Analýza závislosti

Martina Litschmannová

VŠB TECHNICKÁ UNIVERZITA OSTRAVA | FAKULTA ELEKTROTECHNIKY A INFORMATIKY | KATEDRA APLIKOVANÉ MATEMATIKY

Obsah

- **Analýza závislosti dvou kategoriálních proměnných**
  - ✓ Analýza závislosti v kontingečních tabulkách
  - ✓ Analýza závislosti v asociačních tabulkách
  - ✓ Simpsonův paradox
- **Analýza závislosti dvou spojitých proměnných**
  - ✓ Pearsonův korelační koeficient,
  - ✓ Spearmanův korelační koeficient

Analýza závislosti

V praxi často u stat. jednotek zjišťujeme současně řadu znaků. Například

- spotřeba, objem motoru, hmotnost a zrychlení automobilů,
- výše mzdy, velikost IQ, hmotnost a výška mužů,
- školní prospěch a pocit deprese u dětí, apod.

Možnosti vyhodnocení:

- Analýza jednotlivých znaků (každý zvlášť)
- Analýza závislosti, tj. může zajímat, zda existuje závislost mezi
  - ✓ spotřebou automobilu a jeho hmotností,
  - ✓ výši mzdy a velikostí IQ,
  - ✓ pocitem deprese u dětí a školním prospěchem.

Metody analýzy jednostranné závislosti

- Jednostranná závislost - znak X působí na znak Y, avšak znak Y již nepůsobí zpětně na znak X.

		Typ znaku Y (důsledek)	
		kategoriální	kvantitativní
Typ znaku X (příčina)	kategoriální	analýza závislosti v kontingenčních, resp. v asociačních tabulkách	ANOVA
	kvantitativní	diskriminační analýza, logistická regrese…	korelační a regresní analýza

Toto není náplni základního kurzu Pravděpodobnost a statistika!

Analýza závislosti dvou kategoriálních proměnných

Analýza závislosti v kontingenčních tabulkách

Motivační příklad

Pro diferencovaný přístup v personální politice potřebuje vedení podniku vědět, zda spokojenost v práci závisí na tom, jedná-li se o pražský závod či závody mimopražské. Šetření se účastnilo 100 pracovníků z Prahy a 200 pracovníků z venkova. Výsledky šetření jsou v následující tabulce.

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	10	25	50	15
Venkov	20	10	130	40

Výsledky šetření analyzujte.

Motivační příklad

- **V jakém formátu obvykle získáváme tento typ dat?**

standardní datový formát

Místo	Stupeň spokojenosti
Praha	velmi spokojen
Praha	spíše spokojen
Venkov	spíše nespokojen
Praha	spíše spokojen
Venkov	velmi spokojen
⋮	⋮
Venkov	spíše spokojen

kontingenční tabulka

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	10	25	50	15
Venkov	20	10	130	40

- Převod dat ze standardního datového formátu do kontingenční tabulky umožňuje statistický software i většina tabulkových procesorů.

Co je to kontingenční tabulka?

- Dvourozměrná tabulka četností, z jejichž hodnot můžeme usoudit na závislost či nezávislost mezi dvěma kategoriálními proměnnými.

$X \backslash Y$	$y_{[1]}$	$y_{[2]}$	$\cdots$	$y_{[s]}$	Celkem
$x_{[1]}$	n_{11}	n_{12}	$\cdots$	n_{1s}	$n_{1.}$
$x_{[2]}$	n_{21}	n_{22}	$\cdots$	n_{2s}	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
$x_{[r]}$	n_{r1}	n_{r2}	$\cdots$	n_{rs}	$n_{r.}$
Celkem	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.s}$	$\boldsymbol{n}$

Schéma rozšířené kontingenční tabulky

- Rozšířená kontingenční tabulka obsahuje navíc tzv. marginální četnosti (sumární řádek a sloupec).

Jako doplňující informace se často uvádí:

- Sdružené relativní četnosti
- Řádkové relativní četnosti
- Sloupcové relativní četnosti

Doplňující informace ke kont. tabulce (motivační příklad)

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

sdružené relativní četnosti

	v. nespokojen	s. nespokojen	s. spokojen	v. spokojen
Praha	0.03333333	0.08333333	0.1666667	0.0500000
Venkov	0.06666667	0.03333333	0.4333333	0.1333333

řádkové relativní četnosti

	v. nespokojen	s. nespokojen	s. spokojen	v. spokojen
Praha	0.1	0.25	0.50	0.15
Venkov	0.1	0.05	0.65	0.20

sloupcové relativní četnosti

	v. nespokojen	s. nespokojen	s. spokojen	v. spokojen
Praha	0.3333333	0.7142857	0.2777778	0.2727273
Venkov	0.6666667	0.2857143	0.7222222	0.7272727

Jak posoudit intenzitu závislosti mezi kat. proměnnými?

Explorační (orientační) analýza

- Grafická analýza
  - ✓ shlukový sloupcový graf,
  - ✓ kumulativní sloupcový graf,
  - ✓ prostorový sloupcový graf (angl. sky chart),
  - ✓ mozaikový graf,
  - ✓ 100% skládaný pruhový graf

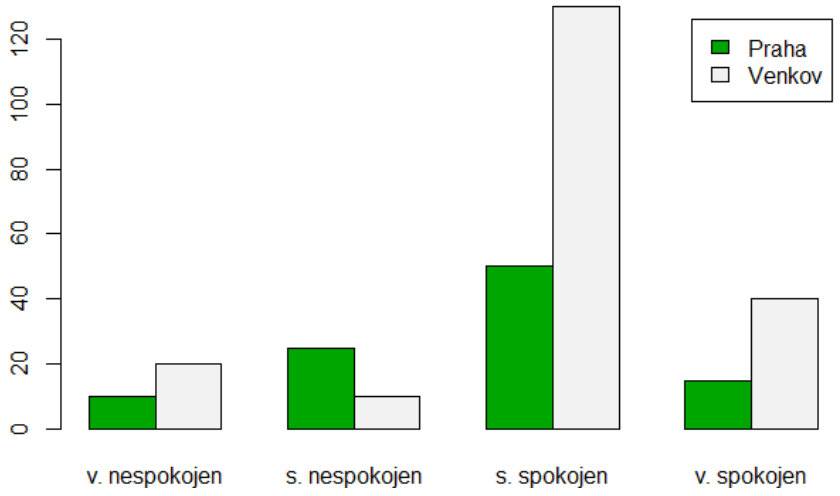
- Míry kontingence
  - ✓ koeficient kontingence (pouze pokud je počet variant obou proměnných je stejný)
  - ✓ korigovaný koeficient kontingence,
  - ✓ Cramerovo V

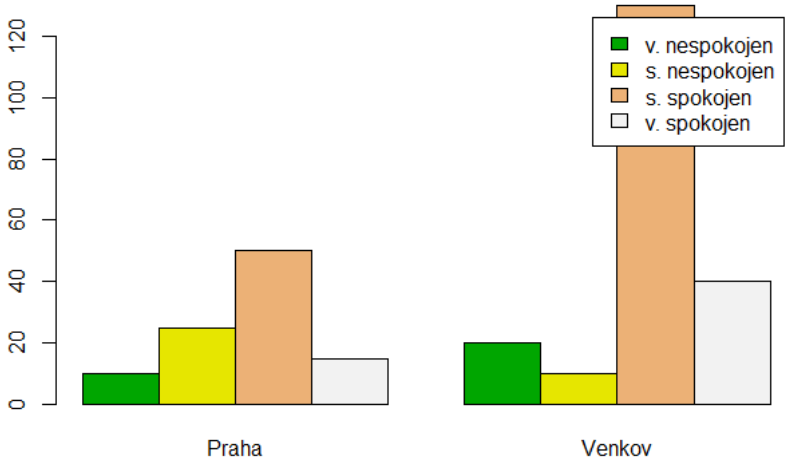
Čím jsou tyto koeficienty blíže 1, tím je závislost mezi proměnnými těsnější.

Grafická analýza

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Shlukový sloupcový graf



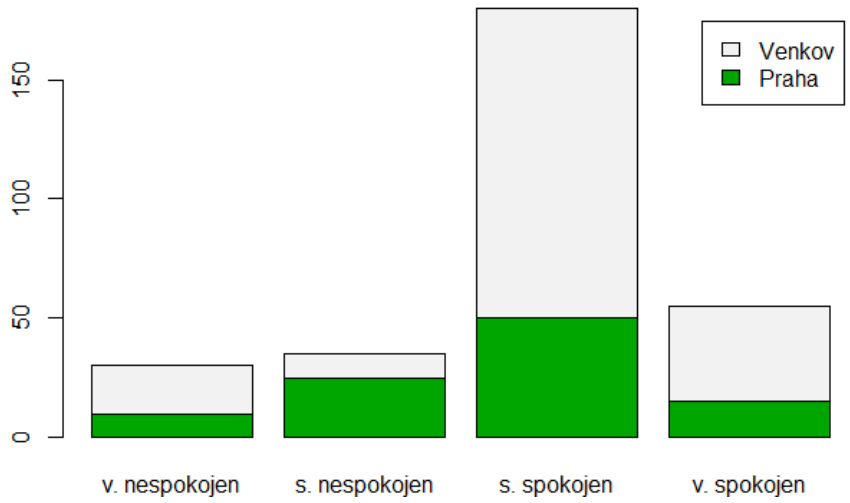


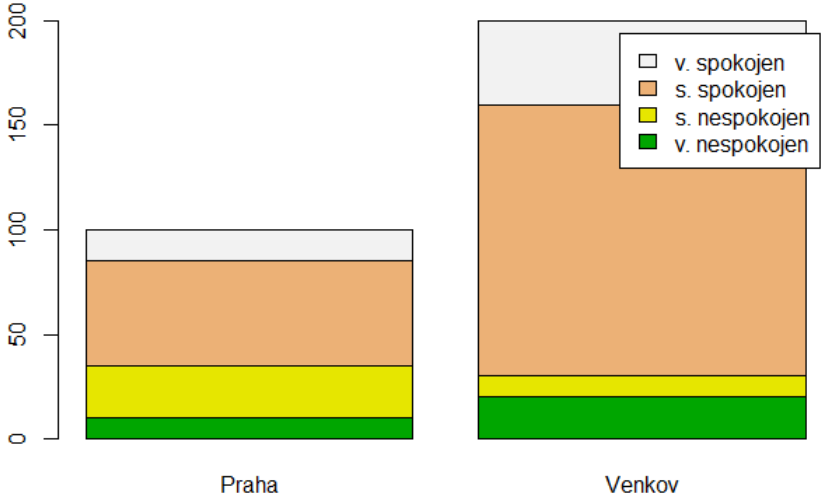
Který z grafů je pro naší analýzu vhodnější? Jak by se graf měl ještě upravit?

Grafická analýza

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Skládaný sloupcový graf



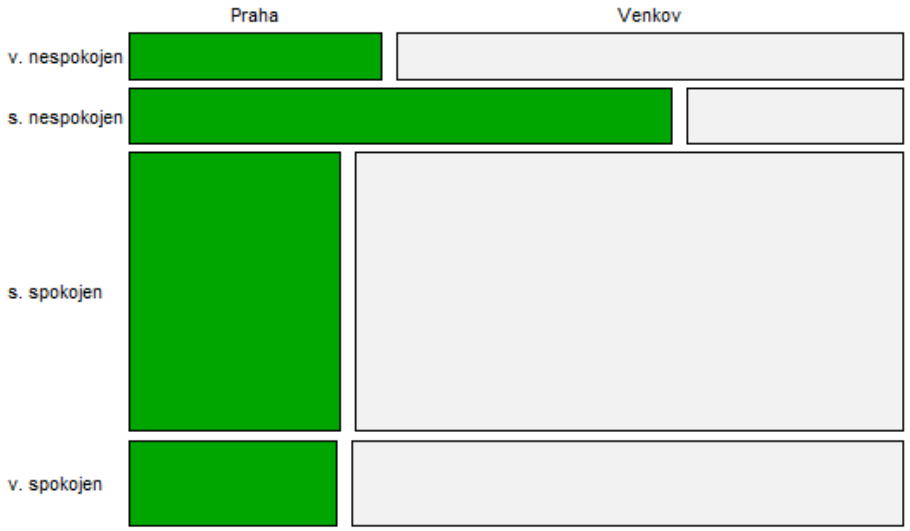


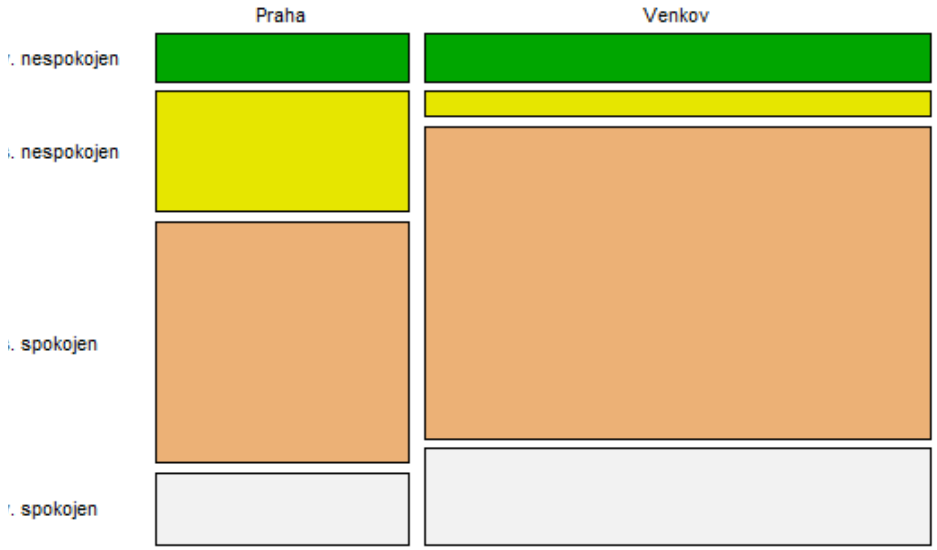
Který z grafů je pro naší analýzu vhodnější? Jak by se graf měl ještě upravit?

Grafická analýza

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Mozaikový graf

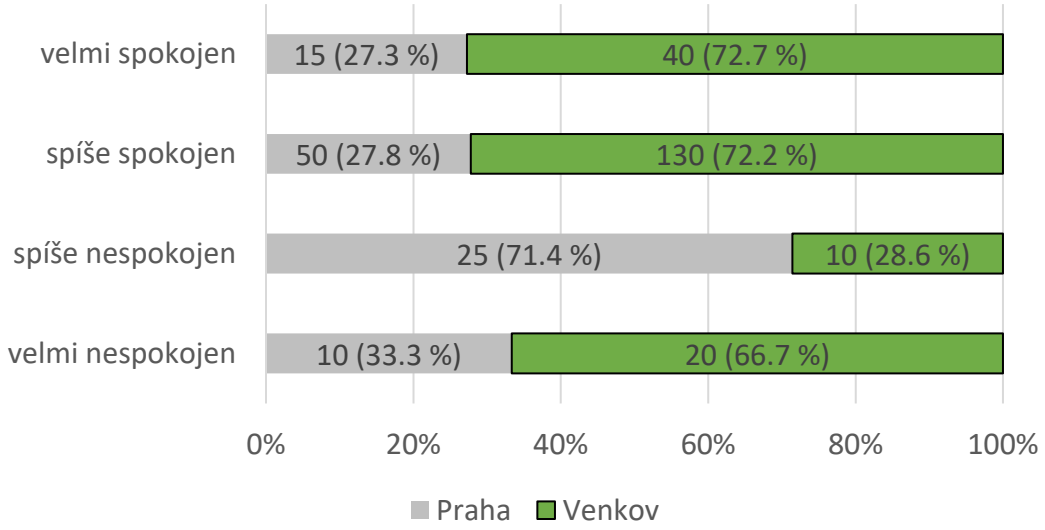




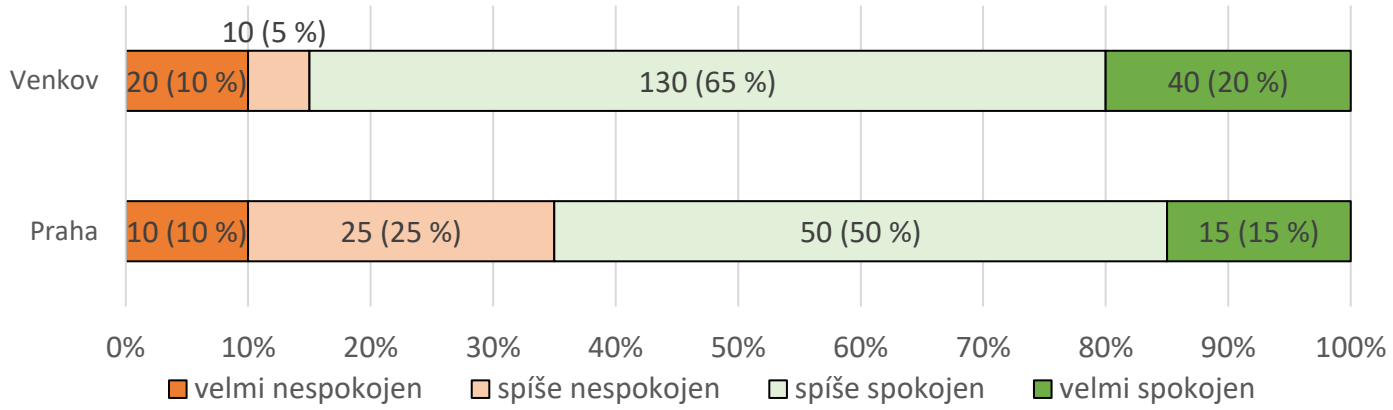
Který z grafů je pro naší analýzu vhodnější? Jak by se graf měl ještě upravit?

Grafická analýza

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**



100% skládaný řádkový graf



Který z grafů je pro naší analýzu vhodnější? Jak by se graf měl ještě upravit?

Míry kontingence

Označme:

- r… počet variant proměnné X, s… počet variant proměnné Y,
- $K = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$, kde O_{ij} jsou pozorované sdružené četnosti zapsané v kontingenční tabulce a E_{ij} jsou očekávané četnosti odpovídající součinu příslušných marginálních relativních četností.
- Koeficient kontingence: $CC = \sqrt{\frac{K}{K+n}}$ (POZOR! : $(r = s) \Rightarrow CC \in \langle 0; 1 \rangle$)
- Korigovaný koeficient kontingence: $CC_{cor} = \frac{CC}{CC_{max}}$, kde $CC_{max} = \sqrt{\frac{min(r;s)-1}{min(r;s)}}$
- Cramerovo V: $V = \sqrt{\frac{K}{n(min(r;s)-1)}}$

Motivační příklad

Pro diferencovaný přístup v personální politice potřebuje vedení podniku vědět, zda spokojenost v práci závisí na tom, jedná-li se o pražský závod či závody mimopražské. Šetření se účastnilo 100 pracovníků z Prahy a 200 pracovníků z venkova. Výsledky šetření jsou v následující tabulce.

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	10	25	50	15
Venkov	20	10	130	40

Výsledky šetření analyzujte.

Míry kontingence

Označme:

- r… počet variant proměnné X, s… počet variant proměnné Y,

- $K = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$, kde O_{ij} jsou pozorované sdružené četnosti zapsané v kontingenční tabulce a E_{ij} jsou očekávané četnosti odpovídající součinu příslušných marginálních relativních četností.

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha				
Venkov				

Očekávané četnosti

Míry kontingence

Označme:

- r… počet variant proměnné X, s… počet variant proměnné Y,
- $K = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$, kde O_{ij} jsou pozorované sdružené četnosti zapsané v kontingenční tabulce a E_{ij} jsou očekávané četnosti odpovídající součinu příslušných marginálních relativních četností.

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

- Jak odhadnout sdružené p-sti při daných marginálních četnostech za předpokladu, že analyzované proměnné jsou nezávislé?

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j) = \frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot j}}{n}$$

Míry kontingence

Označme:

- r… počet variant proměnné X, s… počet variant proměnné Y,
- $K = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, kde O_{ij} jsou pozorované sdružené četnosti zapsané v kontingenční tabulce a E_{ij} jsou očekávané četnosti odpovídající součinu příslušných marginálních relativních četností.

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

- Jak odhadnout sdružené p-sti při daných marginálních četnostech za předpokladu, že analyzované proměnné jsou nezávislé?

např.: $P(X = \text{"Praha"}, Y = \text{"spíše spokojen"}) = P(X = \text{"Praha"}) \cdot P(Y = \text{"spíše spokojen"}) = \frac{100}{300} \cdot \frac{180}{300}$

Míry kontingence

Označme:

- r… počet variant proměnné X, s… počet variant proměnné Y,
- $K = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$, kde O_{ij} jsou pozorované sdružené četnosti zapsané v kontingenční tabulce a E_{ij} jsou očekávané četnosti odpovídající součinu příslušných marginálních relativních četností.

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

- Jak odhadnout očekávané četnosti při daných marginálních četnostech za předpokladu, že analyzované proměnné jsou nezávislé?

$$E_{ij} = n \cdot P(X = x_i, Y = y_j) = n \cdot P(X = x_i) \cdot P(Y = y_j) = n \cdot \frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot j}}{n} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$$

Míry kontingence

Označme:

- r… počet variant proměnné X, s… počet variant proměnné Y,
- $K = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$, kde O_{ij} jsou pozorované sdružené četnosti zapsané v kontingenční tabulce a E_{ij} jsou očekávané četnosti odpovídající součinu příslušných marginálních relativních četností.

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

- Jak odhadnout očekávané četnosti při daných marginálních četnostech za předpokladu, že analyzované proměnné jsou nezávislé?

např.: $E_{\text{"Praha","spíše spokojen"}} = \frac{100 \cdot 180}{300} = 60$

Míry kontingence

Označme:

- r… počet variant proměnné X, s… počet variant proměnné Y,

- $K = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$, kde O_{ij} jsou pozorované sdružené četnosti zapsané v kontingenční tabulce a E_{ij} jsou očekávané četnosti odpovídající součinu příslušných marginálních relativních četností.

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha				
Venkov				

Očekávané četnosti

Míry kontingence

Označme:

- r… počet variant proměnné X, s… počet variant proměnné Y,

- $K = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$, kde O_{ij} jsou pozorované sdružené četnosti zapsané v kontingenční tabulce a E_{ij} jsou očekávané četnosti odpovídající součinu příslušných marginálních relativních četností.

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	$\frac{30 \cdot 100}{300}$	$\frac{35 \cdot 100}{300}$	$\frac{180 \cdot 100}{300}$	$\frac{55 \cdot 100}{300}$
Venkov	$\frac{30 \cdot 200}{300}$	$\frac{35 \cdot 200}{300}$	$\frac{180 \cdot 200}{300}$	$\frac{55 \cdot 200}{300}$

Očekávané četnosti

Míry kontingence

Označme:

- r… počet variant proměnné X, s… počet variant proměnné Y,

- $K = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$, kde O_{ij} jsou pozorované sdružené četnosti zapsané v kontingenční tabulce a E_{ij} jsou očekávané četnosti odpovídající součinu příslušných marginálních relativních četností.

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	10	11,7	60	18,3
Venkov	20	23,3	120	36,7

Očekávané četnosti

Míry kontingence

Označme:

- r… počet variant proměnné X, s… počet variant proměnné Y,
- $K = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$, kde O_{ij} jsou pozorované sdružené četnosti zapsané v kontingenční tabulce a E_{ij} jsou očekávané četnosti odpovídající součinu příslušných marginálních relativních četností.

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	10	25	50	15
Venkov	20	10	130	40

Kontingenční tabulka – pozorované četnosti

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	10	11,7	60	18,3
Venkov	20	23,3	120	36,7

Očekávané četnosti

- $K = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij}-E_{ij})^2}{E_{ij}} = \frac{(10-10)^2}{10} + \frac{(25-11{,}7)^2}{11{,}7} + \frac{(50-60)^2}{60} + \cdots + \frac{(40-36{,}7)^2}{36{,}7} \cong 26{,}27$

Míry kontingence

Označme:

- r… počet variant proměnné X, s… počet variant proměnné Y,
- $K = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$, kde O_{ij} jsou pozorované sdružené četnosti zapsané v kontingenční tabulce a E_{ij} jsou očekávané četnosti odpovídající součinu příslušných marginálních relativních četností.
- Tj. $r = 2, s = 4, n = 300, K = 26{,}27$
- Koeficient kontingence: $CC = \sqrt{\frac{K}{K+n}}$ – v tomto případě nepoužijeme ($r \neq s$)
- Korigovaný koeficient kontingence: $CC_{cor} = \frac{CC}{CC_{max}}$, kde $CC_{max} = \sqrt{\frac{min(r;s)-1}{min(r;s)}}$

$$CC = \sqrt{\frac{K}{K+n}} = \sqrt{\frac{26{,}27}{26{,}27+300}} \cong 0{,}284,\ CC_{max} = \sqrt{\frac{min(r;s)-1}{min(r;s)}} = \sqrt{\frac{2-1}{2}} \cong 0{,}707 \Rightarrow \boldsymbol{CC_{cor}} = \frac{CC}{CC_{max}} \cong \frac{0{,}284}{0{,}707} \cong \mathbf{0{,}402}$$

Míry kontingence

Označme:

- r… počet variant proměnné X, s… počet variant proměnné Y,
- $K = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$, kde O_{ij} jsou pozorované sdružené četnosti zapsané v kontingenční tabulce a E_{ij} jsou očekávané četnosti odpovídající součinu příslušných marginálních relativních četností.
- Tj. $r = 2, s = 4, n = 300, K = 26{,}27$
- Koeficient kontingence: $CC = \sqrt{\frac{K}{K+n}}$ – v tomto případě nepoužijeme ($r \neq s$)
- Korigovaný koeficient kontingence: $CC_{cor} = \frac{CC}{CC_{max}}$, kde $CC_{max} = \sqrt{\frac{min(r;s)-1}{min(r;s)}}$, tj. $CC_{cor} \cong 0{,}402$
- Cramerovo V: $V = \sqrt{\frac{K}{n(min(r;s)-1)}} = \sqrt{\frac{26{,}27}{300(2-1)}} = 0{,}296$

Motivační příklad

Pro diferencovaný přístup v personální politice potřebuje vedení podniku vědět, zda spokojenost v práci závisí na tom, jedná-li se o pražský závod či závody mimopražské. Šetření se účastnilo 100 pracovníků z Prahy a 200 pracovníků z venkova. Výsledky šetření jsou v následující tabulce.

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	10	25	50	15
Venkov	20	10	130	40

Výsledky šetření analyzujte.

Míry kontingence

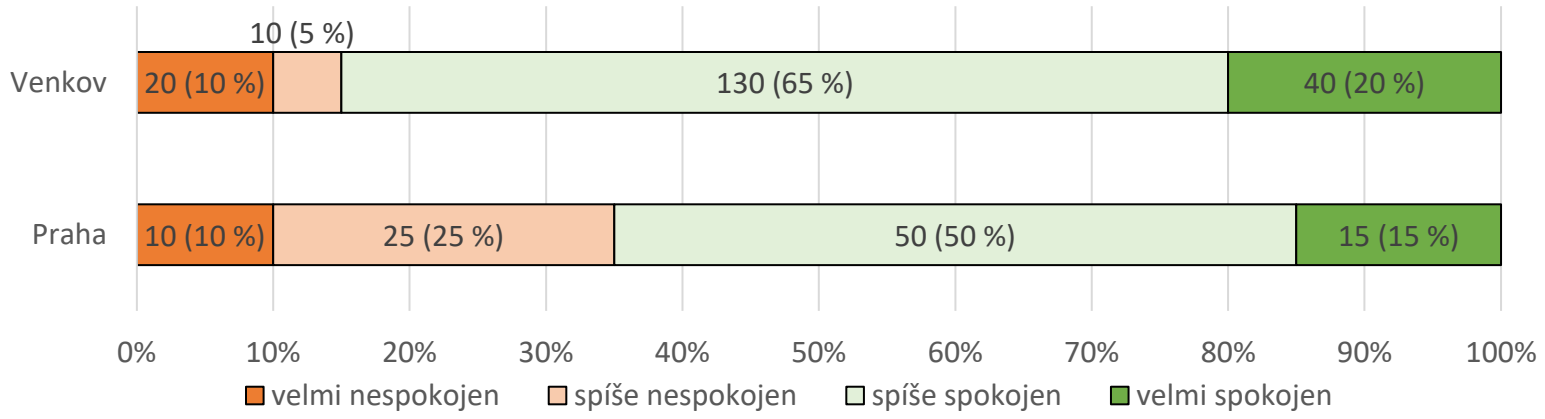
- Korigovaný koeficient kontingence: $0{,}402$, Cramerovo V: $0{,}296$

Dle uvedených měr se zdá, že závislost mezi spokojenosti zaměstnanců a místem zaměstnání je středně silná.

Motivační příklad – výstup explorační analýzy

Pro diferencovaný přístup v personální politice potřebuje vedení podniku vědět, zda spokojenost v práci závisí na tom, jedná-li se o pražský závod či závody mimopražské. Šetření se účastnilo 100 pracovníků z Prahy a 200 pracovníků z venkova. Výsledky šetření analyzujte.

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	celkem
Praha	10 (10 %)	25 (10 %)	50 (10 %)	15 (10 %)	**100**
Venkov	20 (10 %)	10 (5 %)	130 (65 %)	40 (20 %)	**200**
celkem	**30 (10 %)**	**35 (12 %)**	**180 (60 %)**	**55 (18 %)**	**300**



Na základě uvedeného grafu a Cramerova V (0,296) lze závislost mezi umístěním podniku a spokojenosti zaměstnanců hodnotit jako středně silnou.

Metody stat. indukce vhodné pro analýzu závislosti kat. veličin

- **Intervalové odhady vybraných pravděpodobností**

(viz Úvod do statistiky, kap. 4)

Frequency Table

	Velmi nespok	Spíše nespok	Spíše spokoj	Velmi spokoj	Row Total
Praha	10 3,33% 10,00% 33,33%	25 8,33% 25,00% 71,43%	50 16,67% 50,00% 27,78%	15 5,00% 15,00% 27,27%	100 33,33%
Venkov	20 6,67% 10,00% 66,67%	10 3,33% 5,00% 28,57%	130 43,33% 65,00% 72,22%	40 13,33% 20,00% 72,73%	200 66,67%
Column Total	30 10,00%	35 11,67%	180 60,00%	55 18,33%	300 100,00%

Cell contents:
Observed frequency
Percentage of table
Percentage of row
Percentage of column

A to musím počítat intervalové odhady pro „všechny“ pravděpodobnosti, které jsou v té tabulce???

Metody stat. indukce vhodné pro analýzu závislosti kat. veličin

- **Intervalové odhady vybraných pravděpodobností**

(viz Úvod do statistiky, kap. 4)

Frequency Table

	Velmi nespok	Spíše nespok	Spíše spokoj	Velmi spokoj	Row Total
Praha	10 3,33% 10,00% 33,33%	25 8,33% 25,00% 71,43%	50 16,67% 50,00% 27,78%	15 5,00% 15,00% 27,27%	100 33,33%
Venkov	20 6,67% 10,00% 66,67%	10 3,33% 5,00% 28,57%	130 43,33% 65,00% 72,22%	40 13,33% 20,00% 72,73%	200 66,67%
Column Total	30 10,00%	35 11,67%	180 60,00%	55 18,33%	300 100,00%

Cell contents:
Observed frequency
Percentage of table
Percentage of row
Percentage of column

Metody stat. indukce vhodné pro analýzu závislosti kat. veličin

- **χ^2 test nezávislosti v kontingenční tabulce**

H_0: Znaky *X* a *Y* v kontingenční tabulce jsou statisticky **nezávislé.**

H_A: Znaky *X* a *Y* v kontingenční tabulce jsou statisticky závislé.

Předpoklady testu:

- žádná z očekávaných četností E_{ij} nesmí být menší než 2,
- alespoň 80 % očekávaných četností E_{ij} musí být větších než 5.

Testové kritérium: $K = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

p-hodnota: $p\text{–}hodnota = 1 - F_0(x_{OBS})$,

kde $F_0(x)$ je distribuční funkce χ^2 rozdělení s $(r-1) \cdot (s-1)$ stupni volnosti.

Motivační příklad

Pro diferencovaný přístup v personální politice potřebuje vedení podniku vědět, zda spokojenost v práci závisí na tom, jedná-li se o pražský závod či závody mimopražské. Šetření se účastnilo 100 pracovníků z Prahy a 200 pracovníků z venkova. Výsledky šetření jsou v následující tabulce.

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	10	25	50	15
Venkov	20	10	130	40

Výsledky šetření analyzujte.

H_0: Spokojenost v práci **nesouvisí** s umístěním závodu.

H_A: Spokojenost v práci souvisí s umístěním závodu.

Motivační příklad

H_0: Spokojenost v práci **nesouvisí** s umístěním závodu.

H_A: Spokojenost v práci souvisí s umístěním závodu.

- **Ověření předpokladů**

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	10	11,7	60	18,3
Venkov	20	23,3	120	36,7

Očekávané četnosti

Všechny očekávané četnosti jsou větší než 5.

Motivační příklad

H_0: Spokojenost v práci **nesouvisí** s umístěním závodu.

H_A: Spokojenost v práci souvisí s umístěním závodu.

- **Ověření předpokladů**

A co když předpoklady splněny nebudou???

Pokud lze některé varianty proměnné „smysluplně“ sloučit, zkus to udělat. Pokud ne, nelze výsledky z výběrového šetření zobecnit na populaci. Na tento možný problém je vhodné myslet již před výběrovým šetřením (dostatečný rozsah výběru).

Motivační příklad

H_0: Spokojenost v práci **nesouvisí** s umístěním závodu.

H_A: Spokojenost v práci souvisí s umístěním závodu.

- **Pozorovaná hodnota**

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	10	11,7	60	18,3
Venkov	20	23,3	120	36,7

Očekávané četnosti

$$x_{OBS} = K = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}} = \frac{(10-10)^2}{10} + \frac{(25-11{,}7)^2}{11{,}7} + \frac{(50-60)^2}{60} + \cdots + \frac{(40-36{,}7)^2}{36{,}7} \cong 26{,}27$$

Motivační příklad

H_0: Spokojenost v práci **nesouvisí** s umístěním závodu.

H_A: Spokojenost v práci souvisí s umístěním závodu.

- **p-hodnota**

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	10	11,7	60	18,3
Venkov	20	23,3	120	36,7

Očekávané četnosti

$p - hodnota = 1 - F_0(x_{OBS})$, kde $F_0(x)$ je distribuční funkce χ^2 rozdělení s $(r-1)\cdot(s-1)$ stupni volnosti.

Motivační příklad

H_0: Spokojenost v práci **nesouvisí** s umístěním závodu.

H_A: Spokojenost v práci souvisí s umístěním závodu.

- **p-hodnota**

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	10	11,7	60	18,3
Venkov	20	23,3	120	36,7

Očekávané četnosti

$p - hodnota = 1 - F_0(26{,}27)$, kde $F_0(x)$ je distribuční funkce χ^2 rozdělení s $(2-1)\cdot(4-1)$ stupni volnosti.

Motivační příklad

H_0: Spokojenost v práci **nesouvisí** s umístěním závodu.

H_A: Spokojenost v práci souvisí s umístěním závodu.

- **p-hodnota**

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	10	11,7	60	18,3
Venkov	20	23,3	120	36,7

Očekávané četnosti

$p - hodnota = 1 - F_0(26{,}27)$, kde $F_0(x)$ je distribuční funkce χ^2 rozdělení s 3 stupni volnosti.

Motivační příklad

H_0: Spokojenost v práci **nesouvisí** s umístěním závodu.

H_A: Spokojenost v práci souvisí s umístěním závodu.

- **p-hodnota**

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	10	11,7	60	18,3
Venkov	20	23,3	120	36,7

Očekávané četnosti

$p - hodnota \ll 0{,}001$ (*1-pchisq(26.27,3*))

Motivační příklad

H_0: Spokojenost v práci **nesouvisí** s umístěním závodu.

H_A: Spokojenost v práci souvisí s umístěním závodu.

- **Rozhodnutí**

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	**celkem**
Praha	10	25	50	15	**100**
Venkov	20	10	130	40	**200**
celkem	**30**	**35**	**180**	**55**	**300**

Schéma rozšířené kontingenční tabulky

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen
Praha	10	11,7	60	18,3
Venkov	20	23,3	120	36,7

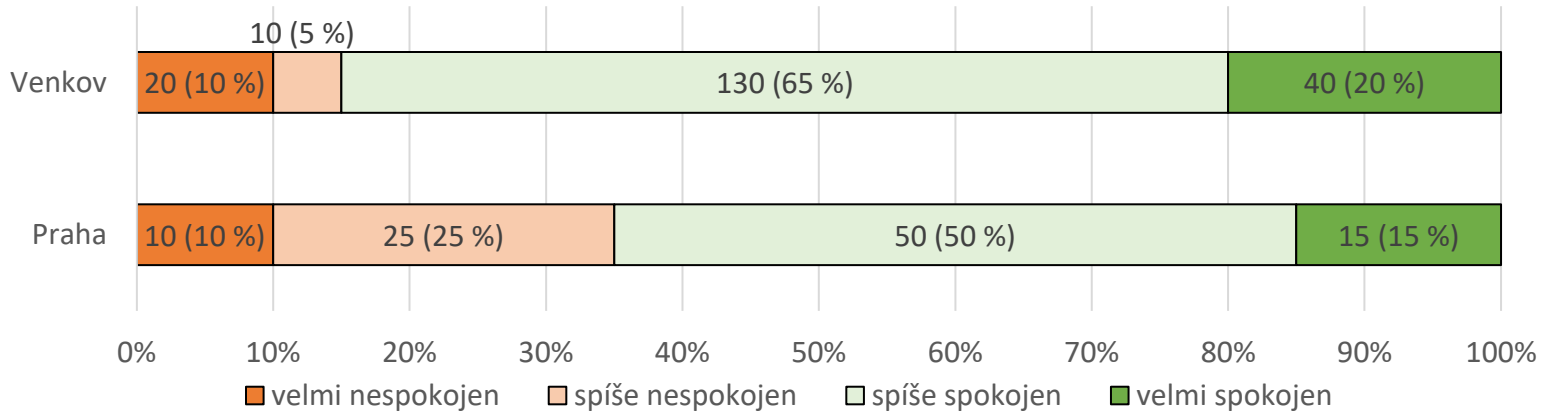
Očekávané četnosti

Na hladině významnosti 0,05 zamítáme nulovou hypotézu (χ^2 test nezávislosti, x_{OBS} = 26,27, df = 3, $p - hodnota \ll 0{,}001$), tj. spokojenost v práci statisticky významně souvisí s umístěním závodu.

Motivační příklad – celkový výstup analýzy

Pro diferencovaný přístup v personální politice potřebuje vedení podniku vědět, zda spokojenost v práci závisí na tom, jedná-li se o pražský závod či závody mimopražské. Šetření se účastnilo 100 pracovníků z Prahy a 200 pracovníků z venkova. Výsledky šetření analyzujte.

místo/stupeň spokojenosti	velmi nespokojen	spíše nespokojen	spíše spokojen	velmi spokojen	celkem
Praha	10 (10 %)	25 (10 %)	50 (10 %)	15 (10 %)	**100**
Venkov	20 (10 %)	10 (5 %)	130 (65 %)	40 (20 %)	**200**
celkem	**30 (10 %)**	**35 (12 %)**	**180 (60 %)**	**55 (18 %)**	**300**



Spokojenost v práci statisticky významně souvisí s umístěním závodu (χ^2 test nezávislosti, x_{OBS} = 26,27, df = 3, $p - hodnota \ll 0{,}001$). Na základě uvedeného grafu a Cramerova V (0,296) lze závislost mezi umístěním podniku a spokojenosti zaměstnanců hodnotit jako středně silnou.

Analýza závislosti dvou kategoriálních proměnných

Analýza závislosti v asociačních tabulkách

Co je to asociační tabulka?

- Asociační tabulka je speciální typ kontingenčních tabulek, které používáme k sledování závislosti dvou dichotomických znaků, tj. kategoriálních znaků nabývajících pouze dvou variant.
- asociace = vztah dvou dichotomických znaků

$X\ (okolnosti) \backslash Y (výskyt\ události)$	$y_{[1]}$ (úspěch)	$y_{[2]}$ (neúspěch)	Celkem
$x_{[1]}$ (I.)	a	b	$a+b$
$x_{[2]}$ (II.)	c	d	$c+d$
Celkem	$a+c$	$b+d$	n

Schéma rozšířené asociační tabulky

$X\ (sledovaný\ faktor) \backslash Y (výskyt\ onemocnění)$	D (ANO)	$\bar{D}$ (NE)	Celkem
E (přítomnost faktoru)	a	b	$a+b$
$\bar{E}$ (nepřítomnost faktoru)	c	d	$c+d$
Celkem	$a+c$	$b+d$	n

Schéma rozšířené asociační tabulky (biomedicínská aplikace)

- Na asociační tabulku lze sice nahlížet jako na speciální případ kontingenčních tabulek a při analýze používat jejich aparát, nicméně vhodnější je využít specifické metody a charakteristiky asociace.

Popisné statistiky pro asociační tabulky

$X\ (okolnosti) \backslash Y (výskyt\ události)$	$y_{[1]}$ (úspěch)	$y_{[2]}$ (neúspěch)	Celkem
$x_{[1]}$ (I.)	a	b	$a+b$
$x_{[2]}$ (II.)	c	d	$c+d$
Celkem	$a+c$	$b+d$	n

Schéma rozšířené asociační tabulky

- **Riziko** (pravděpodobnost) **výskytu události**
  - ✓ $\hat{R}_I = \frac{a}{a+b}$ (p-st, že nastane úspěch v případě okolností I)
  - ✓ $\hat{R}_{II} = \frac{c}{c+d}$ (p-st, že nastane úspěch v případě okolností II)

- **Šance na výskyt události**
  - ✓ $\hat{O}_I = \frac{a}{b}$ (šance, že nastane úspěch v případě okolností I)
  - ✓ $\hat{O}_{II} = \frac{c}{d}$ (šance, že nastane úspěch v případě okolností II)

***Poznámka**: šance = poměr počtu úspěchů k počtu neúspěchu*

Popisné statistiky pro asociační tabulky

exponovaná populace ⇨

$X\ (okolnosti) \backslash Y (výskyt\ události)$	$y_{[1]}$ (úspěch)	$y_{[2]}$ (neúspěch)	Celkem
$x_{[1]}$ (I.)	a	b	$a+b$
$x_{[2]}$ (II.)	c	d	$c+d$
Celkem	$a+c$	$b+d$	n

Schéma rozšířené asociační tabulky

Míry asociace

- **Relativní riziko** (pravděpodobnost) **výskytu události** (angl. Relative Risk)
  - ✓ $\widehat{RR} = \frac{\hat{R}_I}{\hat{R}_{II}} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{a(c+d)}{c(a+b)}$
  - ✓ Rel. riziko udává kolikrát je větší riziko výskytu události u tzv. exponované populace, než u populace neexponované.
  - ✓ $100(1-\alpha)\%$ intervalový odhad RR: $\langle \widehat{RR} \cdot e^{-\sqrt{\frac{b}{a(a+b)}+\frac{d}{c(c+d)}} \cdot z_{1-\frac{\alpha}{2}}}; \widehat{RR} \cdot e^{\sqrt{\frac{b}{a(a+b)}+\frac{d}{c(c+d)}} \cdot z_{1-\frac{\alpha}{2}}} \rangle$

(Katzova metoda)

Popisné statistiky pro asociační tabulky

exponovaná populace

$X\ (okolnosti) \backslash Y (výskyt\ události)$	$y_{[1]}$ (úspěch)	$y_{[2]}$ (neúspěch)	Celkem
$x_{[1]}$ (I.)	a	b	$a+b$
$x_{[2]}$ (II.)	c	d	$c+d$
Celkem	$a+c$	$b+d$	n

Schéma rozšířené asociační tabulky

Míry asociace

- **Poměr šancí na výskyt události** (angl. Odds Ratio)
  - ✓ $\widehat{OR} = \frac{\hat{O}_I}{\hat{O}_{II}} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$
  - ✓ Poměr šancí udává kolikrát je větší šance na výskyt události u tzv. exponované populace, než u populace neexponované.
  - ✓ $100(1-\alpha)\%$ intervalový odhad OR: $\langle \widehat{OR} \cdot e^{-\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} \cdot z_{1-\frac{\alpha}{2}}}; \widehat{OR} \cdot e^{\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} \cdot z_{1-\frac{\alpha}{2}}} \rangle$

(Woolfova metoda)

Motivační příklad pro asociační tabulky

Závisí novorozenecká úmrtnost (do 7 dnů po porodu) na porodní váze? Data odpovídající situaci v New Yorku v roce 1974 jsou uvedena v následující tabulce.

exp. populace, tj. populace s vyšším rizikem události ⟹

porodní váha\novorozenecká úmrtí	ANO	NE	Celkem
nízká	618	4 597	5 215
normální	422	67 093	67 515
Celkem	1 040	71 690	72 730

- **Odhad rizika novorozeneckého úmrtí u obou populací**
  - ✓ $\hat{R}_{nízká} = \frac{618}{5\,215} \cong 0{,}118$, 95% intervalový odhad: $\langle 0{,}109; 0{,}127 \rangle$ (*binom.test(618,5215)*)
    U novorozenců s nízkou porodní hmotností je riziko novorozeneckého úmrtí cca 11,8 %. Dle Clopperova-Pearsonova 95% odhadu očekáváme toto riziko v intervalu $\langle 10{,}9\ \%; 12{,}7\ \% \rangle$.
  - ✓ $\hat{R}_{normální} = \frac{422}{67\,515} \cong 0{,}006$, 95% intervalový odhad: $\langle 0{,}005; 0{,}007 \rangle$ (*binom.test(422,67 515)*)
    U novorozenců s normální porodní hmotností je riziko novorozeneckého úmrtí cca 0,6 %. Dle Clopperova-Pearsonova 95% odhadu očekáváme toto riziko v intervalu $\langle 0{,}5\ \%; 0{,}7\ \% \rangle$.

Motivační příklad pro asociační tabulky

Závisí novorozenecká úmrtnost (do 7 dnů po porodu) na porodní váze? Data odpovídající situaci v New Yorku v roce 1974 jsou uvedena v následující tabulce.

exp. populace, tj. populace s vyšším rizikem události ⟹

porodní váha\novorozenecká úmrtí	ANO	NE	Celkem
nízká	618	4 597	5 215
normální	422	67 093	67 515
Celkem	1 040	71 690	72 730

- **Odhad relativního rizika novorozeneckého úmrtí u novorozenců s nízkou porodní hmotností (u exponované populace)**

✓ $\widehat{RR} = \frac{\hat{R}_{nízká}}{\hat{R}_{normální}} \cong \frac{0{,}118}{0{,}006} \cong 19{,}7$; 95% intervalový odhad: $\langle \widehat{RR} \cdot e^{-\sqrt{\frac{b}{a(a+b)}+\frac{d}{c(c+d)}} \cdot z_{1-\frac{\alpha}{2}}}; \widehat{RR} \cdot e^{\sqrt{\frac{b}{a(a+b)}+\frac{d}{c(c+d)}} \cdot z_{1-\frac{\alpha}{2}}} \rangle$

Motivační příklad pro asociační tabulky

Závisí novorozenecká úmrtnost (do 7 dnů po porodu) na porodní váze? Data odpovídající situaci v New Yorku v roce 1974 jsou uvedena v následující tabulce.

exp. populace, tj. populace s vyšším rizikem události ⟹

porodní váha\novorozenecká úmrtí	ANO	NE	Celkem
nízká	618	4 597	5 215
normální	422	67 093	67 515
Celkem	1 040	71 690	72 730

- **Odhad relativního rizika novorozeneckého úmrtí u novorozenců s nízkou porodní hmotností (u exponované populace)**
  - ✓ $\widehat{RR} = \frac{\hat{R}_{nízká}}{\hat{R}_{normální}} \cong \frac{0{,}118}{0{,}006} \cong 19{,}7$; 95% intervalový odhad: $\langle 17{,}1; 21{,}0 \rangle$

- Jestliže $100(1-\alpha)\%$ intervalový odhad RR nezahrnuje 1, pak na hladině významnosti α zamítáme hypotézu o nezávislosti znaků X a Y.

Motivační příklad pro asociační tabulky

Závisí novorozenecká úmrtnost (do 7 dnů po porodu) na porodní váze? Data odpovídající situaci v New Yorku v roce 1974 jsou uvedena v následující tabulce.

exp. populace, tj. populace s vyšším rizikem události ⟹

porodní váha\novorozenecká úmrtí	ANO	NE	Celkem
nízká	618	4 597	5 215
normální	422	67 093	67 515
Celkem	1 040	71 690	72 730

- **Odhad relativního rizika novorozeneckého úmrtí u novorozenců s nízkou porodní hmotností (u exponované populace)**
  - ✓ $\widehat{RR} = \frac{\hat{R}_{nízká}}{\hat{R}_{normální}} \cong \frac{0{,}118}{0{,}006} \cong 19{,}7$; 95% intervalový odhad: $\langle 17{,}1; 21{,}0 \rangle$

$RR < 1$	Expozice snižuje riziko onemocnění.
$RR = 1$	Mezi expozici a onemocněním neexistuje žádná asociace.
$RR > 1$	Expozice zvyšuje riziko onemocnění.

Motivační příklad pro asociační tabulky

Závisí novorozenecká úmrtnost (do 7 dnů po porodu) na porodní váze? Data odpovídající situaci v New Yorku v roce 1974 jsou uvedena v následující tabulce.

exp. populace, tj. populace s vyšším rizikem události ⟹

porodní váha\novorozenecká úmrtí	ANO	NE	Celkem
nízká	618	4 597	5 215
normální	422	67 093	67 515
Celkem	1 040	71 690	72 730

- **Odhad relativního rizika novorozeneckého úmrtí u novorozenců s nízkou porodní hmotností (u exponované populace)**

  ✓ $\widehat{RR} \cong 19{,}7$; 95% intervalový odhad: $\langle 17{,}1; 21{,}0 \rangle$

  U novorozenců s nízkou porodní hmotností je riziko novorozeneckého úmrtí cca 19,7 krát vyšší riziko novorozeneckého úmrtí než u novorozenců s normální porodní hmotností. 95% intervalový odhad tohoto relativního rizika (Katzova metoda) je $\langle 17{,}1; 21{,}0 \rangle$, tj. souvislost mezi porodní hmotností a novorozeneckým úmrtím je statisticky významná.

Motivační příklad pro asociační tabulky

Závisí novorozenecká úmrtnost (do 7 dnů po porodu) na porodní váze? Data odpovídající situaci v New Yorku v roce 1974 jsou uvedena v následující tabulce.

exp. populace, tj. populace s vyšším rizikem události ⟹

porodní váha\novorozenecká úmrtí	ANO	NE	Celkem
nízká	618	4 597	5 215
normální	422	67 093	67 515
Celkem	1 040	71 690	72 730

- **Odhad šance na novorozenecké úmrtí u obou populací**
  - ✓ $\hat{O}_{nízká} = \frac{618}{4\,597} \cong 0{,}134 = \frac{134}{1\,000}$

    U novorozenců s nízkou porodní hmotností je šance na novorozenecké úmrtí 134:1 000, tj. na 1 000 přeživších novorozenců s nízkou porodní hmotností připadá cca 134 zemřelých.
  - ✓ $\hat{O}_{normální} = \frac{422}{67\,093} \cong 0{,}006 = \frac{6}{1\,000}$

    U novorozenců s normální porodní hmotností je šance na novorozenecké úmrtí 6:1 000, tj. na 1 000 přeživších novorozenců s normální porodní hmotností připadá cca 6 zemřelých.

Motivační příklad pro asociační tabulky

Závisí novorozenecká úmrtnost (do 7 dnů po porodu) na porodní váze? Data odpovídající situaci v New Yorku v roce 1974 jsou uvedena v následující tabulce.

exp. populace, tj. populace s vyšším rizikem události ⟹

porodní váha\novorozenecká úmrtí	ANO	NE	Celkem
nízká	618	4 597	5 215
normální	422	67 093	67 515
Celkem	1 040	71 690	72 730

- **Odhad poměru šancí na novorozenecké úmrtí u novorozenců s nízkou porodní hmotností (u exponované populace)**

✓ $\widehat{OR} = \frac{\hat{O}_{nízká}}{\hat{O}_{normální}} \cong \frac{0{,}134}{0{,}006} \cong 22{,}3$; 95% intervalový odhad: $\langle \widehat{OR} \cdot e^{-\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} \cdot z_{1-\frac{\alpha}{2}}}; \widehat{OR} \cdot e^{\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} \cdot z_{1-\frac{\alpha}{2}}} \rangle$

Motivační příklad pro asociační tabulky

Závisí novorozenecká úmrtnost (do 7 dnů po porodu) na porodní váze? Data odpovídající situaci v New Yorku v roce 1974 jsou uvedena v následující tabulce.

exp. populace, tj. populace s vyšším rizikem události ⟹

porodní váha\novorozenecká úmrtí	ANO	NE	Celkem
nízká	618	4 597	5 215
normální	422	67 093	67 515
Celkem	1 040	71 690	72 730

- **Odhad relativního rizika novorozeneckého úmrtí u novorozenců s nízkou porodní hmotností (u exponované populace)**
  - ✓ $\widehat{OR} = \frac{\hat{O}_{nízká}}{\hat{O}_{normální}} \cong \frac{0{,}134}{0{,}006} \cong$22,3; 95% intervalový odhad: $\langle 19{,}2; 23{,}8 \rangle$

- Jestliže $100(1-\alpha)\%$ intervalový odhad OR nezahrnuje 1, pak na hladině významnosti α zamítáme hypotézu o nezávislosti znaků X a Y.

Motivační příklad pro asociační tabulky

Závisí novorozenecká úmrtnost (do 7 dnů po porodu) na porodní váze? Data odpovídající situaci v New Yorku v roce 1974 jsou uvedena v následující tabulce.

exp. populace, tj. populace s vyšším rizikem události ⟹

porodní váha\novorozenecká úmrtí	ANO	NE	Celkem
nízká	618	4 597	5 215
normální	422	67 093	67 515
Celkem	1 040	71 690	72 730

- **Odhad relativního rizika novorozeneckého úmrtí u novorozenců s nízkou porodní hmotností (u exponované populace)**
  - ✓ $\widehat{OR} = \frac{\hat{O}_{nízká}}{\hat{O}_{normální}} \cong \frac{0{,}134}{0{,}006} \cong 22{,}3$; 95% intervalový odhad: $\langle 19{,}2; 23{,}8 \rangle$

$0R < 1$	U exponované populace (populace vystavené sledovanému faktoru) je nižší šance na výskyt nemoci.
$OR = 1$	Šance výskytu onemocnění u exponované a neexponované populace jsou shodné.
$OR > 1$	U exponované populace je vyšší šance na výskyt nemoci.

Motivační příklad pro asociační tabulky

Závisí novorozenecká úmrtnost (do 7 dnů po porodu) na porodní váze? Data odpovídající situaci v New Yorku v roce 1974 jsou uvedena v následující tabulce.

exp. populace, tj. populace s vyšším rizikem události ⟹

porodní váha\novorozenecká úmrtí	ANO	NE	Celkem
nízká	618	4 597	5 215
normální	422	67 093	67 515
Celkem	1 040	71 690	72 730

- **Odhad relativního rizika novorozeneckého úmrtí u novorozenců s nízkou porodní hmotností (u exponované populace)**

  ✓ $\widehat{OR} \cong 22{,}3$; 95% intervalový odhad: $\langle 19{,}2; 23{,}8 \rangle$

  U novorozenců s nízkou porodní hmotností je šance na novorozeneckého úmrtí cca 22,3 krát vyšší než šance na novorozenecké úmrtí než u novorozenců s normální porodní hmotností. 95% intervalový odhad tohoto poměru šancí (Woolfova metoda) je $\langle 19{,}2; 23{,}8 \rangle$, tj. souvislost mezi porodní hmotností a novorozeneckým úmrtím je statisticky významná.

Simpsonův paradox

Motivační příklad – Simpsonův paradox

V Horních Sádrovicích bylo hospitalizováno 600 „lehkých“ pacientů, z nichž 10 (1,7 %) zemřelo a 400 „těžkých“ pacientů, z nichž zemřelo 190 (47,5 %). Ve Staré Dláze bylo hospitalizováno 900 „lehkých“ pacientů, z nichž 30 (3,2 %) zemřelo a 100 „těžkých“ pacientů, z nichž zemřelo 100 (10,0 %).

Horní Sádrovice			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	10 0,017 (10/600)	590 0,983 (590/600)	**600**
těžký	190 0,475 (190/400)	210 0,525 (210/400)	**400**
celkem	**200** **0,200 (200/1000)**	**800** **0,800 (800/1000)**	**1 000**

Motivační příklad – Simpsonův paradox

Horní Sádrovice			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	10 0,017 (10/600)	590 0,983 (590/600)	**600**
těžký	190 0,475 (190/400)	210 0,525 (210/400)	**400**
celkem	**200** **0,200 (200/1000)**	**800** **0,800 (800/1000)**	**1 000**

Stará Dláha			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	30 0,033 (30/900)	870 0,967 (870/900)	**900**
těžký	70 0, 700 (70/100)	30 0,300 (30/100)	**100**
celkem	**100** **0, 100 (100/1000)**	**900** **0,900 (900/1000)**	**1 000**

Kontingenční tabulky rozšířené o marginální četnosti a řádkové rel. četnosti

Motivační příklad – Simpsonův paradox

Horní Sádrovice			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	10 0,017 (10/600)	590 0,983 (590/600)	**600**
těžký	190 0,475 (190/400)	210 0,525 (210/400)	**400**
celkem	**200** **0,200 (200/1000)**	**800** **0,800 (800/1000)**	**1 000**

Stará Dláha			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	30 0,033 (30/900)	870 0,967 (870/900)	**900**
těžký	70 0, 700 (70/100)	30 0,300 (30/100)	**100**
celkem	**100** **0, 100 (100/1000)**	**900** **0,900 (900/1000)**	**1 000**

Ve kterém městě je u lehkých pacientů nižší riziko úmrtí?

Motivační příklad – Simpsonův paradox

Horní Sádrovice			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	10 **0,017** (10/600)	590 0,983 (590/600)	**600**
těžký	190 0,475 (190/400)	210 0,525 (210/400)	**400**
celkem	**200** **0,200 (200/1000)**	**800** **0,800 (800/1000)**	**1 000**

Stará Dláha			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	30 **0,033** (30/900)	870 0,967 (870/900)	**900**
těžký	70 0, 700 (70/100)	30 0,300 (30/100)	**100**
celkem	**100** **0, 100 (100/1000)**	**900** **0,900 (900/1000)**	**1 000**

Ve kterém městě je u lehkých pacientů nižší riziko úmrtí?

Motivační příklad – Simpsonův paradox

Horní Sádrovice			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	10 0,017 (10/600)	590 0,983 (590/600)	**600**
těžký	190 0,475 (190/400)	210 0,525 (210/400)	**400**
celkem	**200** **0,200 (200/1000)**	**800** **0,800 (800/1000)**	**1 000**

Stará Dláha			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	30 0,033 (30/900)	870 0,967 (870/900)	**900**
těžký	70 0, 700 (70/100)	30 0,300 (30/100)	**100**
celkem	**100** **0, 100 (100/1000)**	**900** **0,900 (900/1000)**	**1 000**

Ve kterém městě je u těžkých pacientů nižší riziko úmrtí?

Motivační příklad – Simpsonův paradox

Horní Sádrovice			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	10 0,017 (10/600)	590 0,983 (590/600)	**600**
těžký	190 **0,475** (190/400)	210 0,525 (210/400)	**400**
celkem	**200** **0,200 (200/1000)**	**800** **0,800 (800/1000)**	**1 000**

Stará Dláha			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	30 0,033 (30/900)	870 0,967 (870/900)	**900**
těžký	70 **0, 700** (70/100)	30 0,300 (30/100)	**100**
celkem	**100** **0, 100 (100/1000)**	**900** **0,900 (900/1000)**	**1 000**

Ve kterém městě je u těžkých pacientů nižší riziko úmrtí?

Motivační příklad – Simpsonův paradox

Horní Sádrovice			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	10 0,017 (10/600)	590 0,983 (590/600)	**600**
těžký	190 0,475 (190/400)	210 0,525 (210/400)	**400**
celkem	**200** **0,200 (200/1000)**	**800** **0,800 (800/1000)**	**1 000**

Stará Dláha			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	30 0,033 (30/900)	870 0,967 (870/900)	**900**
těžký	70 0, 700 (70/100)	30 0,300 (30/100)	**100**
celkem	**100** **0, 100 (100/1000)**	**900** **0,900 (900/1000)**	**1 000**

Ve kterém městě je u pacientů nižší riziko úmrtí?

Motivační příklad – Simpsonův paradox

Horní Sádrovice			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	10 0,017 (10/600)	590 0,983 (590/600)	**600**
těžký	190 0,475 (190/400)	210 0,525 (210/400)	**400**
celkem	**200** **0,200 (200/1000)**	**800** **0,800 (800/1000)**	**1 000**

Stará Dláha			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	30 0,033 (30/900)	870 0,967 (870/900)	**900**
těžký	70 0, 700 (70/100)	30 0,300 (30/100)	**100**
celkem	**100** **0, 100 (100/1000)**	**900** **0,900 (900/1000)**	**1 000**

Srovnání nemocnic			
nemocnice/ úmrtí	**ANO**	**NE**	**celkem**
Horní Sádrovice	200 0,200	800 0,800	**1 000**
Stará Dláha	100 0,100	900 0,900	**1 000**
celkem	**300** **0,150**	**1 700** **0,850**	**2 000**

Ve kterém městě je u pacientů nižší riziko úmrtí?

Motivační příklad – Simpsonův paradox

Horní Sádrovice			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	10 0,017 (10/600)	590 0,983 (590/600)	**600**
těžký	190 0,475 (190/400)	210 0,525 (210/400)	**400**
celkem	**200** **0,200 (200/1000)**	**800** **0,800 (800/1000)**	**1 000**

Stará Dláha			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	30 0,033 (30/900)	870 0,967 (870/900)	**900**
těžký	70 0, 700 (70/100)	30 0,300 (30/100)	**100**
celkem	**100** **0, 100 (100/1000)**	**900** **0,900 (900/1000)**	**1 000**

Srovnání nemocnic			
nemocnice/ úmrtí	**ANO**	**NE**	**celkem**
Horní Sádrovice	200 0,200	800 0,800	**1 000**
Stará Dláha	100 0,100	900 0,900	**1 000**
celkem	**300** **0,150**	**1 700** **0,850**	**2 000**

???

Ve kterém městě je u pacientů nižší riziko úmrtí?

Motivační příklad – Simpsonův paradox

Horní Sádrovice			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	10 0,017 (10/600)	590 0,983 (590/600)	**600**
těžký	190 0,475 (190/400)	210 0,525 (210/400)	**400**
celkem	**200** **0,200 (200/1000)**	**800** **0,800 (800/1000)**	**1 000**

Stará Dláha			
stav pacienta při přijetí/úmrtí	**ANO**	**NE**	**celkem**
lehký	30 0,033 (30/900)	870 0,967 (870/900)	**900**
těžký	70 0, 700 (70/100)	30 0,300 (30/100)	**100**
celkem	**100** **0, 100 (100/1000)**	**900** **0,900 (900/1000)**	**1 000**

Srovnání nemocnic			
nemocnice/ úmrtí	**ANO**	**NE**	**celkem**
Horní Sádrovice	200 0,200	800 0,800	**1 000**
Stará Dláha	100 0,100	900 0,900	**1 000**
celkem	**300** **0,150**	**1 700** **0,850**	**2 000**

Simpsonův paradox

Ve kterém městě je u pacientů nižší riziko úmrtí?

Simpsonův paradox

- Jedná se o situaci, kdy se závislost mezi dvěma znaky kvalitativně změní, jestliže uvážíme vliv znaku třetího (skrytého). (Např. vztah mezi úmrtnosti pacientů a místem léčby (Horní Sádrovice vs. Stará Dláha), vezmeme-li v úvahu stav pacienta při přijetí do nemocnice.)

- Důvodem je silná závislost mezi jedním z dvou analyzovaných znaků a znakem skrytým.

Simpsonův paradox

Zajímavé odkazy

- http://koroptew.blogspot.com/2010/11/zlocin-statistika.html
- Agresti, A. (2002). Categorical Data Analysis, Second Edition. Hoboken: John Wiley and Sons. ISBN 0-471-36093-7.
- Blyth, C. R. (1972). On Simpson's paradox and the sure-thing principle. Journal of the American Statistical Association, 67, 364-366.
- Davis, L. J. (1989). Intersection union tests for strictly collapsibility in three-dimensional contingency tables. Annals of Statistics, 17, 1693-1708.
- Dong, J. (1998). Simpson's paradox. Pp. 4108-4110 in Encyclopedia of Biostatistics, vol. 5. Chichester: John Wiley and Sons.
- Pavlides, M. G., Perlman, M. D. (2009). How likely is Simpson's paradox? The American Statistician, 63, 226-233.
- Samuels, M. L. (1993). Simpson's paradox and related phenomena. Journal of the American Statistical Association, 88, 81-88.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. Journal of the Royal Statistical Society, Series B, 13, 238-241.
- Wagner, C. H. (1982). Simpson's paradox in real life. The American Statistician, 36, 46-48.
- Wardrop, R. L. (1995). Simpson's paradox and the hot hand in basketball. The American Statistician, 49, 24-28.

Analýza závislosti dvou kvantitativních proměnných

Opakování z pravděpodobnosti - kovariance

Kovariance $cov(X,Y)$ je definována jako sdružený centrální moment řádu $(1+1)$.

$$cov(X,Y) = E\Big(\big(X - E(X)\big) \cdot \big(Y - E(Y)\big)\Big)$$

Vlastnosti kovariance:

1. $cov(X,Y) = E(XY) - E(X) \cdot E(Y)$ (výpočetní vztah)
2. $cov(X,X) = D(X)$,
3. $cov(a_1X + b_1, a_2Y + b_2) = a_1a_2cov(X,Y)$,
4. jsou-li X, Y nezávislé náhodné veličiny, pak $cov(X,Y) = 0$.

Opakování z pravděpodobnosti – korelační koeficient

Korelační koeficient $\rho(X,Y)$ je mírou závislosti náhodných veličin X, Y.

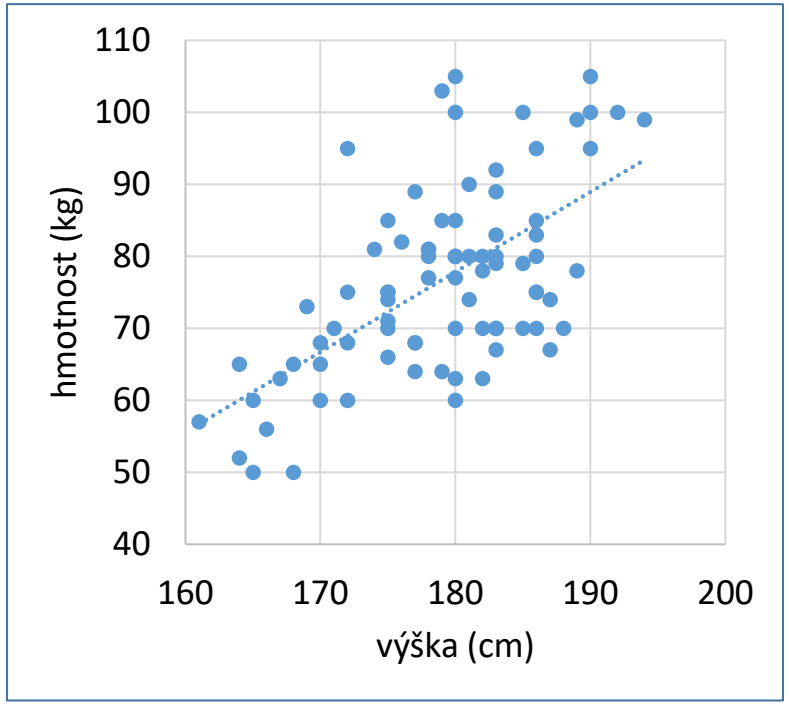
$$\rho(X,Y) = \begin{cases} \dfrac{cov(X,Y)}{\sqrt{DX \cdot DY}}, & DX, DY \neq 0, \\ 0 & jinak. \end{cases}$$

Pearsonův výběrový korelační koeficient je mírou **lineární** závislosti.

$$r = \frac{1}{n-1} \cdot \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{s_X \cdot s_Y}$$

Opakování z pravděpodobnosti – korelační koeficient

- **Vizualizace – bodový graf**

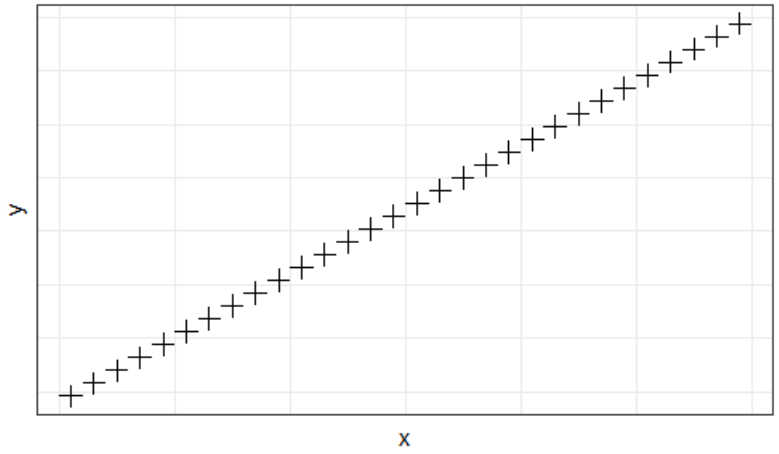


- **Pearsonův výběrový korelační koeficient**

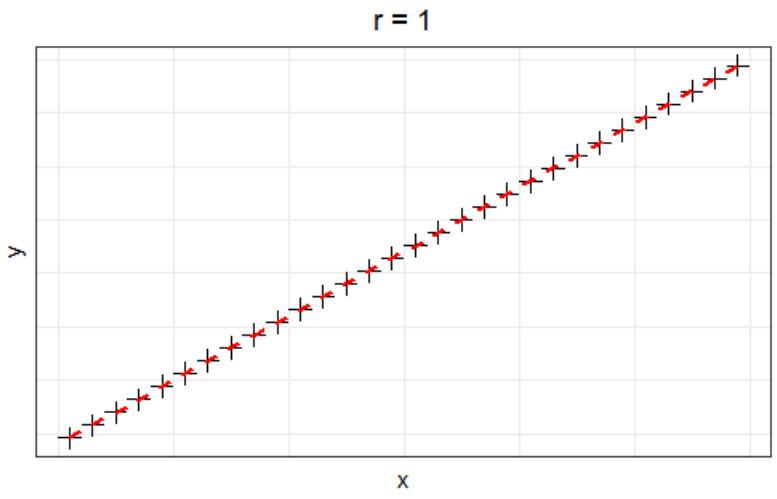
$$r = \frac{1}{n-1} \cdot \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{s_X \cdot s_Y}$$

- $-1 \leq r(X,Y) \leq 1$,
- $r(X,Y) = r(Y,X)$,
- $r(X,X) = 1$,
- je-li $r(X,Y) = 0$, říkáme, že X, Y jsou nekorelované znaky,
- je-li $r(X,Y) > 0$, říkáme, že X, Y jsou pozitivně korelované (s rostoucím X roste Y),
- je-li $r(X,Y) < 0$, říkáme, že X, Y jsou negativně korelované (s rostoucím X klesá Y),
- je-li $|r(X,Y)| = 1$, pak je mezi X a Y lineární závislost.

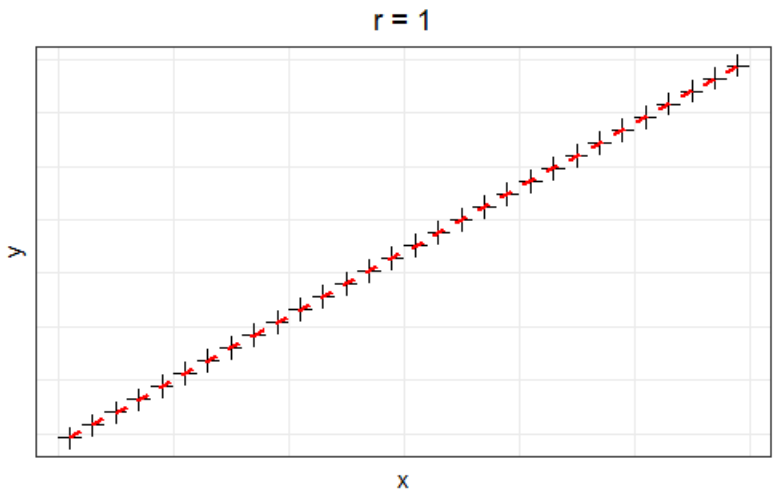
Pearsonův korelační koeficient

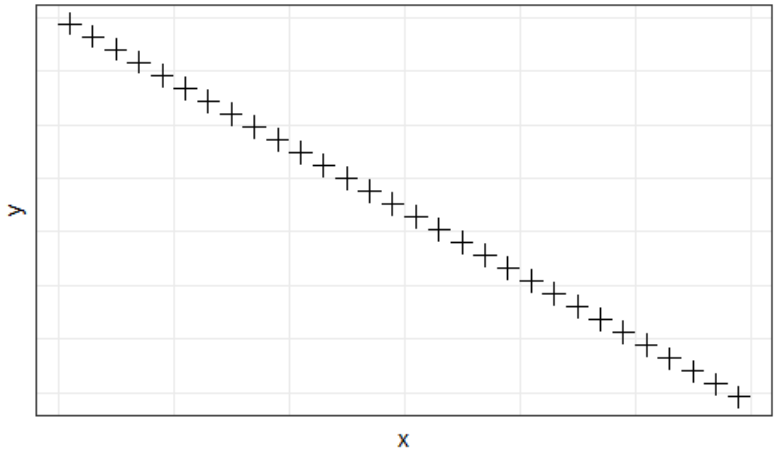


Pearsonův korelační koeficient

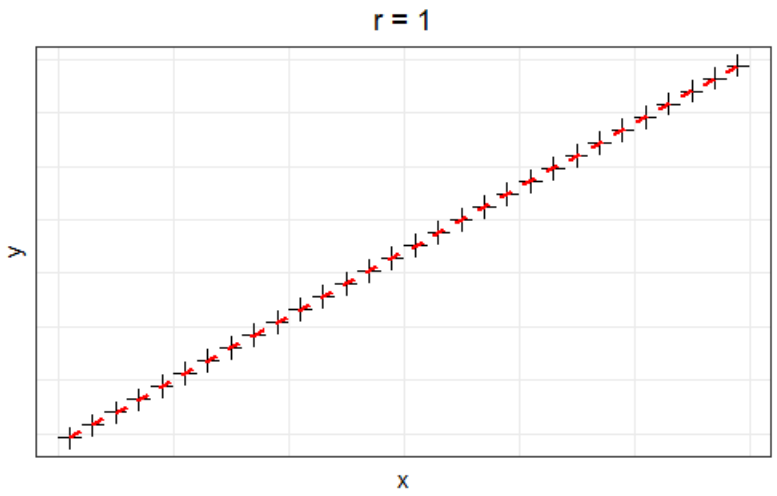


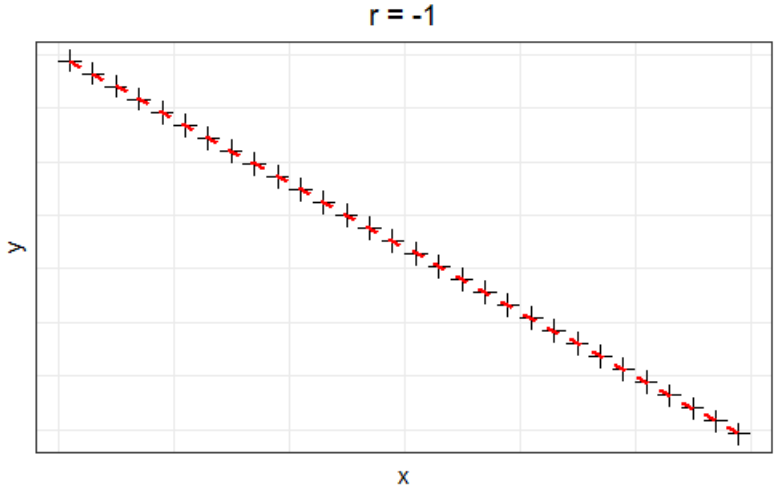
Pearsonův korelační koeficient



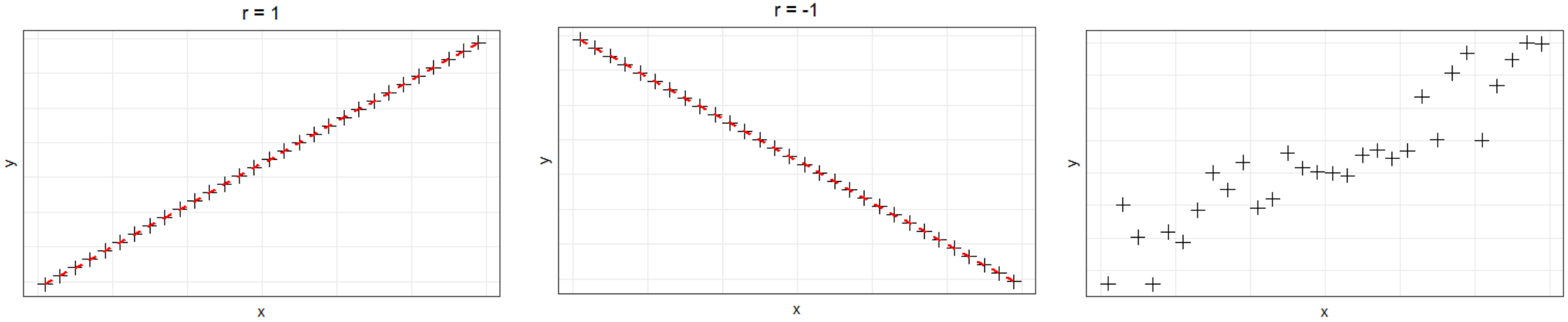


Pearsonův korelační koeficient

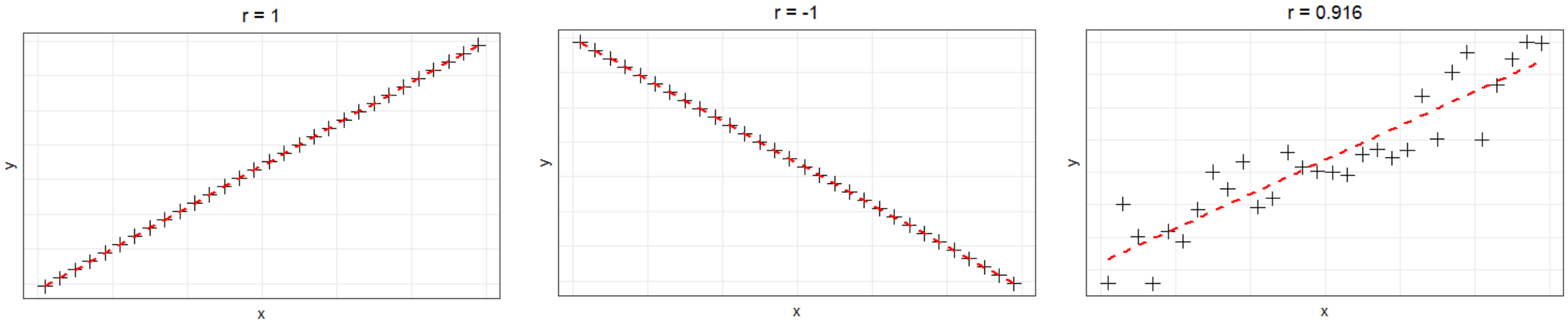




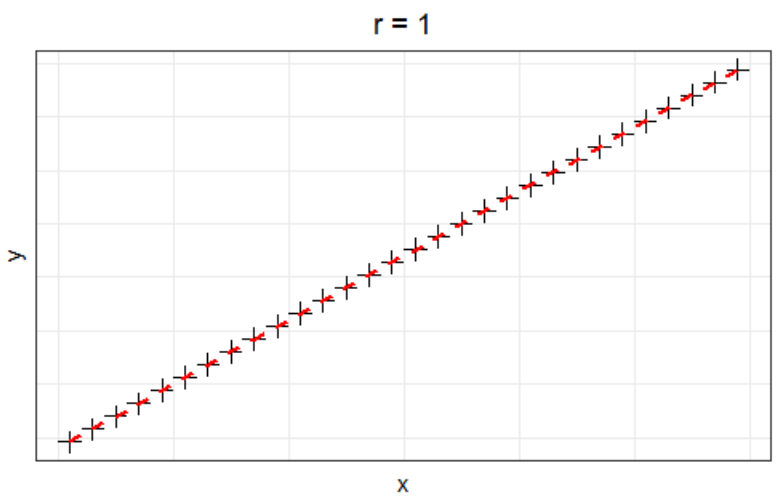
Pearsonův korelační koeficient

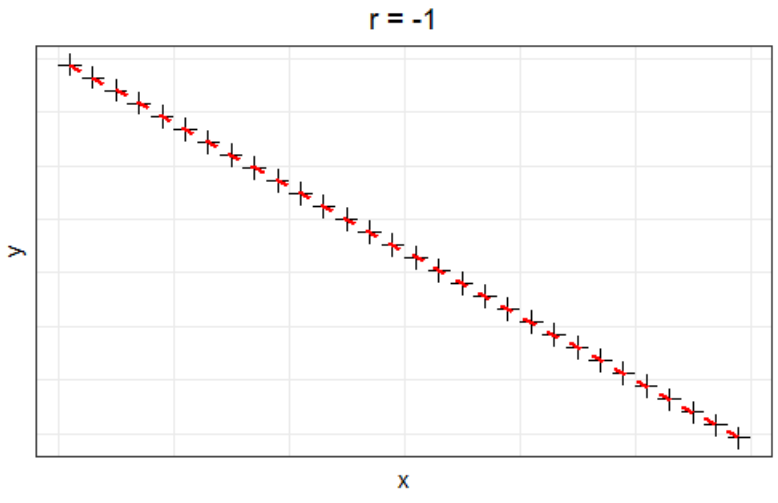


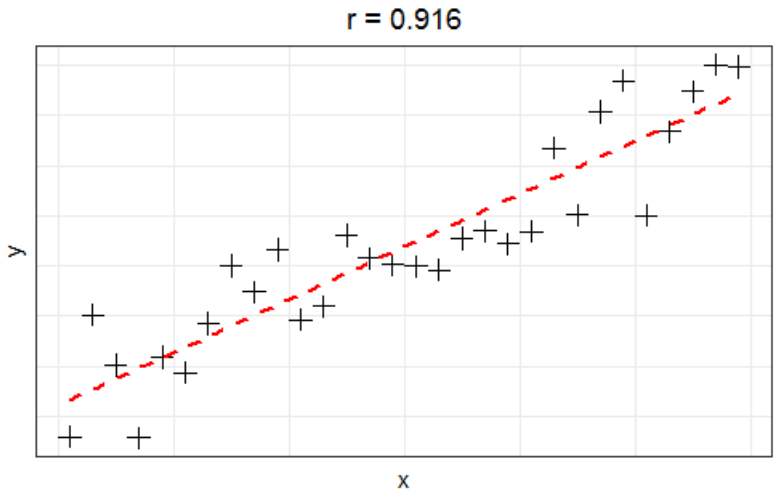
Pearsonův korelační koeficient

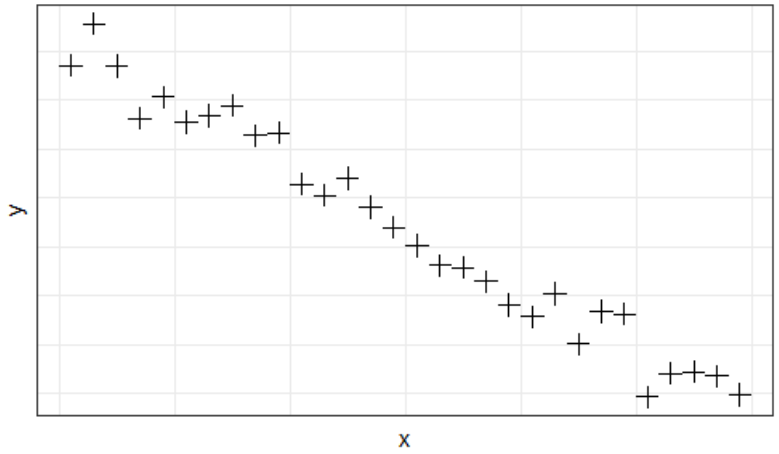


Pearsonův korelační koeficient

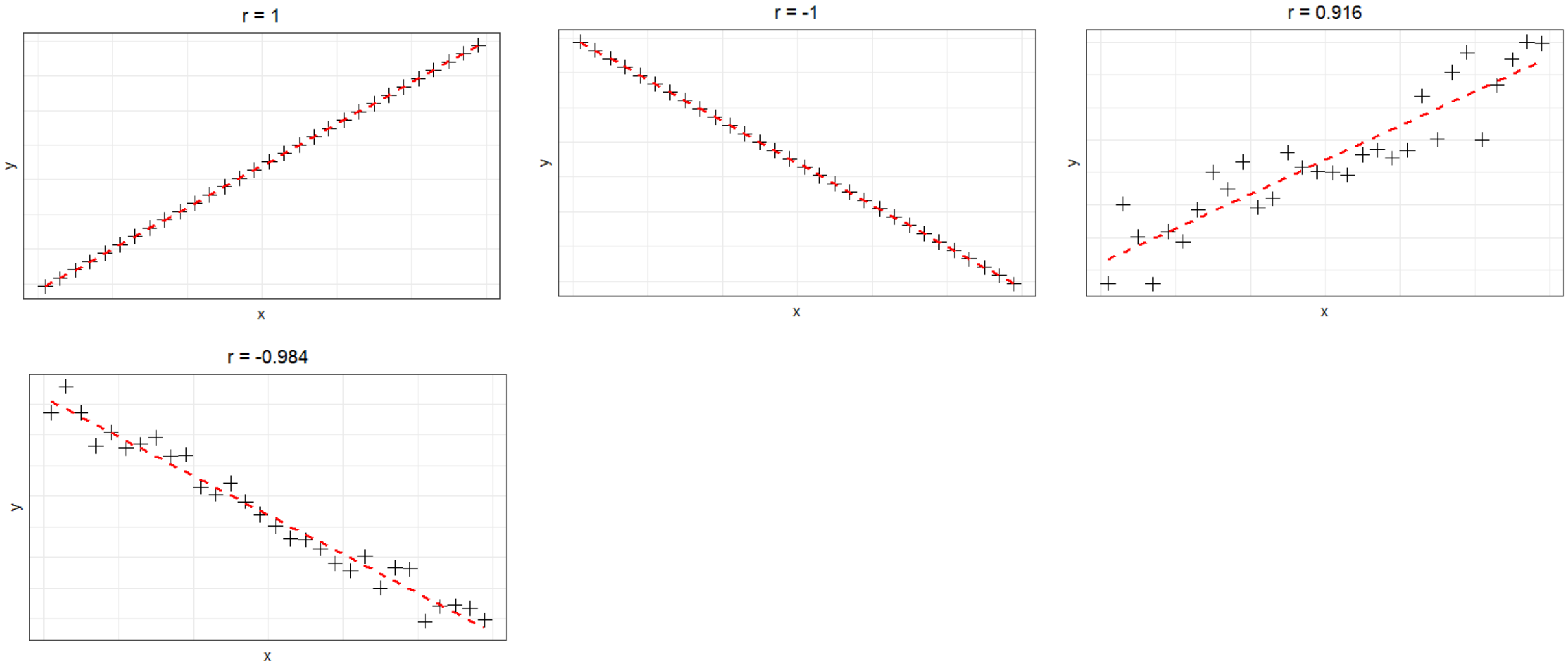




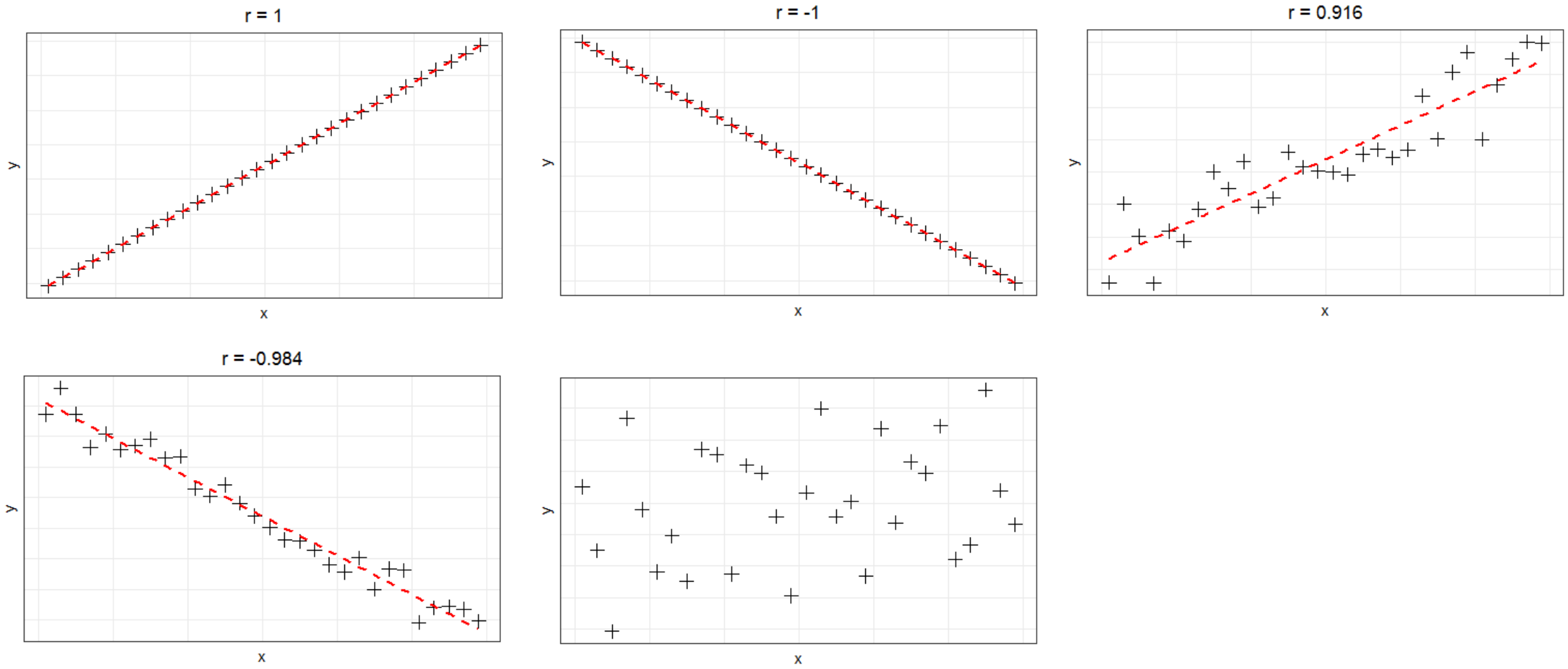




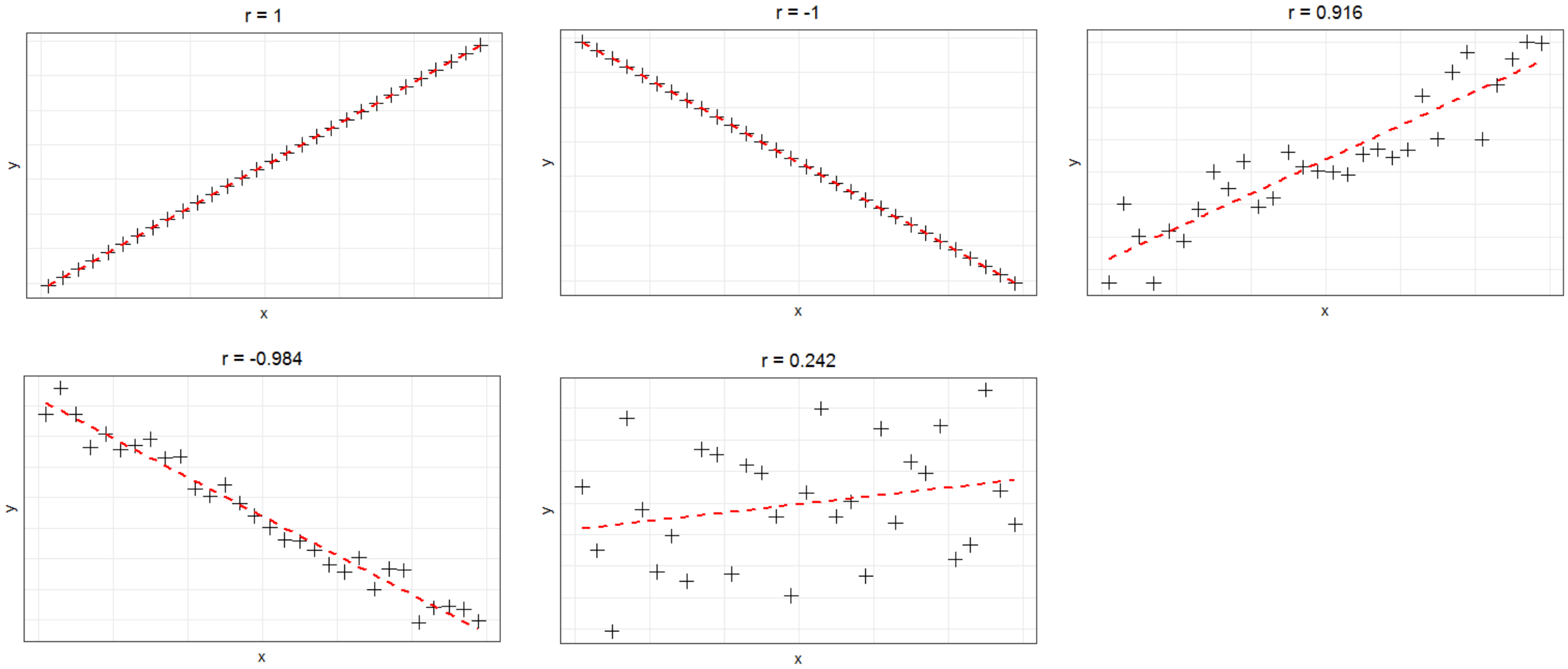
Pearsonův korelační koeficient



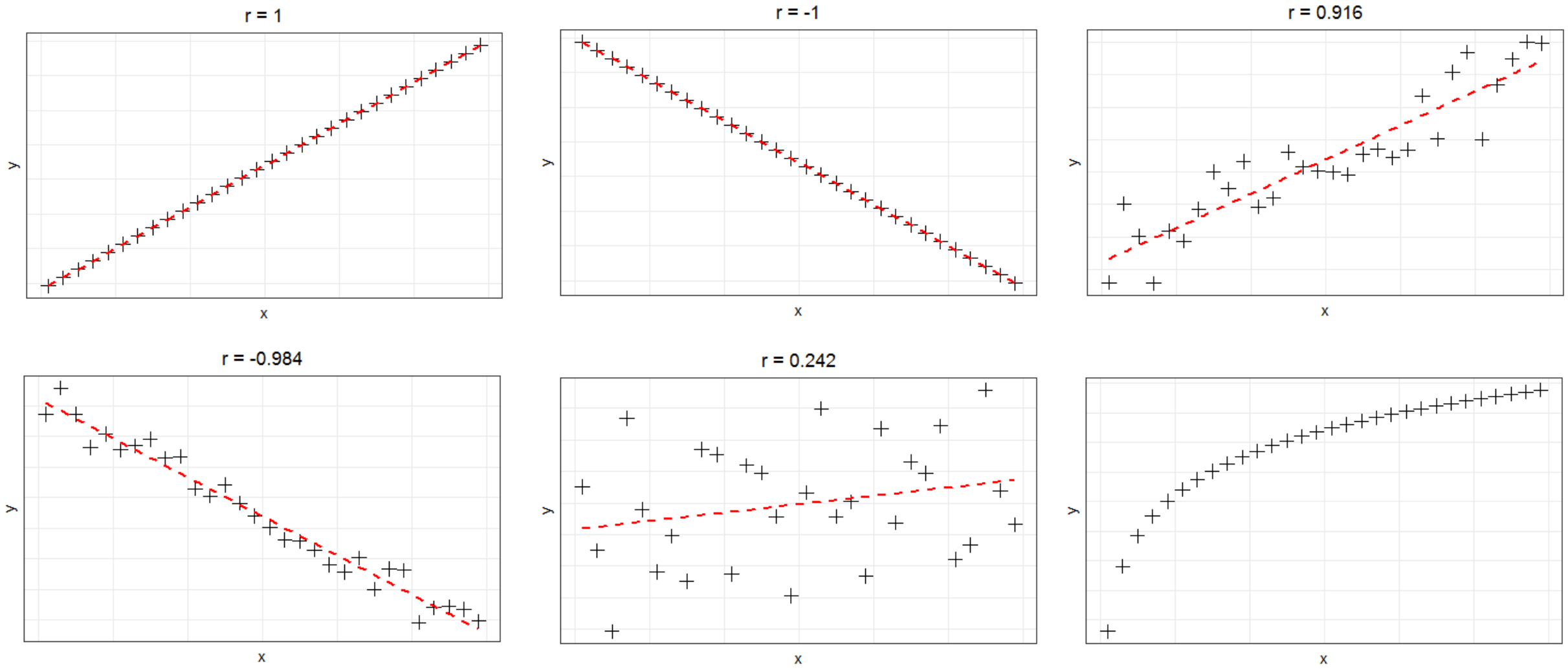
Pearsonův korelační koeficient



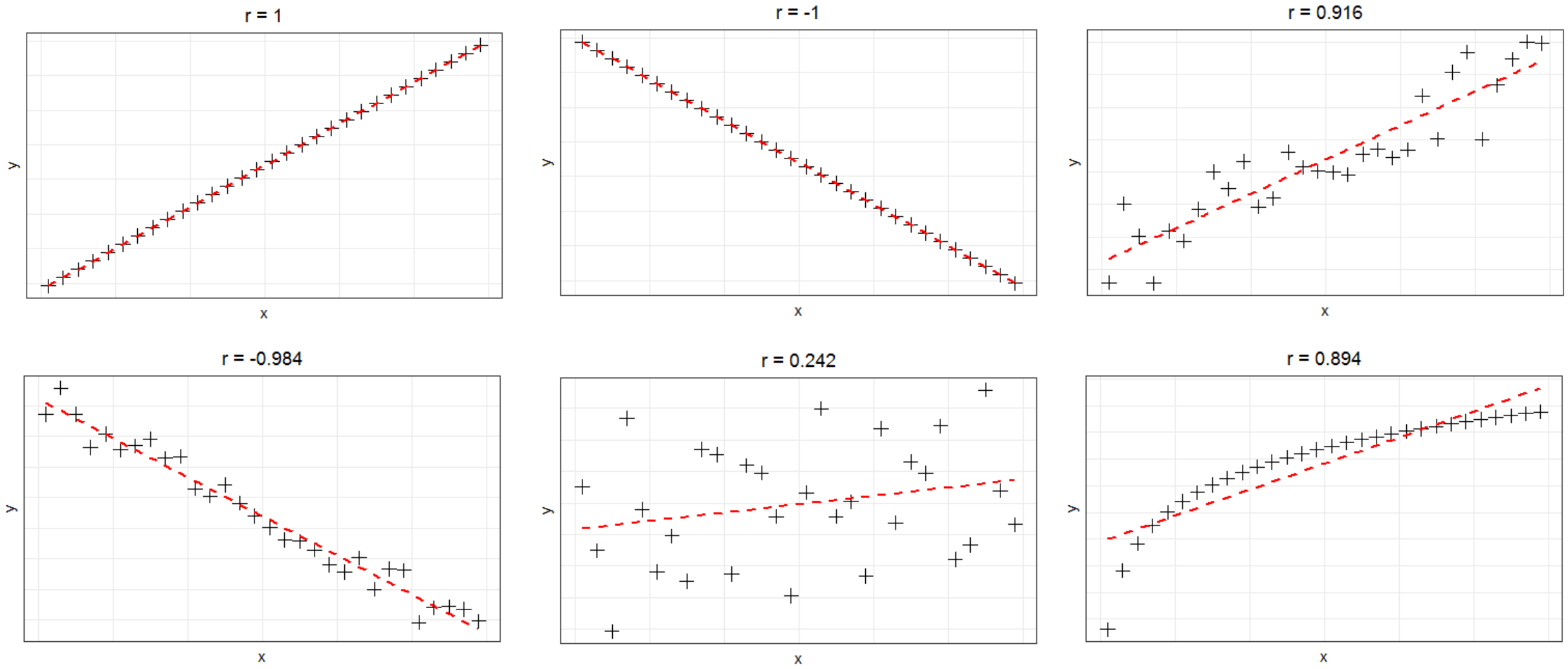
Pearsonův korelační koeficient



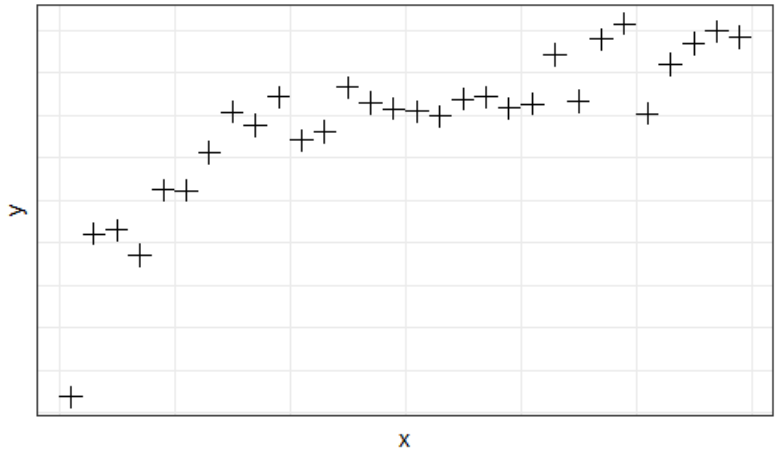
Pearsonův korelační koeficient



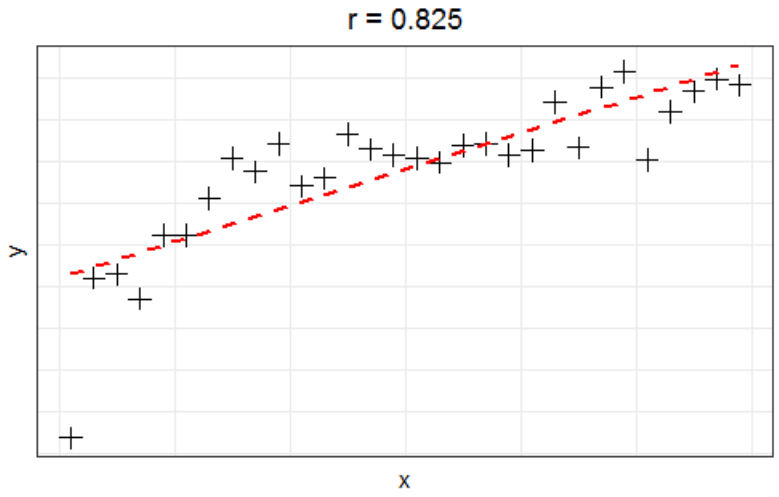
Pearsonův korelační koeficient



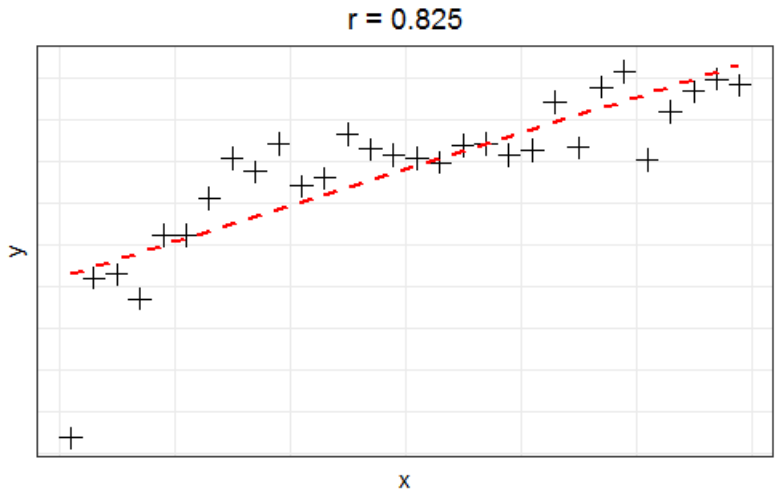
Pearsonův korelační koeficient

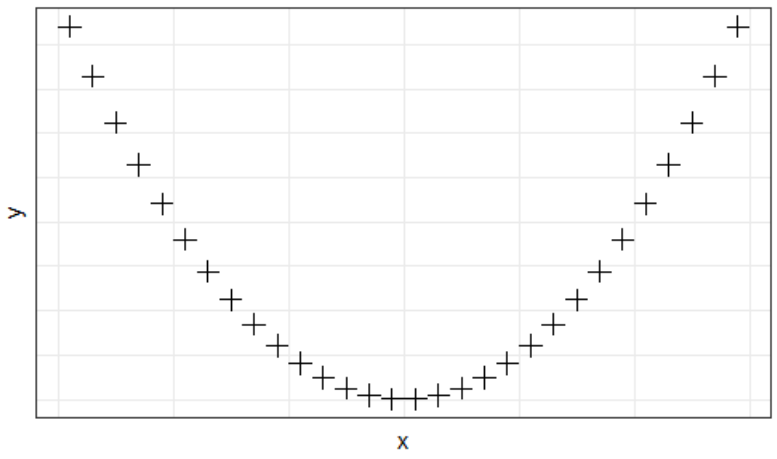


Pearsonův korelační koeficient

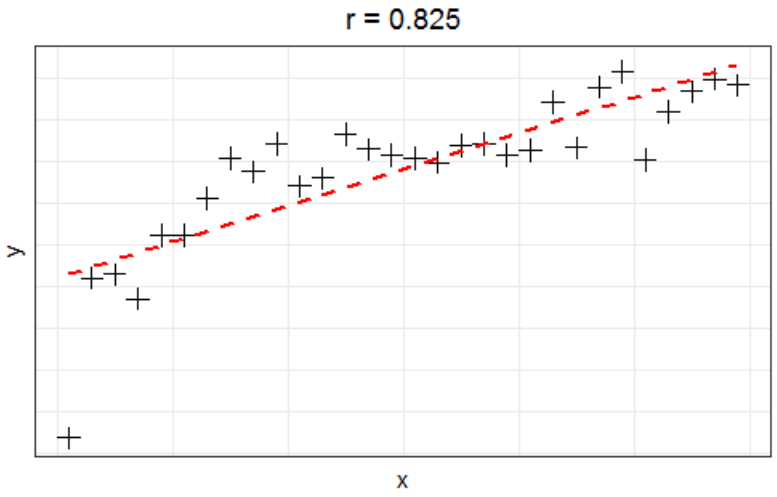


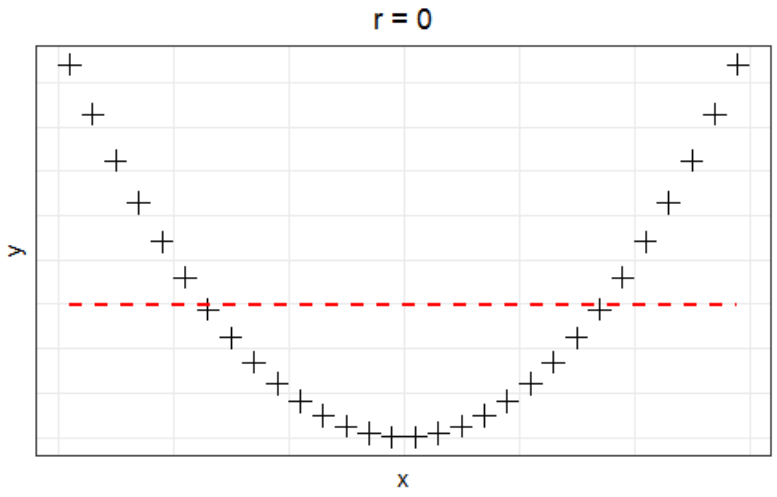
Pearsonův korelační koeficient



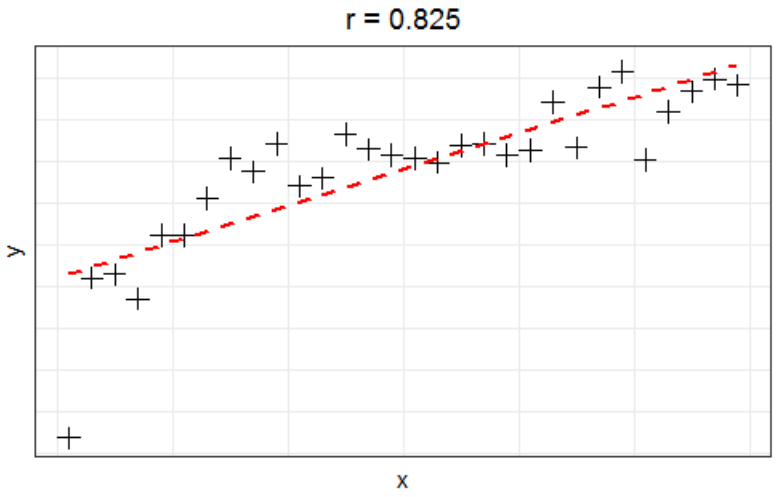


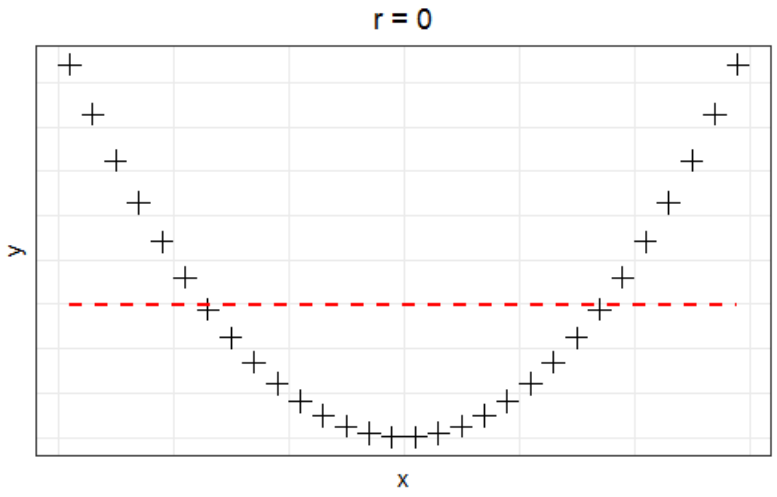
Pearsonův korelační koeficient

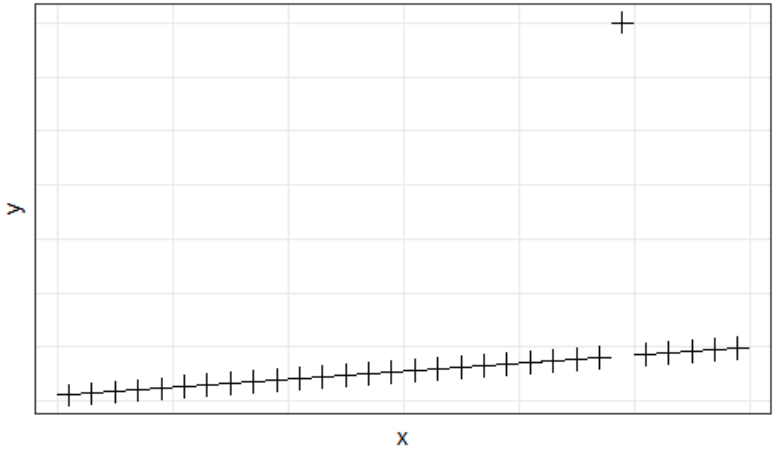




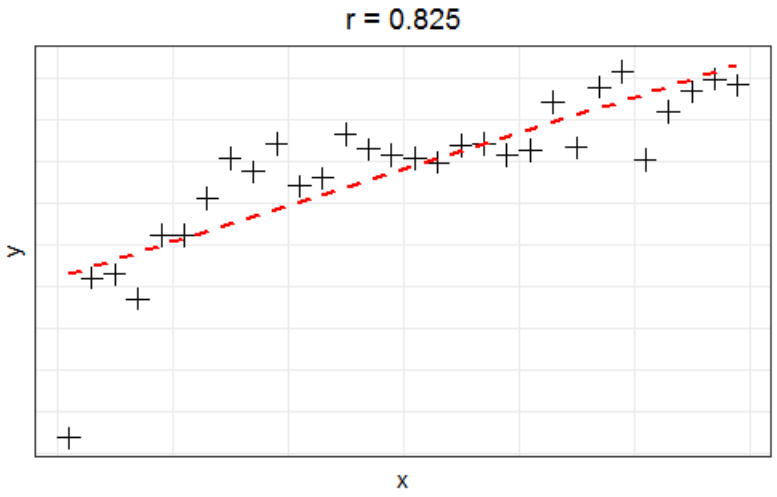
Pearsonův korelační koeficient

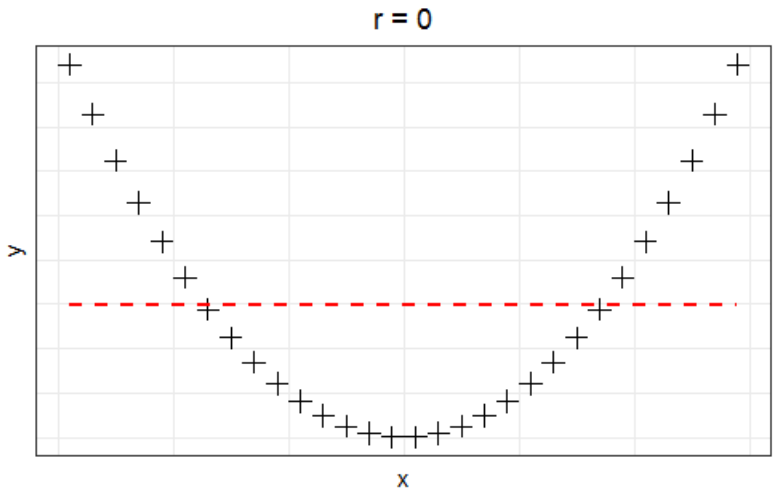


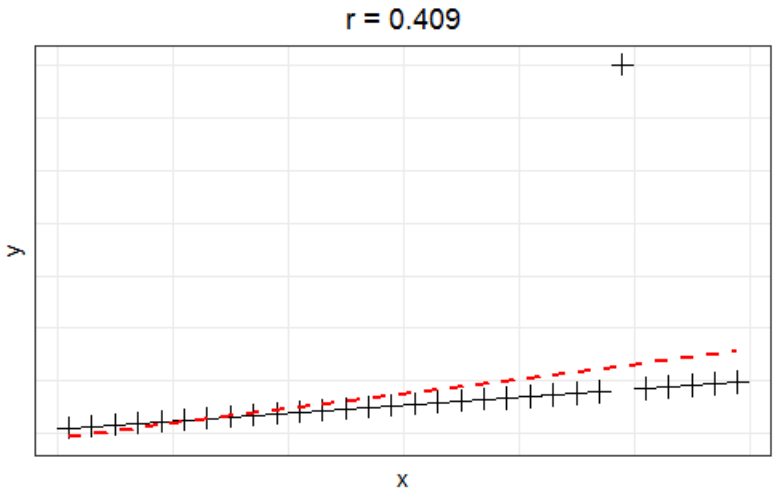




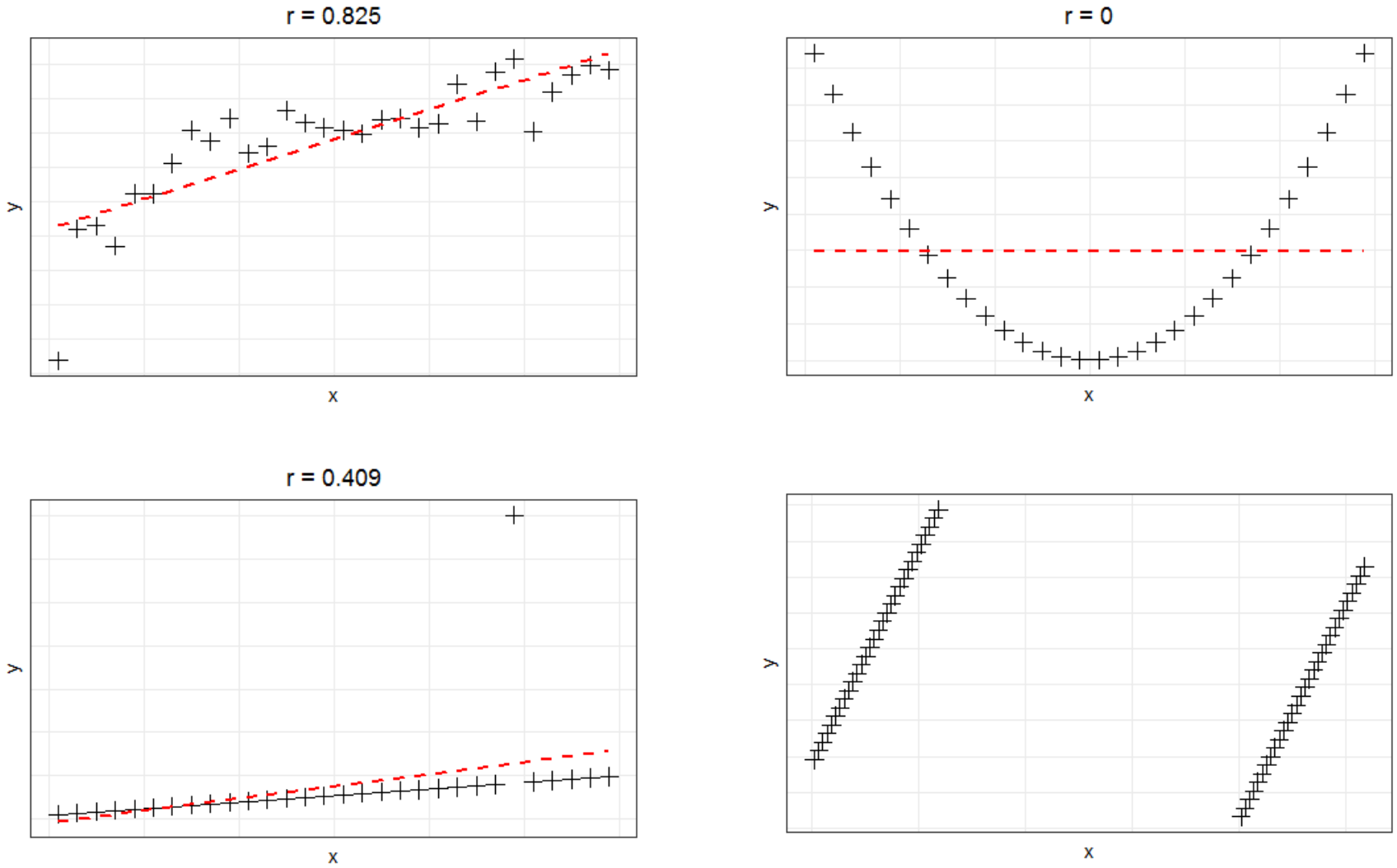
Pearsonův korelační koeficient



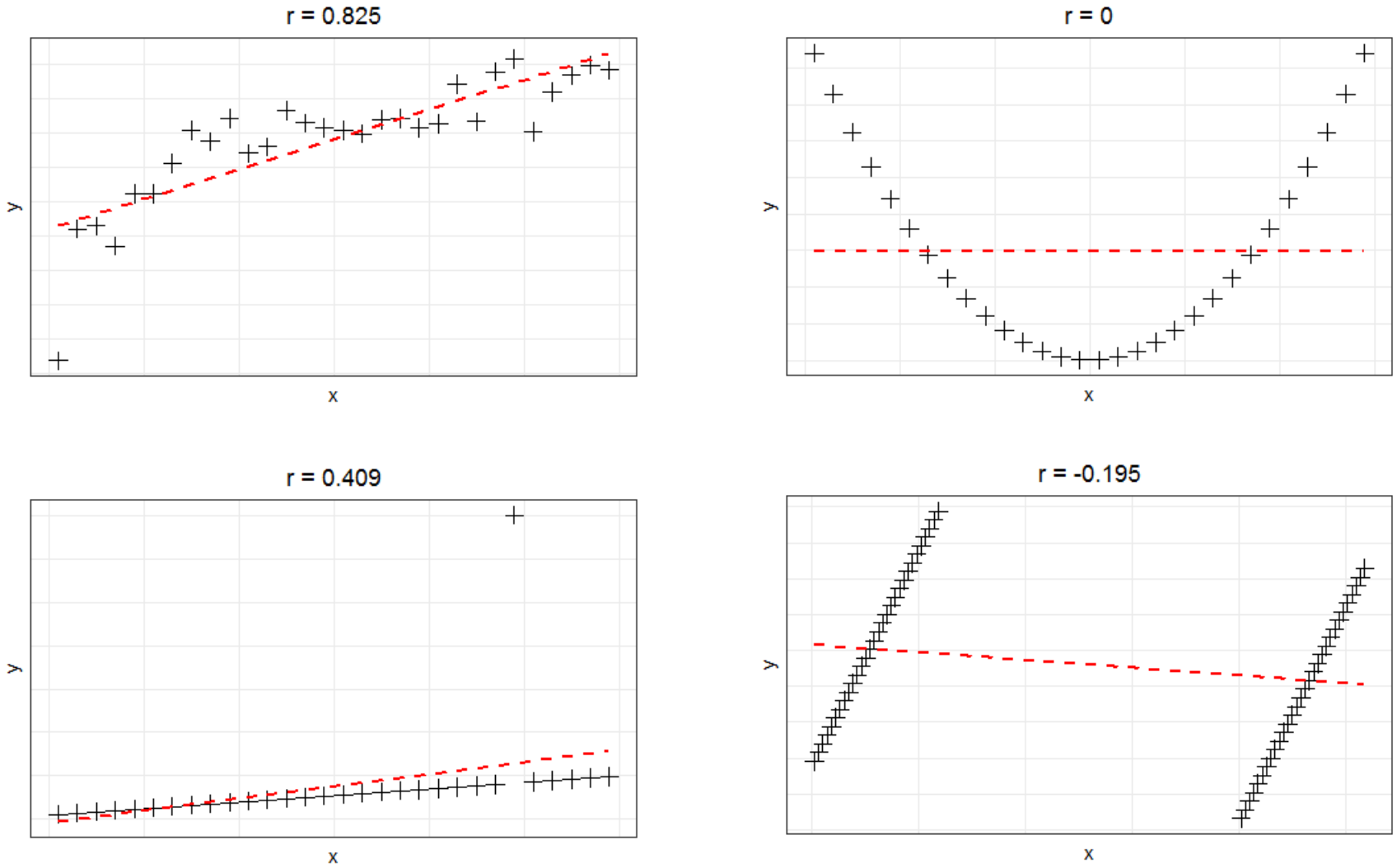




Pearsonův korelační koeficient



Pearsonův korelační koeficient



Pearsonův korelační koeficient

- Korelační koeficient ρ dokážeme určit pouze tehdy, známe-li sdružené rozdělení náhodného vektoru $(X;Y)$.
- Nechť $(X_1;Y_1),\dots,(X_n;Y_n)$ je výběr z **dvourozměrného normálního rozdělení**, tj. z rozdělení, jehož sdružená hustota pravděpodobnosti je dána vztahem

$$f(x;y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2-2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}.$$

Pak lze odhad korelačního koeficientu ρ určit jako

$$r=\begin{cases}\frac{S_{XY}}{\sqrt{S_X^2\cdot S_Y^2}}, & S_X^2,S_Y^2\neq 0,\\ 0 & jinak,\end{cases}$$

kde $S_{XY}=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})(Y_i-\bar{Y})=\frac{\sum_{i=1}^n X_iY_i-n\bar{X}\bar{Y}}{\sqrt{\left(\sum_{i=1}^n X_i^2-n\bar{X}^2\right)\left(\sum_{i=1}^n Y_i^2-n\bar{Y}^2\right)}}.$

Pearsonův korelační koeficient

- Nechť $(X_1; Y_1), \dots, (X_n; Y_n)$ je výběr z **dvourozměrného normálního rozdělení**.
- Zjistíme-li, že výběrový korelační koeficient $r \neq 0$, zpravidla nás zajímá, zda je indikovaná korelace statisticky významná. Chceme testovat nulovou hypotézu

$$H_0: \rho = 0 \text{ vůči alternativě } H_A: \rho \neq 0, \text{ resp. } \rho < 0, \text{ resp. } \rho > 0.$$

- Testová statistika: $T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t(n-2)$

Poznámka: *Jsou-li složky náhodného vektoru $(X; Y)$ s dvourozměrným normálním rozdělením nekorelované, jsou nezávislé. (**POZOR! Obecně to neplatí.**)*

Spearmanův korelační koeficient

- Mějme náhodný výběr $(X_1; Y_1), \dots, (X_n; Y_n)$ z dvourozměrného rozdělení. Nechť $R_{X_1}, \dots, R_{X_n}$ jsou pořadí veličin $X_1, \dots, X_n$ a nechť $R_{Y_1}, \dots, R_{Y_n}$ jsou pořadí veličin $Y_1, \dots, Y_n$.

$$r_S = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} \left(R_{X_1} - R_{Y_1}\right)^2$$

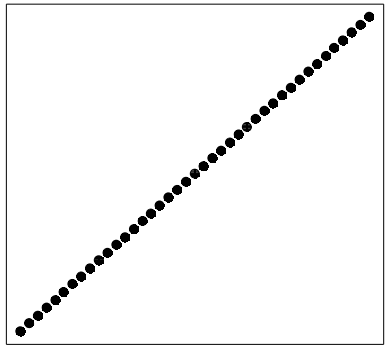
- Pokud se v náhodných výběrech, z nichž je r_S počítán, vyskytuje mnoho shod (tj. stejně velkých pozorování), doporučuje se používat **korigovaný Spearmanův korelační koeficient** $r_{S_{korig}}$.

$$r_{S_{korig}} = 1 - \frac{6}{n^3 - n - T_X - T_Y} \sum_{i=1}^{n} \left(R_{X_1} - R_{Y_1}\right)^2,$$

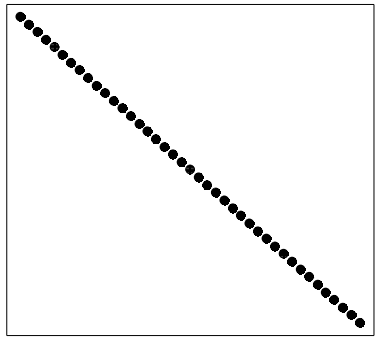
kde $T_X = \frac{1}{2} \sum \left(t_X^3 - t_X\right)$, $T_Y = \frac{1}{2} \sum \left(t_Y^3 - t_Y\right)$,

kde t_X jsou rozsahy skupin stejně velkých X-ových hodnot. Obdobně definujeme t_Y.

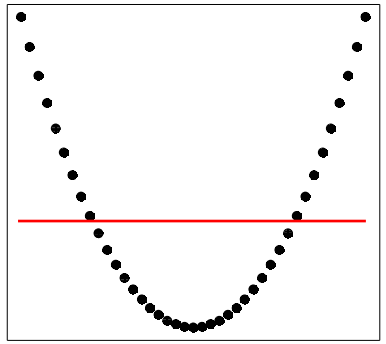
Srovnání Pearsonova a Spearmanova kor. koeficientu



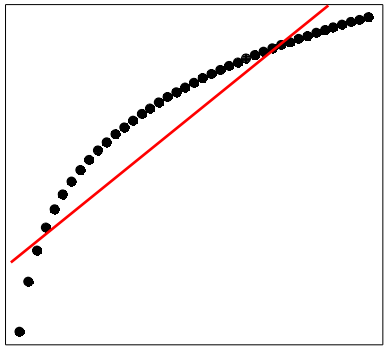
$\rho(X,Y) = 1{,}000$
$\rho_S(X,Y) = 1{,}000$



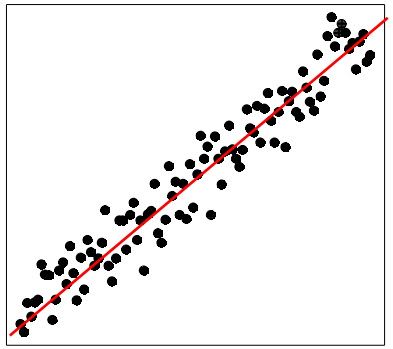
$\rho(X,Y) = -1{,}000$
$\rho_S(X,Y) = -1{,}000$



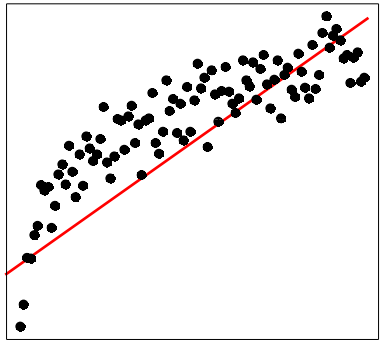
$\rho(X,Y) = 0{,}000$
$\rho_S(X,Y) = 0{,}000$



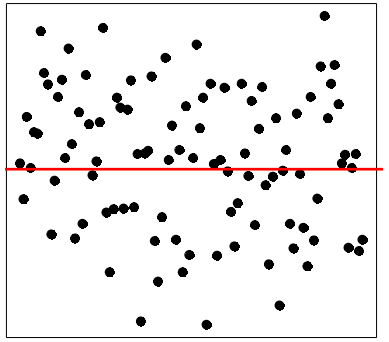
$\rho(X,Y) = 0{,}934$
$\rho_S(X,Y) = 1{,}000$



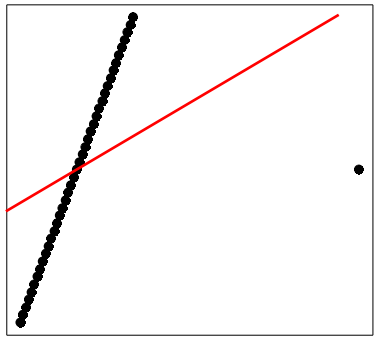
$\rho(X,Y) = 0{,}967$
$\rho_S(X,Y) = 0{,}981$



$\rho(X,Y) = 0{,}857$
$\rho_S(X,Y) = 0{,}893$



$\rho(X,Y) = -0{,}143$
$\rho_S(X,Y) = -0{,}178$



$\rho(X,Y) = 0{,}608$
$\rho_S(X,Y) = 0{,}911$

Spearmanův korelační koeficient

- Je-li hodnota Spearmanova korelačního koeficientu r_S blízká nule, chceme zpravidla testovat, zda je odchylka koeficientu r_S od nuly náhodná či statisticky významná.

H_0: X, Y jsou **nezávislé** náhodné veličiny.

H_A: X, Y jsou závislé náhodné veličiny.

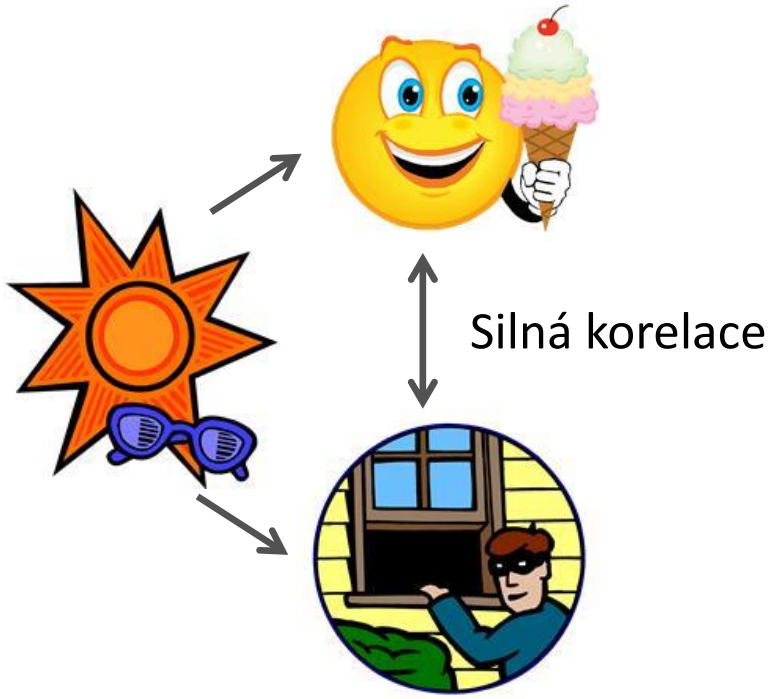
- Testová statistika: r_S
- Nulovou hypotézu zamítáme pokud $|r_S| \geq r_S^*(\alpha)$, kde $r_S^*(\alpha)$ je kritická hodnota Spearmanova korelačního koeficientu.
- Pro rozsah výběru $n \leq 30$ a hladiny významnosti 0,05, resp. 0,01 jsou kritické hodnoty $r_S^*(\alpha; n)$ tabelovány (tabulka T16). Je-li rozsah výběru $n > 30$, pak

$$r_S^*(\alpha; n) = \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n-1}},$$

kde $z_{1-\frac{\alpha}{2}}$ je $\left(1 - \frac{\alpha}{2}\right)$ kvantil normovaného normálního rozdělení.

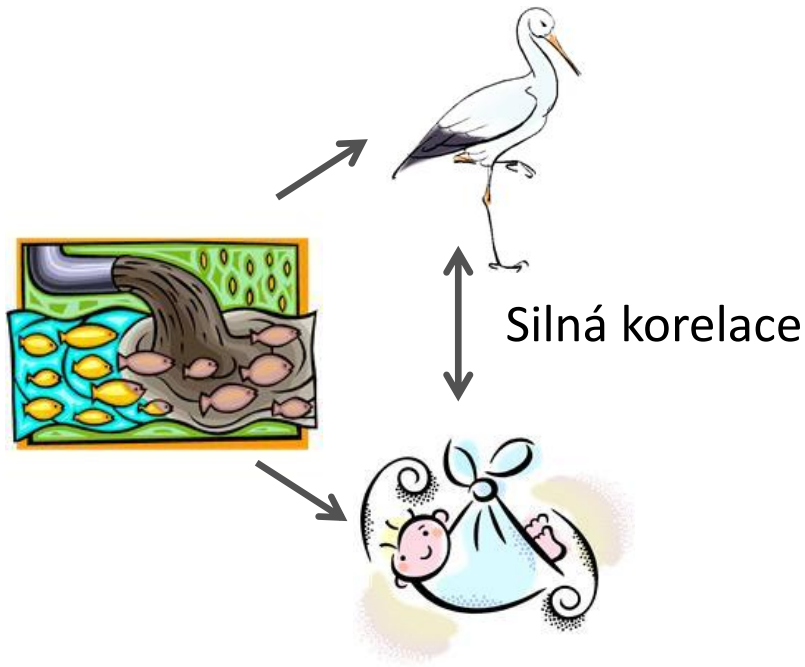
Korelační koeficient

- Pokud jsou dvě náhodné veličiny korelované (Pearsonův korelační koeficient), znamená to pouze to, že jsou lineárně závislé.
- Nelze z toho však ještě usoudit, že by jedna z nich musela být příčinou a druhá následkem. To samotná korelovanost nedovoluje rozhodnout.



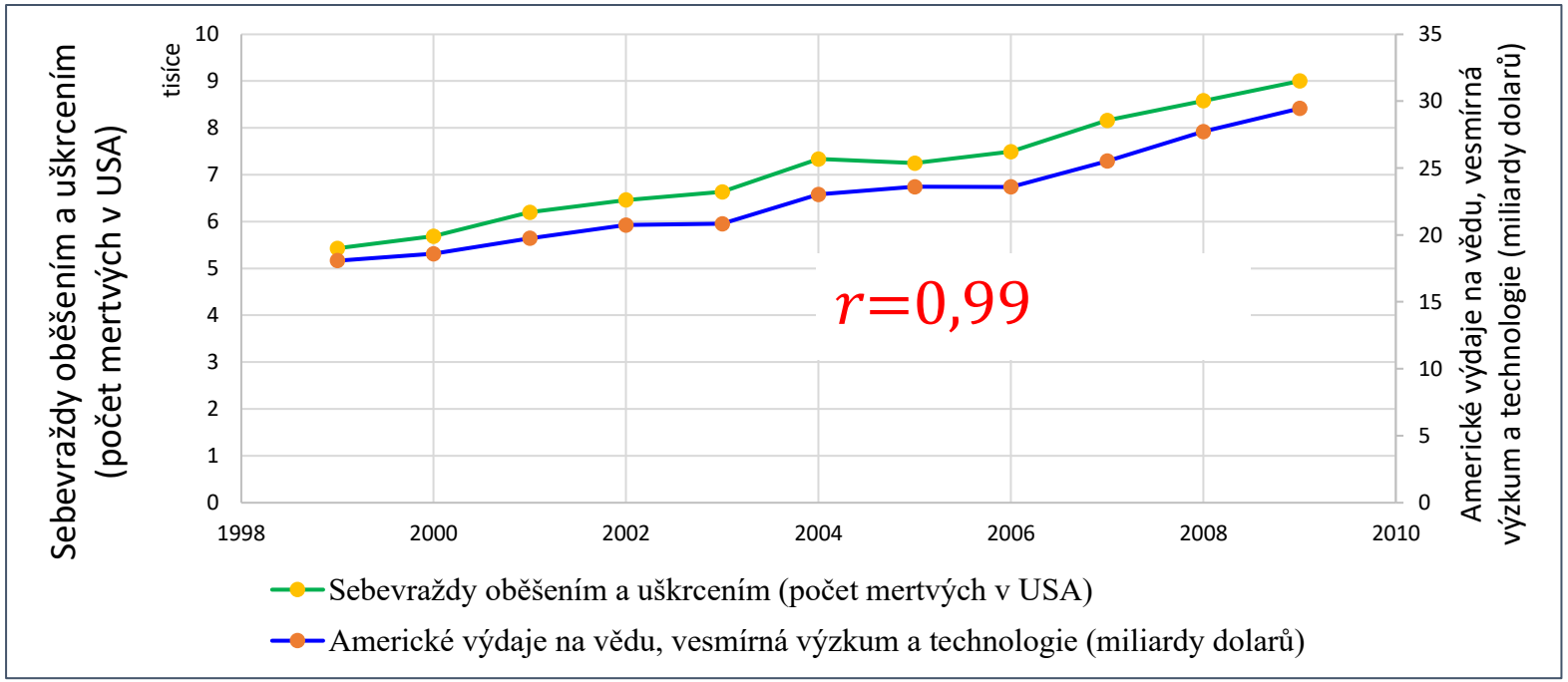
Korelační koeficient

- Pokud jsou dvě náhodné veličiny korelované (Pearsonův korelační koeficient), znamená to pouze to, že jsou lineárně závislé.
- Nelze z toho však ještě usoudit, že by jedna z nich musela být příčinou a druhá následkem. To samotná korelovanost nedovoluje rozhodnout.



Korelační koeficient

- Pokud jsou dvě náhodné veličiny korelované (Pearsonův korelační koeficient), znamená to pouze to, že jsou lineárně závislé.
- Nelze z toho však ještě usoudit, že by jedna z nich musela být příčinou a druhá následkem. To samotná korelovanost nedovoluje rozhodnout.



Korelační koeficient

Aktuálně.cz

ZPRÁVY Domácí Zahraničí

ZPRÁVY / ZAHRANIČÍ

K Nobelově ceně dopomáhá čokoláda, naznačuje studie

12. 10. 2012 10:36 AKTUALIZOVÁNO

Mezi počtem nobelistů v přepočtu na obyvatele a konzumaci čokolády je souvislost

New York - Počet nositelů Nobelovy ceny v přepočtu na jednoho obyvatele se v jednotlivých zemích odvíjí od spotřeby čokolády.

To není úryvek z reklamy na sladkosti, ale závěr studie publikované v jednom z nejprestižnějších světových lékařských časopisů New England Journal of Medicine.

Nejvíce laureátů Nobelovy ceny mají Švýcaři, kteří jsou zároveň největšími jedlíky oblíbené sladké pochutiny.

Zdroj: http://zpravy.aktualne.cz/zahranici/k-nobelove-cene-dopomaha-cokolada-naznacuje-studie/r~i:article:760147/

Odhad korelačního koeficientu

V praxi se zpravidla hodnota korelačního koeficientu interpretuje takto:

Korelační koeficient	**Typ lineární závislosti**
$\lvert r \rvert = 0{,}0$	neexistující
$\lvert r \rvert \in (0{,}0; 0{,}3\rangle$	velmi slabá
$\lvert r \rvert \in (0{,}3; 0{,}7\rangle$	středně silná
$\lvert r \rvert \in (0{,}7; 1{,}0)$	těsná
$\lvert r \rvert = 1{,}0$	funkční

- Mezi proudem a napětím na odporu byl zjištěn korelační koeficient 0,6.
- Mezi školním prospěchem a pocitem deprese u dětí byl zjištěn korelační koeficient 0,6.

Výsledky interpretujte!

Korelační matice

Korelační matice

$$cor(\boldsymbol{X}) = \begin{pmatrix} \rho(X,X) & \rho(X,Y) \\ \rho(Y,X) & \rho(Y,Y) \end{pmatrix} = \begin{pmatrix} 1 & \rho(X,Y) \\ \rho(X,Y) & 1 \end{pmatrix}$$

Děkuji za pozornost!

martina.litschmannova@vsb.cz

VŠB TECHNICKÁ UNIVERZITA OSTRAVA | FAKULTA ELEKTROTECHNIKY A INFORMATIKY | KATEDRA APLIKOVANÉ MATEMATIKY