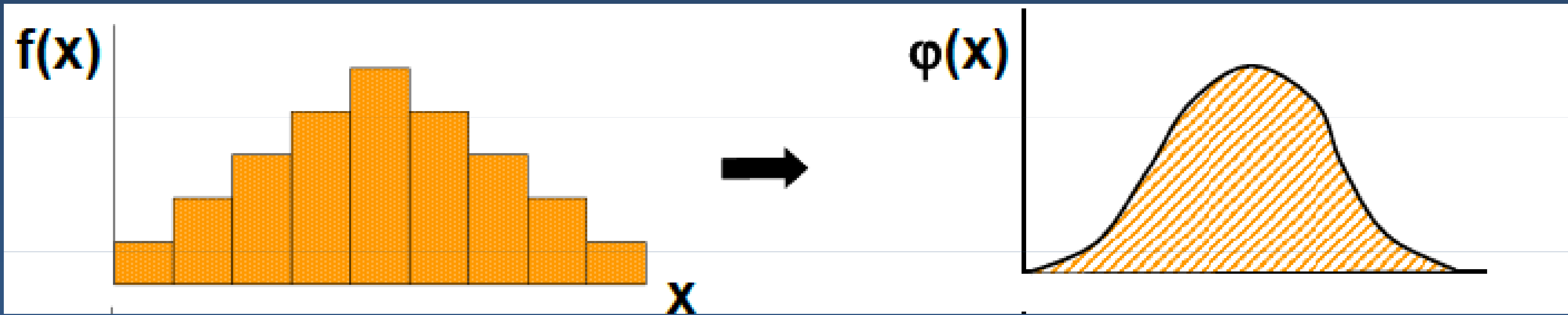
7. Odhady populačních průměrů a pravděpodobností

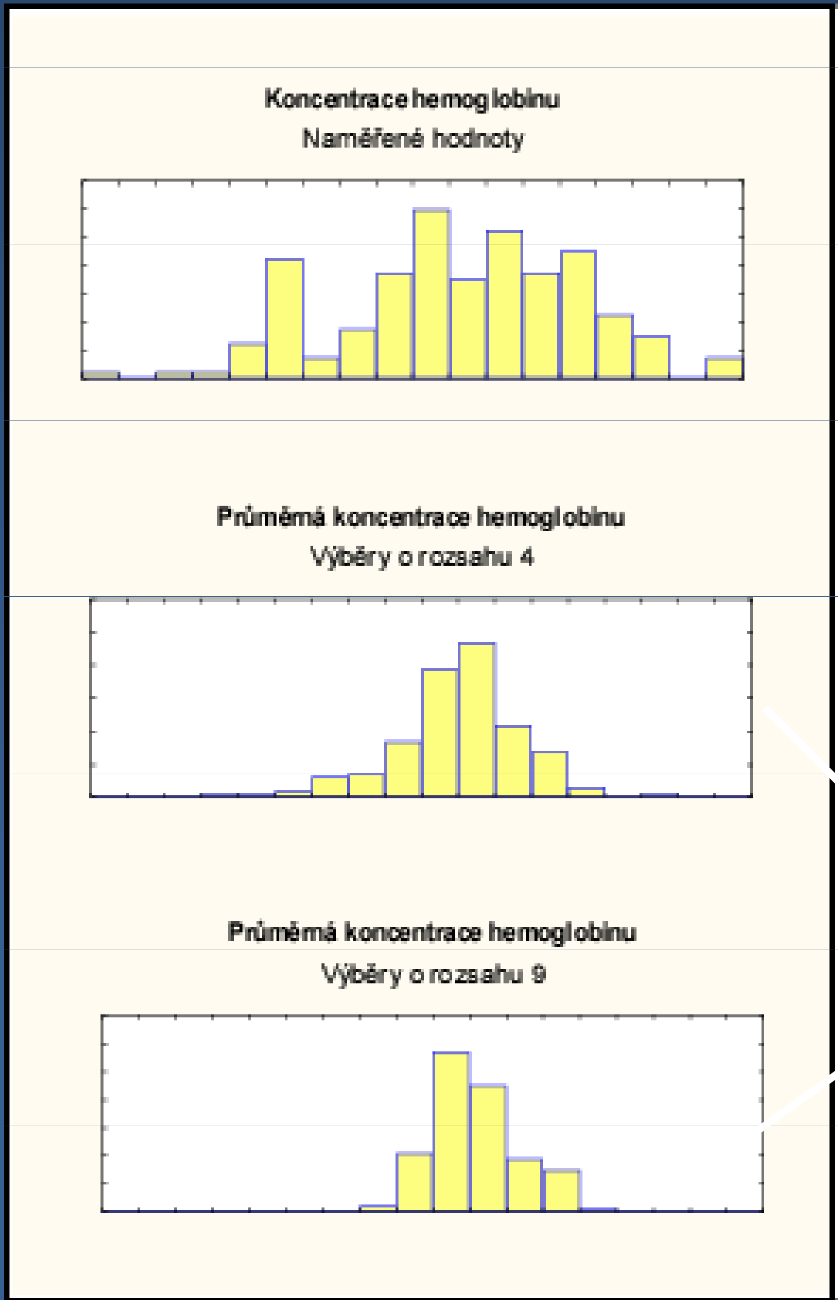
Výběrové a teoretické rozložení



Výběrové rozložení
(N, $\bar{x}$, s)

Teoretické rozložení
(μ, σ)

Motivace



1. všechny tři histogramy kolísají kolem stejného středu
2. čím větší rozsah výběru, tím užší rozdělení
3. rozdělení průměrů pro n = 4 a n = 9 jsou podobnější normálnímu rozdělení než rozdělení původních dat.

Rozdělení výběrového průměru

Odhady

Bodové
(číslo)

Intervalové
(interval pravděpodobných hodnot)

Výběrový průměr $\bar{x}$ je bodovým odhadem parametru μ

Výběrová směrodatná odchylka s je bodovým odhadem parametru σ

Relativní četnost p je bodovým odhadem parametru π

Intervalový odhad průměru

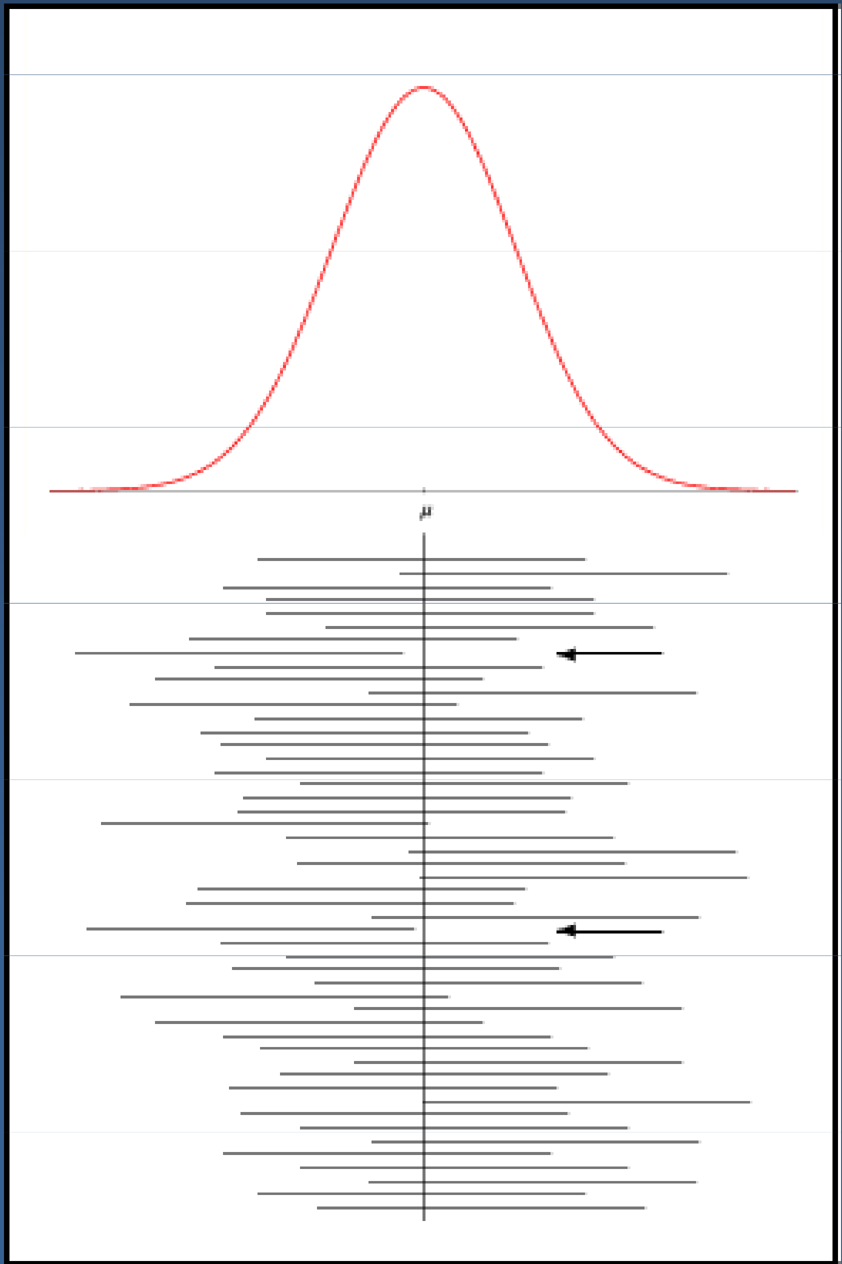
- $X_i \sim N(\mu,\sigma^2) \Rightarrow \bar{X} \sim N(\mu,\sigma^2/n)$
- Provedeme výběr o rozsahu n a vypočteme $\bar{x}$, pak μ leží s pstí 0,95 v intervalu:

$$\bar{x} \pm 1{,}96 \cdot \sigma / \sqrt{n}$$ Interval spolehlivosti pro průměr

- z = 1,96 kritická hodnota standardizovaného normálního rozložení pro koeficient spolehlivosti P = 0,95

$\bar{x} - 1{,}96 \cdot \sigma / \sqrt{n}$ $\bar{x}$ $\bar{x} + 1{,}96 \cdot \sigma / \sqrt{n}$

Simulace



50 výběrů o rozsahu 10
Pouze dva nepokrývají populační průměr

Obecný vzorec pro interval spolehlivosti

$$P(L_1 < Odhad < L_2) \geq 1 - \alpha / 2$$

Odhadovaný parametr ± Kvantil modelového rozložení (pro 1-α/2) x SE (odhadu)

α	0,1	0,05	0,01	0,001
$z_{1-\alpha/2}$	1,645	1,960	2,576	3,290
$z_{1-\alpha}$	1,282	1,645	2,326	3,090
z_{α}	-1,282	-1,645	-2,326	-3,090

Intervalový odhad průměru při neznámé směrodatné odchylce v populaci

- Neznáme směrodatnou odchylku v populaci σ => nahradíme výběrovou směrodatnou odchylkou *s*
- Další nejistota
- Místo kvantilů standardizovaného normálního rozdělení použijeme kvantily studentova rozdělení pro příslušný počet stupňů volnosti (*n*-1)

$\bar{x} \pm t \frac{s}{\sqrt{n}}; t = t_{1-\alpha/2}(n-1); \frac{s}{\sqrt{n}}$ - standardní chyba průměru

Studentovo t-rozdělení

- Podobné standardizovanému normálnímu rozdělení
- Symetrické kolem střední hodnoty $\mu = 0$
- Má pouze 1 parametr:
- Stupně volnosti: $\nu = n-1$

Intervalový odhad populační pravděpodobnosti

- Výběr o rozsahu *n*, danou vlastnost má *r*
- Relativní četnost výskytu vlastnosti ve výběru *p = r/n*
- Pro $n\pi(1-\pi) \geq 9$ má relativní četnost výskytu vlastnosti normální rozdělení
- Průměr = pst výskytu v celé populaci (*π*)
- Standardní chyba = $\sqrt{\pi(1-\pi)/n}$
- 95% interval spolehlivosti pro populační pst

$$p \pm 1{,}96\sqrt{\frac{\pi(1-\pi)}{n}}$$