

Intervaly spolehlivosti

B02907 Informační a komunikační technologie



Lubomír Štěpánek,
Ústav biofyziky a informatiky
1. LF UK



Princip statistických šetření

- statistická pozorování omezená počtem prvků výběru
- statistika se snaží z hodnot znaku na malém výběru usuzovat na typické hodnoty znaku v celé populaci
 - výběr je množina, která náleží množině populace, a je podstatně méně mohutná
- užívají se odhady velikosti hodnoty znaku (střední hodnoty – průměru, četnosti, pravděpodobnosti) v populaci
- odhad je zatížen „náhodou“

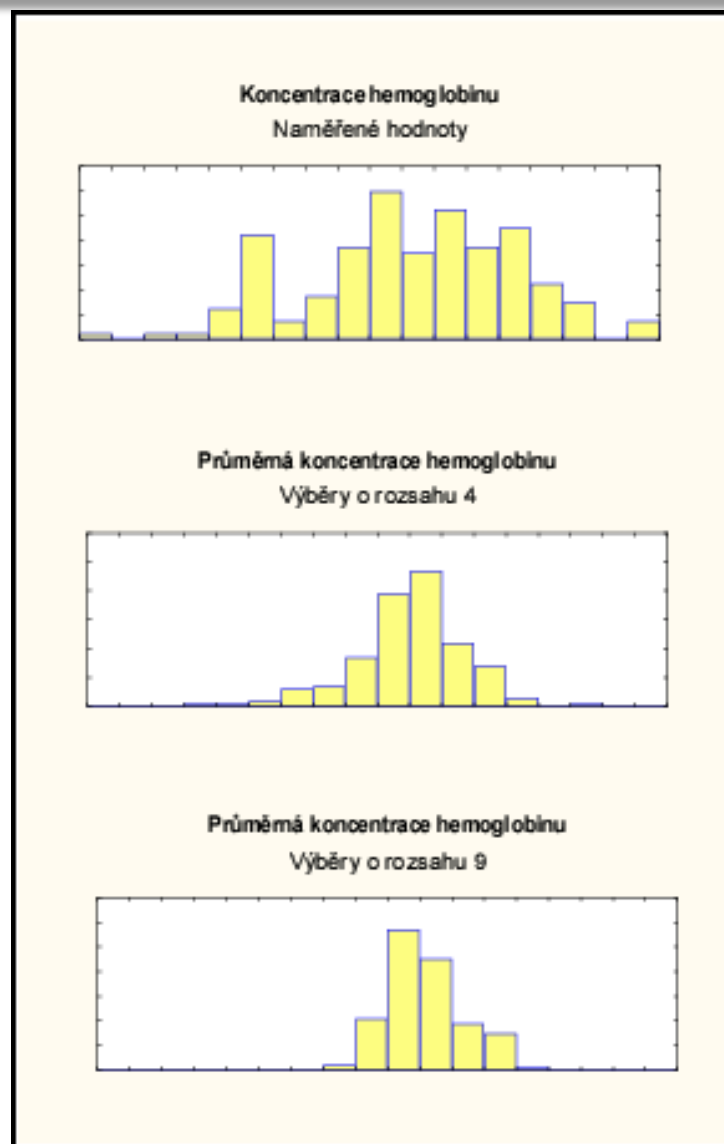
Odhady

- *bodový odhad*
 - předpokládáme, že charakteristická hodnota výběru (průměr, četnost) odpovídá populační hodnotě
 - populační hodnota se pokládá rovna dané charakteristické hodnotě výběru
 - např. „*je-li četnost hypertoniků mezi dvaceti náhodnými pacienty 7, je i četnost hypertoniků v populaci 35 % (=7/20)*“
- *intervalový odhad (interval spolehlivosti = konfidenční interval)*
 - interval, ve kterém leží charakteristická hodnota populace s určitou pravděpodobností (spolehlivostí)
 - sestavuje se pomocí vztahu mezi výběrovými a populačními hodnotami
 - např. „*je-li v souboru sto lidí průměrná výška 175 cm a odchylka 10 cm, pak průměrná výška populace leží s 95 % pravděpodobností v intervalu 173 – 177 cm*“

Výběrový vs. populační ukazatel

výběr	populace
odhad	parametr
relativní četnost	pravděpodobnost
výběrový průměr	populační průměr
výběrová odchylka	populační odchylka

Rozdělení výběrového průměru

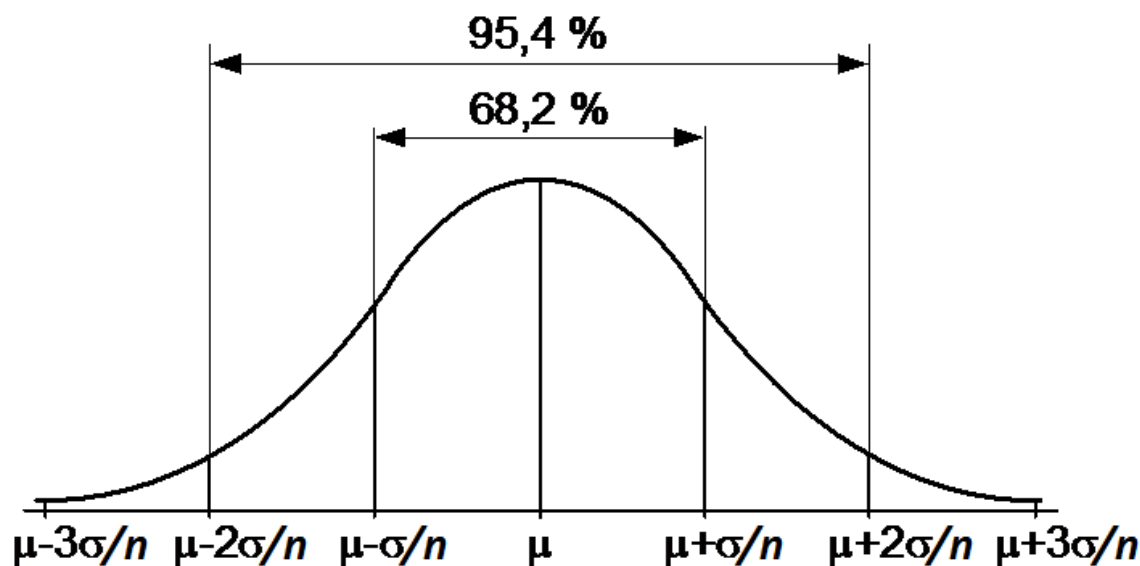


Rozdělení výběrového průměru

- má-li původní rozdělení průměr μ a směrodatnou odchylku σ , má rozdělení průměrů výběrů (o n prvcích) průměr rovněž μ a směrodatnou odchylku σ/\sqrt{n}
- rozdělení průměrů je nezávisle na původním rozdělení normální
- máme-li tedy jeden výběr o rozsahu n , jenž má nějaký průměr x a odchylku σ , lze podle rozdělení výběrových průměrů odhadnout populační průměr μ

Rozdělení výběrového průměru

- interval $(x - \sigma/\sqrt{n}; x + \sigma/\sqrt{n})$ zahrnuje průměr s 68% pravděpodobností
- interval $(x - 2\sigma/\sqrt{n}; x + 2\sigma/\sqrt{n})$ zahrnuje průměr s 95% pravděpodobností



Odhad průměru se zvolenou hladinou spolehlivosti

- obecně máme-li výběr s průměrem \bar{x} a směrodatnou odchylkou σ daného znaku, pak populační průměr μ daného znaku leží s $100(1-\alpha) \%$ pravděpodobností v intervalu

$$\mu \in \left(\bar{x} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}; \bar{x} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

- $z_{1-\alpha/2}$ je kvantil standardního normálního rozdělení, je tabelizován
 - není-li možné použít směrodatnou odchylku výběru jako populační, je nutné místo kvantilů normálního rozdělení použít kvantily Studentova t rozdělení (opět tabelizováno) a zároveň použít pouhý odhad populační odchylky

Odhad průměru se zvolenou hladinou spolehlivosti

- obecně máme-li výběr s průměrem \bar{x} a směrodatnou odchylkou s daného znaku, pak populační průměr μ daného znaku leží s $100(1-\alpha) \%$ pravděpodobností v intervalu

$$\mu \in \left(\bar{x} - t_{1-\alpha/2} \cdot \frac{s}{\sqrt{n-1}} ; \bar{x} + t_{1-\alpha/2} \cdot \frac{s}{\sqrt{n-1}} \right)$$

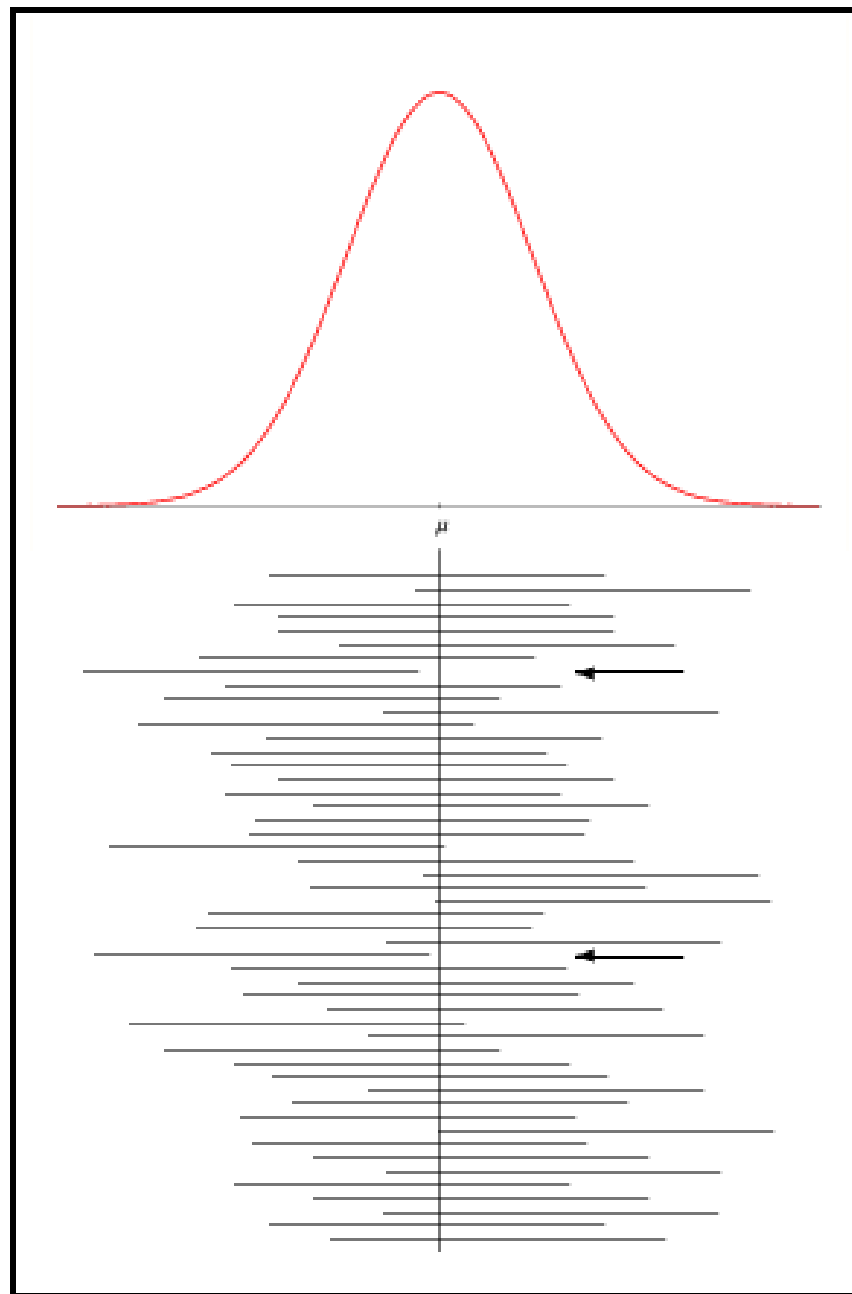
- $t_{1-\alpha/2}$ je kvantil Studentova t -rozdělení, je tabelizován
 - není-li možné použít směrodatnou odchylku výběru jako populační, je nutné místo kvantilů normálního rozdělení použít kvantily Studentova t rozdělení (opět tabelizováno) a zároveň použít pouhý odhad populační odchylky

Postup odhadování průměru

- 1. nasbíráme výběrový soubor o n prvcích
- 2. spočítáme jeho průměr x ; pokud neznáme populační směrodatnou odchylku σ , odhadneme ji (získáme s) – viz dále
- 3. spočítáme interval průměru μ a směrodatnou odchylku σ/\sqrt{n} souboru výběrových průměrů (tzv. výběrovou chybu průměru), výběrový průměr x by měl v intervalu kolem μ ležet, tedy i μ by měl ležet v intervalu kolem x)
- 4. teď už můžeme odhadnout populační průměr μ pomocí x : populační průměr μ náleží s 95 % pravděpodobností intervalu $(x-2\sigma/\sqrt{n}; x+2\sigma/\sqrt{n})$

Postup odhadování průměru

soubor	hodnota průměru	hodnota směrodatné odchylky
populace	μ (neznáme, ale chceme odhadnout)	σ (neznáme, ale odhadneme záhy z výběru)
výběr z populace o n prvcích (ten si sestavíme)	x (lehce spočítáme)	s (lehce spočítáme, může být shodná s populační)
soubor výběrů o n prvcích z populace (slouží spíš jako model)	μ náležící intervalu ($x+2\sigma/\sqrt{n}$; $x-2\sigma/\sqrt{n}$)	σ/\sqrt{n} ; s/\sqrt{n} (spočítáme, protože známe σ nebo s i n výběru)
populace (odhadnutá)	μ náležící intervalu ($x+2\sigma/\sqrt{n}$; $x-2\sigma/\sqrt{n}$) (již jsme odhadli z výběrových hodnot)	σ či s (už známe)



Intermezzo

- v souboru sto lidí je průměrná výška 175 cm a odchylka 10 cm, jaká je s 95 % pravděpodobností průměrná výška populace?

Intermezzo

- v souboru sto lidí je průměrná výška 175 cm a odchylka 10 cm, jaká je s 95 % pravděpodobností průměrná výška populace?

$$\sigma/\sqrt{n} = 10 \text{ cm} / \sqrt{100} = 1 \text{ cm}$$

$$\mu \in (\bar{x} - 2 \cdot \sigma/\sqrt{n}; \bar{x} + 2 \cdot \sigma/\sqrt{n}) = (175 - 2; 175 + 2) \text{ cm}$$

$$\mu \in (173 \text{ cm}; 177 \text{ cm})$$

- podobné úvahy jako pro odhadování průměru
 - v populaci má daná hodnota znaku relativní četnost π a směrodatnou odchylku $\sqrt{\pi(1-\pi)}$
 - vyberme z ní výběry o n prvcích a určíme rozdělení relativní četnosti dané hodnoty znaku – bude vykazovat normální rozdělení s průměrem rovným populační relativní četnosti π a směrodatnou odchylkou $\sqrt{\pi(1-\pi)/n}$

Postup odhadování populační relativní četnosti

- 1. nasbíráme výběrový soubor o n prvcích
- 2. spočítáme relativní četnost p zkoumané hodnoty znaku a směrodatnou odchylku $\sqrt{\pi(1-\pi)}$
- (3. soubor z výběrů o n prvcích, v nichž počítáme relativní četnosti dané hodnoty znaku, má normální rozdělení s intervalem průměru kolem π a směrodatnou odchylkou $\sqrt{\pi(1-\pi)/n}$, p by měla ležet v intervalu kolem π , takže π i by mělo ležet ve stejném intervalu kolem p)
- 4. teď už můžeme odhadnout populační relativní četnost π pomocí p : populační relativní četnost π náleží s 95 % pravděpodobností intervalu $(p-2\sqrt{(\pi(1-\pi)/n)}; p+2\sqrt{(\pi(1-\pi)/n)})$

Shrnutí

- máme-li výběr s průměrem \bar{x} a směrodatnou odchylkou σ daného znaku, pak populační průměr μ daného znaku leží s 95 % pravděpodobností v intervalu

$$\mu \in \left(\bar{x} - 2 \frac{\sigma}{\sqrt{n}}; \bar{x} + 2 \frac{\sigma}{\sqrt{n}} \right)$$

- máme-li výběr s relativní četností p daného znaku, pak populační relativní četnost π daného znaku leží s 95 % pravděpodobností v intervalu

$$\pi \in \left(p - 2 \sqrt{\frac{p(1-p)}{n}}; p + 2 \sqrt{\frac{p(1-p)}{n}} \right)$$