

POPISNÉ STATISTIKY

MARTINA SEBALO VŇUKOVÁ

MARTINA.VNUKOVA@LF1.CUNI.CZ



AGENDA

- QUIZ & RECAP
- POPISNÉ STATISTIKY
- ČISTĚNÍ DAT
- OUTLIERS
- CENTRÁLNÍ TENDENCE
- MÍRY VARIABILITY
- GRAFY

S UŽITEČNOSTÍ KAMPANĚ PROTI OBTĚŽOVÁNÍ V MHD SOUHLASÍ



OSLOVENÝCH

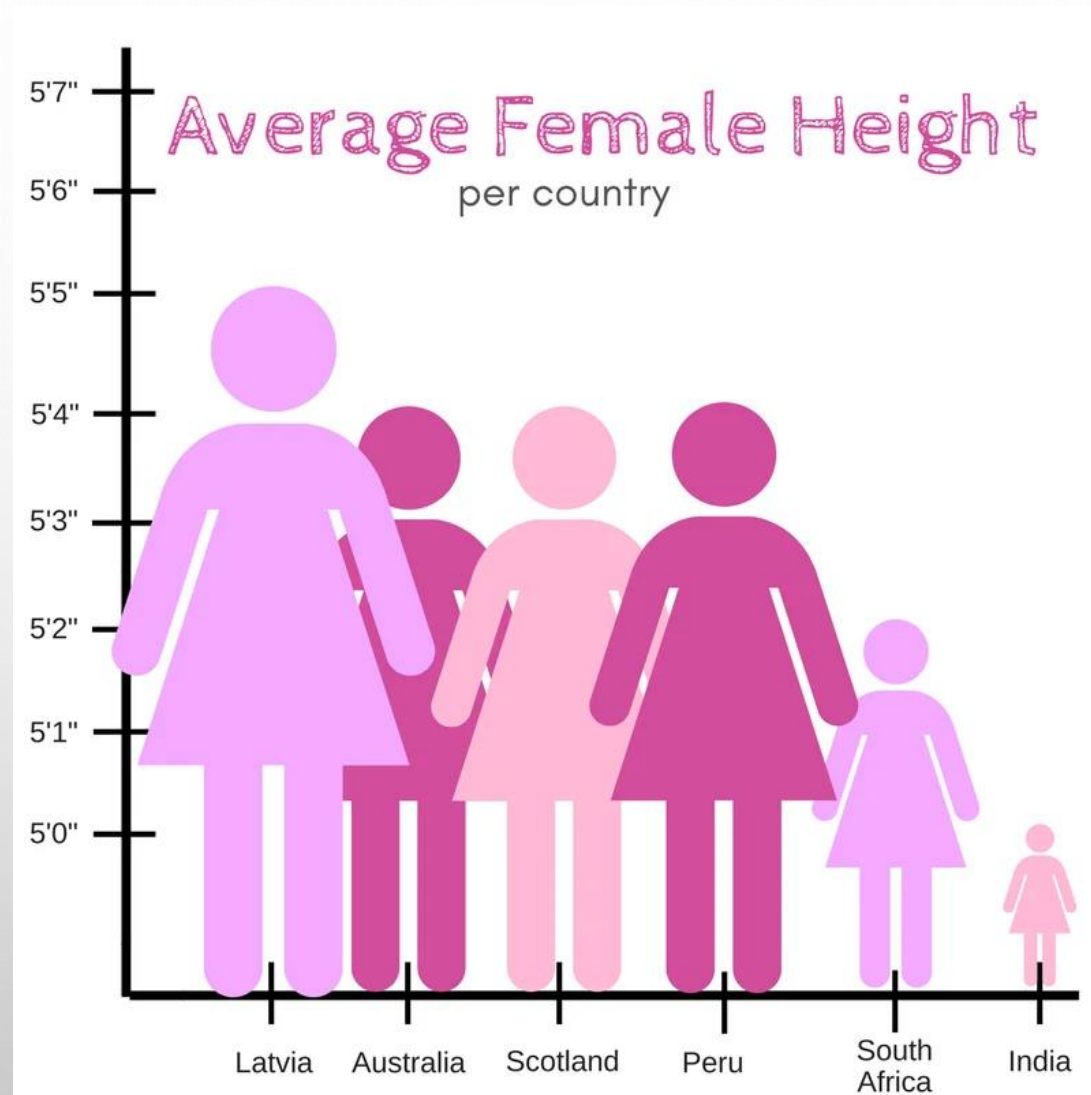


ŽEN



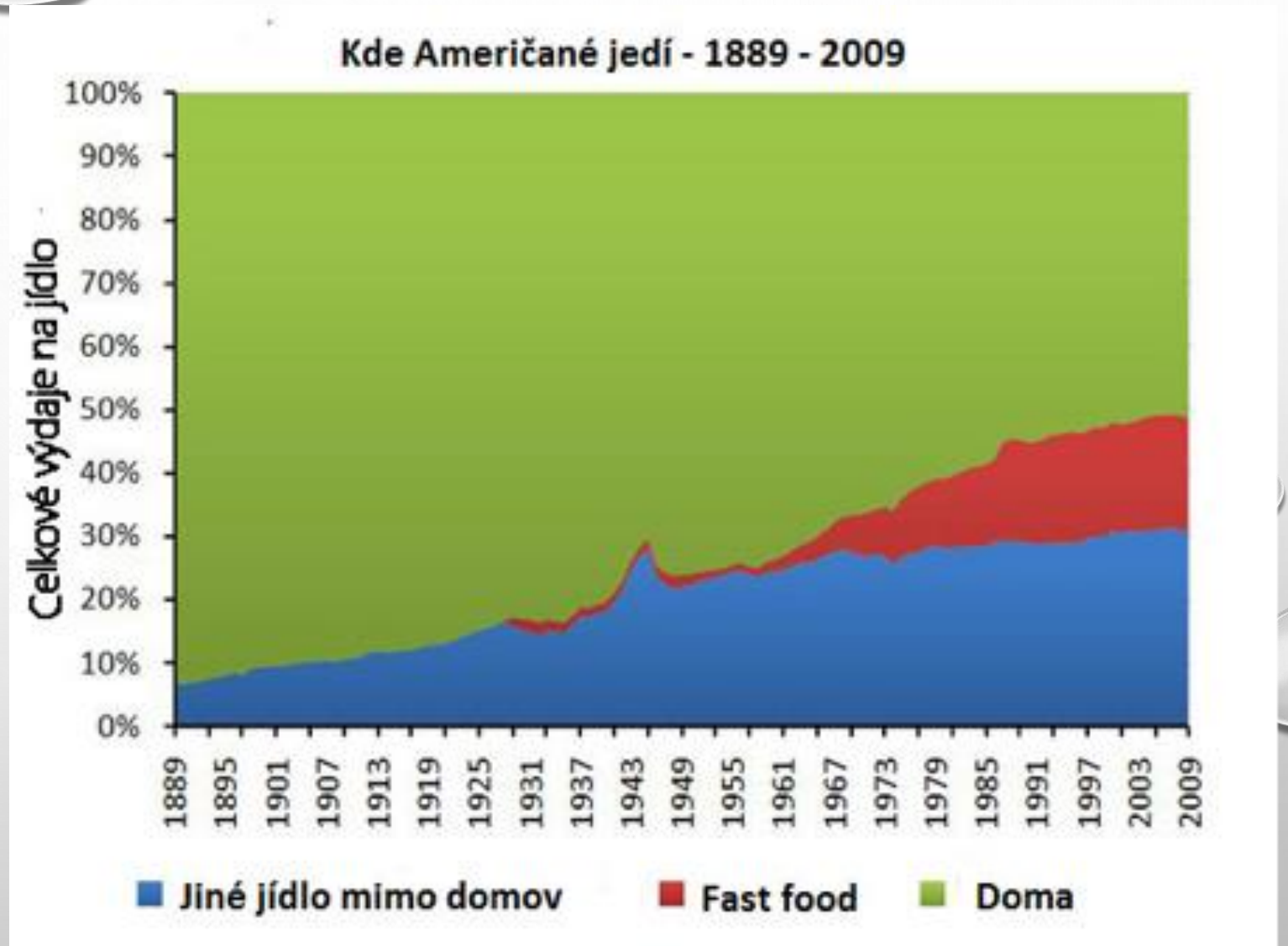
GENERACE Z





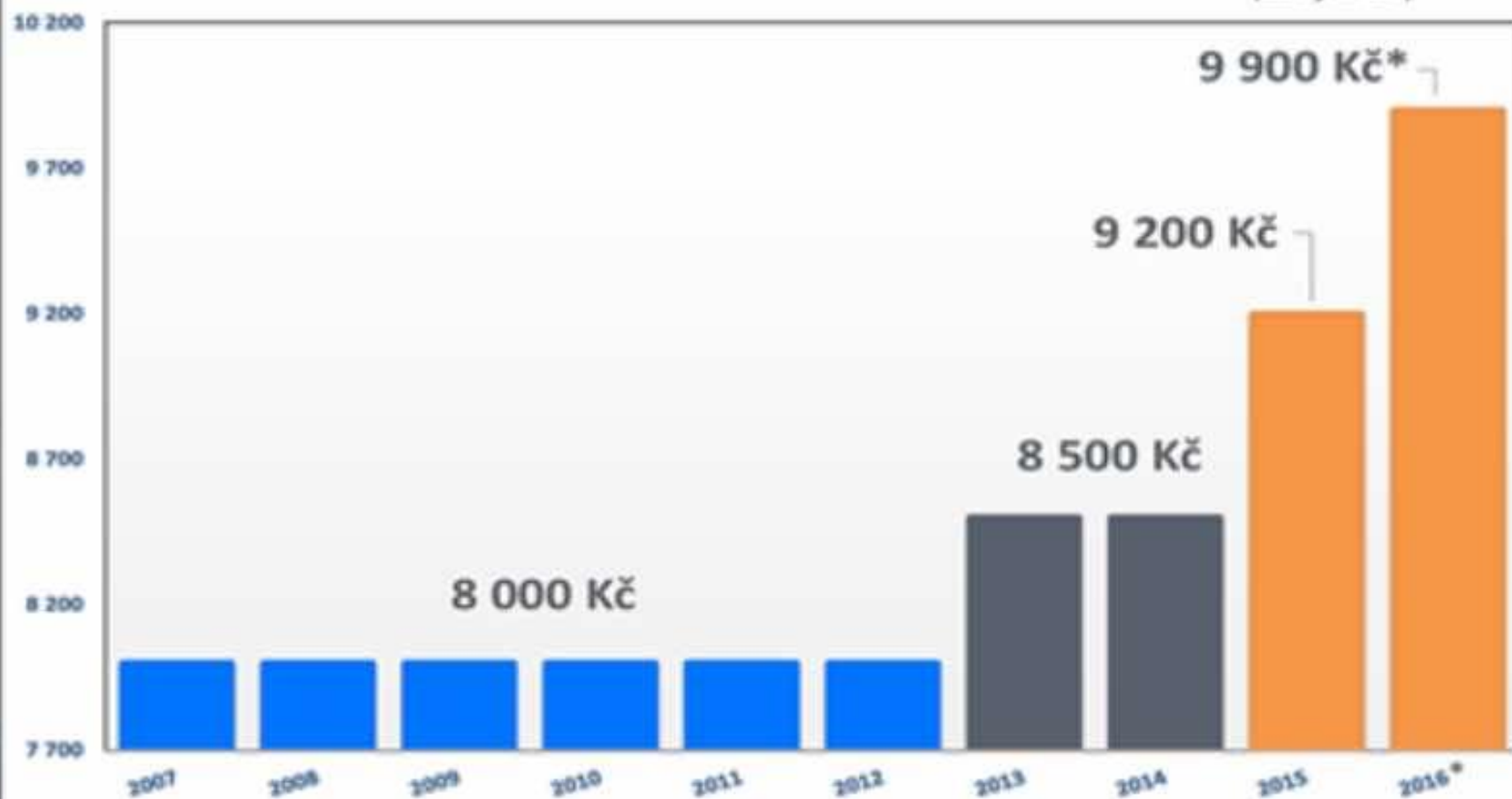
- 1) KATEGORIE: ROZDĚLENÍ NA "JINÉ JÍDLO MIMO DOMOV", "FAST FOOD", A "DOMA" JE PŘÍLIŠ ZJEDNODUŠUJÍCÍ. NAPŘÍKLAD, CO PŘESNĚ Zahrnuje KATEGORIE "JINÉ JÍDLO MIMO DOMOV"? JAKÉ TYPY RESTAURACÍ NEBO JÍDELNÍCH ZAŘÍZENÍ JSOU ZahrnutY?
- 2) NEROZUMÍM OSE Y: CO PŘESNĚ ZNAČÍ % PODÍL?
- 3) BAREVNÉ ROZVRŽENÍ GRAFU MI PŘIJDE MATOUČÍ A NEPŘEHLEDNÉ. VHODNĚJŠÍ BY BYL HISTOGRAM.
- 4) NEVÍME, CO SE DĚLO V LETECH, KTERÁ V GRAFU NEJSOU UVEDENA, NAPŘ. ROKU 1890, 1903, 1995...

<https://www.margit.cz/grafy-lide-tloustnou/>



Vývoj minimální mzdy v ČR v Kč od roku 2007

(zdroj MPSV)



Úřad vlády ČR @strakovka · Aug 20

@SlavekSobotka



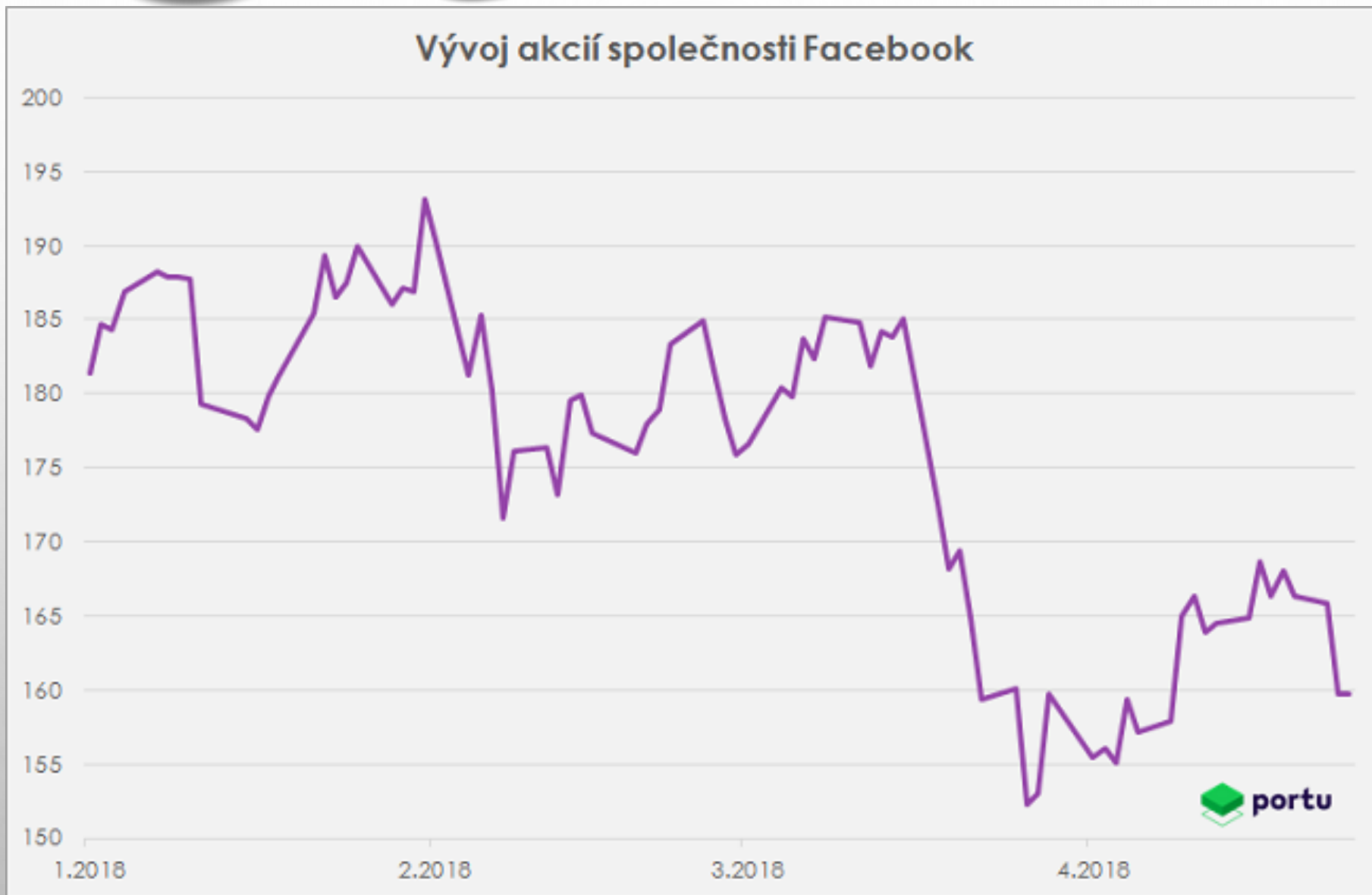
5



8



Vývoj akcií společnosti Facebook – vyřiznutý úsek za rok 2018 ([zdroj](#))

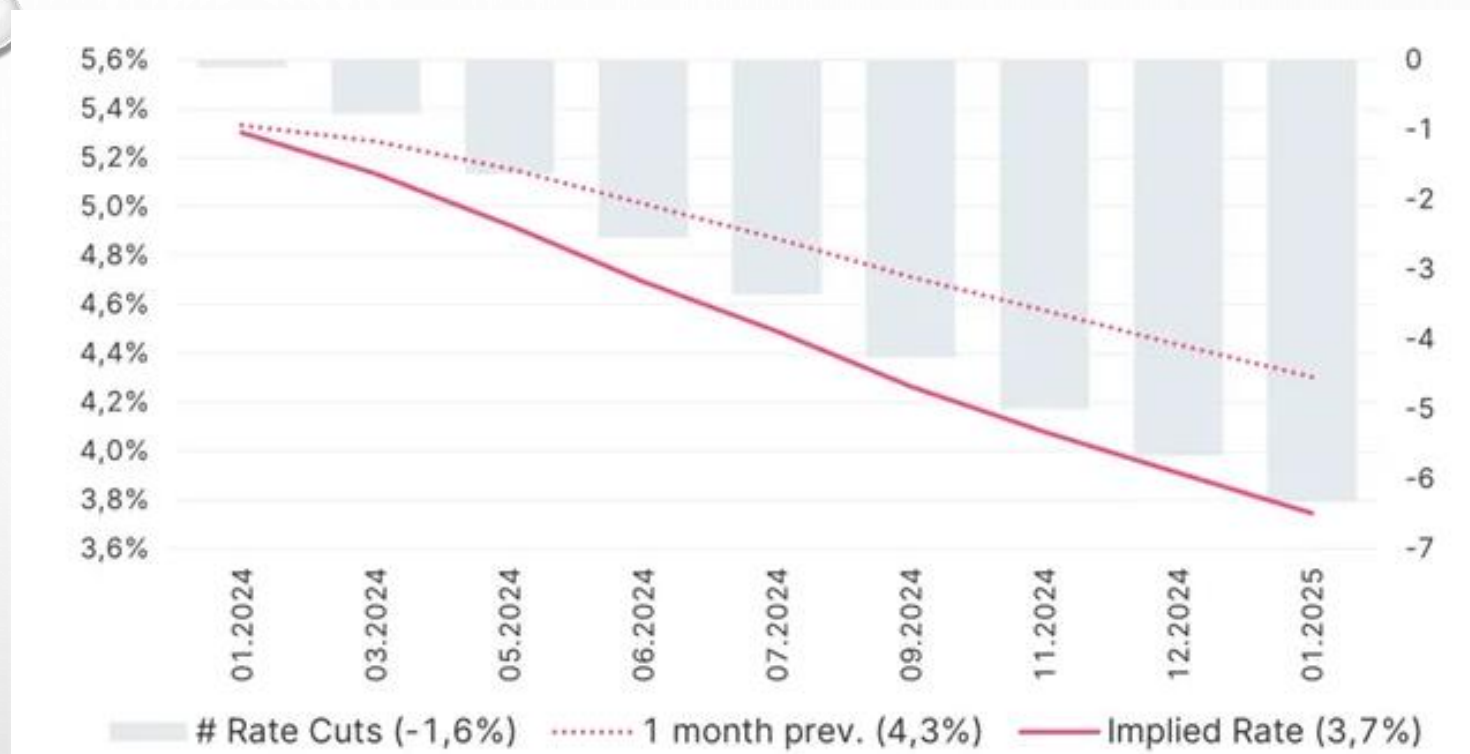


Název článku: To nejlepší z grafů. Co přichází, co odchází a co zůstává?

[Odkaz na článek](#)

Článek pojednává o očekávání vývoje úrokových sazeb v USA.

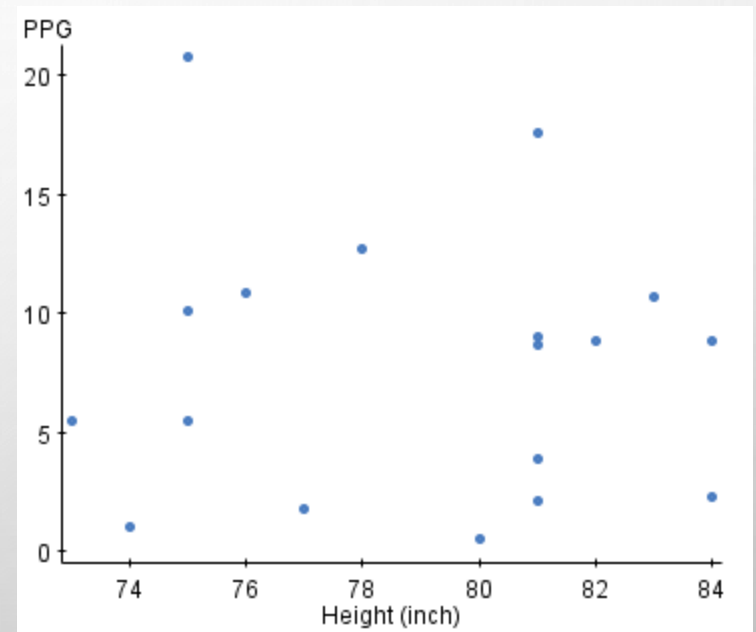
- **Dvě datové řady na ose y**
 - Procentuální hodnoty na jedné datové řadě osy y jsou podseknuté.
 - Z grafu ani článku není pochopitelné čemu odpovídá druhá datová řada vyobrazená v grafu.
- **Data na ose x**
 - Měsíce vypsány na ose x nemají stejné časové rozložení, ale stejné rozestupy.
- **Zavádějící čísla v legendě**
 - Za popisky znázorňují sloupce a osy, jsou procentuální hodnoty, u kterých není pochopitelné, proč jsou zde vypsány.
- **Nevhodně zvolené barvy**
 - Příliš světlé zobrazení sloupců označených v grafu jako Rate Cuts.



Zdroj: Bloomberg; stav k 20.12.2023

RESTRICTION OF RANGE

- KORELACE MEZI VÝŠKOU BASKETBALISTŮ A JEJICH SCHOPNOSTÍ SKÓROVAT (POINTS PER GAME) JE NULOVÁ ($R = -0,05$).
- **VYSVĚTLENÍ:** VZORKEM JSOU PROFESIONÁLNÍ BASKETBALISTI.
 - ROZSAH VÝŠKY JE 185-213 CM
 - PRŮMĚR AMERICKÝCH MUŽŮ = 175 CM
 - MENŠÍ HRÁČI SE DOSTALI DO NBA DÍKY SCHOPNOSTI KOMPENZOvat SVOU VÝŠKU
- **PODOBNÝ PRINCIP:** GRIGGS V. DUKE (1971) USTANOVIL PRECEDENS PROTI POUŽÍVÁNÍ „MATURITY“ A INTELIGENČNÍCH TESTŮ PŘI NAJÍMÁNÍ V USA, PROTOŽE NEKORELOVALY S VÝKONEM PRACOVNÍKŮ ELEKTRÁRNY



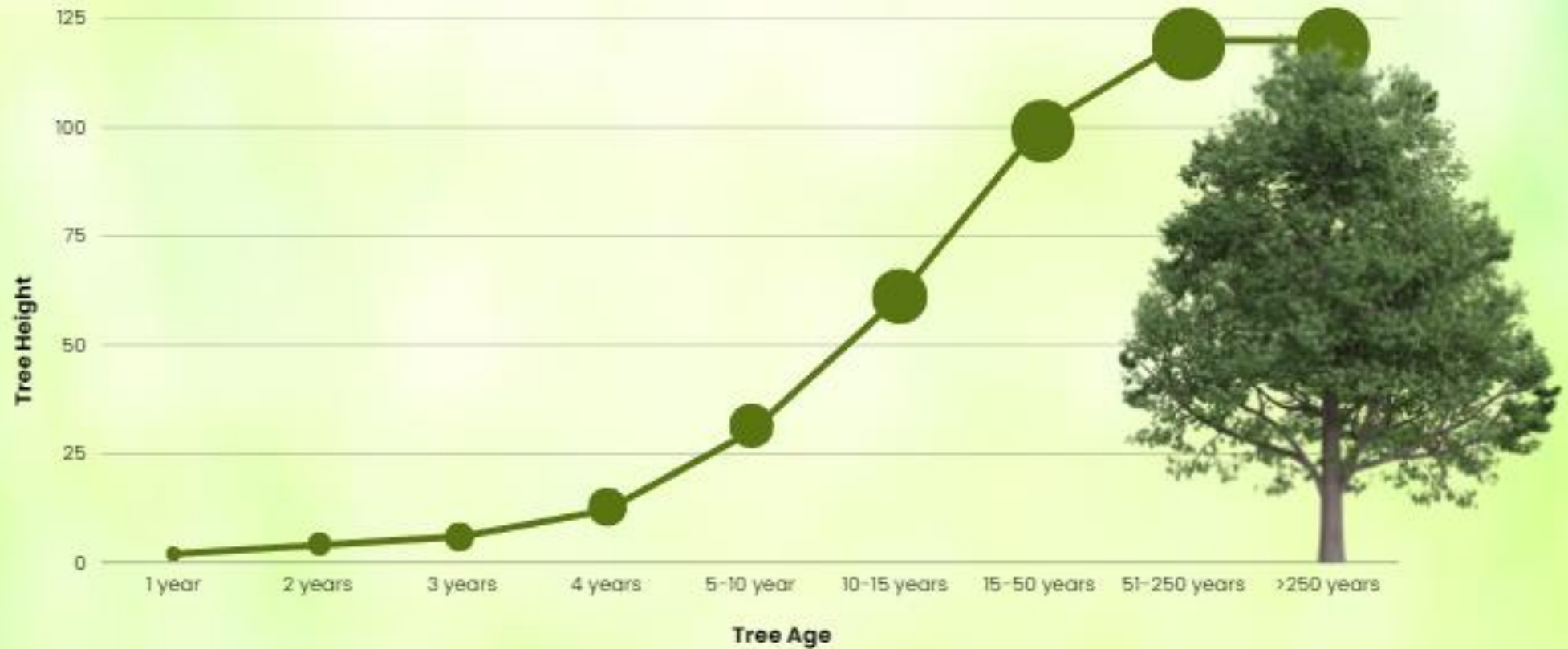
Colgate



Dentist 1	Dentist 2	Dentist 3	Dentist 4	Dentist 5
Colgate	A	D	A	B
A	B	Colgate	F	Colgate
B	Colgate	A	C	D

It is clear that 4/5 dentist recommend Colgate. But 4/5 also recommend Brand A toothpaste.

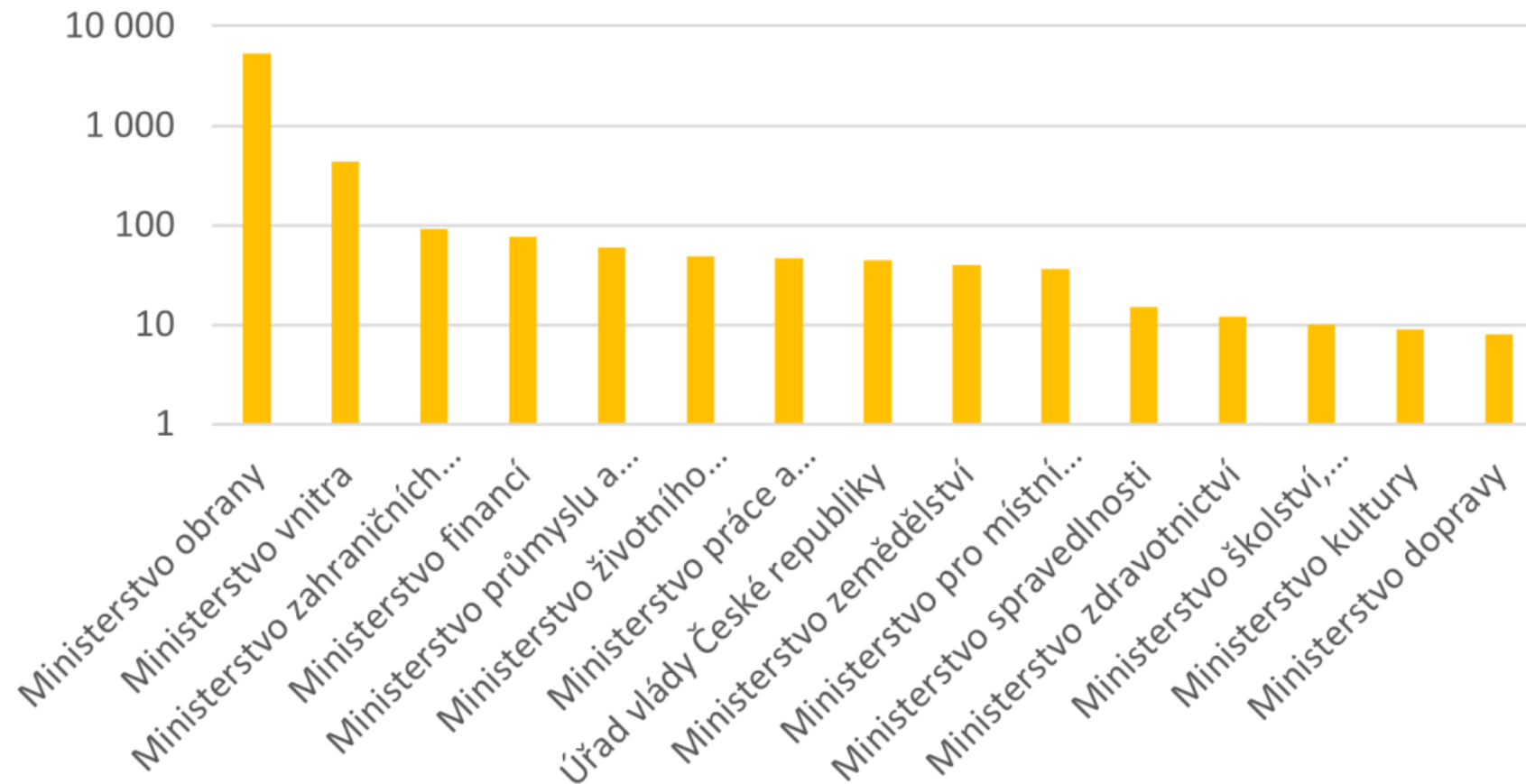
American Sycamore Tree Growth Chart



Average Growth Rate: 2-6 feet per year



Vypsané veřejné zakázky 2022



Polovina Čechů by nezvládla poskytnout první pomoc, ukázal průzkum



13. 9. 2024 17:56

Ad

- V ČLÁNKU SE DOZVÍME, ŽE NADPISEM JE MYŠLENO, ŽE 46 % OBYVATEL ČR NIKDY **NEABSOLVOVALO KURZ PRVNÍ POMOCI – ZNAMENÁ TO, ŽE BY NEZVLÁDLI POMOC? ANI S OPERÁTOREM 155 NA TELEFONU? OPRAVDU BY TO ZVLÁDLI PO KURZU?**
- JAK BYLO DEFINOVÁNO „NEZVLÁDLI POMOC“, NEABSOLVOVÁNÍM KURZU?
- 46 % **NENÍ CELÁ POLOVINA**
- JEDNALO SE O **REPREZENTATIVNÍ VZOREK ČECHŮ**, KDYŽ JE ZOBECNĚNO V NADPISE?

Deprese

> Co je to deprese?

Depresivní porucha (nemoc) je psychické onemocnění, které charakterizuje změna nálady, nevykonnost, snížený pocit energie, ztráta zájmů a aktivit, zhoršená soustředěnost a z toho vyplývající zhoršení paměti; uzavírání se před okolím, přemýšlení o smrti a někdy pokusy o sebevraždu.

> Jak často se vyskytuje deprese?

Deprese se vyskytují v relativně vysoké míře. Celoživotní riziko vzniku depresivní poruchy je u žen v 10–25 %, u mužů v 5–12 %. Nejvíce mužů i žen je postiženo mezi 25–50 lety.

Z uvedených statistik vyplývá, že skoro každá čtvrtá žena a téměř každý osmý muž během svého života onemocní depresivní poruchou.

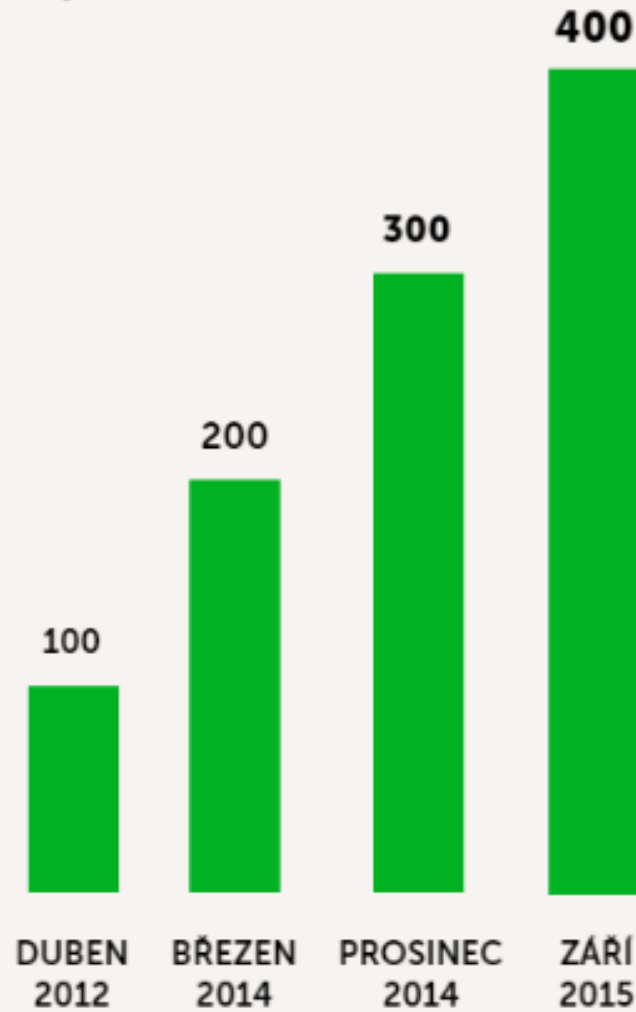
V UVEDENÉ UKÁZCE SE
VYCHÁZÍ Z HORNÍCH
HRANIC UVEDENÝCH
STATISTIK

->

INTERPRETACE
SCHVÁLNĚ ZVĚTŠUJE
DŮRAZ NA
PROBLEMATIKU

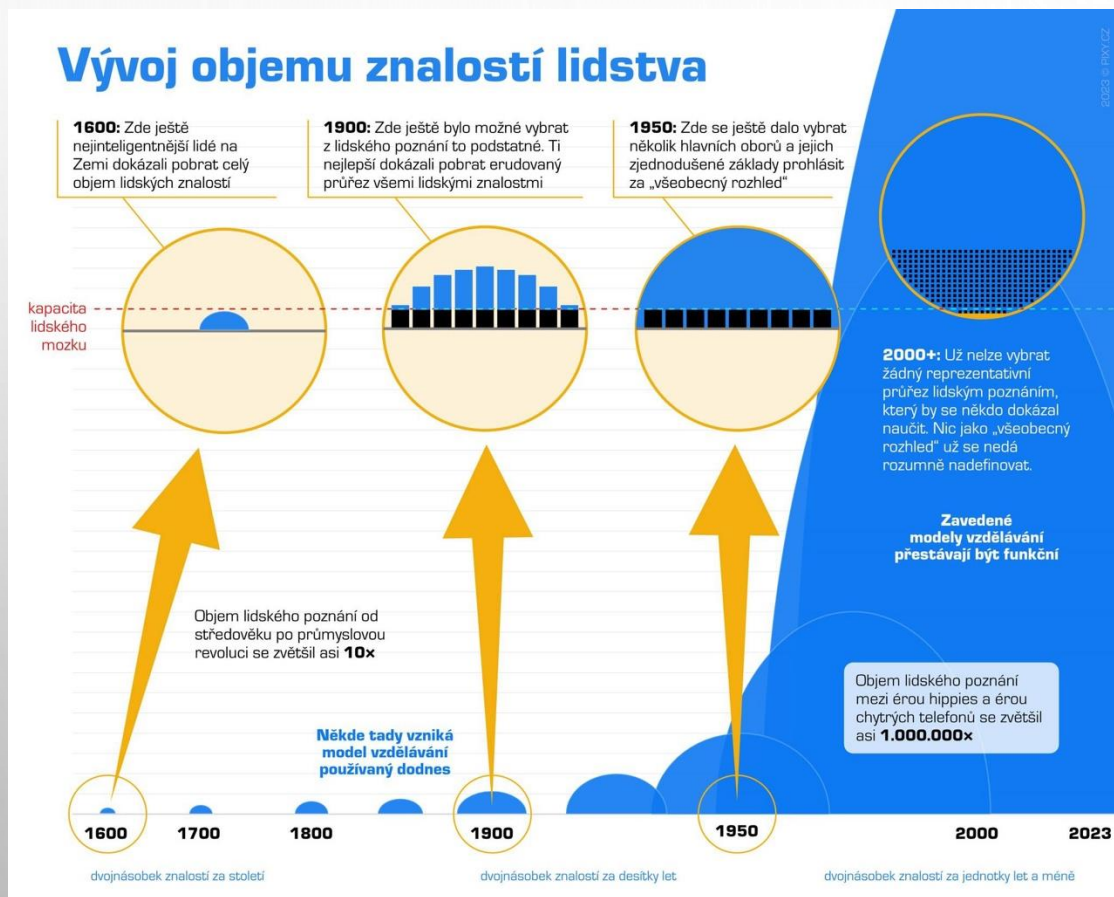
JAK ROSTL INSTAGRAM

milióny aktivních uživatelů



Zdroj: Týdeník Dotyk/Facebook

UKÁZKA STATISTICKÉ MISINTERPRETACE



- DATA POCHÁZÍ Z AUTOROVI SAMOSTATNÉ ANALÝZY INTERNETOVÝCH DISKUZÍ, KTERÝCH SE SÁM ÚČASTNIL. AVŠAK JAKÝM ZPŮSOBEM VYBÍRAL DISKUZE NENÍ ZNÁM. NEDOHLEDALA JSEM ANI ODKAZY NA ZMÍNĚNÉ DISKUZE.
- DOMNÍVÁM SE, ŽE GRAF JE DĚLANÝ JAKO PODPORA PRO KONKRÉTNÍ TVRZENÍ. UŽ JEN Z POHLEDU JAZYKA INTERPRETACE VÝSLEDKŮ.
- VZDÁLENOST PROMĚNNÝCH (ROKY) NA OSE X JE NEROVNOUŠERNÉ.
- DATA NA OSE Y NEJSOU ADEKVÁTNĚ OZNAČENÁ – CHYBÍ KONKRÉTNÍ DATASET, EXPONENCIÁLNÍ RŮST JE VELMI ZVÝRAZNĚNÝ.
- PROMĚNNÉ „KAPACITA LIDSKÉHO MOZKU“ NEROZUMÍM. PROČ JE V GRAFU TAKTO VIZUALIZOVÁNA A PROČ JI AUTOR DO GRAFU VŮBEC UMISŤUJE?
- V GRAFU SE OBJEVUJE FORMULOVANÝ ZÁVĚR AUTORA.



CO TO JE, KDYŽ SE ŘEKNE PRŮMĚRNÁ MZDA. (N.D.). FINANCE.CZ.

[HTTPS://WWW.FINANCE.CZ/ZPRAVY/FINANCE/42752-CO-TO-JE-KDYZ-SE-REKNE-PRUMERNA-MZDA/](https://www.finance.cz/zpravy/finance/42752-co-to-je-kdyz-se-rekne-prumer-na-mzda/)

EXKURZE DO ČESKÝCH BOTNÍKŮ

Zdroj:

Ciborová, K. (2012). *Jak se obouváme? Devět párů ve skříni*. Deník.cz.

https://www.denik.cz/z_domova/jak-se-obouvame-devet-paru-ve-skrini-20120713.html



EXKURZE DO ČESKÝCH BOTNÍKŮ

- graf - počet respondentů?
- zahrnuje graf muže i ženy z vesnic a měst?
- muži x ženy





35 obuvnických firem

s více než 20 zaměstnanci působí v ČR

až **80 %** bot

vyrobených v Česku směřuje do EU

90 % české výroby

se soustřeďuje ve Zlínském kraji

	2002	2006	2009	2011
Výroba obuvi v ČR (páry v mil.) [*]	8	9,2	4,5	4,4
Vývoz (páry v mil.) ^{**}	10,5	30	35,5	51
Dovoz (páry v mil.)	46	102	109	183 ^{***}
Počet zaměstnanců obuvnického průmyslu v ČR	9500	4100	3500	3400

^{*}Před rokem 1989 se u nás vyrábělo více než 70 milionů párů.

^{**}Do vývozu jsou započítány i boty cizí výroby.

^{***}188 milionů párů pocházelo z Číny.

- roky – 2003, 2004, 2005, ... - zkreslení dat
- poznámky pod čarou
- článek se věnuje českému obuvnictví, ale do vývozu byly zahrnuty i boty cizí výroby
- u dovozu nevím přesný počet bot dovezených z Číny v minulých letech, pouze v roce 2011

**Určitý lék snížil riziko určité choroby o
50 %**

Pokud je původní riziko onemocnění velmi malé (např. 0,2 %), snížení o polovinu (na 0,1 %) ve skutečnosti nepředstavuje velký dopad.

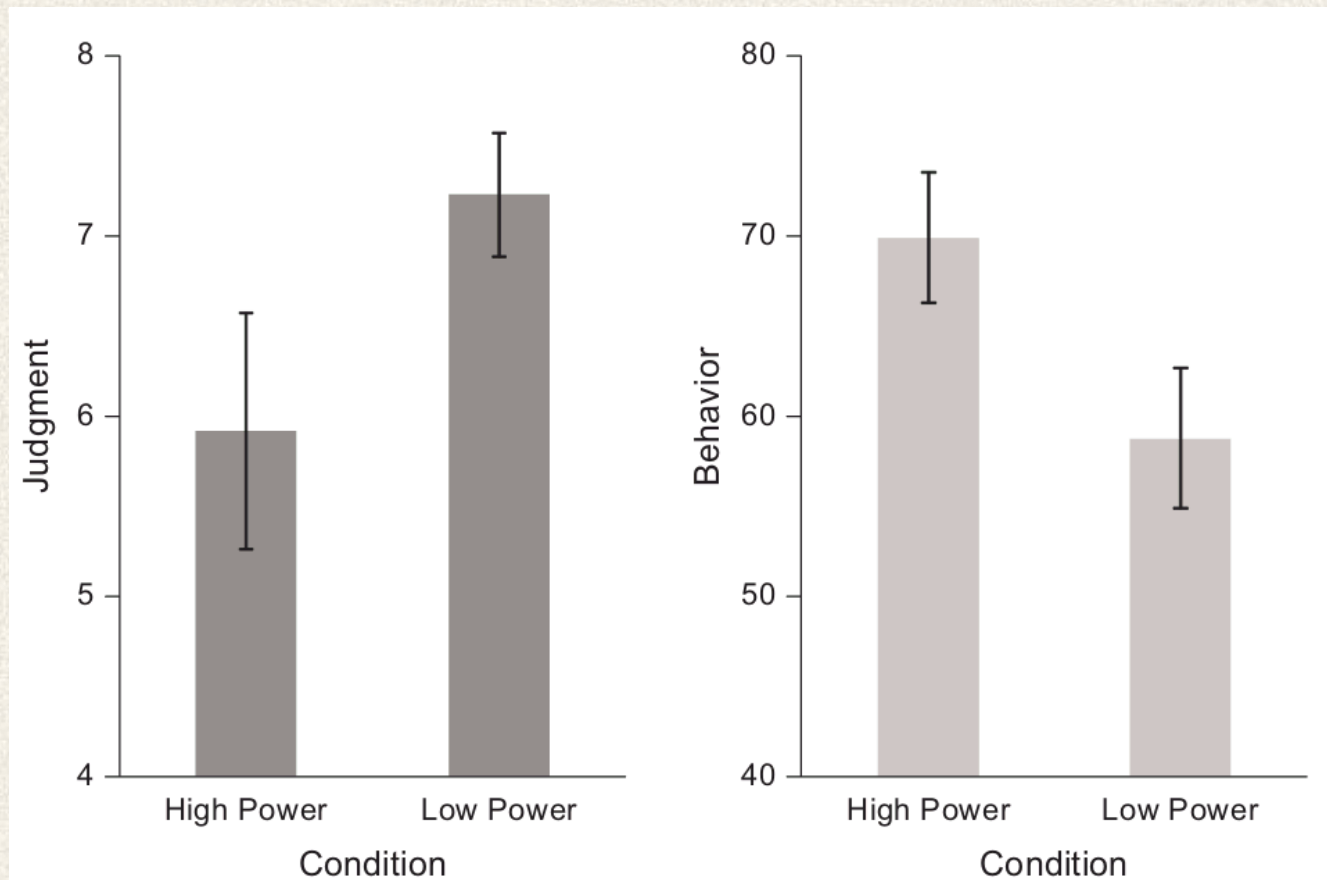
POWER INCREASES HYPOCRISY: MORALIZING IN REASONING, IMMORALITY IN BEHAVIOR

Joris Lammers, Diederik A Stapel, Adam D Galinsky
<https://doi.org/10.1177/0956797610368810>

Výzkum zkoumá vztah mezi mocí a morálním pokrytectvím, tedy jak mocní lidé odsuzují špatné chování u ostatních, ale sami ho praktikují.

Misinterpretace:

- Graf nezačíná od nuly = rozdíl mezi hodnotami se zdá být větší, než ve skutečnosti je
- Chybové úsečky naznačují, že výsledky jsou jednoznačnější a spolehlivější než doopravdy jsou
- Graf se zaměřuje pouze na průměry



ÚVOD DO STATISTIKY

1. KTERÁ Z NÁSLEDUJÍCÍCH JE KVANTITATIVNÍ ÚROVEŇ MĚŘENÍ?

- A) BARVA
- B) HMOTNOST
- C) STÁTNÍ PŘÍSLUŠNOST
- D) KRAJ

2. KTERÁ ÚROVEŇ MĚŘENÍ POUŽÍVÁ JASNÉ POŘADÍ, ALE ROZDÍLY MEZI PROMĚNNÝMI NEJSOU EXAKTNĚ MĚŘITELNÉ?

- A) NOMINÁLNÍ
- B) ŘADOVÁ
- C) INTERVALOVÁ
- D) POMĚROVÁ

3. S JAKÝM TYPEM PROMĚNNÉ SE V EXPERIMENTU MANIPULUJE?

- A) ZÁVISLÝM
- B) NEZÁVISLÝM
- C) ŘADOVÝM
- D) NOMINÁLNÍM

4. POKUD VÝZKUMNÍK ROZDĚLÍ POPULACI DO MENŠÍCH SKUPIN A VYBERE VZORKY Z KAŽDÉ SKUPINY, JAKÝ TYP METODY VÝBĚRU POUŽÍJE?

- A) JEDNODUCHÝ NÁHODNÝ VÝBĚR
- B) STRATIFIKOVANÝ NÁHODNÝ VÝBĚR
- C) POHODLNÉ VZORKOVÁNÍ
- D) VÝBĚR METODOU SNĚHOVÉ KOULE

5. ÚPLNOST JEDNOTLIVCŮ NEBO POZOROVÁNÍ, KTERÁ MÁ STUDIE ZA CÍL POPSAT, SE NAZÝVÁ:

- A) VZOREK
- B) PROMĚNNÁ
- C) POPULACE
- D) SOUBOR DAT

6. PROČ JE V PSYCHOLOGII POCHOPENÍ STATISTIKY KLÍČOVÉ?

- A) ABYCHOM MOHLI PROVÁDĚT EXPERIMENTY
- B) NENÍ
- C) ABYCHOM MOHLI INTERPRETOVAT A VYHODNOCOVAT VÝSLEDKY VÝZKUMU
- D) ABY PSYCHOLOGOVÉ MOHLI PŘEDEPSAT LÉKY

7. KTERÁ NENÍ KVANTITATIVNÍ ÚROVEŇ MĚŘENÍ?

- A) VĚK
- B) PŘÍJEM
- C) POHLAVÍ
- D) TEPLOTA

8. O JAKOU ÚROVEŇ MĚŘENÍ SE JEDNÁ, KDYŽ JSOU POJMENOVÁNY KATEGORIE, ALE NEEXISTUJE ŽÁDNÉ POŘADÍ NEBO POŘADÍ?

- A) POMĚROVOU
- B) NOMINÁLNÍ
- C) ŘADOVOU
- D) INTERVALOVOU

9. KTERÁ ÚROVEŇ MĚŘENÍ MÁ ABSOLUTNÍ NULOVÝ BOD?

- A) ŘADOVÁ
- B) NOMINÁLNÍ
- C) POMĚROVÁ
- D) INTERVALOVÁ

10. KTERÁ Z NÁSLEDUJÍCÍCH MOŽNOSTÍ JE TYPEM NENÁHODNÉHO VÝBĚRU?

- A) JEDNODUCHÝ NÁHODNÝ VÝBĚR
- B) SYSTEMATICKÝ VÝBĚR
- C) CLUSTEROVÝ VÝBĚR
- D) NAMÁTKOVÝ VÝBĚR

POPISNÁ STATISTIKA

Třídít

Popisovat

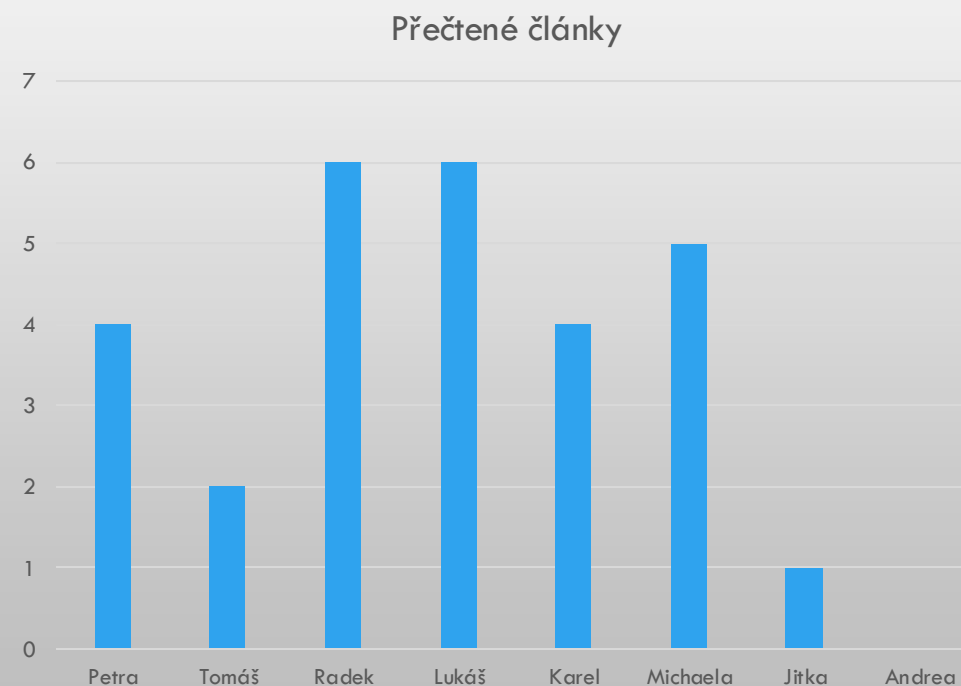
Zobrazovat

Jak přehledně vyjádřit
nepřehledné?

V KNIHOVNĚ

PETRA PŘEČETLA MINULÝ TÝDEN 4 ČLÁNKY , TOMÁŠ JENOM 2 ,
RADEK PŘEČETL 6 ČLÁNKŮ A LUKÁŠ TAKÉ 6, KAREL 4 ČLÁNKY,
MICHAELA 5, JITKA JENOM 1 A ANDREA NEPŘEČETLA NIC

- CO NÁM TENTO TEXT ŘÍKÁ?
- 8 STUDENTŮ PŘEČETLO 28 ČLÁNKŮ
- V PRŮMĚRU PŘEČTOU STUDENTI 3,5 ČLÁNKU
- MAXIMUM PŘEČETNÝCH ČLÁNKŮ BYLO 6, MINIMUM 0



PROČ?

Stručný přehled dat

Zjednodušení

Identifikace vzorců a trendů

Porovnávání

Určení distribuce dat

Identifikace odlehlých hodnot

základ pro inferenční statistiku

Vizualizace dat

Zajištění transparentnosti a replikovatelnosti výzkumu

Kontrola kvality

ČIŠTĚNÍ DAT



ZÁKLAD PRO
ROZHODOVÁNÍ



EFEKTIVITA ZDROJŮ



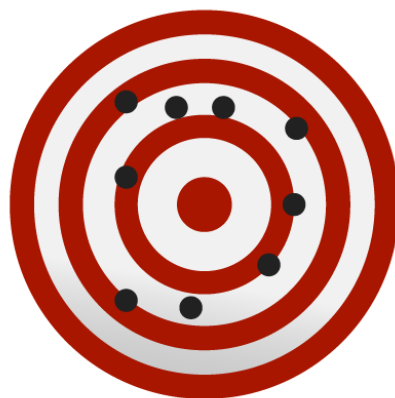
VÝKON MODELU

PŘESNOST , PLATNOST A SPOLEHLIVOST

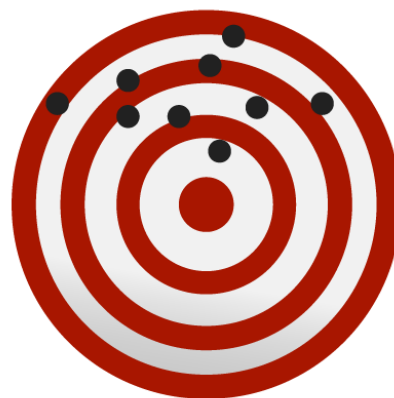
Reliability and Validity



Reliable
Not valid



Low validity
Low reliability



Not reliable
Not valid



Both reliable
and valid

BĚŽNÉ PROBLÉMY S DATY

- CHYBĚJÍCÍ HODNOTY

- V DLOUHODOBÉ STUDII O ÚČINCÍCH TRAUMATU Z DĚTSTVÍ NA DUŠEVNÍ ZDRAVÍ DOSPĚLÝCH SE OD ÚČASTNÍKŮ OČEKÁVÁ, ŽE BUDOU KAŽDÝ MĚSÍC VYPLŇOVAT SEBEHODNOTÍCÍ DOTAZNÍKY. POKUD ÚČASTNÍK ZAPOMENE NEBO SE ROZHODNE NEVYPLNIT DOTAZNÍK PRO KONKRÉTNÍ MĚSÍC, BUDOU PRO TENTO ČASOVÝ BOD CHYBĚT HODNOTY.

- DUPLIKÁTY

- V DATABÁZI ÚČASTNÍKŮ STUDIE O KOGNITIVNÍM VÝVOJI JE TENTÝŽ ÚČASTNÍK, „JAN NOVÁK“, UVEDEN DVAKRÁT SE STEJNÝMI ODPOVĚĐMI NA STEJNOU SADU OTÁZEK KVŮLI CHYBĚ PŘI ZADÁVÁNÍ DAT.

- NEKONZISTENCE

- DATUM NAROZENÍ JANA NOVÁKA JE V JEDNOM SYSTÉMU UVEDENO JAKO „05/10/1990“ A V JINÉM „5. ŘÍJNA 1988“

- ODLEHLÉ HODNOTY (OUTLIERS)

- V EXPERIMENTU, KDE JSOU ÚČASTNÍCI TESTOVÁNI NA POČET SLOV, KTERÁ SI PAMATUJÍ ZE SEZNAMU 20, SI VĚTŠINA ÚČASTNÍKŮ PAMATUJE 5-15 SLOV. JEDEN ÚČASTNÍK SI VŠAK PAMATUJE POUZE 1 SLOVO

- CHYBY

- BĚHEM SLOVNÍHO HODNOCENÍ MŮŽE ÚČASTNÍK ZMÍNIT, ŽE PRODĚLAL „DYSTHYMIÍ“, ALE PŘEPISOVATEL OMYLEM ZAPÍŠE „DYSFORII“.

- IRELEVANTNÍ DATA

- VE STUDII ANALYZUJÍCÍ PSYCHOLOGICKÉ ÚČINKY SPÁNKOVÉ DEPRIVACE BY ÚDAJE O OBLÍBENÝCH BARVÁCH ÚČASTNÍKŮ BYLY IRELEVANTNÍ. PODOBNĚ PŘI ANALÝZE ÚDAJŮ O ÚČINNOSTI KONKRÉTNÍ TERAPEUTICKÉ METODY PRO DEPRESI BY ZNALOST VELIKOSTI BOT ÚČASTNÍKŮ NEBYLA RELEVANTNÍ.

JAK IDENTIFIKOVAT PROBLÉMY



Vizuální kontrola

Po shromáždění dat z průzkumu o úrovni stresu u vysokoškolských studentů si výzkumník mohl projít odpovědi. Mohou si všimnout řady položek „999“ nebo „N/A“, které indikují potenciální chybějící nebo nesprávně zaznamenaná data.



Softwarové nástroje

Pomocí softwaru jako SPSS nebo R může psycholog nahrát data z inventáře osobnosti. Software lze naprogramovat tak, aby zvýraznil nebo označil potenciální problémy, jako jsou duplicitní položky nebo hodnoty mimo očekávané rozsahy.



Statistické testy

Ve studii měření reakčních časů by mohl být Kolmogorov-Smirnovův test použit k určení, zda se data odchylojí od normální distribuce, což naznačuje potenciální anomálie.



Popisná statistika

Ve studii o skóre štěstí, které si sami uvedli (na stupnici 1–10), by střední skóre „50“ okamžitě signalizovalo chybu, protože je mimo očekávaný rozsah. Podobně extrémně vysoká směrodatná odchylka může naznačovat velkou variabilitu, což naznačuje odlehlé hodnoty nebo široký rozsah odpovědí.



Vizualizace dat

Ve studii srovnávající skóre úzkosti před intervencí a po intervenci lze použít bodový graf. Pokud jsou skóre po intervenci konzistentně a významně vyšší než skóre před intervencí (mimo očekávaný terapeutický účinek), může to znamenat problém – buď s intervencí, nebo s procesem záznamu dat.

ŘEŠENÍ PROBLÉMŮ



Imputace - doplnění / nahrazení

V longitudinální studii sledující depresivní symptomy v průběhu času, pokud účastník vynechá hodnocení jednoho měsíce, chybějící hodnota může být přičtena pomocí průměru jejich skóre za předchozí a následující měsíc. Pokročilejší metody mohou používat regresní modely k předpovědi chybějící hodnoty na základě jiných proměnných.



Vymazání

V průzkumu o úrovních stresu, pokud respondent ponechá více než polovinu otázek nezodpovězenou, může být celá jeho odpověď vymazána, aby byla zajištěna kvalita datového souboru.



Oprava

Pokud si psycholog během vizuální kontroly dat všimne skóre „15“ na stupnici 1–10, může tuto chybu opravit tím, že se vrátí k původnímu průzkumu nebo jej nahradí věrohodnější hodnotou.



Standardizace

Pokud sbíráte data mezinárodně, úrovně stresu mohou být zaznamenávány na různých škálách (např. 1-5 v jedné zemi, 1-10 v jiné). Standardizace těchto dat na společné měřítko (např. 1-10) zajišťuje konzistenci analýzy.



ověření

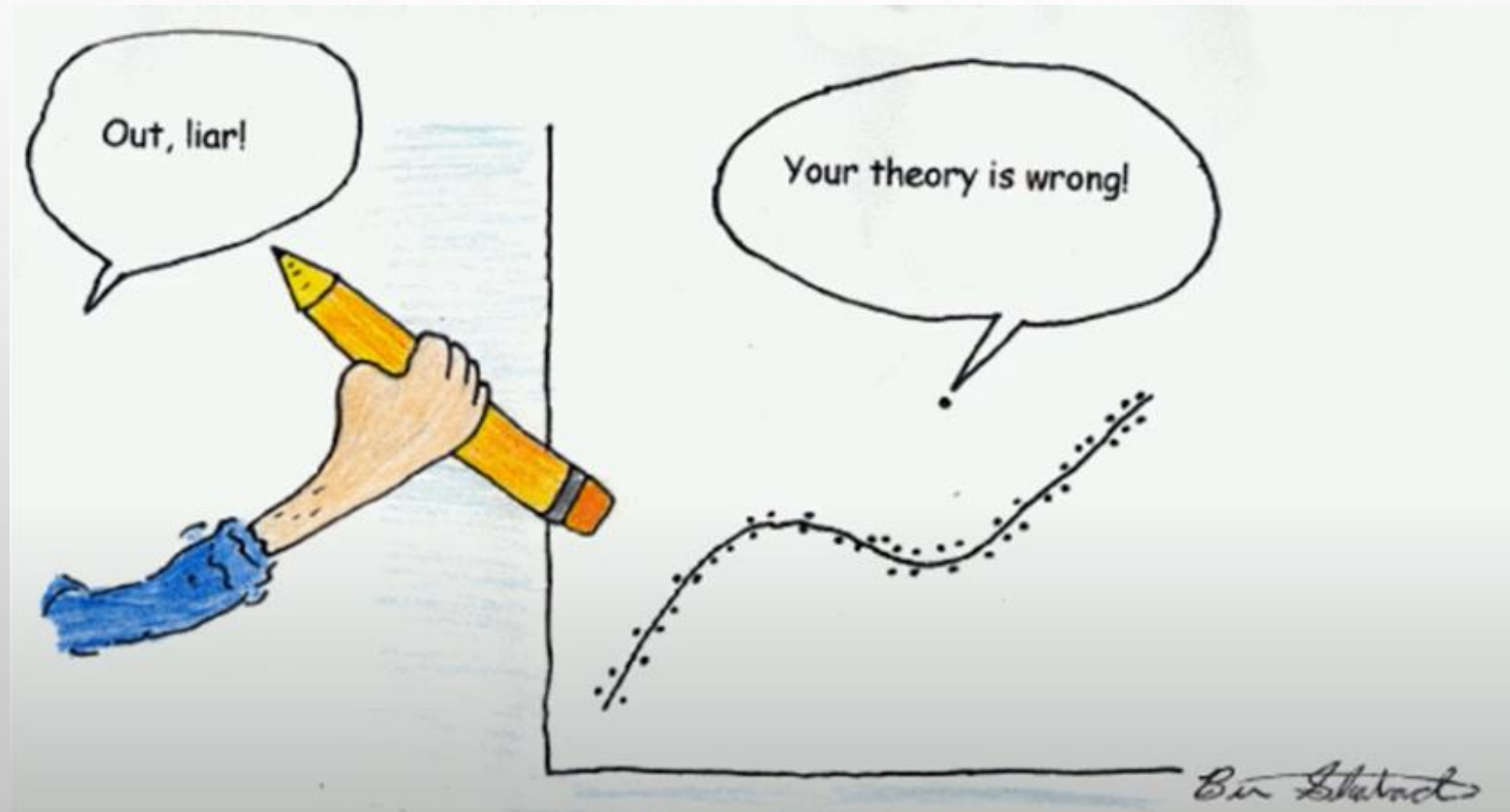
Pokud se provádí studie o prevalenci určité psychologické poruchy v populaci, výsledky lze ověřit podle zavedených lékařských záznamů nebo předchozího výzkumu, aby se zajistilo, že jsou v přijatelném rozmezí.



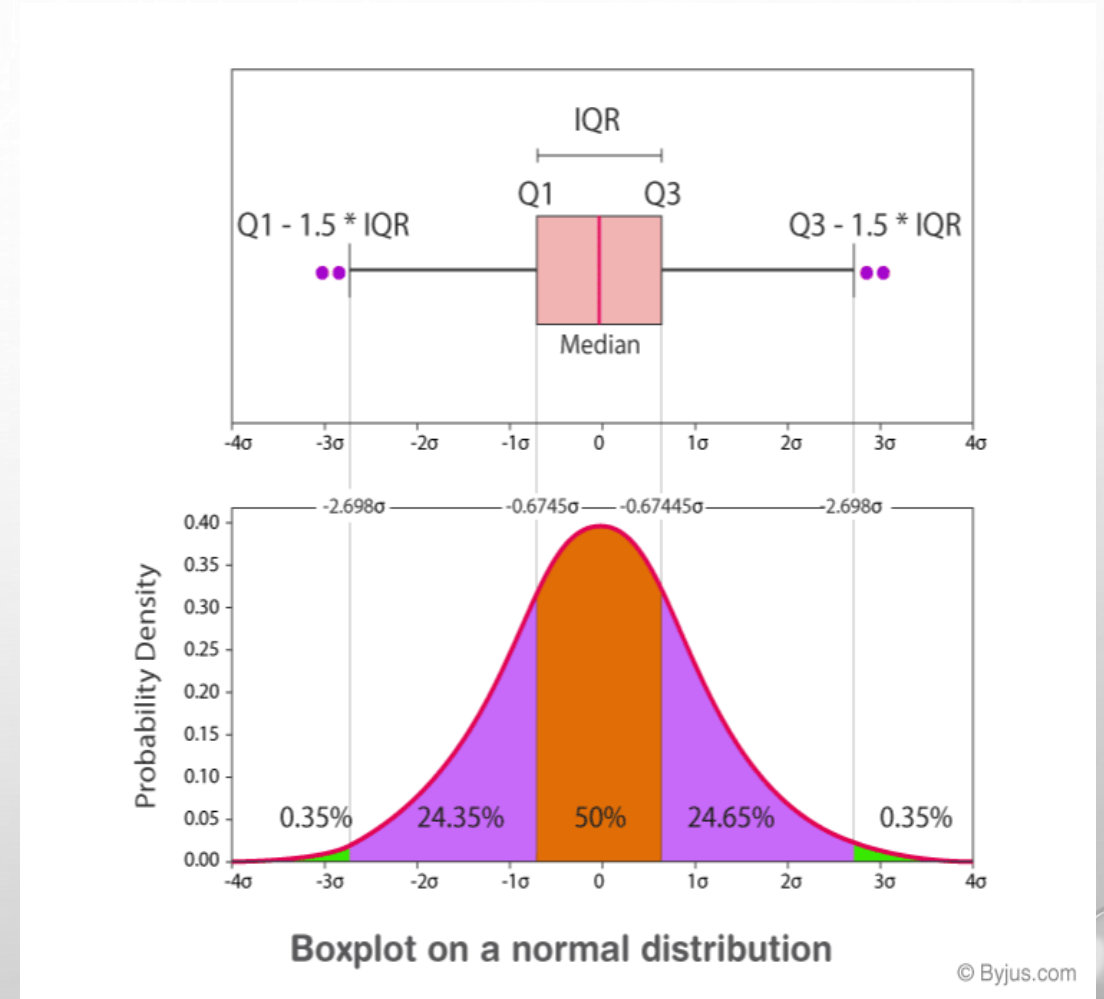
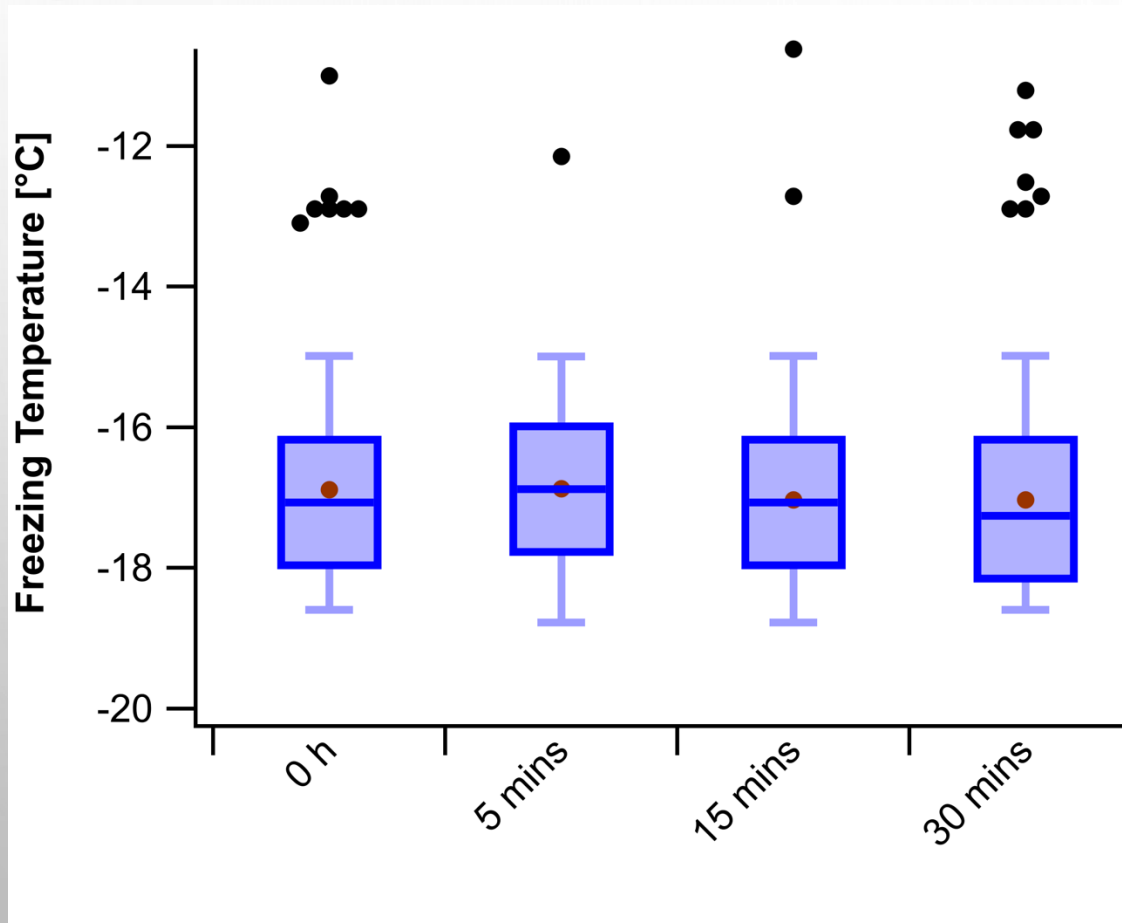
Deduplikace

Pokud se ve studii využívající online průzkum zjistí, že stejný účastník odeslal odpovědi dvakrát (možná jednou ze svého mobilu a jednou z počítače), bude nutné jeden ze záznamů odstranit, aby se předešlo nadbytečnosti.

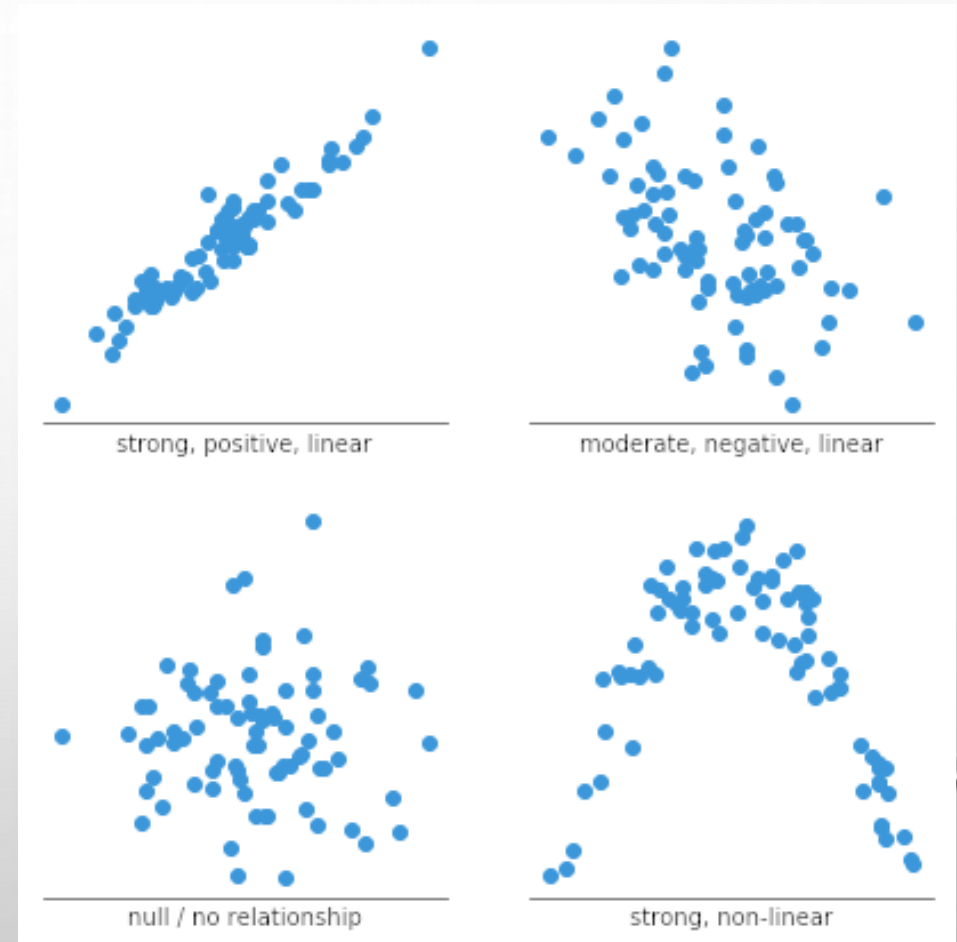
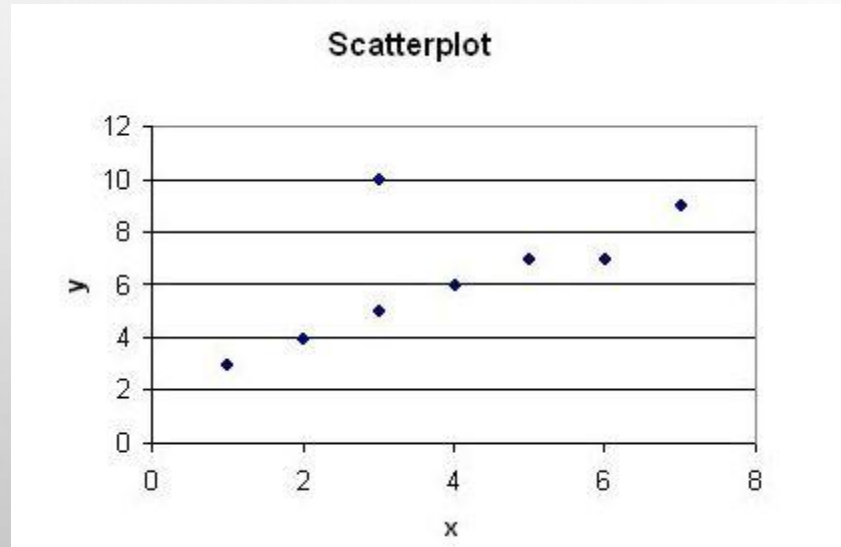
ODLEHLÉ HODNOTY



PŘÍKLADY



PŘÍKLADY



DOPAD NA ANALÝZU

- ZEŠIKMENÍ
- ZAVÁDĚJÍCÍ PRŮMĚR
- SNÍŽENÁ PLATNOST VÝSLEDKŮ

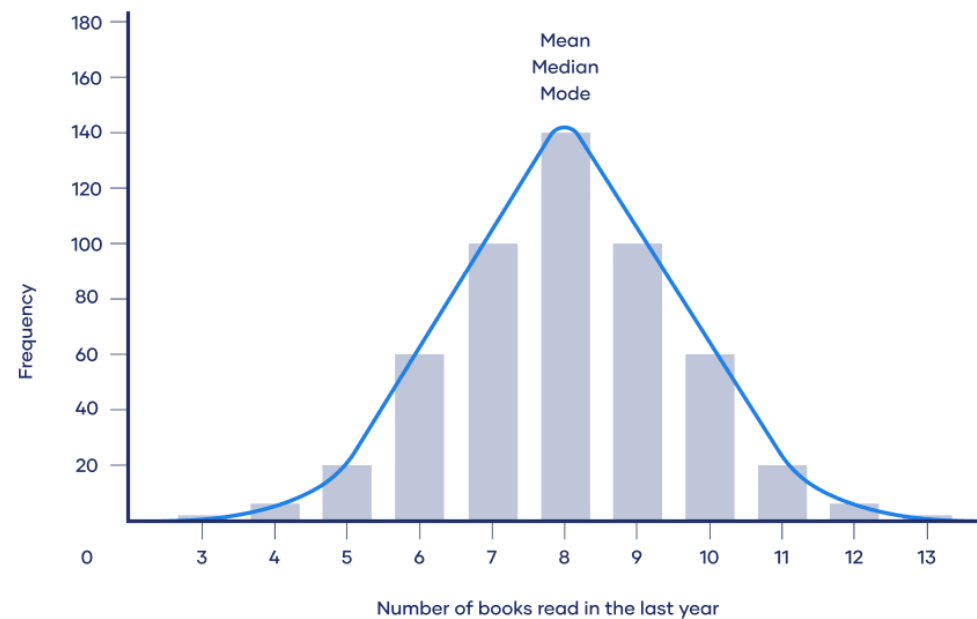


ŘEŠENÍ

- VYLOUČENÍ EXTRÉMNÍCH HODNOT
- TRANSFORMACE
- ROBUSTNÍ STATISTIKA

CENTRÁLNÍ TENDENCE

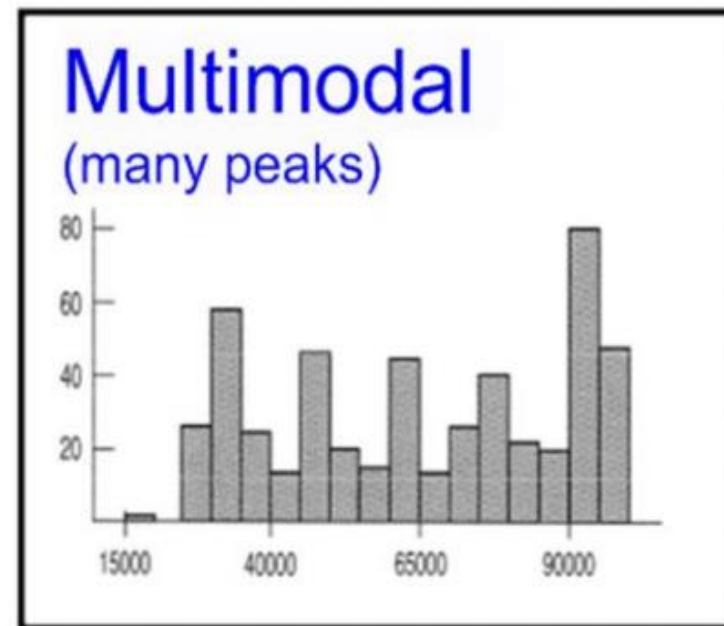
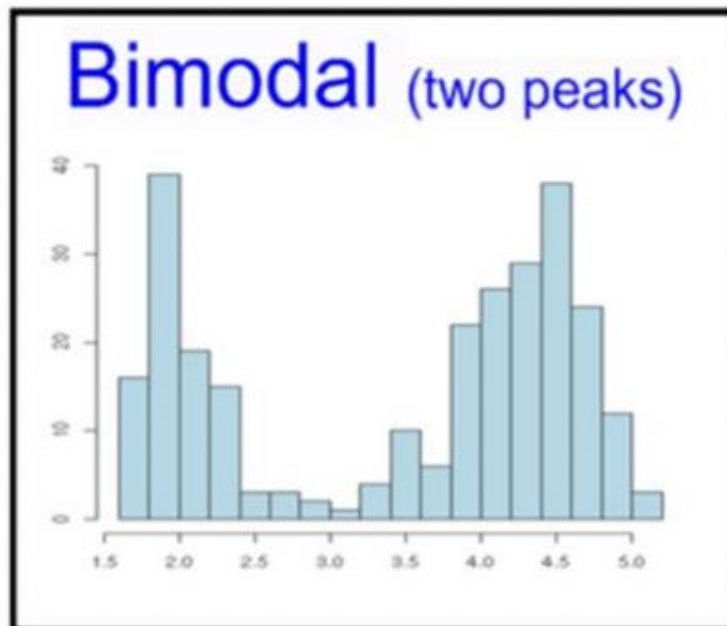
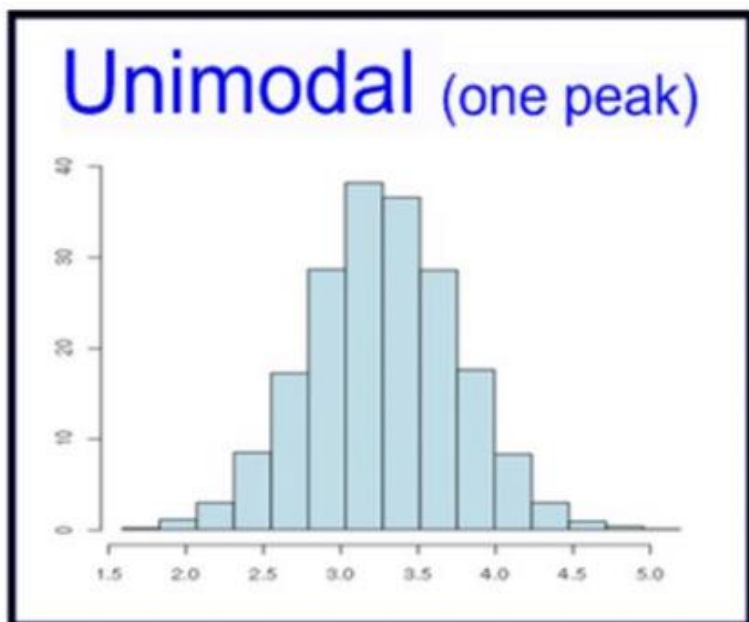
Normal distribution: Number of books read in the last year



CENTRÁLNÍ TENDENCE - MODUS

- JE HODNOTA NEBO HODNOTY, KTERÉ SE V DATOVÉ SADĚ OBJEVUJÍ NEJČASTĚJI.
- V PRŮZKUMU DOTAZUJÍCÍM SE ÚČASTNÍKŮ NA JEJICH OBLÍBENÝ TYP TERAPEUTICKÉHO PŘÍSTUPU (NAPŘ. CBT, PSYCHODYNAMICKÝ, HUMANISTICKÝ ATD.), REŽIM UVEDE NEJPREFEROVANĚJŠÍ METODU MEZI ÚČASTNÍKY.
- VHODNÝ PRO **NOMINÁLNÍ, ORDINÁLNÍ, INTERVALOVÉ A POMĚROVÉ** PROMĚNNÉ
- 1,2,2,2,2,3,3,4,5,5,5
- ČERVENÁ, MODRÁ, FIALOVÁ, ZELENÁ, MODRÁ , ZELENÁ
- LEDEN, BŘEZEN, LEDEN, ČERVEN, ČERVEN, SRPEN, ÚNOR, PROSINEC, SRPEN

CENTRÁLNÍ TENDENCE - MODUS



CENTRÁLNÍ TENDENCE - MEDIÁN

- STŘEDNÍ HODNOTA V USPOŘÁDANÉ DATOVÉ SADĚ. POKUD EXISTUJE SUDÝ POČET HODNOT, JE TO PRŮMĚR DVOU STŘEDNÍCH HODNOT.
- ZVAŽTE TERAPEUTICKOU STUDII, KDE ÚČASTNÍCI UVÁDĚJÍ POČET DNÍ, KDY V POSLEDNÍM MĚSÍCI CÍTILI ÚZKOST. POKUD NĚKOLIK ÚČASTNÍKŮ POCIŤOVALO ÚZKOST TĚMĚŘ KAŽDÝ DEN, ZATÍMCO VĚTŠINA POCIŤOVALA ÚZKOST POUZE NĚKOLIK DNÍ, MŮŽE MEDIÁN POSKYTNOUT REPREZENTATIVNĚJŠÍ CENTRÁLNÍ HODNOTU NEŽ PRŮMĚR.
- MEDIÁN JE VHODNÝ PRO ORDINÁLNÍ, INTERVALOVÉ A POMĚROVÉ PROMĚNNÉ.
- 1, 27,29,27,28,5,5,5,4,3,2,6,4,3,2 -----> 1,2,2,3,3,4,4,**5**,5,5,6,27,27,28,29

CENTRÁLNÍ TENDENCE - ARITMETICKÝ PRŮMĚR

$$\bar{X} = \frac{\sum X}{N}$$

- PŘEDSTAVTE SI, ŽE BYSTE PROVEDLI STUDII, KTERÁ BY MĚŘILA SKÓRE SEBEVĚDOMÍ TEENAGERŮ POMOCÍ ROSENBERGOVY ŠKÁLY SEBEÚCTY, KTERÁ DÁVÁ SKÓRE MEZI 10 A 40. PRŮMĚRNÉ SKÓRE POSKYTNE PRŮMĚRNOU ÚROVEŇ SEBEÚCTY U VYBRANÝCH TEENAGERŮ.
- PRŮMĚR SE NEJLÉPE POUŽÍVÁ PRO INTERVALOVÉ A POMĚROVÉ PROMĚNNÉ.
- 5, 10, 15, 5, 10 -> 9

MÍRY VARIABILITY (ROZPTÝLENOSTI) - ROZPĚTÍ

- JE NEJJEDNODUŠŠÍ MÍROU VARIABILITY. POČÍTÁ SE JAKO ROZDÍL NEJVĚTŠÍ A NEJMENŠÍ HODNOTY SOUBORU
- PŘEDPOKLÁDEJME, ŽE PSYCHOLOG PROVEDE PRŮZKUM STRESU SKUPINĚ PACIENTŮ SE SKÓRE V ROZMEZÍ OD 1 (NEJMÉNĚ STRESOVANÍ) DO 10 (NEJVÍCE STRESOVANÍ). POKUD SE SKÓRE POHYBUJE OD 3 DO 9, JE ROZPĚTÍ 6. TO PSYCHOLOGOVI ŘÍKÁ O CELKOVÉM ROZLOŽENÍ ÚROVNÍ STRESU VE SKUPINĚ, ALE NEPOSKYTUJE PODROBNÉ INFORMACE O TOM, JAK JSOU SKÓRE V TOMTO ROZMEZÍ DISTRIBUOVÁNA.
- LZE POUŽÍT PRO ORDINÁLNÍ, INTERVALOVÉ A POMĚROVÉ PROMĚNNÉ.
- **MEZIKVARTILOVÉ ROZPĚTÍ**, TĚŽ **MEZIKVARTILOVÁ ŠÍŘE** (ANGL. *INTERQUARTILE RANGE*, ZKRATKA **IQR**) PŘEDSTAVUJE ROZDÍL MEZI TŘETÍM A PRVNÍM KVARTILEM (TEDY MEZI 75. A 25. PERCENTILEM). REPREZENTUJE TEDY OBLAST HODNOT, KTERÉ MÁ STŘEDNÍCH 50 % HODNOT PROMĚNNÉ

MÍRY VARIABILITY – SMĚRODATNÁ ODCHYLKA

- JE DRUHÁ ODMOCNINA ROZPTYLU
- POSKYTUJE MĚŘÍTKO PRŮMĚRNÉ VZDÁLENOSTI KAŽDÉ HODNOTY OD PRŮMĚRU A POSKYTUJE PŘEHLED O ŠÍŘENÍ DAT.
- SMĚRODATNÁ ODCHYLKA UMOŽŇUJE PSYCHOLOGŮM POROZUMĚT ŠÍŘENÍ SKÓRE KOLEM PRŮMĚRU. POKUD JE NAPŘÍKLAD V TESTU IQ PRŮMĚR 100 A SMĚRODATNÁ ODCHYLKA 15, ZNAMENÁ TO, ŽE VĚTŠINA SKÓRE LEŽÍ MEZI 85 ($100 - 15$) A 115 ($100 + 15$).
- STEJNĚ JAKO ROZPTYL JE STANDARDNÍ ODCHYLKA NEJVÍCE POUŽITELNÁ PRO INTERVALOVÉ A POMĚROVÉ PROMĚNNÉ.

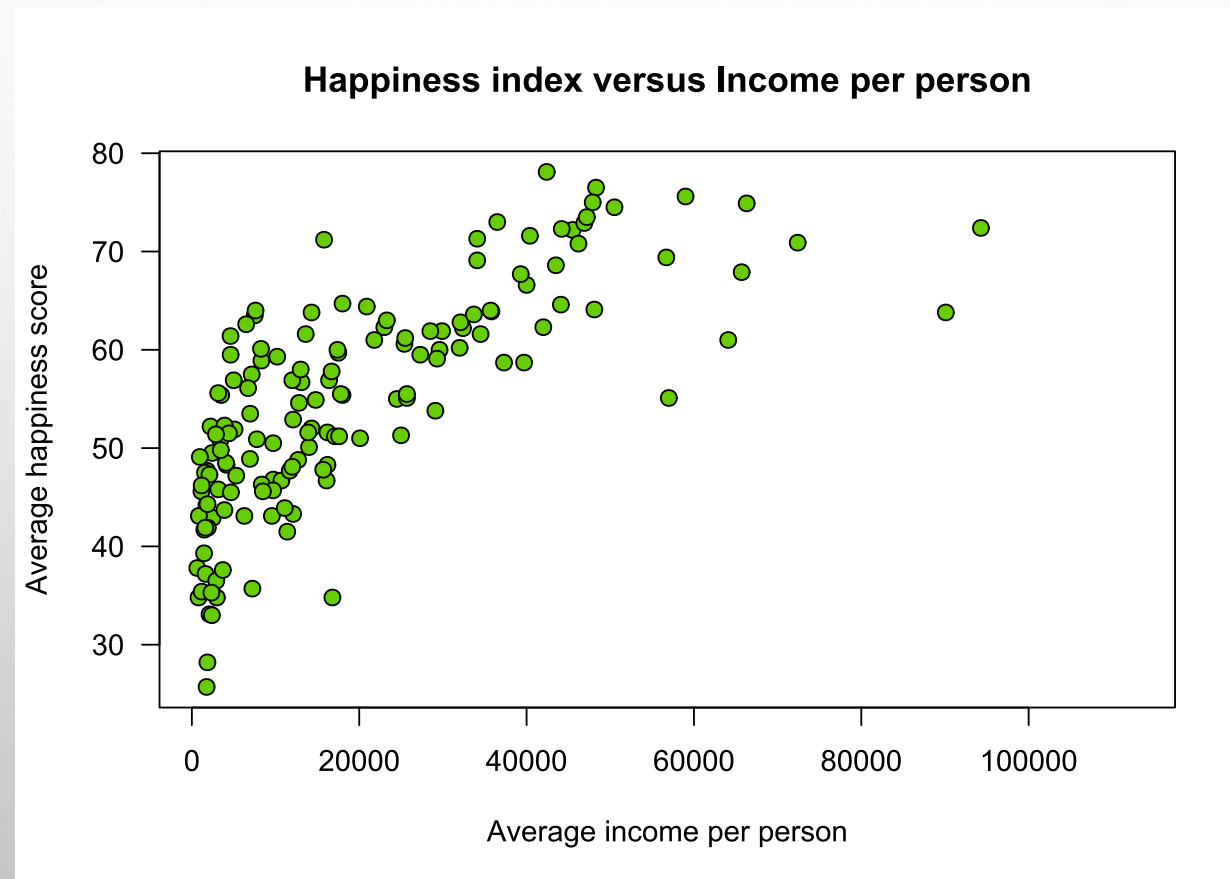
MÍRY VARIABILITY – ROZPTYL (VARIANCE)

- MĚŘÍ, JAK DALEKO JE KAŽDÁ HODNOTA V DATOVÉ MNOŽINĚ OD PRŮMĚRU, A TEDY OD VŠECH OSTATNÍCH HODNOT V SADĚ. POSKYTUJE PŘEDSTAVU O ŠÍŘENÍ NEBO ROZPTYLU POZOROVANÝCH HODNOT.
- PRŮMĚRNÁ ODCHYLKA NA DRUHOU
- POCHOPENÍ ROZPTYLU JE ZÁSADNÍ PŘI HODNOCENÍ KONZISTENCE CHOVÁNÍ NEBO REAKCÍ. NAPŘÍKLAD PŘI HODNOCENÍ ÚČINNOSTI TERAPIE, POKUD ÚČASTNÍCI VYKAZUJÍ PODOBNÉ VÝSLEDKY, JE ROZPTYL NÍZKÝ. POKUD JSOU VÝSLEDKY VELMI ODLIŠNÉ, JE ROZPTYL VYSOKÝ, COŽ UKAZUJE NA RŮZNÉ ÚČINKY TERAPIE.
- ROZPTYL SE OBVYKLE POUŽÍVÁ PRO INTERVALOVÉ A POMĚROVÉ PROMĚNNÉ.

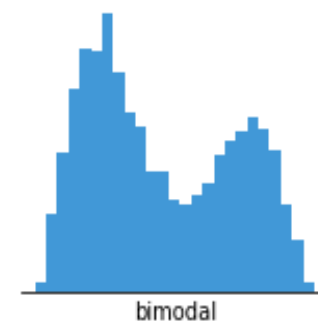
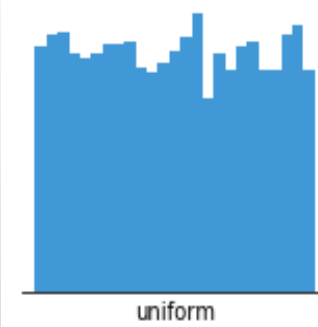
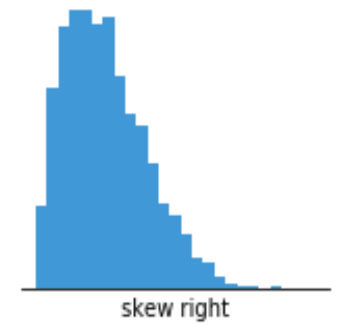
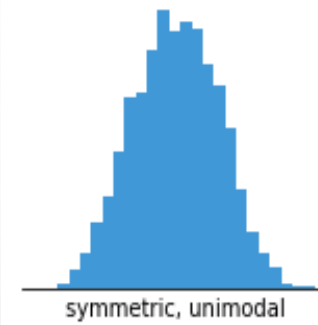
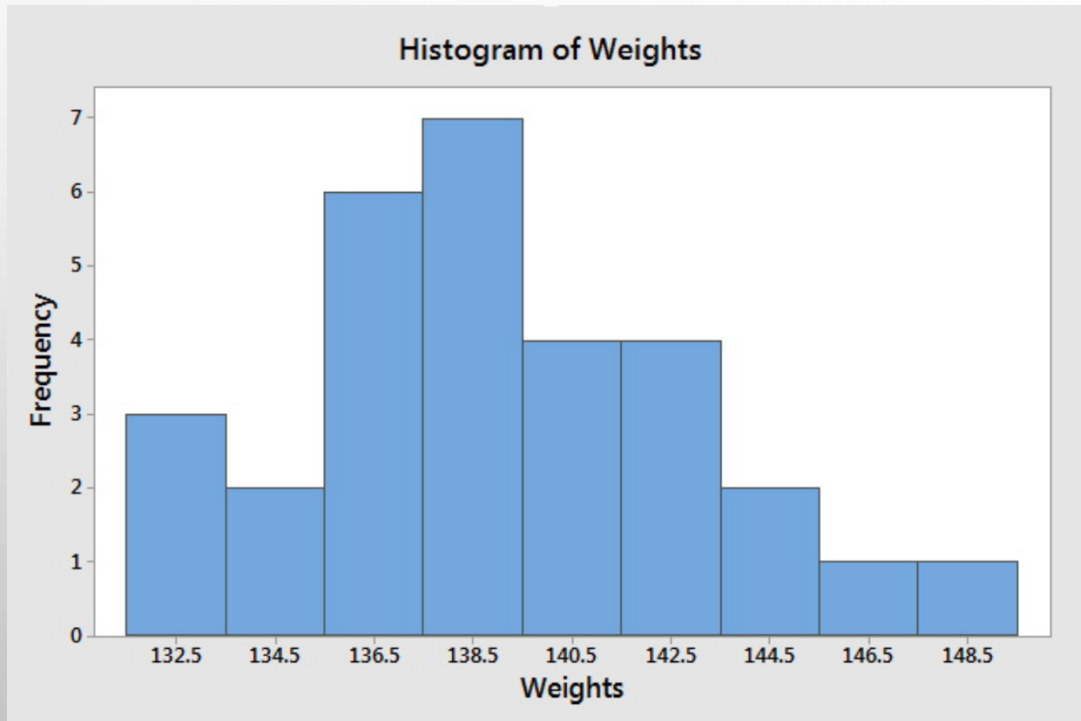
The image features a light gray gradient background with several realistic water droplets of various sizes scattered in the corners. The droplets have highlights and shadows, giving them a three-dimensional appearance. The word "GRAFY" is centered in the upper half of the image.

GRAFY

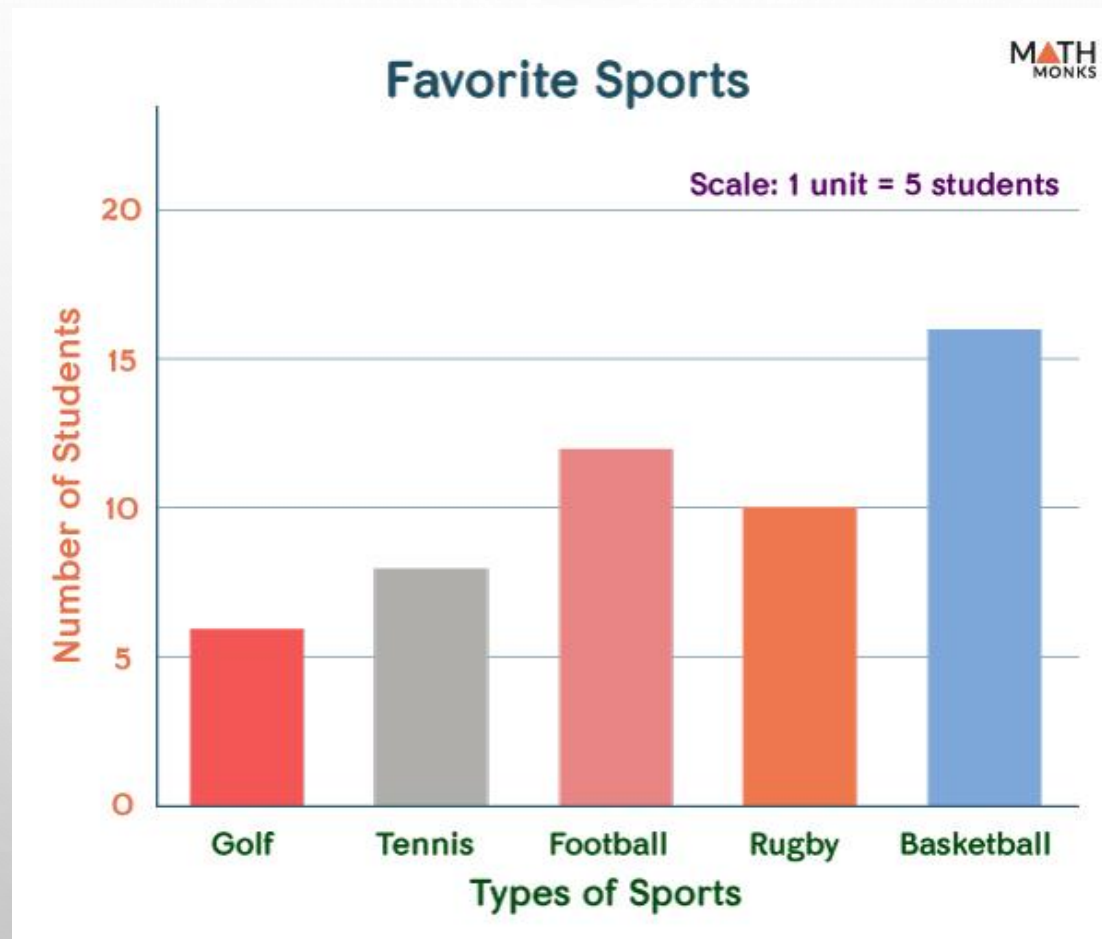
BODOVÝ GRAF



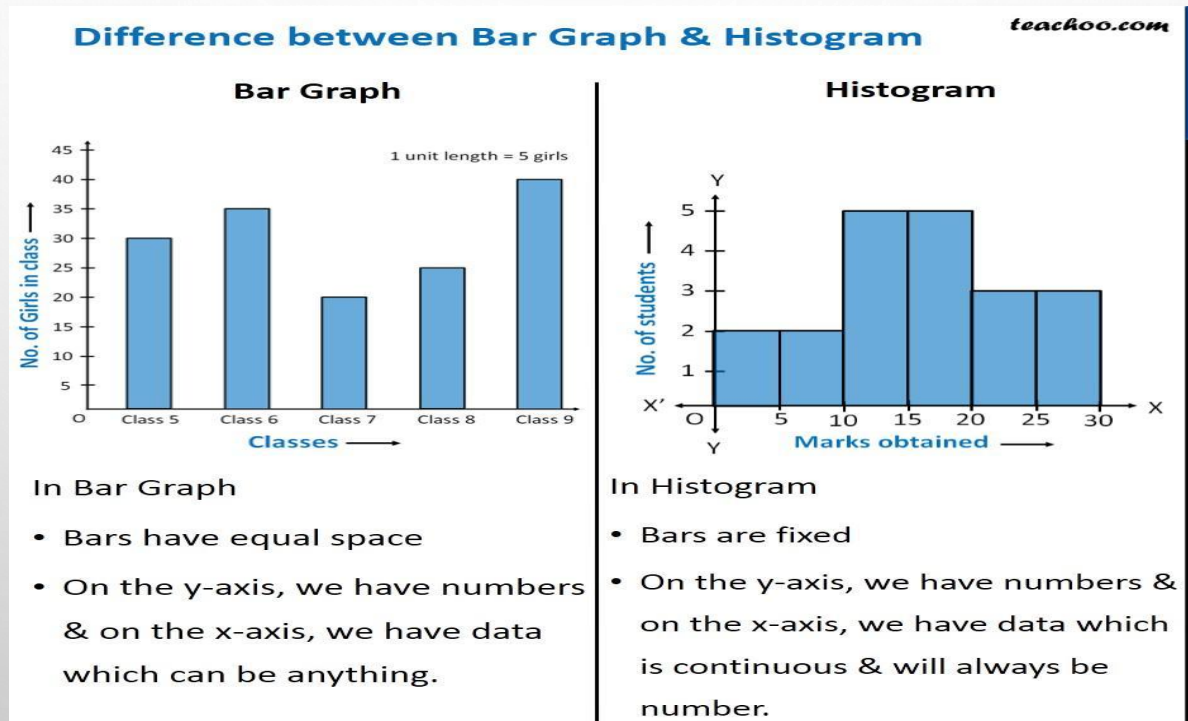
HISTOGRAM



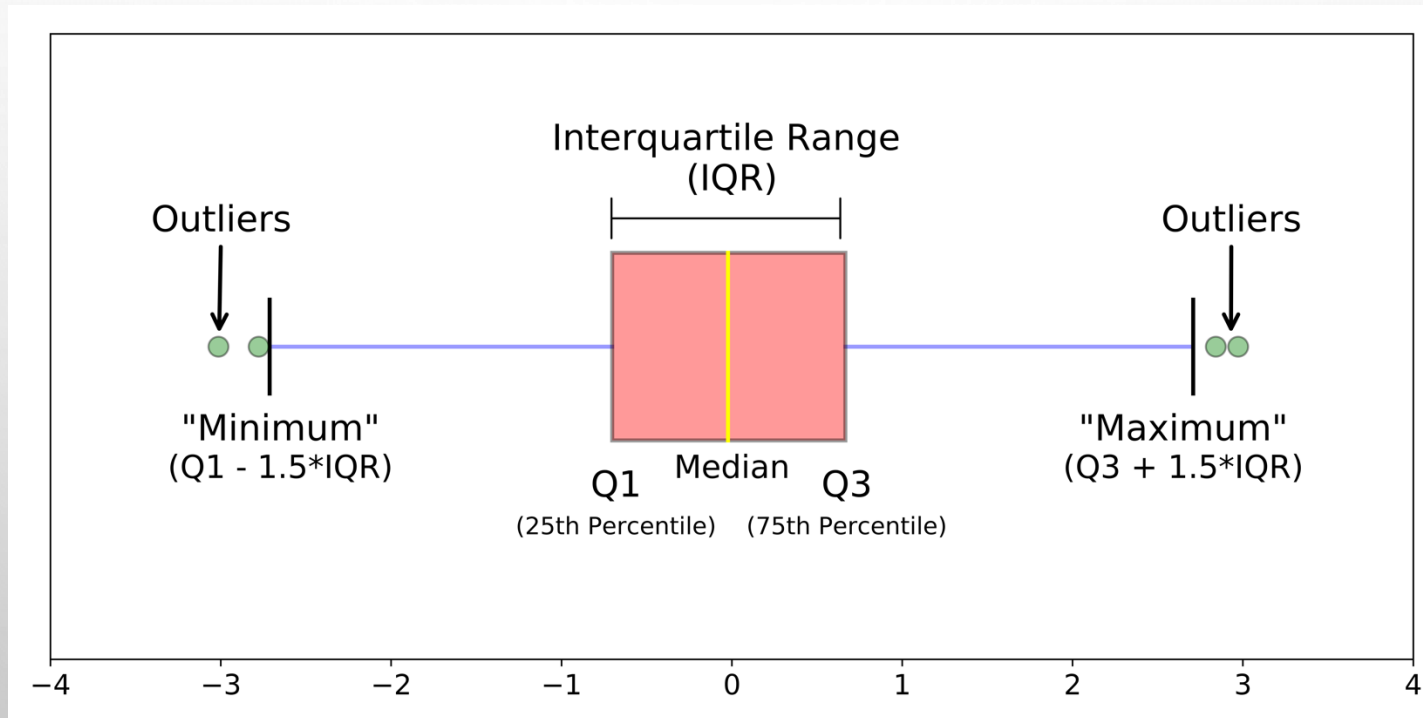
SLOUPCOVÝ GRAF



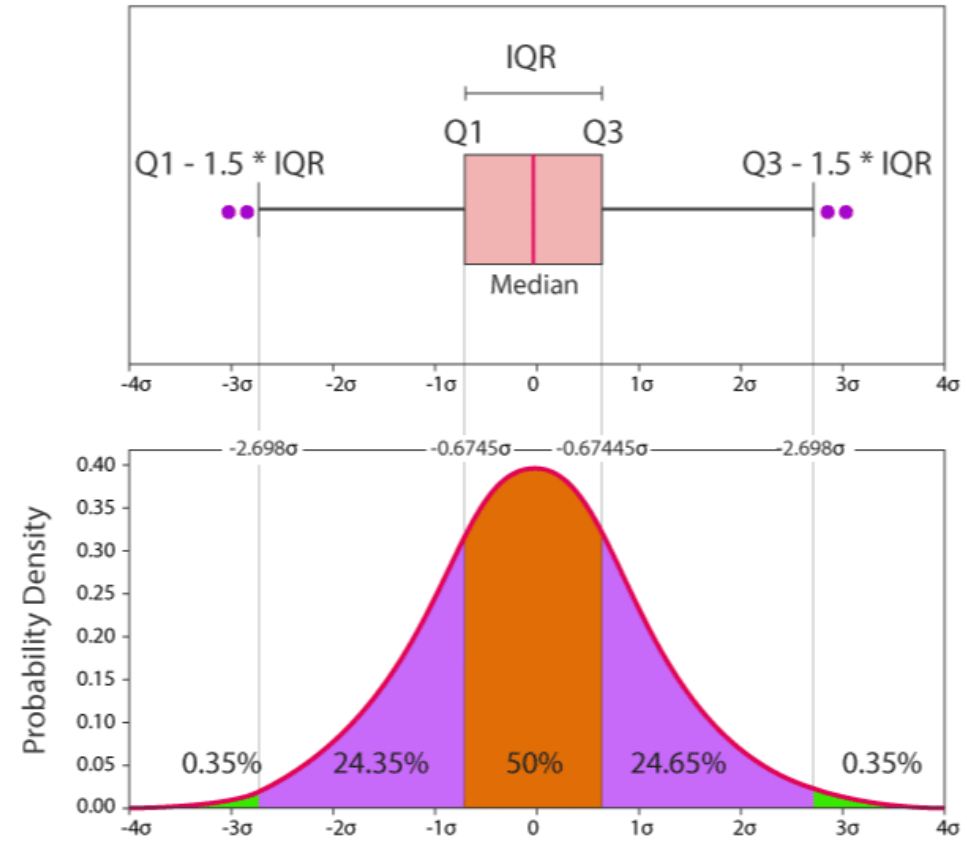
HISTOGRAM VS SLOUPCOVÝ GRAF



BOXPLOT



BOXPLOT



Boxplot on a normal distribution

ZÁVĚREM

- **POPISNÁ STATISTIKA**
 - ZÁKLADNÍ PRO JAKOUKOLI ANALÝZU DAT, POSKYTUJE ZPŮSOB, JAK SHRNOU A POPSAT HLAVNÍ RYSY DATOVÉ SADY, NABÍZÍ JEDNODUCHÝ PŘEHLED BEZ SLOŽITÝCH INTERPRETACÍ.
- **ČIŠTĚNÍ DAT**
 - VŽDY JE DŮLEŽITÉ ZAČÍT S ČISTOU DATOVOU SADOU. DISKUTOVANÉ TECHNIKY, JAKO JE IMPUTACE, MAZÁNÍ, KOREKCE A STANDARDIZACE.
- **ODLEHLÉ HODNOTY**
 - , JAK MOHOU ODLEHLÉ HODNOTY ZKRESLIT VÝSLEDKY A POTENCIÁLNĚ VÉST K NESPRÁVNÝM ZÁVĚRŮM. PROZKOUMANÉ METODY DETEKCE A MANIPULACE.
- **CENTRÁLNÍ TENDENCE**
 - MÍRY CENTRÁLNÍ TENDENCE POMÁHAJÍ STANOVIT ZÁKLADNÍ LINII NEBO PRŮMĚRNOU ODPOVĚĎ, KTERÁ MŮŽE BÝT KLÍČOVÁ PŘI VYVOZOVÁNÍ ZÁVĚRŮ. NAPŘÍKLAD PŘI TESTOVÁNÍ ÚČINNOSTI TERAPEUTICKÉ INTERVENCE MŮŽE PRŮMĚRNÉ SKÓRE PO INTERVENCI VE SROVNÁNÍ S PRŮMĚREM PŘED INTERVENCÍ POSKYTNOUT NÁHLED NA ÚČINNOST INTERVENCE.
 - V PROSTŘEDÍ VÝZKUMU UMOŽŇUJE POCHOPENÍ CENTRÁLNÍ HODNOTY SROVNÁNÍ MEZI SKUPINAMI NEBO PODMÍNKAMI. NAPŘÍKLAD POROVNÁVÁNÍ STŘEDNÍCH ODPOVĚDÍ KONTROLNÍ SKUPINY S ODPOVĚĎMI EXPERIMENTÁLNÍ SKUPINY MŮŽE BÝT V EXPERIMENTÁLNÍ PSYCHOLOGII KLÍČOVÉ
- **VARIABILITA**
 - MÍRY ROZPTYLU, ZEJMÉNA STANDARDNÍ ODCHYLKA, POSKYTUJÍ POHLED NA TO, JAK JSOU SKÓRE ROZLOŽENA KOLEM CENTRÁLNÍ HODNOTY. POKUD JE NAPŘÍKLAD V TESTU IQ PRŮMĚRNÉ IQ 100 SE SMĚRODATNOU ODCHYLKOU 15, ZNAMENÁ TO, ŽE VĚTŠINA JEDINCŮ DOSÁHNE SKÓRE V ROZMEZÍ 85 AŽ 115.
 - VYSOKÝ STUPEŇ ROZPTYLU MŮŽE ZNAMENAT RŮZNORODOU SKUPINU S RŮZNÝMI REAKCEMI, ZATÍMCO NÍZKÝ ROZPTYL ZNAMENÁ HOMOGENNĚJŠÍ SKUPINU.
- **GRAFY**
 - GRAFY A GRAFY POSKYTUJÍ INTUITIVNÍ A OKAMŽITÉ POCHOPENÍ DATOVÝCH TRENDŮ, ROZLOŽENÍ A VZORCŮ. NEŽ SE PONOŘÍME DO KOMPLEXNÍCH STATISTICKÝCH ANALÝZ, JEDNODUCHÝ GRAF MŮŽE ČASTO UKÁZAT, ZDA EXISTUJÍ ZNATELNÉ ROZDÍLY MEZI SKUPINAMI, POTENCIÁLNÍ ODLEHLÉ HODNOTY NEBO ZAJÍMAVÉ VZORCE

1) KTERÁ Z NÁSLEDUJÍCÍCH NENÍ MĚŘÍTKEM CENTRÁLNÍ TENDENCE?

- A) ROZPTYL
- B) PRŮMĚR
- C) MEDIÁN
- D) MODUS

2) KTERÉ GRAFICKÉ ZNÁZORNĚNÍ SE NEJLÉPE POUŽÍVÁ K POZOROVÁNÍ ROZLOŽENÍ KONTINUÁLNÍ/SPOJITÉ

- A) BODOVÝ GRAF
- B) KRABICOVÝ POZEMEK
- C) HISTOGRAM
- D) SLOUPCOVÝ GRAF

3) KTERÁ ČÁRA V KRABICOVÉM GRAFU PŘEDSTAVUJE MEDIÁN?

- A) HORNÍ ČÁST KRABICE
- B) SPODNÍ ČÁST KRABICE
- C) ČÁRA UVNITŘ KRABICE
- D) VOUS

4) ROZPĚTÍ/ROZSAH JE DEFINOVÁN JAKO:

- A) MAXIMÁLNÍ HODNOTA - MINIMÁLNÍ HODNOTA
- B) PRŮMĚR - MODUS
- C) MEDIÁN - MEDIÁN
- D) PRŮMĚR - MEDIÁN

• 5) KTERÁ Z NÁSLEDUJÍCÍCH JE DRUHÁ ODMOCNINA ROZPTYLU? A) ROZSAH

- A) ROZSAH
- B) MEDIÁN
- C) MODUS
- D) SMĚRODATNÁ ODCHYLKA

6) KTERÁ MÍRA CENTRÁLNÍ TENDENCE JE NEJVÍCE OVLIVNĚNA EXTRÉMNÍMI HODNOTAMI?

- A) MODUS
- B) MEDIÁN
- C) PRŮMĚR

7) POKUD MÁTE BIMODÁLNÍ DISTRIBUCI, KOLIK MÁTE VRCHOLŮ?

- A) 0
- B) 1
- C) 2
- D) 3

8) ČIŠTĚNÍ DAT POMÁHÁ PŘEDEVŠÍM V:

- A) ZLEPŠENÍ DESIGNU EXPERIMENTU
- B) ZVĚTŠENÍ VELIKOSTI VZORKU
- C) SNÍŽENÍ ZKRESLENÍ PŘI INTERPRETACI DAT
- D) ABY DATA VYPADALA KRÁSNĚJI

9) KTERÝ VÝRAZ OZNAČUJE HODNOTU V SOUBORU DAT, KTERÁ SE VÝRAZNĚ LIŠÍ OD OSTATNÍCH HODNOT?

- A) PRŮMĚR
- B) ODLEHLÁ HODNOTA
- C) SMĚRODATNÁ ODCHYLKA
- D) MEDIÁN

10) POKUD MÁ DATOVÁ SADA VÍCE NEŽ DVA MODUSY, OZNAČUJE SE JAKO:

- A) BIMODÁLNÍ
- B) TRIMODÁLNÍ
- C) MULTIMODÁLNÍ

11) KTERÁ Z NÁSLEDUJÍCÍCH NENÍ KROKEM V ČIŠTĚNÍ DAT?

- A) IMPUTACE
- B) VÝPOČET PRŮMĚRU
- C) VYMAZÁNÍ
- D) OPRAVA

12) JAKÁ MÍRA CENTRÁLNÍ TENDENCE JE NEJVHODNĚJŠÍ PRO ORDINÁLNÍ DATA?

- A) PRŮMĚR
- B) MEDIÁN
- C) MODUS

13) KTERÉ Z NÁSLEDUJÍCÍCH GRAFICKÝCH ZNÁZORNĚNÍ MŮŽE UKÁZAT POTENCIÁLNÍ ODLEHLÉ HODNOTY?

- A) HISTOGRAM
- B) BODOVÝ GRAF
- C) KRABICOVÝ GRAF/ BOX PLOT

14) MODUS JE NEJLÉPE POPSÁN TAKTO:

- A) PRŮMĚR SOUBORU DAT
- B) STŘEDNÍ HODNOTA, KDYŽ JSOU DATA USPOŘÁDÁNA VZESTUPNĚ
- C) NEJČASTĚJI SE VYSKYTUJÍCÍ HODNOTA V SOUBORU DAT