

NĚKOLIK TIPŮ NA OŽIVENÍ VÝUKY STATISTIKY ANEB POTŘEHY JEDNÉ DATOVÉ ANALYTIČKY

Martina Litschmannová
UPVM 2017



Katedra
aplikované matematiky

MATH 4 U

Spolufinancováno
z programu Evropské unie
Erasmus+



Obsah

- Datová analýza
- Mezioborové přesahy a vazby (Matematika, ICT, Fyzika, Mediální výchova)
- Základní terminologie
- Analýza kvalitativního znaku – rozdělení četnosti, vizualizace
- Analýza kvantitativního znaku – míry polohy, míry variability, vizualizace
- Analýza závislosti dvou kvantitativních znaků – koeficient korelace, vizualizace

Motivace

- Dotazník pro studenty (např. pomocí Google Apps) - <http://goo.gl/forms/Z289s0ALPY>

Dotazník pro výukové účely - Brigáda při studiu

Vážené studentky, vážení studenti,

prosím Vás o vyplnění krátkého anonymního dotazníku na téma brigáda při studiu. Výsledky dotazníkového šetření budou použity pro prezentaci analýzy dat v rámci výuky.

Předem moc děkuji za váš čas a ochotu,
Martina Litschmannová

*Povinné pole

Identifikace respondenta

1. Pohlaví *

Označte jen jednu elipsu.

- žena
 muž

2. Výška (cm) *

.....

3. Váha (kg) *

.....

4. Přivyděláváte si v rámci prezenčního studia na brigádách? *

Označte jen jednu elipsu.

- ano *Přeskočte na otázku 5.*
 ne *Přeskočte na otázku 8.*

Brigáda při studiu

5. Jak často brigádu máte? *

Označte jen jednu elipsu.

- každý pracovní den
 pouze o víkendu
 nepravidelně
 Jiné:

6. Jak byste svou brigádu charakterizoval(a)? *

Označte jen jednu elipsu.

- praxe v oboru během studia
 "kancelářská" práce mimo obor mého studia
 manuální práce
 Jiné:

7. Kolik času týdně obvykle věnujete brigádě? *

Uvedte čas v hodinách.

.....

Studium

8. Kolik času týdně obvykle věnujete studiu? *

Uvedte čas v hodinách.

.....

Základní pojmy

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	kancelářská práce a na ní navazující práce manuální při realizaci projektů	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během studia	5	36
1.4.2016 10:45	žena	166	56	ano	Různě, 2-3 týdně	Hlídní dětí	12	10
1.4.2016 10:52	žena	188	70	ano	3 dny v týdnu	praxe v oboru během studia	24	26

- Statistika se zabývá zkoumáním **hromadných jevů**.
- **Statistický soubor** – soubor osob, věcí, událostí, časových období, ...
- **Statistická jednotka** – prvek statistického souboru (v dotazníkovém šetření nazýváme statistické jednotky respondenty)
- **Statistický znak** – atribut statistické jednotky (to, co u stat. jednotek zkoumáme)

Základní pojmy

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	kancelářská práce a na ní navazující práce manuální při realizaci projektů	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během studia	5	36
1.4.2016 10:45	žena	166	56	ano	Různě, 2-3 týdně	Hlídní dětí	12	10
1.4.2016 10:52	žena	188	70	ano	3 dny v týdnu	praxe v oboru během studia	24	26

- **Statistický znak** – atribut statistické jednotky (to, co u stat. jednotek zkoumáme)

Základní pojmy

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	kancelářská práce a na ní navazující práce manuální při realizaci projektů	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během studia	5	36
1.4.2016 10:45	žena	166	56	ano	Různě, 2-3 týdně	Hlídní dětí	12	10
1.4.2016 10:52	žena	188	70	ano	3 dny v týdnu	praxe v oboru během studia	24	26

- **Statistický znak** – atribut statistické jednotky (to, co u stat. jednotek zkoumáme)
 - **Kvantitativní znak** – znak, jehož varianty nabývají číselných hodnot

Základní pojmy

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	kancelářská práce a na ní navazující práce manuální při realizaci projektů	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během studia	5	36
1.4.2016 10:45	žena	166	56	ano	Různě, 2-3 týdně	Hlídní dětí	12	10
1.4.2016 10:52	žena	188	70	ano	3 dny v týdnu	praxe v oboru během studia	24	26

- **Statistický znak** – atribut statistické jednotky (to, co u stat. jednotek zkoumáme)
 - **Kvantitativní znak** – znak, jehož varianty nabývají číselných hodnot

Základní pojmy

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	kancelářská práce a na ní navazující práce manuální při realizaci projektů	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během studia	5	36
1.4.2016 10:45	žena	166	56	ano	Různě, 2-3 týdně	Hlídní dětí	12	10
1.4.2016 10:52	žena	188	70	ano	3 dny v týdnu	praxe v oboru během studia	24	26

- **Statistický znak** – atribut statistické jednotky (to, co u stat. jednotek zkoumáme)
 - **Kvantitativní znak** – znak, jehož varianty nabývají číselných hodnot
 - **Kvalitativní znak** – znak, jehož varianty se liší kvalitou (může jít i o číselné hodnoty – např. známka z matematiky)

Základní pojmy

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	kancelářská práce a na ní navazující práce manuální při realizaci projektů	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během studia	5	36
1.4.2016 10:45	žena	166	56	ano	Různě, 2-3 týdně	Hlídní dětí	12	10
1.4.2016 10:52	žena	188	70	ano	3 dny v týdnu	praxe v oboru během studia	24	26

- **Statistický znak** – atribut statistické jednotky (to, co u stat. jednotek zkoumáme)
 - **Kvantitativní znak** – znak, jehož varianty nabývají číselných hodnot
 - **Kvalitativní znak** – znak, jehož varianty se liší kvalitou (může jít i o číselné hodnoty – např. známka z matematiky)

Základní pojmy

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	kancelářská práce a na ní navazující práce manuální při realizaci projektů	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během studia	5	36
1.4.2016 10:45	žena	166	56	ano	Různě, 2-3 týdně	Hlídní dětí	12	10
1.4.2016 10:52	žena	188	70	ano	3 dny v týdnu	praxe v oboru během studia	24	26

- **Statistický znak** – atribut statistické jednotky (to, co u stat. jednotek zkoumáme)
 - **Kvantitativní znak** – znak, jehož varianty nabývají číselných hodnot
 - **Kvalitativní znak** – znak, jehož varianty se liší kvalitou (může jít i o číselné hodnoty – např. známka z matematiky)
 - **Alternativní znak** – kvalitativní znak, který nabývá pouze dvou možných variant

Základní pojmy

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	kancelářská práce a na ní navazující práce manuální při realizaci projektů	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během studia	5	36
1.4.2016 10:45	žena	166	56	ano	Různě, 2-3 týdně	Hlídní dětí	12	10
1.4.2016 10:52	žena	188	70	ano	3 dny v týdnu	praxe v oboru během studia	24	26

- **Statistický znak** – atribut statistické jednotky (to, co u stat. jednotek zkoumáme)
 - **Kvantitativní znak** – znak, jehož varianty nabývají číselných hodnot
 - **Kvalitativní znak** – znak, jehož varianty se liší kvalitou (může jít i o číselné hodnoty – např. známka z matematiky)
 - **Alternativní znak** – kvalitativní znak, který nabývá pouze dvou možných variant

Popisná statistika
aneb
Jak jednoduše a přehledně prezentovat výsledky šetření?

Část I.
Kvalitativní znak

Popisná statistika – kvalitativní znak

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	kancelářská práce a na ní navazující práce manuální při realizaci	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během		
1.4.2016 10:45	žena	166	56	ano	Různě, 2-3 týdně	Hlídní dětí		
1.4.2016 10:52	žena	188	70	ano	3 dny v týdnu	praxe v oboru během		

Brigáda při studiu

5. Jak často brigádu máte? *

Označte jen jednu elipsu.

- každý pracovní den
 pouze o víkendu
 nepravidelně
 Jiné:

6. Jak byste svou brigádu charakterizoval(a)? *

Označte jen jednu elipsu.

- praxe v oboru během studia
 "kancelářská" práce mimo obor mého studia
 manuální práce
 Jiné:

Popisná statistika – kvalitativní znak

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	kancelářská práce a na ní navazující práce manuální při realizaci	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během		
1.4.2016 10:45	žena	166	56	ano	Různě, 2-3 týdně	Hlídní dětí		
1.4.2016 10:52	žena	188	70	ano	3 dny v týdnu	praxe v oboru během		

- POZOR na nutnost **čištění databáze!**

Brigáda při studiu

5. Jak často brigádu máte? *

Označte jen jednu elipsu.

- každý pracovní den
 pouze o víkendu
 nepravidelně
 Jiné:

6. Jak byste svou brigádu charakterizoval(a)? *

Označte jen jednu elipsu.

- praxe v oboru během studia
 "kancelářská" práce mimo obor mého studia
 manuální práce
 Jiné:

Popisná statistika – kvalitativní znak

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	jinak	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během		
1.4.2016 10:45	žena	166	56	ano	jinak	jinak		
1.4.2016 10:52	žena	188	70	ano	jinak	praxe v oboru během		

- POZOR na nutnost **čištění databáze!**

Brigáda při studiu

5. Jak často brigádu máte? *

Označte jen jednu elipsu.

- každý pracovní den
 pouze o víkendu
 nepravidelně
 Jiné:

6. Jak byste svou brigádu charakterizoval(a)? *

Označte jen jednu elipsu.

- praxe v oboru během studia
 "kancelářská" práce mimo obor mého studia
 manuální práce
 Jiné:

Popisná statistika – kvalitativní znak

- Rozdělení četnosti

TABULKA ROZDĚLENÍ ČETNOSTI		
Varianty x_i	Absolutní četnosti n_i	Relativní četnosti p_i
x_1	n_1	$p_1 = n_1/n$
x_2	n_2	$p_2 = n_2/n$
x_k	n_k	$p_k = n_k/n$
Celkem:	$n_1 + n_2 + \dots + n_k = n$	1

+ **Modus** (název nejčetnější varianty)

Popisná statistika – kvalitativní znak

- Rozdělení četnosti

TABULKA ROZDĚLENÍ ČETNOSTI		
Pohlaví	Absolutní četnosti	Relativní četnosti
muž	66	0,776470588
žena	19	0,223529412
celkem	85	1,000000000

Popisná statistika – kvalitativní znak

▪ Rozdělení četnosti

TABULKA ROZDĚLENÍ ČETNOSTI		
Pohlaví	Absolutní četnosti	Relativní četnosti (%)
muž	66	77,64705882
žena	19	22,35294118
celkem	85	100,00000000

Jak zaokrouhlovat relativní četnost?

1% ... 0,85 osob
0,1% ... 0,085 osob



Popisná statistika – kvalitativní znak

▪ Rozdělení četnosti

TABULKA ROZDĚLENÍ ČETNOSTI		
Pohlaví	Absolutní četnosti	Relativní četnosti (%)
muž	66	78
žena	19	22
celkem	85	100

} POZOR na
zaokrouhlovací chybu!

Jak zaokrouhlovat relativní četnost?

1% ... 0,85 osob
0,1% ... 0,085 osob



Popisná statistika – kvalitativní znak

▪ Rozdělení četnosti

TABULKA ROZDĚLENÍ ČETNOSTI		
Pohlaví	Absolutní četnosti	Relativní četnosti (%)
muž	66	78
žena	19	22
celkem	85	100

POZOR na
zaokrouhlovací chybu!
Dopočet do 100 %!

Jak zaokrouhlovat relativní četnost?

1 % ... 0,85 osob
0,1 % ... 0,085 osob



Popisná statistika – kvalitativní znak

- Rozdělení četnosti

TABULKA ROZDĚLENÍ ČETNOSTI		
Pohlaví	Absolutní četnosti	Relativní četnosti (%)
muž	?	78
žena	?	22
celkem	85	100

Relativní četnosti uvádějme vždy pouze jako doplněk absolutních četností, nikoliv samostatně!



Popisná statistika – kvalitativní znak

- Rozdělení četnosti

TABULKA ROZDĚLENÍ ČETNOSTI		
Pohlaví	Absolutní četnosti	Relativní četnosti (%)
muž	66	78
žena	19	22
celkem	85	100

- Modus = muž

(Mezi respondenty převažovali muži.)

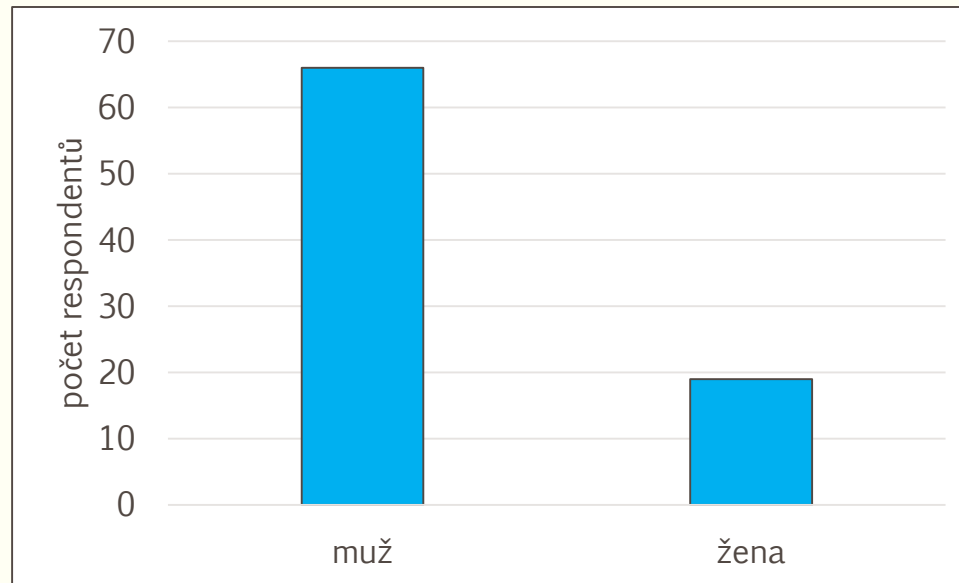
Popisná statistika – kvalitativní znak

- Jak výsledky vizualizovat?

TABULKA ROZDĚLENÍ ČETNOSTI		
Pohlaví	Absolutní četnosti	Relativní četnosti (%)
muž	66	78
žena	19	22
celkem	85	100

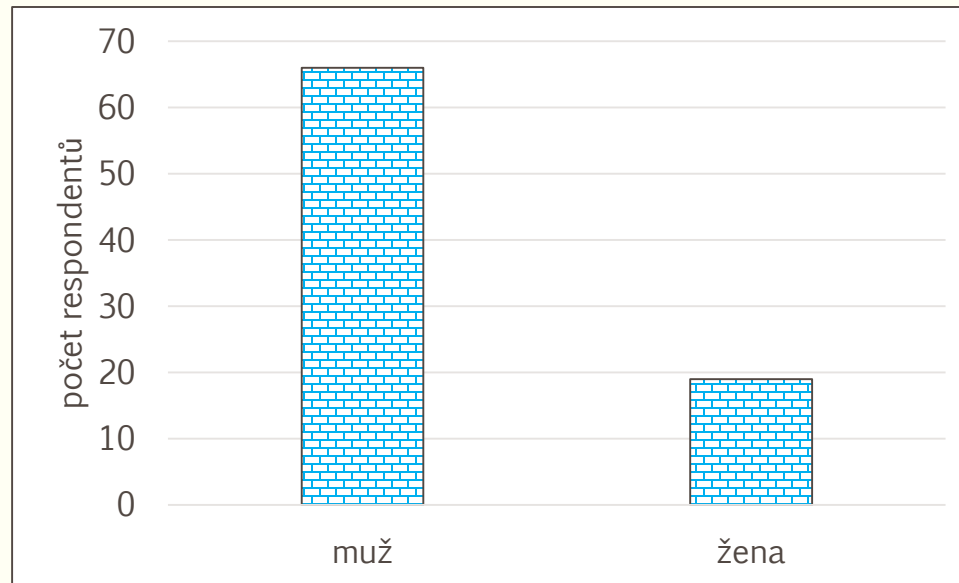
Popisná statistika – kvalitativní znak

- Sloupcový graf (Bar Chart)



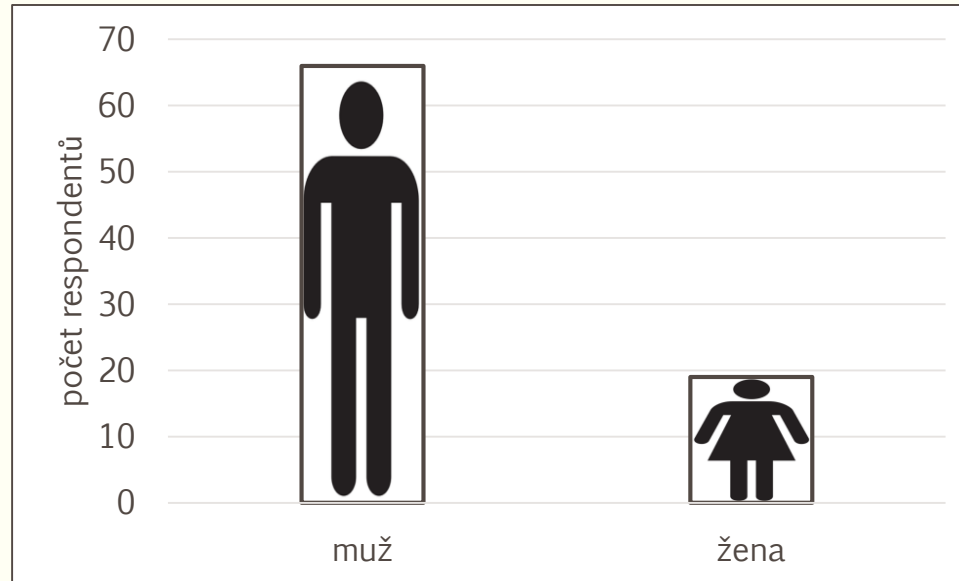
Popisná statistika – kvalitativní znak

- Sloupcový graf (Bar Chart)



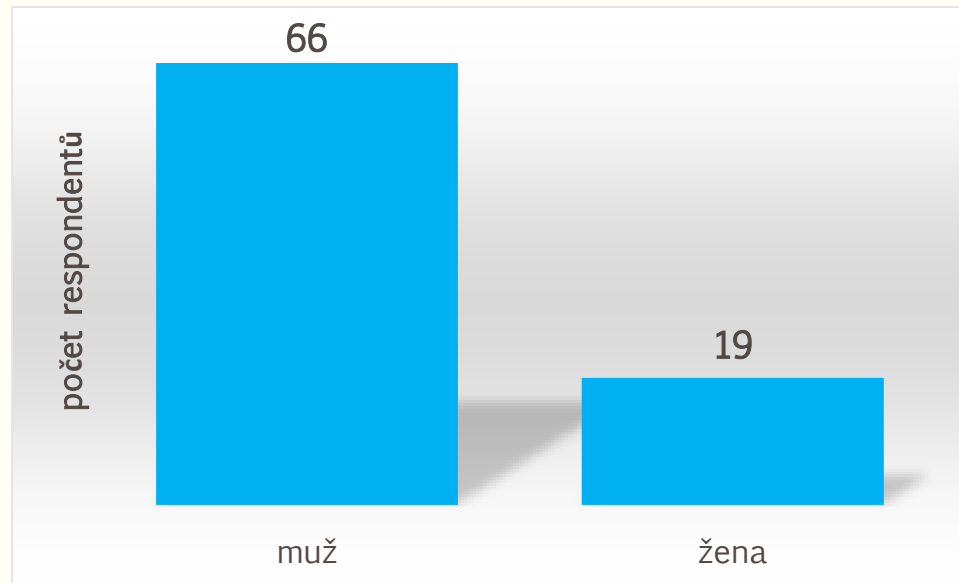
Popisná statistika – kvalitativní znak

- Sloupcový graf (Bar Chart)



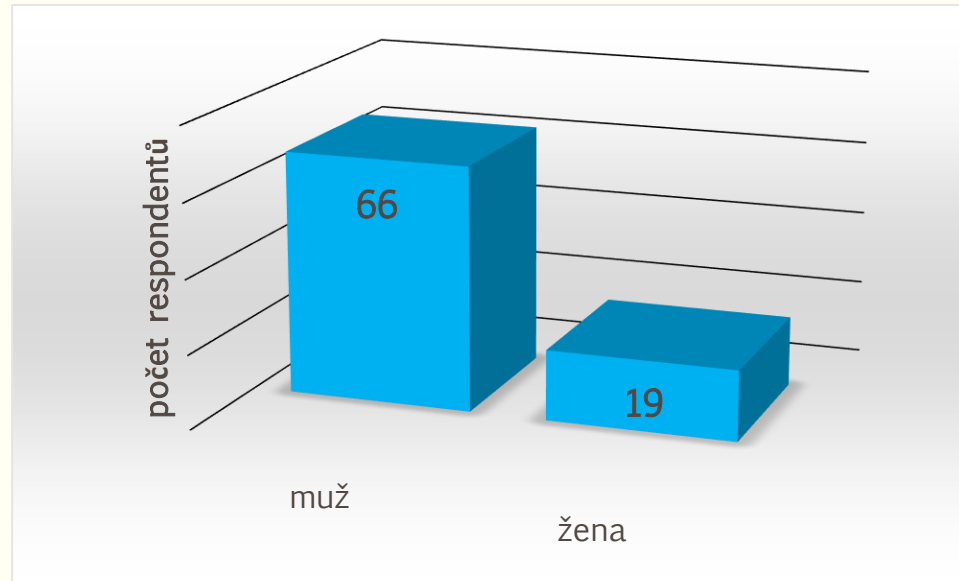
Popisná statistika – kvalitativní znak

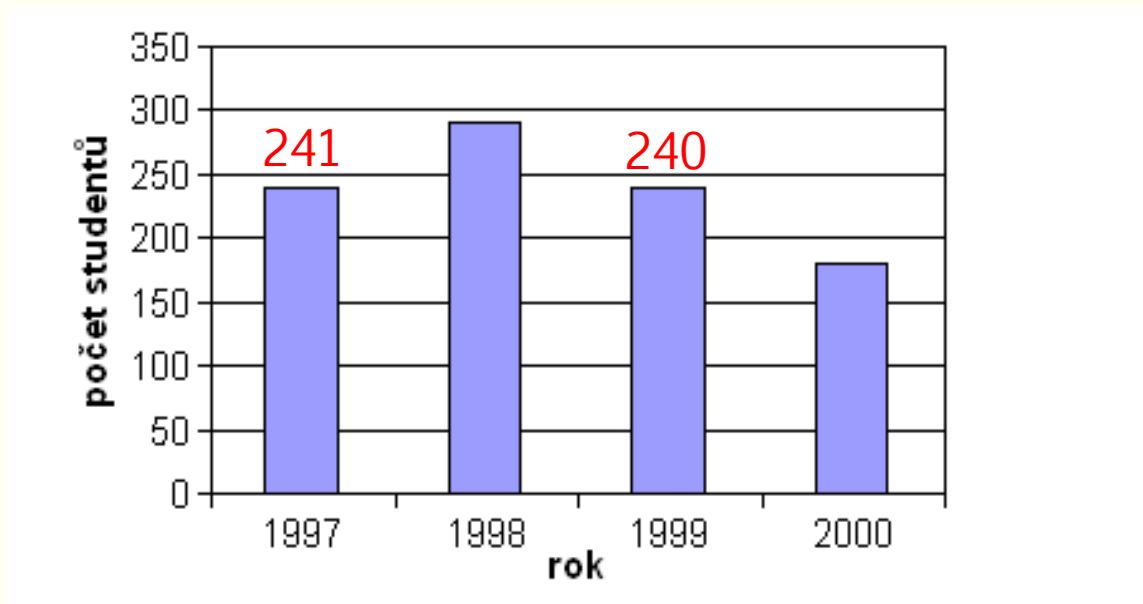
- Sloupcový graf (Bar Chart)



Popisná statistika – kvalitativní znak

- Sloupcový graf (Bar Chart)



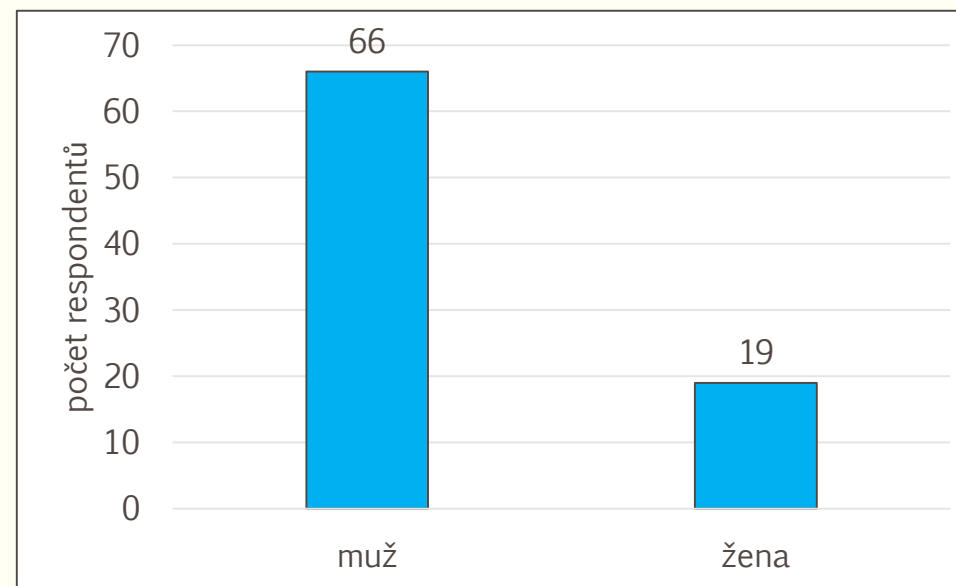
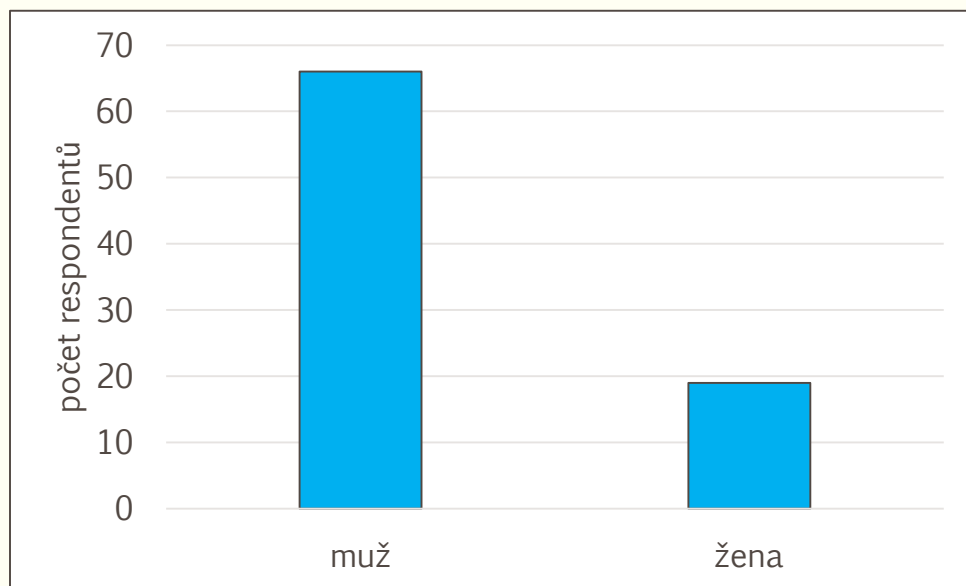


Určete pravdivost tvrzení:
V žádných dvou letech nebyl počet studentů stejný.

Zdroj: Testové příklady určené žákům 9. tříd.

Popisná statistika – kvalitativní znak

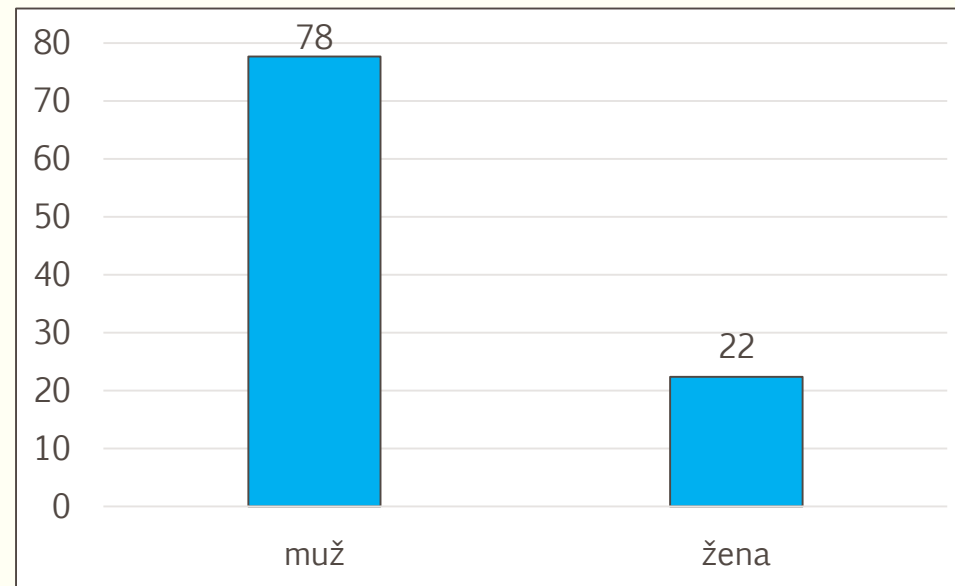
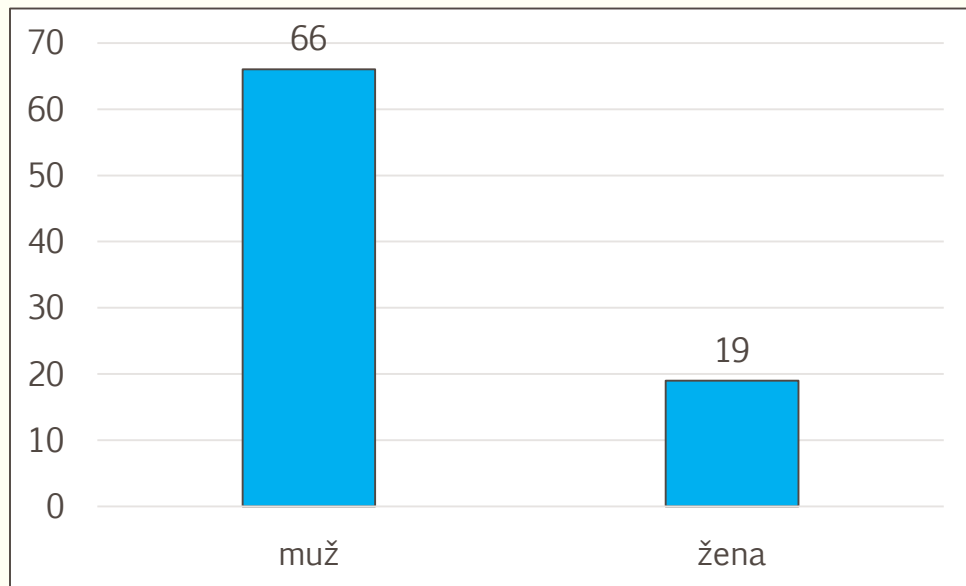
- Sloupcový graf (Bar Chart)



- Nejsou-li v grafu uvedeny absolutní četnosti, obvykle je nedokážeme „od oka“ přesně odečíst.

Popisná statistika – kvalitativní znak

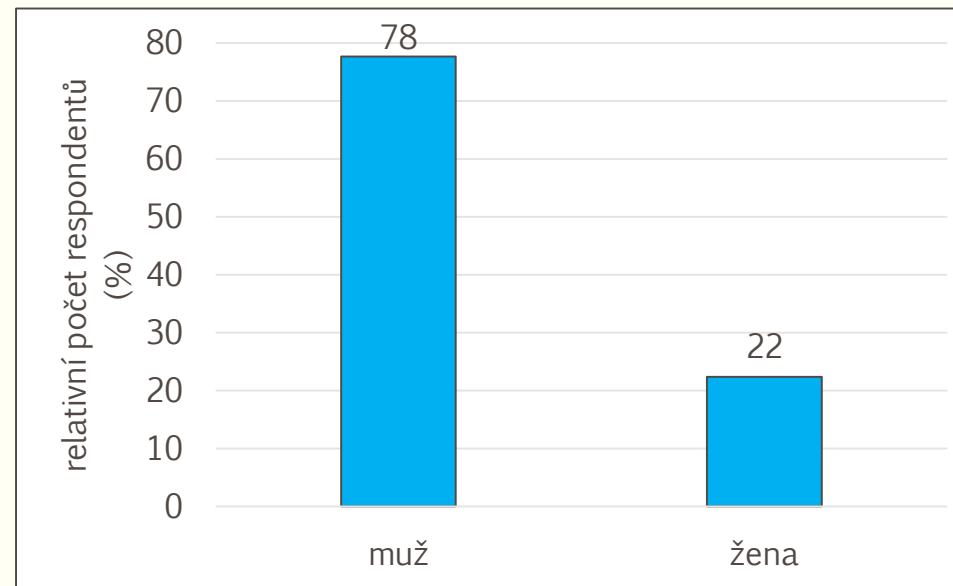
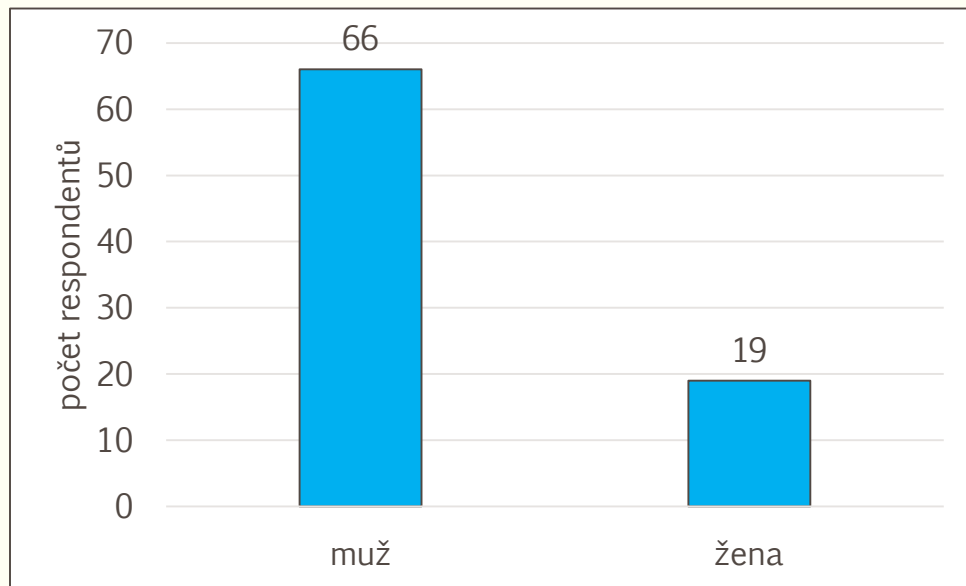
- Sloupcový graf (Bar Chart)



- Nejsou-li v grafu uvedeny absolutní (relativní) četnosti, obvykle je nedokážeme „od oka“ přesně odečíst.

Popisná statistika – kvalitativní znak

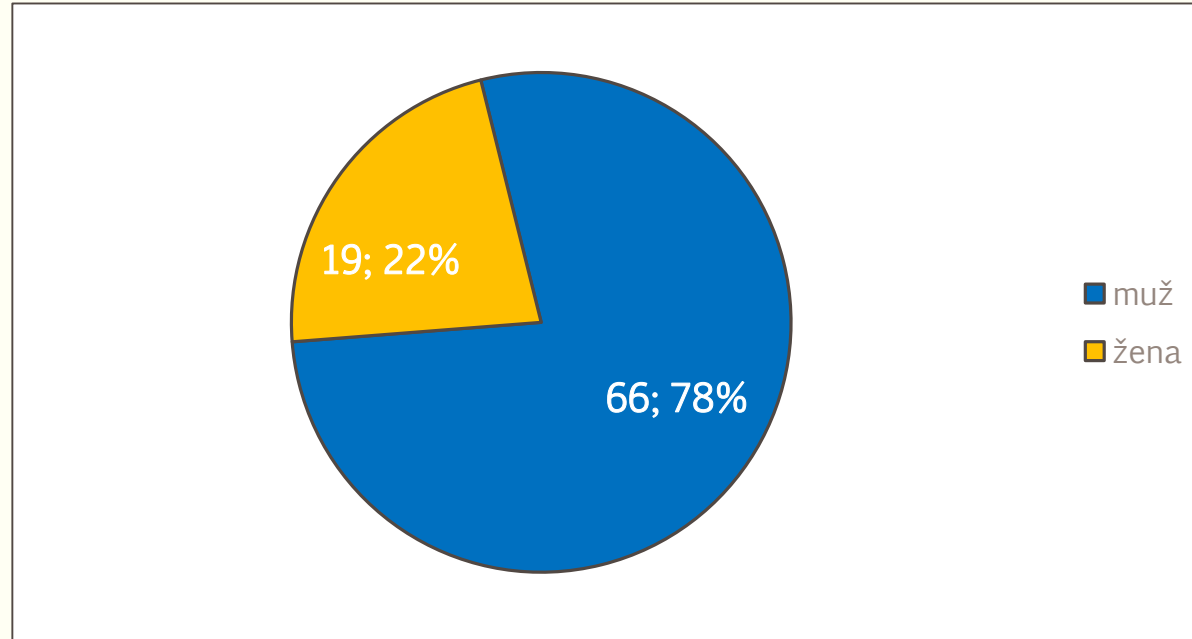
- Sloupcový graf (Bar Chart)



- Nejsou-li v grafu uvedeny absolutní (relativní) četnosti, obvykle je nedokážeme „od oka“ přesně odečíst.
- Pozor na uvádění popisu os!

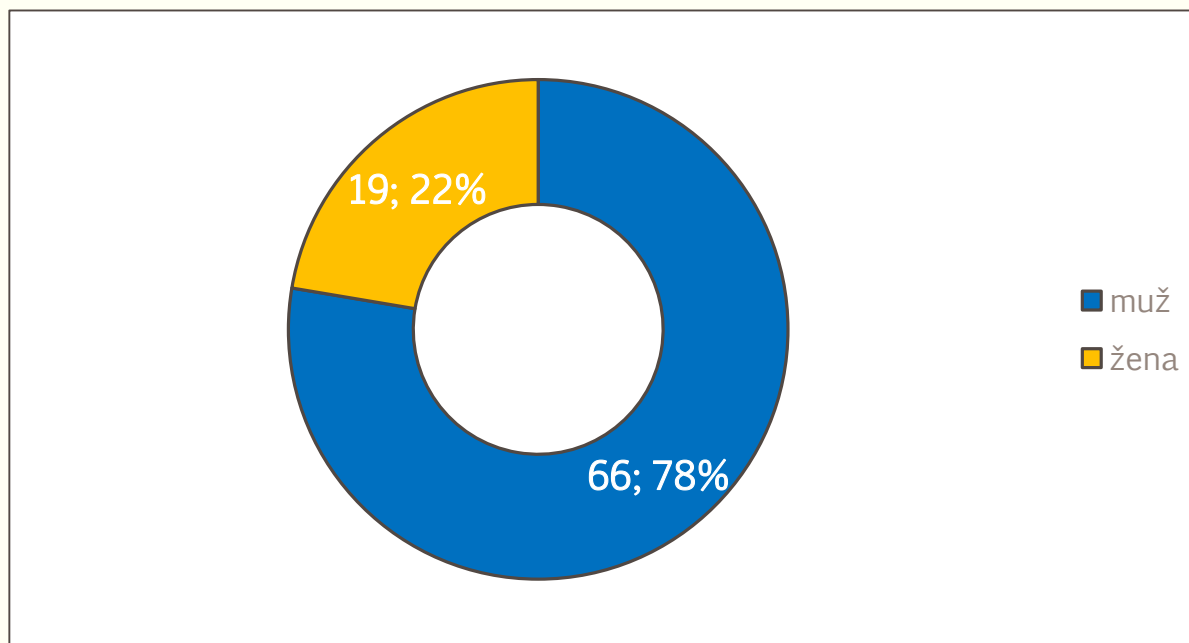
Popisná statistika – kvalitativní znak

- Výsečový graf



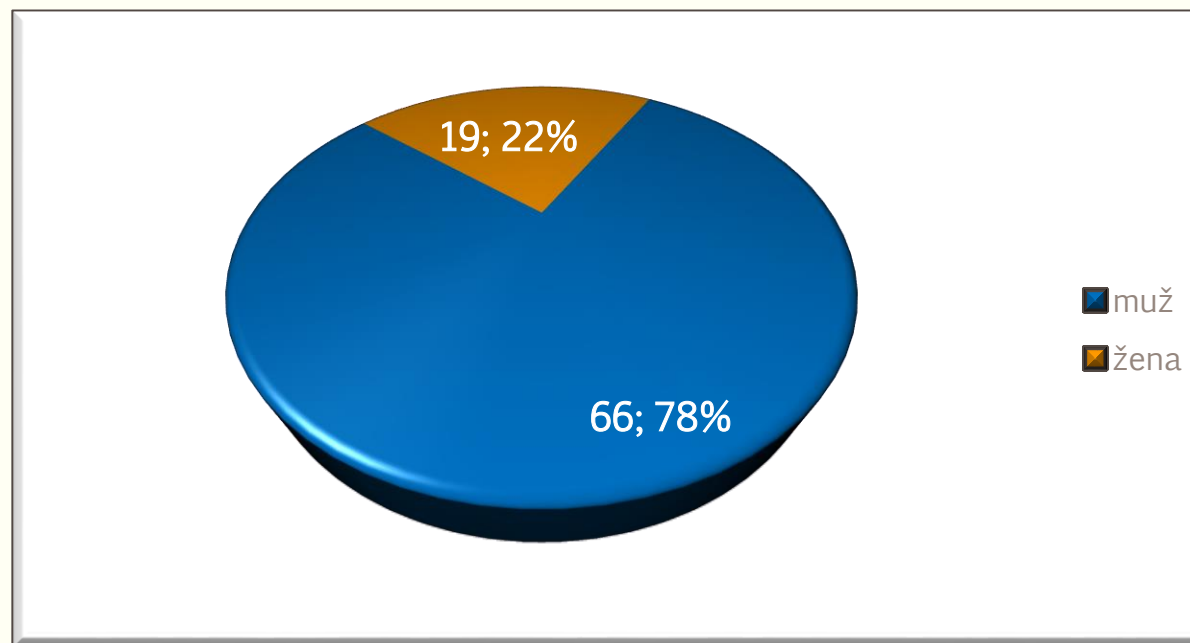
Popisná statistika – kvalitativní znak

- Prstencový graf



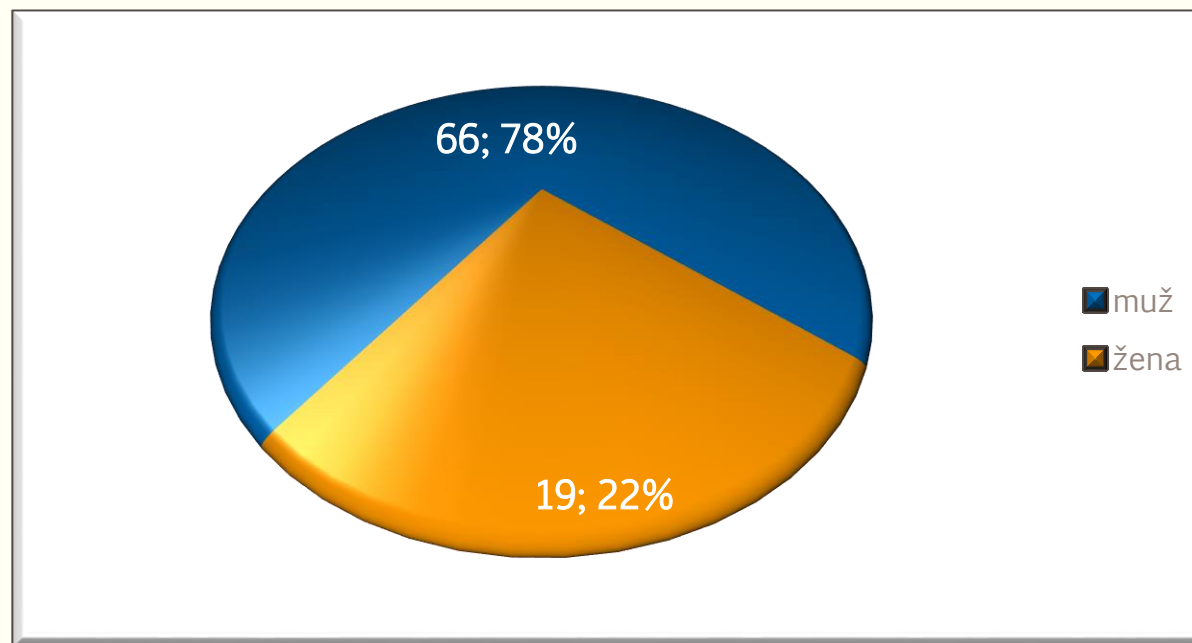
Popisná statistika – kvalitativní znak

- Výsečový graf



Popisná statistika – kvalitativní znak

- Výsečový graf

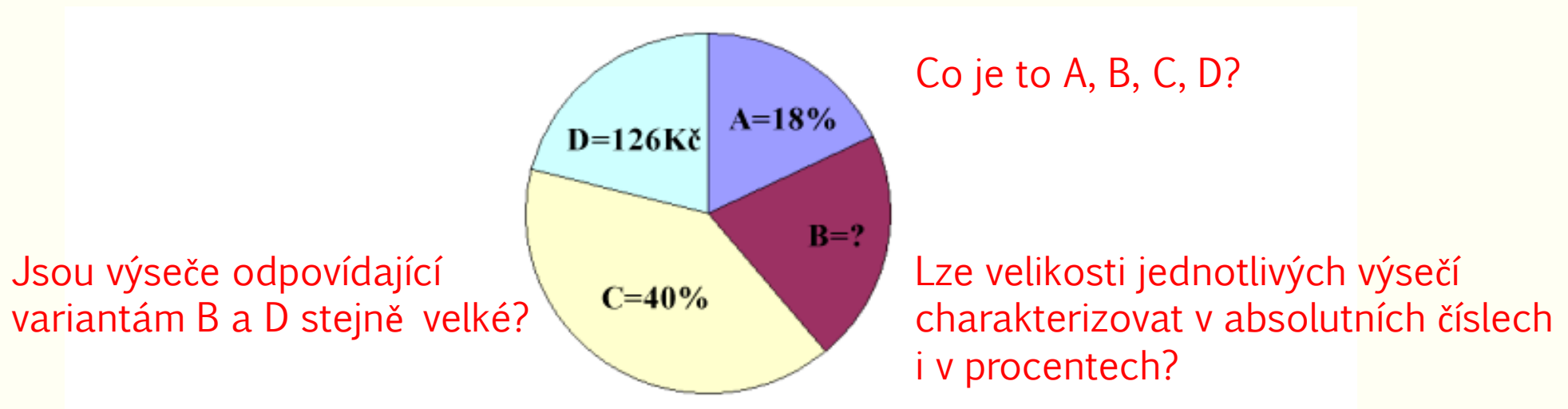


Anketa

Jste pro navýšení hodinové dotace matematiky?



Zdroj: Testové příklady určené žákům 9. tříd.



Určete pravdivost tvrzení:

- a) Místo otazníku patří 20%.
- b) Místo otazníku patří 126 Kč.
- c) Část C je dvojnásobkem části D.

Rozdělení četností **kvalitativního znaku** se znázorňuje kruhovým diagramem, kde různým hodnotám znaku odpovídají kruhové výseče, jejichž plošné obsahy jsou úměrné četnostem. (Prometheus)

Popisná statistika
aneb
Jak jednoduše a přehledně prezentovat výsledky šetření?

Část II.
Kvantitativní znak

Popisná statistika – kvantitativní znak

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	jinak	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během studia	5	36
1.4.2016 10:45	žena	166	56	ano	jinak	jinak	12	10
1.4.2016 10:52	žena	188	70	ano	jinak	praxe v oboru během studia	24	26

Popisná statistika – kvantitativní znak

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	jinak	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během studia	5	36
1.4.2016 10:45	žena	166	56	ano	jinak	jinak	12	10
1.4.2016 10:52	žena	188	70	ano	jinak	praxe v oboru během studia	24	26

- Míry polohy (aritmetický průměr, kvantily)
- Míry variability (rozptyl, směrodatná odchylka, variační koeficient)
- Odlehlá pozorování

Popisná statistika – kvantitativní znak

- **Míry polohy**

- Aritmetický průměr: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

POZOR!

- Průměr je číslo, které nemusí patřit do definičního oboru analyzovaného znaku.
(např. průměrný počet dětí jedné ženy)

Popisná statistika – kvantitativní znak

- **Míry polohy**

- Aritmetický průměr: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

V malé vesnici někde v Americe žije 6 lidí, jejichž roční plat je uveden níže.

\$25 000 \$27 000 \$29 000
\$35 000 \$37 000 \$38 000

Určete průměrný plat obyvatel této vesnice.

(\$31 830)

Do vesnice se přistěhoval Bill Gates, jehož roční příjem je \$40 000 000.

\$25 000 \$27 000 \$29 000
\$35 000 \$37 000 \$38 000 \$40 000 000

Určete průměrný plat obyvatel této vesnice.

(\$5 741 571)

Popisná statistika – kvantitativní znak

- **Míry polohy**

- Aritmetický průměr: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

POZOR!

- Průměr je číslo, které nemusí patřit do definičního oboru analyzovaného znaku. (např. průměrný počet dětí jedné ženy)
 - Průměr není rezistentní vůči odlehlým pozorováním.

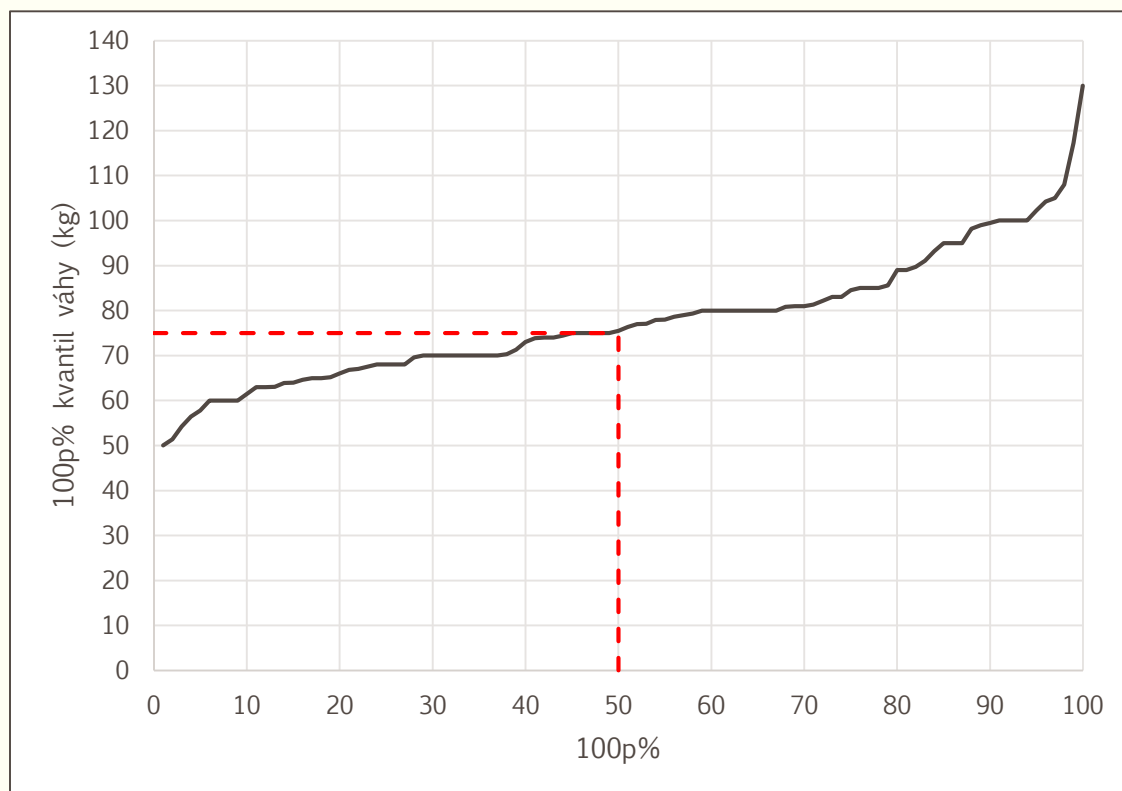
Popisná statistika – kvantitativní znak

▪ Míry polohy

- Aritmetický průměr: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
- Medián *Med* (50% kvantil – 50% hodnot je menších nebo rovných mediánů)
- Dolní kvartil *Q1* (25% kvantil – 25% hodnot je menších nebo rovných dolnímu kvartilu)
- Horní kvartil *Q3* (75% kvantil – 75% hodnot je menších nebo rovných hornímu kvartilu)
- 100p% kvantil – 100p% hodnot je menších nebo rovných 100p% kvantilu

Popisná statistika – kvantitativní znak

■ Míry polohy

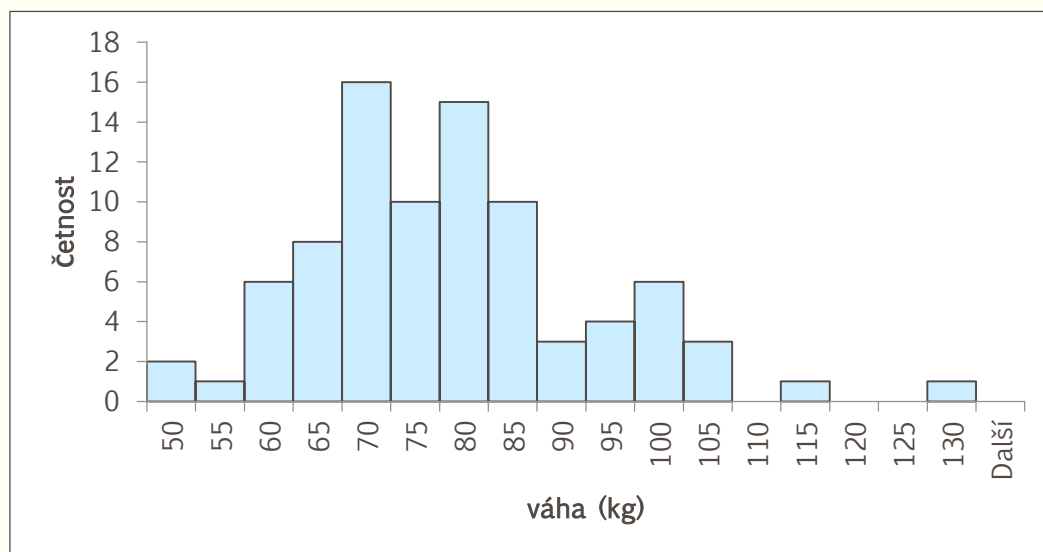


Kvantilová funkce

Míry polohy	Váha (kg)
minimum	50
dolní kvartil	68
průměr	78
medián	76
horní kvartil	85
maximum	130

Popisná statistika – kvantitativní znak

■ Míry polohy

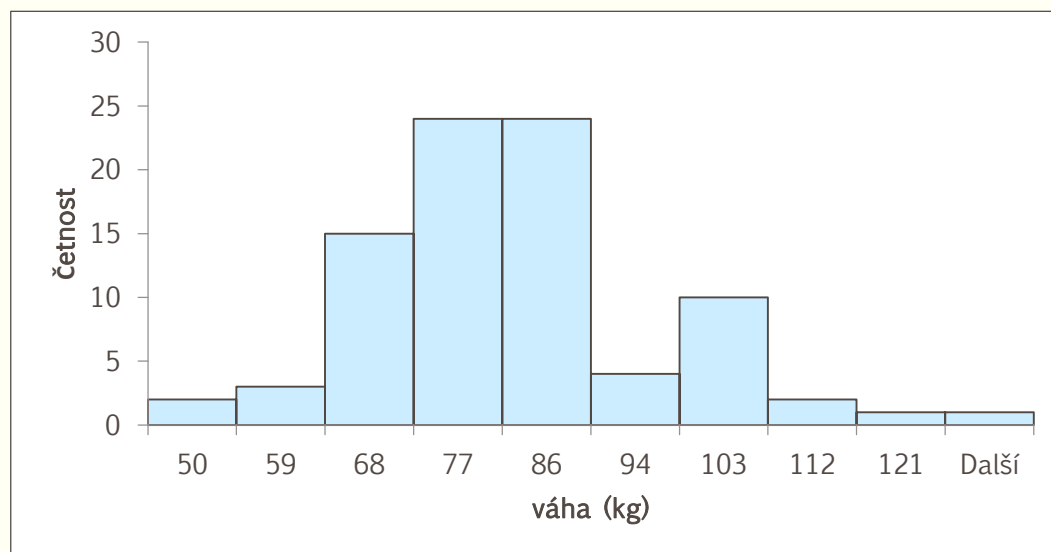


Histogram

Míry polohy	Váha (kg)
minimum	50
dolní kvartil	68
průměr	78
medián	76
horní kvartil	85
maximum	130

Popisná statistika – kvantitativní znak

▪ Míry polohy



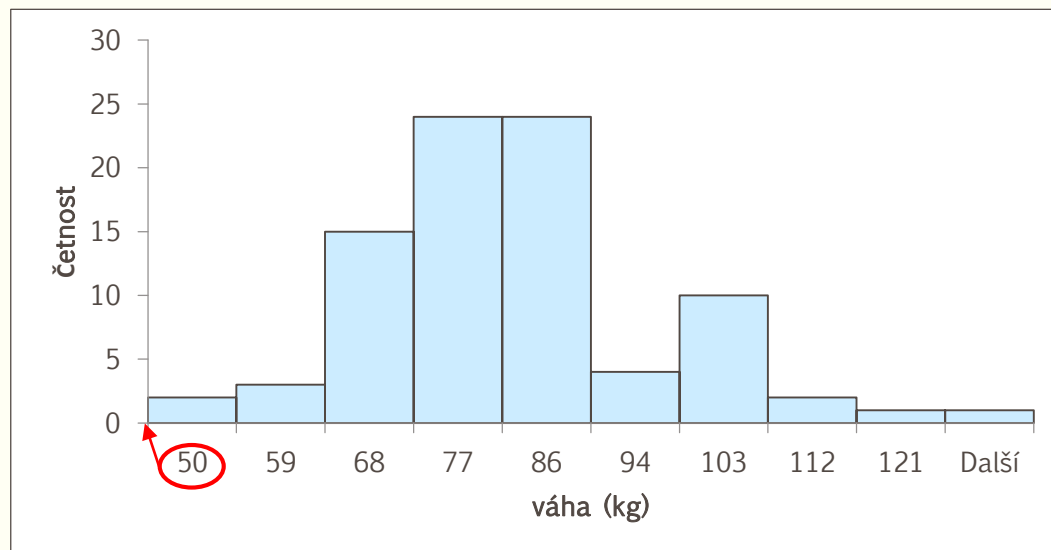
Histogram

Míry polohy	Váha (kg)
minimum	50
dolní kvartil	68
průměr	78
medián	76
horní kvartil	85
maximum	130

Tvar histogramu závisí na počtu tříd („sloupečků“).

Popisná statistika – kvantitativní znak

■ Míry polohy

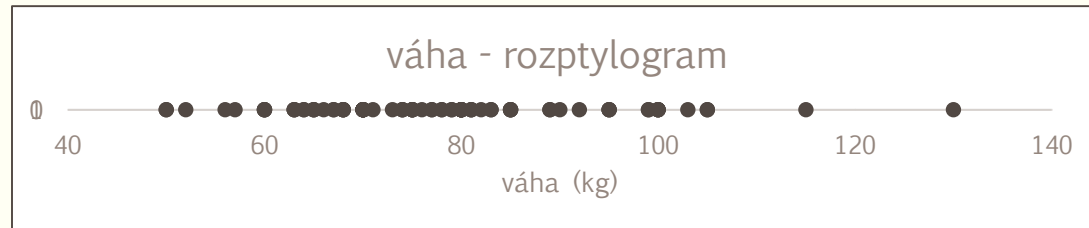


Histogram

Míry polohy	Váha (kg)
minimum	50
dolní kvartil	68
průměr	78
medián	76
horní kvartil	85
maximum	130

Popisná statistika – kvantitativní znak

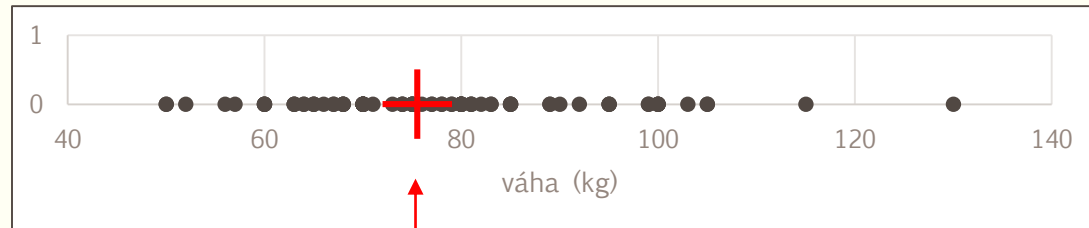
- Odlehlá pozorování



Míry polohy	Váha (kg)
minimum	50
dolní kvartil	68
průměr	78
medián	76
horní kvartil	85
maximum	130

Popisná statistika – kvantitativní znak

- Odlehlá pozorování

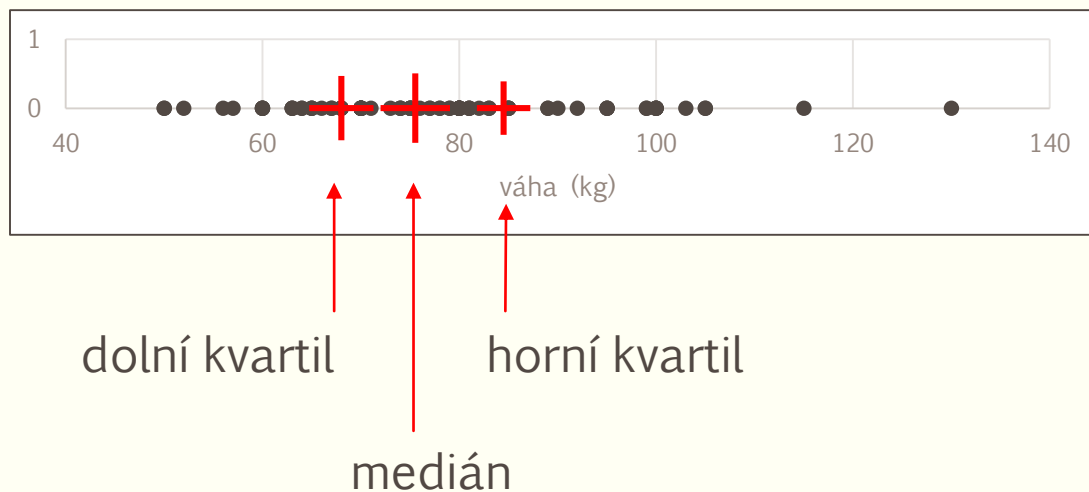


medián

Míry polohy	Váha (kg)
minimum	50
dolní kvartil	68
průměr	78
medián	76
horní kvartil	85
maximum	130

Popisná statistika – kvantitativní znak

▪ Odlehlá pozorování



Míry polohy	Váha (kg)
minimum	50
dolní kvartil	68
průměr	78
medián	76
horní kvartil	85
maximum	130

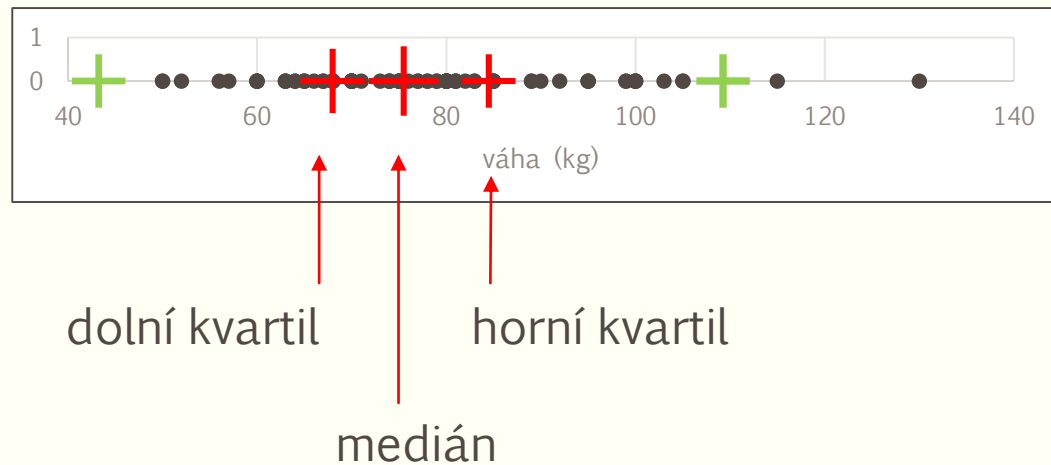
Mezikvartilové rozpětí: $IQR = Q3 - Q1$

$[(x < Q1 - 1,5 \cdot IQR) \vee (x > Q3 + 1,5 \cdot IQR)] \Rightarrow x$ je odlehlé pozorování

vnitřní hrady

Popisná statistika – kvantitativní znak

▪ Odlehlá pozorování



Míry polohy	Váha (kg)
minimum	50
dolní kvartil	68
průměr	78
medián	76
horní kvartil	85
maximum	130

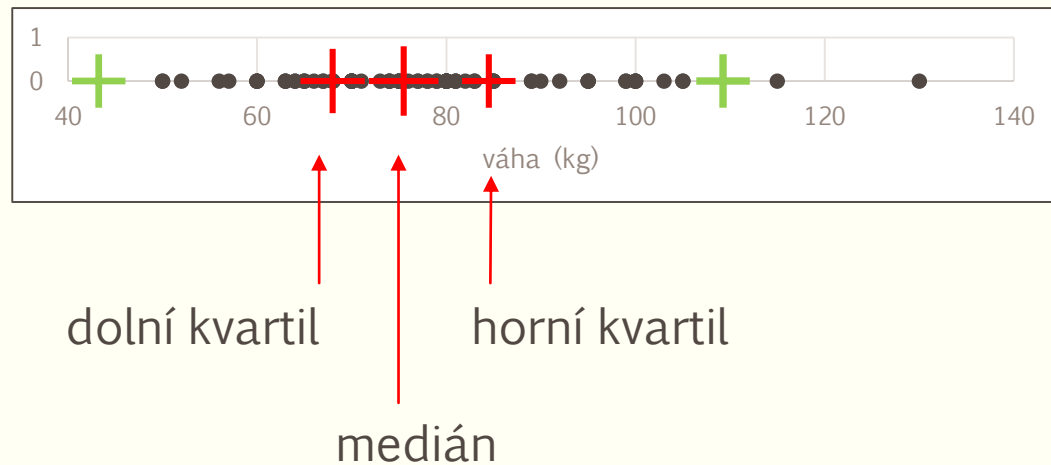
Mezikvartilové rozpětí: $IQR = Q3 - Q1$

$[(x < Q1 - 1,5 \cdot IQR) \vee (x > Q3 + 1,5 \cdot IQR)] \Rightarrow x$ je odlehlé pozorování

vnitřní hradby

Popisná statistika – kvantitativní znak

▪ Odlehlá pozorování



Míry polohy	Váha (kg)	Váha* (kg)
minimum	50	50
dolní kvartil	68	68
průměr	78	77
medián	76	75
horní kvartil	85	83
maximum	130	105

*po odstranění odlehlých pozorování

Mezikvartilové rozpětí: $IQR = Q3 - Q1$

$[(x < Q1 - 1,5 \cdot IQR) \vee (x > Q3 + 1,5 \cdot IQR)] \Rightarrow x$ je odlehlé pozorování

vnitřní hrady

Popisná statistika – kvantitativní znak

▪ Míry variability

- Výběrový rozptyl: $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$
 - POZOR! – Jednotka rozptylu je kvadrátem jednotky analyzovaného znaku.
- Výběrová směrodatná odchylka: $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$
 - Neumožňuje srovnání variability znaků s různými jednotkami.
- Variační koeficient: $V = \frac{s}{|\bar{x}|} \cdot 100 (\%)$
 - Čím nižší var. koeficient, tím homogennější soubor.
 - $V > 50\%$ značí silně rozptýlený soubor.

Popisná statistika – kvantitativní znak

Míry variability	Váha (kg)	Váha* (kg)
rozptyl	215,3893	169,5506311
směrodatná odchylka	14,67615	13,0211609
variační koeficient (%)	18,90576	17,0078917

*po odstranění odlehlých pozorování

Jak zaokrouhlovat výběrové charakteristiky?

Směrodatnou odchylku jakožto míru nejistoty měření zaokrouhlujeme **nahoru** na jednu, maximálně dvě platné cifry a míry polohy (průměr, kvantily...) zaokrouhlujeme tak, aby nejnižší zapsaný řád odpovídal nejnižšímu zapsanému řádu směrodatné odchylky.

Popisná statistika – kvantitativní znak

Jak zaokrouhlovat výběrové charakteristiky?

Směrodatnou odchylku jakožto míru nejistoty měření zaokrouhlujeme **nahoru** na jednu, maximálně dvě platné cifry a míry polohy (průměr, kvantily...) zaokrouhlujeme tak, aby nejnižší zapsaný řád odpovídal nejnižšímu zapsanému řádu směrodatné odchylky.

Míry polohy	Váha (kg)	zaokrouhleno
minimum	50	
dolní kvartil	68	
průměr	77,62791	
medián	75,5	
horní kvartil	84,5	
maximum	130	
Míry variability		
směrodatná odchylka	14,67615	
variační koeficient (%)	18,90576	

Popisná statistika – kvantitativní znak

Jak zaokrouhlovat výběrové charakteristiky?

Směrodatnou odchylku jakožto míru nejistoty měření zaokrouhlujeme **nahoru** na jednu, maximálně dvě platné cifry a míry polohy (průměr, kvantily...) zaokrouhlujeme tak, aby nejnižší zapsaný řád odpovídal nejnižšímu zapsanému řádu směrodatné odchylky.

Míry polohy	Váha (kg)	zaokrouhleno
minimum	50	
dolní kvartil	68	
průměr	77,62791	
medián	75,5	
horní kvartil	84,5	
maximum	130	
Míry variability		
směrodatná odchylka	14,67615	15
variační koeficient (%)	18,90576	

Popisná statistika – kvantitativní znak

Jak zaokrouhlovat výběrové charakteristiky?

Směrodatnou odchylku jakožto míru nejistoty měření zaokrouhlujeme **nahoru** na jednu, maximálně dvě platné cifry a míry polohy (průměr, kvantily...) zaokrouhlujeme tak, aby nejnižší zapsaný řád odpovídal nejnižšímu zapsanému řádu směrodatné odchylky.

Míry polohy	Váha (kg)	zaokrouhleno
minimum	50	50
dolní kvartil	68	68
průměr	77,62791	78
medián	75,5	76
horní kvartil	84,5	85
maximum	130	130
Míry variability		
směrodatná odchylka	14,67615	15
variační koeficient (%)	18,90576	

Popisná statistika – kvantitativní znak

Jak zaokrouhlovat výběrové charakteristiky?

Směrodatnou odchylku jakožto míru nejistoty měření zaokrouhlujeme **nahoru** na jednu, maximálně dvě platné cifry a míry polohy (průměr, kvantily...) zaokrouhlujeme tak, aby nejnižší zapsaný řád odpovídal nejnižšímu zapsanému řádu směrodatné odchylky.

Míry polohy	Váha (kg)	zaokrouhleno
minimum	50	50
dolní kvartil	68	68
průměr	77,62791	78
medián	75,5	76
horní kvartil	84,5	85
maximum	130	130
Míry variability		
směrodatná odchylka	14,67615	15
variační koeficient (%)	18,90576	18,9

zaokrouhlujeme na desetiny %



Popisná statistika
aneb
Jak jednoduše a přehledně prezentovat výsledky šetření?

Část III.
Analýza závislosti dvou kvantitativních znaků

Analýza závislosti dvou kvantitativních znaků

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	jinak	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během studia	5	36
1.4.2016 10:45	žena	166	56	ano	jinak	jinak	12	10
1.4.2016 10:52	žena	188	70	ano	jinak	praxe v oboru během studia	24	26

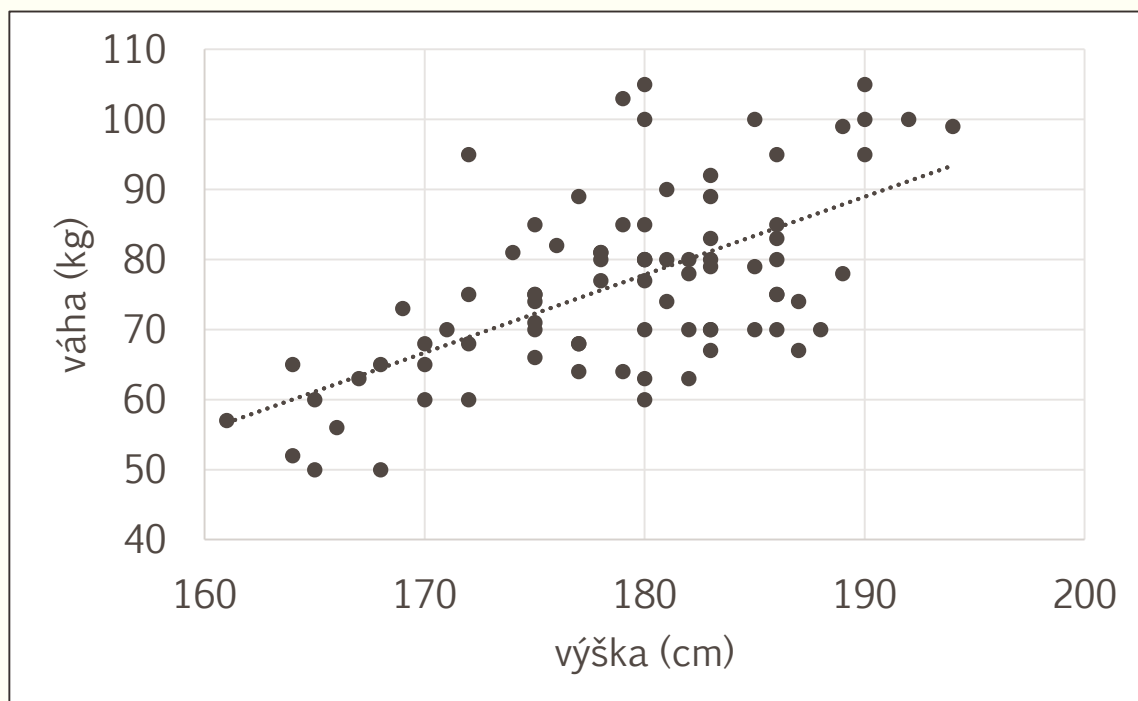
Analýza závislosti dvou kvantitativních znaků

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	jinak	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během studia	5	36
1.4.2016 10:45	žena	166	56	ano	jinak	jinak	12	10
1.4.2016 10:52	žena	188	70	ano	jinak	praxe v oboru během studia	24	26

- Vizualizace – bodový graf
- Korelační koeficient

Analýza závislosti dvou kvantitativních znaků

- Vizualizace – bodový graf

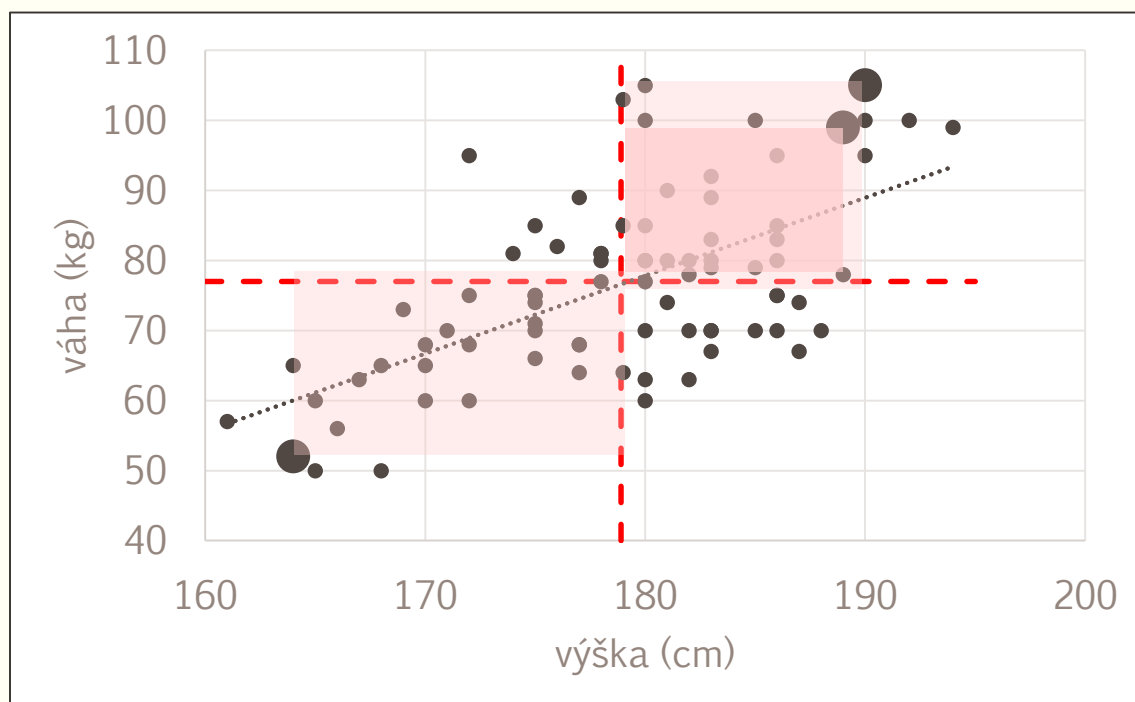


Výběrový korelační koeficient:

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

Analýza závislosti dvou kvantitativních znaků

- Vizualizace – bodový graf

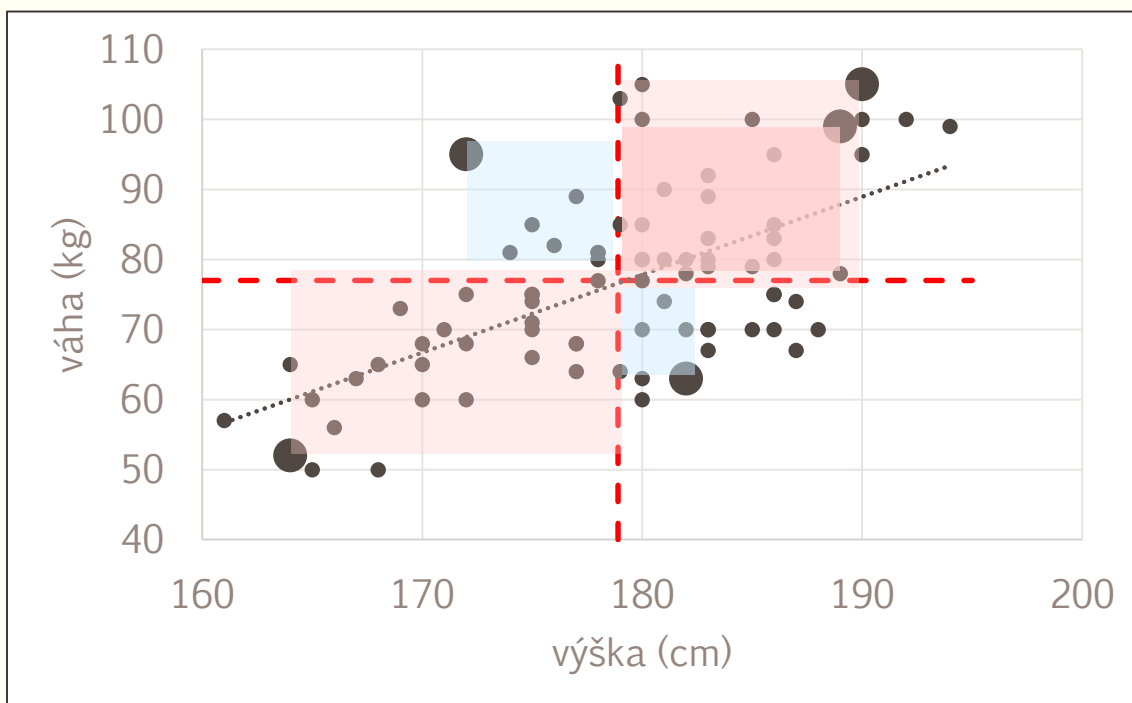


Výběrový korelační koeficient:

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

Analýza závislosti dvou kvantitativních znaků

- Vizualizace – bodový graf



Výběrový korelační koeficient:

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

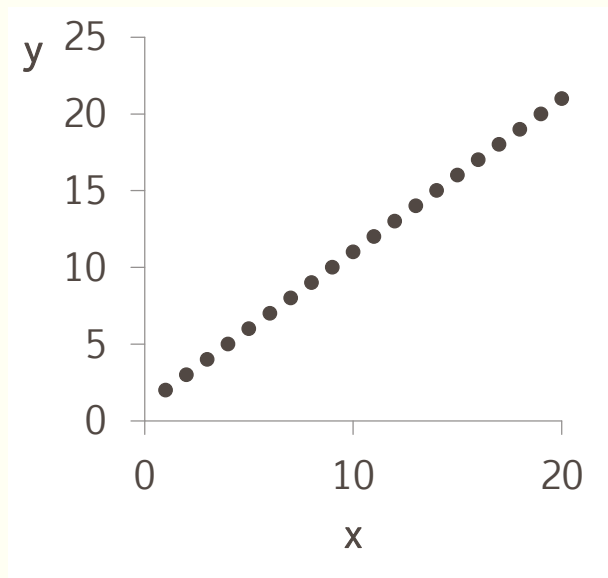
Analýza závislosti dvou kvantitativních znaků

- Vizualizace – bodový graf

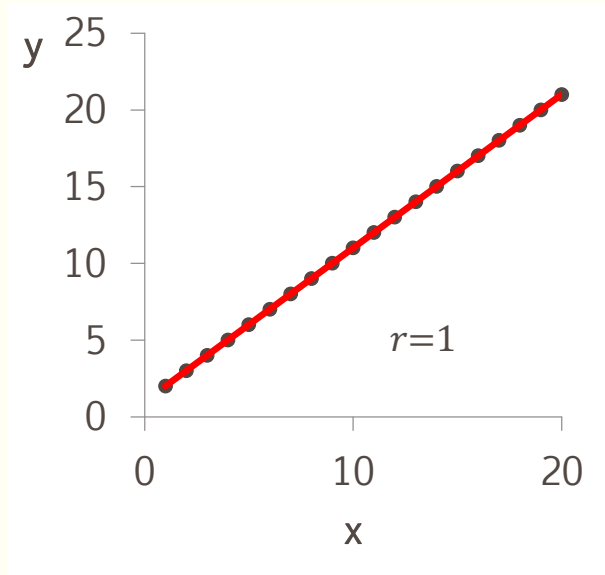
- **Výběrový korelační koeficient:** $r(X, Y) = \frac{1}{n-1} \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_X \cdot s_Y}$

1. $-1 \leq r(X, Y) \leq 1$,
2. $r(X, Y) = r(Y, X)$,
3. $r(X, X) = 1$,
4. je-li $r(X, Y) = 0$, říkáme, že X, Y jsou **nekorelované** náhodné veličiny,
5. je-li $r(X, Y) > 0$, říkáme, že X, Y jsou **pozitivně korelované** (s rostoucím X roste Y),
6. je-li $r(X, Y) < 0$, říkáme, že X, Y jsou **negativně korelované** (s rostoucím X klesá Y),
7. je-li $|r(X, Y)| = 1$, pak je mezi X a Y **lineární závislost**.

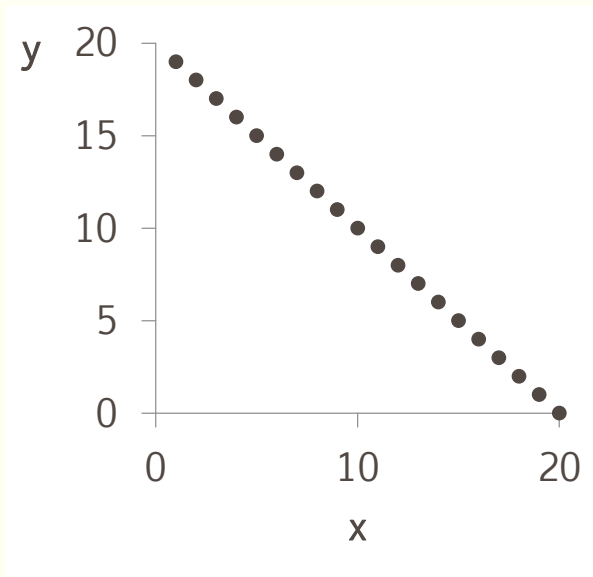
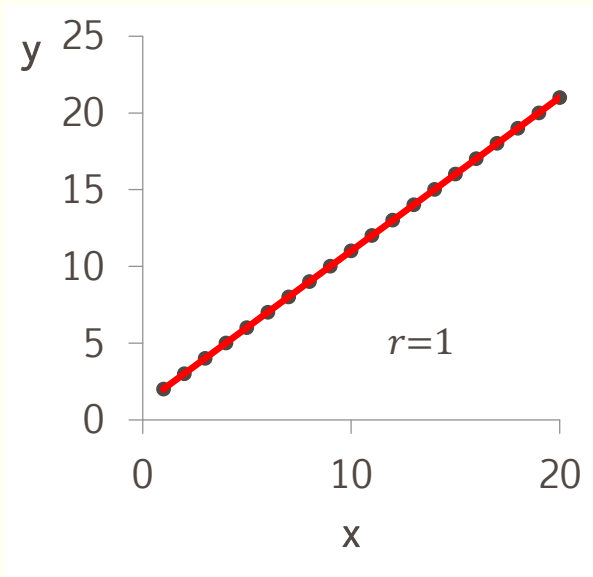
Korelační koeficient



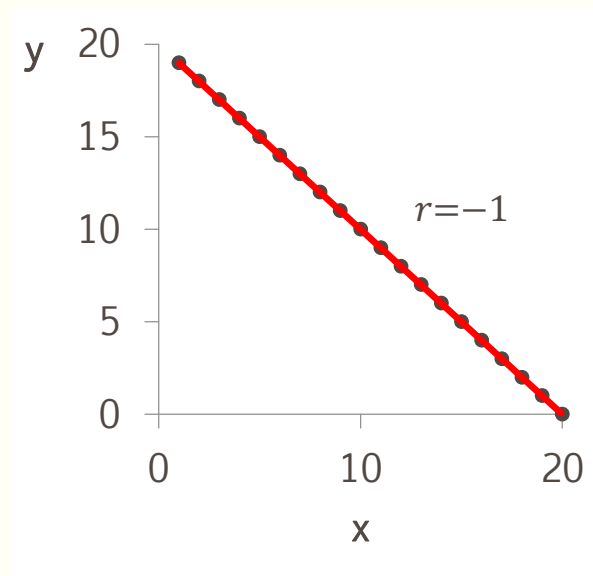
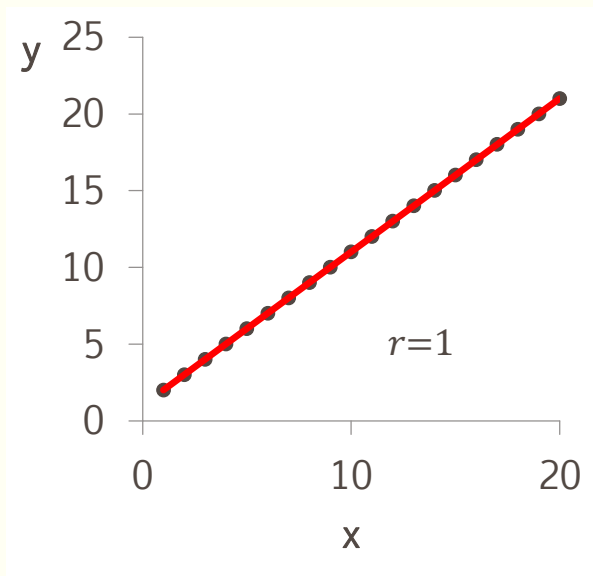
Korelační koeficient



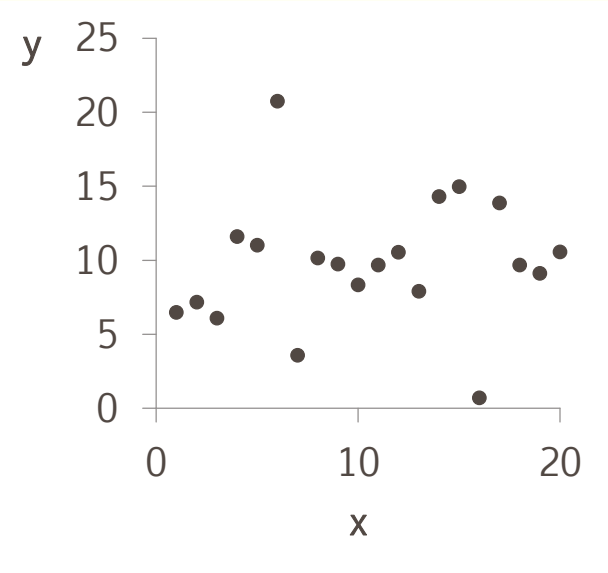
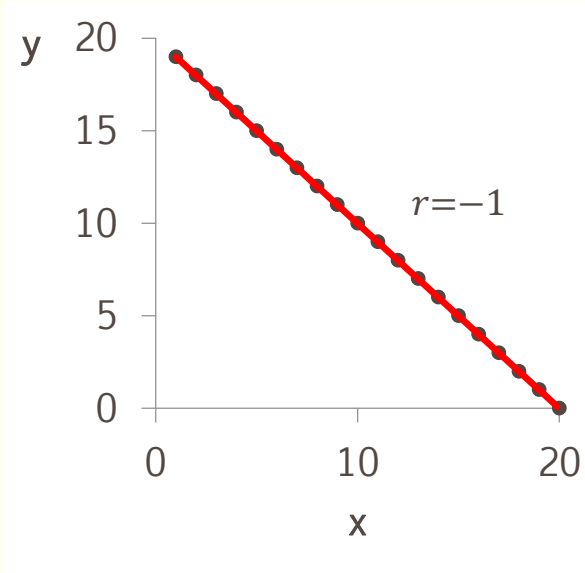
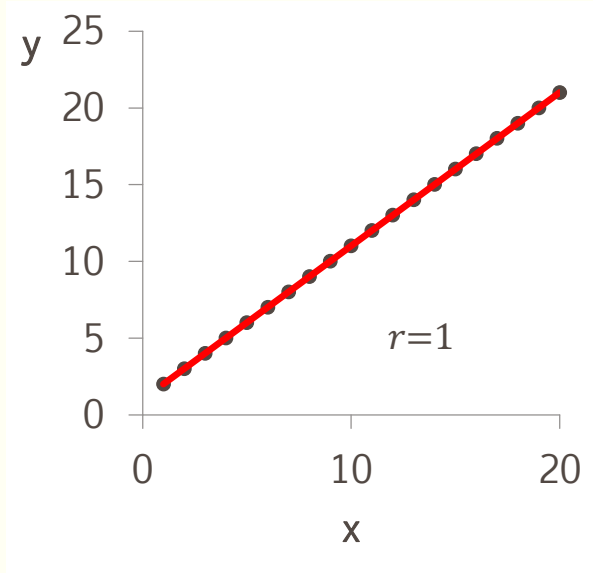
Korelační koeficient



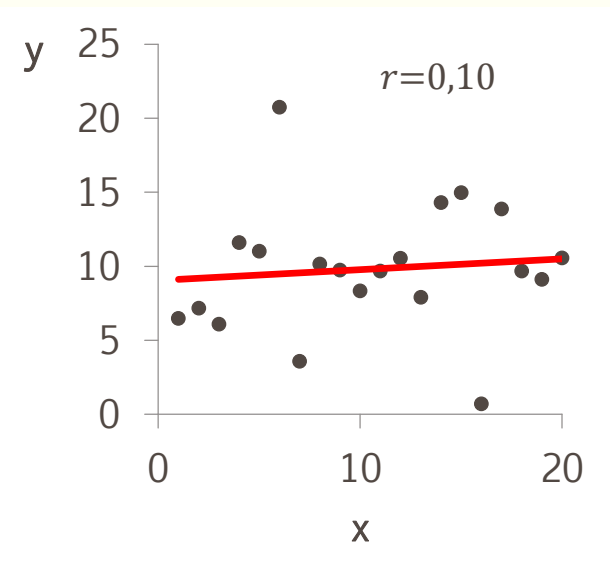
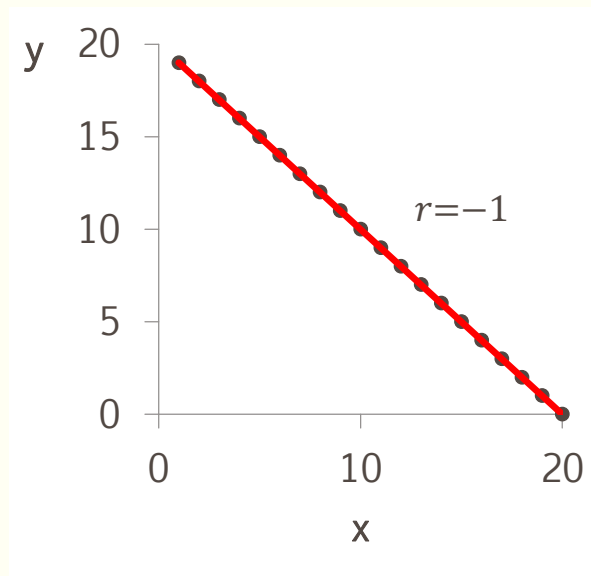
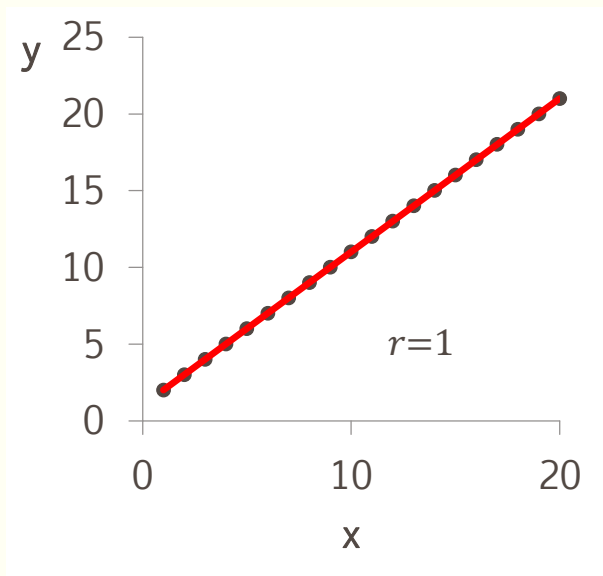
Korelační koeficient



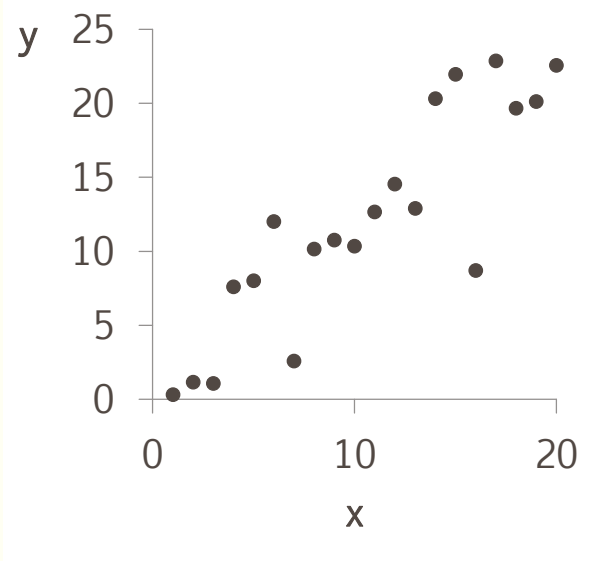
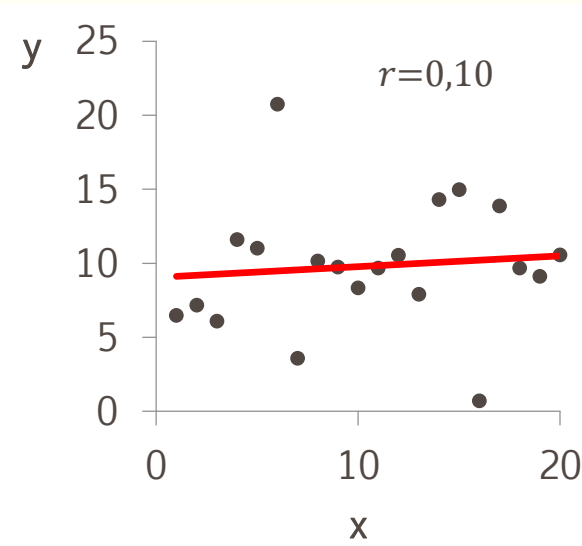
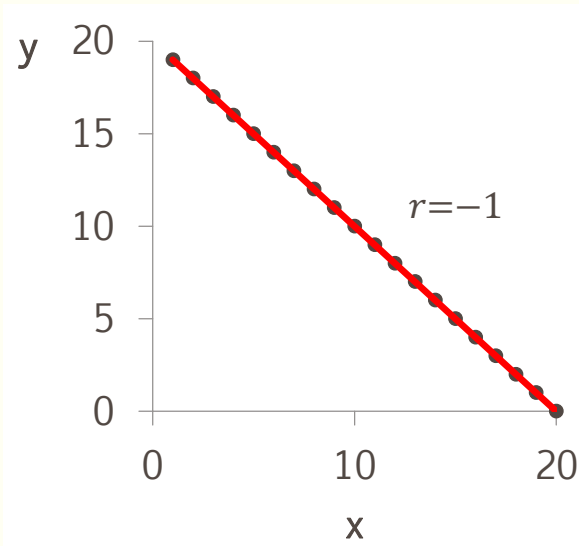
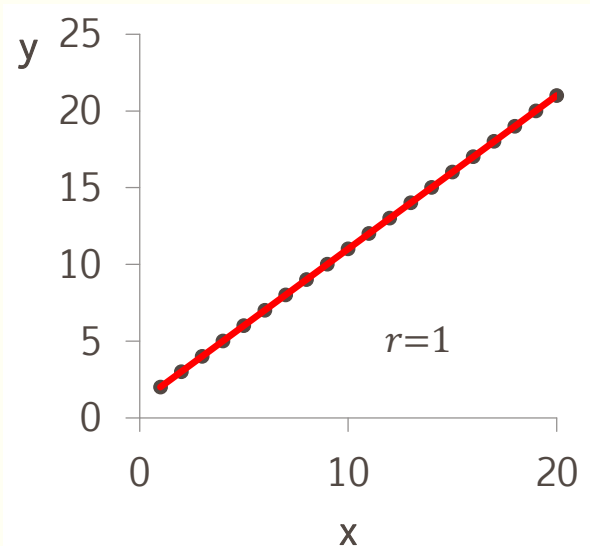
Korelační koeficient



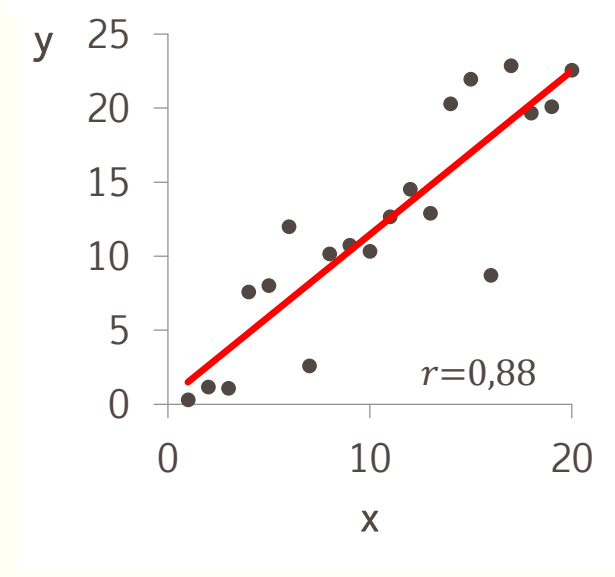
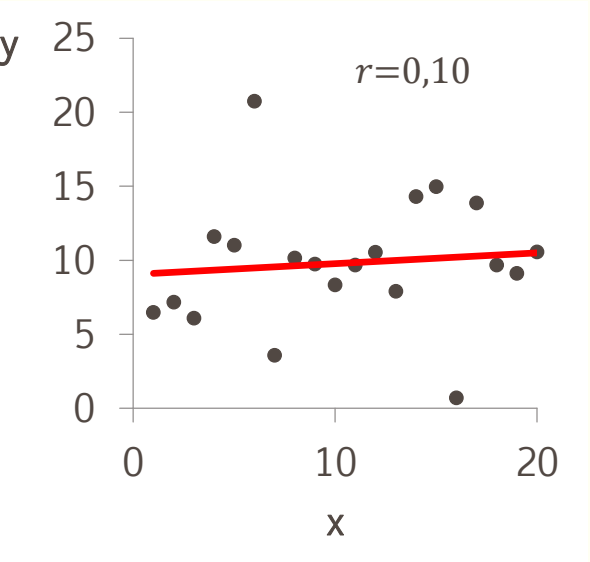
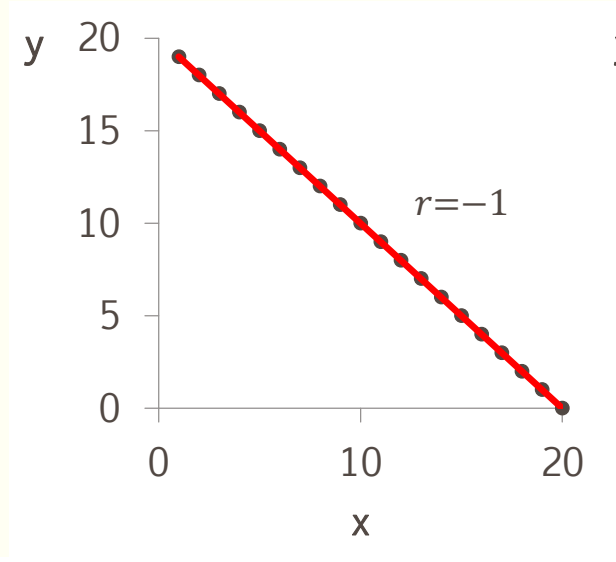
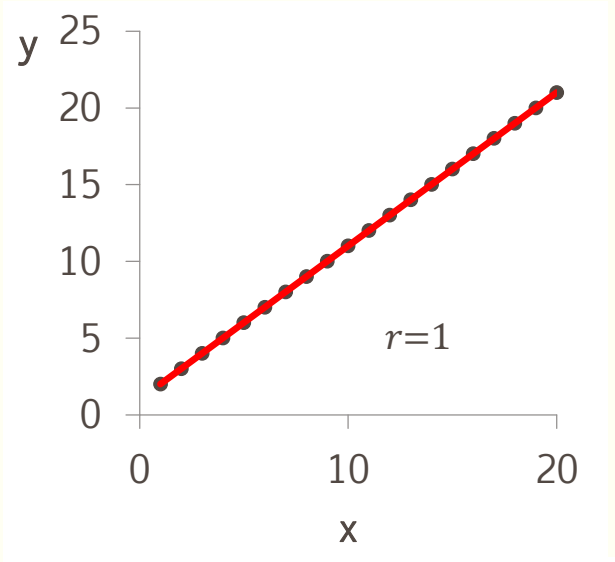
Korelační koeficient



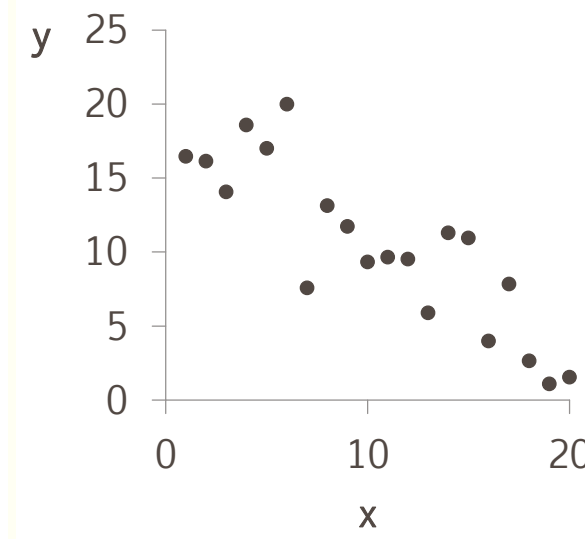
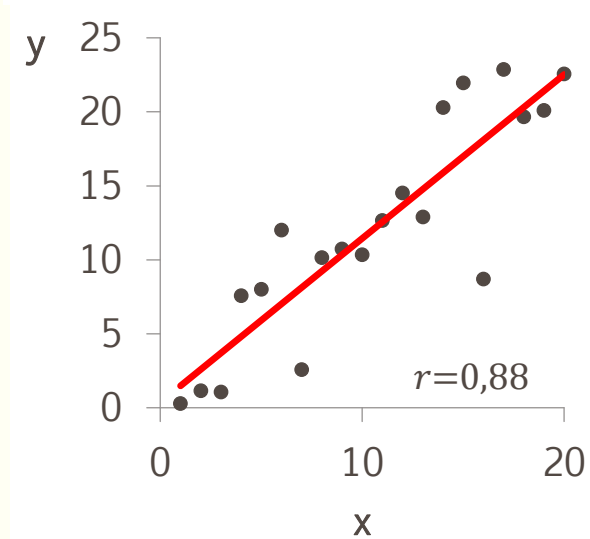
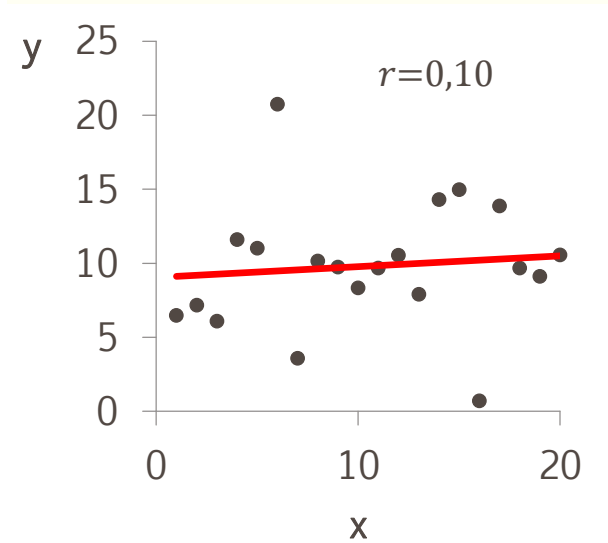
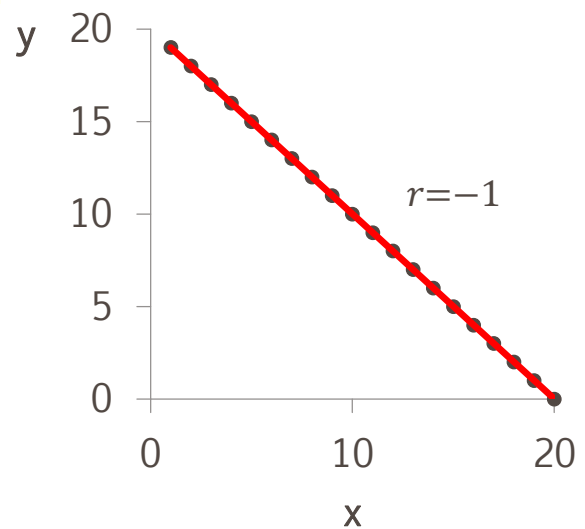
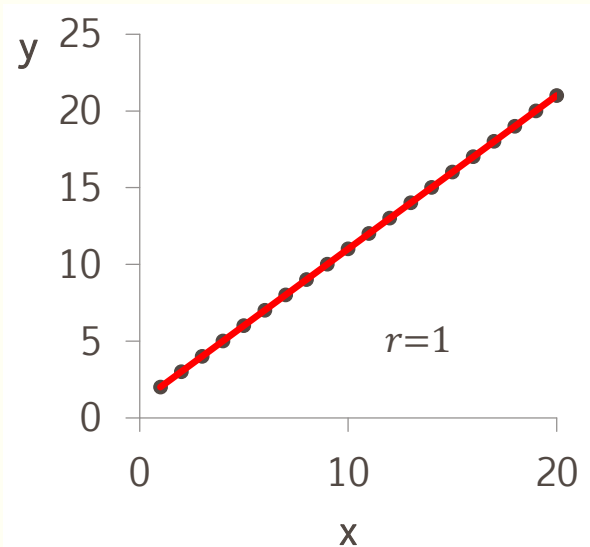
Korelační koeficient



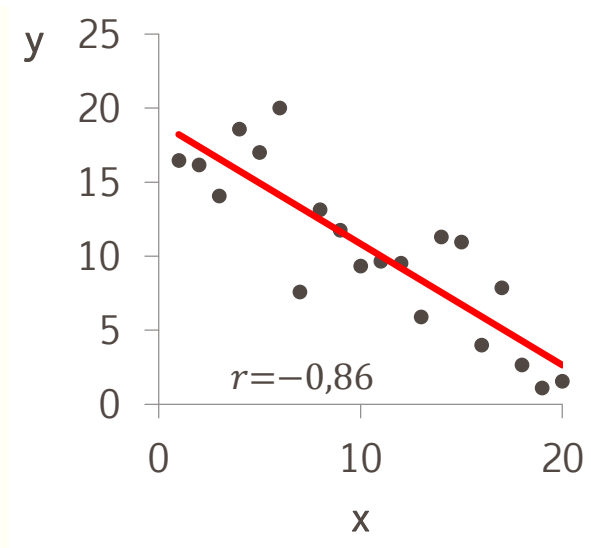
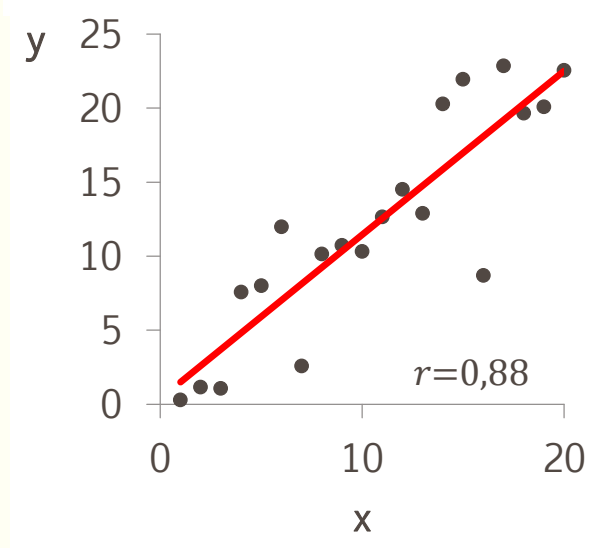
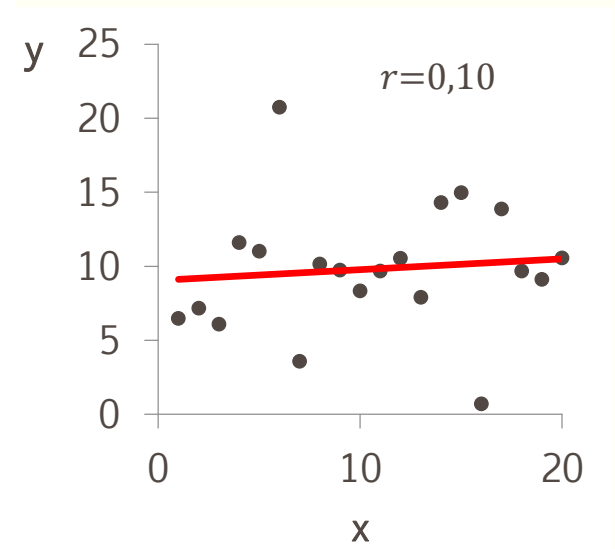
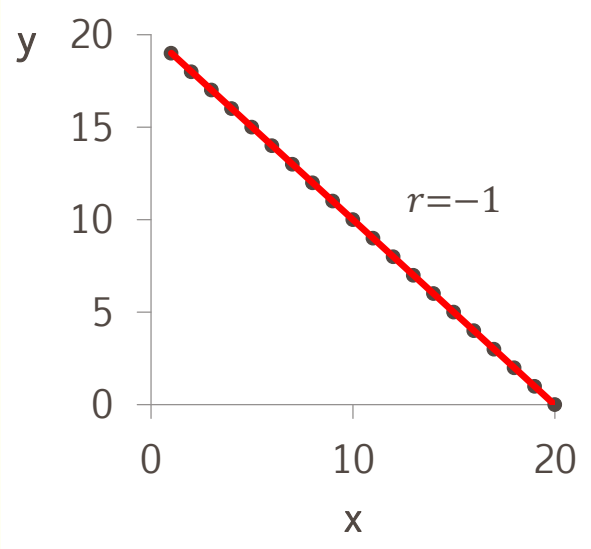
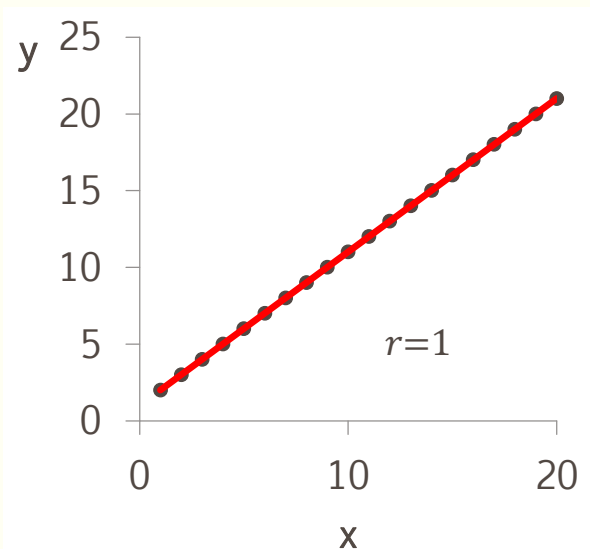
Korelační koeficient



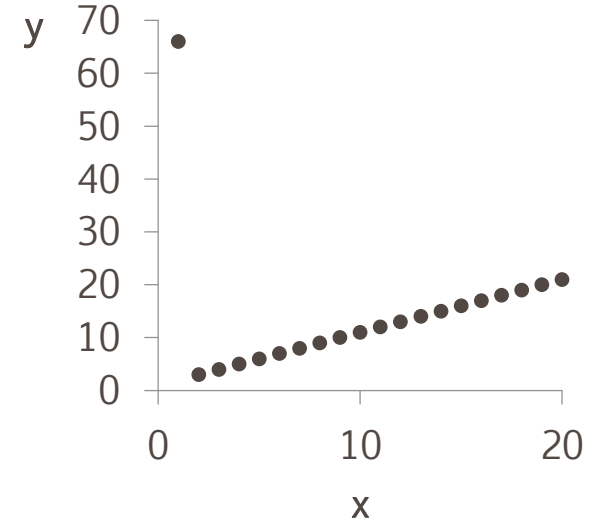
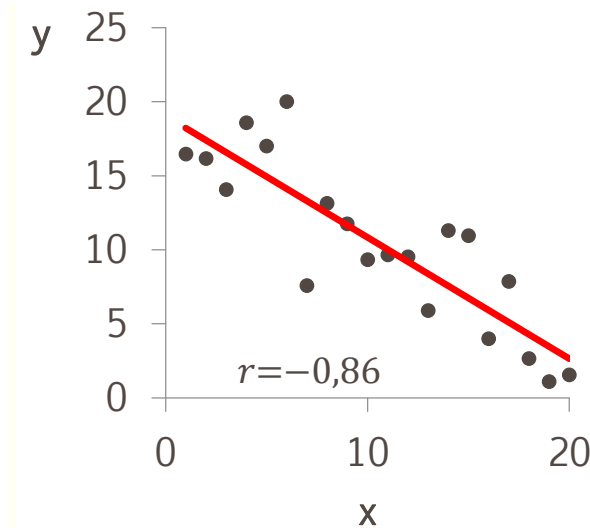
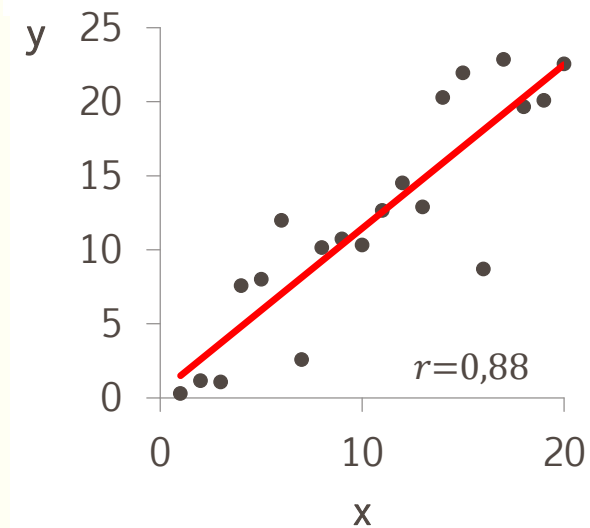
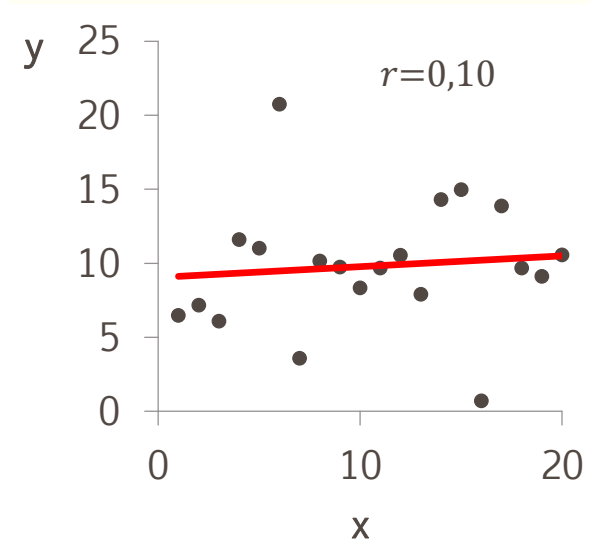
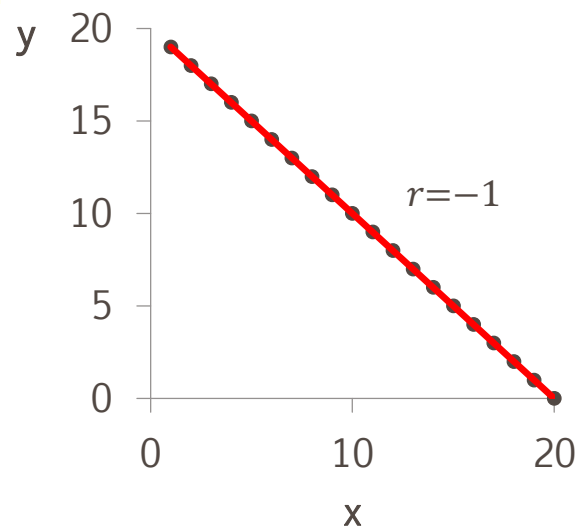
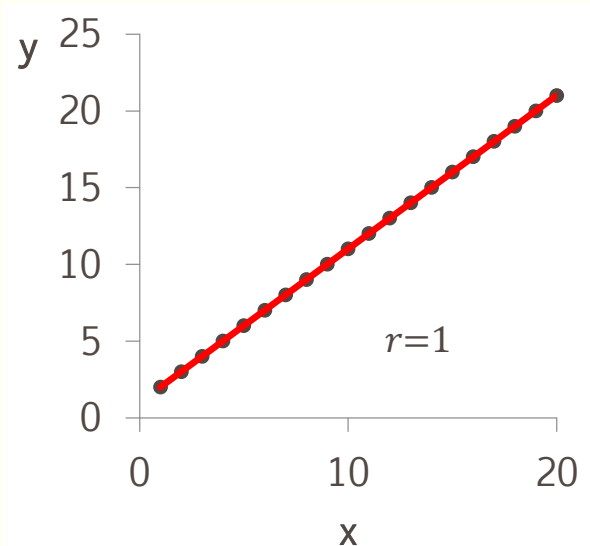
Korelační koeficient



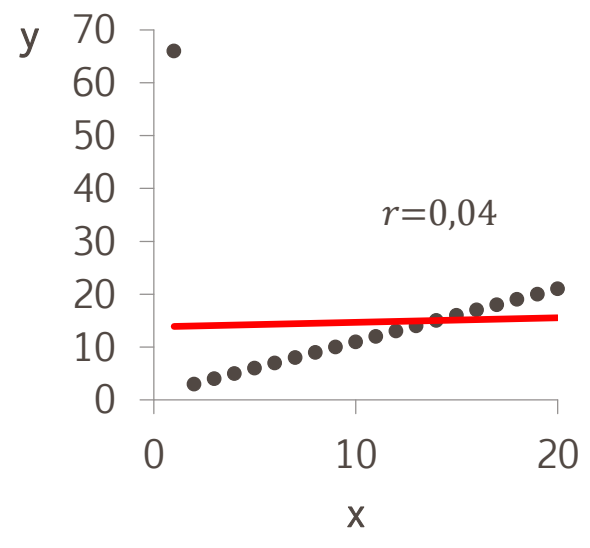
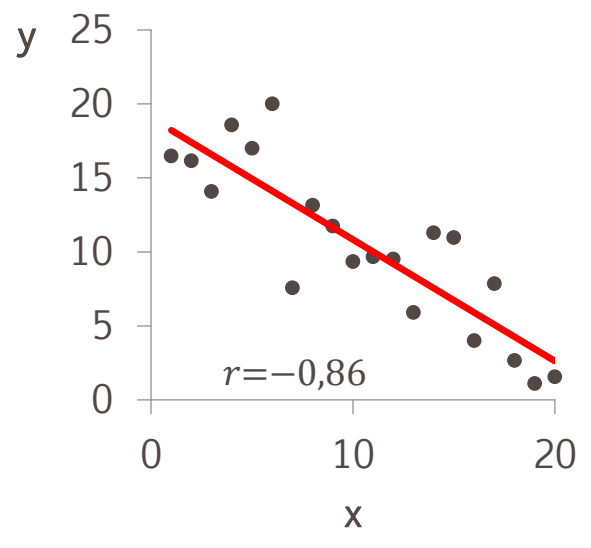
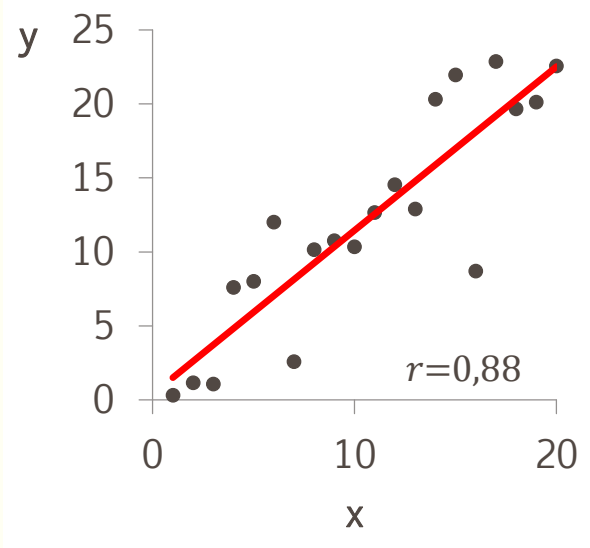
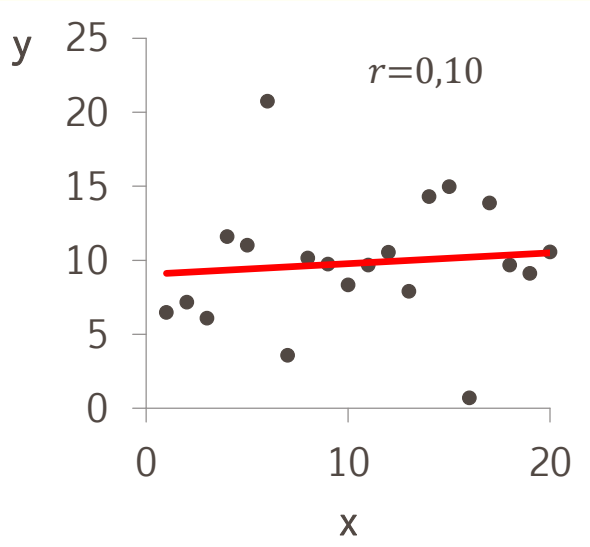
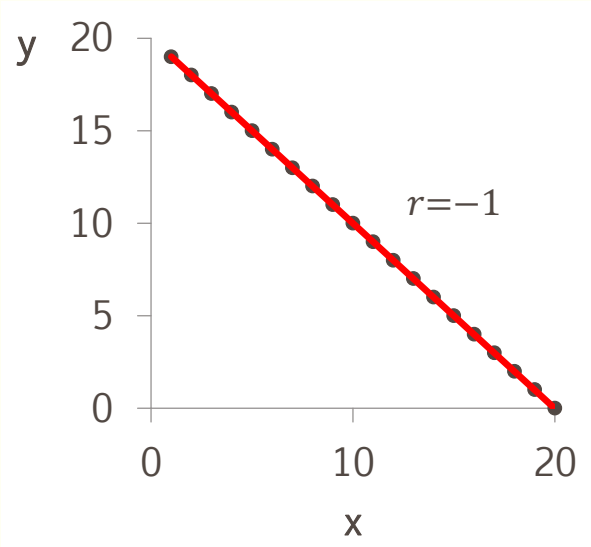
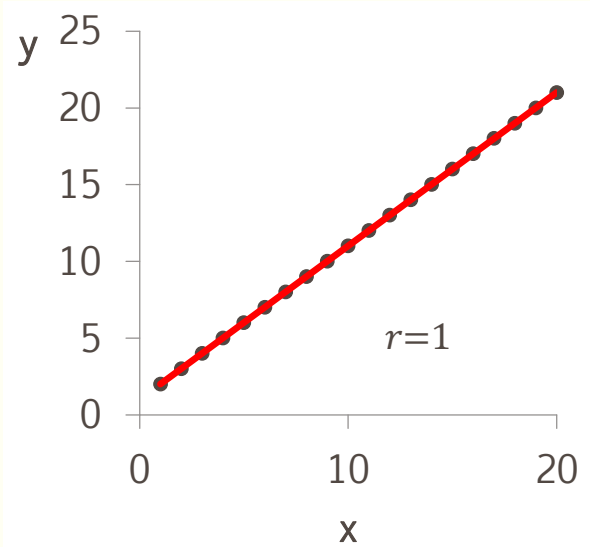
Korelační koeficient



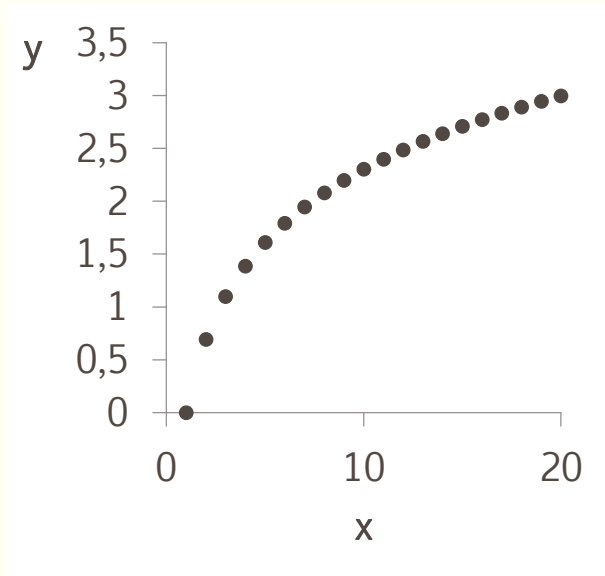
Korelační koeficient



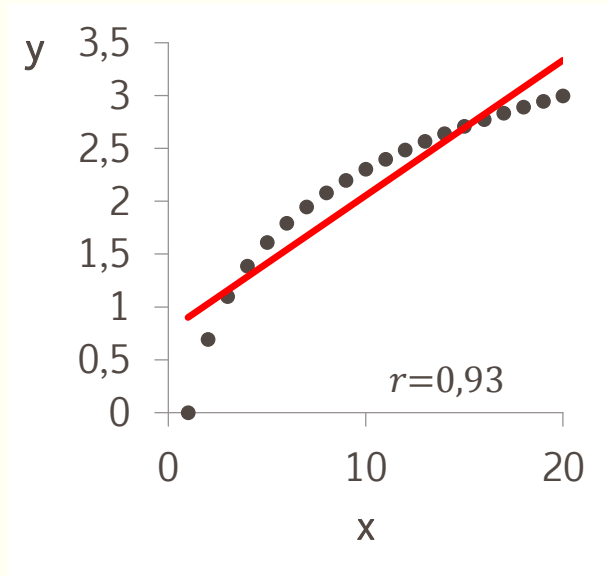
Korelační koeficient



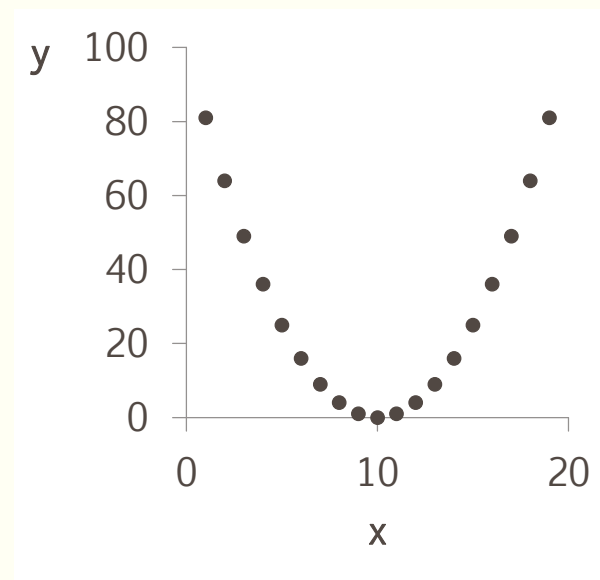
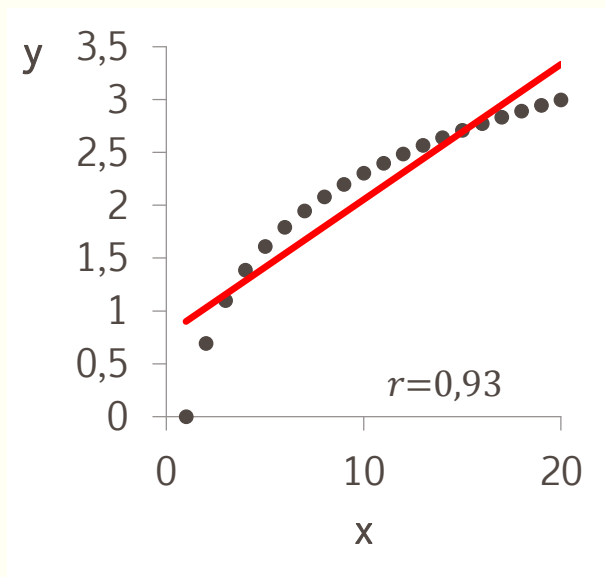
Korelační koeficient



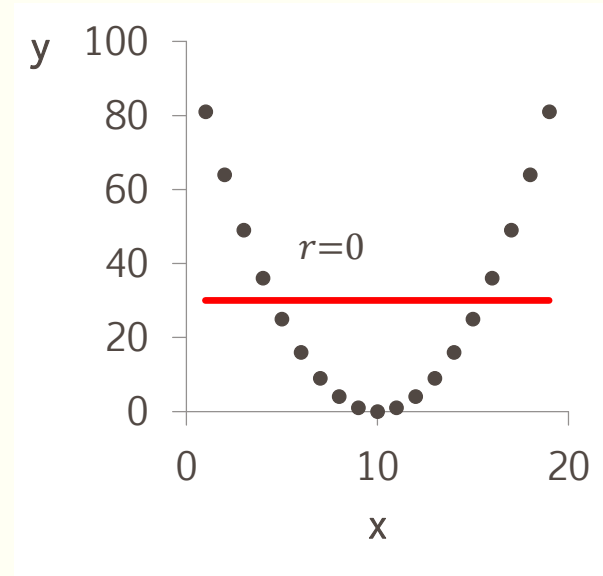
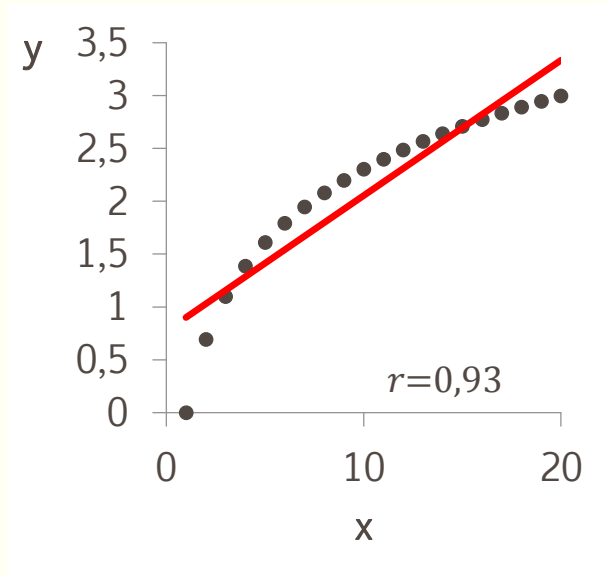
Korelační koeficient



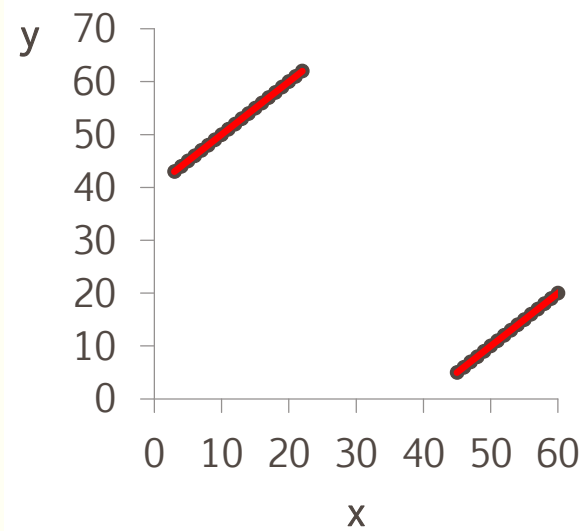
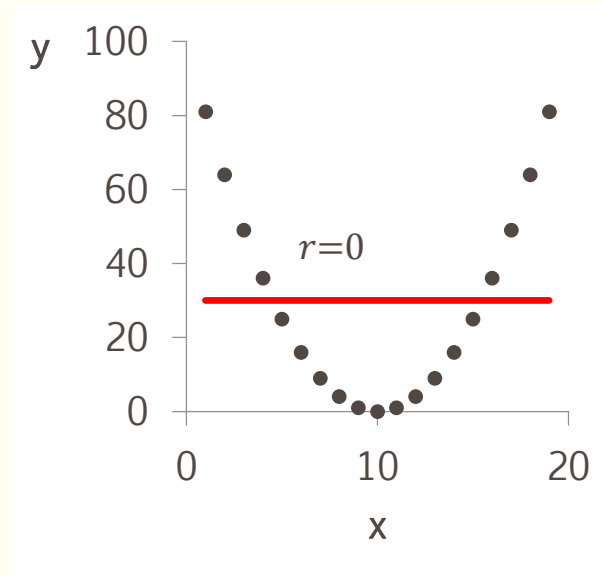
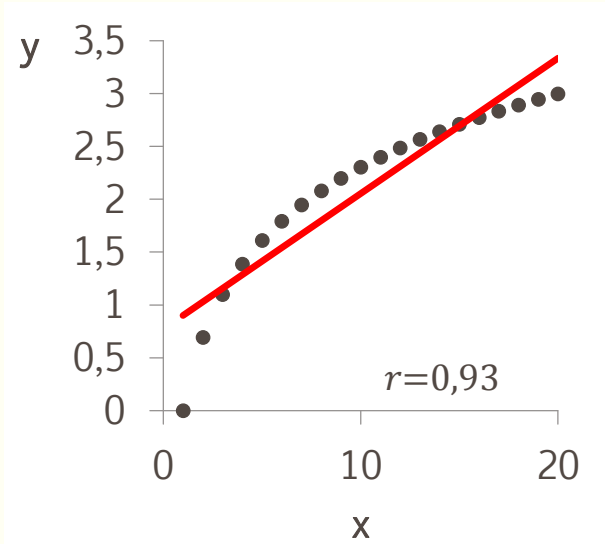
Korelační koeficient



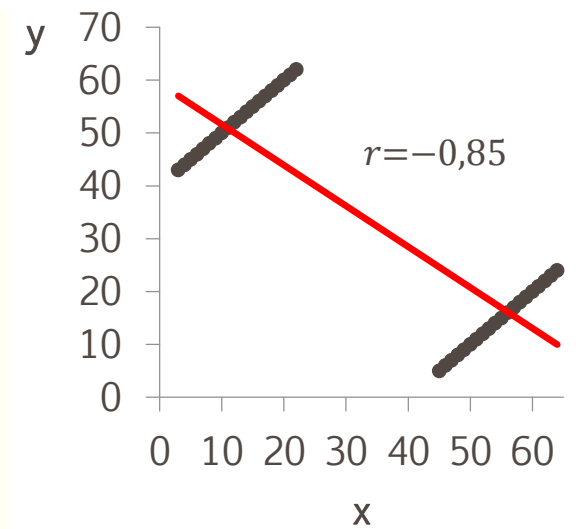
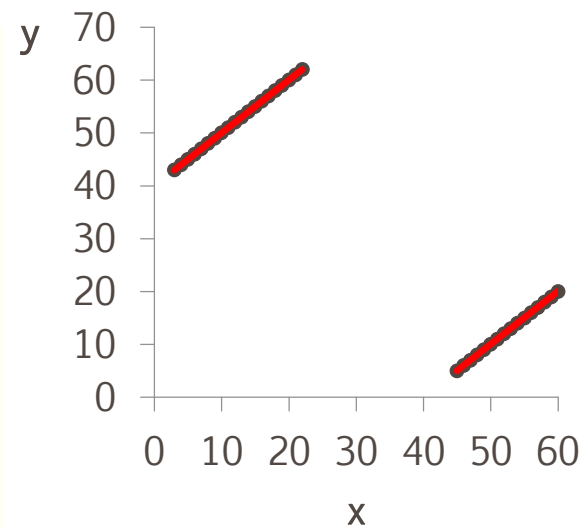
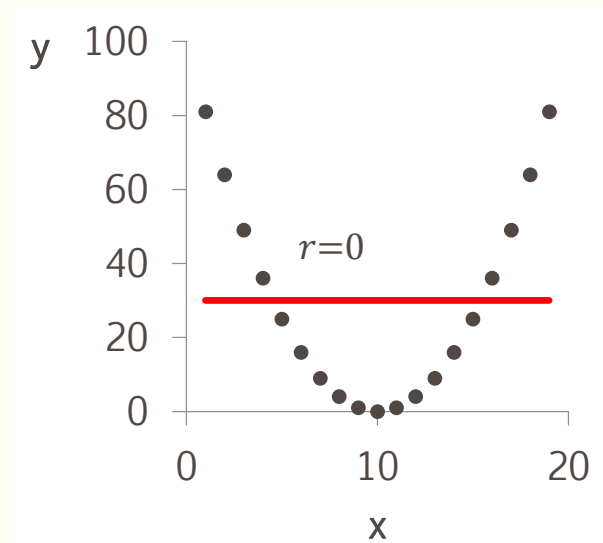
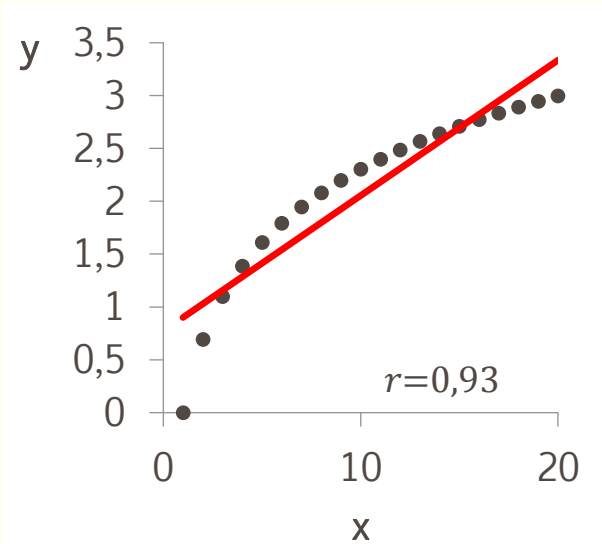
Korelační koeficient



Korelační koeficient

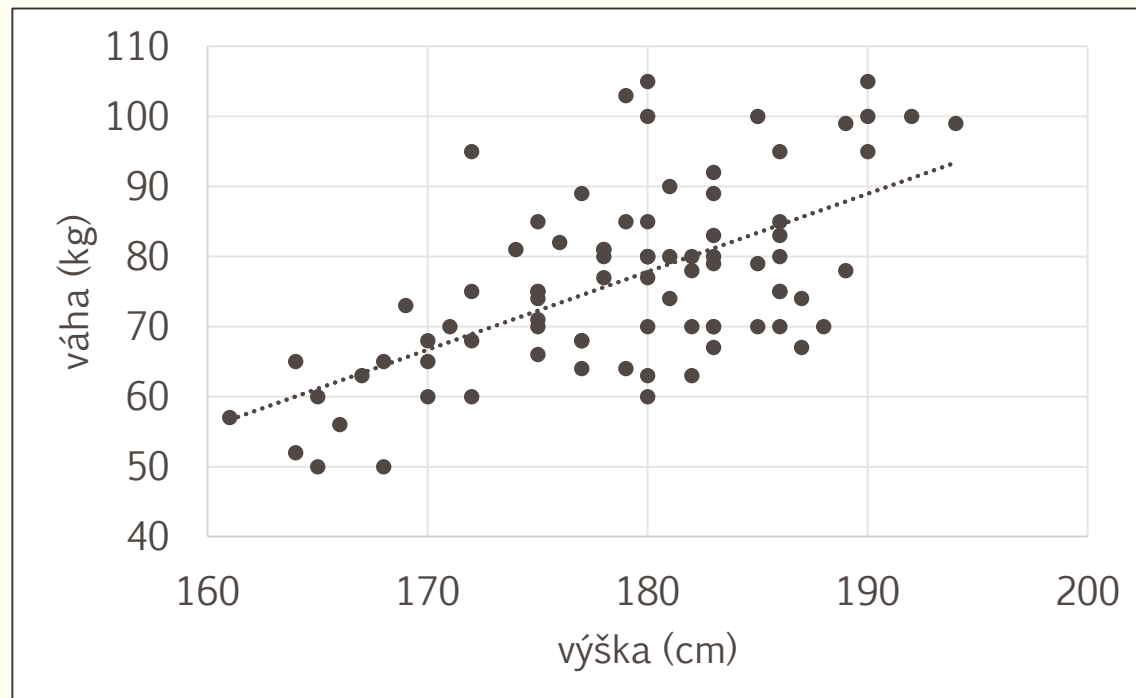


Korelační koeficient



Analýza závislosti dvou kvantitativních znaků

- Vizualizace – bodový graf

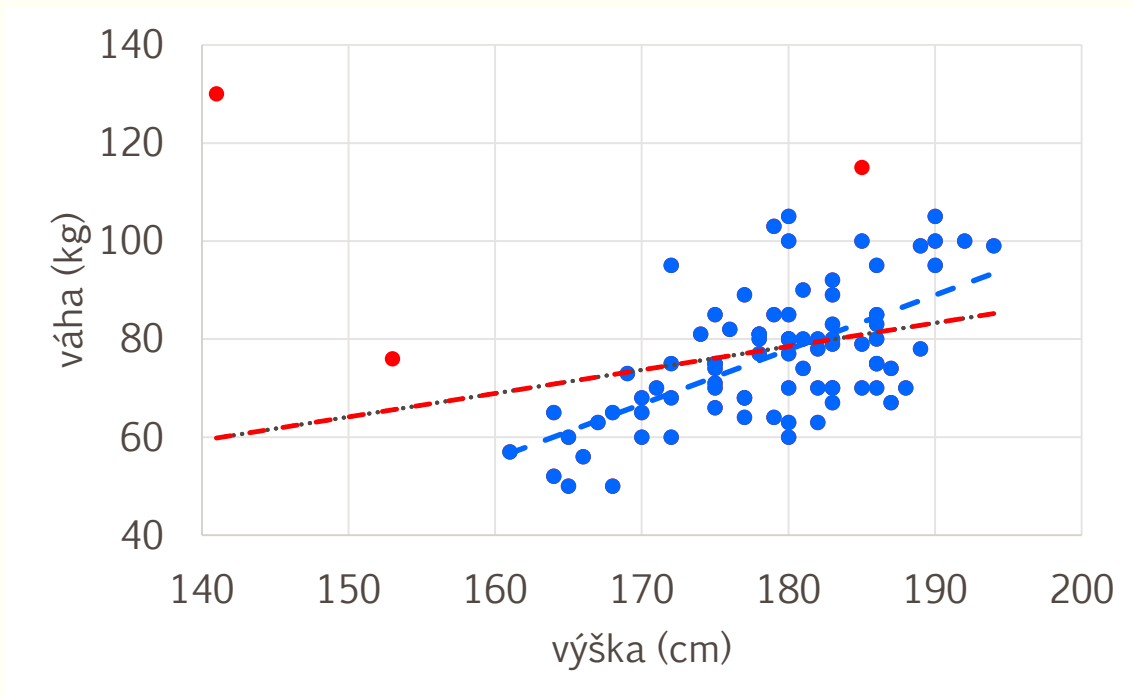


Výběrový korelační koeficient:

$$r = 0,6146$$

Analýza závislosti dvou kvantitativních znaků

- Vizualizace – bodový graf



Výběrový korelační koeficient:

$r = 0,6146$ vs. $r = 0,2832$

Analýza závislosti dvou kvantitativních znaků

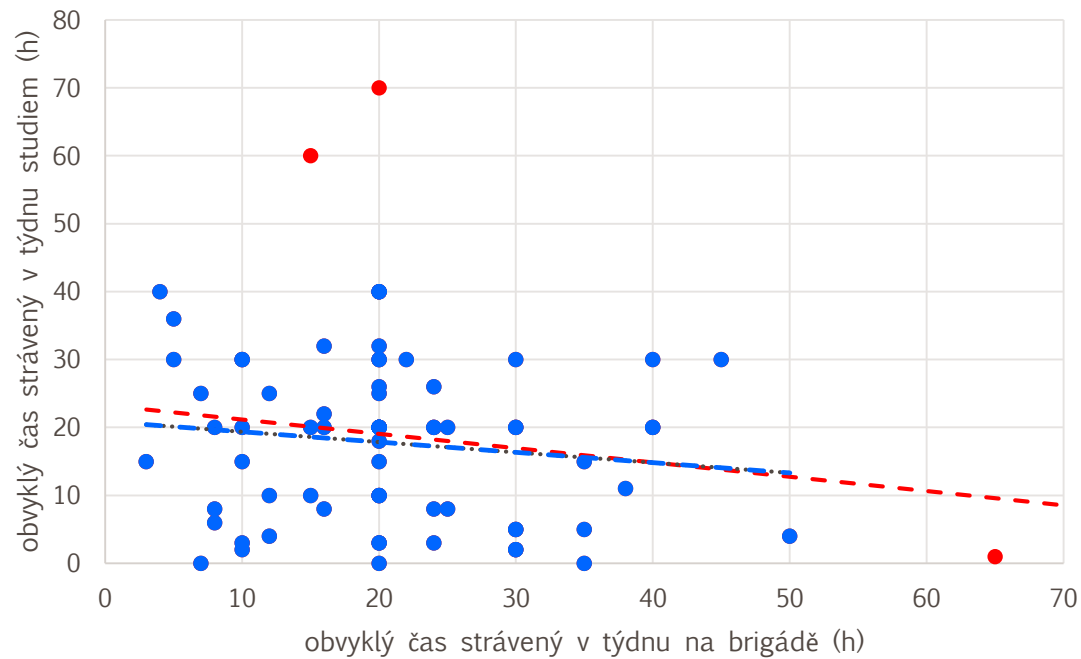
Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	jinak	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během studia	5	36
1.4.2016 10:45	žena	166	56	ano	jinak	jinak	12	10
1.4.2016 10:52	žena	188	70	ano	jinak	praxe v oboru během studia	24	26

Analýza závislosti dvou kvantitativních znaků

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	jinak	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během studia	5	36
1.4.2016 10:45	žena	166	56	ano	jinak	jinak	12	10
1.4.2016 10:52	žena	188	70	ano	jinak	praxe v oboru během studia	24	26

Analýza závislosti dvou kvantitativních znaků

- Vizualizace – bodový graf

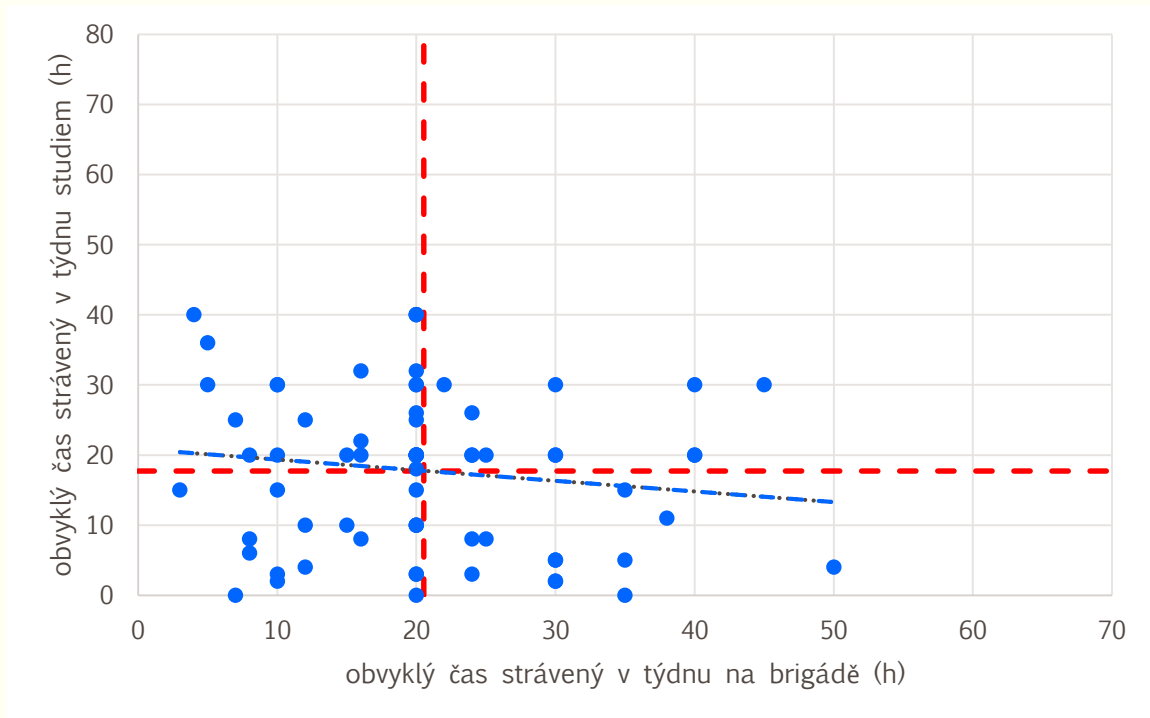


Výběrový korelační koeficient:

$$r = -0,1370 \text{ vs. } r = 0,2027$$

Analýza závislosti dvou kvantitativních znaků

- Vizualizace – bodový graf



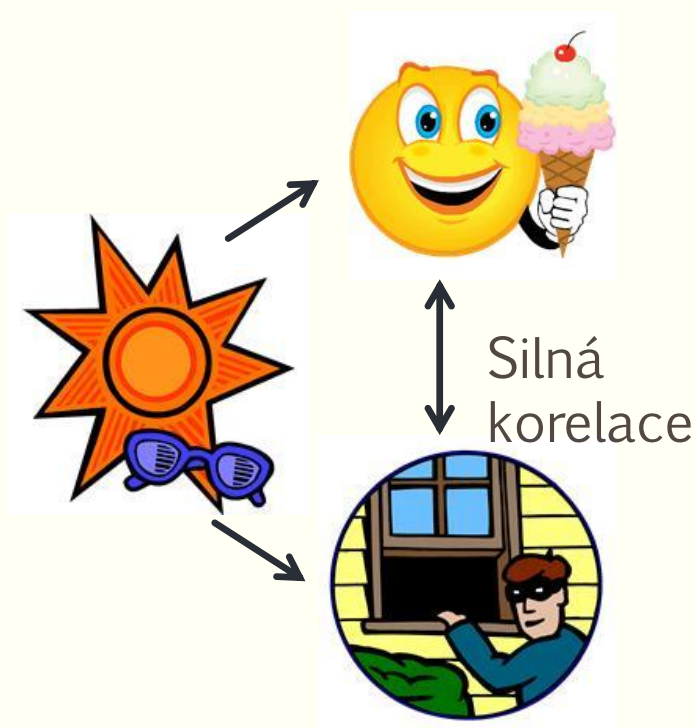
Výběrový korelační koeficient:

$$r = -0,1370$$

Korelační koeficient

Pokud jsou dvě náhodné veličiny korelované, znamená to pouze to, že jsou lineárně závislé.

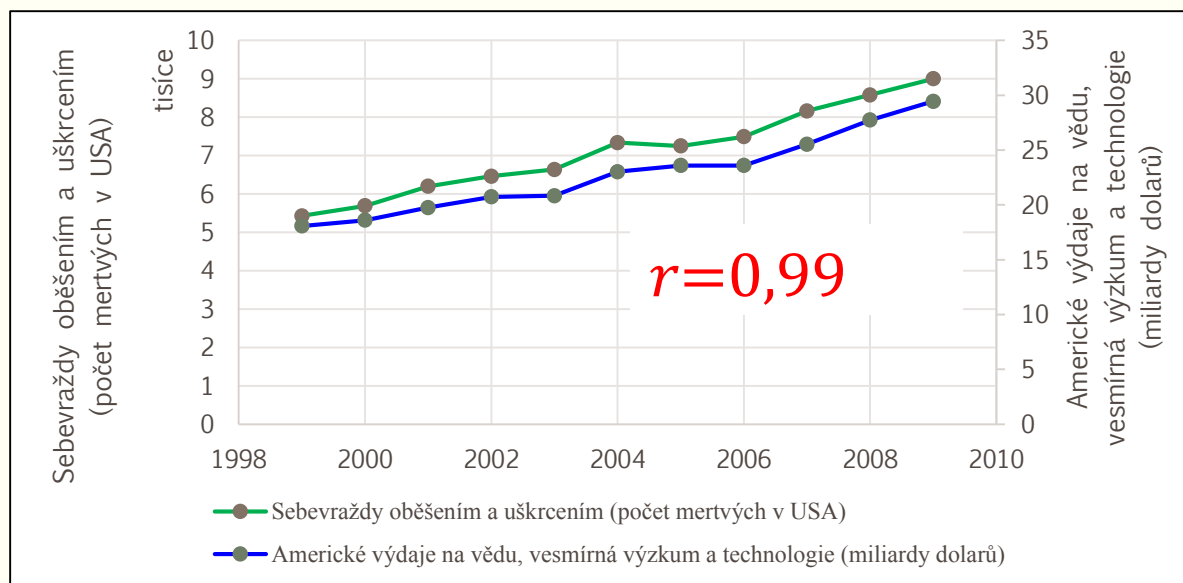
Nelze z toho však ještě usoudit, že by jedna z nich musela být **příčinou** a druhá **následkem**. To samotná korelovanost nedovoluje rozhodnout.



Korelační koeficient

Pokud jsou dvě náhodné veličiny korelované, 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.



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/>



Korelační koeficient

V praxi se zpravidla hodnota koeficientu korelace interpretuje takto:

Korelační koeficient	Typ lineární závislosti
$ r = 0,0$	neexistující
$ r \in (0,0; 0,3)$	velmi slabá
$ r \in (0,3; 0,7)$	středně silná
$ r \in (0,7; 1,0)$	těsná
$ r = 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!

DĚKUJI ZA
POZORNOST!

