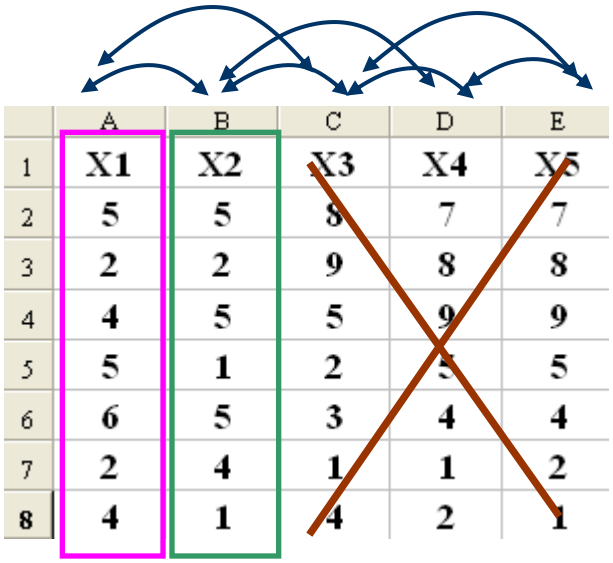
Nástroje ⇒Analýza dat ⇒Regrese



Regresní statistika	
Násobné R	0.74577
Hodnota spolehlivosti R	0.556173
Nastavená hodnota spolehlivosti R	-0.33148
Chyba stř. hodnoty	1.762609
Pozorování	7

PARCIÁLNÍ KORELAČNÍ KOEFICIENT

Používá se k posouzení síly závislosti **dvou veličin** ve vícerozměrném souboru **při vyloučení vlivu ostatních veličin.**



	A	B	C	D	E
1	X1	X2	X3	X4	X5
2	5	5	8	7	7
3	2	2	9	8	8
4	4	5	5	9	9
5	5	1	2	5	5
6	6	5	3	4	4
7	2	4	1	1	2
8	4	1	4	2	1

Podle počtu „vyloučených“ proměnných se stanovují řády parciálního *R* v příkladu vlevo to je parciální korelace III. řádu (3 „vyloučené“ proměnné)

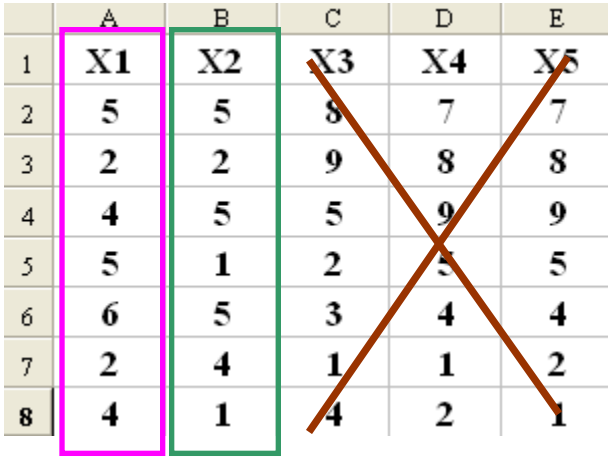
PARCIÁLNÍ KORELAČNÍ KOEFICIENT
výpočet

„Klasický“ výpočet je velmi zdlouhavý – vychází se z korelační matice, poté se počítají parciální korelace I. řádu (s jednou vyloučenou proměnnou), z nich II. řádu (dvě vyloučené proměnné), atd. až do potřebného řádu.

Při využití Excelu je možné využít vzorce

$$R_{ij(1,2,\ldots,m)} = \frac{(-1)^{j} \cdot \det(\mathbf{R}_{(ij)})}{\sqrt{\det(\mathbf{R}_{(ii)}) \cdot \det(\mathbf{R}_{(jj)})}}$$

PARCIÁLNÍ KORELAČNÍ KOEFICIENT
numerický výpočet v Excelu



	A	B	C	D	E
1	X1	X2	X3	X4	X5
2	5	5	8	7	7
3	2	2	9	8	8
4	4	5	5	9	9
5	5	1	2	5	5
6	6	5	3	4	4
7	2	4	1	1	2
8	4	1	4	2	1

$$R_{ij(1,2,\ldots,m)} = \frac{(-1)^{j} \cdot \det(\mathbf{R}_{(ij)})}{\sqrt{\det(\mathbf{R}_{(ii)}) \cdot \det(\mathbf{R}_{(jj)})}}$$

$$R_{ij(1,2,\ldots,m)} = \frac{(-1)^{2} \cdot \det(R_{(12)})}{\sqrt{\det(R_{(11)}) \cdot \det(R_{(22)})}}$$

	X1	*X2*	*X3*	*X4*	*X5*
X1	1	0.23	-0.15	0.07	0
X2	0.23	1	0.08	0.25	0.34
X3	-0.15	0.08	1	0.73	0.67
X4	0.07	0.25	0.73	1	0.98
X5	0	0.34	0.67	0.98	1

PARCIÁLNÍ KORELAČNÍ KOEFICIENT
numerický výpočet v Excelu

	X1	*X2*	*X3*	*X4*	*X5*
X1	1	0.23	-0.15	0.07	0
X2	0.23	1	0.08	0.25	0.34
X3	-0.15	0.08	1	0.73	0.67
X4	0.07	0.25	0.73	1	0.98
X5	0	0.34	0.67	0.98	1

$\det(R_{(11)}) = 0.010715$

	X1	*X2*	*X3*	*X4*	*X5*
X1	1	0.23	-0.15	0.07	0
X2	0.23	1	0.08	0.25	0.34
X3	-0.15	0.08	1	0.73	0.67
X4	0.07	0.25	0.73	1	0.98
X5	0	0.34	0.67	0.98	1

$\det(R_{(12)}) = 0.006086$

	X1	*X2*	*X3*	*X4*	*X5*
X1	1	0.23	-0.15	0.07	0
X2	0.23	1	0.08	0.25	0.34
X3	-0.15	0.08	1	0.73	0.67
X4	0.07	0.25	0.73	1	0.98
X5	0	0.34	0.67	0.98	1

$\det(R_{(22)}) = 0.010248$

PARCIÁLNÍ KORELAČNÍ KOEFICIENT
numerický výpočet v Excelu

$$R_{12(3,4,5)} = \frac{(-1)^2 \cdot \det(R_{(12)})}{\sqrt{\det(R_{(11)}) \cdot \det(R_{(22)})}} = \frac{1 \cdot 0.00608}{\sqrt{0.01071 \cdot 0.01025}} = 0.58082$$

Parciální korelační koeficient III. řádu pro závislost proměnných x_1 a x_2 (při vyloučení vlivu proměnných x_3, x_4 a x_5) je 0.58.

TESTY VÝZNAMNOSTI V KORELAČNÍ A REGRESNÍ ANALÝZE

- test významnosti korelačního koeficientu
- test významnosti modelu jako celku
- test významnosti jednotlivých regresních parametrů
- test shody lineárních regresních modelů

a mnoho dalších …..

TEST VÝZNAMNOSTI *R*

Test významnosti odpoví, zda je korelace mezi výběrovými proměnnými *R* natolik silná, abychom ji mohli považovat za dostatečně prokázanou i pro základní soubor ρ.

Pro párový *R*: $t_R = \frac{R \cdot \sqrt{n-2}}{\sqrt{1-R^2}}$ KH $t_{\alpha,n-2}$ **n – počet hodnot výběru**

Pro násobný *R*: $F_R = \frac{R^2(n-m)}{(1-R^2)(m-1)}$ $t_{\alpha,n-m}$ **m – počet proměnných**

Pro parciální *R*: $t_R = \frac{R \cdot \sqrt{n-k-2}}{\sqrt{1-R^2}}$ $t_{\alpha,n-k-2}$ **k – počet „vyloučených“ proměnných**

Příklad 7.1 *Odvození teoretické regrese pro standardizované normálně rozdělené náhodné veličiny*

Předpokládejme, že náhodné veličiny ξ_1 a ξ_2 mají normované normální rozdělení N(0, 1) s nulovými středními hodnotami a jednotkovými rozptyly. Sdružené rozdělení těchto veličin nechť je také normální, definované hustotou pravděpodobnosti

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp\left[- \frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1 - \rho^2)} \right]$$

kde $\rho = \rho(\xi_1, \xi_2)$ je korelační koeficient mezi náhodnými veličinami. Odvoďte teoretickou regresi $E(\xi_2/x_1)$.

Řešení: Nejprve je třeba vypočítat podmíněnou hustotu pravděpodobnosti $f(x_2/x_1)$. Platí, že

$$f(x_2/x_1) = \frac{f(x_1, x_2)}{f(x_1)}$$

Po dosazení a úpravách vyjde

$$f(x_2/x_1) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{(x_2-\rho x_1)^2}{2(1-\rho^2)}\right]$$

Po dosazení do definičního vztahu a analytické integraci dostaneme

$$E(\xi_2/x_1) = \int_{-\infty}^{\infty} \frac{x_2}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{(x_2-\rho x_1)^2}{2(1-\rho^2)}\right] dx_2 = \rho x_1$$

Závěr: Teoretická regrese je pro tyto náhodné veličiny lineární s nulovým úsekem a směrnicí odpovídající korelačnímu koeficientu ρ.

Postačuje nahradit střední hodnoty μ_1 a μ_2 aritmetickými průměry $\bar{x}_1$ a $\bar{x}_2$, dále rozptyly σ_1^2 a σ_2^2 výběrovými rozptyly s_1^2 a s_2^2 a konečně korelační koeficient ρ výběrovým korelačním koeficientem

$$R = \frac{\sum_{i=1}^{n} (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sqrt{\sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2 \sum_{i=1}^{n} (x_{2i} - \bar{x}_2)^2}}$$

Směrnice b_1 a úsek b_2 regrese $E(\xi_2/x_1)$ odpovídají odhadům určeným metodou nejmenších čtverců a $D(\xi_2/x_1)$ odpovídá reziduálnímu součtu čtverců odchylek.

Příklad 7.4 *Vícenásobný korelační koeficient pro dvě vysvětlující proměnné*
Určete vícenásobný korelační koeficient $R_{1(2,3)}$ mezi proměnnou ξ_1 a proměnnými ξ_2, ξ_3.

Řešení: Pro korelační matice **R** a $\mathbf{R}_{11}$ můžeme v tomto případě psát

$$\mathbf{R} = \begin{bmatrix} 1 & R_{12} & R_{13} \\ R_{12} & 1 & R_{23} \\ R_{13} & R_{23} & 1 \end{bmatrix} \qquad \mathbf{R}_{11} = \begin{bmatrix} 1 & R_{23} \\ R_{23} & 1 \end{bmatrix}$$

V těchto výrazech je již využito symetrie $R_{ij} = R_{ji}$ párových korelačních koeficientů. Po dosazení a úpravách snadno určíme, že

$$R_{1(2,3)} = \sqrt{\frac{R_{12}^2 + R_{13}^2 - 2\,R_{12}\,R_{13}\,R_{23}}{1 - R_{23}^2}}$$

Rovnice ukazuje, že párové korelační koeficienty nemohou nabývat libovolných hodnot v rozmezí $-1 \le R_{ij} < 1$, ale jsou vzájemně spjaty podmínkou, že $R_{1(2,3)} \le 1$.

Pokud je $R_{23} = 0$, tj. vysvětlující proměnné jsou vzájemně nekorelované, je

$$R_{1(2,3)}^2 = R_{12}^2 + R_{13}^2$$

Závěr: Vícenásobný korelační koeficient je možné určit přímo z definičního vztahu jako funkci párových korelačních koeficientů. Pro případ, že jsou vysvětlující proměnné $\xi_2, \ldots, \xi_m$ vzájemně nekorelované, je čtverec vícenásobného korelačního koeficientu součtem čtverců párových korelačních koeficientů.

Z jednotlivých parciálních korelačních koeficientů všech řádů je možné vyčíslit také ***vícenásobný korelační koeficient***

$$R^2_{1(2,\ldots,m)} = 1 - (1 - R^2_{1,2})(1 - R^2_{1,3(2)})(1 - R^2_{1,4(2,3)}) \ldots$$

$$\ldots (1 - R^2_{1,m(2,3,\ldots,m-1)})$$

Výpočet ***parciálních korelačních koeficientů*** se provede

$$R_{1i(2,3,\ldots,m)} = \frac{(-1)^i \det(\mathbf{R}_{1,i})}{\sqrt{\det(\mathbf{R}_{11}) \det(\mathbf{R}_{i,i})}}$$

kde $\mathbf{R}$ je ***korelační matice*** odpovídající vektoru ξ

a $\mathbf{R}_{i,j}$ je ***matice vzniklá vynecháním i-tého řádku a j-tého sloupce*** matice $\mathbf{R}$.

Parciální korelační koeficienty prvního řádu $R_{1,3(2)}$ odpovídají ***párovému korelačnímu koeficientu mezi rezidui***

$$\varepsilon_2 = \xi_1 - E(\xi_1/x_2)$$

a rezidui

$$\varkappa_2 = \xi_3 - E(\xi_3/x_2)$$

a

$$R_{1,3(2)} = \frac{R_{13} - R_{12}\,R_{23}}{\sqrt{(1 - R_{12}^2)\,(1 - R_{23}^2)}}$$

Analogicky: **parciální korelační koeficienty** $R_{1i(j)}$ **prvního řádu** jsou ***párové korelační koeficienty mezi rezidui***

$$\varepsilon_j = \xi_1 - E(\xi_1/x_j)$$

a rezidui

$$\varkappa_j = \xi_i - E(\xi_i/x_j)$$

a

$$R_{1,i(j)} = \frac{R_{1i} - R_{1j}\,R_{ij}}{\sqrt{(1 - R_{1i}^2)\,(1 - R_{ij}^2)}}$$

Parciální korelační koeficienty druhého řádu $R_{1i(j,k)}$ jsou vlastně ***párové korelační koeficienty reziduí***

$$\varepsilon_{j,k} = \xi_1 - E(\xi_1/(x_j, x_k))$$

a reziduí

$$\varkappa_{j,k} = \xi_i - E(\xi_i/(x_j, x_k))$$

a

$$R_{1i(j,k)} = \frac{R_{1i(j)} - R_{1j(k)}\, R_{ij(k)}}{\sqrt{(1 - R^2_{1j(k)})\,(1 - R^2_{ij(k)})}}$$

Příklad 7.7 *Parciální korelační koeficienty prvního řádu.*
Pro případ tří náhodných proměnných ξ_1, ξ_2, ξ_3 vyčíslete s využitím rovnice parciální korelační koeficienty $R_{1,2(3)}$ a $R_{1,3(2)}$.

Řešení:
Pro výpočet korelačních koeficientů jsou potřebné matice $\mathbf{R}$, $\mathbf{R}_{11}$, $\mathbf{R}_{12}$, $\mathbf{R}_{22}$, $\mathbf{R}_{13}$ a $\mathbf{R}_{33}$. Určíme

$$\mathbf{R} = \begin{bmatrix} 1 & R_{12} & R_{13} \\ R_{12} & 1 & R_{23} \\ R_{13} & R_{23} & 1 \end{bmatrix} \qquad \mathbf{R}_{11} = \begin{bmatrix} 1 & R_{23} \\ R_{23} & 1 \end{bmatrix}$$

$$\mathbf{R}_{12} = \begin{bmatrix} R_{12} & R_{23} \\ R_{13} & 1 \end{bmatrix} \qquad \mathbf{R}_{33} = \begin{bmatrix} 1 & R_{12} \\ R_{12} & 1 \end{bmatrix}$$

$$\mathbf{R}_{13} = \begin{bmatrix} R_{12} & 1 \\ R_{13} & R_{23} \end{bmatrix} \qquad \mathbf{R}_{22} = \begin{bmatrix} 1 & R_{13} \\ R_{13} & 1 \end{bmatrix}$$

Po dosazení

$$R_{1,2(3)} = \frac{R_{12} - R_{23}\, R_{13}}{\sqrt{(1 - R_{23}^2)\,(1 - R_{13}^2)}}$$

resp.

$$R_{1,3(2)} = \frac{R_{13} - R_{12}\, R_{23}}{\sqrt{(1 - R_{23}^2)\,(1 - R_{12}^2)}}$$

Závěr: Parciální korelační koeficienty lze vyčíslit přímo.

Příklad 7.8 *Parciální korelace mezi obsahem dusíku v obilí a půdě*
Pro data uvedená v Příkladu 7.6 stanovte parciální korelační koeficienty mezi obsahem dusíku v obilí a obsahem anorganického dusíku v půdě $R_{1,2(3)}$ a obsahem organického dusíku v půdě $R_{1,3(2)}$.

Řešení: Přímým dosazením do rovnice vyjde

$$R_{1,2(3)} = \frac{0.6934 - 0.4616 \cdot 0.3545}{\sqrt{(1 - 0.4616^2)\,(1 - 0.3545^2)}} = 0.6386$$

a z rovnice

$$R_{1,3(2)} = 0.05325$$

Závěr:

Téměř nulová hodnota $R_{1,3(2)}$ ukazuje na zanedbatelný vliv organického dusíku v půdě na obsah dusíku v obilí.

Poměrně vysoká hodnota párového korelačního koeficientu $R_{13} = 0.3545$ je silně ovlivněna korelací $R_{23} = 0.462$ mezi organickým a anorganickým dusíkem v půdě.

Při detailnější analýze se odhalí bod č. 17 jako silně vybočující, a proto by bylo třeba opakovat analýzu bez tohoto bodu.

Úlohy na výstavbu korelačního modelu

Korelace

Postup analýzy úloh:

1) Graf regresní křivky.
2) Vyšetřete graf rezidua vs. predikce.
3) R, D, s(e).
4) Fisher-Snedecorův test celkové regrese.
5) Odhady parametrů přímky: úsek a směrnice.

Úloha B7.01 *Vliv množství farmaka na dobu práce pacienta*

Zadání: Byl sledován účinek množství podpůrného farmaka na organismus v době, ve které je pacient schopen provést standardní manuální výkon.

Úkoly:

Rozhodněte, zda existuje korelace mezi oběma proměnnými x_2 a x_1 a nalezněte lineární stochastickou vazbu k vyjádření doby manuální práce x_2 na množství farmaka x_1. Co v tomto případě rozumíme pod pojmem míra lineární stochastické vazby?

Data: Množství farmaka x_1 [mg], doba práce x_2 [min]:

x_1	x_2
15	48
...	...
75	200

Úloha B7.02 *Vliv úniku radioaktivního odpadu na růst úmrtnosti na rakovinu*

Zadání: Při úniku radioaktivního odpadu ze skládky v Hanfordu do řeky Columbia bylo vystaveno radioaktivitě obyvatelstvo v 9 okresech. Byla sledována úmrtnost na rakovinu x_1 (úmrtí na 100000 lidí v letech 1959-64) v různých vzdálenostech od Hanfordu x_2.

Úkoly:

1) Účelem je zjistit, zda existuje korelace mezi úmrtností a ozářením, vyjádřeným vzdáleností od skládky.

2) Popište možné korelační modely pro dvě náhodné veličiny.

Data: Úmrtnost na rakovinu x_1 [počet], vzdálenost od radioaktivní skládky x_2 [km]:

x_1	x_2
1.20	120
...	...
11.6	210

Úloha B7.03 *Spotřeba cigaret a úmrtí na rakovinu plic*

Zadání: Z náhodného výběru v šesti státech USA byla zjištěna spotřeba cigaret na obyvatele x_1 a roční míra úmrtnosti na 100 000 lidí následkem rakoviny plic x_2.

Úkoly:

1) Vyšetřete, zda existuje korelace mezi oběma proměnnými x_1 a x_2 na hladině významnosti $\alpha = 0.05$.
2) Uveďte druhy korelačních modelů.

Data: Spotřeba cigaret x_1 [četnost], úmrtnost x_2 [četnost]:

x_1	x_2
3400	24
...	...
2100	20

Úloha B7.04 *Závislost věku žen a koncentrace cholesterolu v krvi*

Zadání: Z náhodného výběru 50 amerických žen byla zjištěna následující data o věku x_1 a koncentraci cholesterolu v krvi [g/l] x_2 u prvních pěti žen.

Úkoly:

1) Vyšetřete míru korelace mezi oběma proměnnými x_1 a x_2.
2) Jaká je příčinná souvislost s korelací dvou veličin?

Data: Věk žen x_1 [roky], koncentrace cholesterolu v krvi x_2 [g/l]:

x_1	x_2
30	1.6
...	...
50	2.7

Úloha B7.05 *Obsahu dehtu, nikotinu a CO v cigaretách*

Zadání: Federální komise obchodu USA posuzuje domácí cigarety dle obsahu dehtu x_1 [mg], nikotinu x_2 [mg] a hmotnosti cigarety x_3 [g] a konečně i obsahu oxidu uhelnatého CO x_4 [mg] v uvolněném cigaretovém kouři. Hlavní hygienik USA totiž považuje faktory x_1, x_2 a x_4 za vysoce nebezpečné pro zdraví člověka. Poslední studie ukázaly, že zvyšující se obsah dehtu a nikotinu spolu nesou i zvýšení obsahu oxidu uhelnatého.

Úkoly:

1) Vyšetřete, zda existuje na hladině výynamnosti $\alpha = 0.05$ korelace mezi proměnnými (a) x_1 a x_4, dále (b) x_2 a x_4, a (c) x_3 a x_4.
2) Vysvětlete pět základních vlastností vícenásobného korelačního koeficientu pro více náhodných veličin.

Data: Obsah dehtu x_1 [mg], obsah nikotinu x_2 [mg], hmotnost cigarety x_3 [g], obsah oxidu uhelnatého CO x_4 [mg]:

Druh cigaret	x_1	x_2	x_3	x_4
Alpine	14.1	0.86	0.9853	13.6
...	...	...	...	...
Winston L.	12.0	0.82	1.1184	14.9