



Statistika nuda ~~je~~ **NENÍ!!!**

MÁ VŠAK CENNÉ ÚDAJE!

Proč nemám rád citáty

*„Nevěřím žádné statistice,
kterou jsem sám nezfalšoval.“*

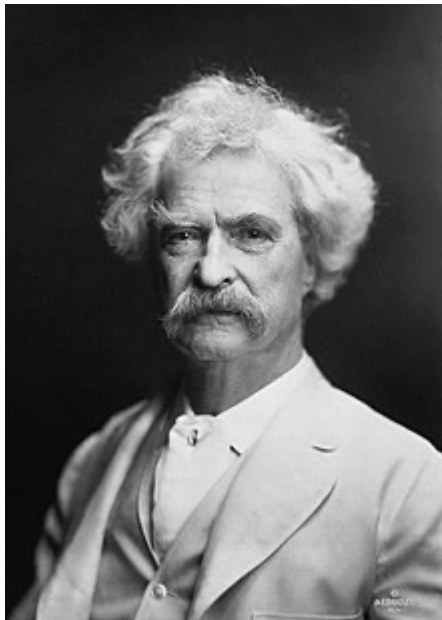
~~Winston Churchill~~

Joseph Goebbels



Proč nemám rád citáty 2

*„Jsou tři typy lží.
Lež, hnusná lež a statistika.“*



~~Mark Twain~~
~~Benjamin Disraeli~~

?



Když se řekne „statistika“

Sbírání čísel

Matematika

Numerická fakta

Složité vzorce

Sumarizace

Evidence

Zkoumání, zpracovávání a vyhodnocování dat

Vy nevíte co je statistika?



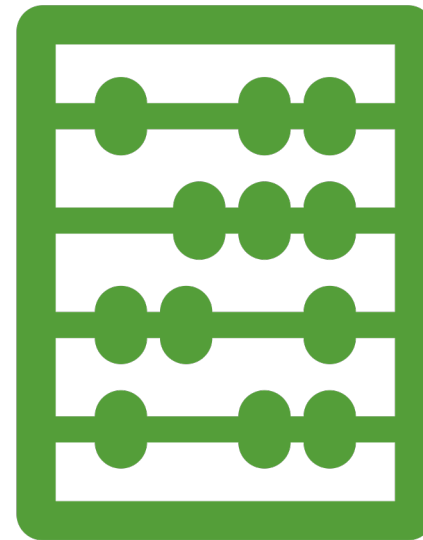
Statistika

- Z latinského ***status*** – stav
- V anglicky mluvících zemích také ***data science***
- Zabývá se studiem zákonitostí hromadných jevů

Jak pohlížet na statistiku



věda



nástroj

Dvě statistiky

statistika
POPISNÁ
(deskriptivní)



popis hromadných jevů

statistika
INFERENČNÍ
(induktivní)



zobecňování na základě
teorie pravděpodobnosti

Život mezi nulou a jedničkou

- Statistika je o **pravděpodobnosti**
 - Pravděpodobnost = šance, že nastane daný jev
- **Náhodný jev** – v závislosti na náhodě může, ale nemusí nastat
- **Náhoda** – souhrn drobných, nezjistitelných příčin
- Pravděpodobnost uskutečnění daného jevu je vyjádřena číslem od 0 do 1 (resp. 0 % až 100 %)

Definice pravděpodobnosti

KLASICKÁ DEFINICE:

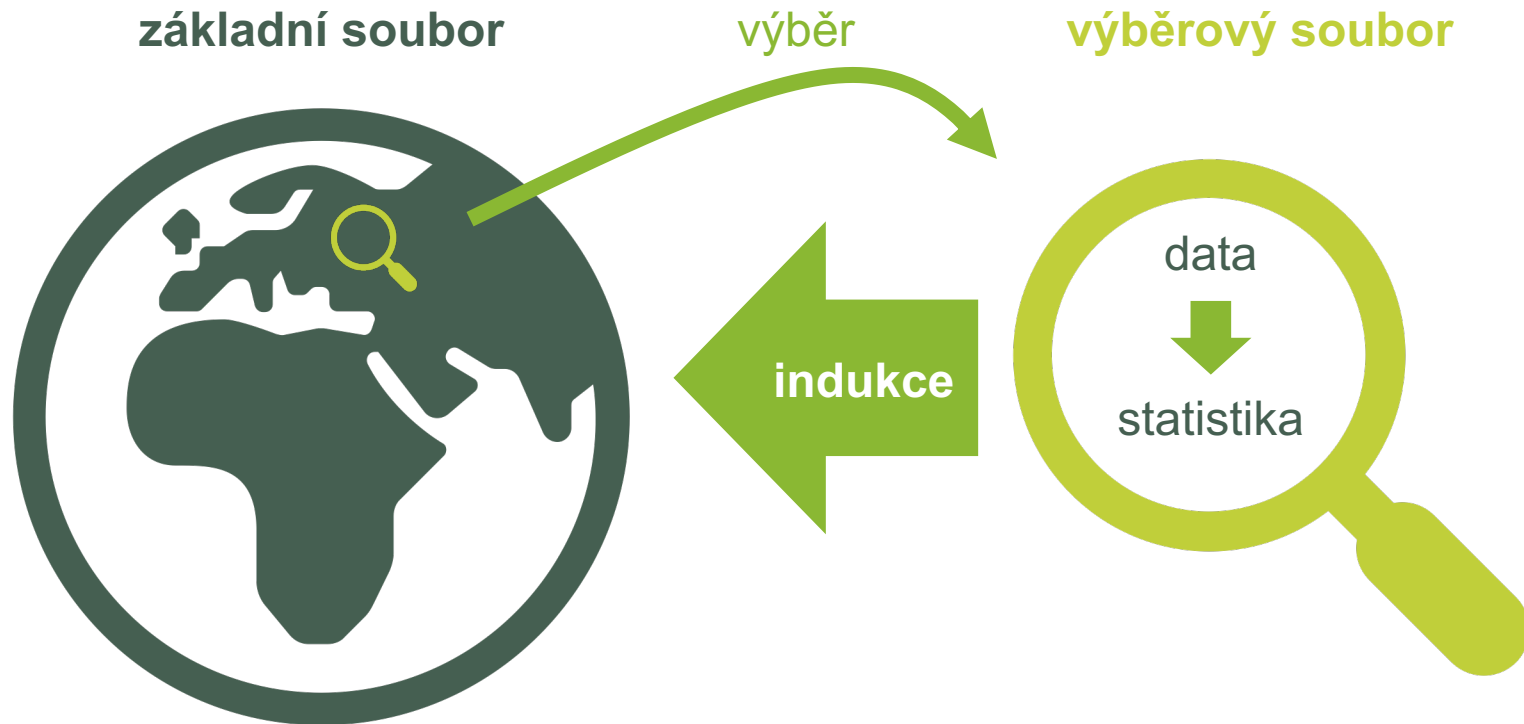
$$P(A) = \frac{m}{n}$$

STATISTICKÁ DEFINICE:

$$P(A) = \lim_{N \rightarrow \infty} \frac{M}{N}$$

m, M ... počet příznivých výsledků
 n, N ... celkový počet pokusů

Statistická indukce



Je statisticky dokázáno ...

- Za dobu, co se člověk naučil myslet, není zaznamenán případ, kdy by nevyšlo slunce.
- Pravděpodobnost východu slunce se tedy limitně blíží 100%.

Je statisticky dokázáno, že slunce vyjde každé ráno, a i když je tma jako v ranci, noc nemá celkem žádnou šanci.

Je statisticky dokázáno ...

Když drak si z nosu síru pouští
a Honza na něj číhá v houšti,
pak statistika předpovídá,
že nestvůra už neposnídá.

- V minulosti, kdykoliv byl drak konfrontován s českým Honzou, odcházel ze střetnutí téměř vždy s hlavami sklopenými (pokud mu nějaké zbyly).
- Pravděpodobnost, že drak souboj vyhraje, se tedy limitně blíží 0%.

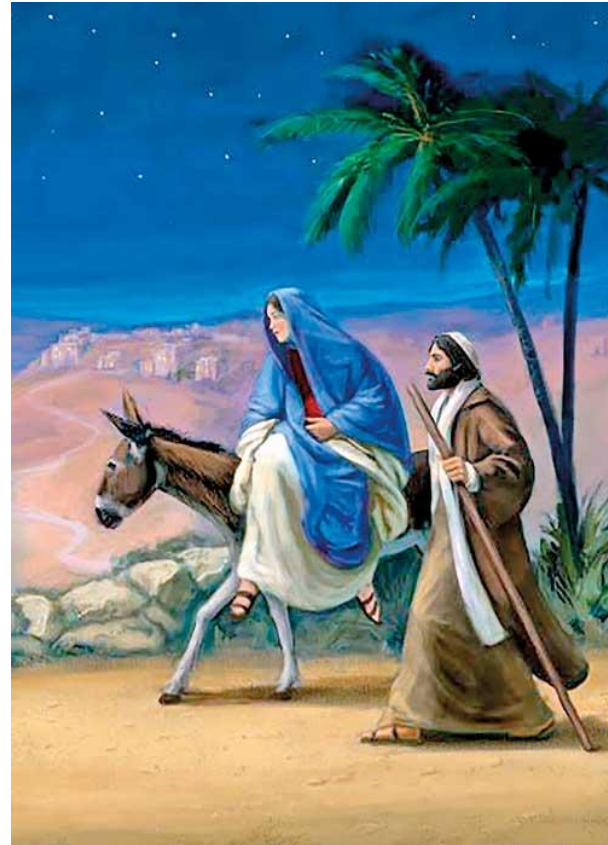
Dějiny statistiky

- Potřeba zaznamenávat hromadná data
- 3800 př. n. l. – sčítání lidu v Babylonii
- Starověký Egypt – soupis dobytka každý druhý rok
- Dále např. Řecko, Čína, Indie
- Starověký Řím – propracovaný cenzus
 - Cenzus = proces získávání údajů o jednotlivých příslušnících obyvatelstva
 - Cenzor – speciální úředník určený k takové činnosti

Statistika v Bibli

„Stalo se v oněch dnech, že vyšlo nařízení od císaře Augusta, aby byl po celém světě proveden soupis lidu. Tento první majetkový soupis se konal, když Sýrii spravoval Quirinius. Všichni se šli dát zapsat, každý do svého města. Také Josef se vydal z Galileje, z města Nazareta, do Judska, do města Davidova, které se nazývá Betlém, poněvadž byl z domu a rodu Davidova, aby se dal zapsat s Marií, která mu byla zasnoubena a čekala dítě. Když tam byli, naplnily se dny a přišla její hodina.“

Lukáš 2, 1-6



Dějiny statistiky

- 13. stol. – první církevní matriky
- 17. stol. – vznik teorie pravděpodobnosti
- 18.-19. stol. – rozvoj statistického myšlení
 - Pierre Simon de Laplace (1749-1827)
 - Carl Friedrich Gauss (1777-1855)
- 19. stol. – vznik matematické statistiky
 - Karl Pearson (1857-1936)

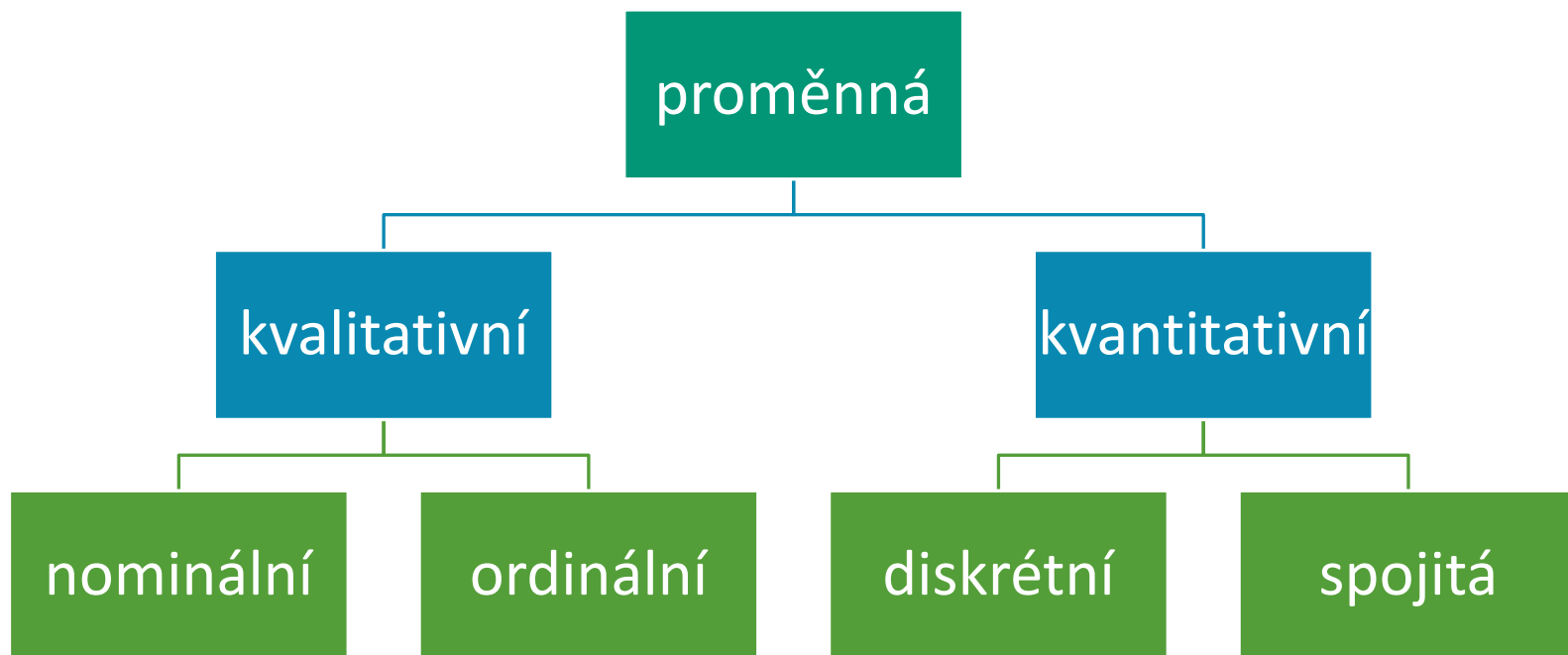
Statistika v českých zemích

- 1753 – patent o každoročním sčítání lidu
- 1787 – první statistické svazky
 - Josef Antonín rytíř Riegger (1742-1795)
- 1897 – Zemský statistický úřad
- 1919 – Státní úřad statistický
- 1969 – Český statistický úřad

Základní statistické pojmy

- Typy proměnných
- Míry polohy
- Míry variability

Typy proměnných



Průměr (aritmetický)

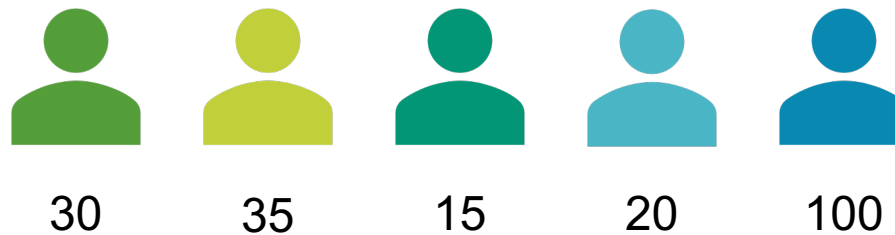
- Vyjadřuje typickou hodnotu popisující soubor

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Medián

- Hodnota, jež dělí řadu na dvě stejné poloviny

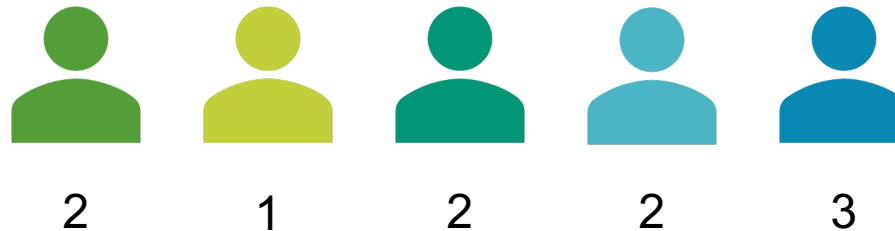
Průměrný plat ... 40 tisíc



Medián ... 30 tisíc

Modus

- Nejčastěji se vyskytující hodnota



Směrodatná odchylka

- Jak se od sebe liší jednotlivé případy



2

2

2

Průměr = 2
Směrodatná odchylka = 0



1

2

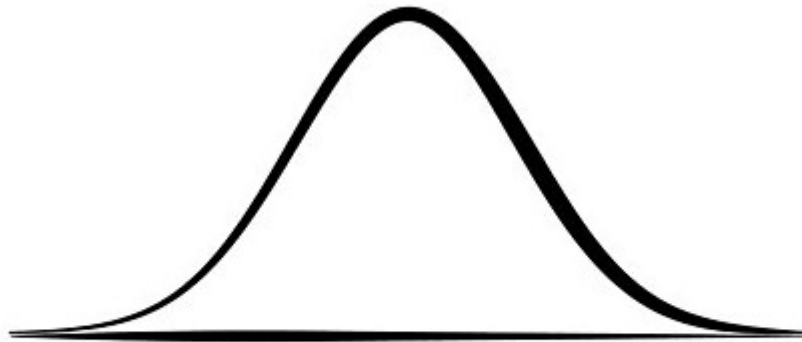
3

Průměr = 2
Směrodatná odchylka = 1

Pozor na chyby

- Průměr vs. medián
- Podíl vs. poměr
- Procento vs. procentní bod
- Meziroční vs. meziměsíční

Rozdělení pravděpodobnosti



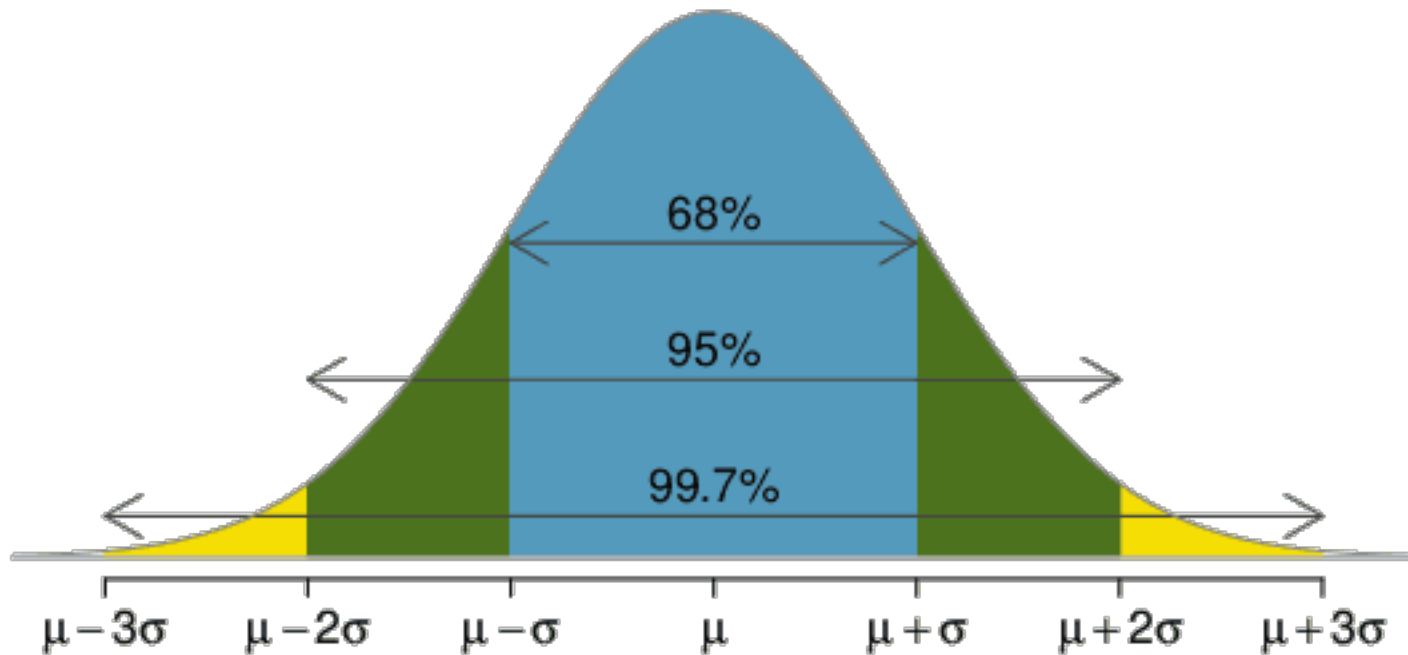
Normální rozdělení



Paranormální rozdělení

Normální rozdělení

- Název normální ve smyslu „řídící se předpisem“



Statistické testování

- Posouzení, zda experimentálně získaná data vyhovují danému předpokladu
- Postup:
 1. Nulová a alternativní hypotéza
 2. Hladina významnosti
 3. Volba vhodného testu
 4. Výpočet testového kritéria
 5. Vymezení kritického oboru
 6. Rozhodnutí
 7. Interpretace

Statistické testování – příklad

Chemický výzkum má za cíl zjistit, zda mají použítá hnojiva vliv na produkci obilí. Na základě údajů v následující tabulce zjistěte, zda existuje průkazný rozdíl produkce obilí na těch polích, kde bylo použito hnojivo A, a na polích, kde bylo použito hnojivo B.

	Produkce obilí									
hnojivo A	492	914	594	487	496	793	598	614	704	722
hnojivo B	506	507	634	425	419	625	427	325	549	494

Statistické testování – příklad

- Průměr hnojivo A ... 641,4
- Průměr hnojivo B ... 491,1

- Dvouvýběrový t-test:
$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

- Testové kritérium ... 2,779
- p-hodnota ... 1,2 %

Senioři v ČR (2021)

- Počet obyvatel nad 65 – 2,17 milionu (20,6 %)
- 57,9 % žen
- Osoby pobírající starobní důchod – 2,38 milionu
- Průměrná výše starobního důchodu – 15 453 Kč

Senioři v ČR (2021)

- Stěhování – 7,5 z 1 000 seniorů
- Sňatky
 - 1 žena z 1 000 nevdaných
 - 5 mužů z 1 000 neženatých
- Rozvody
 - 0,8 žen z 1 000 vdaných
 - 1,2 mužů z 1 000 ženatých



Děkuji za pozornost!