KORELACE

typy podle počtu korelovaných znaků

- **Jednoduchá** popisuje vztah dvou znaků,

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

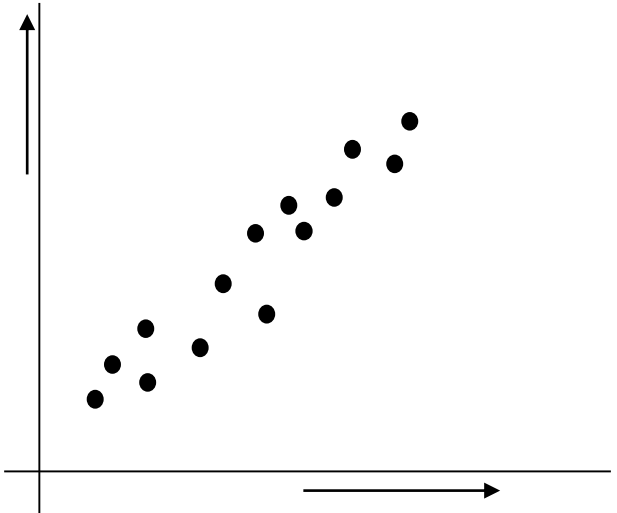
- **Parciální** popisuje závislost dvou znaků ve vícerozměrném statistickém souboru při vyloučení vlivu ostatních znaků na tuto závislost·

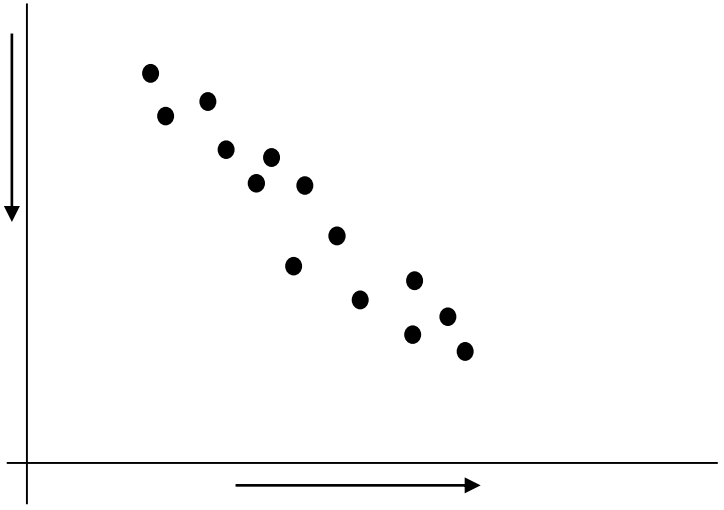
KORELACE

typy podle smyslu změny hodnot

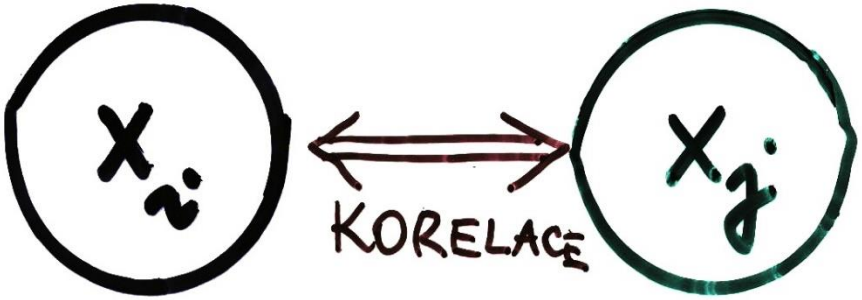
Kladná značí, že se zvyšováním hodnot jednoho znaku se zvyšují i hodnoty druhého znaku,

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



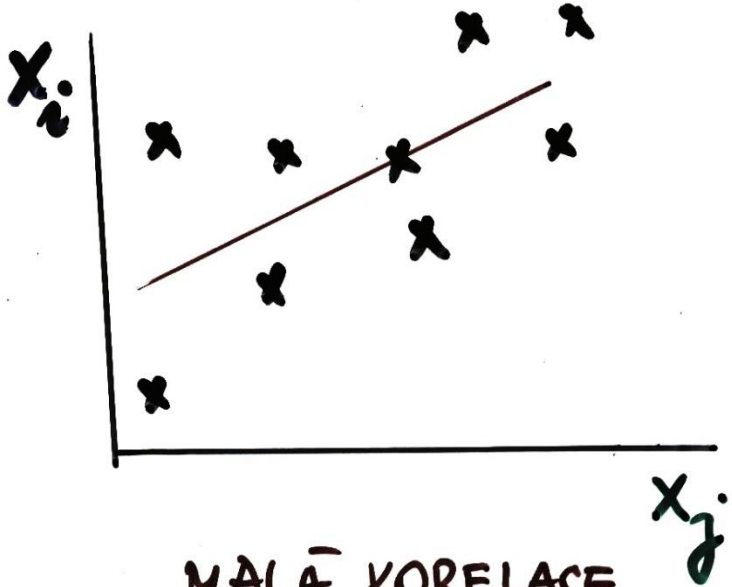


KORELAČNÍ CHARAKTERISTIKY = míry lineární závislosti mezi dvěma či více náhodnými veličinami



VELKÁ KORELACE

$r_{ij} \approx 1$



MALÁ KORELACE

$r_{ij} \approx 0.2$

KORELAČNÍ POČET

Korelační analýza

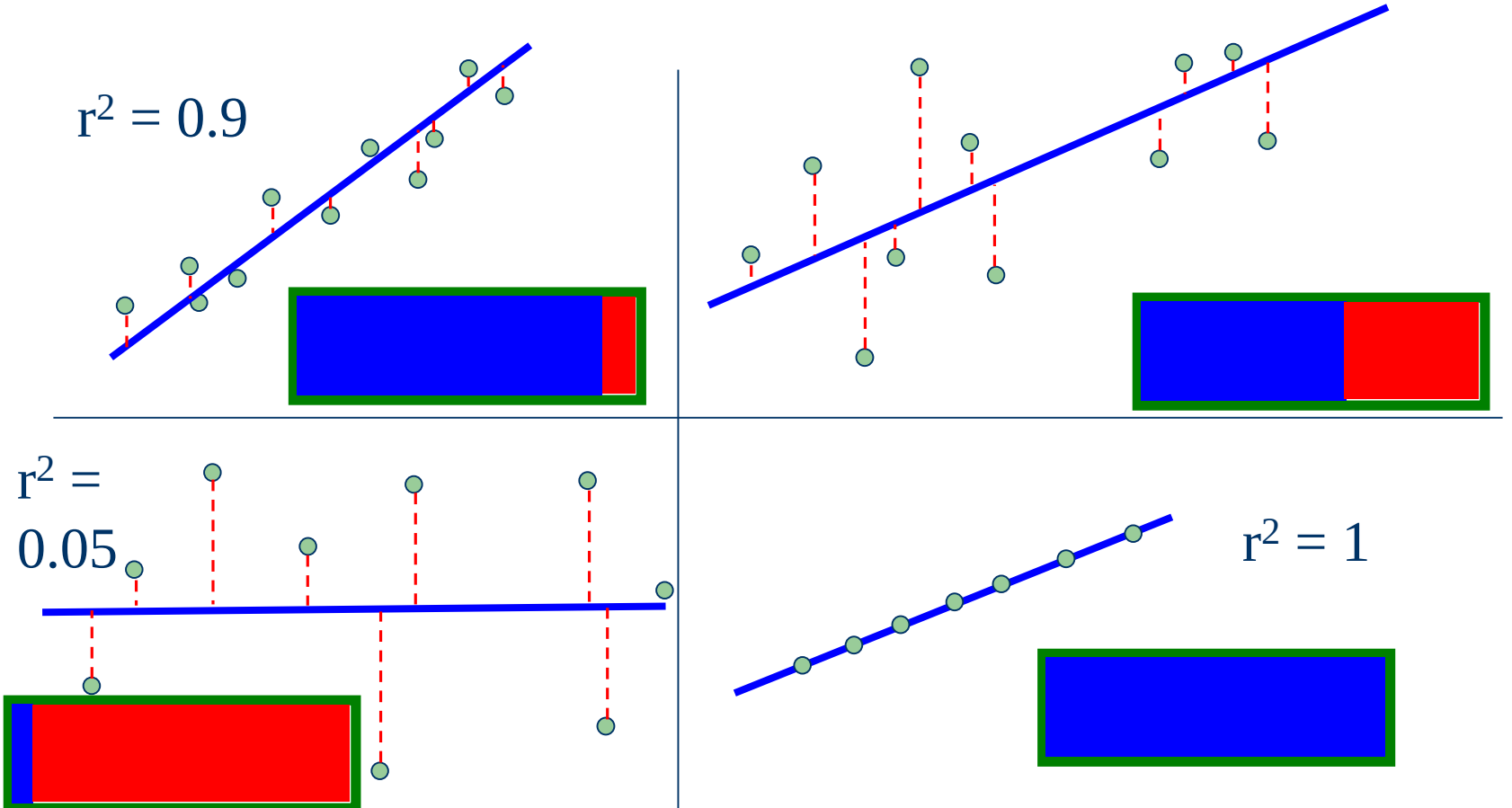
- zjišťuje ***existenci závislosti*** a její druhy,
- měří ***těsnost závislosti***,
- ověřuje ***hypotézy o statistické významnosti*** *závislosti*;

Regresní analýza

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

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**,**

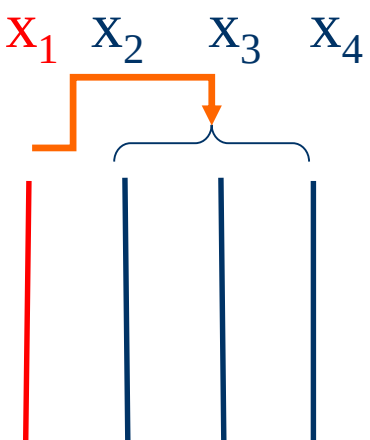
- **Pearsonův**
- **Spearmanův (korelace pořadí)**

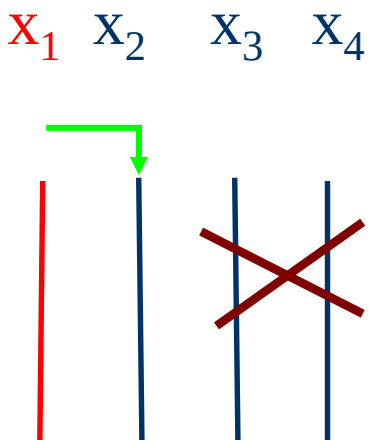
KORELAČNÍ KOEFICIENT

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 $\boldsymbol{x}$

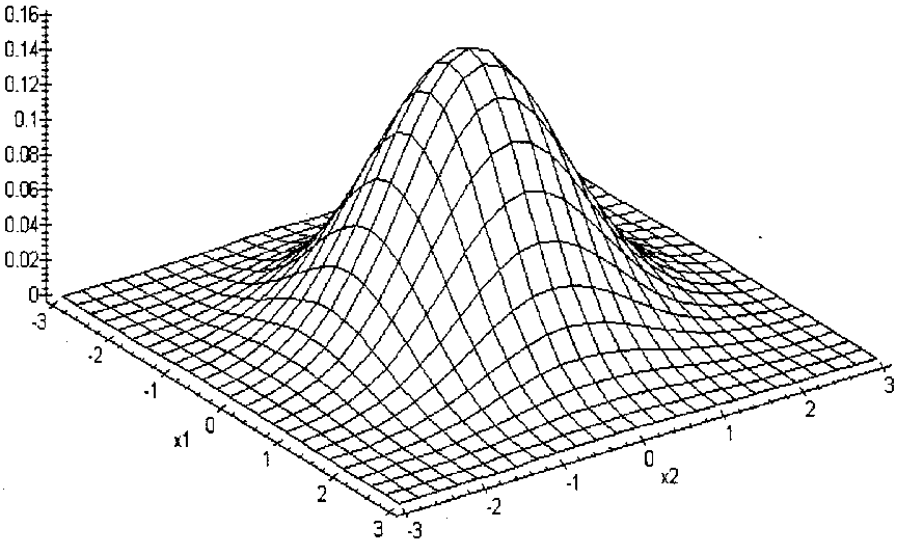
Parciální definuje míru lineární stochastické závislosti mezi náhodnými veličinami x_i a x_j při skonstantnění ostatních složek vektoru $\boldsymbol{x}$





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**

$$\mathrm{cov}_{x_1 x_2} = \frac{1}{n}\sum_{i=1}^{n}\left(x_{1i} - \bar{x}_1\right)\cdot\left(x_{2i} - \bar{x}_2\right)$$

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.

SPEARMANŮV KORELAČNÍ KOEFICIENT

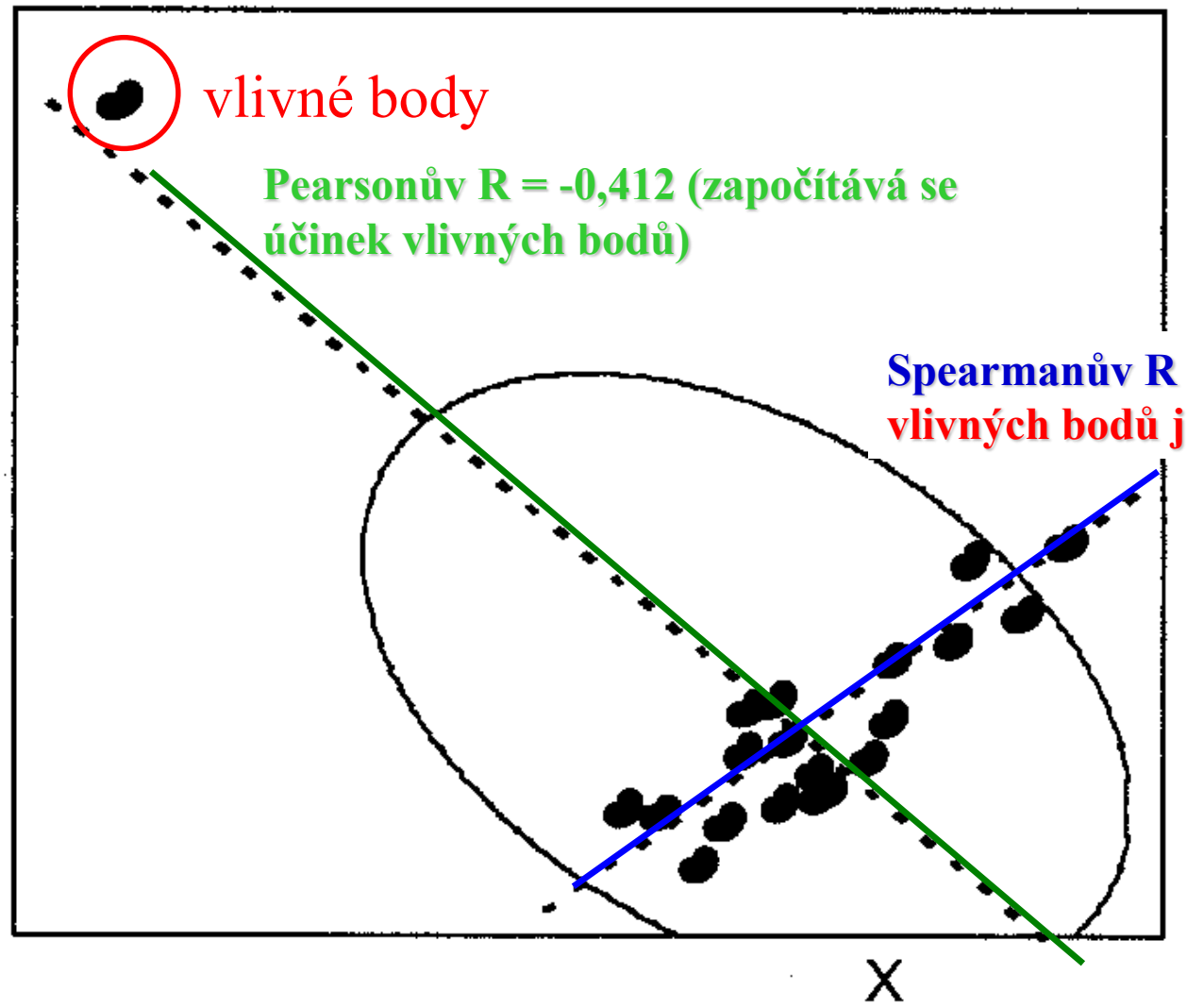
Neparametrický korelační koeficient, vycházející nikoli z hodnot, ale z jejich **pořadí.**

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

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

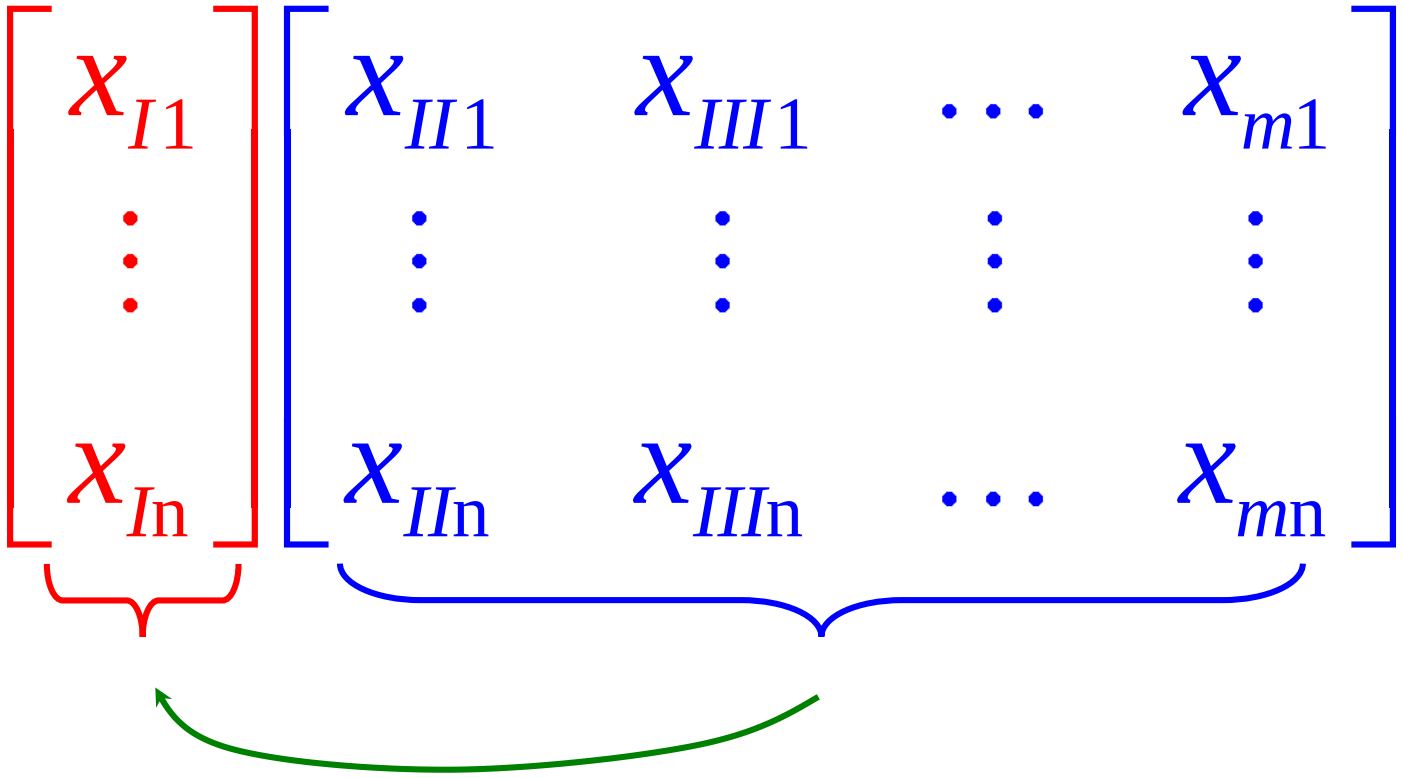
Diference mezi pořadími hodnot *x* a *y* v jednom řádku

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** $x_2, \ldots, x_m$.

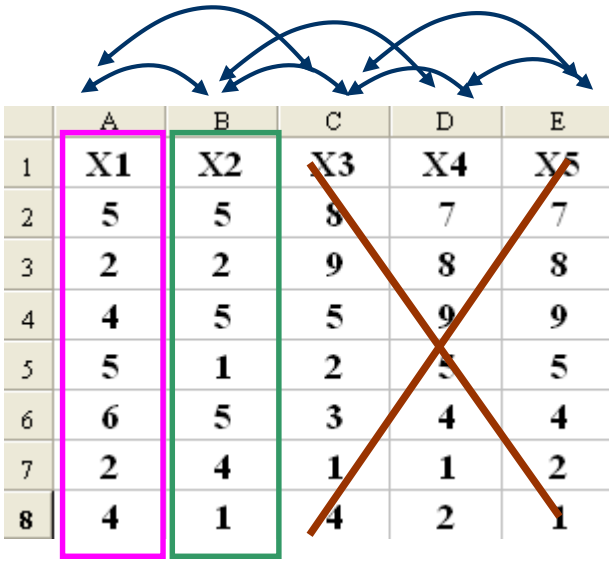
c) Pokud je $R = 0$**, potom jsou také všechny párové korelační koeficienty nulové.**

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

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

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)

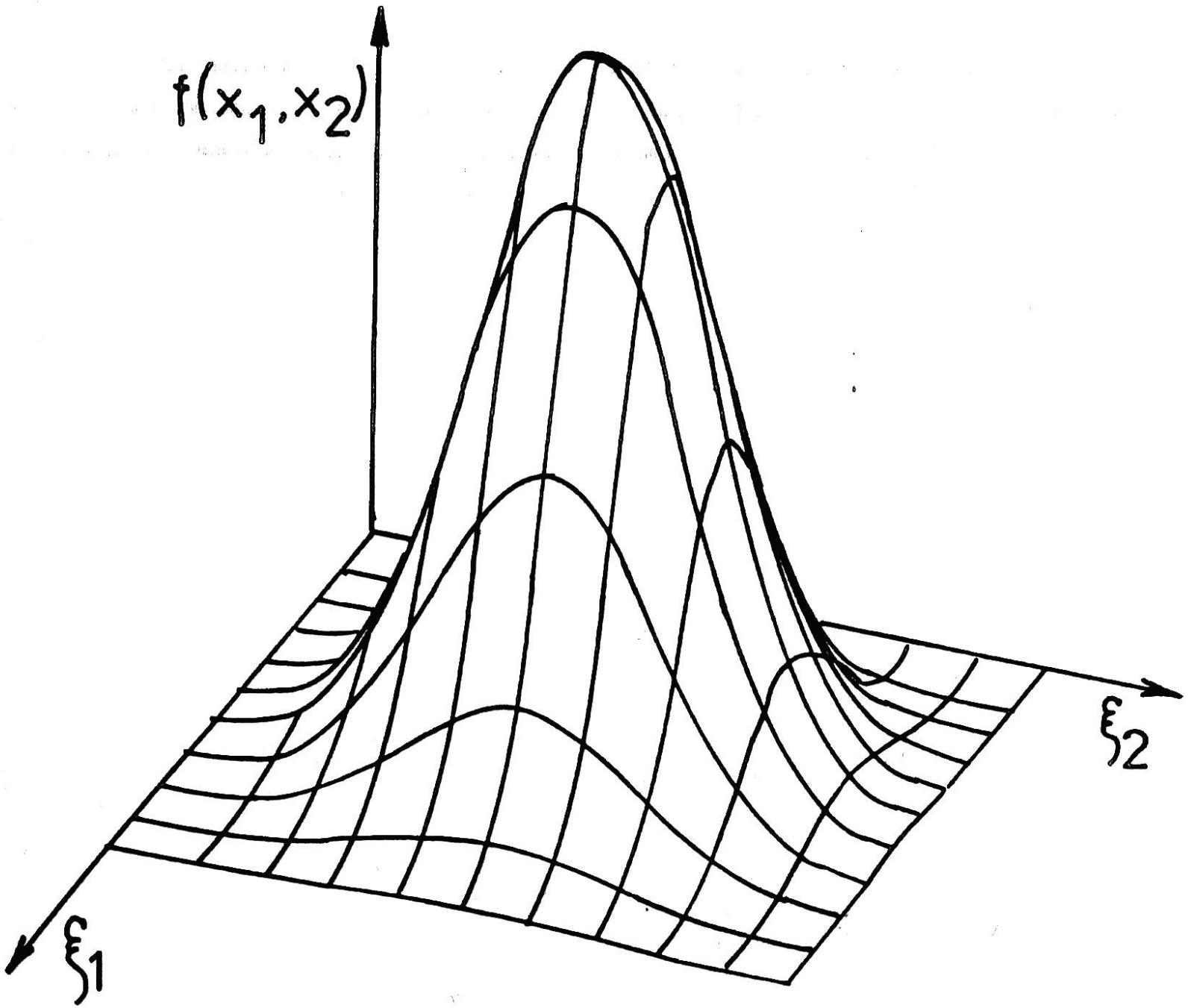
TEST VÝZNAMNOSTI *R*

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

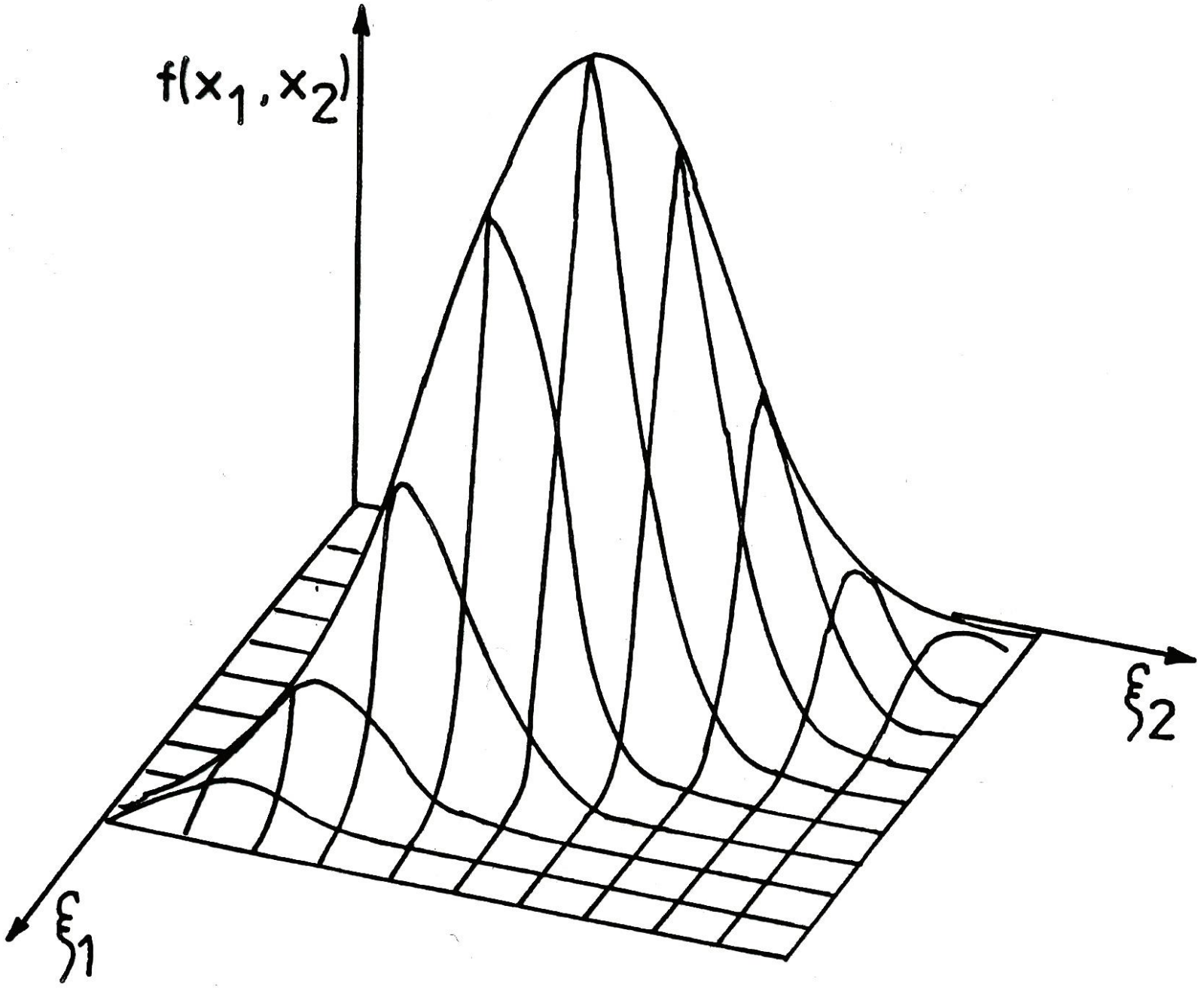
Pro párový *R*: $t_R = \frac{R \cdot \sqrt{n-2}}{\sqrt{1-R^2}}$ KH $t_{\alpha,n-2}$ **n – počet hodnot výběru**

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**



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



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

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

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

F1=nápověda F2=uložení F3=čtení F4=numerický mód F5=zoom ESC=konec

Řádek 1 Sloupec 5 Insert NUM B701.txt

```
B701x  B701y
.00000  5.33000
.00000  5.75000
.00000
.00000
.00000
.00000  5.78000
.00000  5.90000
.00000  6.23000
.00000  7.28000
.00000  7.06000
.00000  7.60000
.00000  7.45000
.00000  8.23000
.00000  8.50000
.00000  9.38000
```

Čtení ze souboru:

B701.txt

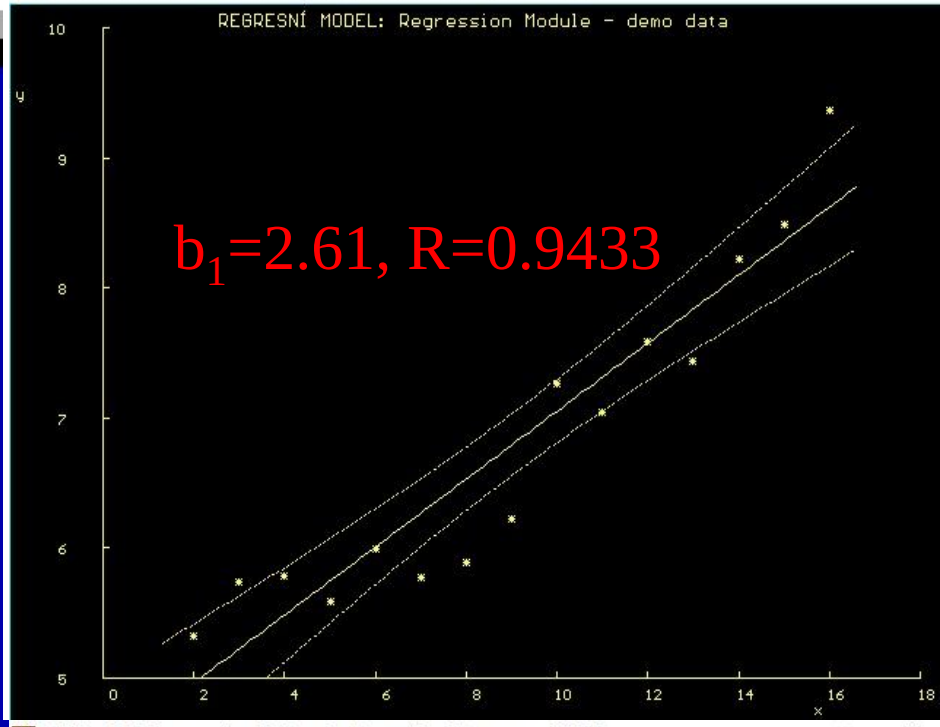
(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test H0: B[j] = 0 vs. HA: B[j] <> 0 t-kriterium	hypoteza H0 je	Hlad. výz.
B[0]	4.4417E+00	2.5444E-01	1.7457E+01	Zamítnuta	0.000
B[1]	2.6121E-01	2.5487E-02	1.0249E+01	Zamítnuta	0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 9.4333E-01
Koeficient determinace, R^2	: 8.8987E-01
Predikovaný korelační koeficient, Rp^2	: 9.1940E-01
Střední kvadratická chyba predikce, MEP	: 2.2144E-01
Akaikeho informační kritérium, AIC	:-2.3712E+01

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG



DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

V Ý S L E D K Y

(6) TESTOVÁNÍ REGRESNÍHO TRIPLETU (DATA + MODEL + METODA):

Fisher-Snedocorův test významnosti regrese,F	: 1.0504E+02
Tabulkový kvantil, F(1-alpha,m-1,n-m)	: 4.6672E+00
Závěr: Navržený model je přijat jako významný.	
Spočtená hladina významnosti	: 0.000
Scottovo kriterium multikolinearity, M	: 2.0291E-16
Závěr: Navržený model je korektní.	
Cook-Weisbergův test heteroskedasticity, Sf	: 3.9798E+01
Tabulkový kvantil, Chi^2(1-alpha,1)	: 3.8415E+00
Závěr: Rezidua vykazují heteroskedasticitu.	
Spočtená hladina významnosti	: 0.000
Jarque-Berraův test normality reziduí, L(e)	: 5.1475E-01
Tabulkový kvantil, Chi^2(1-alpha,2)	: 5.9915E+00
Závěr: Normalita je přijata.	
Spočtená hladina významnosti	: 0.773
Waldův test autokorelace, Wa	: 4.4330E+00
Tabulkový kvantil, Chi^2(1-alpha,1)	: 3.8415E+00
Závěr: Rezidua jsou autokorelována.	
Spočtená hladina významnosti	: 0.035

Nápověda-F1 Řádek: 108 - 130 Celkem: 196 Délka: 10766

Závěry o testu korelace

Otázka	Odpověď
Graf regresní křivky.	
Graf rezidua vs. predikce.	
R, D, s(e).	
Fisher-Snedecorův test celkové regrese	
Odhady parametrů přímky: úsek a směrnice	

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

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

F1=nápověda F2=uložení F3=čtení F4=numerický mód F5=zoom ESC=konec

Řádek 1 Sloupec 5 Insert NUM B702.txt

B702x	B702y
1.20000	120.00000
2.50000	150.00000
1.60000	
8.30000	
6.40000	
3.40000	130.00000
3.80000	170.00000
2.60000	130.00000
11.60000	210.00000

Čtení ze souboru:

B702.txt

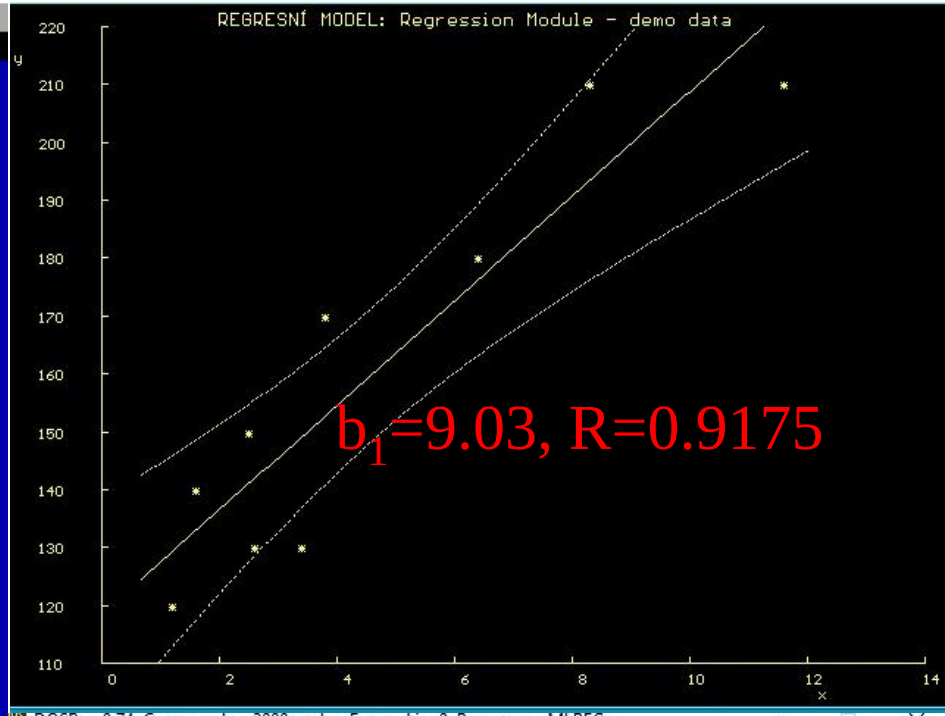
(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test H0: B[j] = 0 vs. HA: B[j] <> 0 t-kriterium	hypoteza H0 je	Hlad. výz.
B[0]	1.1845E+02	8.3646E+00	1.4161E+01	Zamítnuta	0.000
B[1]	9.0328E+00	1.4801E+00	6.1026E+00	Zamítnuta	0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 9.1749E-01
Koeficient determinace, R^2	: 8.4178E-01
Predikovaný korelační koeficient, Rp^2	: 8.1878E-01
Střední kvadratická chyba predikce, MEP	: 3.4424E+02
Akaikeho informační kritérium, AIC	: 4.9967E+01

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG



DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

V Ý S L E D K Y

(6) TESTOVÁNÍ REGRESNÍHO TRIPLETU (DATA + MODEL + METODA):

Fisher-Snedocorův test významnosti regrese,F	: 3.7242E+01
Tabulkový kvantil, F(1-alpha,m-1,n-m)	: 5.5914E+00
Závěr: Navržený model je přijat jako významný.	
Spočtená hladina významnosti	: 0.000
Scottovo kriterium multikolinearity, M	: 0.0000E+00
Závěr: Navržený model je korektní.	
Cook-Weisbergův test heteroskedasticity, Sf	: 1.1795E+01
Tabulkový kvantil, Chi^2(1-alpha,1)	: 3.8415E+00
Závěr: Rezidua vykazují heteroskedasticitu.	
Spočtená hladina významnosti	: 0.001
Jarque-Berrauv test normality reziduí, L(e)	: 8.6074E-01
Tabulkový kvantil, Chi^2(1-alpha,2)	: 5.9915E+00
Závěr: Normalita je přijata.	
Spočtená hladina významnosti	: 0.650
Waldův test autokorelace, Wa	: 4.6005E-01
Tabulkový kvantil, Chi^2(1-alpha,1)	: 3.8415E+00
Závěr: Rezidua nejsou autokorelována.	
Spočtená hladina významnosti	: 0.498

Napověda-F1 Řádek: 98 - 120 Celkem: 168 Délka: 8779

Závěry o testu korelace

Otázka	Odpověď
Graf regresní křivky.	
Graf rezidua vs. predikce.	
R, D, s(e).	
Fisher-Snedecorův test celkové regrese	
Odhady parametrů přímky: úsek a směrnice	

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

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

F1=nápověda F2=uložení F3=čtení F4=numerický mód F5=zoom ESC=konec

Řádek 1 Sloupec 5 Insert NUM B703.txt

Textový editor

```
B703x  B703y
3400.00000  24.00000
2600.00
2200.00
2400.00
2900.00000  26.00000
2100.00000  20.00000
```

Čtení ze souboru:

B703.txt

Hlášení

Vyhledávání chyb: F6=první F7=předchozí F8=následující F9=poslední

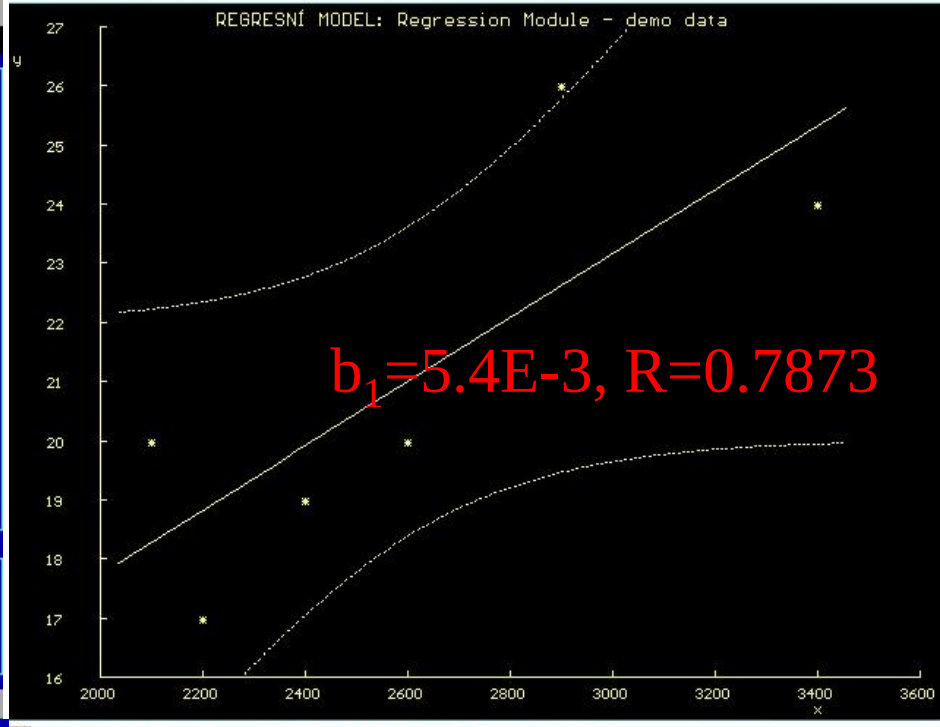
(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test H0: B[j] = 0 vs. HA: B[j] <> 0 t-kriterium	hypoteza H0 je	Hlad. výz.
B[0]	6.8983E+00	5.6014E+00	1.2315E+00	Akceptována	0.286
B[1]	5.4237E-03	2.1237E-03	2.5539E+00	Akceptována	0.063

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 7.8731E-01
Koeficient determinace, R^2	: 6.1985E-01
Predikovaný korelační koeficient, Rp^2	: 0.0000E+00
Střední kvadratická chyba predikce, MEP	: 9.7080E+00
Akaikeho informační kritérium, AIC	: 1.1598E+01

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG



b_1=5.4E-3, R=0.7873

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

V Ý S L E D K Y

(6) TESTOVÁNÍ REGRESNÍHO TRIPLETU (DATA + MODEL + METODA):

Fisher-Snedocorův test významnosti regrese,F	: 6.5223E+00
Tabulkový kvantil, F(1-alpha,m-1,n-m)	: 7.7086E+00
Závěr: Navržený model není přijat jako významný.	
Spočtená hladina významnosti	: 0.063
Scottovo kriterium multikolinearity, M	: 1.3618E-16
Závěr: Navržený model je korektní.	
Cook-Weisbergův test heteroskedasticity, Sf	: 3.2968E+00
Tabulkový kvantil, Chi^2(1-alpha,1)	: 3.8415E+00
Závěr: Rezidua vykazují homoskedasticitu.	
Spočtená hladina významnosti	: 0.069
Jarque-Berraův test normality reziduí, L(e)	: 9.1080E-01
Tabulkový kvantil, Chi^2(1-alpha,2)	: 5.9915E+00
Závěr: Normalita je přijata.	
Spočtená hladina významnosti	: 0.634
Waldův test autokorelace, Wa	: 1.2144E+00
Tabulkový kvantil, Chi^2(1-alpha,1)	: 3.8415E+00
Závěr: Rezidua nejsou autokorelována.	
Spočtená hladina významnosti	: 0.270

Napověda-F1 Řádek: 95 - 117 Celkem: 156 Délka: 7804

Závěry o testu korelace

Otázka	Odpověď
Graf regresní křivky.	
Graf rezidua vs. predikce.	
R, D, s(e).	
Fisher-Snedecorův test celkové regrese	
Odhady parametrů přímky: úsek a směrnice	

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

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

F1=nápověda F2=uložení F3=čtení F4=numerický mód F5=zoom ESC=konec

Řádek 1 Sloupec 5 Insert NUM B704.txt

Textový editor

```
3.000000 E+01      1.600000
6.000000 E+01      2.500000
4.0
2.0
5.0
```

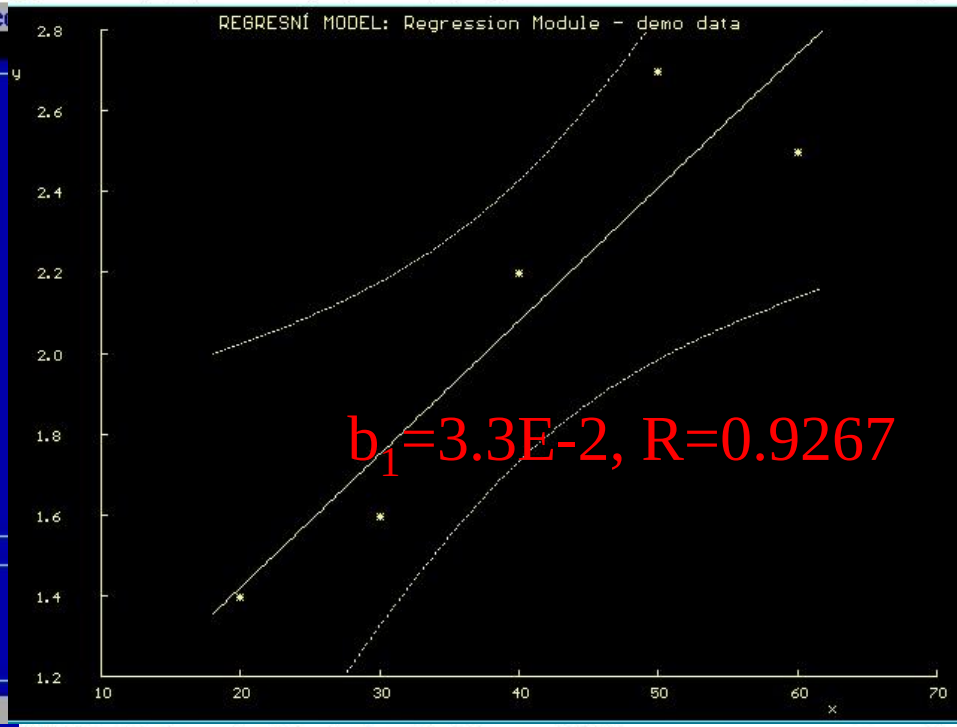
Čtení ze souboru:

B704.txt

Hlášení

Vyhledávání chyb: F6=první F7=předchozí F8=následující F9=poslední

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG



(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test H0: B[j] = 0 vs. HA: B[j] <> 0 t-kriterium	hypoteza H0 je	Hlad. výz.
B[0]	7.6000E-01	3.2772E-01	2.3191E+00	Akceptována	0.103
B[1]	3.3000E-02	7.7244E-03	4.2722E+00	Zamítnuta	0.024

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 9.2673E-01
Koeficient determinace, R^2	: 8.5883E-01
Predikovaný korelační koeficient, Rp^2	: 7.2443E-01
Střední kvadratická chyba predikce, MEP	: 1.2051E-01
Akaikeho informační kritérium, AIC	:-1.2649E+01

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

VÝSLEDKY

(6) TESTOVÁNÍ REGRESNÍHO TRIPLETU (DATA + MODEL + METODA):

Fisher-Snedocorův test významnosti regrese,F	: 1.8251E+01
Tabulkový kvantil, F(1-alpha,m-1,n-m)	: 1.0128E+01
Závěr: Navržený model je přijat jako významný.	
Spočtená hladina významnosti	: 0.024
Scottovo kriterium multikolinearity, M	:-9.7334E-17
Závěr: Navržený model je korektní.	
Cook-Weisbergův test heteroskedasticity, Sf	: 1.2054E+00
Tabulkový kvantil, Chi^2(1-alpha,1)	: 3.8415E+00
Závěr: Rezidua vykazují homoskedasticitu.	
Spočtená hladina významnosti	: 0.272
Jarque-Berraův test normality reziduí, L(e)	: 3.9220E-01
Tabulkový kvantil, Chi^2(1-alpha,2)	: 5.9915E+00
Závěr: Normalita je přijata.	
Spočtená hladina významnosti	: 0.822
Waldův test autokorelace, Wa	: 5.5527E-04
Tabulkový kvantil, Chi^2(1-alpha,1)	: 3.8415E+00
Závěr: Rezidua nejsou autokorelována.	
Spočtená hladina významnosti	: 0.981

Napověda-F1 Řádek: 90 - 112 Celkem: 148 Délka: 7439

Závěry o testu korelace

Otázka	Odpověď
Graf regresní křivky.	
Graf rezidua vs. predikce.	
R, D, s(e).	
Fisher-Snedecorův test celkové regrese	
Odhady parametrů přímky: úsek a směrnice	

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

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

F1=nápověda F2=uložení F3=čtení F4=numerický mód F5=zoom ESC=konec

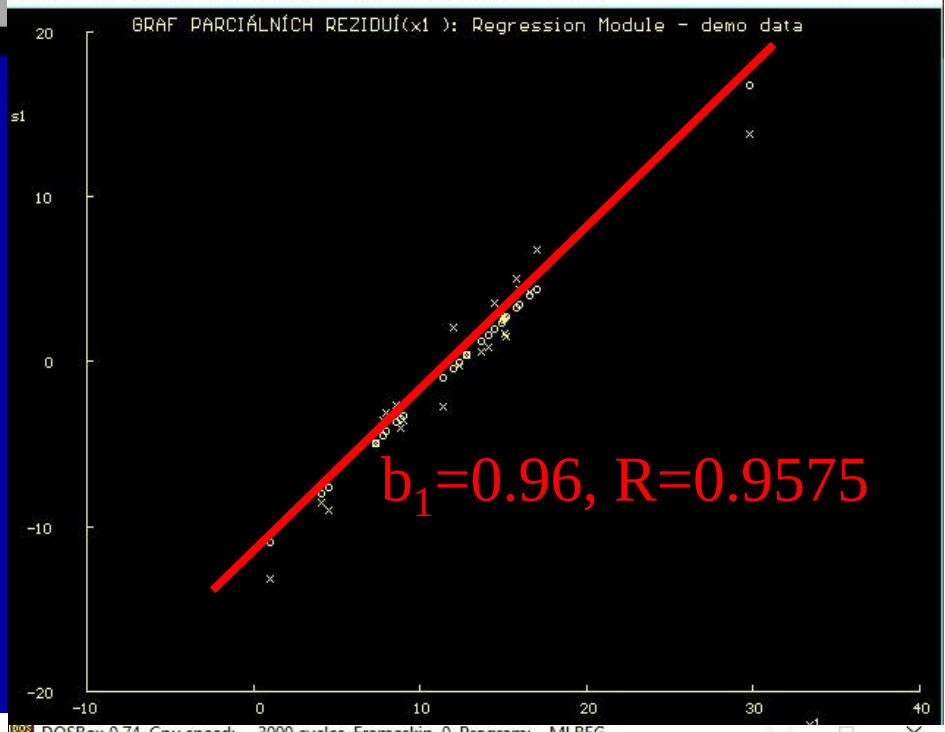
Řádek 1 Sloupec 5 Insert NUM B705.txt

B705x1	B705x2	B705x3	B705y
14.10000	0.86000	0.98530	13.60000
16.00000	1.06000	1.09380	16.60000
29.80000			
B.00000			
4.10000			
15.00000	1.04000	0.88850	15.00000
8.80000	0.76000	1.02670	9.00000
12.40000	0.95000	0.92250	12.30000
16.60000	1.12000	0.93720	16.30000
14.90000	1.02000	0.88580	15.40000
13.70000	1.01000	0.96430	13.00000
15.10000	0.90000	0.93160	14.40000
7.80000	0.57000	0.97050	10.00000
11.40000	0.78000	1.12400	10.20000
9.00000	0.74000	0.85170	9.50000
1.00000	0.13000	0.78510	1.50000
17.00000	1.26000	0.91860	18.50000
12.80000	1.08000	1.03950	12.60000
15.80000	0.96000	0.95730	17.50000
4.50000	0.42000	0.91060	4.90000
14.50000	1.01000	1.00700	15.90000
7.30000	0.61000	0.98060	8.50000

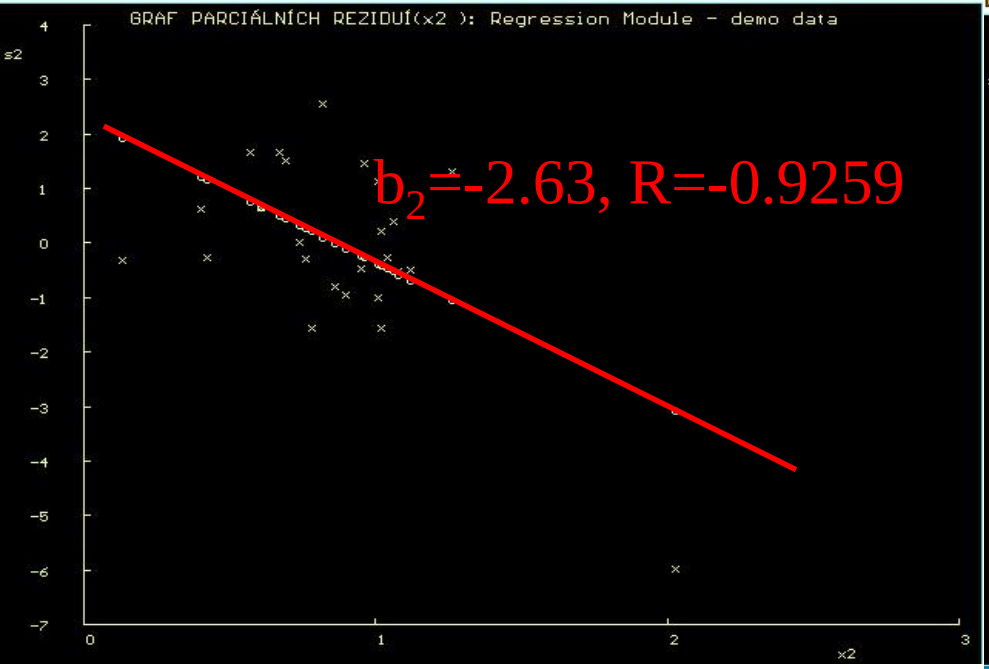
Čtení ze souboru:

B705.txt

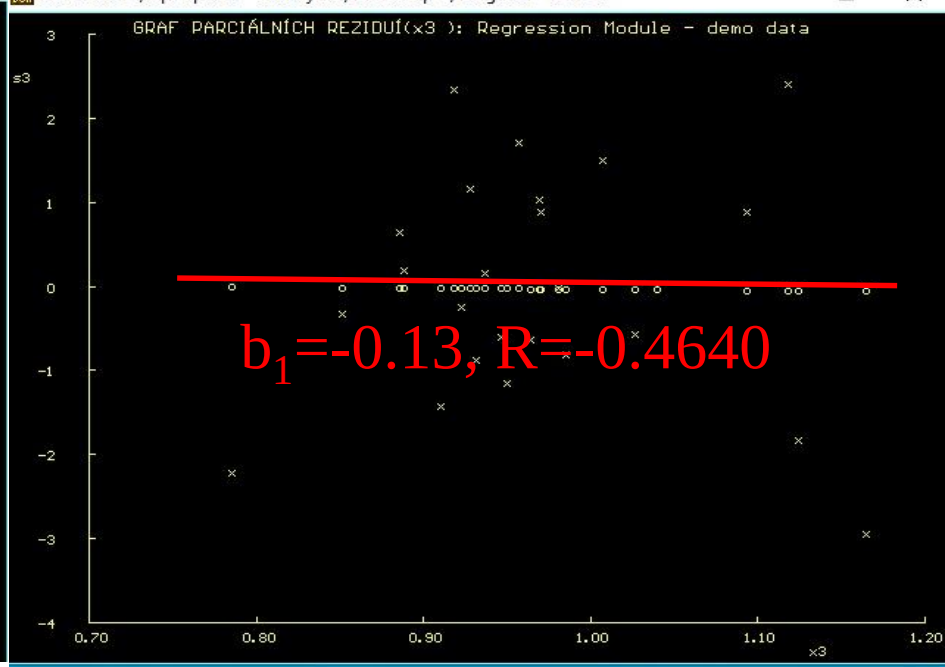
DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG



DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG



DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG



DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

V Ý S L E D K Y

(1) PŘEDBĚŽNÁ STATISTICKÁ ANALÝZA:

Proměnná	Průměr	Směrodatná odchylka	Párový korelační koeficient	Spočtená hladina výz.
y	1.2528E+01	4.7397E+00	1.0000	-----
x1	1.2216E+01	5.6658E+00	0.9575	0.000
x2	8.7640E-01	3.5406E-01	0.9259	0.000
x3	9.7028E-01	8.7721E-02	0.4640	0.019

Párové korelační koeficienty mezi dvojicemi vysvětlujících proměnných		Spočtená hladina významnosti
x1 versus x2 :	9.7661E-01	0.000
x1 versus x3 :	4.9077E-01	0.013
x2 versus x3 :	5.0018E-01	0.011

(2) INDIKACE MULTIKOLINEARITY:

Č [j]	Vlastní čísla korel. matice l[j]	Čísla podmíněnosti K[j]	Variance inflation factor VIF[j]	Vícenás.korel. koef pro X[j]
1	2.3331E-02	1.0040E+02	2.1631E+01	0.9766
2	6.3429E-01	3.6929E+00	2.1900E+01	0.9769
3	2.3424E+00	1.0000E+00	1.3339E+00	0.5003

Maximální číslo podmíněnosti K : 1.0040E+02
(K[j], K > 1000 indikuje silnou multikolinearitu)

Napověda-F1 Řádek: 64 - 86 Celkem: 263 Délka: 15522

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test H0: B[j] = 0 vs. HA: B[j] <> 0 t-kriterium	hypoteza H0 je	Hlad. výz.
B[0]	3.2022E+00	3.4618E+00	9.2502E-01	Akceptována	0.365
B[1]	9.6257E-01	2.4224E-01	3.9736E+00	Zamítnuta	0.001
B[2]	-2.6317E+00	3.9006E+00	-6.7469E-01	Akceptována	0.507
B[3]	-1.3048E-01	3.8853E+00	-3.3583E-02	Akceptována	0.974

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R : 9.5843E-01
Koeficient determinace, R^2 : 9.1859E-01
Predikovaný korelační koeficient, Rp^2 : 9.1326E-01
Střední kvadratická chyba predikce, MEP : 3.5791E+00
Akaikeho informační kritérium, AIC : 2.2072E+01

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

V Ý S L E D K Y

[j] korel. matice l[j] | ěnosti K[j] | factor VIF[j] | koef pro X[j]
1 2.3331E-02 | 1.0040E+02 | 2.1631E+01 | 0.9766
2 6.3429E-01 | 3.6929E+00 | 2.1900E+01 | 0.9769
3 2.3424E+00 | 1.0000E+00 | 1.3339E+00 | 0.5003
Maximální číslo podmíněnosti K : 1.0040E+02
(K[j], K > 1000 indikuje silnou multikolinearitu)
(VIF[j] > 10 indikuje silnou multikolinearitu)

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr | Odhad | Směrodatná odchylka | Test H0: B[j] = 0 vs. HA: B[j] <> 0 t-kriterium | hypoteza H0 je | Hlad. výz.
B[0] | 3.2022E+00 | 3.4618E+00 | 9.2502E-01 | Akceptována | 0.365
B[1] | 9.6257E-01 | 2.4224E-01 | 3.9736E+00 | Zamítnuta | 0.001
B[2] | -2.6317E+00 | 3.9006E+00 | -6.7469E-01 | Akceptována | 0.507
B[3] | -1.3048E-01 | 3.8853E+00 | -3.3583E-02 | Akceptována | 0.974

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:
Vícenásobný korelační koeficient, R : 9.5843E-01
Koeficient determinace, R^2 : 9.1859E-01
Predikovaný korelační koeficient, Rp^2 : 9.1326E-01
Střední kvadratická chyba predikce, MEP : 3.5791E+00
Akaikeho informační kritérium, AIC : 2.2072E+01

Napověda-F1 Řádek: 81 - 103 Celkem: 263 Délka: 15522

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

V Ý S L E D K Y

(6) TESTOVÁNÍ REGRESNÍHO TRIPLETU (DATA + MODEL + METODA):

Fisher-Snedocorův test významnosti regrese,F : 7.8984E+01
Tabulkový kvantil, F(1-alpha,m-1,n-m) : 3.0725E+00
Závěr: Navržený model je přijat jako významný.
Spočtená hladina významnosti : 0.000

Scottovo kriterium multikolinearity, M : 8.7168E-01
Závěr: Navržený model není korektní.

Cook-Weisbergův test heteroskedasticity, Sf : 1.0589E+02
Tabulkový kvantil, Chi^2(1-alpha,1) : 3.8415E+00
Závěr: Rezidua vykazují heteroskedasticitu.
Spočtená hladina významnosti : 0.000

Jarque-Berraův test normality reziduí, L(e) : 2.2290E-01
Tabulkový kvantil, Chi^2(1-alpha,2) : 5.9915E+00
Závěr: Normalita je přijata.
Spočtená hladina významnosti : 0.895

Waldův test autokorelace, Wa : 1.3061E+01
Tabulkový kvantil, Chi^2(1-alpha,1) : 3.8415E+00
Závěr: Rezidua jsou autokorelována.
Spočtená hladina významnosti : 0.000

Napověda-F1 Řádek: 145 - 167 Celkem: 263 Délka: 15522

Závěry o testu korelace

Otázka	Odpověď
Graf regresní křivky.	
Graf rezidua vs. predikce.	
R, D, s(e).	
Fisher-Snedecorův test celkové regrese	
Odhady parametrů přímky: úsek a směrnice	

Úlohy na korelační analýzu

U následujících úloh vyšetřete především:

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 B7.06 *Vliv počtu vypitých skleniček rumu na obsahu alkoholu v krvi*
Vyšetřete, zda obsah alkoholu v krvi x_2 koreluje s počtem vypitých skleniček rumu x_1. Jaká je příčinná souvislost s korelací dvou veličin? Vysvětlete párový korelační koeficient a uveďte rovněž jeho vlastnosti.
Data: Počet vypitých skleniček rumu x_1, obsah alkoholu v krvi x_2:

$x1$	$x2$
2	0.05
...	...
8	0.22

Úloha C7.04 *Vliv množství hnojiva na dosažený výnos*
Analýzou následujících dat vyšetřete, zda koreluje výnos pšenice x_1 [bušel/akr] s množstvím použitého hnojiva x_2 [libra/akr]. Vysvětlete párový korelační koeficient a uveďte rovněž jeho vlastnosti, 1 lb = 0.454 kg, 1 bušel = 36.37 l, 1 akr = 40 468 m^2
Data: Výnos pšenice x_1 [bušel/akr], množství hnojiva x_2 [libra/akr]:

$x1$	$x2$
100	40
...	...
700	80

Úloha E7.02 *Vliv nadmořské výšky na hektarový výnos ječmene*
Jsou dány údaje o 27 vybraných pozemcích, na nichž zemědělské závody pěstují v určité oblasti ozimý ječmen. Nadmořskou výšku pozemku v metrech označíme $x1$, hektarový výnos ječmene v t/ha $x2$. Nalezněte regresní model popisující vliv nadmořské výšky $x1$ na výnos ječmene $x2$. Jde o lineární vztah? Popište korelační koeficient dvou proměnných. Jaké vlastnosti musí vykazovat obě proměnné? Jak otestujeme statistickou významnost ještě existující korelace? K čemu je vhodné užít interval spolehlivosti korelačního koeficientu?
Data: Nadmořská výška pozemku $x1$ [m], hektarový výnos ječmene $x2$ [t/ha]

$x1$	$x2$
215	6.3
...	...
489	3.4