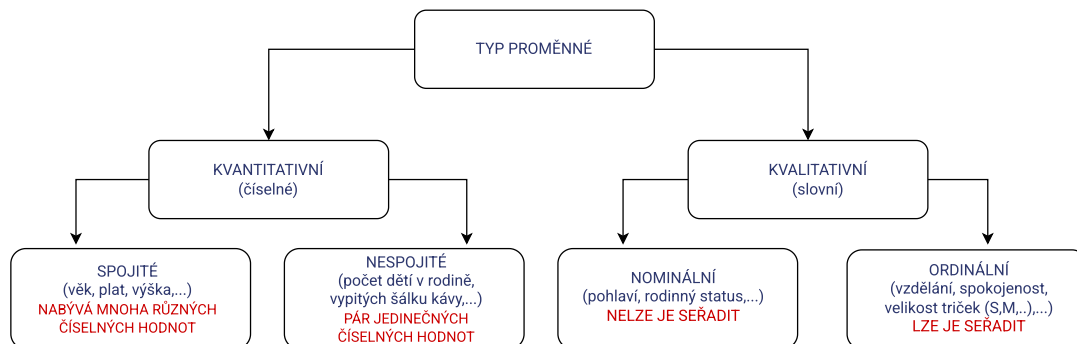


STATISTIKA V KOSTCE

POPISNÁ STATISTIKA

TYPY PROMĚNNÝCH



ČETNOSTI

Absolutní četnost (n_i) nám říká kolik máme v souboru námi sledovaných znaků.

Relativní četnosti (p_i) nám dá stejnou informaci jako absolutní četnost, akorát relativně, nebo-li procentuálně.

Kumulativní četnosti - postupné (na)sčítání četností hodnot stat. znaku, hodnoty by měly být seřazeny vzestupně.

CHARAKTERISTIKY POLOHY A VARIABILITY

POLOHA (ÚROVEŇ):

Průměr – průměrná hodnota v datech (citlivý na odlehlé hodnoty)

Medián – prostřední hodnota, 50% hodnot menší, 50% hodnot vyšší

1. kvartil (dolní) – 25% hodnot menší, 75% vyšší

3. kvartil (horní) – 75% hodnot menší, 25% vyšší

Modus – nejčastěji se vyskytující hodnota

x_i	n_i	p_i	kumulativní četnosti	
			absolutní	relativní
1	4	0,2	4	0,2
2	5	0,25	9	0,45
3	6	0,3	15	0,75
4	5	0,25	20	1
celkem	20	1	—	—

VARIABILITA:

ABSOLUTNÍ:

Rozpětí – vzdálenost mezi minimem a maximem

Kvartilové rozpětí – vzdálenost mezi dolním a horním kvartilem

Rozptyl – střední hodnota kvadrátů odchylek od střední hodnoty

Směrodatná odchylka – odchylka od střední hodnoty, která má jednotky stejné jako zkoumaná proměnná (odmocnina rozptylu)

MAD – mediánová absolutní odchylka. Udává typickou odchylku od mediánu.

RELATIVNÍ:

Variační koeficient – směrodatná odchylka dělená průměrem. K porovnání variability.

VLASTNOSTI PRŮMĚRU, SMĚRODATNÉ ODCHYLKY, ROZPTYLU, VAR. KOEFICIENTU

PŘÍČTENÍ/ODEČTENÍ KONSTANTY (např. 5)

Průměr: změní se o konstantu (průměr: +5), Směrodatná odchylka, rozptyl: **nemění se**

Var. koeficient: **změní se** (protože se změní průměr, nutnost spočítat).

NÁSOBENÍ / DĚLENÍ KONSTANTOU (např. 5)

Průměr: *5, směrodatná odchylka: *5, rozptyl: *5²

Var. koeficient: **nemění se**

ZVÝŠENÍ / SNÍŽENÍ O PROCENTA (např. ± 15 %)

Průměr: * 1,15 (nebo *0,85), směrodatná odchylka: *1,15 (nebo *0,85), rozptyl: *1,15² (nebo *0,85²)

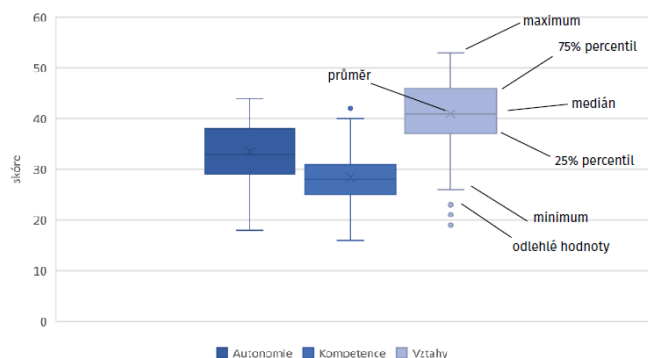
Var. koeficient: **nemění se**

STATISTIKA V KOSTCE

GRAFY PRO ČÍSELNÉ PROMĚNNÉ

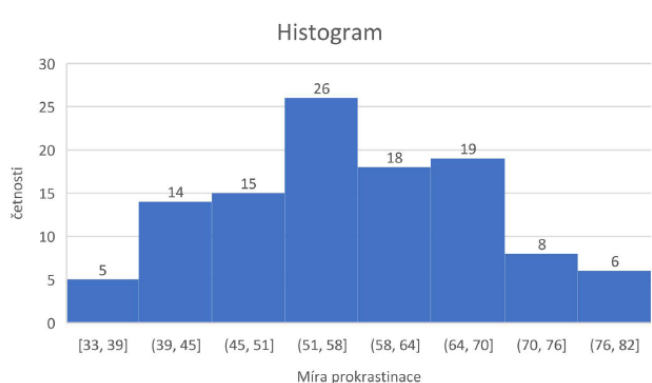
KRABICOVÝ GRAF (BOX-PLOT)

Graf vhodný pro kvantitativní data, umožňuje porovnávat více souborů mezi sebou a identifikovat odlehlá pozorování.



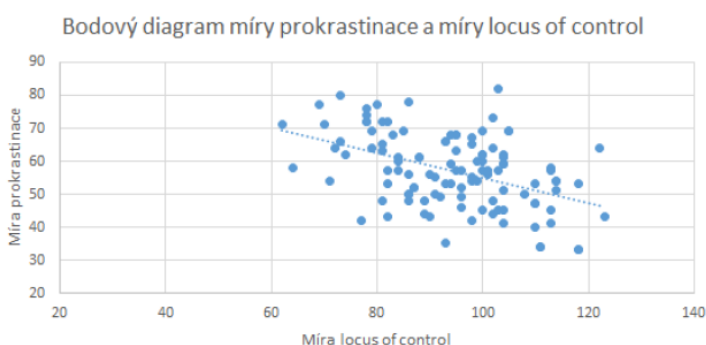
HISTOGRAM

- Graf vhodný pro kvantitativní data.
- Sloupce jsou "nalepené" na sobě a vyjadřují stejnou šířku intervalu.
- Počet intervalů je daný dle Sturgesova pravidla $k = 1 + 3,32 * \log(n)$



BODOVÝ GRAF

- Graf vhodný pro vyjádření vztahu mezi dvěma veličinami (např. vztah mezi cenou a počtem najetých km). Jedna veličina je na svislé (Y), druhá na vodorovné ose (X).
- Neplést si bodový a spojnicový graf. U spojnicového grafu jsou na vodorovné ose zobrazeny seskupení dat.

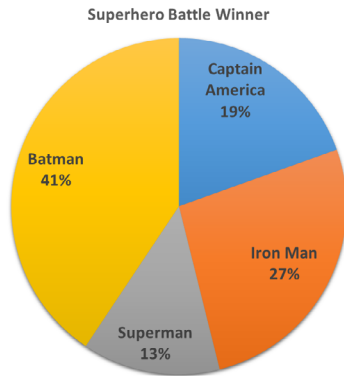


STATISTIKA V KOSTCE

GRAFY PRO SLOVNÍ PROMĚNNÉ

VÝSEČOVÝ (KOLÁČOVÝ)

znázornění kvalitativních proměnných. Ideálně s nižším počtem obměn. Jedná se o vůbec nejzákladnější graf, s jeho použitím to nepřehánějte.



SLOUPCOVÝ

- Porovnání, která kategorie je nejčastější nebo nejméně častá.
- Kategorie ordinální proměnné je vhodné seřadit.



DALŠÍ TYPY GRAFŮ:

- Pruhový graf
- Skupinový (clustered) sloupcový graf (2 slovní)
- Skládáný sloupcový graf (2 slovní)
- Mozaikový graf (2 slovní)

DEJ POZOR NA:

Při tvorbě grafů nezapomeňte na název grafu, popisky os, dle typu dat je vhodné uvést jednotky a samozřejmě neuškodí legenda, či popisek grafu.

Žádná manipulace! Například změnou měřítka osy nebo volbou velikosti rozestupů mezi značkami se mohou stejná data jevit úplně jinak. Pozor tedy na manipulování, ať už vědomé či nevědomé.



STATISTIKA V KOSTCE

PRAVDĚPODOBNOST

ZÁKLADNÍ POJMY

Jev - událost, která může nastat (A, B, ...).

Pravděpodobnost jevu:

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

m – počet příznivých výsledků, n – počet všech možných výsledků.

Doplňek jevu:

$$P(\bar{A}) = 1 - P(A)$$

Vhodný k výpočtu opačné pravděpodobnosti, např. pr. že alespoň 1 výrobek je vadný. Spočítám pr. pro 0 (žádný) a odečtu od 1.

SLUČITELNOST A NESLUČITELNOST

Slučitelné jevy – mohou nastat současně. Vzorec: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ **Sjednocení jevů**

Neslučitelné jevy – nemohou nastat současně, průnik je 0. Vzorec: $P(A \cup B) = P(A) + P(B)$ **Sjednocení jevů**

NEZÁVISLÉ A ZÁVISLÉ JEvy

Nezávislé jevy – nastání jednoho neovlivní druhý. Vzorec: $P(A \cap B) = P(A) P(B)$ **Průnik jevů**

Závislé jevy – nastání jednoho ovlivní druhý. Vzorec: $P(A \cap B) \neq P(A) P(B)$

PODMÍNĚNÁ PRAVDĚPODOBNOST

Pravděpodobnost jevu A za předpokladu, že nastal jev B.

$$P(A | B) = \frac{P(A \cap B) - \text{průnik jevů}}{P(B) - \text{podmínka}}$$

POMOCNÁ PRAVIDLA PRO VÝPOČET

- **NEBO** -> Sčítám (sjednocení jevů)
- **A / A ZÁROVEŇ** -> Násobím (průnik jevů)
- **ZA PŘEDPOKLADU / POKUD NASTAL** -> Podmíněná pravděpodobnost

ZÁKLADNÍ KONTROLA VÝSLEDKŮ

- Pravděpodobnost musí být vždy v intervalu **0 až 1**.
- Pokud výsledek vychází mimo interval 0 až 1 -> výpočet/úvaha je špatně.
- U sjednocení jevů: výsledek nesmí být větší než 1.
- U průniku jevů: výsledek nesmí být větší než menší z pravděpodobností P(A), P(B).



STATISTIKA V KOSTCE

NESPOJITÉ (DISKRÉTNÍ) NV

CO JE DISKRÉTNÍ NÁHODNÁ VELIČINA

- Náhodná veličina, která může nabývat jen **určitých samostatných hodnot** (obvykle 0,1,2,3,...).
- **Typické příklady:** počet dětí, počet chyb, počet úspěchů v sérii pokusů, počet reklamací, počet příchozích hovorů za minutu.

ZÁKLADNÍ POJMY

- **$P(x)$** – pravděpodobnost, že X nabude hodnoty x .
- **$F(x)$** – distribuční funkce: pravděpodobnost, že X je $\leq x$.
- **Střední hodnota ($E[x]$)** – dlouhodobě očekávaná hodnota (průměr)
- **Rozptyl, směrodatná odchylka** – míra variability hodnot kolem průměru.

ALTERNATIVNÍ (BERNOULLIHO) ROZDĚLENÍ

Modeluje jediný pokus, ve kterém může nastat $\rightarrow$ úspěch (1) nebo neúspěch (0).

Excel: pravděpodobnost přesně x : **=BINOM.DIST(1; 1; p; NEPRAVDA)**

BINOMICKÉ ROZDĚLENÍ

Modeluje počet úspěchů v **n nezávislých pokusech**, kde každý má **stejnou pravděpodobnost úspěchu π** .

Typické příklady: Počet šestek při 20 hodech kostkou, počet vadných výrobků z várky 100 ks, kde pr. vady je např. 0,2. Házím mincí 10x, jaká je pravděpodobnost, že padne 4x panna?

Excel:

přesně x úspěchů: **=BINOM.DIST(x; n; π ; NEPRAVDA)**

max x úspěchů (kumulativně): **=BINOM.DIST(x; n; π ; PRAVDA)**

alespoň 5 úspěchů **=1-BINOM.DIST(4; n; π ; PRAVDA)**, počítám pr. do 4 a výsledek odečtu od 1.



STATISTIKA V KOSTCE

SPOJITÉ NV

CO JE SPOJITÁ NÁHODNÁ VELIČINA

- Může nabývat libovolných hodnot v intervalu (např. výška, čas, teplota).
- Pravděpodobnost nastání konkrétní jediné hodnoty je **0**, počítá se jen pravděpodobnost intervalu.

ZÁKLADNÍ POJMY

- **Hustota pravděpodobnosti $f(x)$** – popisuje „jak často“ se hodnoty vyskytují.
- **$F(x)$** – distribuční funkce: pravděpodobnost, že **X je $\leq x$** .
- **Střední hodnota ($E[x]$)** – dlouhodobě očekávaná hodnota (průměr)
- **Rozptyl, směrodatná odchylka** – míra variability hodnot kolem průměru.

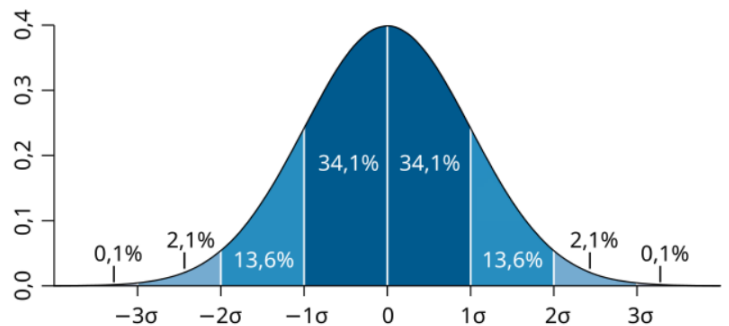
NORMOVANÉ NORMÁLNÍ ROZDĚLENÍ ($U \sim N(0,1)$)

- Základní „standardní“ normální rozdělení se střední hodnotou 0 a směrodatnou odchylkou 1.
- Slouží jako referenční rozdělení.
- Tabulky a funkce pracují právě s tímto rozdělením.

Excel:

distribuce: **=NORM.S.DIST(u; PRAVDA)**

kvantil: **=NORM.S.INV(p)**



NORMÁLNÍ ROZDĚLENÍ ($X \sim N(\mu, \sigma)$)

Klasické „Gaussovo“ rozdělení, zvonová křivka.
Většina hodnot je v intervalu $\mu \pm 1\sigma$
skoro všechny v $\mu \pm 3\sigma$

Typické příklady: Výška studentů má $\mu = 175$ cm, $\sigma = 7$ cm. Jaká je pravděpodobnost, že student měří méně než 180 cm?

Excel:

náhodná veličina je menší než x : **=NORM.DIST(x ; μ ; σ ; PRAVDA)**

náhodná veličina leží v intervalu x_1 až x_2 : **=NORM.DIST(x_2 ; μ ; σ ; PRAVDA) - NORM.DIST(x_1 ; μ ; σ ; PRAVDA)**

náhodná veličina je větší než x : **= 1 - NORM.DIST(x ; μ ; σ ; PRAVDA)**

nalezení kvantilu: **=NORM.INV(p ; μ ; σ)**

DALŠÍ SPOJITÁ ROZDĚLENÍ

Chí-kvadrát rozdělení: Používá se pro testy variability, Chí-kvadrát test dobré shody a nezávislosti.

Kvantil: **=CHISQ.INV(p ; v)**

Studentovo t-rozdělení: Používá se pro malé vzorky ($n \leq 30$) při testování (t-testy), test o korelaci, regresních koeficientech,... Kvantil: **=T.INV(p ; v)**

F-rozdělení: Používá se při porovnání dvou rozptylů nebo při analýze rozptylu (ANOVA), F-testu.

Kvantil: **=F.INV(p ; v_1 ; v_2)**

