



POPISNÉ STATISTIKY

MARTINA SEBALO VŇUKOVÁ

MARTINA.VNUKOVA@LF1.CUNI.CZ



AGENDA

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

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) ORDINÁLNÍ
- 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 NEEEXISTUJE ŽÁDNÉ POŘADÍ NEBO POŘADÍ?

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

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

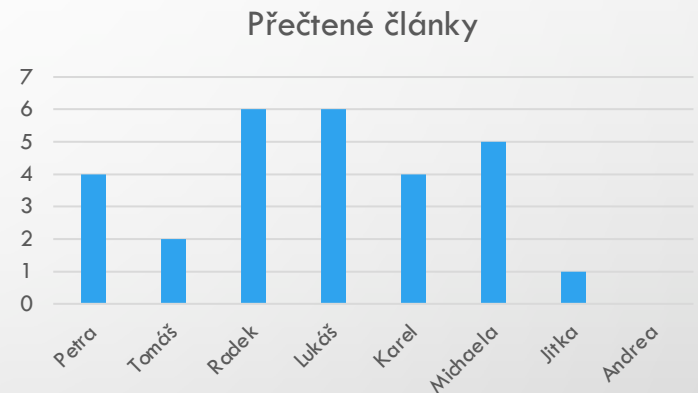
POPISNÁ STATISTIKA

- TŘÍDIT
 - 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



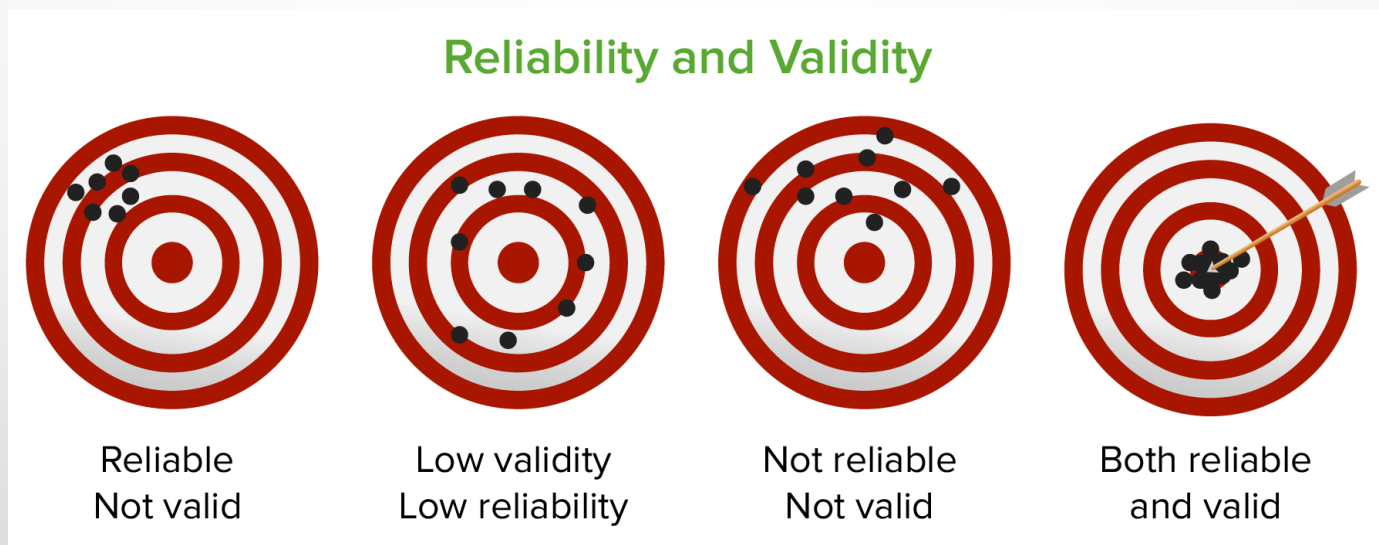
PROČ?

1. STRUČNÝ PŘEHLED DAT
2. ZJEDNODUŠENÍ
3. IDENTIFIKACE VZORCŮ A TRENDŮ
4. POROVNÁVÁNÍ
5. URČENÍ DISTRIBUCE DAT
6. IDENTIFIKACE ODLEHLÝCH HODNOT
7. ZÁKLAD PRO INFERENČNÍ STATISTIKU
8. VIZUALIZACE DAT
9. ZAJIŠTĚNÍ TRANSPARENTNOSTI A REPLIKOVATELNOSTI VÝZKUMU
10. KONTROLA KVALITY

ČIŠTĚNÍ DAT

- ZÁKLAD PRO ROZHODOVÁNÍ
- EFEKTIVITA ZDROJŮ
- VÝKON MODELU

● PŘESNOST , PLATNOST A SPOLEHLIVOST



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 „DYSFORIÍ“.
- 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.

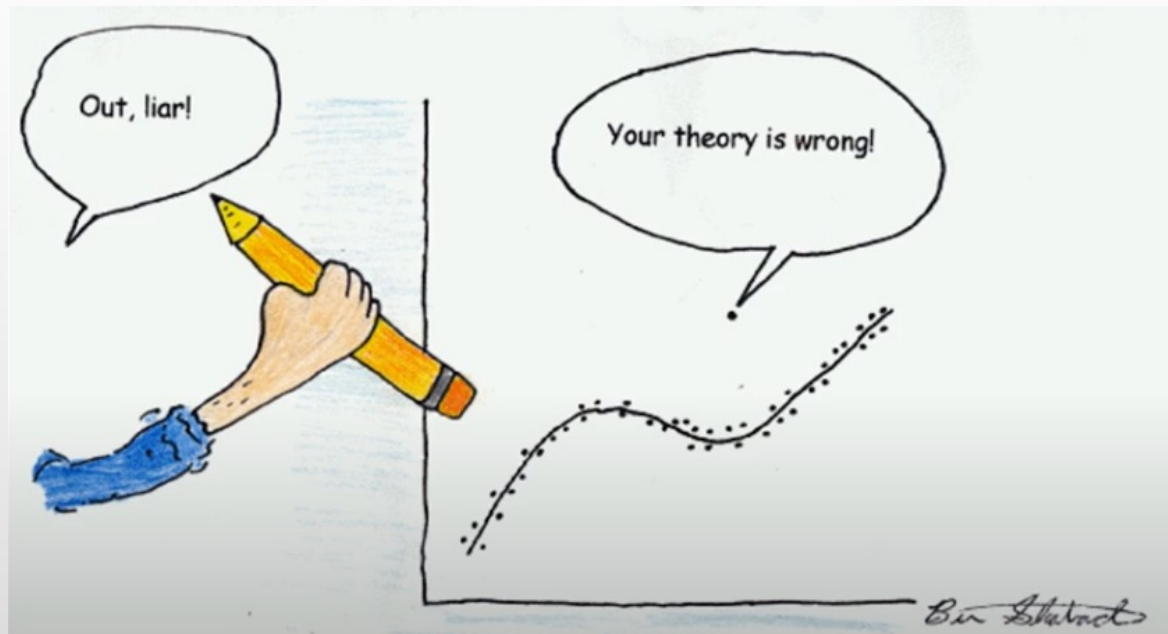
JAK IDENTIFIKOVAT PROBLÉMY

- 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.
- SOFTWAREVÉ NÁSTROJE
 - POMOCÍ SOFTWARE 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 ODCHYLJÍ 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.

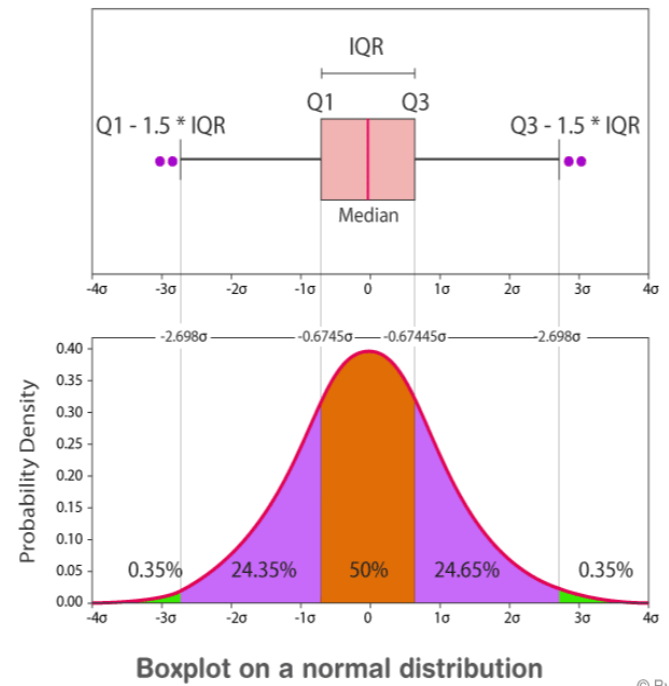
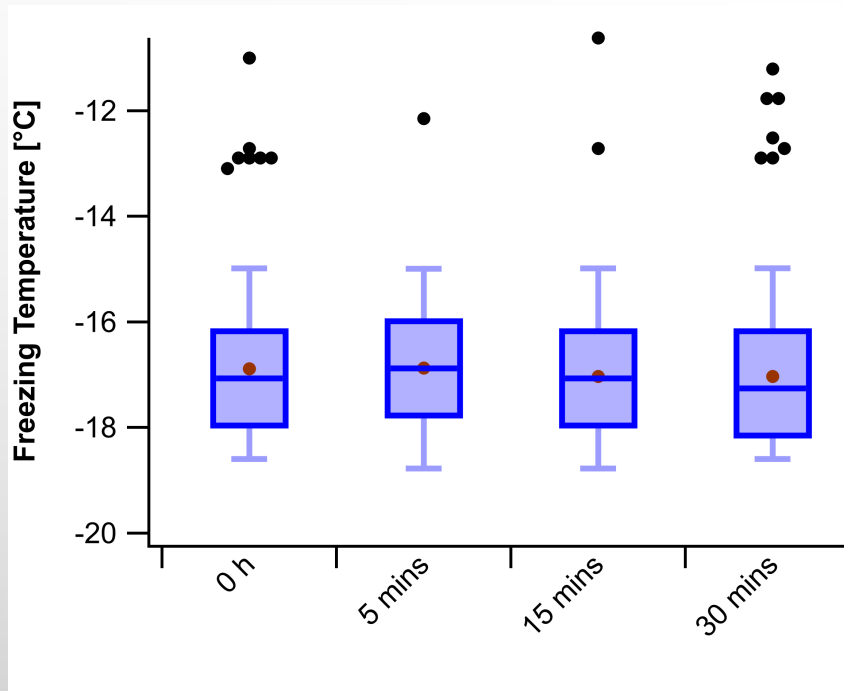
ŘEŠENÍ PROBLÉMŮ

- IMPUTACE - DOPLNĚNÍ / NAHRAZENÍ
 - V LONGITUDINÁLNÍ STUDII SLEDUJÍCÍ DEPRESIVNÍ SYMPTOMY V PRŮBĚHU ČASU, POKUD ÚČASTNÍK VYNECHÁ HODNOCENÍ JEDNOHO MĚSÍCE, CHYBĚJÍCÍ HODNOTA MŮŽE BÝT PŘÍČTENA 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.

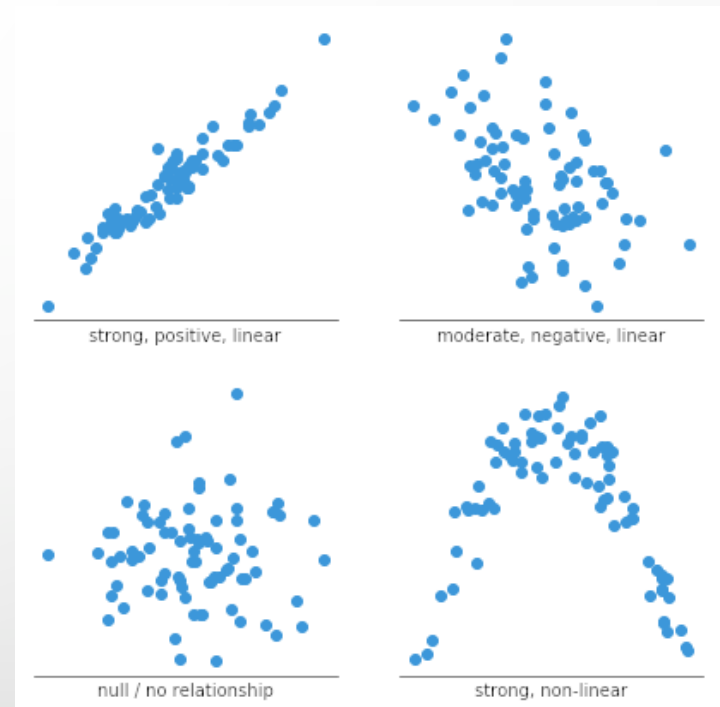
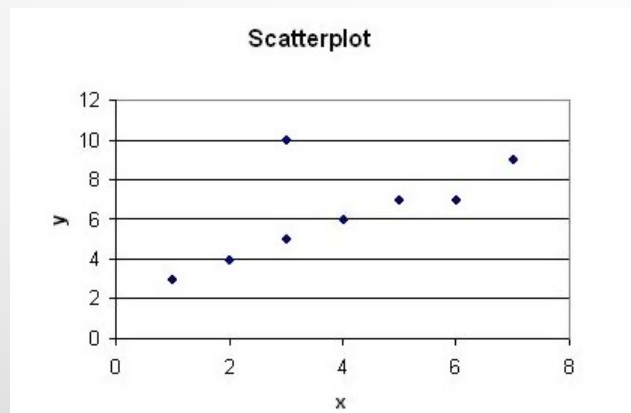
ODLEHLÉ HODNOTY



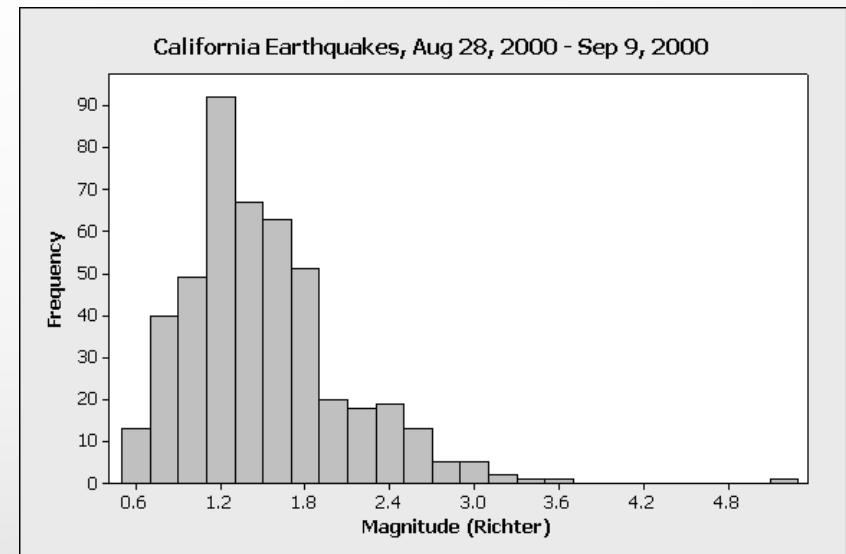
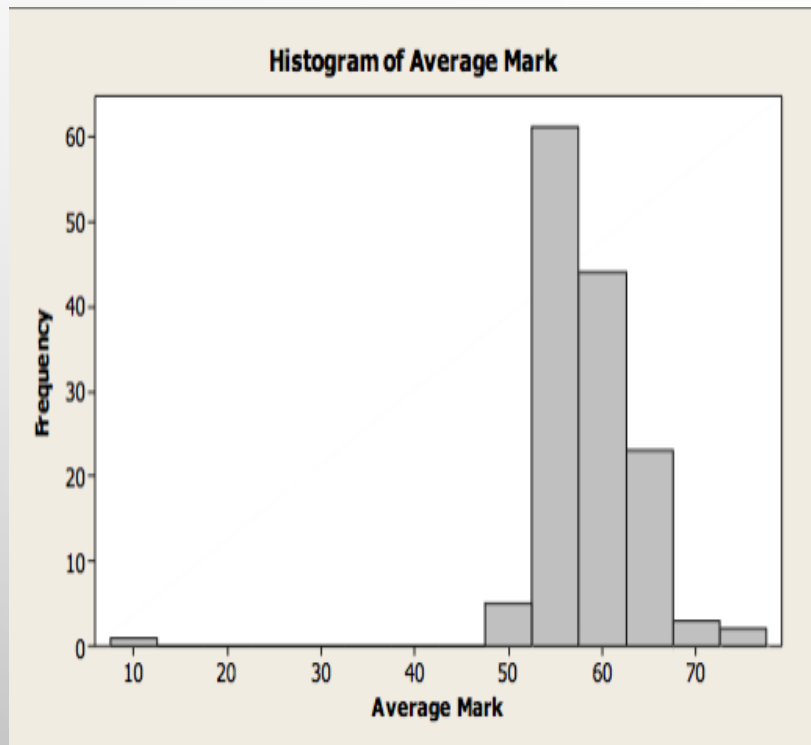
PŘÍKLADY



PŘÍKLADY



PŘÍKLADY



DOPAD NA ANALÝZU

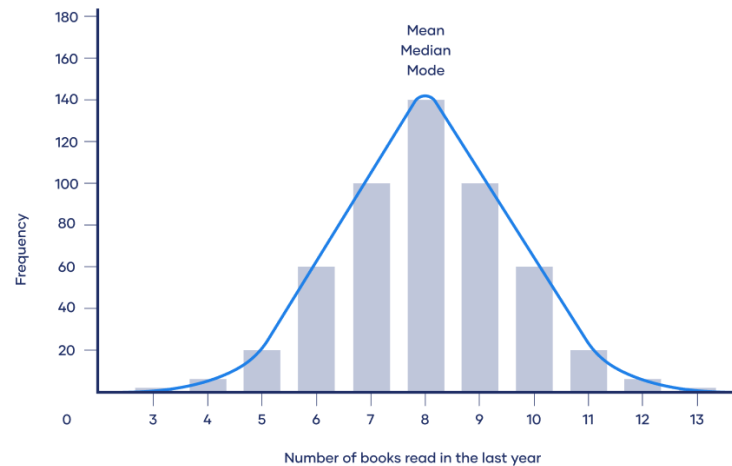
- ZEŠIKMENÍ
- ZAVÁDĚJÍCÍ PRŮMĚR
- SNÍŽENÁ PLATNOST VÝSLEDKŮ

ŘEŠENÍ

- VYLOUČENÍ EXTRÉMNÍCH HODNOT
- TRANSFORMACE
- ROBUSTNÍ STATISTIKA

CENTRÁLNÍ TENDENCE

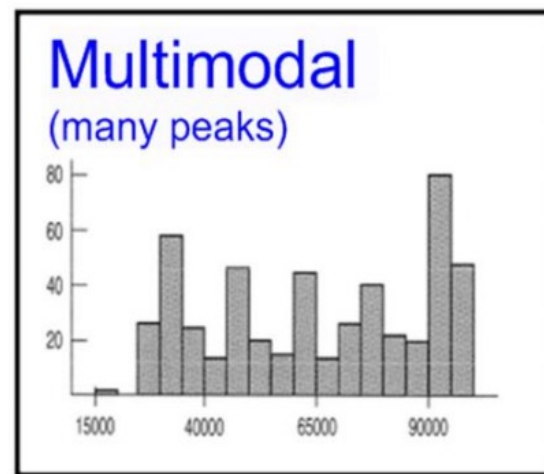
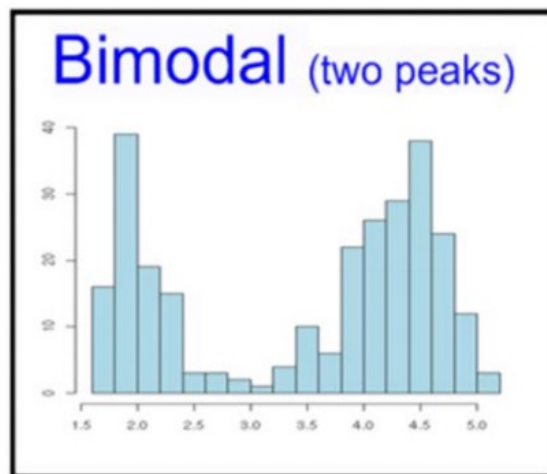
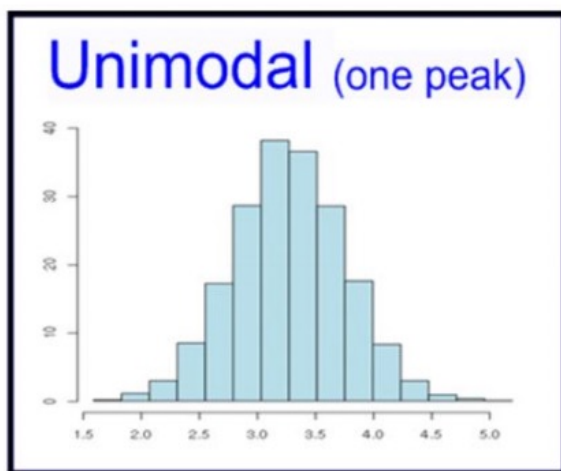
Normal distribution: Number of books read in the last year



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 CÍLILI ÚZKOST. POKUD NĚKOLIK ÚČASTNÍKŮ POCIŤOVALO ÚZKOST TÉMĚŘ 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

MÍRY VARIABILITY (ROZPTÝLENOSTI) - ROZPĚTÍ

- JE NEJEDNODUŠŠÍ 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**, TÉŽ **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 – 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.

MÍRY VARIABILITY – SMĚRODATNÁ ODCHYLKA

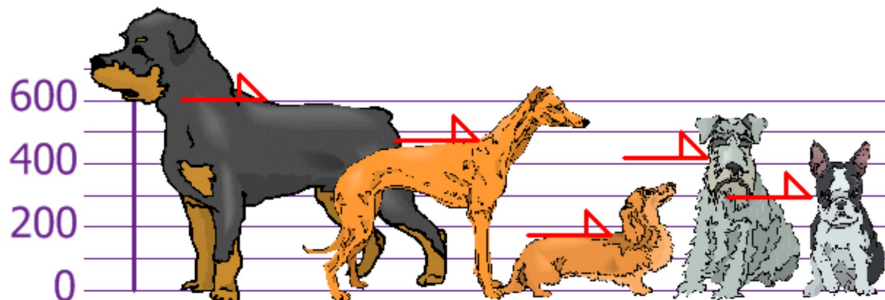
- JE DRUHÁ ODMOCNINA ROZPTYLU
- POSKYTUJE MĚŘÍTKO PRŮMĚRNÉ VZDÁLENOSTI KAŽDÉ HODNOTY OD PRŮMĚRU A POSKYTUJE PŘEHLED O ŠÍŘENÍ DAT.
- SMĚRODATNÁ ODCHYLKA UMOŽŇUJE PSYCHOLOGŮM POROZUMĚT ŠÍŘENÍ SKÓRE KOLEM PRŮMĚRU. POKUD JE NAPŘÍKLAD V TESTU IQ PRŮMĚR 100 A SMĚRODATNÁ ODCHYLKA 15, ZNAMENÁ TO, ŽE VĚTŠINA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

ROZPTYL VS. SMĚRODATNÁ ODCHYLKA



SMĚRODATNÁ ODCHYLKA MĚŘÍ, JAK DALEKO JSOU OD SEBE ČÍSLA V SADĚ DAT. NA DRUHÉ STRANĚ ROZPTYL UDÁVÁ SKUTEČNOU HODNOTU TOHO, JAK MOC SE ČÍSLA V SOUBORU DAT LIŠÍ OD PRŮMĚRU.

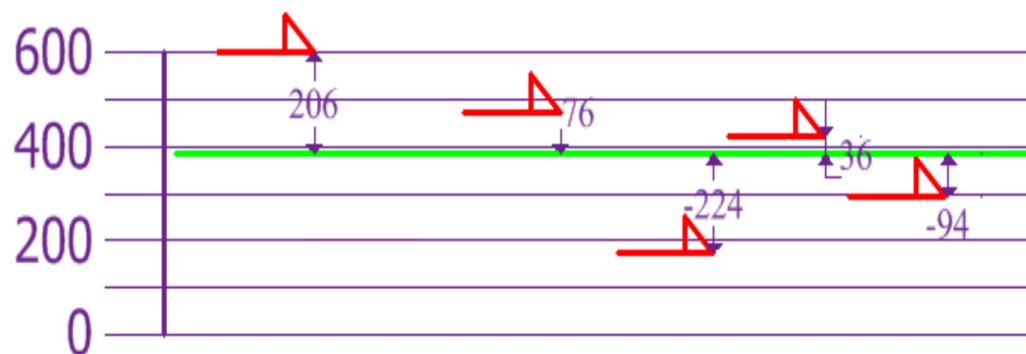
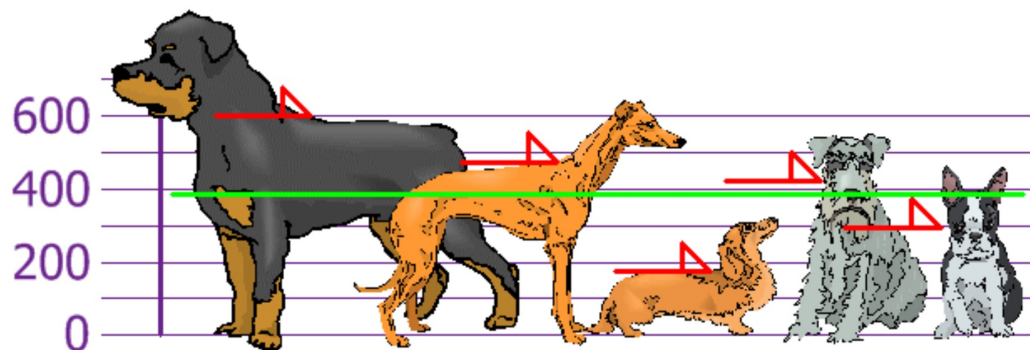
ROZPTYL VS. SMĚRODATNÁ ODCHYLKA



600mm, 470mm, 170mm, 430mm a 300mm

PRŮMĚR

$$\begin{aligned} &= (600 + 470 + 170 + 430 + 300) / 5 \\ &= 1970 / 5 \\ &= 394 \end{aligned}$$



Variance

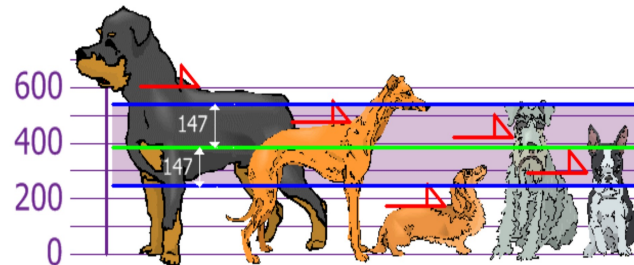
$$\begin{aligned}\sigma^2 &= \frac{206^2 + 76^2 + (-224)^2 + 36^2 + (-94)^2}{5} \\ &= \frac{42436 + 5776 + 50176 + 1296 + 8836}{5} \\ &= \frac{108520}{5} \\ &= 21704\end{aligned}$$

So the Variance is **21.704**

And the Standard Deviation is just the square root of Variance, so:

Standard Deviation

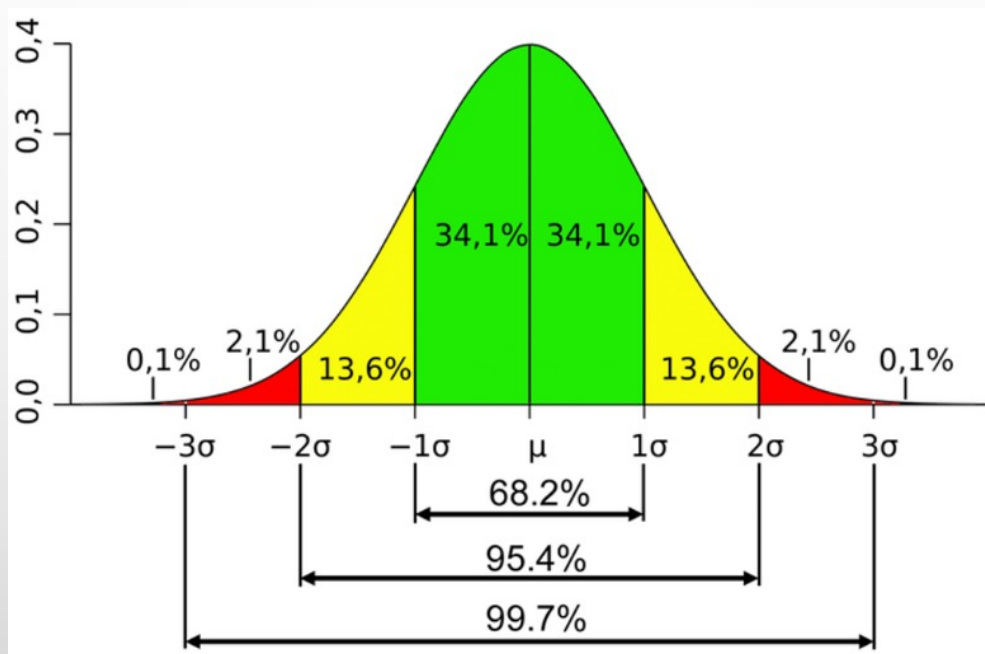
$$\begin{aligned}\sigma &= \sqrt{21704} \\ &= 147,32... \\ &= \mathbf{147} \text{ (to the nearest mm)}\end{aligned}$$



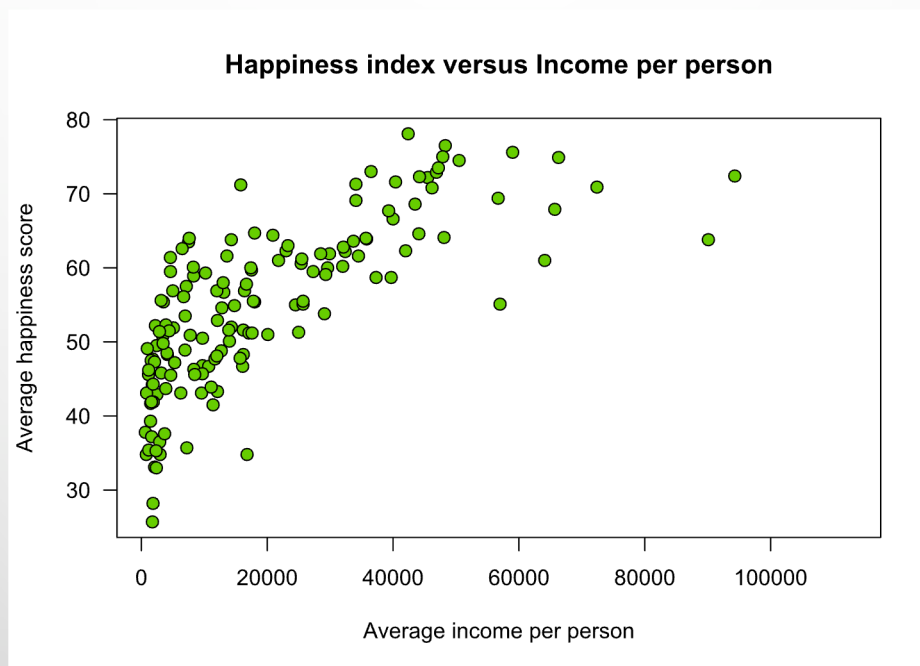
So, using the Standard Deviation we have a "standard" way of knowing what is normal, and what is extra large or extra small.

Rottweilers **are** tall dogs. And Dachshunds **are** a bit short, right?

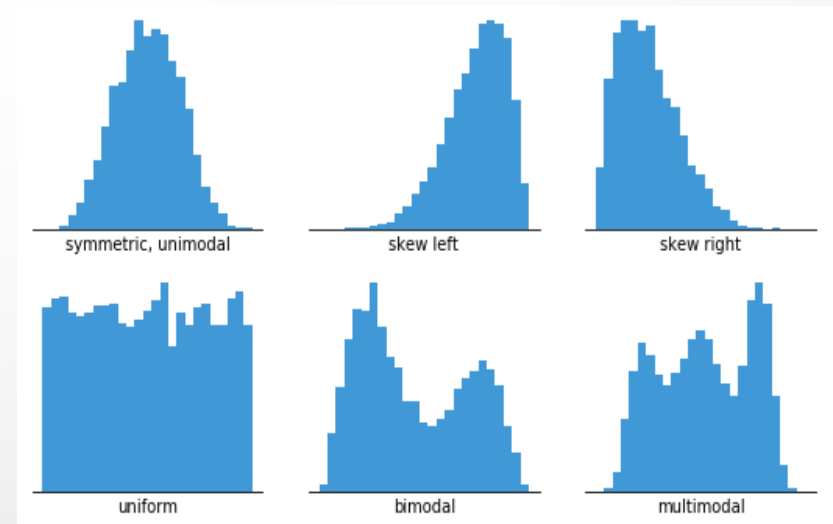
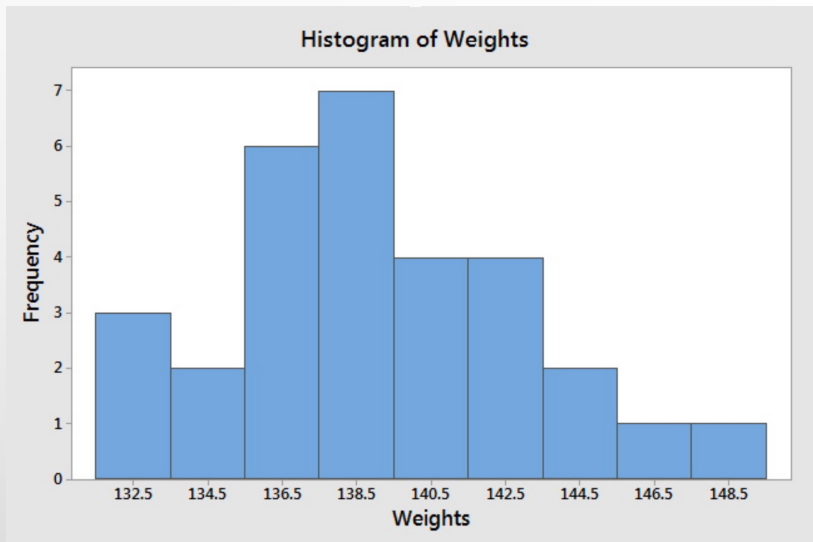
NORMÁLNÍ ROZDĚLENÍ (GAUSSOVA KŘIVKA)



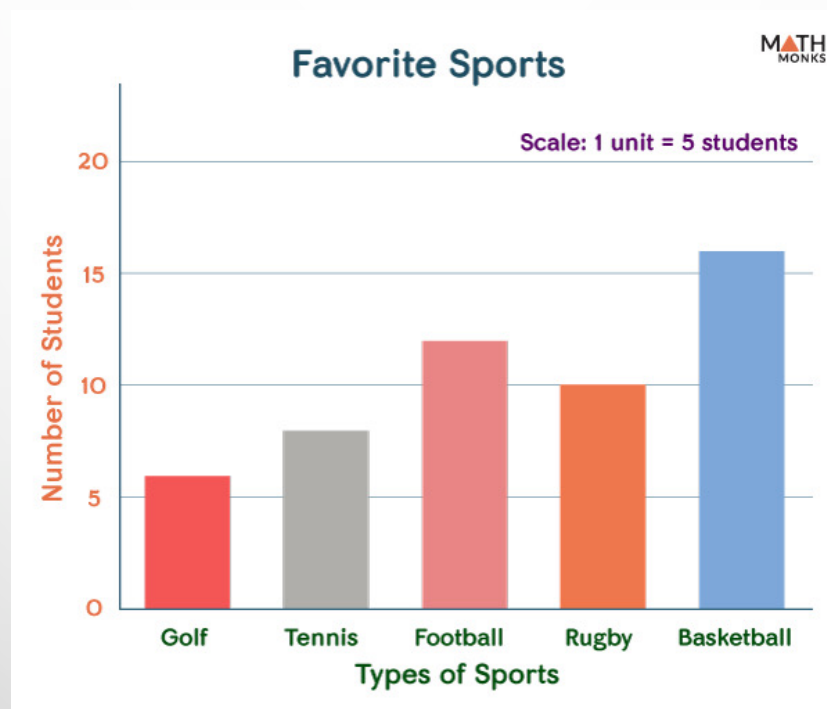
BODOVÝ GRAF



HISTOGRAM



SLOUPCOVÝ GRAF

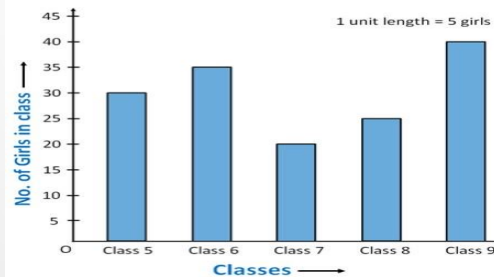


HISTOGRAM VS SLOUPCOVÝ GRAF

Difference between Bar Graph & Histogram

teachoo.com

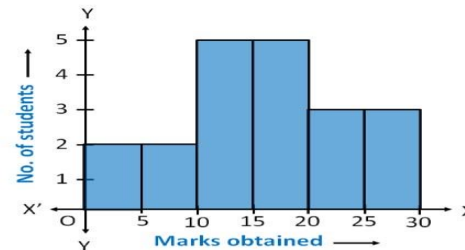
Bar Graph



In Bar Graph

- Bars have equal space
- On the y-axis, we have numbers & on the x-axis, we have data which can be anything.

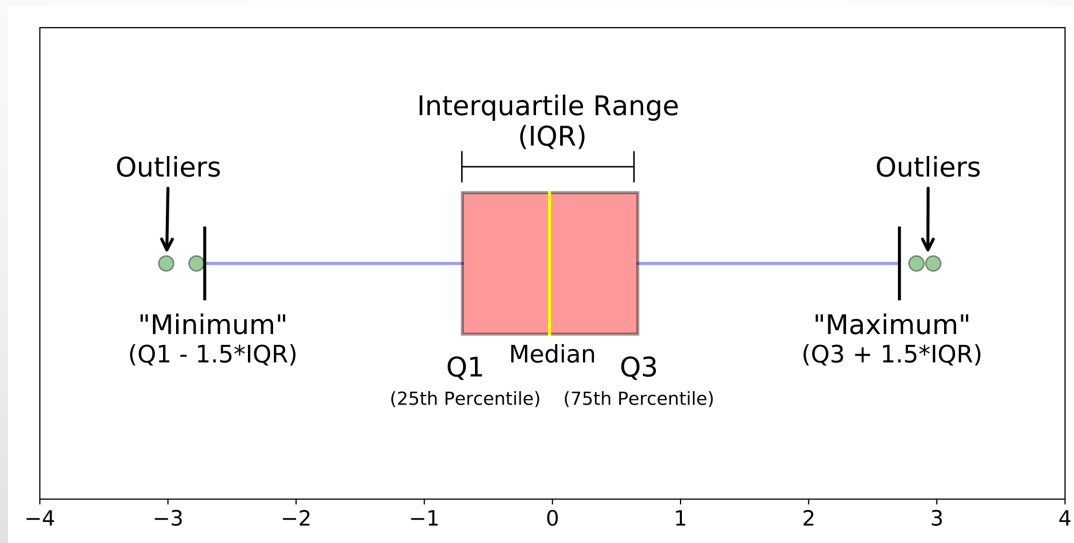
Histogram



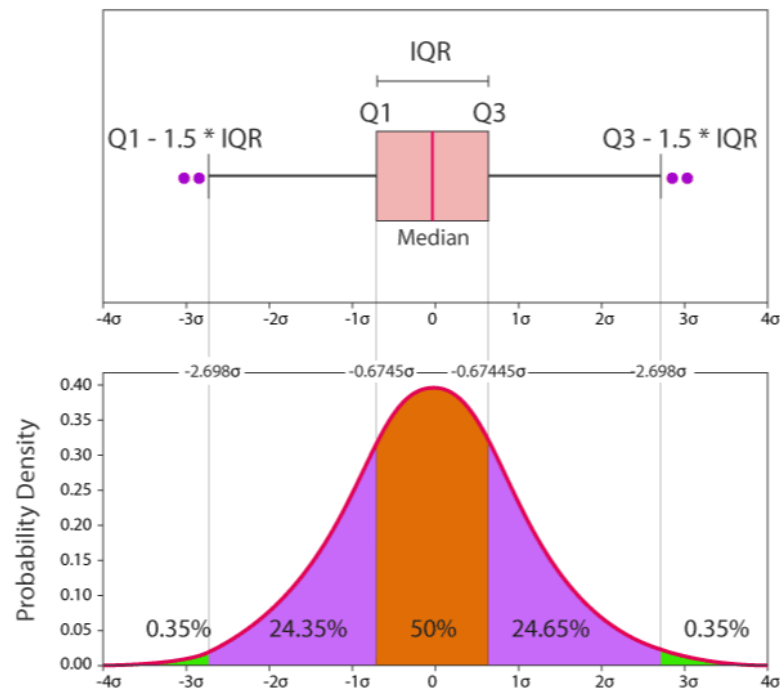
In Histogram

- Bars are fixed
- On the y-axis, we have numbers & on the x-axis, we have data which is continuous & will always be number.

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 ODLÉHLÉ 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 ODPOVĚDÍMI 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 ODLÉHLÉ 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
- 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) KTERÝ 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