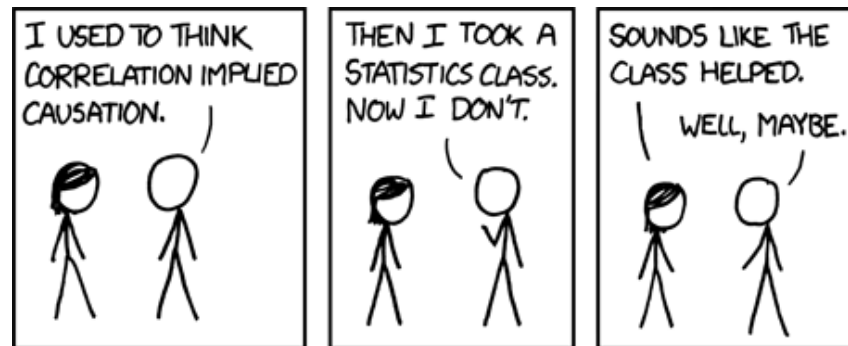


Kvantitativní metodologie



➤ Zpracovat **závěrečnou zprávu** z vašeho výzkumného projektu (2-8 stran); přednes dlouhý max 15 min.

1. Návrh projektu, formulování hypotéz, v jakém softwaru budete zpracovávat data?
2. Popište způsob sběru dat (od kdy do kdy, jak, kolik respondentů bylo osloveno, kolik dotazníků se vrátilo, atd. – pokud je k dispozici)
3. Popisné statistiky a grafy jednotlivých proměnných
 - Volte vhodné charakteristiky k proměnným – frekvenční tabulky u kategoriálních, průměr, směrodatná odchylka, ... u spojitých
 - Volte vhodné grafy, interpretujte je (slovy popište, co v grafech vidíte)
4. Testování hypotéz (2-3) + jejich interpretace
 - Otestujte normalitu spojitých proměnných, interpretujte výsledek
 - Proved'te alespoň jeden t-test či analýzu rozptylu (ANOVA), interpretujte výsledek
 - Proved'te alespoň jednu regresní analýzu (lineární nebo logistickou), interpretujte výsledek
5. Sepište krátký závěr svého výzkumu – na co jste přišli?
6. Pošlete závěrečnou zprávu vyučujícímu před prezentací, připojte i data, co jste dostali

Odkazy na databáze s výzkumnými daty

- https://openpsychometrics.org/_rawdata/
- <https://osf.io/9zhpj/>

Čím se vyznačují vědecké otázky?

- Disponují ženy vyšší mírou empatie než muži?
 - Tato otázka je poměrně vágní
 - Pokud se neupřesní termíny, pak z pohledu jazyka vědy → **nedává smysl** – proč?

Emotion Review

isre International Society for
Research on Emotion

3.780

Impact Factor

5-Year Impact Factor 5.129

Journal Indexing & Metrics »

Journal Home

Browse Journal ▾

Submit Paper ↗

About ▾

Subscribe ↗

Article Menu

Close ^

Download PDF



Full Article

Content List

Abstract

Identifying Areas of Confusion

Distinguishing Empathy From Other
Concepts

Cognitive or Affective?



Figures & Tables



Article Metrics



Empathy: A Review of the Concept

Benjamin M.P. Cuff, Sarah J. Brown, Laura Taylor, more...

Show all authors ▾

First Published December 1, 2014 | Research Article

<https://doi.org/10.1177/1754073914558466>

Article information ▾

Altmetric

6



Abstract

The inconsistent definition of empathy has had a negative impact on both research and practice. The aim of this article is to review and critically appraise a range of definitions of empathy and, through considered analysis, to develop a new conceptualisation. From the examination of 43 discrete definitions, 8 themes relating to the nature of empathy emerged: "distinguishing empathy from other concepts"; "cognitive or affective?"; "congruent or incongruent?"; "subject to other stimuli?"; "self/other distinction or merging?"; "trait or state influences?"; "has a behavioural outcome?"; and "automatic or controlled?" The relevance and validity of each theme is assessed and a new conceptualisation of empathy is offered. The benefits of employing a more consistent and complete definition of empathy are discussed.

Keywords

affective empathy, cognitive empathy, definition, empathy, empathy assessment

Article available in:

Vol 8, Issue 2, 2016

Related Articles

Similar Articles:

View all >

[The Many Faces of Empathy:
Parsing Empathic Phenomena
through a Proximate, Dynamic-
Systems View...](#)

Show details ▾

[Dissecting the Neural Mechanisms
Mediating Empathy](#)

Show details ▾

[Social Cognitive Neuroscience of
Empathy: Concepts, Circuits, and
Genes](#)

Show details ▾



PDF

Help

SAGE Recommends

Čím se vyznačují vědecké otázky?

- Disponují ženy vyšší mírou empatie než muži?
 - Tato otázka je poměrně vágní
 - Aby tato otázka byla přesnější, bylo je třeba upřesnit klíčový termín této otázky (i. e. empatie)
- Disponují ženy vyšší mírou emocionální empatie v porovnání s muži?
 - Je dobré, když v otázce zazní metodologický přístup, který použijeme k jejímu zodpovězení:
- Disponují ženy, statisticky významně vyšší mírou emocionální empatie než muži?
- Mohou být formulovány jak v kvalitativním tak i kvantitativním výzkumu
- Příklady vědeckých otázek:
 - Jaké druhy volnočasových aktivit převládají u vysokoškoláků v dané lokalitě?
 - Jaké neuroanatomické podklady jsou spojeny s mírou emocionální empatie?

Dva typy výzkumu: kvalitativní a kvantitativní

<i>Dle Hendl (2012)</i>	Kvantitativní	Kvalitativní
<u>Cíl</u>	testování hypotéz	vytváření teorie
<u>Informace</u>	málo o mnoha jedincích	mnoho o málo jedincích
<u>Generalizace</u>	snadná	problematická
<u>Standardizace</u>	silná -> vysoká reliabilita, nízká validita	slabá -> nízká reliabilita, vysoká validita
<u>Logika</u>	Deduktivní	Induktivní

Výzkumné hypotézy

- Řecký termín „*hypo-thesis*“ znamená předpoklad
 - Očekávaný vztah mezi proměnnými
 - Příklady hypotéz:
 - Respondenti, kteří dosahují vyšší hodnoty pozitivních emocí, budou vykazovat nižší míru existenciální frustrace (1).
 - Existuje statisticky významný rozdíl v neurální aktivitě orbitofrontální mozkové kůry u lidí co zažili trauma v porovnání s lidmi, kteří trauma nezažili při sledování experimentálního videa.
 - Derivuje či odvozují se z teorie
 - Musí být ověřitelné či testovatelné!
 - Na čem závisí to, zdali jsou hypotézy ověřitelné?
 - na tom, zdali pojmy obsažené v hypotézách jsou empiricky ukotvitelné →
- Operacionalizovatelné**



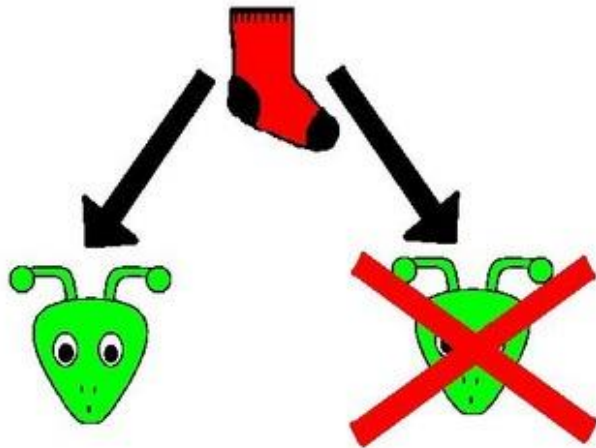
Výzkumné hypotézy

- Bývají formulovány oznamovací větou
- Odpovídá se na ně ano/ne resp. „hypotéza byla zamítnuta“ nebo „hypotézu nebylo možné zamítnout“
- Hypotézy se dělí na dva druhy:
 - Nulová hypotéza (H_0)
 - Alternativní hypotéza (H_1)

Výzkumné hypotézy

Příklad č.1: krádež ponožek mimozemšťany

Q. Where have all my socks gone?



Alternate Hypothesis

=

Extra-terrestrial beings have transported themselves into my house in order to steal my socks.

Null Hypothesis

=

Aliens are not to blame. There is some other explanation for the disappearing socks.

Příklad č.2: tělesná výška u studentů a učitelů:

H0: Studenti a učitelé se **neliší** v průměru tělesné výšky

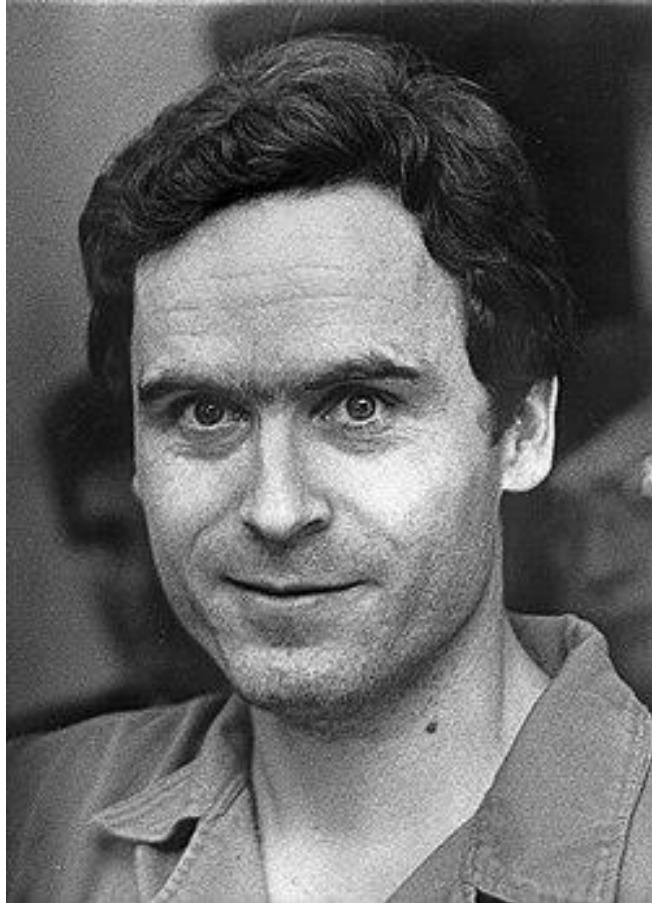
H1: Studenti a učitelé se **liší** v průměru tělesné výšky

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Co když nulovou hypotézu nezamítneme? –
znamená to, že nulová hypotéza platí?

Výzkumné hypotézy



Příklad č.3: násilné trestné činy Teda Bundyho

H0: Ted Bundy se **nedopouštěl** násilných trestných činů

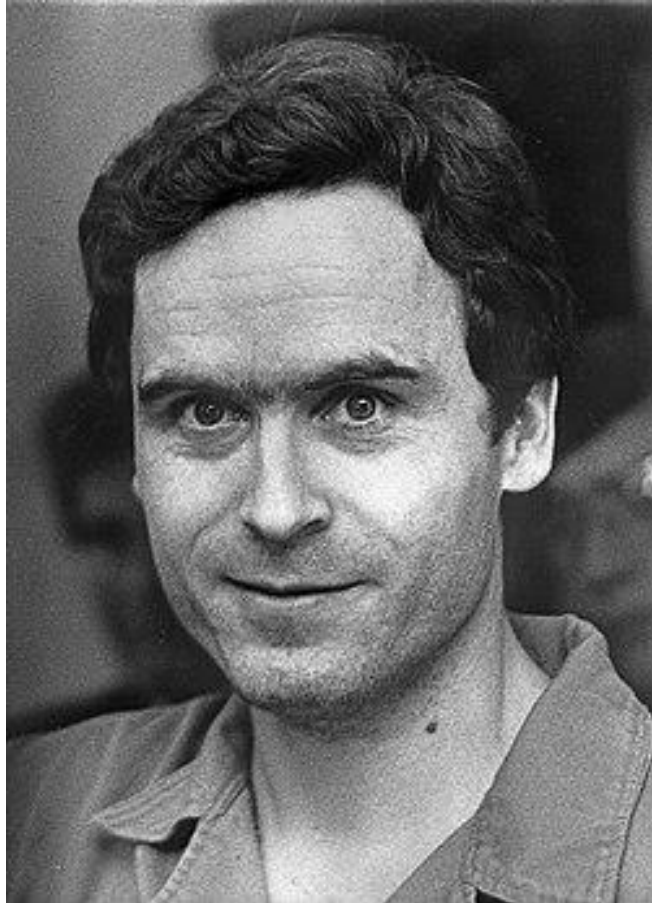
H1: Ted Bundy se **dopouštěl** násilných trestných činů.

Co když nemáme dostatek důkazů? Znamená to, že se násilných trestných činů nedopouštěl?

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Výzkumné hypotézy



Příklad č.3: násilné trestné činy Teda Bundyho

H0: Ted Bundy se **nedopouštěl** násilných trestných činů

H1: Ted Bundy se **dopouštěl** násilných trestných činů.

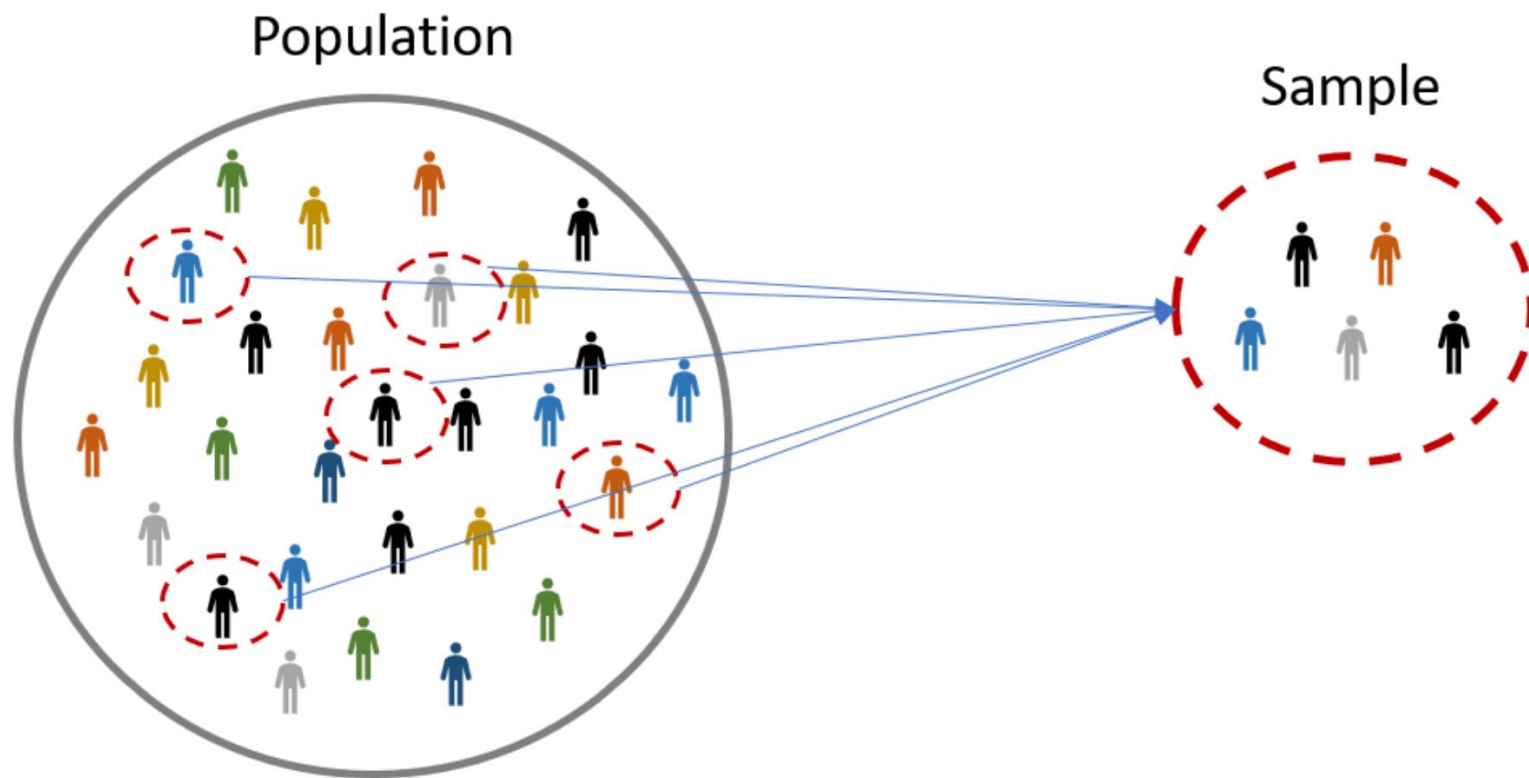
Co když nemáme dostatek důkazů? Znamená to, že se násilných trestných činů nedopouštěl?

↙
Z toho, že nemáme dostatek evidence K zamítnutí nulové hypotézy nevyplývá, že nulová hypotéza platí.

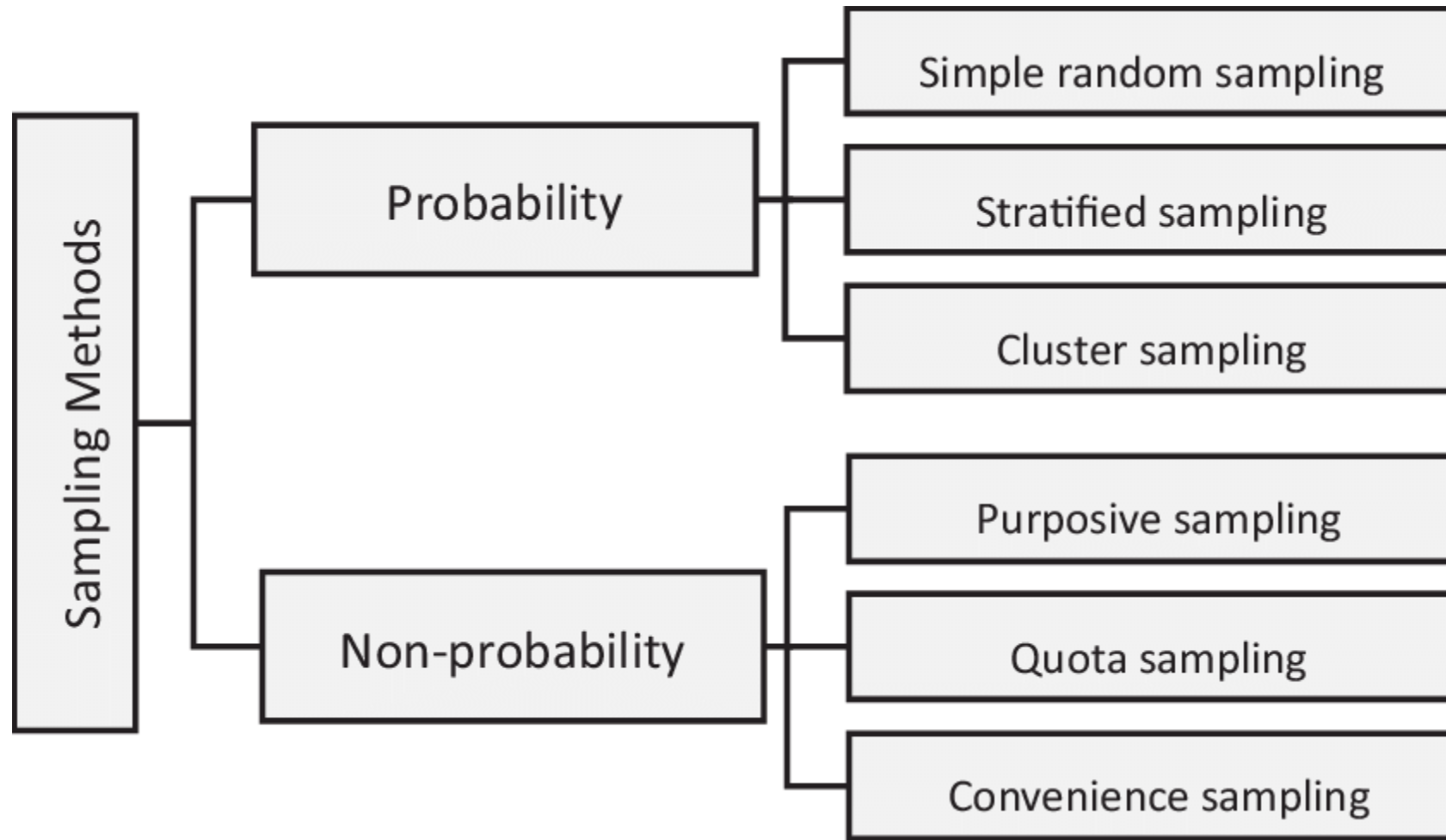
$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Výzkumný soubor vs Populace



Statistický výběr



Fáze kvantitativního výzkumu

= testování předem formulovaných hypotéz (potvrzování/falzifikace teorie)

Fáze výzkumu:

1) Co budeme zkoumat

- Určení výzkumného problému, otázek
- Stanovení hypotéz

2) Rozhodnutí o metodologii

- Metoda sběru dat, metoda výběru...

3) Vyjasnění pojmů, příprava výzkumu

- Konceptualizace a operacionalizace
- Vytvoření výzkumného nástroje (obvykle dotazníku)
- Výběr respondentů

4) Sběr dat

- Sbíráme data, která se vztahují k výzkumným otázkám a hypotézám

5) Analýza dat

- Potvrzení nebo falzifikace hypotéz

6) Výzkumná zpráva

- Informace o bodech 1-5
- Výsledky analýzy, závěry a interpretace dat

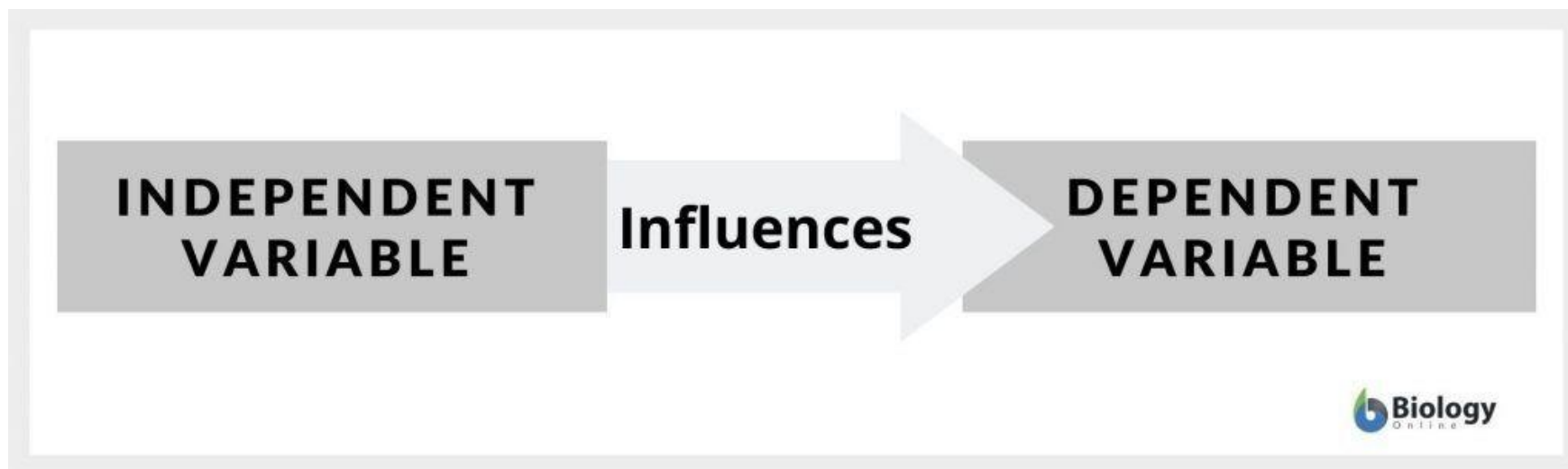
Tab. 4 Základní rysy statistického a teoretického výběru

	statistický výběr (kvantitativní zkoumání)	teoretický výběr (kvalitativní zkoumání)
základní soubor	musí být definován	není definován
rozsah výběru	je předem znám	není předem znám
znaky základního souboru	jejich rozložení je známé nebo je lze odhadnout	nejsou předem známé
provedení výběru	pouze jednou, podle předem zvoleného plánu	opakovaně, podle dosahovaných výsledků
ukončení výběru	po dosažení zvoleného rozsahu	po dosažení teoretického nasycení

Poznámka: srv. Hendl, 1997, s. 69 (též Hendl, 2005, s. 152).

Typy proměnných

- **Nezávislá proměnná (independent Proměnná (Variable), prediktor)**
 - Proměnná, která má vysvětlit jinou proměnnou
 - Pohlaví, věk, vzdělání
- **Závislá proměnná (dependant Proměnná (Variable), outcome)**
 - Proměnná, kterou chceme vysvětlit nějakým prediktorem (nezávislou proměnnou)



Příklady nezávislých a závislých proměnných v
hypotézách

H1: Množství zkonzumovaného cukru za den souvisí s množstvím tělesného tuku v těle.

Nezávislá proměnná



Závislá proměnná



H1: Množství zkonsumovaného cukru za den souvisí s množstvím tělesného tuku v těle.

H1: Existuje statisticky významný rozdíl v tělesné výšce mezi studenty a pedagogy

Nezávislá proměnná



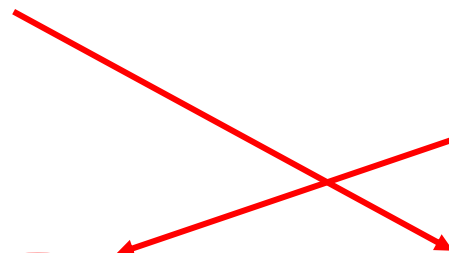
Závislá proměnná

H1: Existuje statisticky významný rozdíl v tělesné výšce mezi studenty a pedagogy

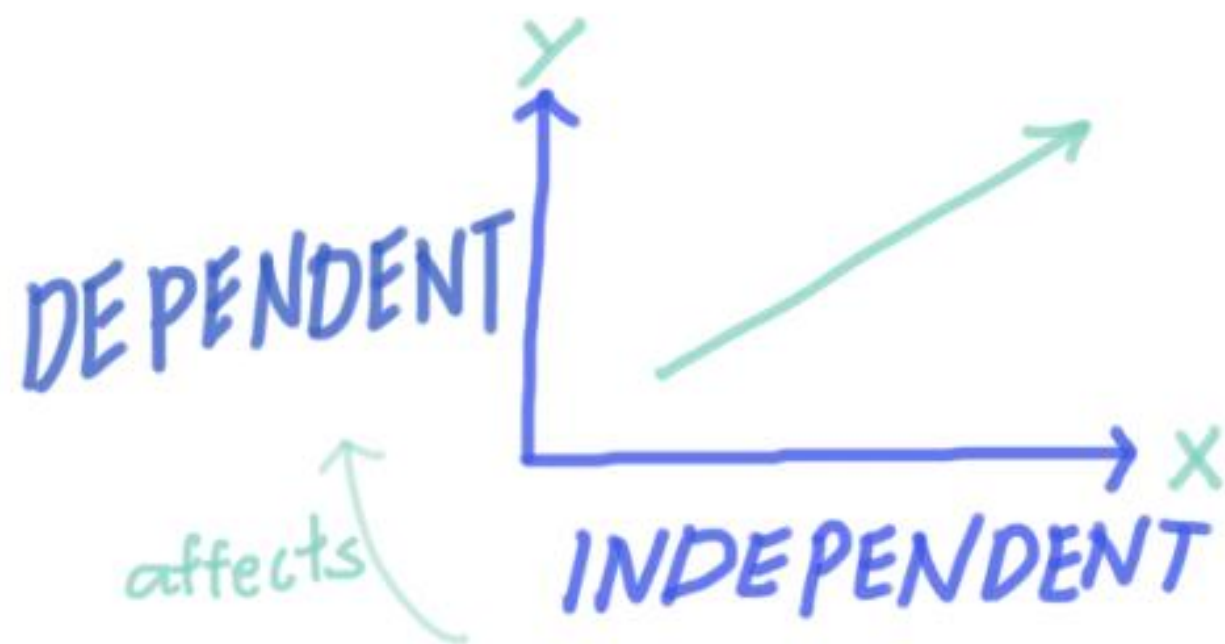
Nezávislá proměnná



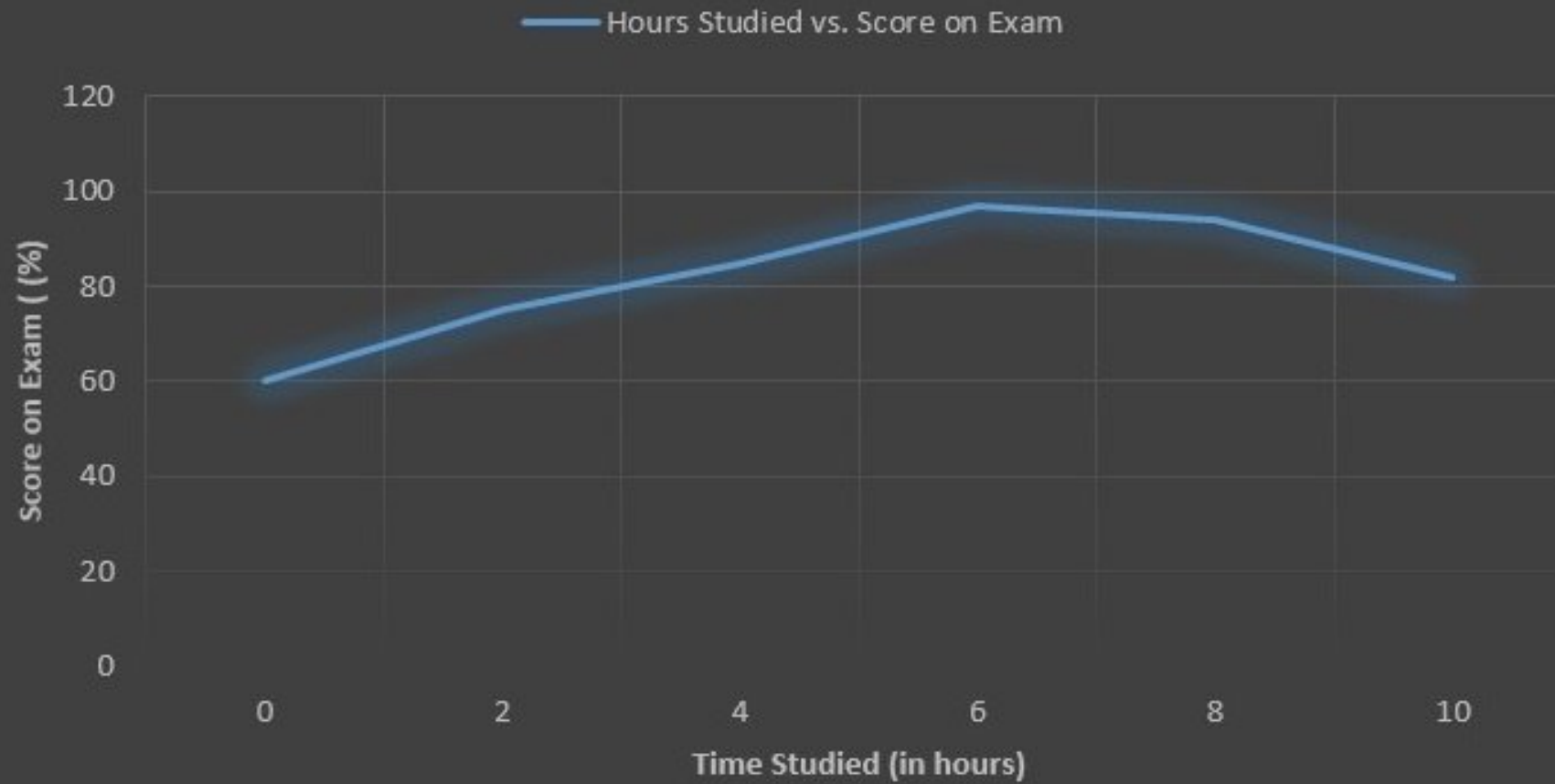
Závislá proměnná



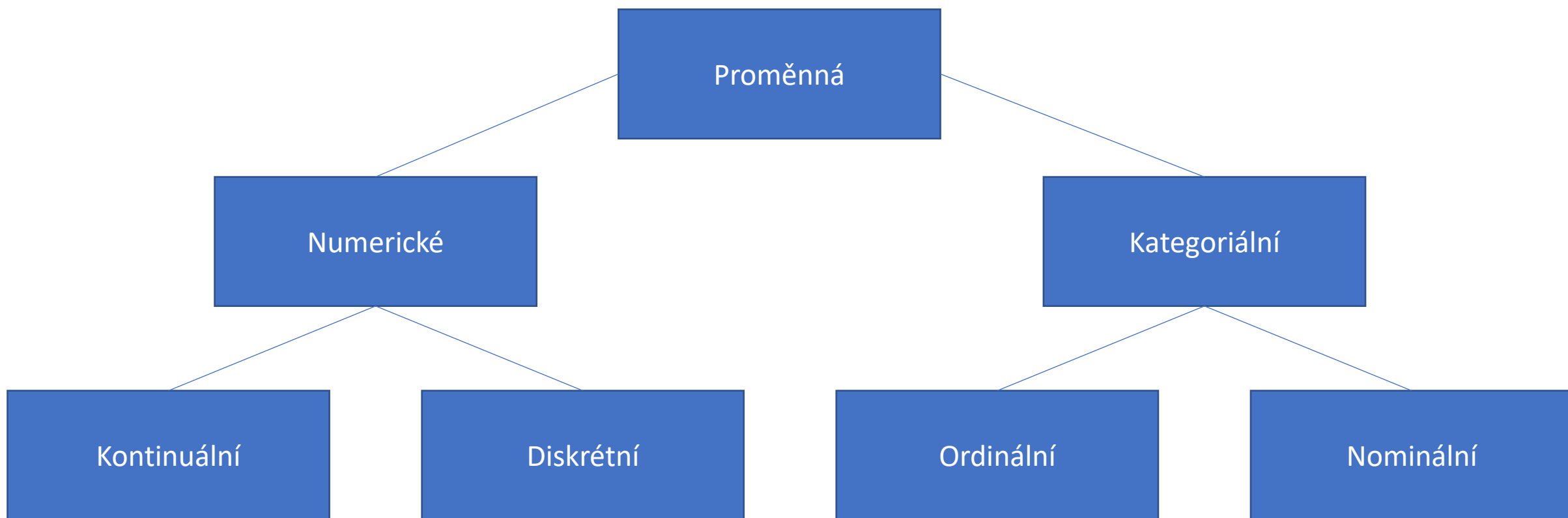
H1: Existuje statisticky významný rozdíl v tělesné výšce mezi studenty a pedagogy



Hours Studied vs. Score on Exam



Typy proměnných



Typy proměnných- numerické

Kontinuální



Diskrétní



Typy proměnných- kategorické

Ordinální



Nominální



Typ proměnných může být změněn

Věk (kontinuální)

29.12, 47.86, 94.12, 25.70



• Věk (diskrétní)

29, 48, 94, 26



• Věk ordinální

mladý / starý

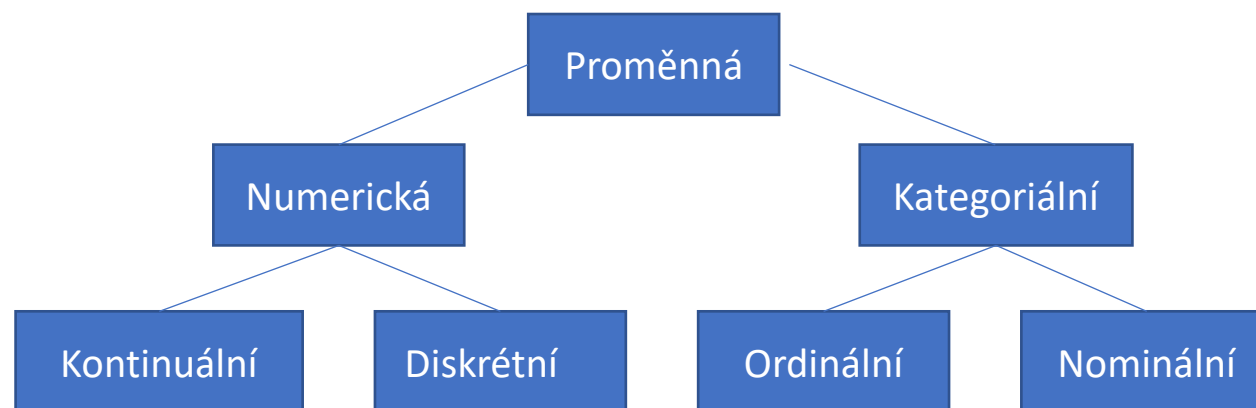
Statistický software

- K testování hypotéz a statistickým analýzám
- Existuje několik známých softwarů:
- Excel, SPSS, Stata, R, ...



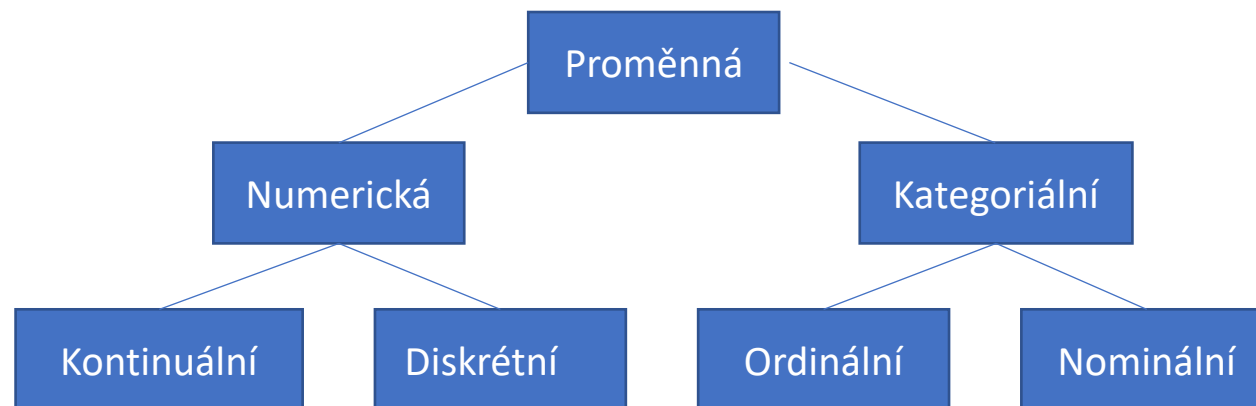
Test

Proměnná (Variable)	Typ
Pohlaví	
Věk	
Známky ve škole	
Rozdělení mezi kontrolní a experimentální skupinu	
Skóre v testu	
Jeden stupeň Likertovy škály	
Počet sourozenců	
Nejvyšší dosažené vzdělání	



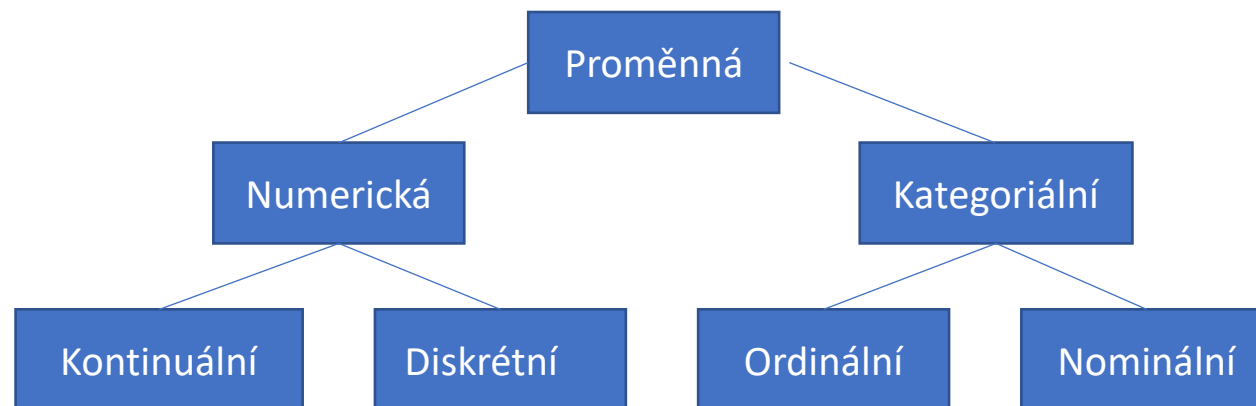
Test

Proměnná (Variable)	Typ
Pohlaví	Nominální
Věk	
Známky ve škole	
Rozdělení mezi kontrolní a experimentální skupinu	
Skóre v testu	
Jeden stupeň Likertovy škály	
Počet sourozenců	
Nejvyšší dosažené vzdělání	



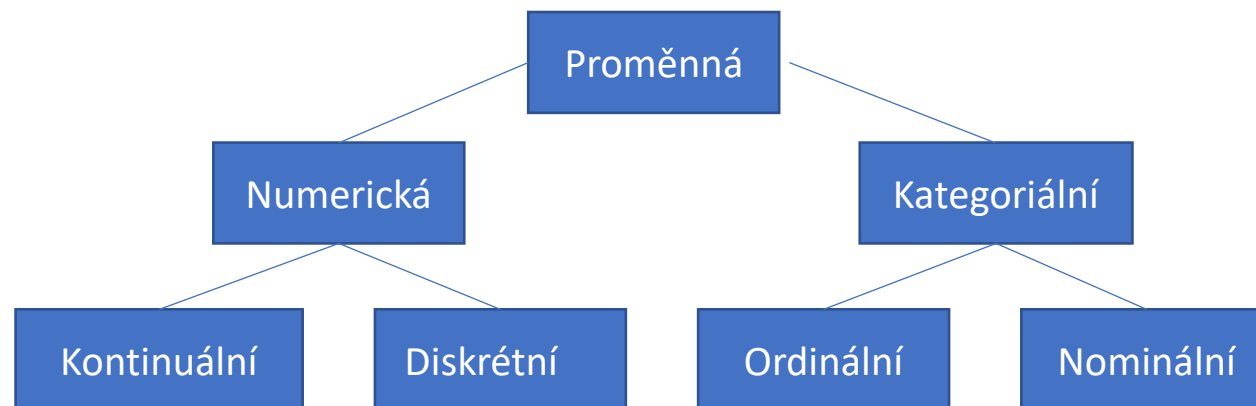
Test

Proměnná (Variable)	Typ
Pohlaví	Nominální
Věk	Kontinuální
Známky ve škole	
Rozdělení mezi kontrolní a experimentální skupinu	
Skóre v testu	
Jeden stupeň Likertovy škály	
Počet sourozenců	
Nejvyšší dosažené vzdělání	



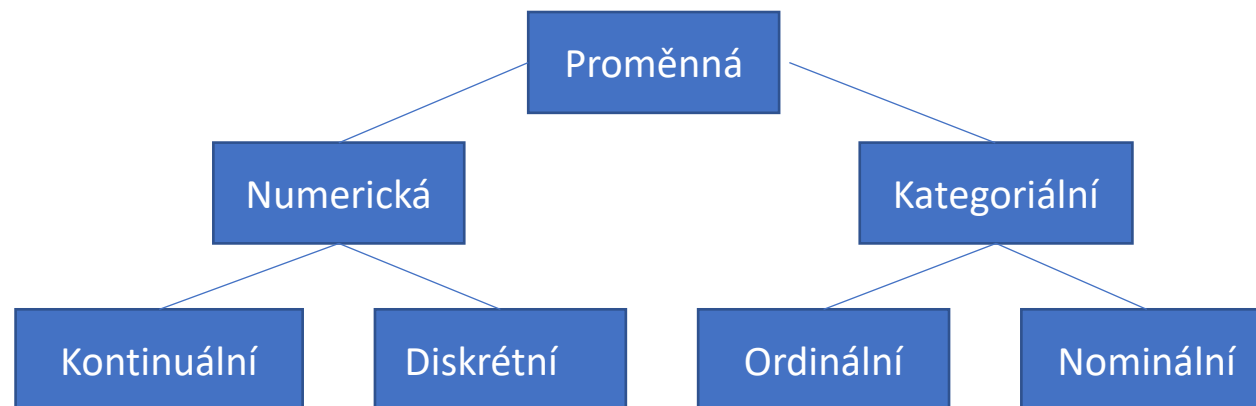
Test

Proměnná (Variable)	Typ
Pohlaví	Nominální
Věk	Kontinuální
Známky ve škole	Ordinální
Rozdělení mezi kontrolní a experimentální skupinu	
Skóre v testu	
Jeden stupeň Likertovy škály	
Počet sourozenců	
Nejvyšší dosažené vzdělání	



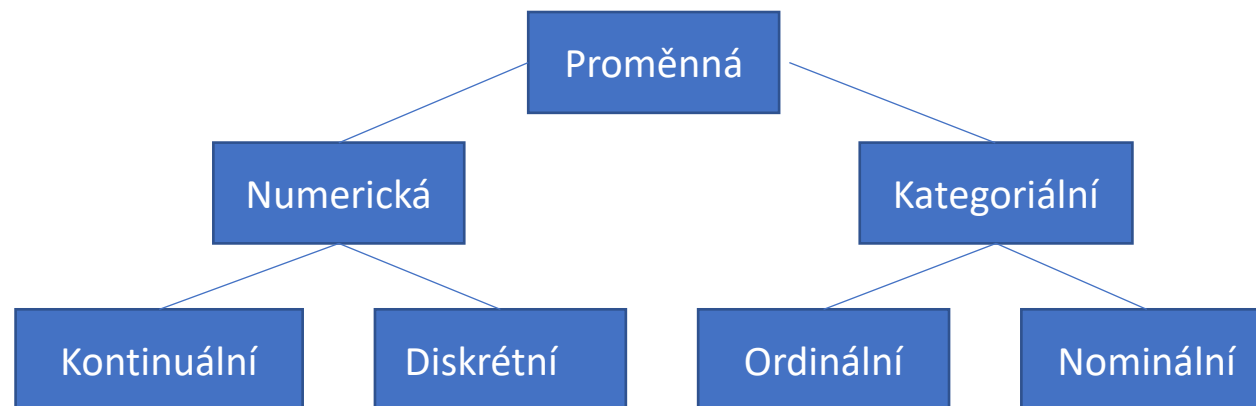
Test

Proměnná (Variable)	Typ
Pohlaví	Nominální
Věk	Kontinuální
Známky ve škole	Ordinální
Rozdělení mezi kontrolní a experimentální skupinu	Nominální
Skóre v testu	
Jeden stupeň Likertovy škály	
Počet sourozenců	
Nejvyšší dosažené vzdělání	



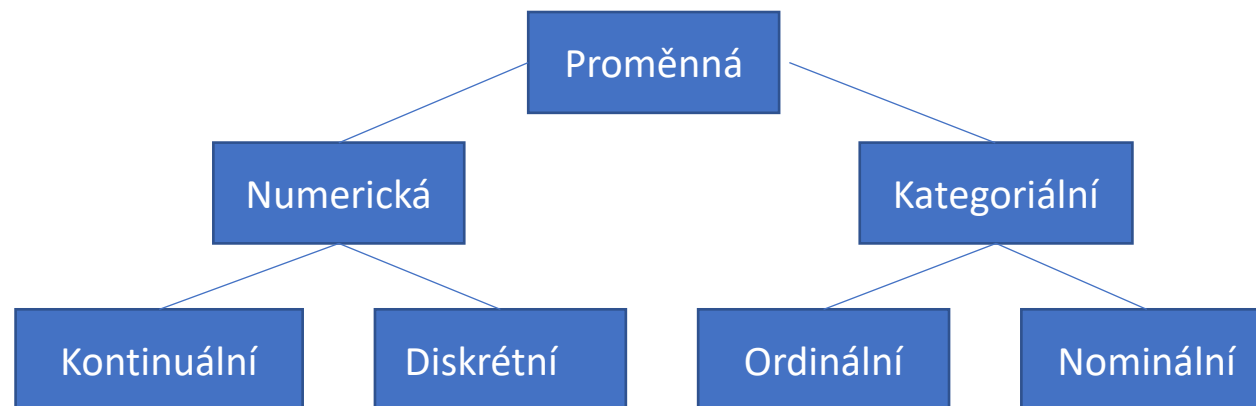
Test

Proměnná (Variable)	Typ
Pohlaví	Nominální
Věk	Kontinuální
Známky ve škole	Ordinální
Rozdělení mezi kontrolní a experimentální skupinu	Nominální
Skóre v testu	Diskrétní
Jeden stupeň Likertovy škály	
Počet sourozenců	
Nejvyšší dosažené vzdělání	



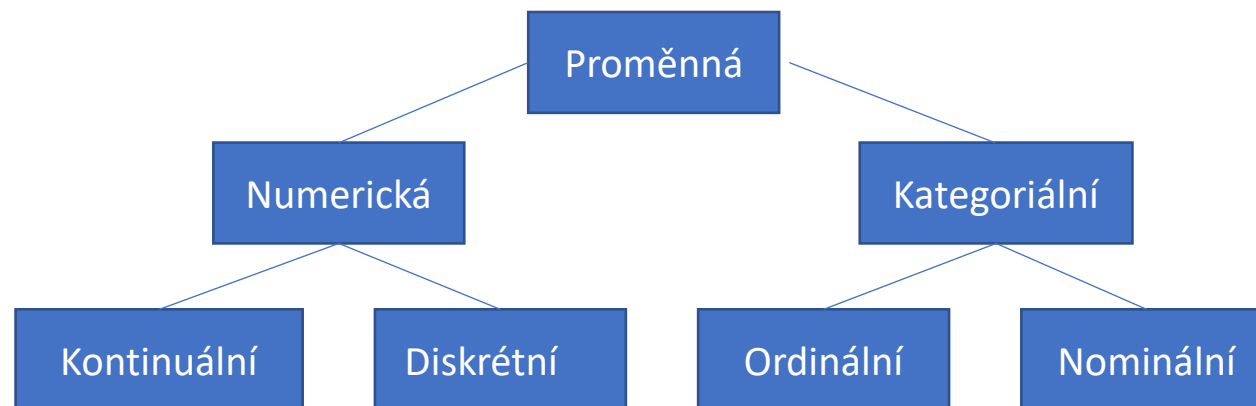
Test

Proměnná (Variable)	Typ
Pohlaví	Nominální
Věk	Kontinuální
Známky ve škole	Ordinální
Rozdělení mezi kontrolní a experimentální skupinu	Nominální
Skóre v testu	Diskrétní
Jeden stupeň Likertovy škály	Ordinální
Počet sourozenců	
Nejvyšší dosažené vzdělání	



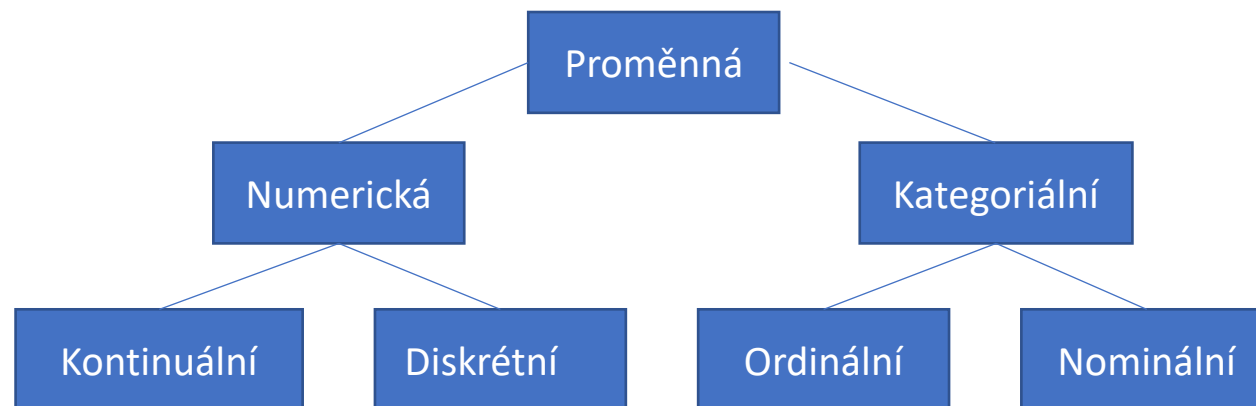
Test

Proměnná (Variable)	Typ
Pohlaví	Nominální
Věk	Kontinuální
Známky ve škole	Ordinální
Rozdělení mezi kontrolní a experimentální skupinu	Nominální
Skóre v testu	Diskrétní
Jeden stupeň Likertovy škály	Ordinální
Počet sourozenců	Diskrétní
Nejvyšší dosažené vzdělání	



Test

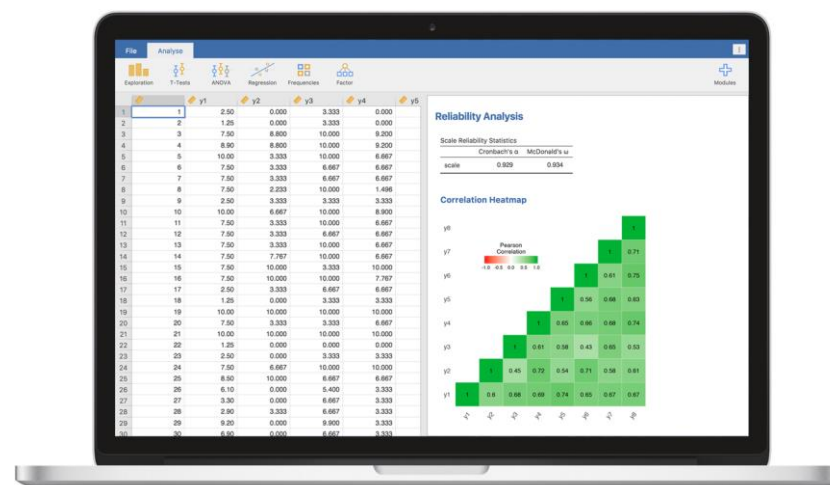
Proměnná (Variable)	Typ
Pohlaví	Nominální
Věk	Kontinuální
Známky ve škole	Ordinální
Rozdělení mezi kontrolní a experimentální skupinu	Nominální
Skóre v testu	Diskrétní
Jeden stupeň Likertovy škály	Ordinální
Počet sourozenců	Diskrétní
Nejvyšší dosažené vzdělání	Ordinální



Excel	SPSS	R
+ jednoduchý k použití	+ relativně jednoduchý	-není jednoduché (ze začátku)
- Omezené možnosti	+ obsahuje řadu metod	+ nejnovější a etablované metody
	- Řada metod zastaralých	+ soustavný update
- Placený	- Je drahý	+ open source (=zdarma)

Excel	SPSS	R
+ jednoduchý k použití	+ relativně jednoduchý	- není jednoduché (ze začátku)
- Omezené možnosti	+ obsahuje řadu metod	+ nejnovější a etablované metody
	- Řada metod zastaralých	+ soustavný update
- Placený	- Je drahý	+ open source (=zdarma)

- A pak jamovi



První kroky při zpracování a analýze dat

1. Čištění, kontrola a překódování výzkumných dat
2. Stanovení závislé a nezávislé proměnné
3. Zvolení vhodného statistického testu
4. Testování předpokladů
5. Pokud jsou předpoklady ok, použijí se parametrické metody
6. Pokud předpoklady nejsou ok, použijí se neparametrické metody

1. Čištění, kontrola a překódování výzkumných dat
2. Stanovení závislé a nezávislé proměnné
3. Zvolení vhodného statistického testu
4. Testování předpokladů
5. Pokud jsou předpoklady ok, použijí se parametrické metody
6. Pokud předpoklady nejsou ok, použijí se neparametrické metody

Co přesně se myslí termínem „výzkumná data“?

Dva možné významy:

Výzkumná data v širokém pojetí:

- Analytické skripty
- Obrázky, fotografie, filmy
- Protokoly

• Výzkumná data v úzkém pojetí:

- Datová tabulka neboli datový set

Sloupce zobrazují
Jednotlivé proměnné
Např. jméno, věk,
příjem

	A	B	C	D	E
1	Jméno	Příjmení	Věk	Příjem	
2	Petr	Novák	25	15450	
3	Jan	Dvořák	19	18000	
4	Markéta	Janů	36	26500	
5	Tomáš	Štědrý	33	33000	
6	Martina	Svobodová	54	10300	
7	Radek	Tomášek	40	29560	
8					
9					

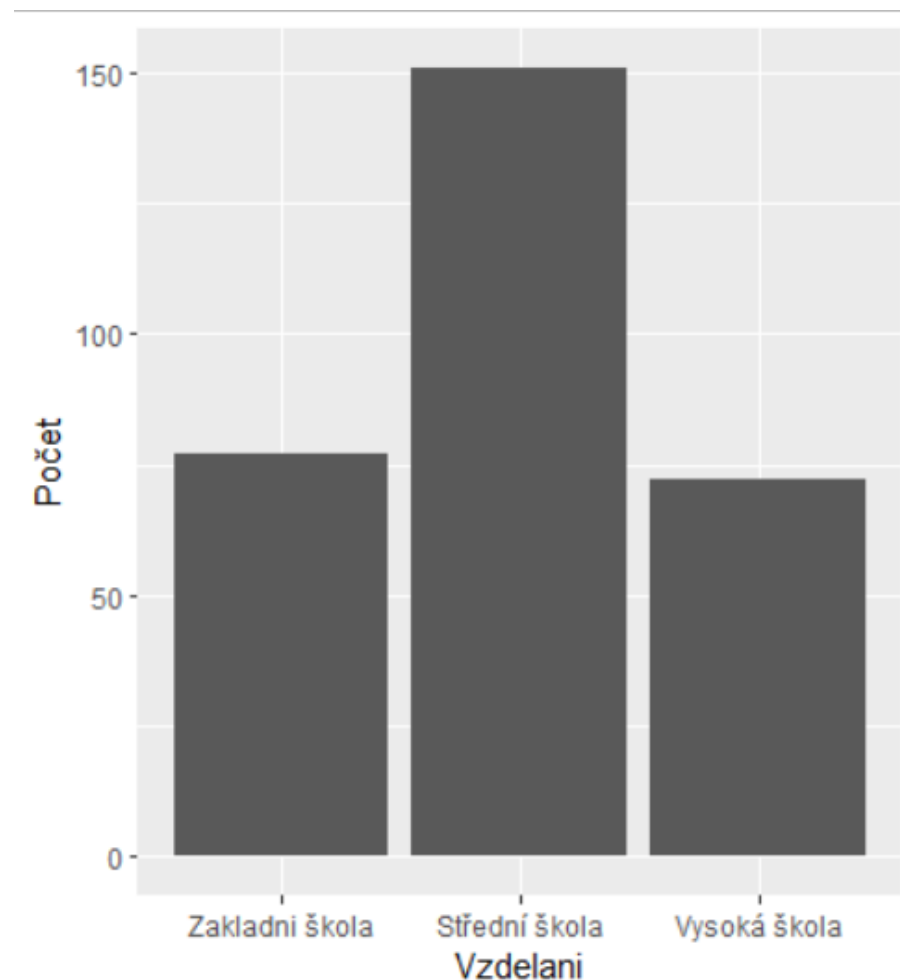
Jednotlivé řádky
představují
respondenty

Čištění a kontrola výzkumných dat

- Formální chyby (např. muž uvádí, že má menstruační obtíže)
- Odlehlé hodnoty – detekce pomocí grafů a IQR funkce v Jamovi

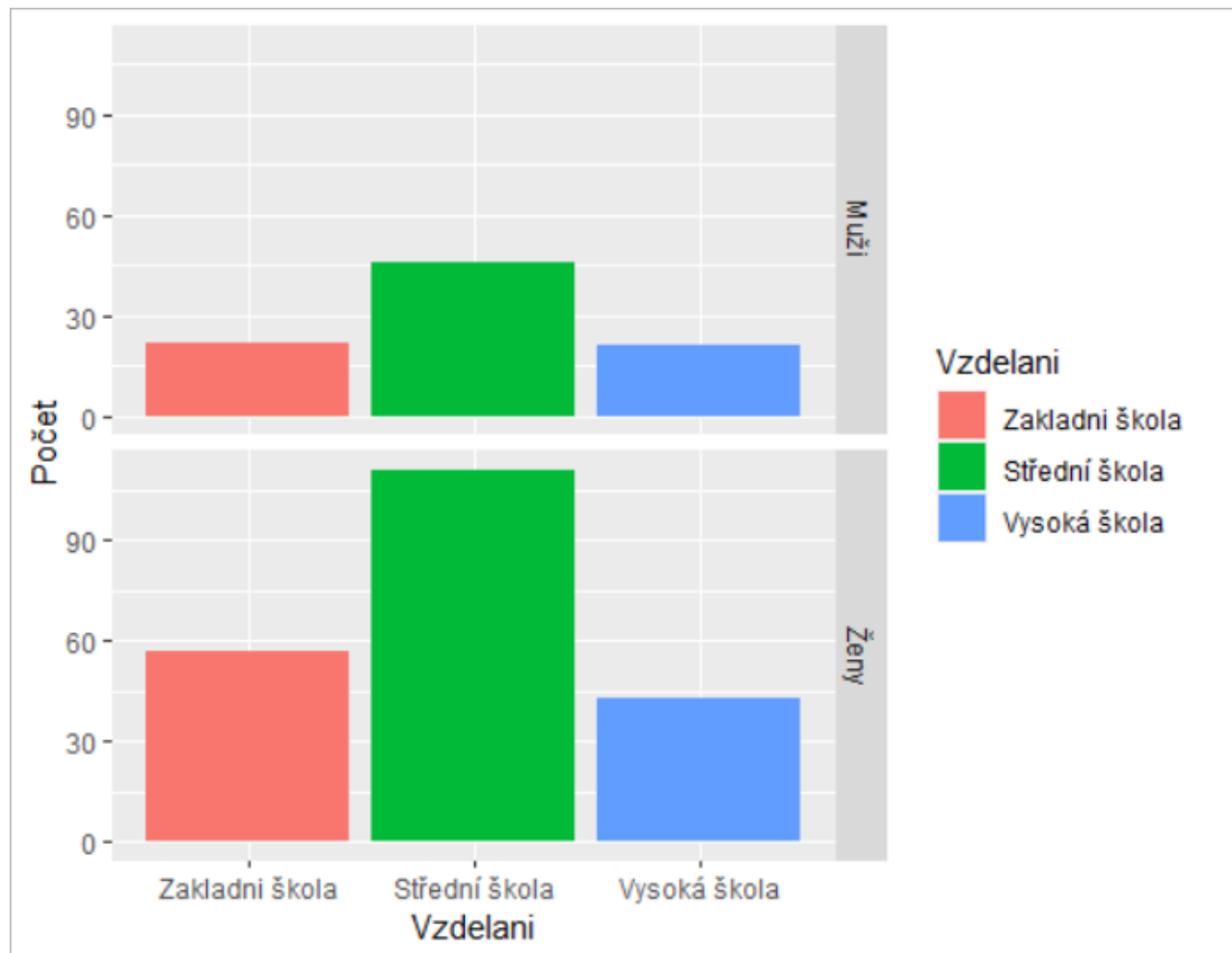
Sloupcový graf (Bar Graph)

- Výška sloupce znázorňuje četnost daného jevu
- Sloupce lze vykreslovat svisle či vodorovně



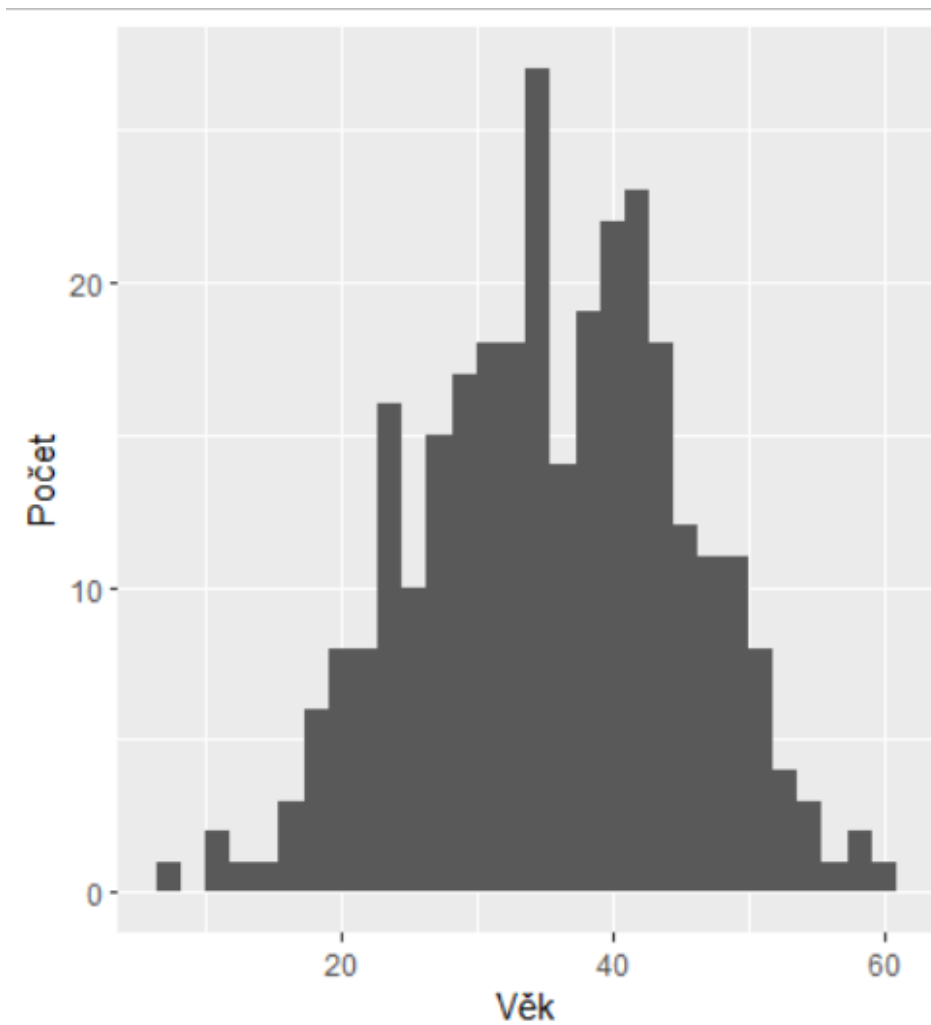
Sloupcový graf (Bar Graph)

- Lze srovnávat i více znaků v jednom grafu

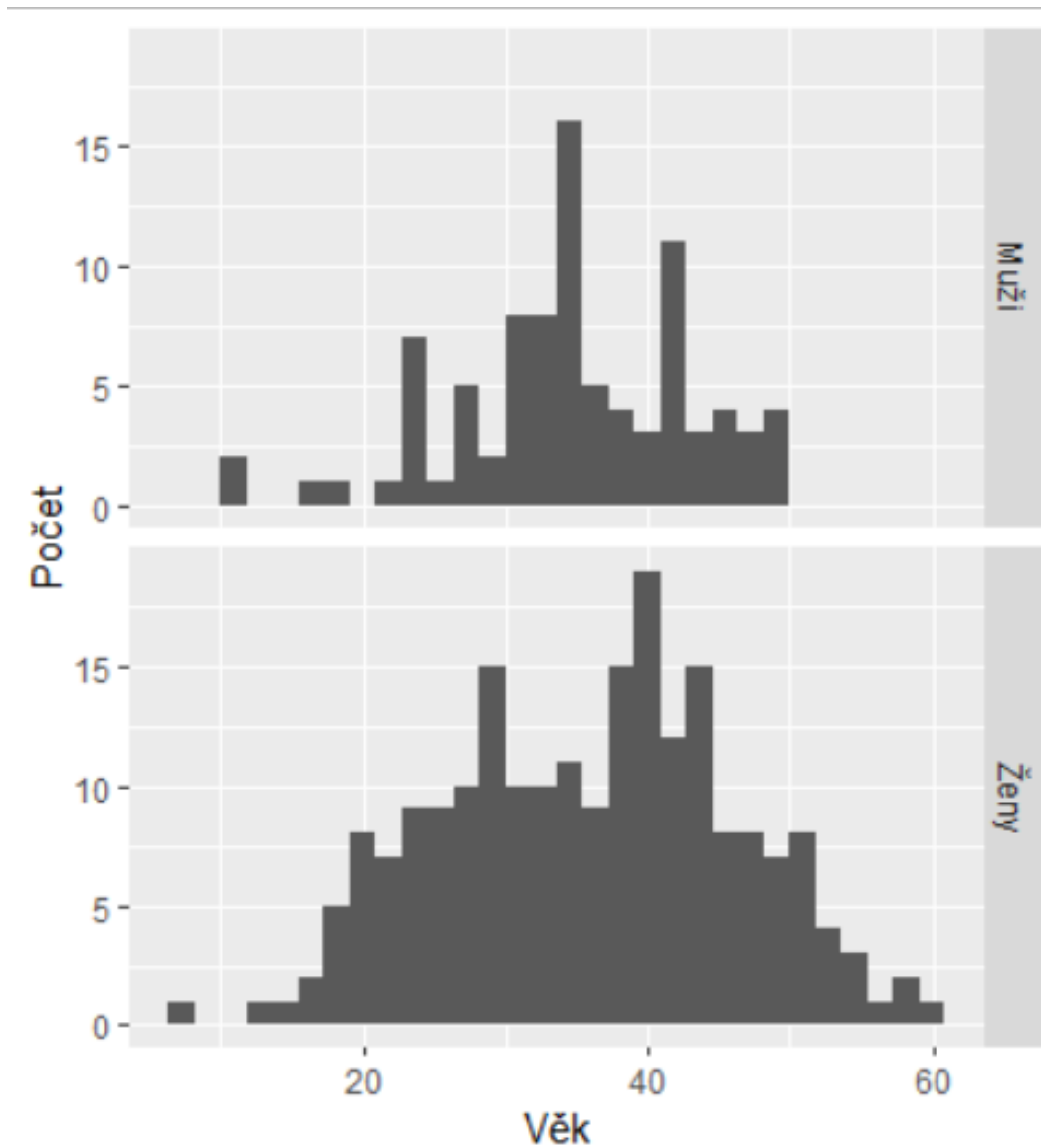


Histogram

- Rozdělení četností spojité veličiny

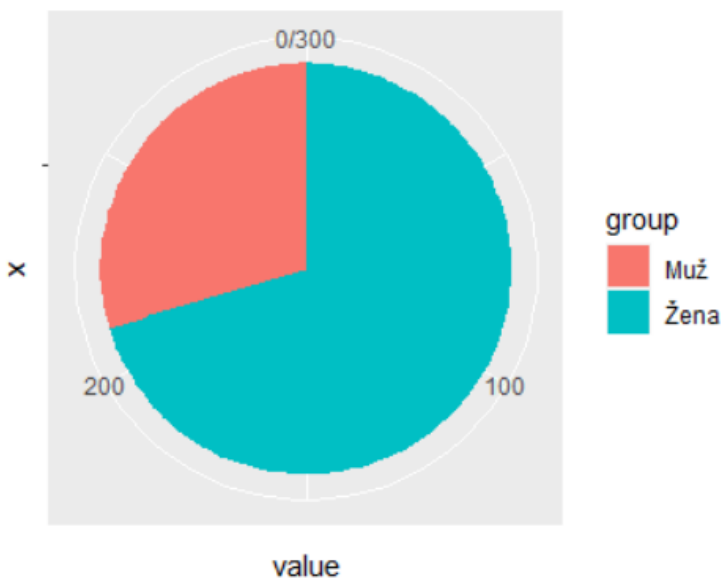


Histogram



- Více znaků
v jednom grafu

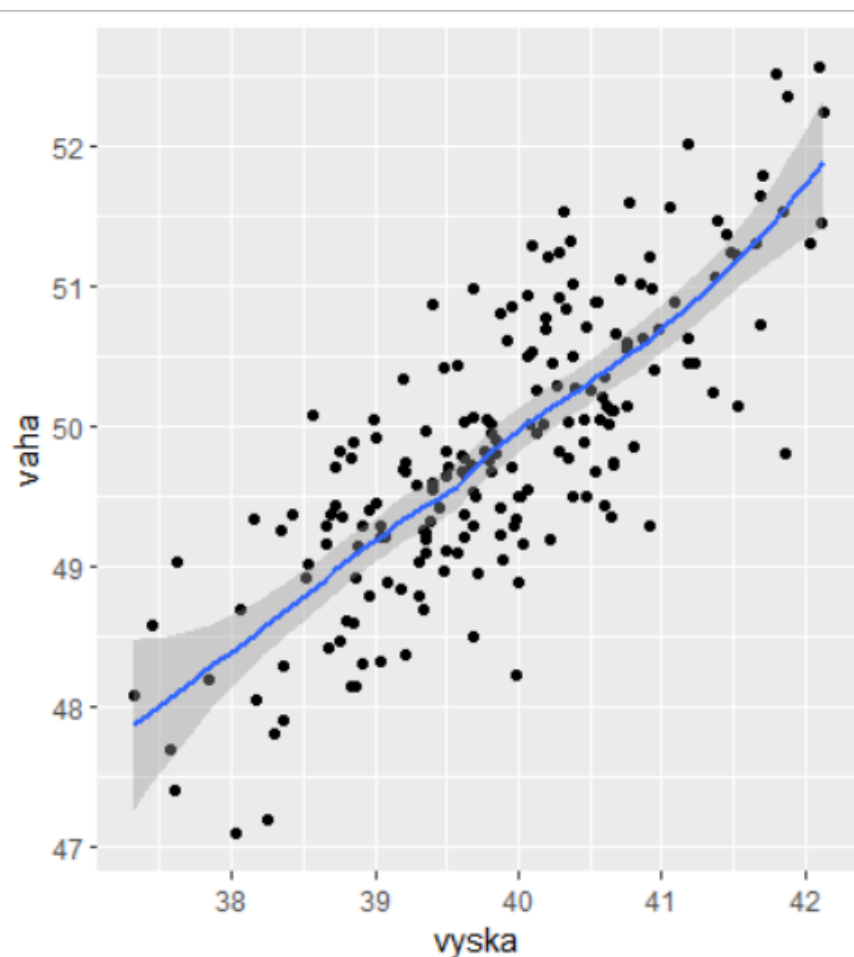
Koláčový graf, kruhový diagram (Pie Chart)



Pohlaví

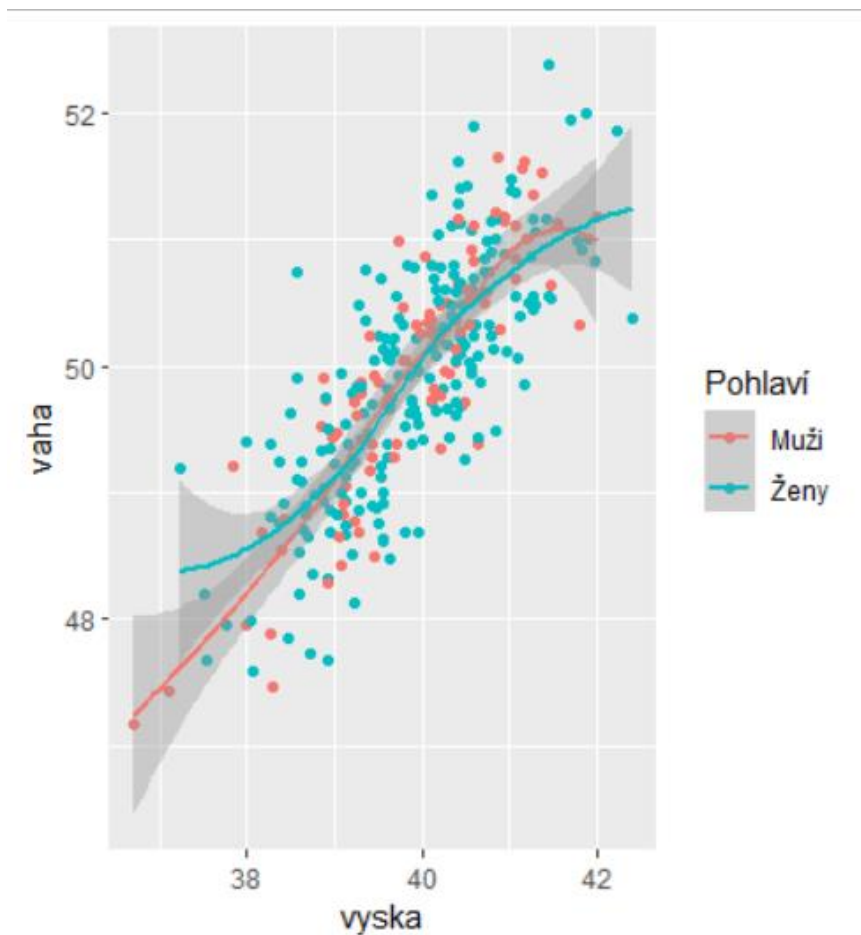
- Plocha kruhu představuje celý soubor
- Jednotlivé části souboru znázorněny výsečemi

Bodový graf (Scatter Plot)



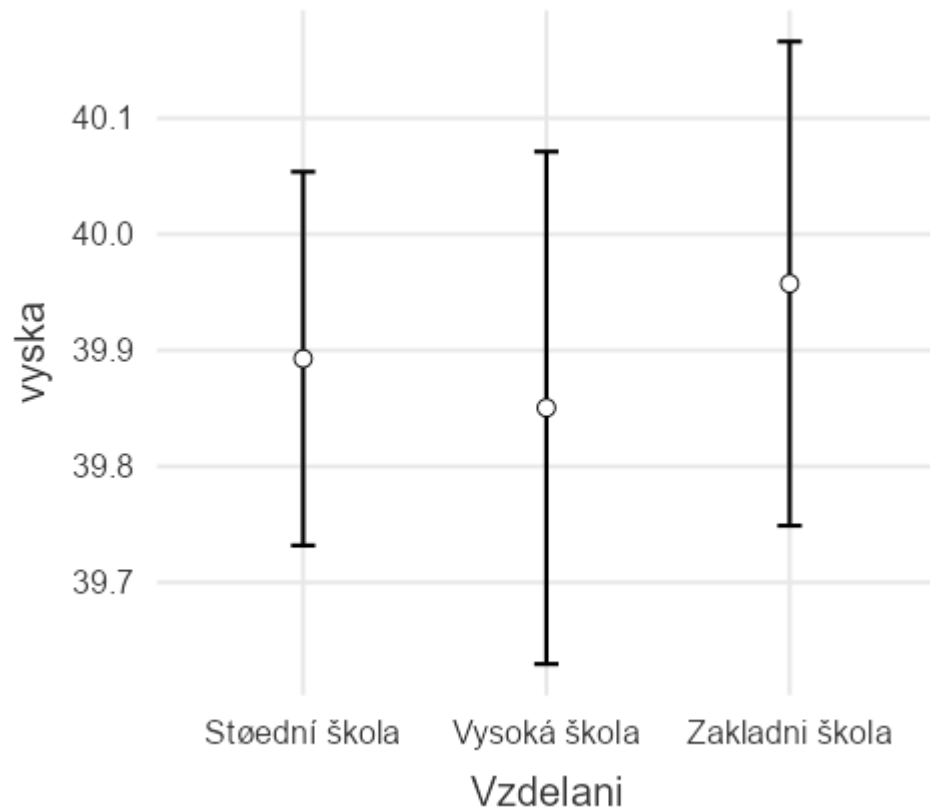
- Zejména pro znázornění závislosti dvou znaků
- Vždy přemýšlejte o smysluplném umístění proměnných
 - x ... prediktor
 - y ... závisí na x

Bodový graf (Scatter Plot)



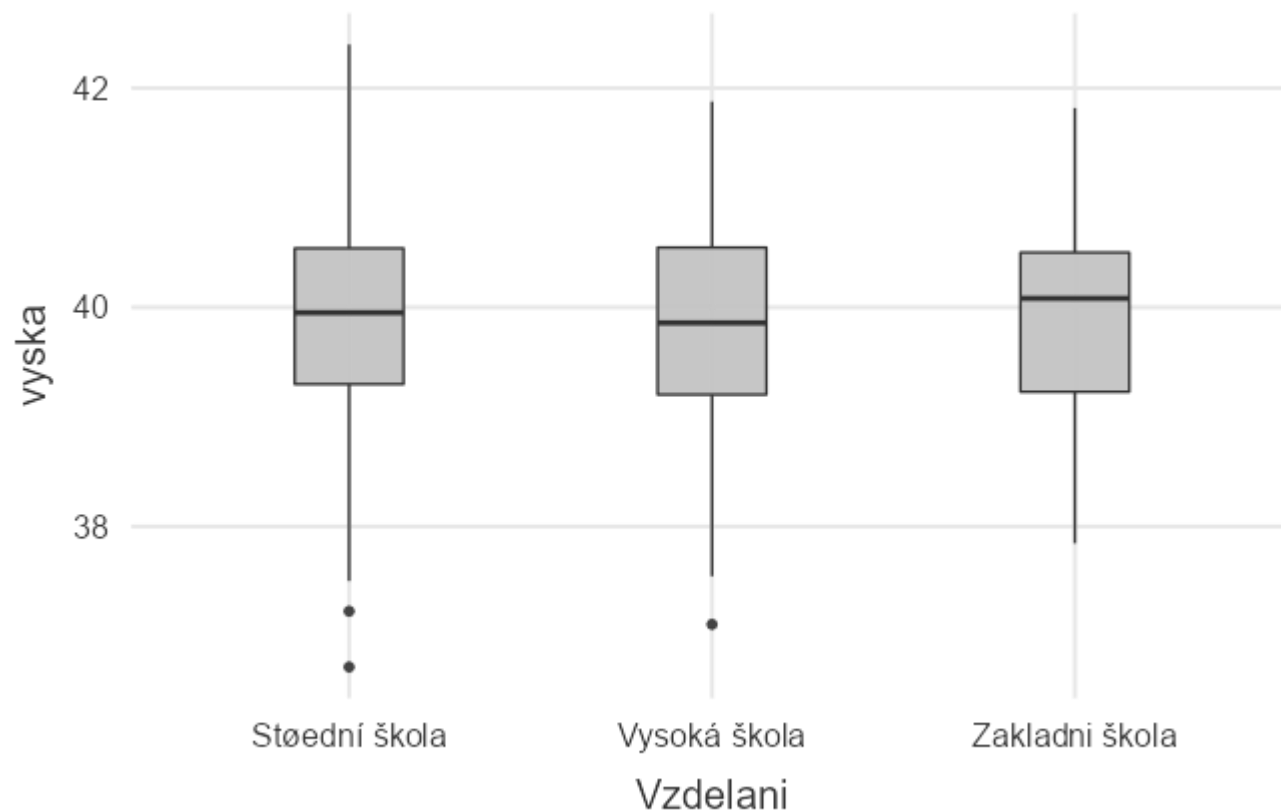
- Více znaků v jednom grafu
- Co zvládneme z takového grafu vyčíst?

Vykreslení průměrů ve skupinách



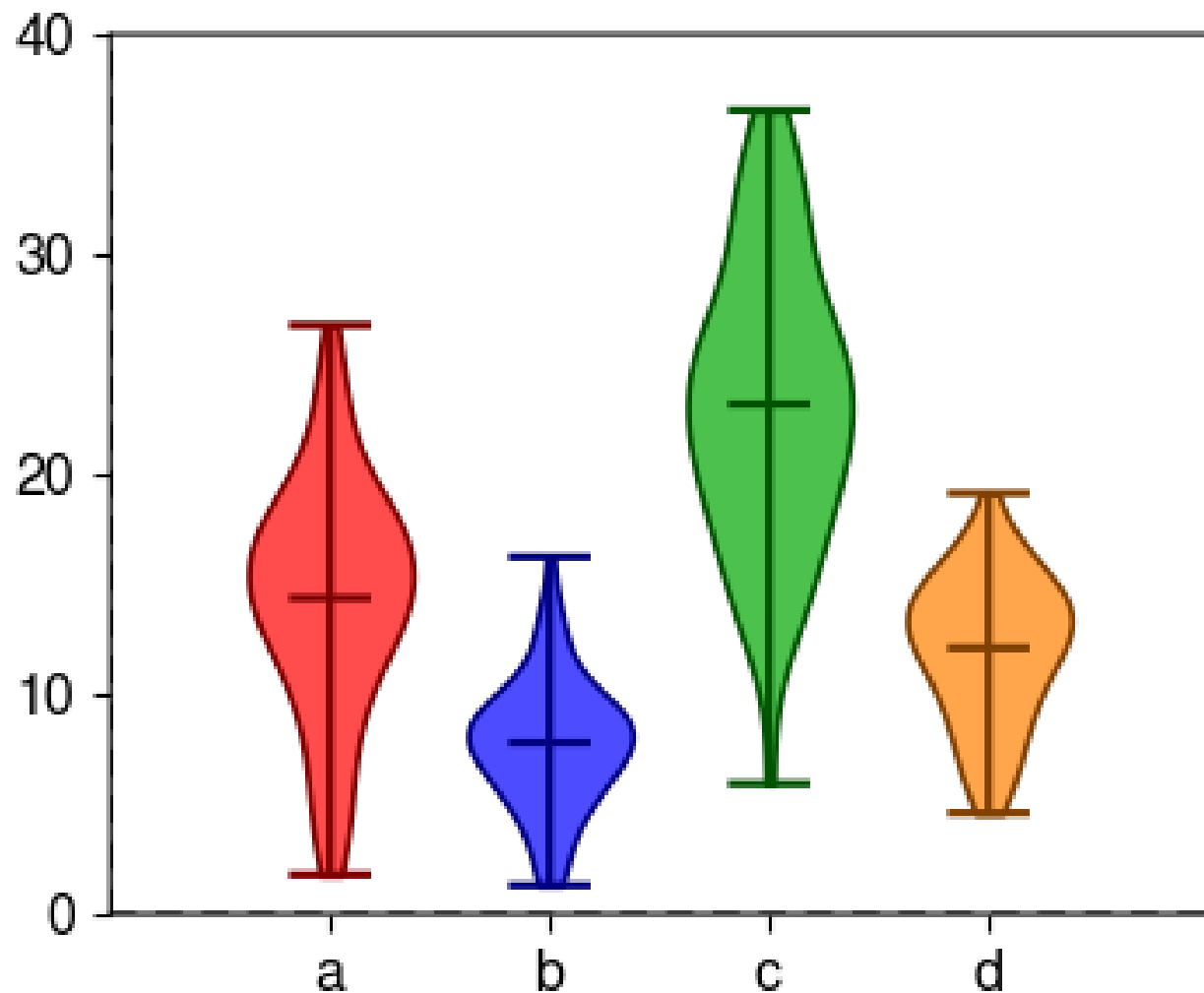
- Jsou vysokoškoláci vyšší než lidé s nižším vzděláním?

Krabicový graf (Box and Whisker Plot)



- Kompletní informace o celé proměnné (výška) v rámci skupin (vzdělání)
- Zadává se tzv. **závislá** proměnná (spojitá, u nás výška) a **grupovací** proměnná (skupiny, kódované v 1 sloupci)

Houslový graf (Violin Plot)



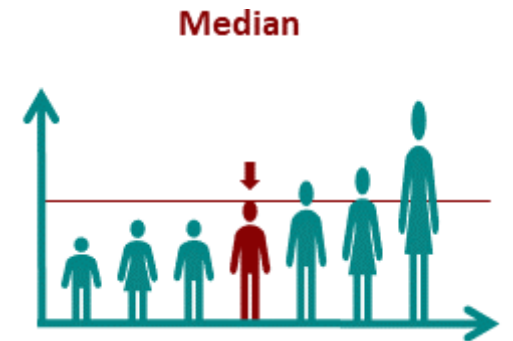
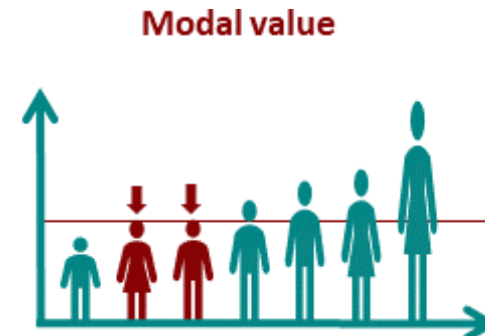
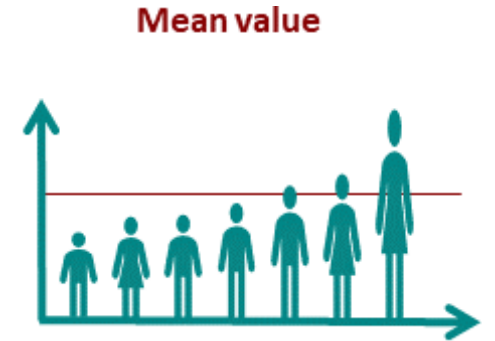
- Kompletní informace o rozdělení proměnné
- Vizualizace, jak se spojitá proměnná (na ose y) liší mezi skupinami (a, b, c, d)
- Umí vykreslit jen některé stat. softwary (R, kód v SPSS)

- **Přemýšlejte o svých grafech.**
- **Jeden obrázek může vydat za tisíc slov!**



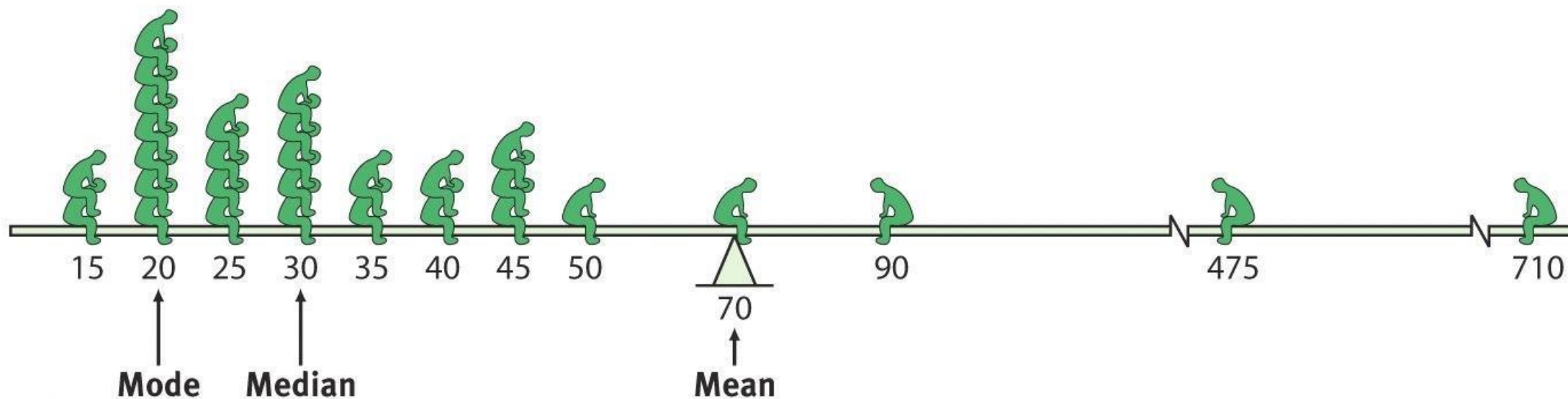
Popisné statistiky

- Nástroj pro organizaci, sumarizaci a porovnání dat
- **Průměr** = střední hodnota
součet všech hodnot vydělený jejich počtem, $\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n)$
- velmi citlivý na odlehlé hodnoty
- **Medián** = prostřední hodnota, dělí soubor na dvě stejně velké skupiny není ovlivněn extrémními hodnotami
- **Modus** = nejčastější hodnota
funguje i pro nominální proměnné



! Pozor na extrémní hodnoty

Měsíční příjem v CZK

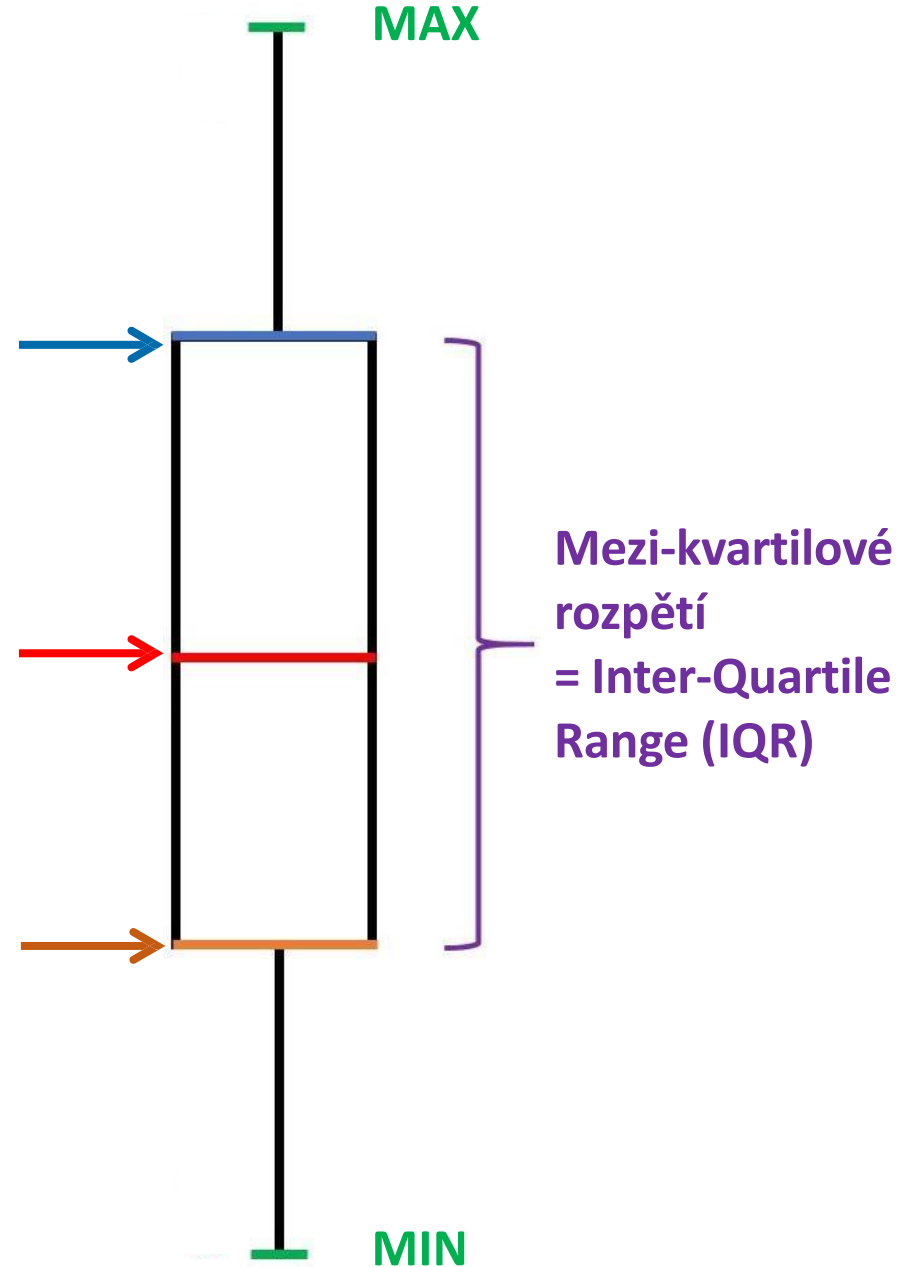


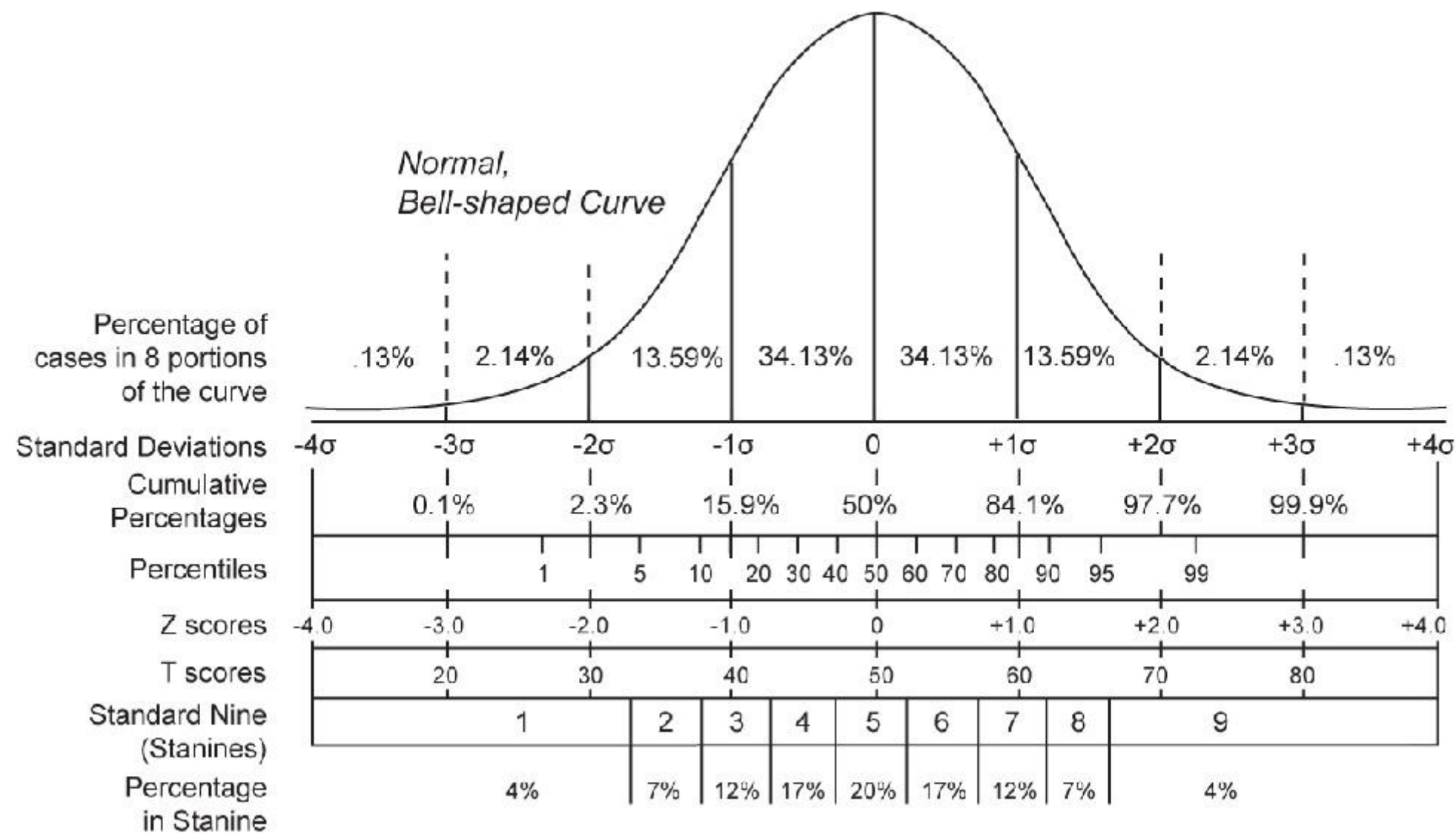
Další popisné statistiky

75. percentil = horní kvartil = Q3

50. percentil = MEDIÁN

25. percentil = dolní kvartil = Q1

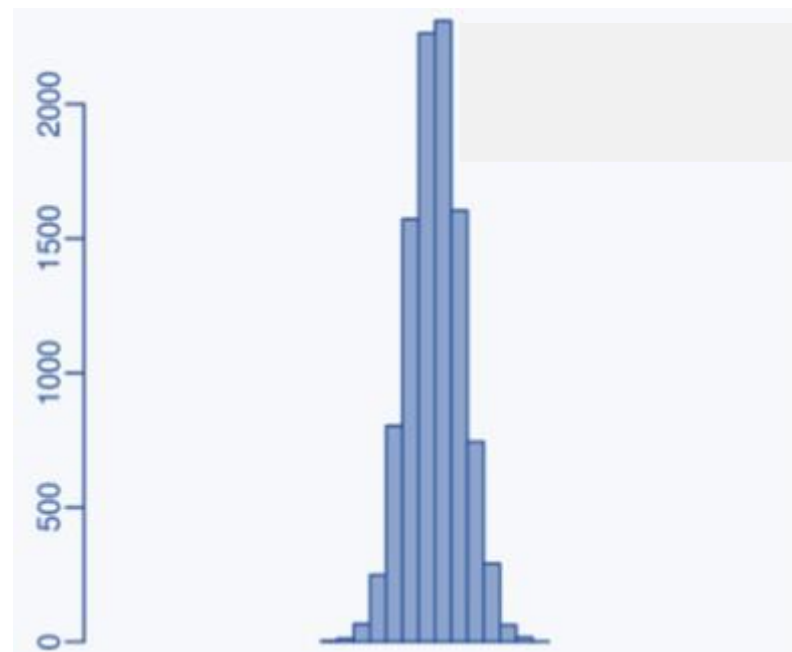
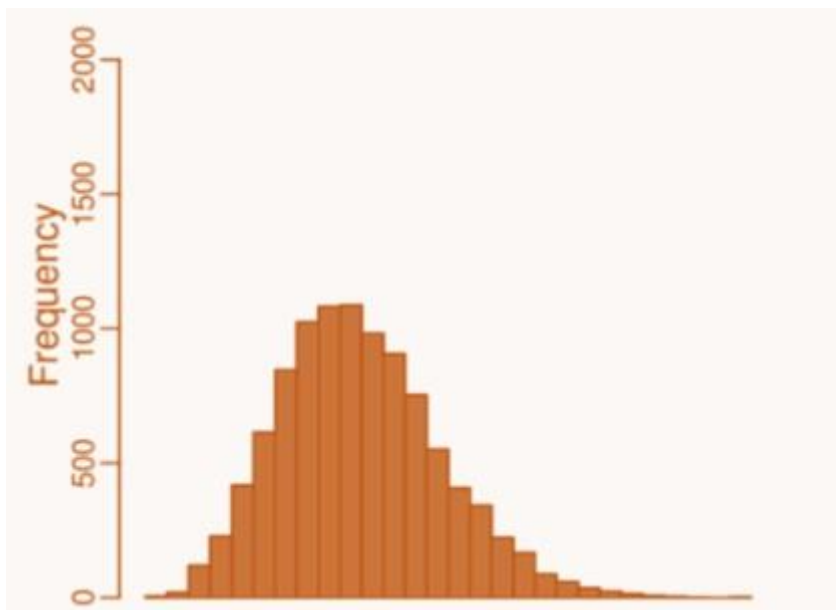




Popisné statistiky – míry variability

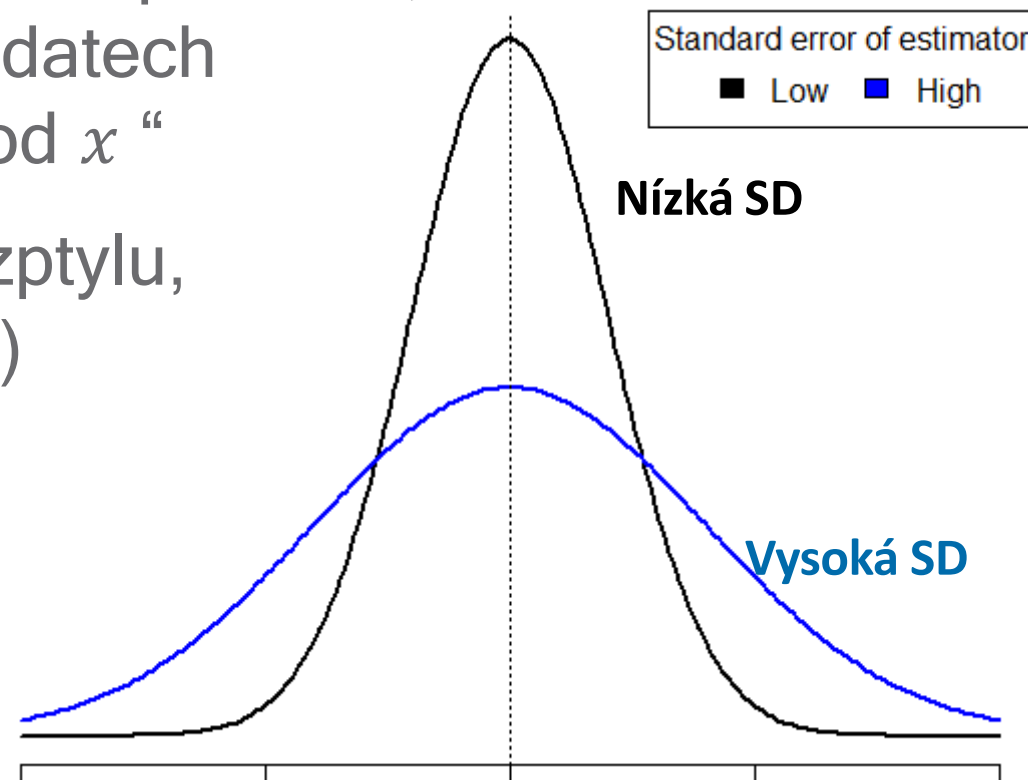


Který z následujících histogramů popisuje data s větší variabilitou?



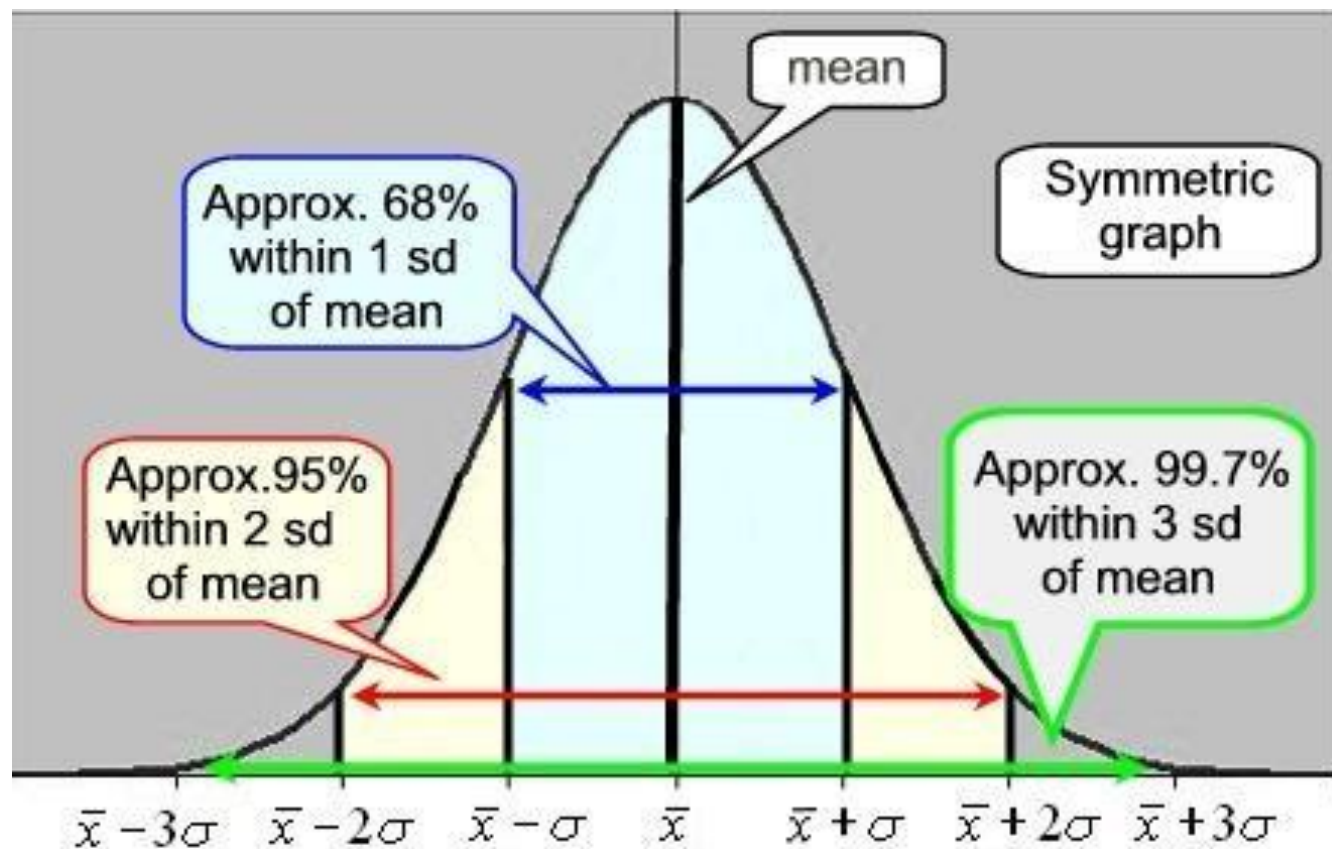
Popisné statistiky – míry variability

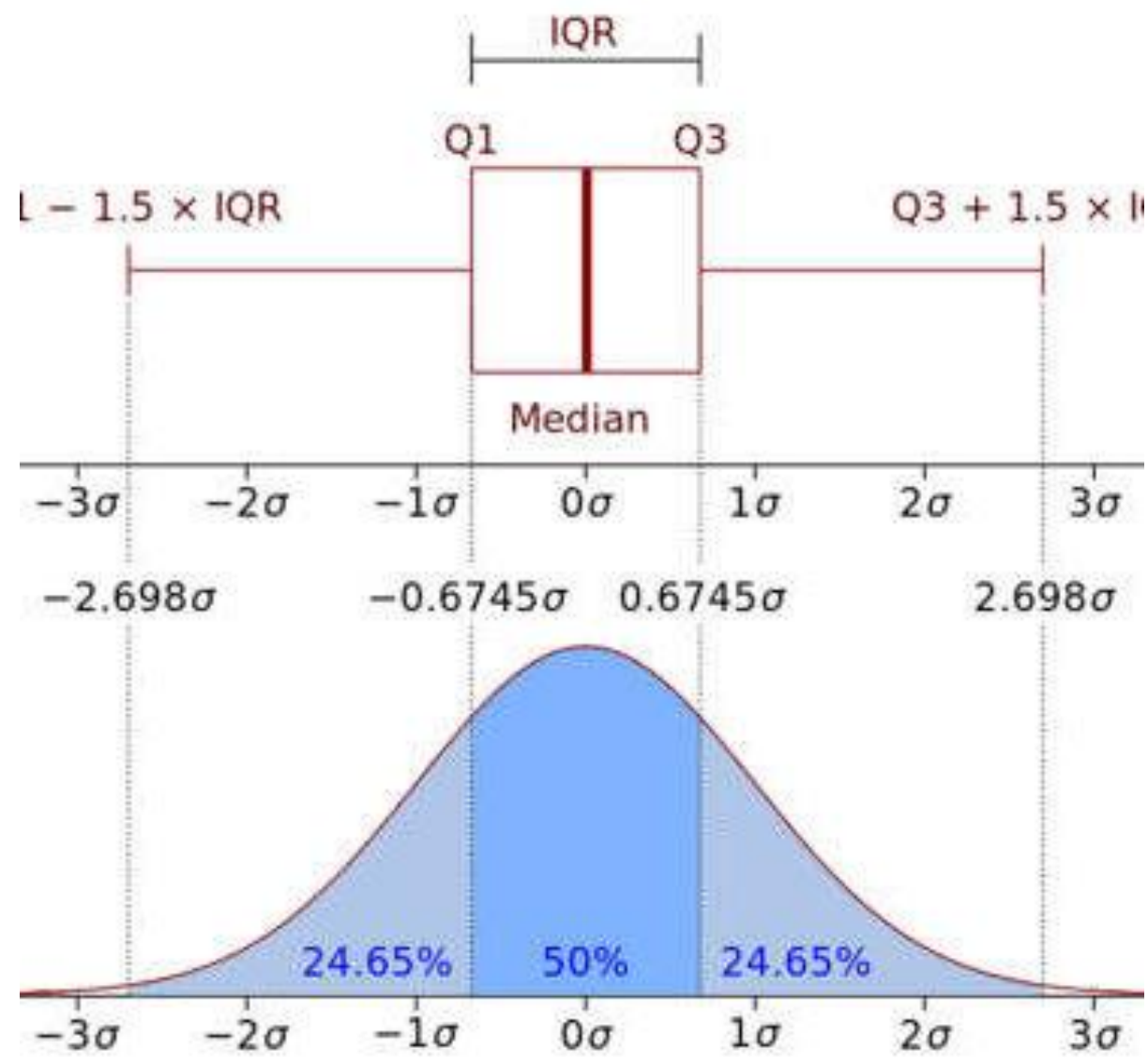
- **Rozptyl** = jak daleko je každé pozorování od průměru, a tedy od každého dalšího pozorování v datech
matematicky: „průměr čtverců odchylek od $\bar{x}$ “
- **Směrodatná odchylka** = odmocnina z rozptylu, značí se σ nebo SD (Standard Deviation)



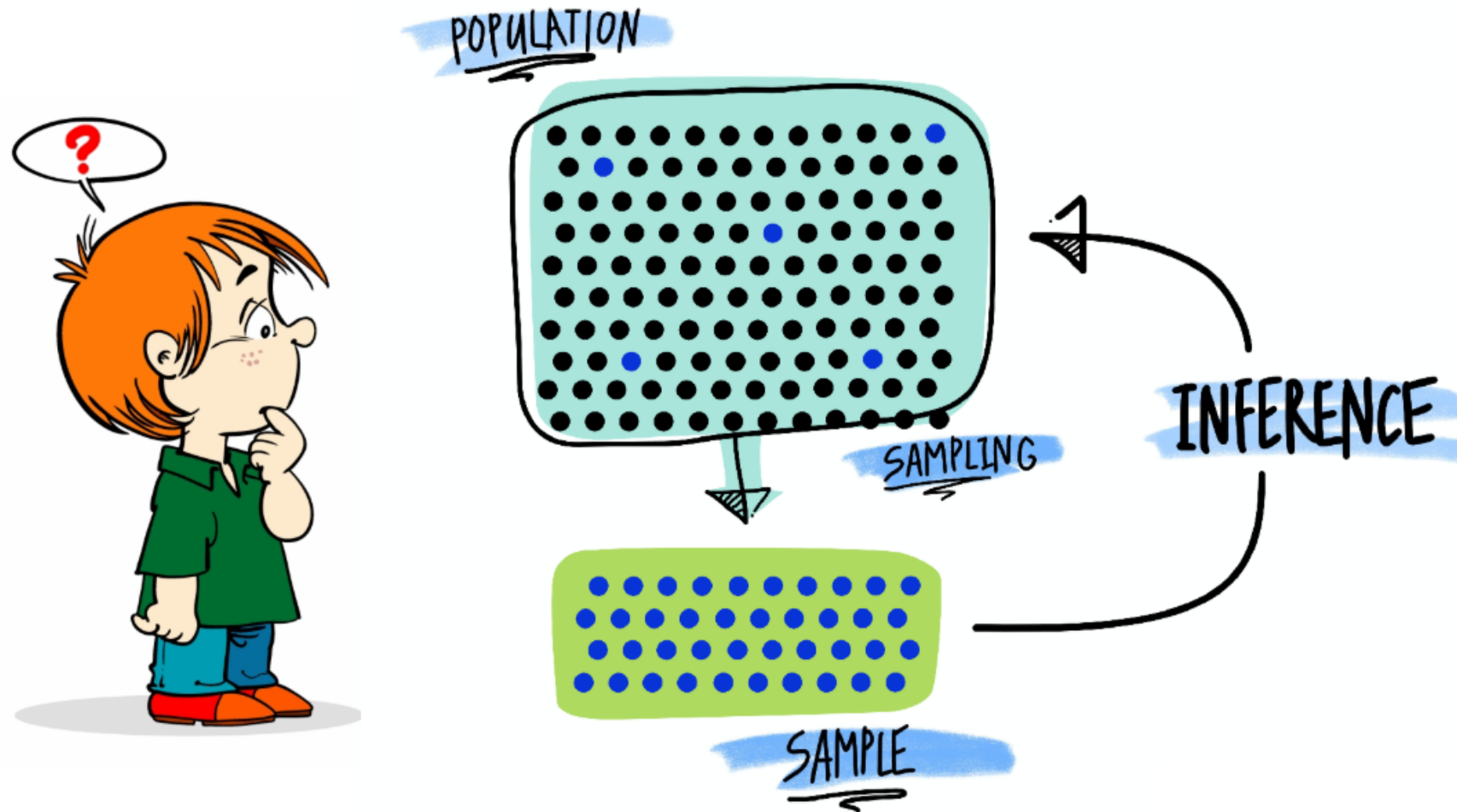
K čemu to je?

- Pokud máme „normální“ data, tj. data s normálním (Gaussovým) rozdělením



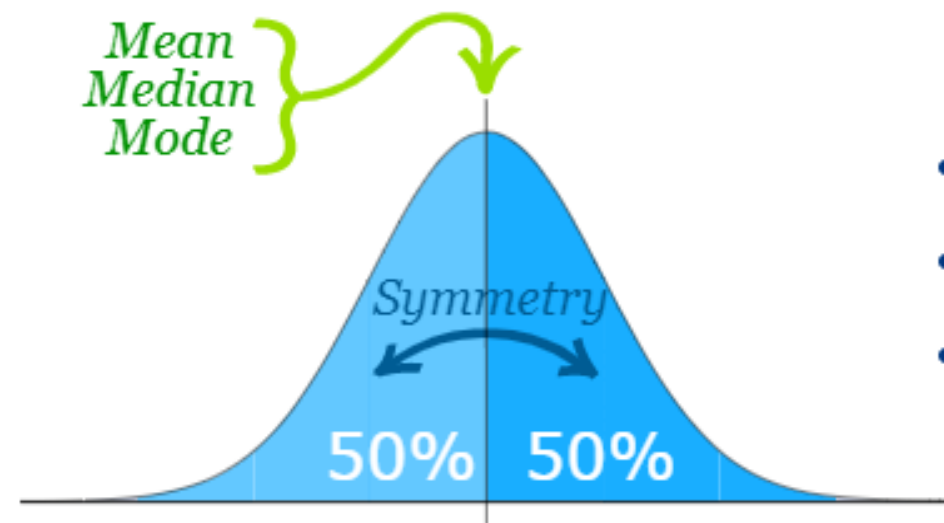


Statistická praxe

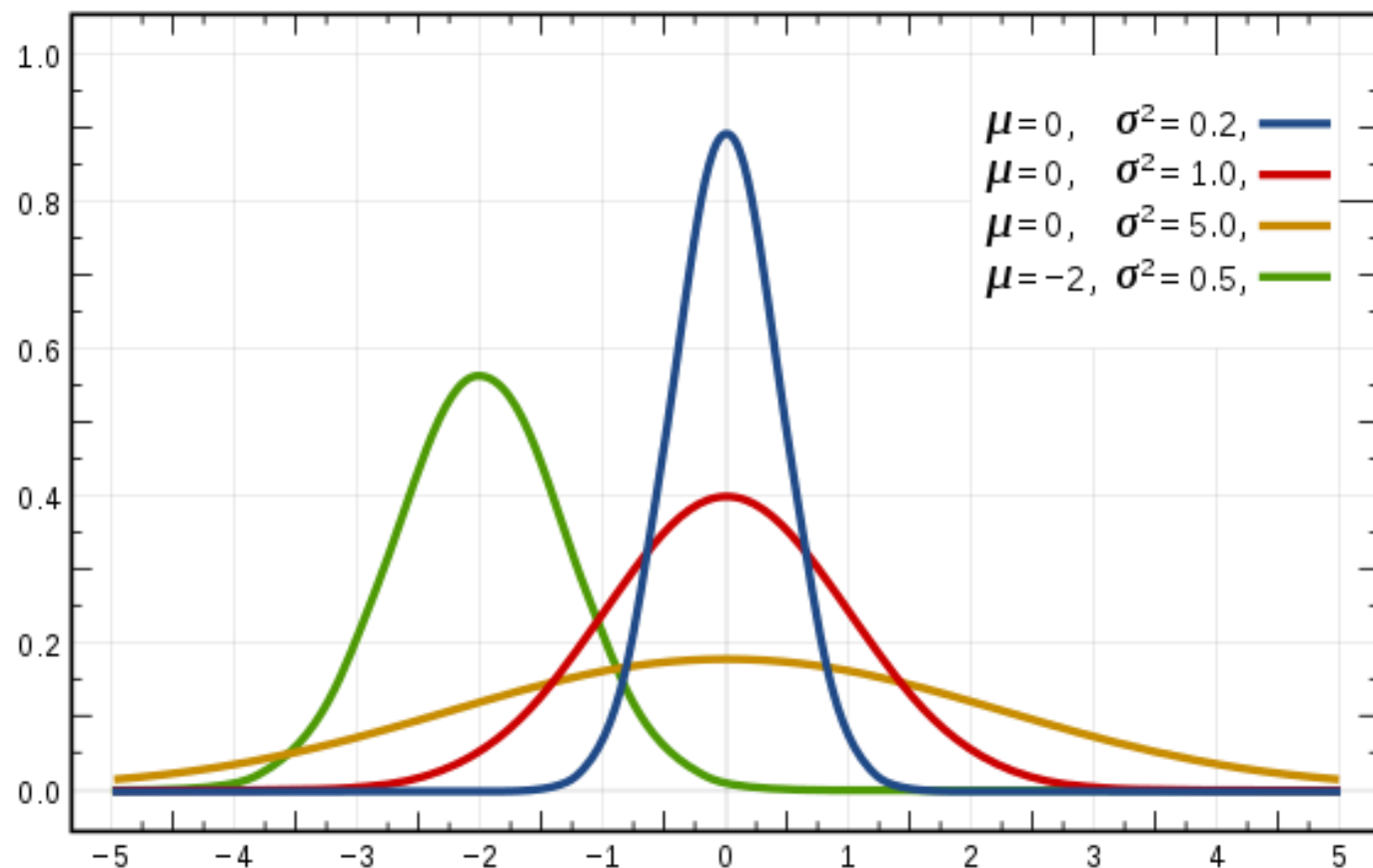


Normální (Gaussovo) rozdělení

- Je symetrické
- Průměr = medián = modus
- Polovina dat je větší než průměr, polovina dat je menší než průměr



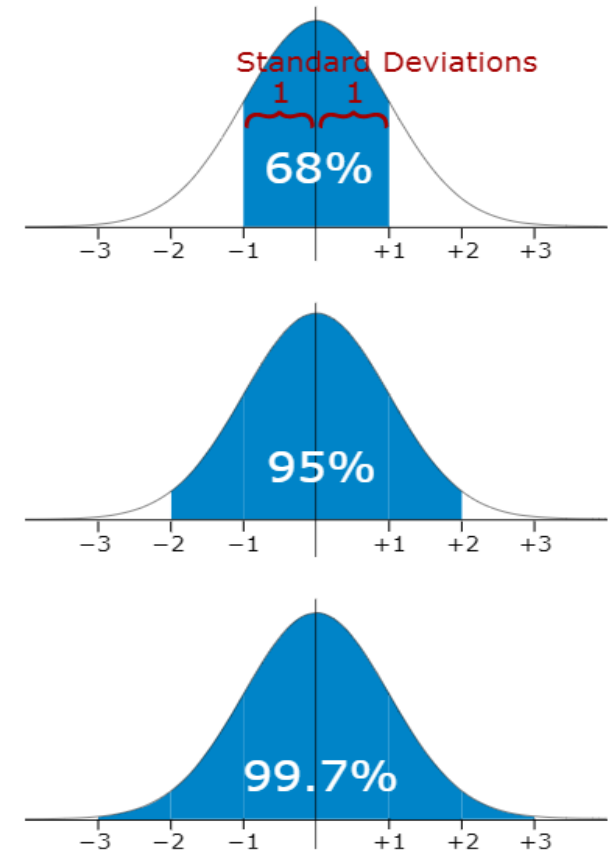
Normální (Gaussovo) rozdělení



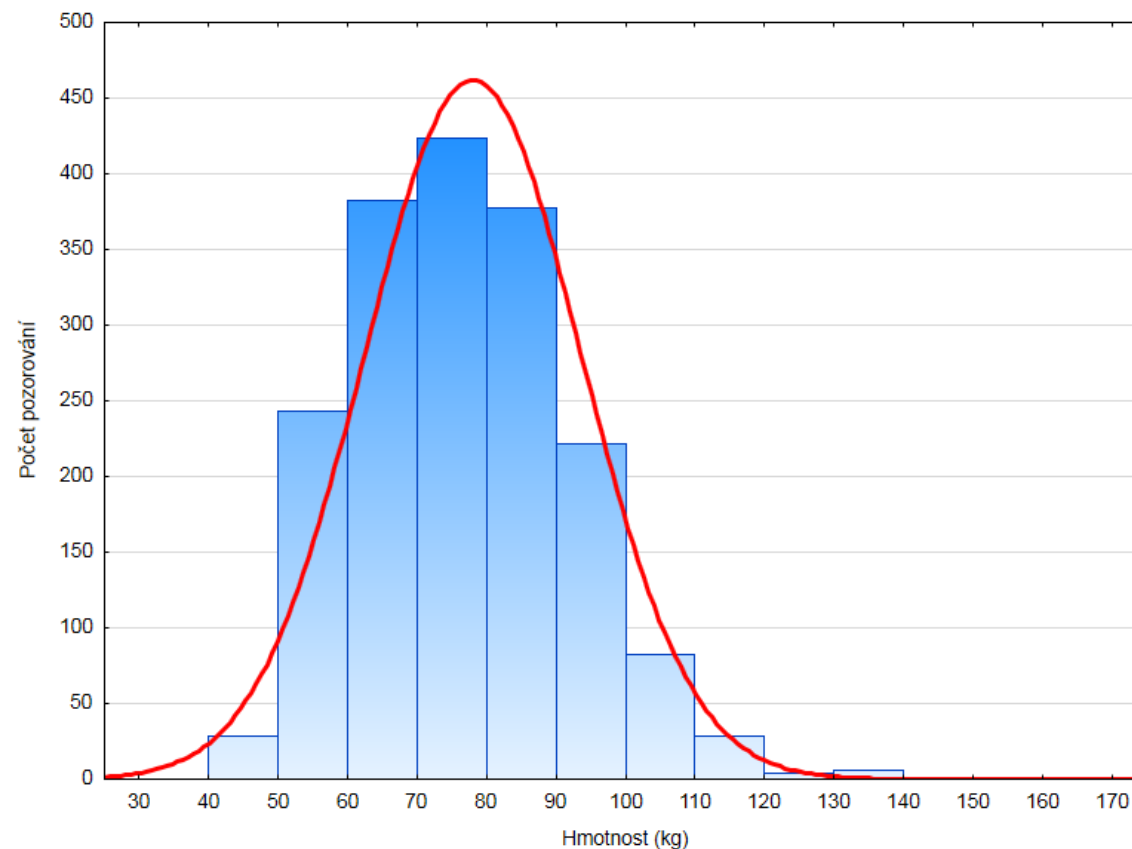
Normální (Gaussovo) rozdělení

- Pokud máme „normální“ data:

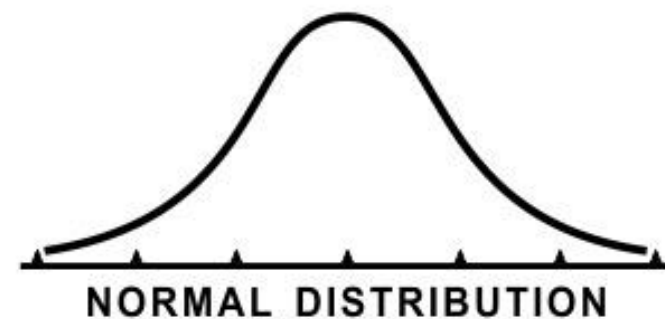
- 68% hodnot leží v intervalu $\bar{x} \pm 1 SD$
- 95% hodnot leží v intervalu $\bar{x} \pm 2 SD$
- 99.7% hodnot leží v intervalu $\bar{x} \pm 3 SD$



V praxi většina dat **nemá** normální
rozdělení



V praxi většina dat nemá normální rozdělení



Co s ne-normálním rozdělením?

- Normalita dat je důležitým předpokladem mnoha statistických metod

- Co když „normální“ data nemáme?



- **Centrální limitní věta**: Když máme hodně respondentů, naše data „se chovají normálně“
- **Robustní metody**: použít metody, které nejsou citlivé na normalitu dat
- **Transformace dat**: matematické kouzlení s původním rozdělením dat
- **Neparametrické metody**: fungují bez předpokladu normality dat

Testování hypotéz

1. Návrh výzkumného projektu

✓ Vymezení cíle a účelu výzkumu.

✓ Formulace hypotéz, které lze ověřit experimentem.

•
⋮

- **Nulová hypotéza:** hypotéza o „nulovém“ (tedy žádném) efektu
 - H_0 : není rozdíl mezi průměrnou výškou mužů a žen
 - H_0 : lidé s VŠ mají průměrně stejný počet dětí jako lidé s nižším vzděláním
- **Alternativní hypotéza:** opak H_0 , neboli efekt není nulový
 - H_1 : průměrná výška žen $\neq$ průměrná výška mužů
 - H_1 : lidé s VŠ mají průměrně méně/více dětí než lidé s nižším vzděláním
- Při každém testu hypotéz musíme mít formulovány obě hypotézy H_0 i H_1

- H_0 (nulová hypotéza): dvě skupiny se ve své tělesné výšce neliší
- H_A : (alternativní hypotéza): dvě skupiny se ve své tělesné výšce liší

• Při ověřování uvedených hypotéz mohou nastat celkově 4 možné

Scénáře:

		Skutečný stav světa	
		H_A je pravdivá	H_0 je pravdivá
Naše zjištění	H_A platí		
	H_0 platí		

- H_0 (nulová hypotéza): dvě skupiny se ve své tělesné výšce neliší
- H_A : (alternativní hypotéza): dvě skupiny se ve své tělesné výšce liší

- Při ověřování uvedených hypotéz mohou nastat celkově 4 možné

Scénáře:

		Skutečný stav světa	
		H_A je pravdivá	H_0 je pravdivá
Naše zjištění	H_A platí	Trefa! :--)	
	H_0 platí		

- H_0 (nulová hypotéza): dvě skupiny se ve své tělesné výšce neliší
- H_A : (alternativní hypotéza): dvě skupiny se ve své tělesné výšce liší

• Při ověřování uvedených hypotéz mohou nastat celkově 4 možné

Scénáře:

		Skutečný stav světa	
		H_A je pravdivá	H_0 je pravdivá
Naše zjištění	H_A platí	Trefa! :--)	
	H_0 platí		Trefa! :--)

- H_0 (nulová hypotéza): dvě skupiny se ve své tělesné výšce neliší
- H_A : (alternativní hypotéza): dvě skupiny se ve své tělesné výšce liší

- Při ověřování uvedených hypotéz mohou nastat celkově 4 možné

Scénáře:

		Skutečný stav světa	
		H_A je pravdivá	H_0 je pravdivá
Naše zjištění	H_A platí	Trefa! :--)	
	H_0 platí	Vedle :--(((	Trefa! :--)

- H_0 (nulová hypotéza): dvě skupiny se ve své tělesné výšce neliší
- H_A : (alternativní hypotéza): dvě skupiny se ve své tělesné výšce liší

- Při ověřování uvedených hypotéz mohou nastat celkově 4 možné

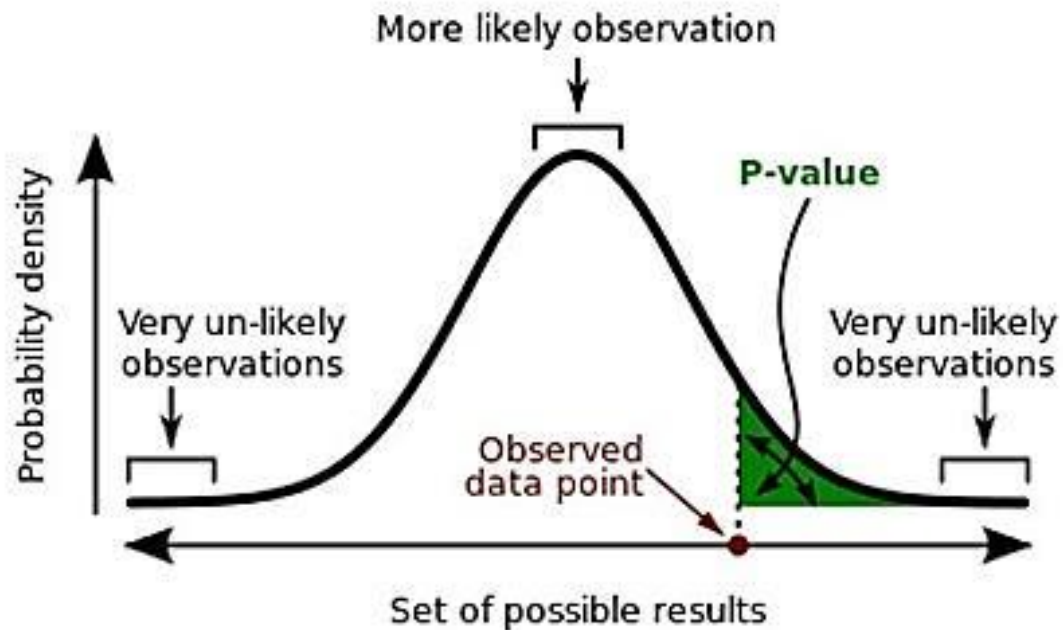
Scénáře:

		Skutečný stav světa	
		H_A je pravdivá	H_0 je pravdivá
Naše zjištění	H_A platí	Trefa! :--)	Falešný alarm :--(((
	H_0 platí	Vedle :--(((	Trefa! :--)

Princip testování hypotéz

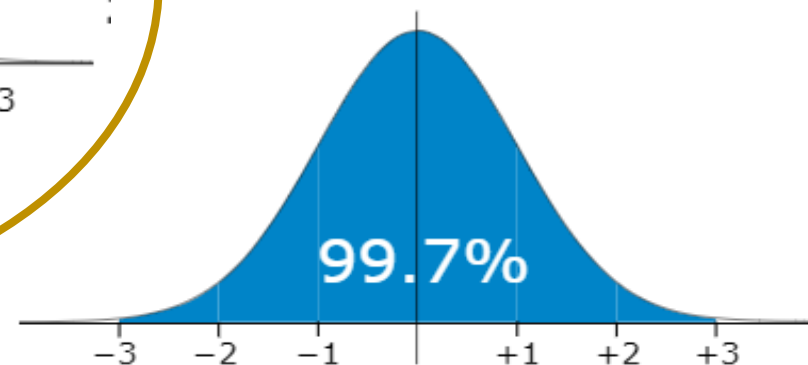
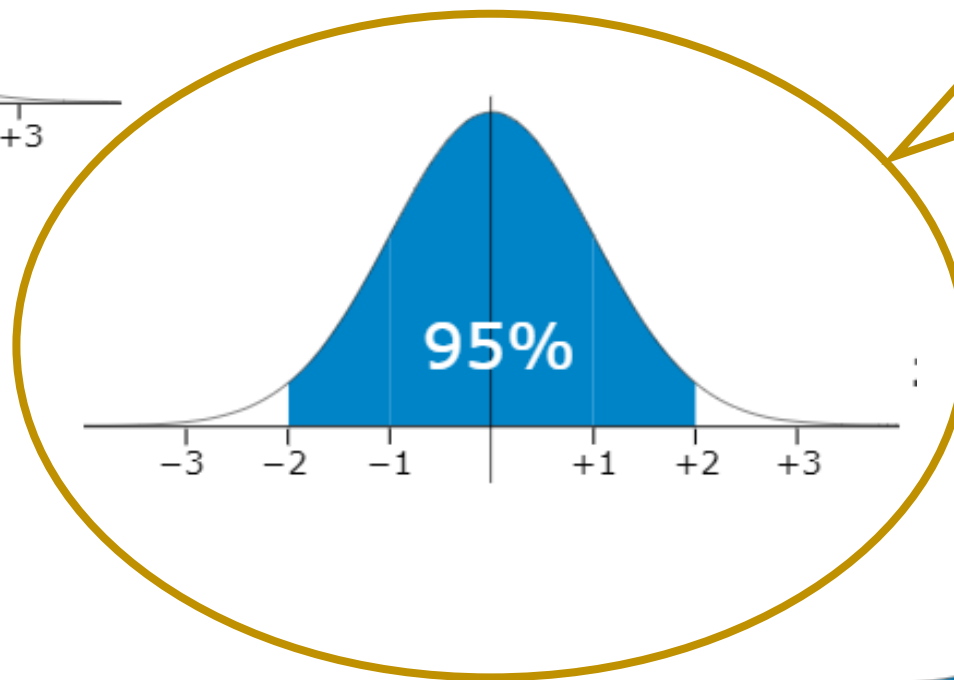
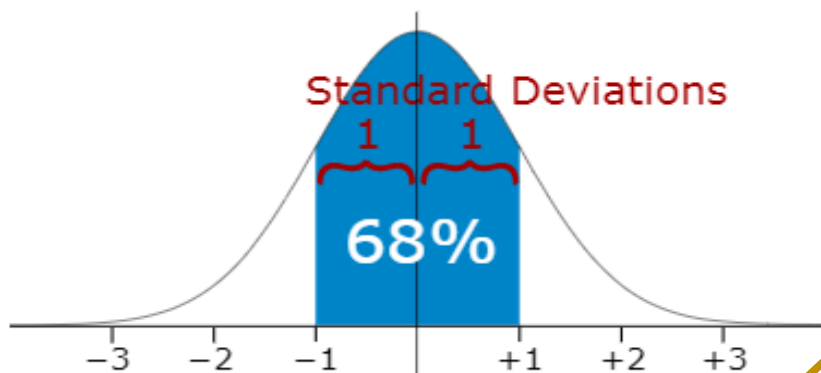
1. Zformulujeme hypotézy
2. Nasbíráme data
3. **Vypočítáme** pravděpodobnost, s jakou bychom mohli obdržet pozorovaná data nebo data ještě více odporující nulové hypotéze, **za předpokladu, že H_0 je pravdivá – pravděpodobnost „falešného alarmu“**
 - ***P-hodnota = (p-value)***
 - Čím menší je p , tím méně důvěryhodná je H_0
 - Obvykle, je-li $p < 0,05$, zamítáme hypotézu H_0 a přikláníme se k alternativní hypotéze H_1

Princip testování hypotéz - graficky



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

Připomenutí normálního rozdělení



Toho využíváme
při testování hypotéz

Příklad

- H_0 : **není rozdíl** mezi průměrnou tělesnou výškou mužů a žen v ČR
- H_1 : průměrná výška žen $\neq$ průměrná výška mužů
- Nasbíráme data
- Nakreslíme **vhodný** graf, vyčíslíme p-value
- Interpretujeme výsledek

<https://rpsychologist.com/d3/nhst/>

Testování normality

- H_0 : testovaná proměnná **má** normální rozdělení
- H_1 : testovaná proměnná **nemá** normální rozdělení

➤ Testy: Shapiro-Wilk, Kolmogorov-Smirnov

➤ Pokud vyjde $p < 0,05$, zamítáme hypotézu H_0 , čili zamítáme normalitu dat

➤ Grafy:

- **histogram** s naitovanou křivkou normálního rozdělení
- **Q-Q plot**: v případě normálního rozdělení tvoří data v QQ grafu více-méně rovnou přímkou

Spojitá proměnná, porovnání skupin

- Např. věk, výška, váha – je rozdíl mezi skupinami?
- H_0 : průměrná hodnota **je stejná** ve všech skupinách
- H_1 : průměrná hodnota **není stejná** ve všech skupinách

➤ Testy:

- Data jsou normální: **t-test** (porovnání 2 skupin), **ANOVA** (3 a více skupin)
- Data nejsou normální: **Mann-Whitney** (porovnání 2 skupin), **Kruskal-Wallis** (3 a více skupin)

➤ Grafy: **Box plot** (krabicový graf), **Violin plot**

Analýza kategoriálních dat

- H_0 : hodnoty jedné proměnné **nemají vztah** k hodnotám druhé proměnné
- H_1 : hodnoty jedné proměnné **mají vztah** k hodnotám druhé proměnné
- Např:

	Otec		
Matka	Obézní	Neobézní	Celkem
Obézní	15	9	24
Neobézní	7	19	26
Celkem	22	28	50

p = 0.011

**Závěr: Zamítáme hypotézu H_0 ,
a přikláníme se k závěru, že obezita
rodičů spolu významně souvisí.**

- Hodnoty: **Kontingenční tabulka** (Crosstabulation)
- Testy: χ^2 (chi-square, chí-kvadrát) – není předpoklad normality dat, ale v políčkách tabulky nesmí být 0
- Grafy: **Sloupcový graf** dělený podle skupin

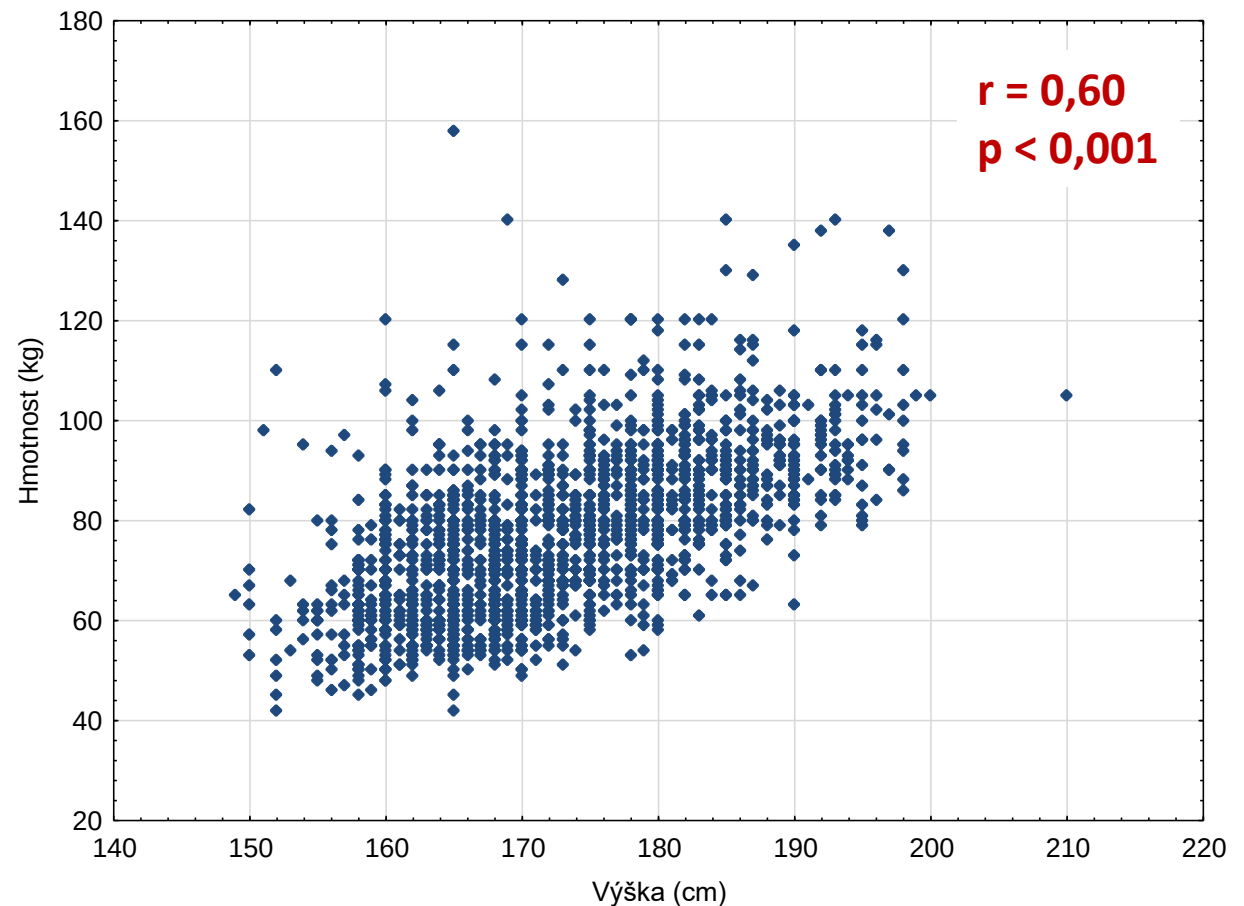
Korelační koeficient

- Vztah mezi dvěma spojitými proměnnými
- H_0 : **není** vztah mezi proměnnými
 - $r = 0$
- H_1 : **je** vztah mezi proměnnými
 - $r \neq 0$

➤ Korelační koeficienty:

- Data jsou normální: **Pearsonův** (r)
- Data nejsou normální: **Spermanův** (ρ)

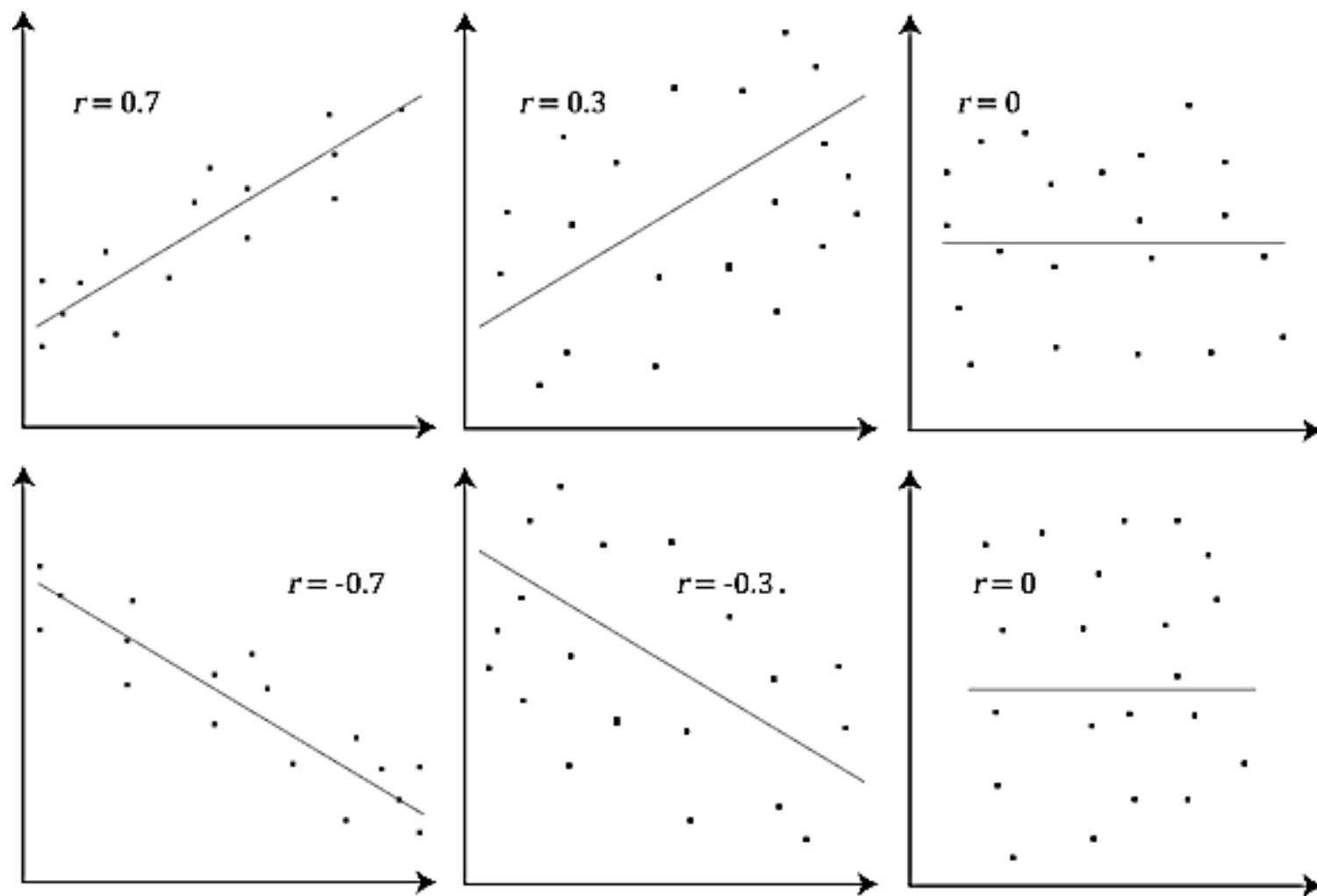
➤ Grafy: **Bodový graf**



Korelační koeficient

- Nabývá hodnot **od -1 do +1**
- Znaménko korelačního koeficientu:
 - **Kladná korelace** = přímá úměra (čím větší x , tím větší y)
 - **Záporná korelace** = nepřímá úměra (čím větší x , tím menší y)
 - **Nula** = není vztah mezi proměnnými
- Čím větší absolutní hodnota korelačního koeficientu, tím silnější vztah mezi proměnnými
 - -0,30 až 0 nebo 0 až 0,30 ... **slabý nebo žádný** vztah
 - -0,3 až -0,50 nebo 0,30 až 0,50 ... **středně silný** vztah
 - -1 až -0,50 nebo 0,50 až 1 ... **silný** vztah

Korelační koeficient



T-test

- Slouží k porovnání 2 skupin v numerické proměnné
- Např. rozdíl v tělesné výšce mezi muži a ženami
- Existuje varianta t-testu pro opakovaná měření (např. míra deprese před a po psychoterapii)
- Má několik předpokladů:
 - Normální rozložení
 - Homogenita rozptylů

t-value

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\bar{x}_1$: Mean value of the first group

$\bar{x}_2$: Mean value of the second group

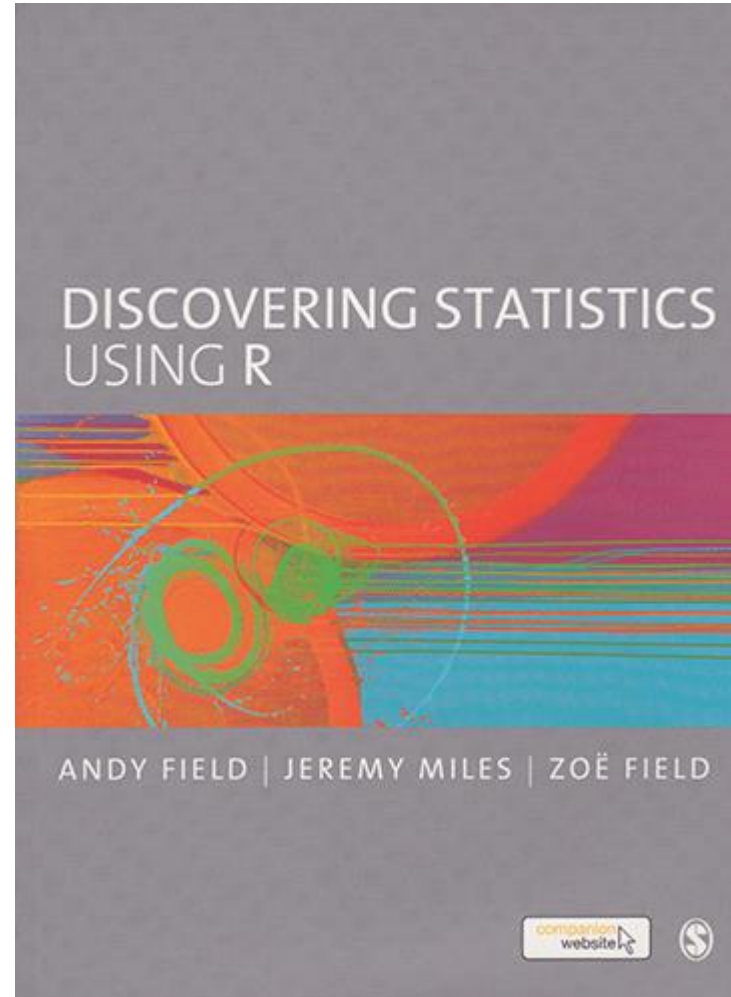
n_1 : Size of the first group

n_2 : Size of the second group

s_1 : Standard deviation of the first group

s_2 : Standard deviation of the second group

Dataset – viagra (z knihy A. Fielda)



Dataset – viagra (z knihy A. Fielda)

- Třem skupinám byla administrována viagra: první skupině byla podána nízká dávka
- Druhé skupině vysoká dávka a třetí skupině bylo podáváno placebo
- Závislou proměnnou tvořila sexuální touha

placebo	nízká dávka	vysoká dávka
3	5	7
2	2	4
1	4	5
1	2	3
4	3	6

Deskriptivní statistika

- Vždy začínáme deskriptivní statistikou!
- placebo: $M = 2.2$, $SD = 1.3$
- nízká dávka: $M = 3.2$, $SD = 1.3$
- vysoká dávka: $M = 5.0$, $SD = 1.58$

Reportování výsledků:

- Výsledky se raportují následovně:

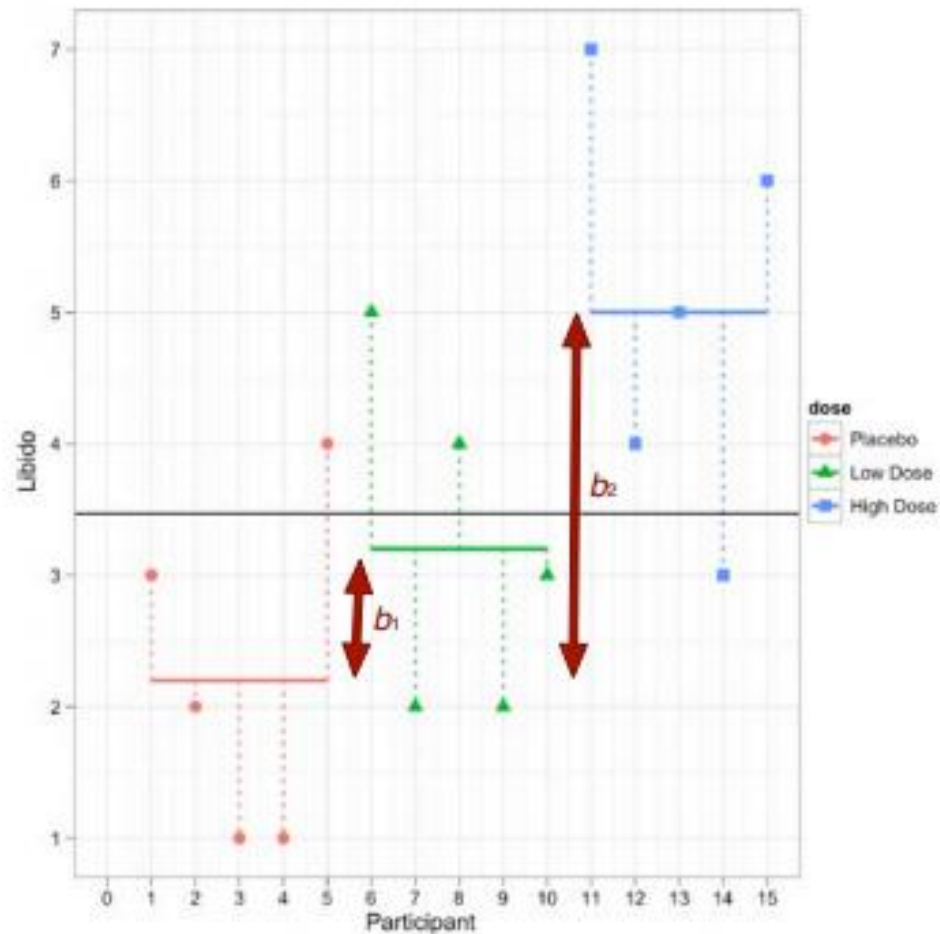
Byl zjištěn statisticky významný rozdíl ($t(8) = -3.06$, $p = .016$, $d = -1.93$) v míře sexuální touhy mezi respondenty konzumující vysoké dávky ($M = 5.0$, $SD = 1.58$) viagry a mezi respondenty užívající placebo ($M = 2.2$, $SD = 1.3$).

Analýza rozptylu (ANOVA)

- Statistický test umožňující porovnat více než 2 skupiny
- Má několik předpokladů:
 - Normální rozložení
 - Homogenita rozptylů

Vizualizace dat

- Horizontální přímká znázorňuje k průměr jednotlivých skupin



Reportování výsledků ANOVy

One-Way ANOVA

person

Dependent Variables

libido

Grouping Variable

dose

Variances

☐ Don't assume equal (Welch's)

☒ Assume equal (Fisher's)

Missing Values

☒ Exclude cases analysis by analysis

☐ Exclude cases listwise

Additional Statistics

☐ Descriptives table

☐ Descriptives plots

Assumption Checks

☐ Normality (Shapiro-Wilk)

☐ Normality (Q-Q plot)

☐ Equality of variances

Post-Hoc Tests

ANOVA

ANOVA

	Sum of Squares	df	Mean Square	F	p	η^2p
dose	20.1	2	10.07	5.12	0.025	0.460
Residuals	23.6	12	1.97			

$F(2, 12) = 5.12, p = .025, \eta_p^2 = .46$

Nalezli jsme **velký** **signifikantní** efekt podávané dávky viagry na sexuální touhu pokusných osob

Měření závislosti

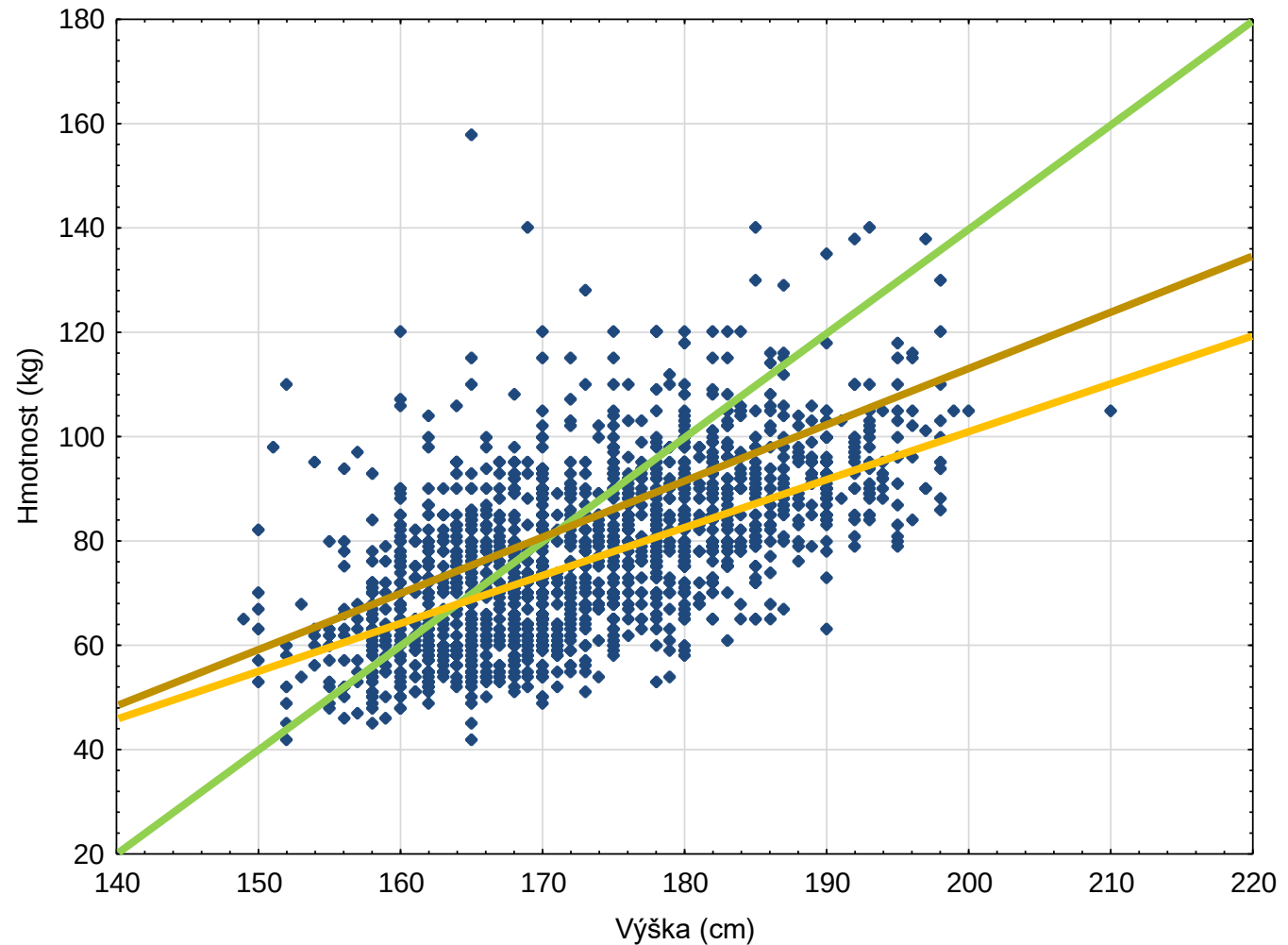
- Závislá proměnná je numerická
- **Nezávislá** proměnná **X** (prediktor) má vliv na **závislou** proměnnou **Y** (outcome)
- Na rozdíl od korelačního koeficientu („mezi proměnnými je vztah“) hledáme **funkční závislost** mezi proměnnými (kauzalitu)
- Podle typu dat rozeznáváme závislost lineární, logaritmickou, kvadratickou, ...

Lineární regrese

- Data prokládáme přímkou

Která přímka je „správně“?

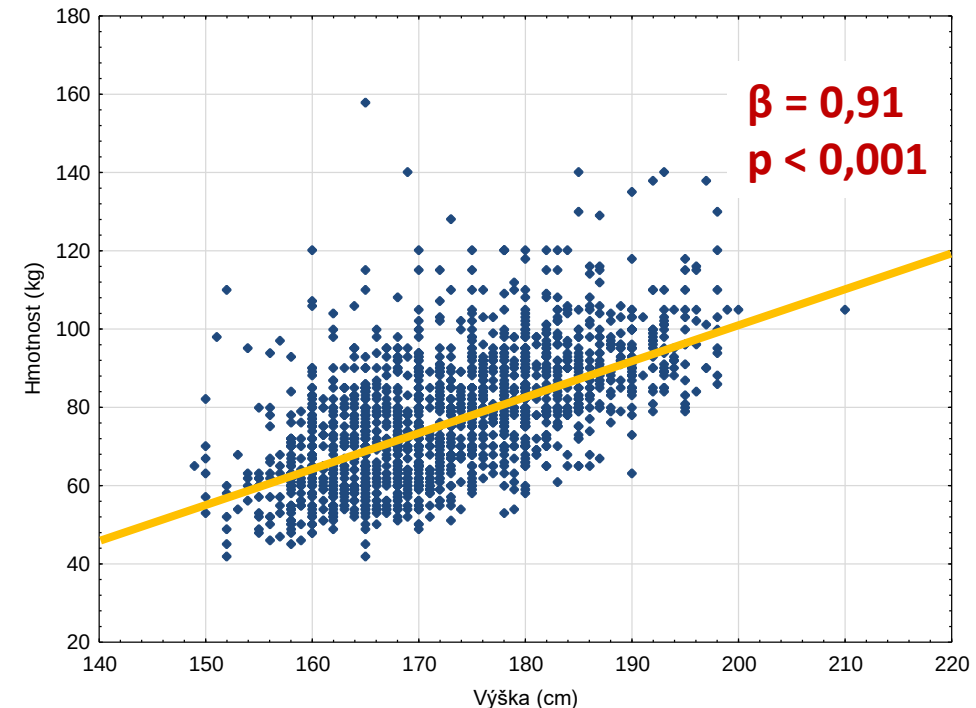
➤ Metoda nejmenších čtverců



Lineární regrese

- Y závisí na X: $y = \beta_0 + \beta_1 x + \epsilon$
- Neboli: $hmotnost = \beta_0 + \beta_1 \cdot výška + náhodná složka$
- H_0 : není lineární vztah mezi proměnnými
 - $\beta_1 = 0$
- H_1 : je lineární vztah mezi proměnnými
 - $\beta_1 \neq 0$

Závěr: Zamítáme hypotézu H_0 a přikláníme se k alternativě, že hmotnost lineárně závisí na výšce

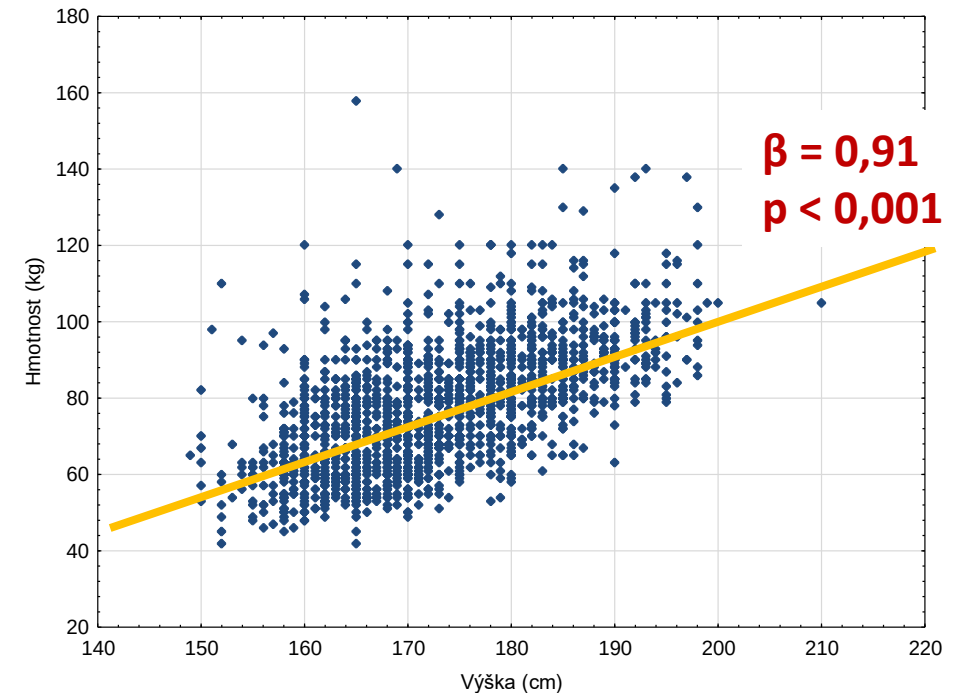


Lineární regrese

- Co nám říká hodnota β ?

➤ $hmotnost = \beta_0 + \beta_1 \cdot výška + náhodná složka$
 $= \beta_0 + 0,91 \cdot výška + "něco"$

- Když se zvýší **výška o 1 cm**, zvýší se **hmotnost o 0,91·1 kg**
- Když se zvýší **výška o 10 cm**, zvýší se **hmotnost o 0,91·10 = 9,1 kg**



Lineární regrese

- Jaké hodnoty používáme k interpretaci lineární regrese?
 - Je β kladné nebo záporné?
 - Je p-hodnota $< 0,05$?
- Ověření předpokladů:
 - Závislá proměnná Y má normální rozdělení pro každou hodnotu X
 - Rozptyl závislé proměnné Y je stejný pro každou hodnotu X
 - Závislost Y na X je lineární

Vícerozměrná lineární regrese

- Y závisí na více prediktorech $X_1, X_2, X_3, \dots$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \epsilon$$

- Např.

$$hmotnost = \beta_0 + \beta_1 \cdot výška + \beta_2 \cdot pohlaví + \beta_3 \cdot věk + \dots$$

- H_0 : Y není lineárně závislá na $X_1, X_2, X_3, \dots$ současně
 - $\beta_1 = 0$ a současně $\beta_2 = 0$ a současně $\beta_3 = 0$ atd.
- H_1 : Y je lineárně závislá alespoň na jednom z prediktorů $X_1, X_2, X_3, \dots$
 - Alespoň jedno $\beta_i \neq 0$

Reportování výsledků

- Model předpovídající míru podvádění na základě spokojenosti v partnerském vztahu vysvětlil 8% hodnot závislé proměnné. S každým jedním bodem nárůstu spokojenosti v partnerském vztahu, došlo ke snížení počtu nevěr o 0.84 ($b = -0.84$, 95% CI $[-1.07, -0.60]$, $p < 0.001$).

Model Fit Measures

Model	R	R ²	Overall Model Test			
			F	df1	df2	p
1	0.280	0.0781	50.8	1	599	< .001

Omnibus ANOVA Test

	Sum of Squares	df	Mean Square	F	p
rating	510	1	510.1	50.8	< .001
Residuals	6019	599	10.0		

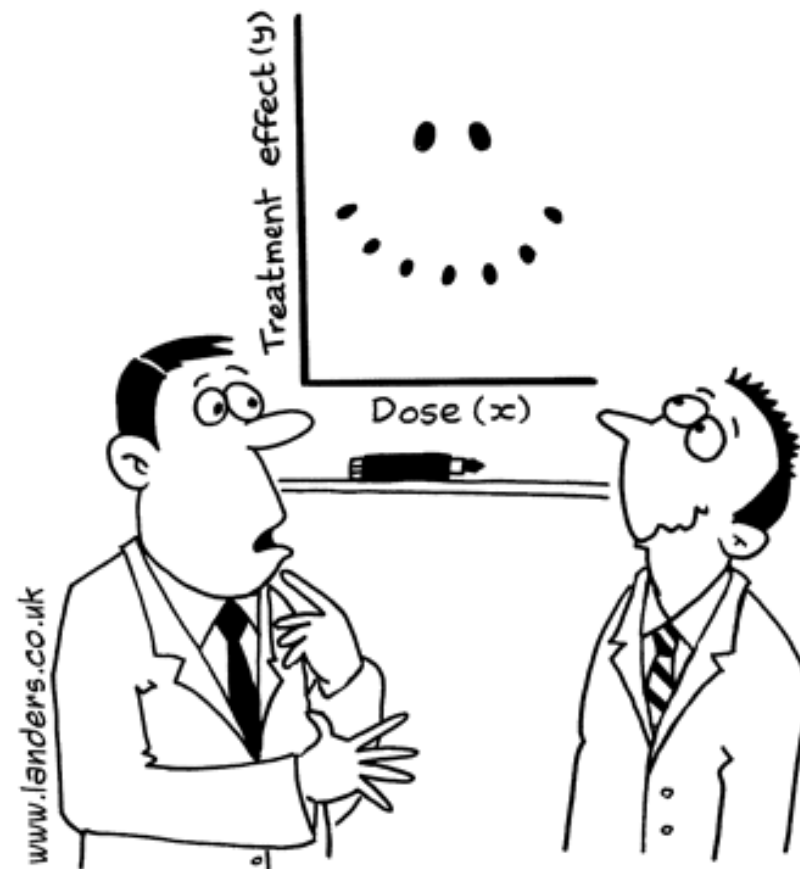
Note. Type 3 sum of squares

[3]

Model Coefficients - affairs

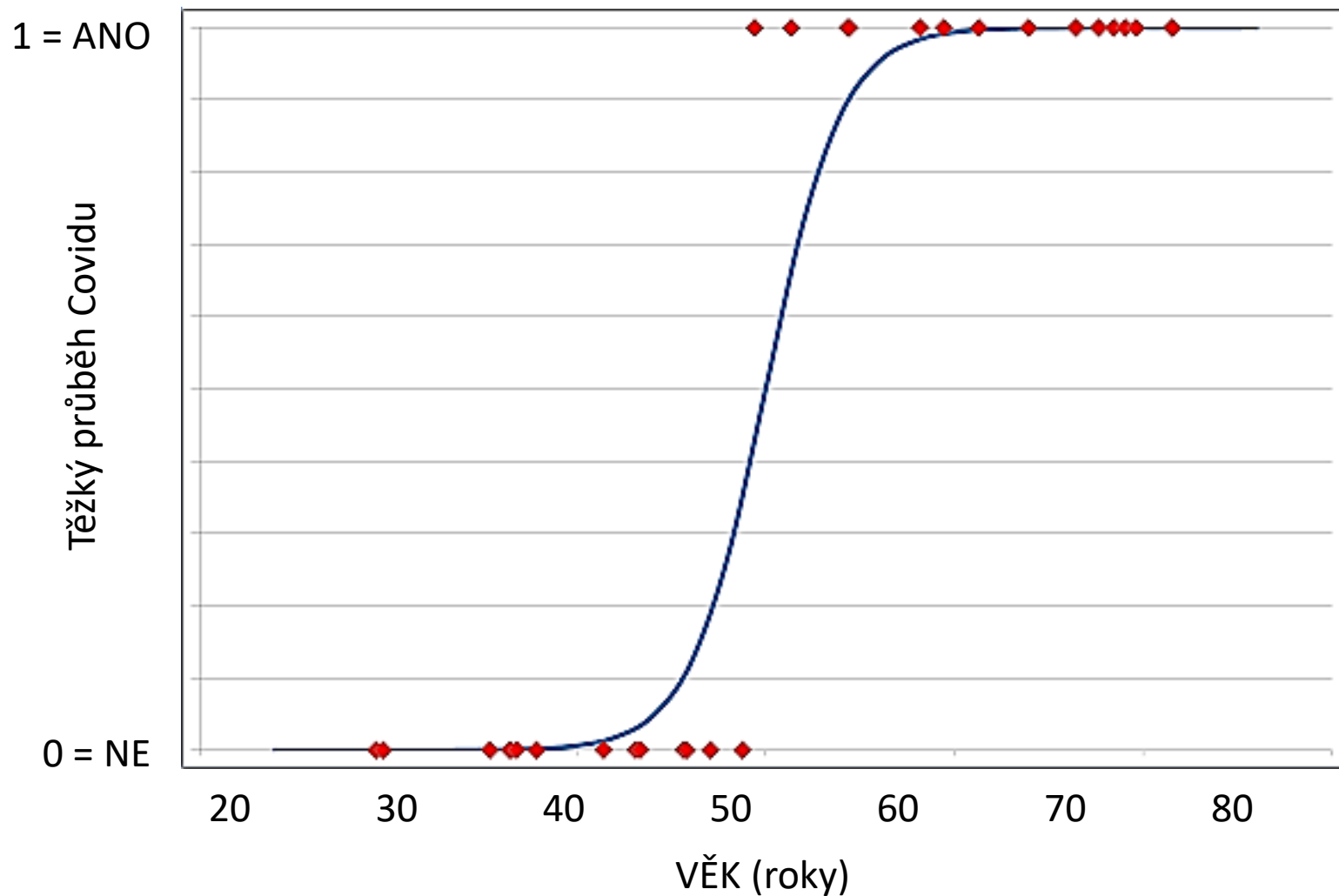
Predictor	Estimate	SE	95% Confidence Interval		t	p
			Lower	Upper		
Intercept	4.742	0.479	3.80	5.683	9.90	< .001
rating	-0.836	0.117	-1.07	-0.605	-7.12	< .001

Logistická regrese – nelineární!



"It's a non-linear pattern with outliers.....but for some reason I'm very happy with the data."

Logistická regrese



- Má věk vliv na těžký průběh Covidu?

Příklad

- H_0 : Věk neovlivňuje pravděpodobnost výskytu chronické bolesti zad ($b = 0$)
- H_1 : Věk ovlivňuje pravděpodobnost výskytu chronické bolesti zad ($b \neq 0$)

- Jak interpretovat výsledek ze softwaru



Model Coefficients - Back_pain							
Predictor	Estimate	SE	Z	p	Odds ratio	95% Confidence Interval	
						Lower	Upper
Intercept	-2.233	0.159	-14.0	< .001	0.107	0.0784	0.146
Věk	0.033	0.003	11.0	< .001	1.034	1.0281	1.041

Note. Estimates represent the log odds of "Back_pain = 1" vs. "Back_pain = 0"

Prediktor

Vlastnosti OR

- OR nabývá hodnot > 0
- Je-li $0 < OR < 1$, tj. β je záporné, je šance („riziko“) outcome nižší
- Je-li $OR > 1$, tj. β je kladné, je šance („riziko“) outcome vyšší

Model Coefficients - Back_pain							
Predictor	Estimate	SE	Z	p	Odds ratio	95% Confidence Interval	
						Lower	Upper
Intercept	-0.525	0.0681	-7.71	< .001	0.591	0.517	0.676
Pohlavi_m:							
1 – 0	-0.190	0.0990	-1.92	0.055	0.827	0.681	1.004
Note. Estimates represent the log odds of "Back_pain = 1" vs. "Back_pain = 0"							

Závěr: Hypotézu H_0 nelze zamítnout ($p=0,055$), vliv pohlaví na výskyt bolestí zad nebyl prokázán

Příklad

Model Coefficients - Back_pain							
Predictor	Estimate	SE	Z	p	Odds ratio	95% Confidence Interval	
						Lower	Upper
Intercept	-2.233	0.159	-14.0	< .001	0.107	0.0784	0.146
Vek	0.033	0.003	11.0	< .001	1.034	1.0281	1.041
Note. Estimates represent the log odds of "Back_pain = 1" vs. "Back_pain = 0"							

**Závěr: Zamítáme hypotézu H_0 ($p < 0,001$), věk má vliv na výskyt bolestí zad.
Konkrétně, s každým rokem věku se šance na výskyt bolestí zad zvýší 1,034 krát (OR=1,034).**

Proč nepoužít t-test?

- Mohli bychom porovnat placebo vs nízká dávka, placebo vs vysoká dávka a vysoká vs nízká dávka
- nízká vs vysoká dávka: $t(8) = -1.96$, $p = .085$, $d = -1.24$
- Placebo vs nízká dávka: $t(8) = -1.21$, $p = .260$, $d = -0.77$
- Placebo vs vysoká dávka: $t(8) = -3.06$, $p = .016$, $d = -1.93$