

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0$$

$\leq \frac{\sigma^2}{n \cdot \varepsilon^2}$

PF: Měřená hmotnost má chybu $\sim N(0, \sigma^2)$, $\sigma = 1g$ (předpokládáme 2. měření)

Chceme $\bar{X}_n$ s chybou $< 0,1g$ s jistotou 95%

→ Moc polární věc

Postup 1: Použití sluhého známého velkého čísel:

Chceme $\varepsilon = 0,1, \sigma = 1$, n t.j. $\frac{\sigma^2}{n \cdot \varepsilon^2} < 0,05 \rightarrow n > 2000$ ✓

Postup 2: Vlastnosti normálního rozdělení:

$$X_1 + \dots + X_n \sim N(0, n \cdot \sigma^2), \text{ tedy } \bar{X}_n \sim N(0, \frac{\sigma^2}{n})$$

množina n hodnot dělíme n-krát

$$P(|\bar{X}_n - 0| < 2 \sqrt{\frac{\sigma^2}{n}}) = 0,95$$

$< 0,1$

$$n > 20^2 = 400$$

→ Těhle díky normálnímu rozdělení

Proč je 1. větší? Protože první postup je obecný a platí pro všechna rozdělení!

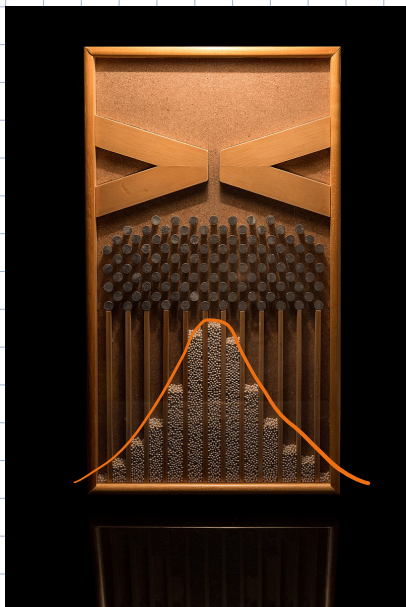
Thm: $X_1, X_2, \dots$ n.n.v., všechny stejné rozdělení, $E X_i = \mu$, neplynou σ^2

$$Y_n = \frac{(X_1 + \dots + X_n) - n \cdot \mu}{\sigma \sqrt{n}} \quad \text{pak} \quad Y_n \xrightarrow{d} N(0,1),$$

což znamená: $\lim_{n \rightarrow \infty} F_{Y_n}(t) = \Phi(t) = \lim_{n \rightarrow \infty} P(Y_n \leq t) = \Phi(t)$

$$Y_n = \text{norm}(X_1 \dots X_n), \quad \text{kde normalizace je } \text{norm}(s) = \frac{s - ES}{\sqrt{\text{Var}(s)}}$$

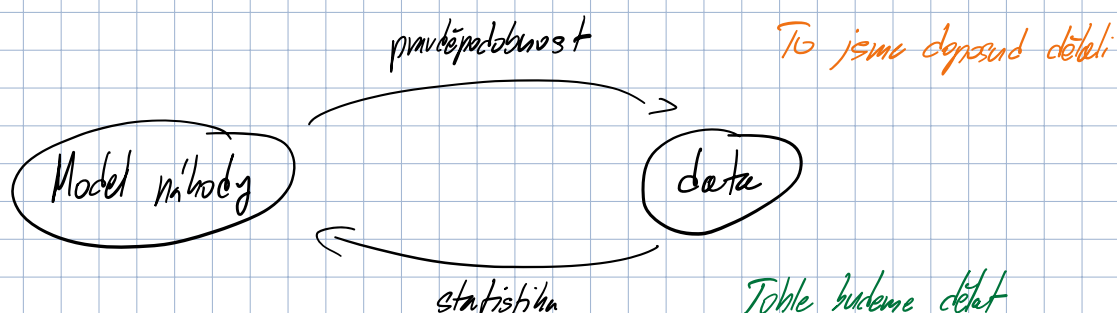
$$\text{norm}(s) \quad \text{má} \quad E = 0 \quad \text{var} = 1$$



Y_{10} popisuje polohu zvrhů na lince

$$P(Y_{10} < t) = \Phi(t)$$

$$P(a < Y_n < b) = \Phi(b) - \Phi(a)$$



Populace ... obvyklé ČR (SZ)

Vzorok ... podmnožina populace

zájem nás je jaké vlastnosti
té populace a její výskyt.

zároveň nás zajímají ty vlastnosti,
kteřé se neobjeví všechny projt výzkum

popisná statistika ... explorativní analýza

mat. statistika ... konfirmativní analýza

inferenční statistika

Problémy:

- nejasná otázka
- volba vzorku
- neúplná data
- outliers

→ chceme reprezentativní vzorek, tedy aby náš vzorek byl = část populace.

Tedy podle zákona o velkých číslech musí být uniformně náhodné.

IQR - interquartile range : $Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)$

Jaké úlohy statistika řeší?

- bodové úlohy ... odhad nějakého čísla (průměr něčeho v populaci etc.)
- intervalové odhady ... interval (a, b) , který obsahuje průměr s velkou pravděpodobností
- testování hypotéz ...

klauzura vs. ústní

- přirozený postup ... mib. výběr bez opak.
- mat. heur. postup ... -11 s opakováním