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,

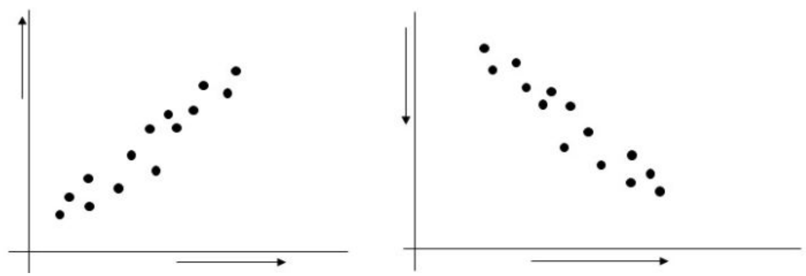
◈**Parciální** popisuje závislost dvou znaků ve vícerozměrném statistickém souboru při vyloučení vlivu ostatních znaků na tuto závislost·

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

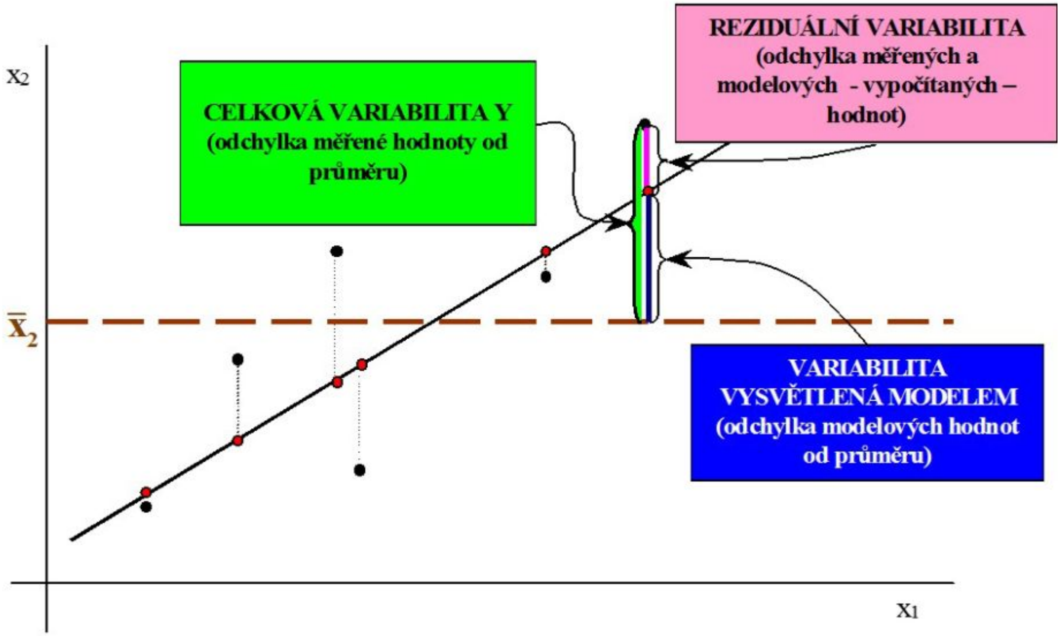
Korelační analýza

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

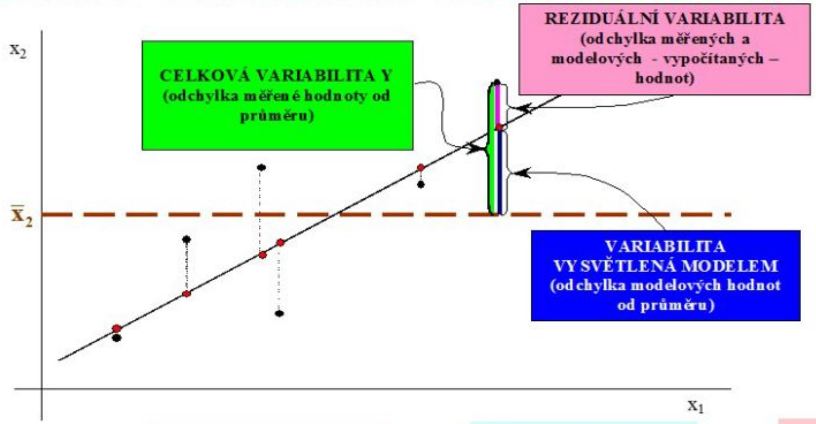
Regresní analýza

- 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

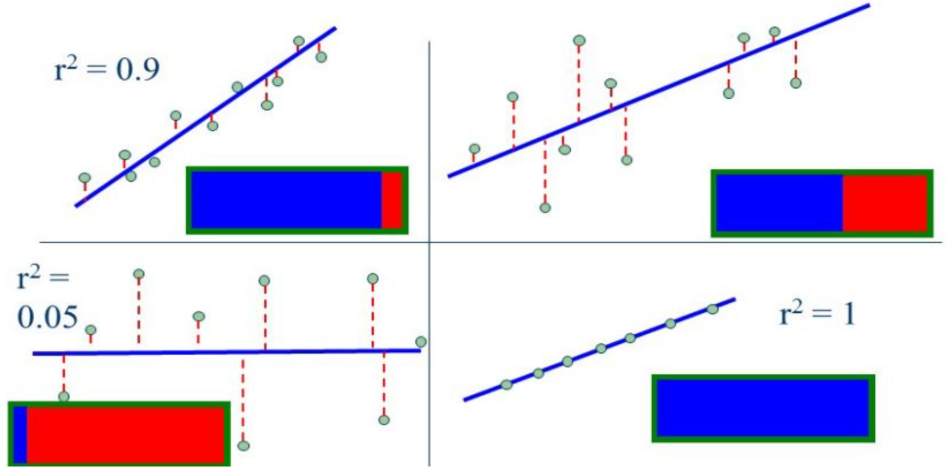
$$R^2 = \frac{S^2_{x'_2}}{S^2_{x_2}} = 1 - \frac{S^2_{x_1 x_2}}{S^2_{x_2}}$$

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,

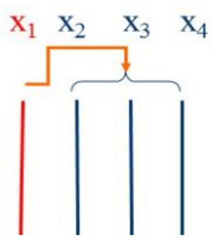
- **Pearsonův**
- **Spearmanův (korelace pořadí)**

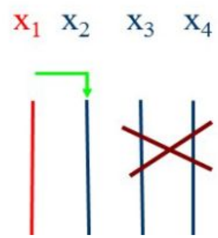
KORELAČNÍ KOEFICIENT

Pro vícenásobnou korelaci:

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

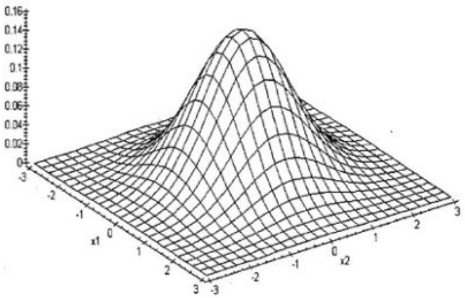
Parciální definuje míru lineární stochastické závislosti mezi náhodnými veličinami x_i a x_j při skonstantnění ostatních složek vektoru $\boldsymbol{x}$





PEARSONŮV KORELAČNÍ KOEFICIENT *r*

Podmínkou je dodržení dvourozměného normálního rozdělení



normovaná kovariance

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

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**

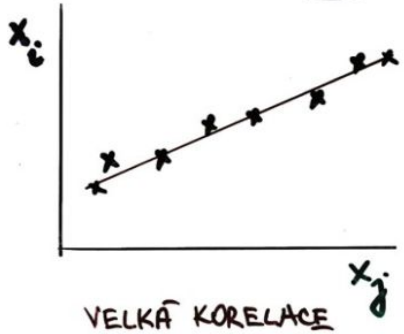
$$cov_{x_1x_2} = \frac{1}{n}\sum_{i=1}^{n}\left(x_{1i} - \overline{x}_1\right)\cdot\left(x_{2i} - \overline{x}_2\right)$$

PEARSONŮV KORELAČNÍ KOEFICIENT *r*

Základní vlastnosti Pearsonova korelačního koeficientu:

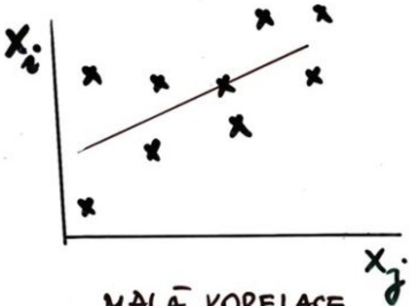
- je to **bezrozměrná** míra lineární korelace;
- nabývá hodnoty **0 – 1 pro kladnou korelaci, 0 – (-1) pro zápornou korelaci**;
- hodnota **0** znamená, že mezi posuzovanými veličinami **není žádný lineární vztah** (může být nelineární) nebo tento vztah zůstal na základě dat, které máme k dispozici,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

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



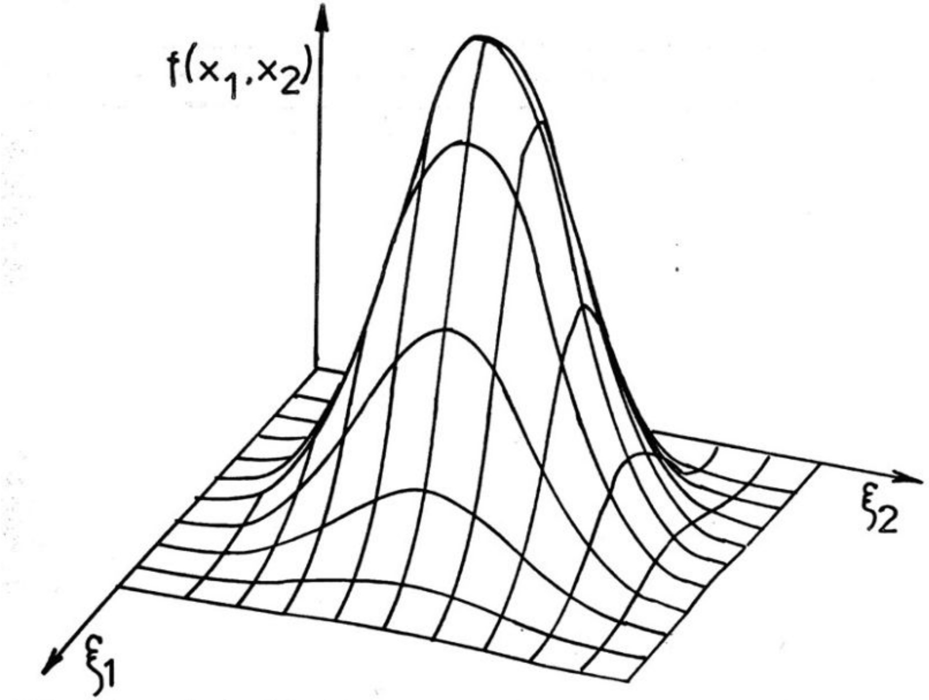
VELKÁ KORELACE

$r_{ij} \approx 1$

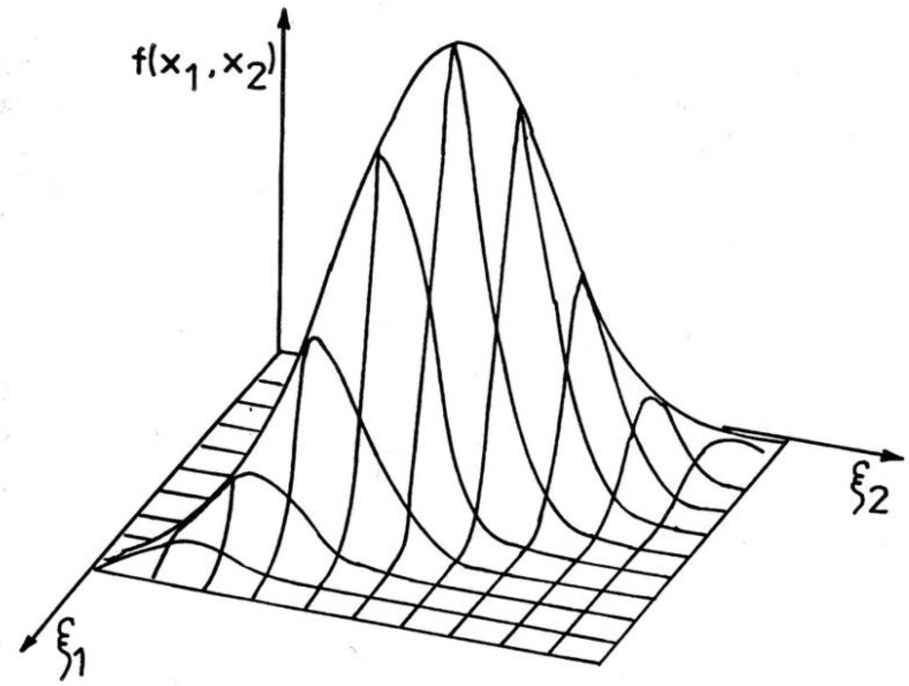


MALÁ KORELACE

$r_{ij} \approx 0.2$

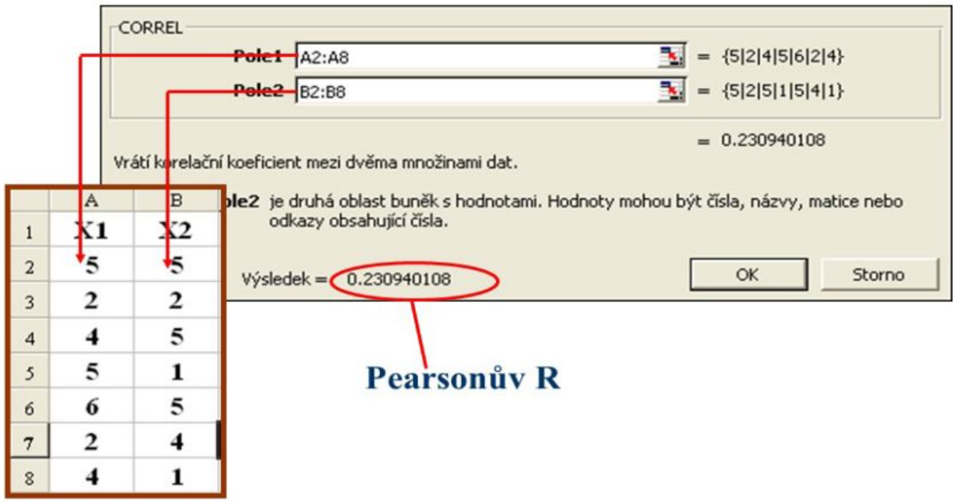


Obr. 4.3 Povrch simultánní hustoty pravděpodobnosti pro $\varrho_{12} = 0$



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

PEARSONŮV KORELAČNÍ KOEFICIENT *r*
výpočet v Excelu



SPEARMANŮV KORELAČNÍ KOEFICIENT

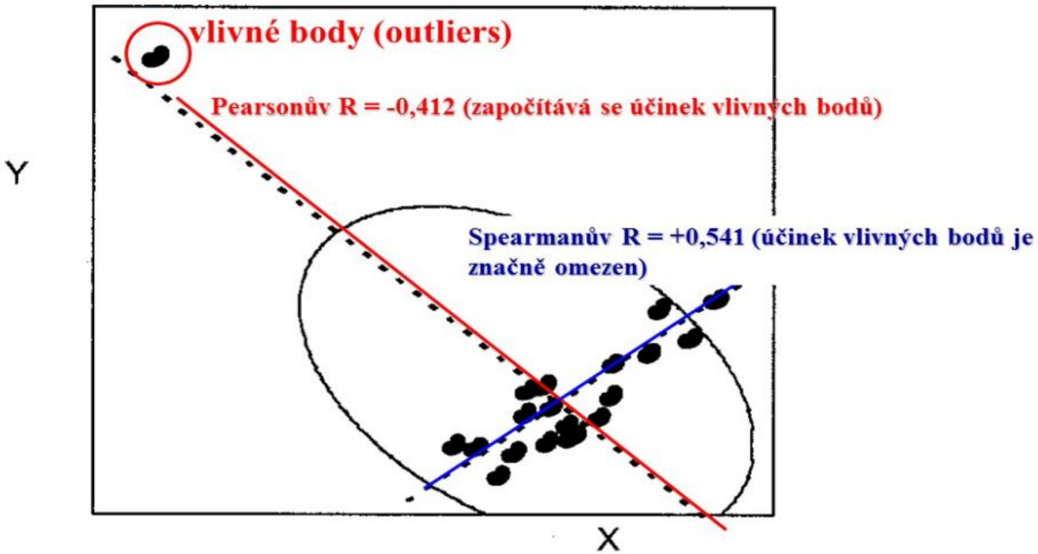
Neparametrický korelační koeficient, vycházející nikoli z hodnot, ale z jejich **pořadí.**

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**

$$\underbrace{\begin{bmatrix} x_{I1} \\ \vdots \\ x_{In} \end{bmatrix}}\underbrace{\begin{bmatrix} x_{II1} & x_{III1} & \cdots & x_{m1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{IIn} & x_{IIIn} & \cdots & x_{mn} \end{bmatrix}}$$

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

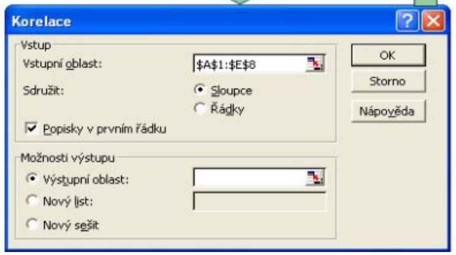
MNOHONÁSOBNÝ KORELAČNÍ KOEFICIENT

$$R_{1(2,3,\dots,m)} = \frac{\begin{bmatrix} 1 & R_{12} & \cdots & R_{1i} & \cdots & R_{1m} \\ R_{21} & 1 & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & 1 & \cdots & \cdots & \cdots \\ R_{i1} & \cdots & \cdots & 1 & \cdots & R_{im} \\ \cdots & \cdots & \cdots & \cdots & 1 & \cdots \\ R_{m1} & R_{m2} & \cdots & R_{mi} & \cdots & 1 \end{bmatrix} \Rightarrow \det(\mathbf{R})}{\begin{bmatrix} \cancel{1} & \cancel{R_{12}} & \cdots & \cancel{R_{1i}} & \cdots & \cancel{R_{1m}} \\ \cancel{R_{21}} & 1 & \cdots & \cdots & \cdots & \cdots \\ \cancel{\cdots} & \cdots & 1 & \cdots & \cdots & \cdots \\ \cancel{R_{i1}} & \cdots & \cdots & 1 & \cdots & R_{im} \\ \cancel{\cdots} & \cdots & \cdots & \cdots & 1 & \cdots \\ \cancel{R_{m1}} & R_{m2} & \cdots & R_{mi} & \cdots & 1 \end{bmatrix} \Rightarrow \det(\mathbf{R}_{(11)})} \Rightarrow \sqrt{1-\frac{\det(\mathbf{R})}{\det(\mathbf{R}_{(11)})}}$$

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

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

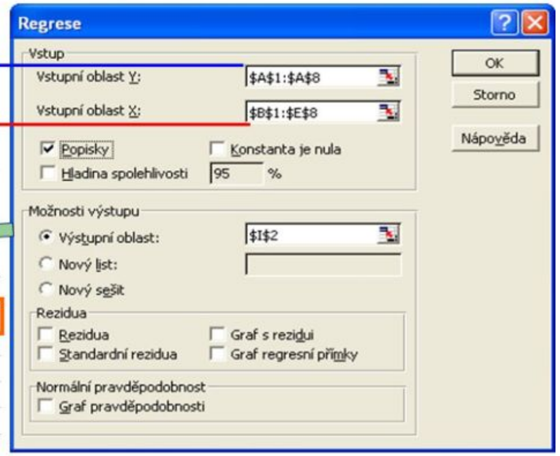
$$= \sqrt{1-\frac{0.004755585}{0.010714947}} = 0.74577$$

MNOHONÁSOBNÝ KORELAČNÍ KOEFICIENT
numerický výpočet v Excelu

❷

	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.

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

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*

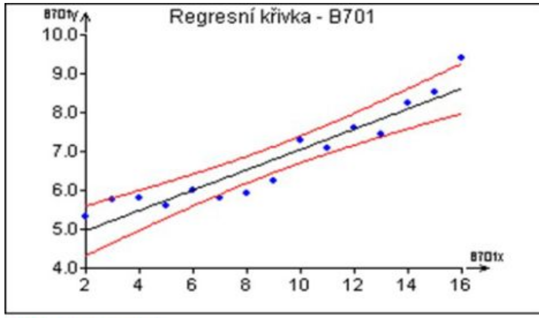
Zadání: Byl sledován účinek množství podpůrného farmaka na organismus v době, ve které je pacient schopen provést standardní manuální výkon.

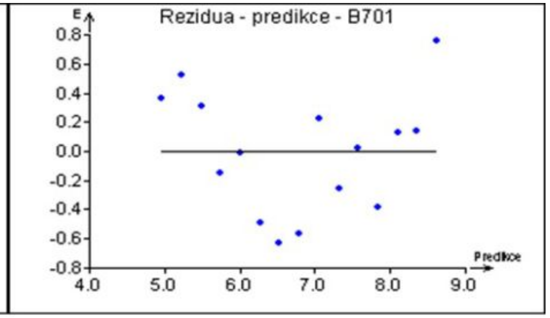
Úkoly:

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





Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	4,441738095	0,2544439571	Významný	2,097402252E-010	3,892045345	4,991430845
B701x	0,2612142857	0,02548690935	Významný	1,35161917E-007	0,2061531656	0,3162754058

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R :	0,9433286393
Koeficient determinace R^2 :	0,8898689218
Predikovaný korelační koeficient Rp :	0,7145109878
Střední kvdratická chyba predikce MEP :	0,2214419698
Akaikeho informační kritérium :	-23,71237813

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F :	105,0411579
Kvantil F (1-alfa, m-1, n-m) :	4,667192732
Pravděpodobnost :	1,351619169E-007
Závěr :	Model je významný

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

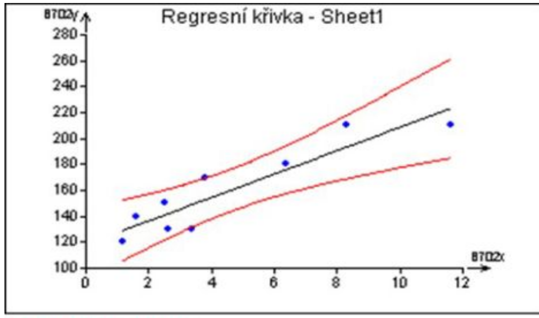
Zadání: Při úniku radioaktivního odpadu ze skládky v Hanfordu do řeky Columbia bylo vystaveno radioaktivitě obyvatelstvo v 9 okresech. Byla sledována úmrtnost na rakovinu x_1 (úmrtí na 100000 lidí v letech 1959-64) v různých vzdálenostech od Hanfordu x_2.

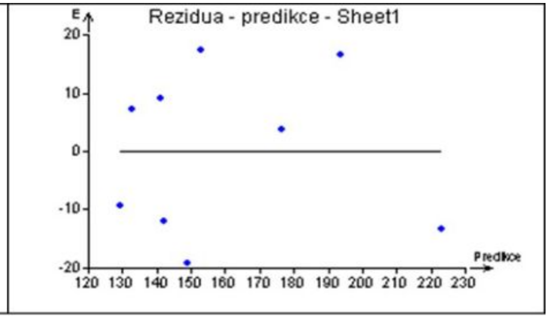
Úkoly:

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





Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	118,4491648	8,364555613	Významný	2,078618235E-006	98,67013372	138,2281958
B702x	9,032790266	1,480149678	Významný	0,0004898156953	5,53279244	12,53278809

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R :	0,9174852467
Koeficient determinace R^2 :	0,841779178
Predikovaný korelační koeficient Rp :	0,4494447646
Střední kvdratická chyba predikce MEP :	344,2421862
Akaikeho informační kritérium :	49,96729099

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F :	37,24196456
Kvantil F (1-alfa, m-1, n-m) :	5,591447851
Pravděpodobnost :	0,0004898156953
Závěr :	Model je významný

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

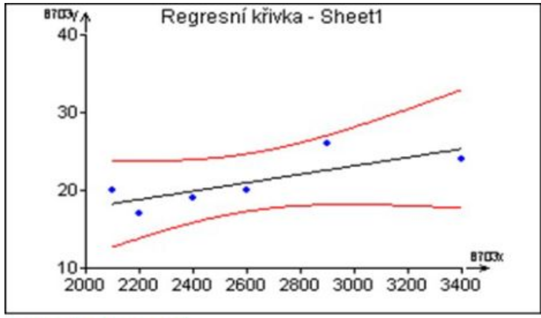
Zadání: Z náhodného výběru v šesti státech USA byla zjištěna spotřeba cigaret na obyvatele x_1 a roční míra úmrtnosti na 100 000 lidí následkem rakoviny plic x_2.

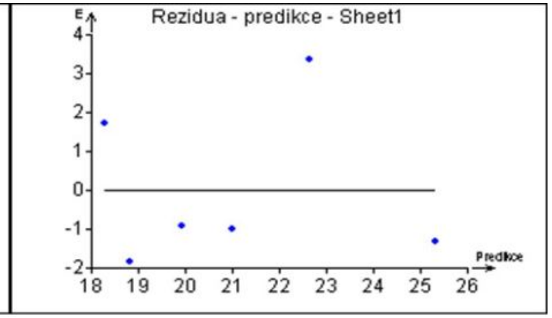
Úkoly:

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





Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	6,898305085	5,601423534	Nevýznamný	0,2855783298	-8,653739869	22,45035004
B703x	0,005423728814	0,002123722726	Nevýznamný	0,06304565971	-0,0004726707551	0,01132012838

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R :	0,7873085301
Koeficient determinace R^2 :	0,6198547215
Predikovaný korelační koeficient Rp :	0,001611171995
Střední kvadratická chyba predikce MEP :	9,707967799
Akaikeho informační kritérium :	11,5983426

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F :	6,522292994
Kvantil F (1-alfa, m-1, n-m) :	7,708647422
Pravděpodobnost :	0,06304565971
Závěr :	Model je nevýznamný

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

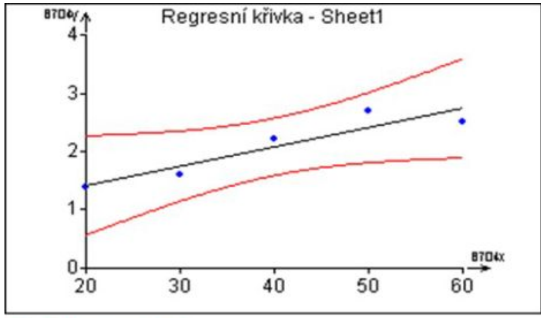
Zadání: Z náhodného výběru 50 amerických žen byla zjištěna následující data o věku x_1 a koncentraci cholesterolu v krvi [g/l] x_2 u prvních pěti žen.

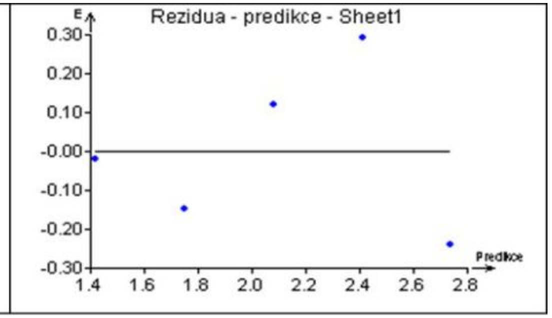
Úkoly:

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





Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	0,76	0,3277193922	Nevýznamný	0,103173012	-0,2829493688	1,802949369
B704x	0,033	0,007724420151	Významný	0,02354350042	0,008417447631	0,05758255237

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R :	0,9267323279
Koeficient determinace R^2 :	0,8588328076
Predikovaný korelační koeficient Rp :	0,2754171753
Střední kvdratická chyba predikce MEP :	0,1205102041
Akaikeho informační kritérium :	-12,64903693

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F :	18,25139665
Kvantil F (1-alfa, m-1, n-m) :	10,12796449
Pravděpodobnost :	0,02354350042
Závěr :	Model je významný

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

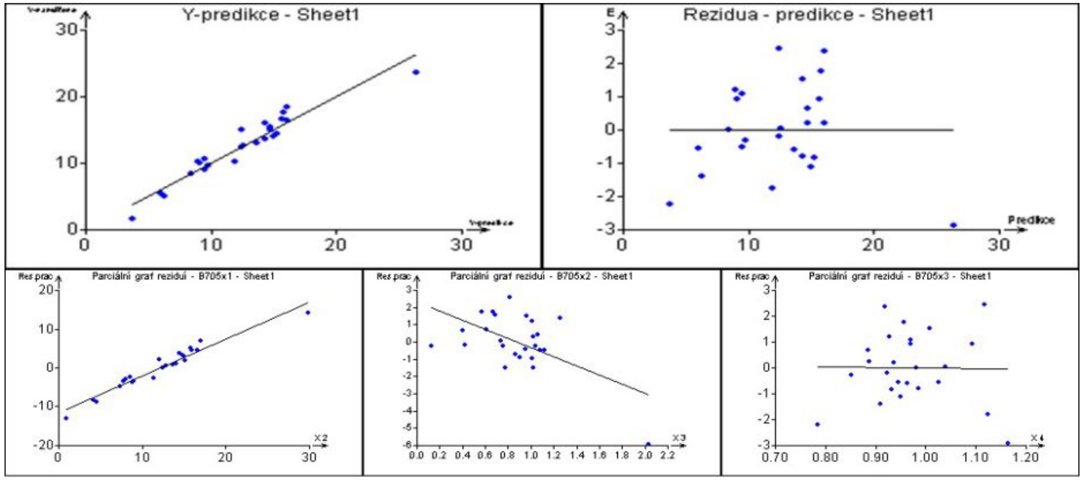
Zadání: Federální komise obchodu USA posuzuje domácí cigarety dle obsahu dehtu x_1 [mg], nikotinu x_2 [mg] a hmotnosti cigarety x_3 [g] a konečně i obsahu oxidu uhelnatého CO x_4 [mg] v uvolněném cigaretovém kouři. Hlavní hygienik USA totiž považuje faktory x_1, x_2 a x_4 za vysoce nebezpečné pro zdraví 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 výýnamnosti $\alpha = 0.05$ 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



Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	3,202190019	3,461754727	Nevýznamný	0,3654642718	-3,996923038	10,40130308
B705x1	0,9625738566	0,2422443598	Významný	0,0006920652485	0,4587991322	1,466348581
B705x2	-2,631661111	3,900557446	Nevýznamný	0,5072342597	-10,74331438	5,479992156
B705x3	-0,1304818533	3,88534182	Nevýznamný	0,9735267532	-8,210492493	7,949528787

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R :	0,9584306693
Koeficient determinace R^2 :	0,9185893479
Predikovaný korelační koeficient Rp :	0,6956222677
Střední kvdratická chyba predikce MEP :	3,579101884
Akaikeho informační kritérium :	22,07173971

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F :	78,9838341
Kvantil F (1-alfa, m-1, n-m) :	3,072466985
Pravděpodobnost :	1,328810092E-011
Závěr :	Model je významný