

Odhady

- odhady
 - bodové a intervalové odhady
 - konstrukce intervalu spolehlivosti pro průměr
 - odhady podílů (kategoriální proměnné)
-

Rozdělení výběrových průměrů

- **tvar** – RVP se při dostatečně velkém vzorku (30 a více) blíží **normálnímu rozdělení**
 - **průměr** – průměr průměrů všech teoretických výběrů je roven průměru populace
 - **variabilita** – směrodatná odchylka RVP se označuje jako **výběrová chyba** (standard error) průměru
-

Rozdělení výběrových průměrů

- velikost výběrové chyby je dána dvěma charakteristikami:
 - **variabilita znaku v populaci**: čím je vyšší, tím je vyšší i variabilita výběrových průměrů
 - **velikost výběru** – čím větší výběr (n), tím lépe jeho průměr reprezentuje průměr populace
-

Rozdělení výběrových průměrů

- vzorec pro výpočet výběrové chyby:

$$\sigma_x = \sigma / \sqrt{n}$$

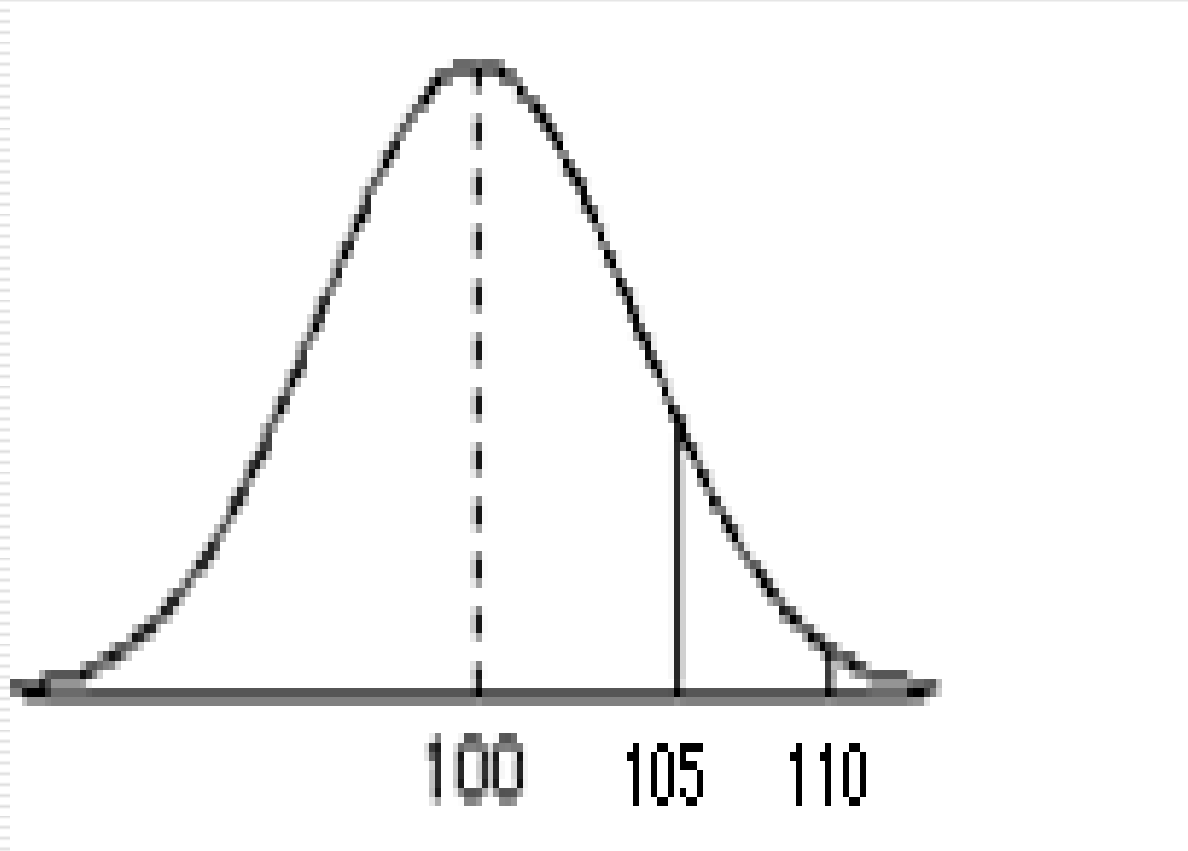
Rozdělení výběrových průměrů

- **příklad:** když vybereme z populace náhodně vzorek 9 osob, jaká je pravděpodobnost, že jejich průměrné IQ bude větší nebo rovno 105?
-

Rozdělení výběrových průměrů

- **příklad:** když vybereme z populace náhodně vzorek 9 osob, jaká je pravděpodobnost, že jejich průměrné IQ bude větší nebo rovno 105?
 - $\mu = 100, \sigma = 15, n = 9$
 - výpočet **výběrové chyby:**
 $\sigma_x = \sigma / \sqrt{n} = 15 / 3 = 5$
-

Rozdělení výběrových průměrů



Rozdělení výběrových průměrů

- **příklad:** když vybereme z populace náhodně vzorek 9 osob, jaká je pravděpodobnost, že jejich průměrné IQ bude větší nebo rovno 105?
 - $\mu = 100, \sigma_x = 5$
 - $z = (105-100)/\sigma_x = 5/5 = \mathbf{1.0}$
-

Rozdělení výběrových průměrů

- **příklad:** když vybereme z populace náhodně vzorek 9 osob, jaká je pravděpodobnost, že jejich průměrné IQ bude větší nebo rovno 112?
 - $\mu = 100, \sigma_x = \sigma/\sqrt{n} = 15/3 = 5$
 - **$z = 1.0$**
 - z tabulky **$P(Z \geq 1.0) = 0.1587$**
-

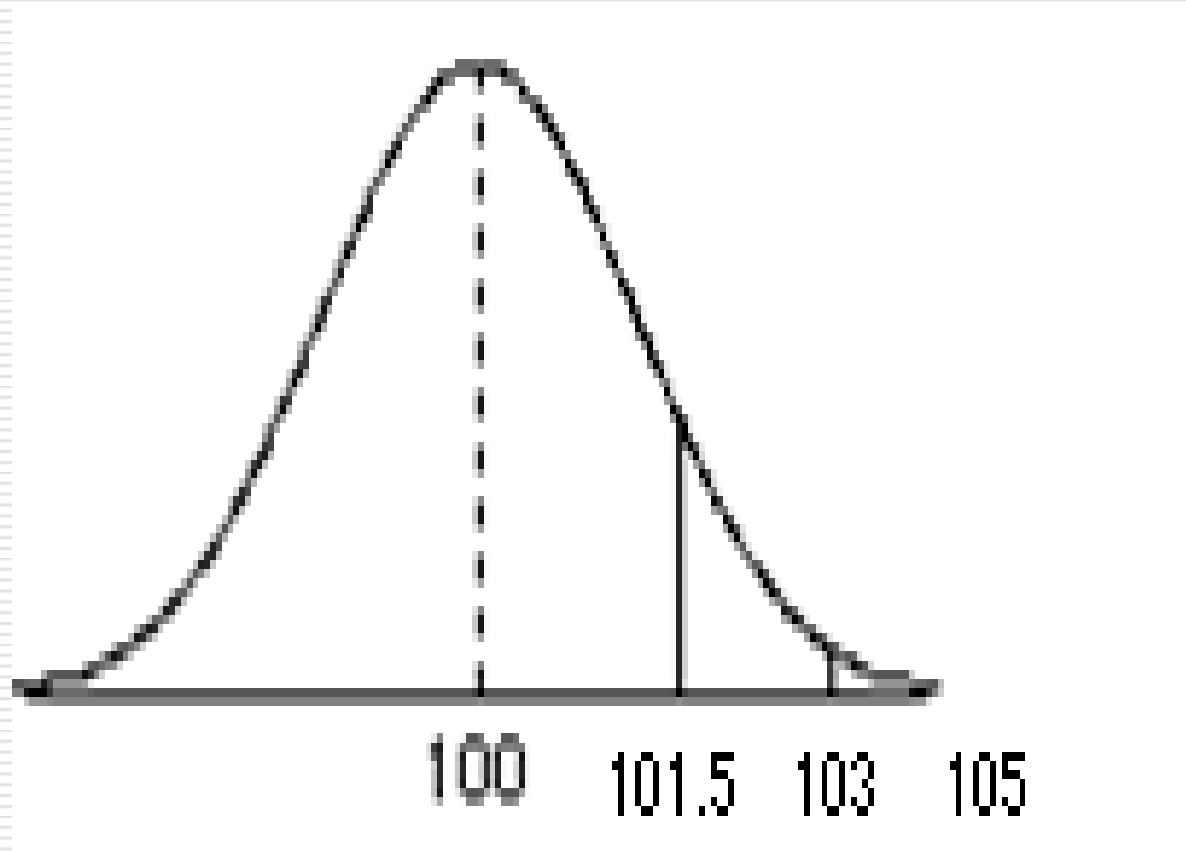
Rozdělení výběrových průměrů

- **příklad:** když vybereme z populace náhodně vzorek 100 osob, jaká je pravděpodobnost, že jejich průměrné IQ bude větší nebo rovno 105?
-

Rozdělení výběrových průměrů

- **příklad:** když vybereme z populace náhodně vzorek 100 osob, jaká je pravděpodobnost, že jejich průměrné IQ bude větší nebo rovno 105?
 - $\mu = 100, \sigma = 15, n = 100$
 - výpočet **výběrové chyby:**
 $\sigma_x = \sigma / \sqrt{n} = 15 / 10 = 1.5$
-

Rozdělení výběrových průměrů



Rozdělení výběrových průměrů

- **příklad:** když vybereme z populace náhodně vzorek 100 osob, jaká je pravděpodobnost, že jejich průměrné IQ bude větší nebo rovno 105?
 - $\mu = 100, \sigma_x = 1.5$
 - $z = (105-100)/\sigma_x = 5/1.5 = \mathbf{3.33}$
-

Rozdělení výběrových průměrů

- **příklad:** když vybereme z populace náhodně vzorek 100 osob, jaká je pravděpodobnost, že jejich průměrné IQ bude větší nebo rovno 112?
 - $\mu = 100, \sigma_x = 1.5$
 - **$z = 3.33$**
 - z tabulky $P(Z \geq 3.33) = \mathbf{0.0005}$
-

Příklad 1

- IQ ($\mu=100$, $\sigma=15$)
 - jaká je **pravděpodobnost**, že **průměr výběru** o velikosti $n=25$ bude mezi **hodnotami 94 a 106?**
-

Příklad 1

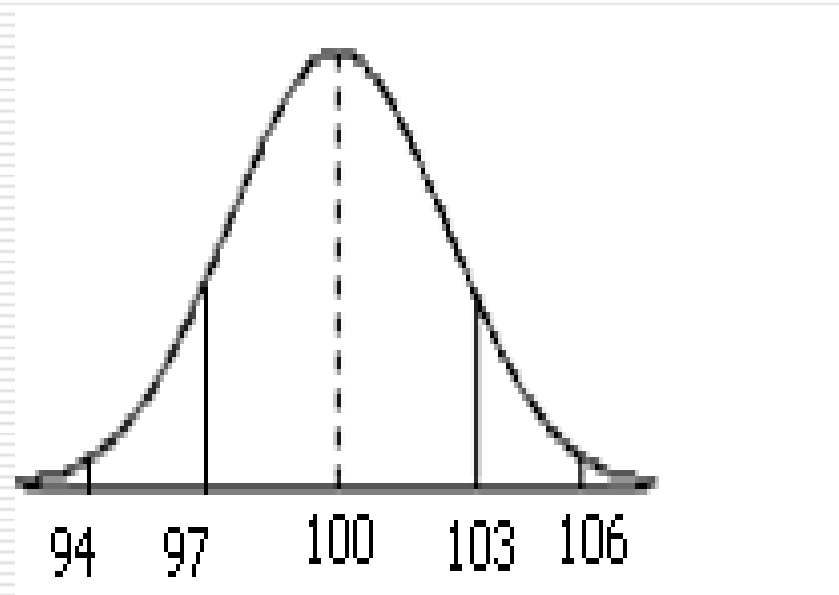
- IQ ($\mu=100$, $\sigma=15$)
 - jaká je pravděpodobnost, že průměr výběru o velikosti $n=25$ bude mezi hodnotami 94 a 106?
 - $\mu = 100$, $\sigma = 15$, $n=25$
 - výpočet **výběrové chyby**:
 $\sigma_x = \sigma/\sqrt{n} = 15/5 = 3$
-

Příklad 1

- IQ ($\mu=100$, $\sigma=15$)
 - jaká je pravděpodobnost, že průměr výběru o velikosti $n=25$ bude mezi hodnotami 94 a 106?
 - $z = (94-100) / 3 = -6/3 = -2$
 - $z = (106-100) / 3 = 6/3 = 2$
-

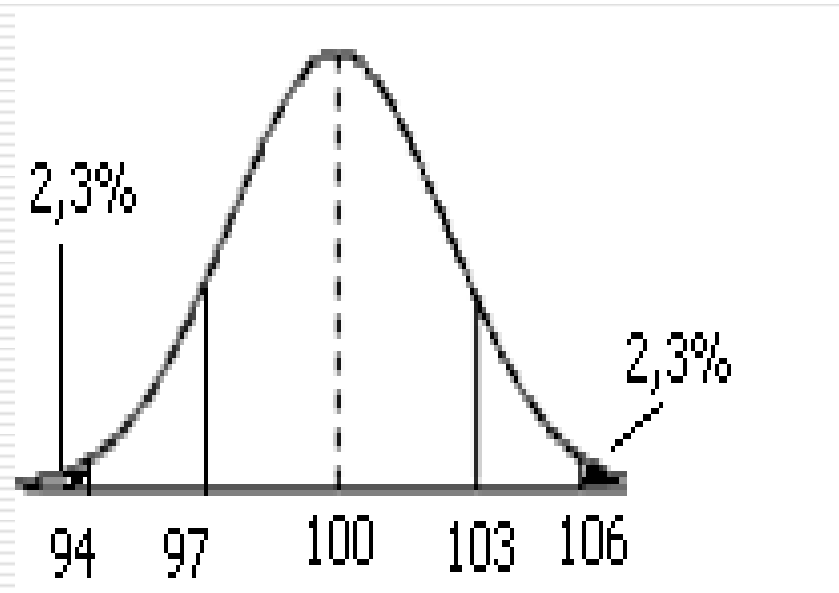
Příklad 1

- najdeme v tabulce normovaného normálního rozdělení hodnotu pravděpodobnosti pro $z=2$ a $z=-2$



Příklad 1

- hodnota je 0,023
- dohromady 0,046
- odečteme od 1,00
- výsledek **0,954**



Příklad 1

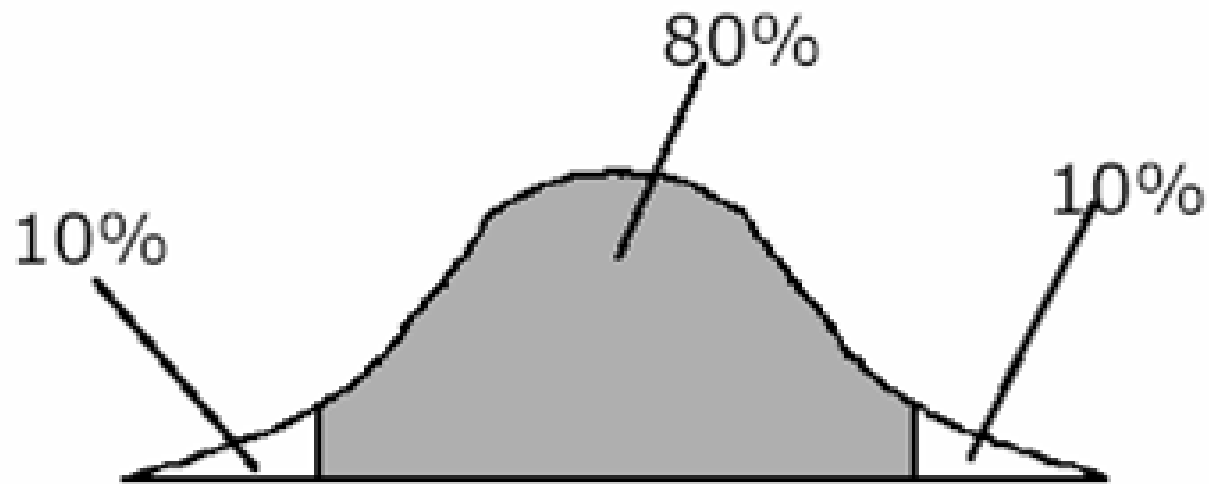
- IQ ($\mu=100$, $\sigma=15$)
 - jaká je pravděpodobnost, že průměr výběru o velikosti $n=25$ bude mezi hodnotami 94 a 106?
 - **pravděpodobnost takového průměru je 95,4%**
-

Příklad 2

- IQ ($\mu=100$, $\sigma=15$)
 - v jakém **rozsahu hodnot** bude pravděpodobně **80% všech průměrů výběrů** o velikosti $n=25$?
-

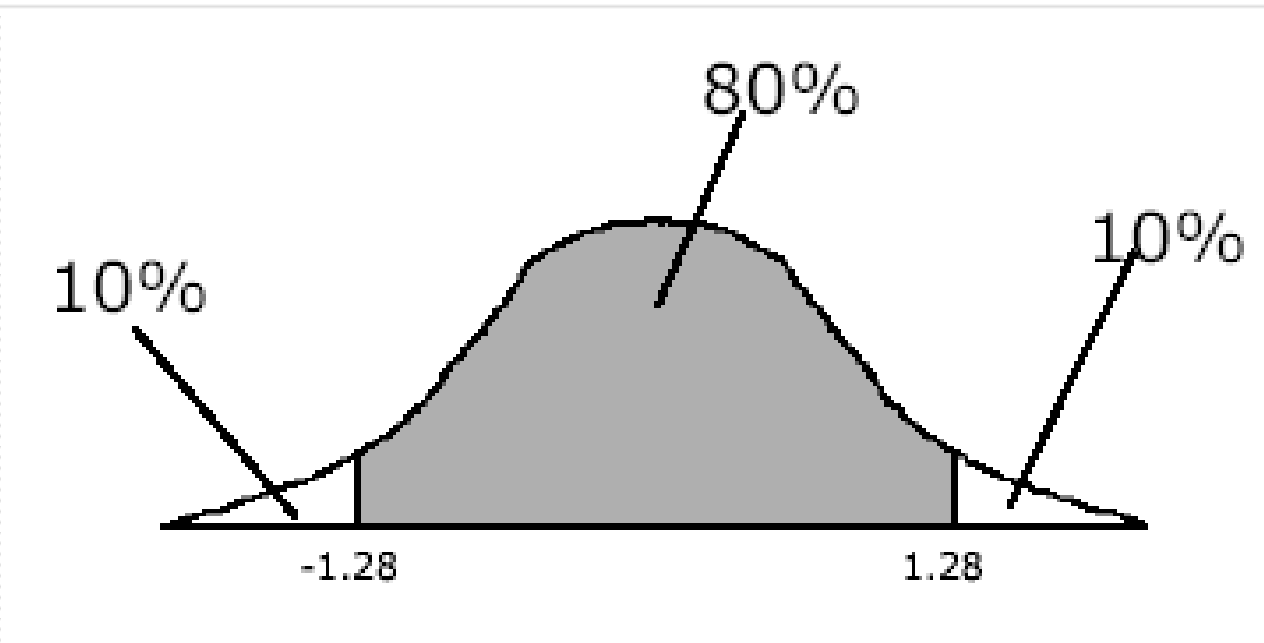
Příklad 2

- IQ ($\mu=100$, $\sigma=15$)
 - v jakém **rozsahu hodnot** bude pravděpodobně **středních 80% všech průměrů výběrů** o velikosti $n=25$?
 - potřebujeme zjistit hodnotu z , která odděluje pravděpodobnost 10% na obou stranách rozdělení
-



Příklad 2

- IQ ($\mu=100$, $\sigma=15$)
 - v jakém **rozsahu hodnot** bude pravděpodobně **středních 80% všech průměrů výběrů** o velikosti $n=25$?
 - $z = 1,28$ a $-1,28$ převedeme na hodnoty IQ pro rozdělení výběrových průměrů
-



Příklad 2

$$\bar{x} = \mu + z^*(\sigma/\sqrt{n})$$

$$\bar{x} = 100 + 1,28*(15/\sqrt{25}) = 100+1,28(3) = \mathbf{103,84}$$

$$\bar{x} = 100 + (-1,28)*(15/\sqrt{25}) = 100-1,28(3) = \mathbf{96,16}$$

Příklad 2

- IQ ($\mu=100$, $\sigma=15$)
 - v jakém rozsahu hodnot bude pravděpodobně středních 80% všech průměrů výběrů o velikosti $n=25$?
 - **80% všech výběrových průměrů bude v rozsahu hodnot 96,16 - 103,84**
-

Odhady

- v předchozích příkladech jsme znali hodnoty průměru a rozptylu populace
 - obvykle tomu ale bývá přesně naopak: **známe hodnoty (statistiky) výběru a neznáme hodnoty (parametry) populace**
 - ty chceme z výběru **odhadnout**
-

Odhady

- 2 typy odhadů: bodové a intervalové
 - **bodový odhad**: použijeme průměr vzorku a odhadneme, že se rovná průměru populace
-

Bodový odhad

- ❑ bodový odhad je problematický v tom, že dva různé výběry nám mohou dát dva různé odhady
 - ❑ bodový odhad **neobsahuje** žádnou **informaci** o jeho **přesnosti** či **spolehlivosti**
 - ❑ na čem závisí přesnost odhadu?
-

Bodový odhad

přesnost odhadu závisí na dvou charakteristikách

- **velikost výběru** (čím větší n , tím menší výběrová chyba)
 - **variabilita hodnot v populaci** (čím vyšší, tím vyšší i výběrová chyba)
-

Intervalový odhad

- poskytuje rozsah (interval) hodnot, který s určitou pravděpodobností obsahuje hledanou hodnotu parametru
-

Intervalový odhad

je založen na:

- bodovém odhadu
 - velikosti výběru
 - variabilitě znaku v populaci (známé nebo rovněž odhadované)
-

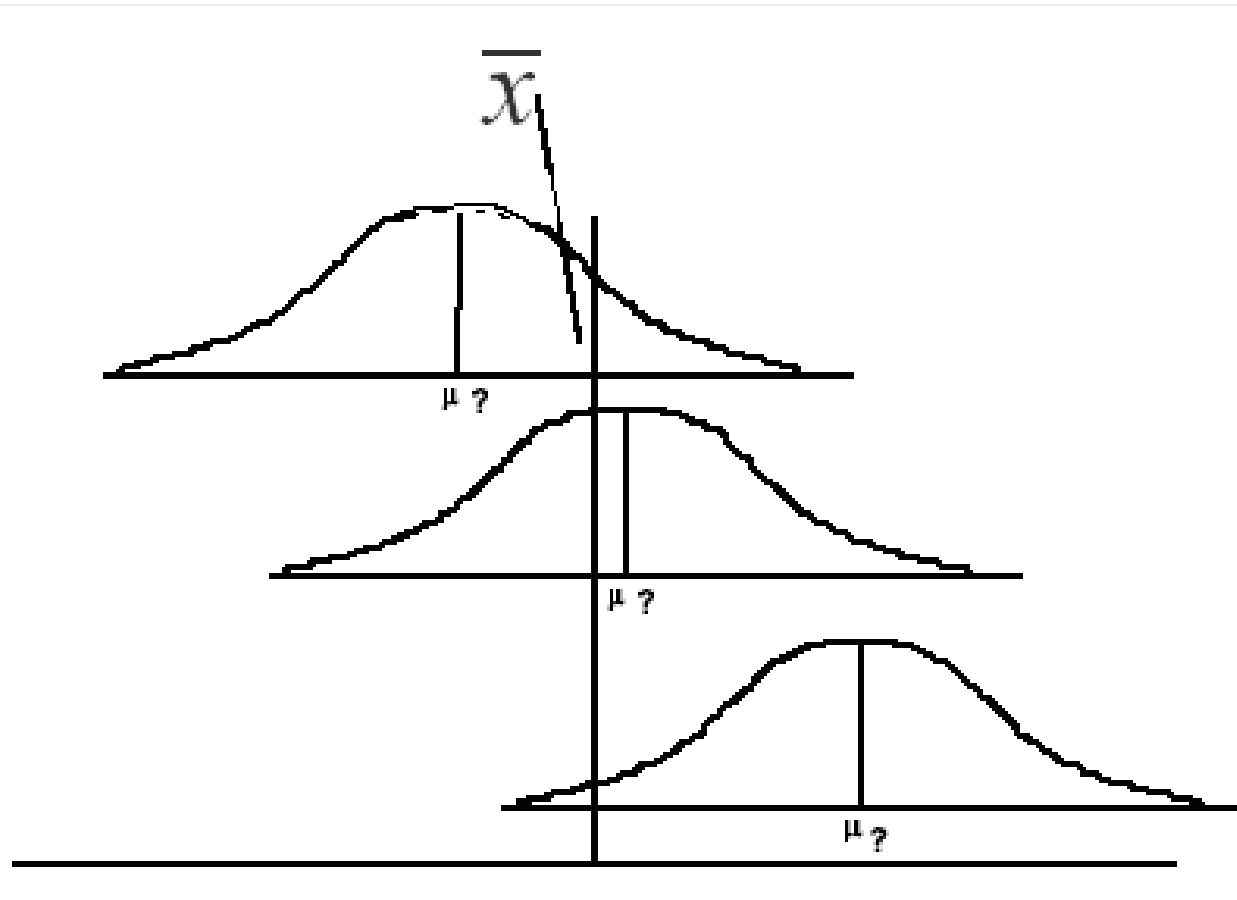
Intervalový odhad

□ ptáme se: **jaká je hodnota μ ?**

Intervalový odhad

- ptáme se: **jaká je hodnota μ ?**
 - výběrový průměr určité hodnoty může pocházet z populací o různých průměrech
 - proto **nemůžeme jednoznačně určit hodnotu μ**
-

Intervalový odhad



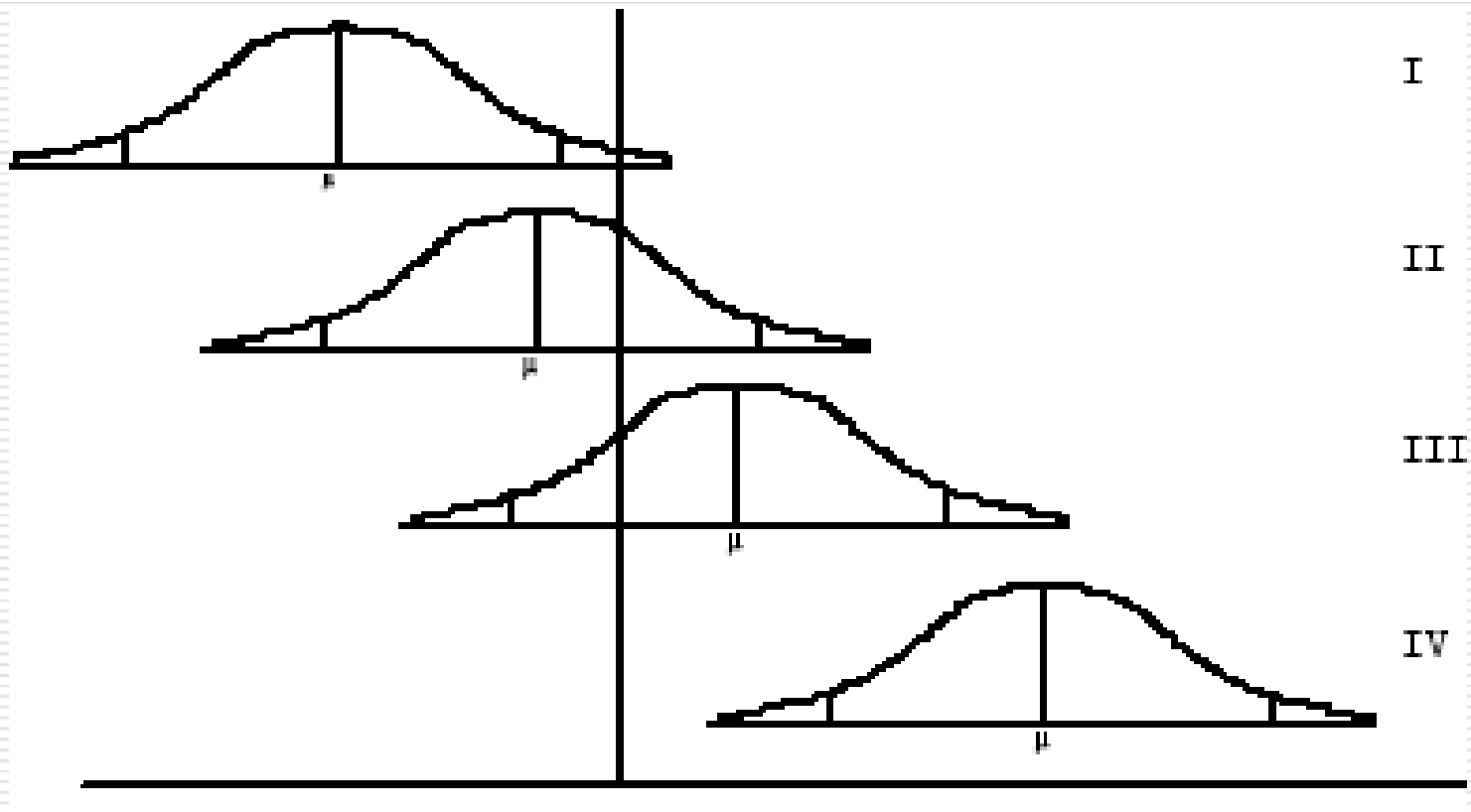
Intervalový odhad

- takže se místo toho snažíme určit, jaký je **možný rozsah hodnot μ**
 - jaké populace (tj. s jakou hodnotou průměru) by mohly být pravděpodobným zdrojem našeho vzorku?
-

Intervalové odhady

- ze které populace nejpravděpodobněji pochází výběr, jehož průměr je v následujícím grafu naznačen svislou čarou?

RVP pro populace I-IV



Intervalové odhady

- výběr pochází
 - nejpravděpodobněji z populace II nebo III
 - méně pravděpodobně z populace I
 - a velmi málo pravděpodobně z populace IV
-

Intervalové odhady

- intervalový odhad spočívá v konstrukci tzv. **intervalu spolehlivosti** (confidence interval) – rozsahu hodnot, ve kterém s určitou pravděpodobností leží průměr populace
-

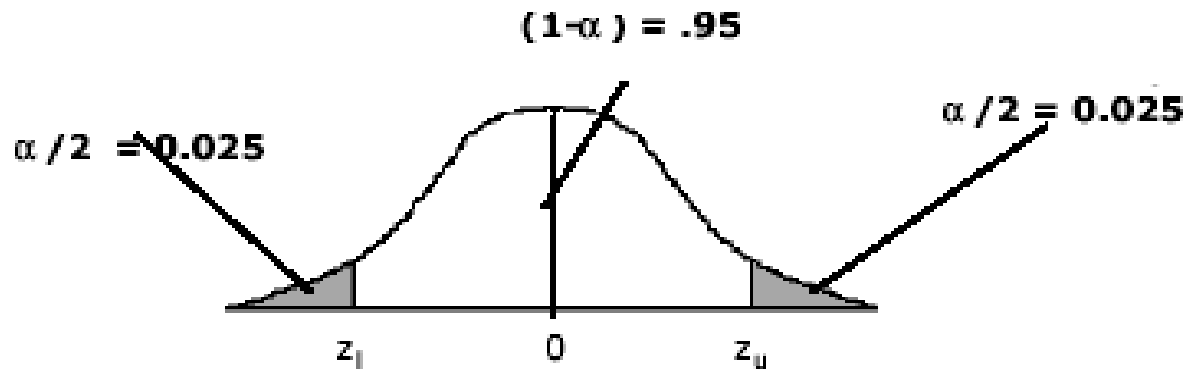
Interval spolehlivosti

- nejprve je třeba si **stanovit tuto pravděpodobnost** – tj. úroveň přesnosti(spolehlivosti);
 - obvyklá je např. **95%** - snažíme se najít interval hodnot, ve kterém s 95% pravděpodobností leží průměr populace
 - pak jde o tzv. **95% interval spolehlivosti**
-

Interval spolehlivosti

- poté **najít hodnotu z pro tuto pravděpodobnost** – tj. rozsah, ve kterém bude ležet středních 95% hodnot (výběrových průměrů)
 - 2,5% na každé straně rozdělení
-

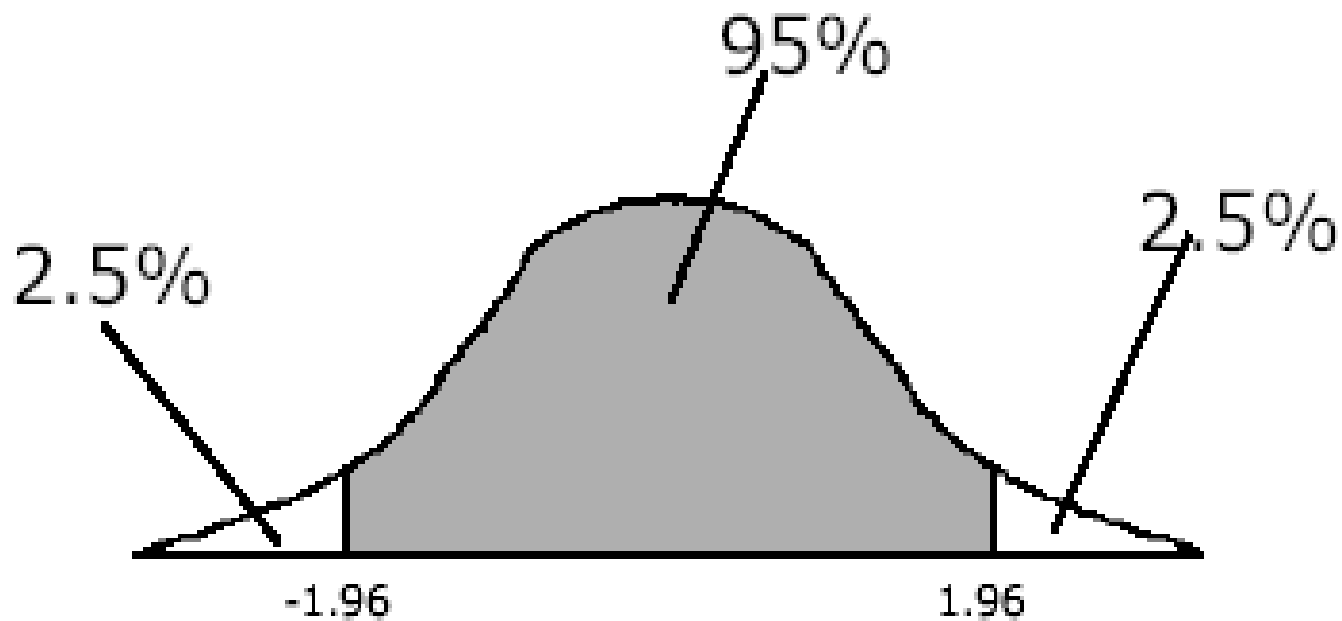
Interval spolehlivosti



Interval spolehlivosti

- tomu odpovídají hodnoty
 $z = -1,96$
 $z = 1,96$

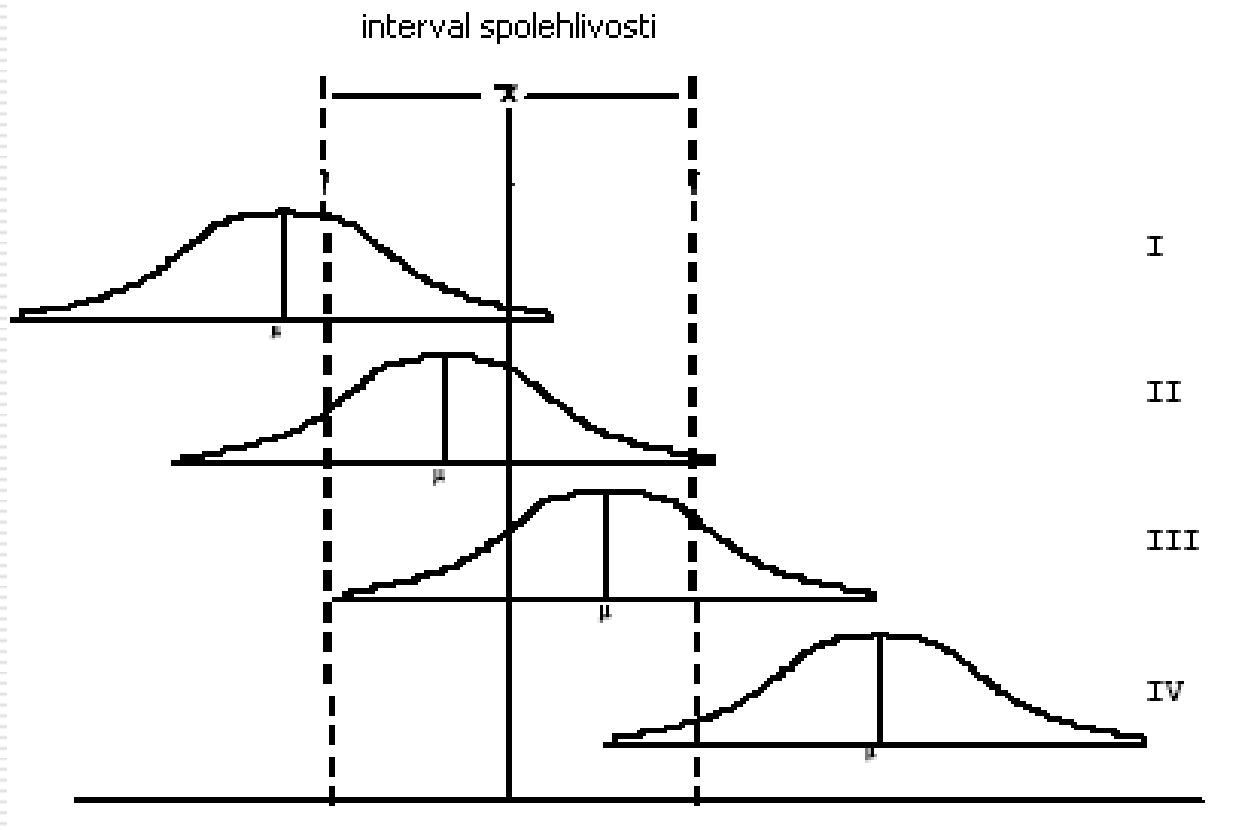
Interval spolehlivosti



Interval spolehlivosti - výpočet

$$\bar{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Interval spolehlivosti



Interval spolehlivosti

- interpretace intervalu spolehlivosti: máme 95% pravděpodobnost, že se v tomto intervalu nachází průměr populace
 - pokud bychom z populace vybrali 100 náhodných výběrů o velikosti n a pro každý z nich sestrojili tento interval, 95 intervalů by obsahovalo průměr populace a 5 nikoliv
-

Interval spolehlivosti

- oblíbený omyl:
 - v 95% intervalu spolehlivosti leží 95% hodnot populace (NEPLATÍ!)

 - kromě 95% intervalu spolehlivosti se používá také např. 99% a 90% pravděpodobnost
-

Příklad 3

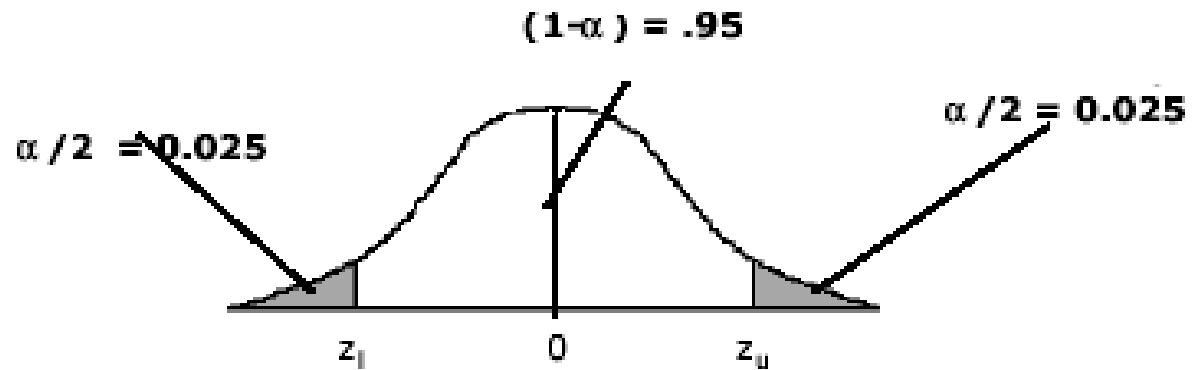
- náhodný výběr 36 dětí, průměrné IQ vzorku = 96
 - na základě tohoto zjištění odhadněte průměrné IQ populace, ze které děti pocházejí (sestavte 95% interval spolehlivosti)
-

Příklad 3

□ Postup:

- bodový odhad: $\mu=96$
 - výpočet výběrové chyby (směrodatné odchyly RVP):
$$\sigma/\sqrt{n} = 15/\sqrt{36} = 15/6 = 2,5$$
 - stanovení úrovně spolehlivosti: 95%
 - najít hodnotu z pro 95% pravděpodobnost
-

Příklad 3



Příklad 3

- v tabulce normálního rozdělení najdeme hodnoty z
 - hodnoty z pro 95% : 1,96 a -1,96
-

Příklad 3

- k výběrovému průměru přičteme (pro horní hranici intervalu) a odečteme (pro spodní hranici) výběrovou chybu, vynásobenou hodnotou z
-

Příklad 3

□ pro 95% :

$$\mu = 96 + 1,96 * 2,5 = 100,90$$

$$\mu = 96 - 1,96 * 2,5 = 91,10$$

**95% interval spolehlivosti je 91,1 – 100,9
bodů IQ**

Příklad 3

- pro 99% interval spolehlivosti
 - $z = 2,57$
-

Příklad 3

□ pro 99% :

$$\mu = 96 + 2,57 * 2,5 = 102,43$$

$$\mu = 96 - 2,57 * 2,5 = 89,58$$

**99% interval spolehlivosti je 89,6 – 102,4
bodů IQ**

Interval spolehlivosti

□ **hodnoty z** pro nejčastěji užívané pravděpodobnosti:

■ 90% (zbývá 5% + 5%) $z = +/- 1,645$

■ 95% (zbývá 2,5% + 2,5%) $z = +/- 1,96$

■ 99% (zbývá 0,5% + 0,5%) $z = +/- 2,57$

Interval spolehlivosti

- v předchozích příkladech jsme předpokládali, že **známe hodnotu variability znaku v populaci**
 - ve skutečnosti je tomu tak však **zřídka**
 - je proto nutno **odhadnout** zároveň s průměrem i hodnotu směrodatné odchylky
-

Interval spolehlivosti

pro známé hodnoty směrodatné odchylky v populaci:

$$\bar{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Studentovo rozdělení

- pokud **za σ nahradíme s** (směr. odchylku výběrového průměru), pak musíme při konstrukci intervalu spolehlivosti místo z rozdělení použít tzv. **Studentovo t rozdělení**
-

Interval spolehlivosti

pro **neznámé** hodnoty směrodatné odchyly v populaci:

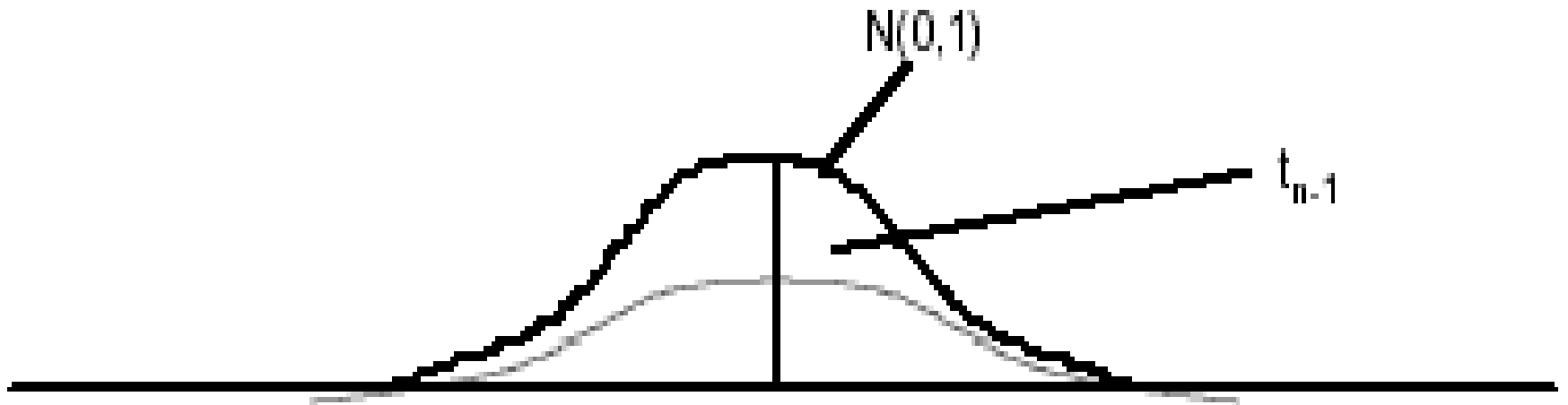
$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Studentovo rozdělení

- má také zvonovitý tvar, ale je více ploché než normální rozdělení
 - je symetrické kolem průměru (0)
 - pro každou velikost výběru (počet stupňů volnosti, df) existuje odlišné t rozdělení
 $df = n - 1$
-

Studentovo rozdělení

srovnání s normálním rozdělením



Studentovo rozdělení

- srovnání s normálním rozdělením:
 - t rozdělení má vyšší variabilitu
 - více plochy na okrajích, méně ve středu
 - vzhledem k vyšší variabilitě budou intervaly spolehlivosti širší než u normálního rozdělení
 - jsou uváděny df obvykle jen do 100, protože pro $n=100$ se t rozdělení blíží normálnímu rozdělení
-

Studentovo rozdělení

□ tabulka t-rozdělení:

- každý řádek udává hodnoty t pro celé rozdělení pro daný počet stupňů volnosti (tj. $n-1$)
 - sloupce pro nejdůležitější percentily
-

Studentovo rozdělení

d.f.	t_{90}	t_{95}	t_{975}	t_{99}	t_{995}
1	3.078	6.3138	12.706	31.821	63.657
2	1.886	2.92	4.3027	6.965	9.9248
3	1.638	2.3534	3.1825	4.541	5.8409

Odhady podílů

- u kategoriálních proměnných nemůžeme počítat průměry
 - odhadujeme proto **podíly** jednotlivých kategorií proměnné
-

Odhady podílů

- např. podíl kuřáků v populaci českých adolescentů
 - podíl pacientů s rakovinou plic, kteří přežijí 5 let od diagnózy
 - podíl chlapců mezi dětmi s poruchou pozornosti
-

Odhady podílů

- pokud zkoumáme místo celé populace pouze výběr z ní, nezajímá nás tolik, jaký je podíl kategorií proměnné ve výběru (četnost \mathbf{p})
 - ale spíše jaký je skutečný podíl v populaci – četnost π
-

Odhady podílů

- při dostatečně velkém n platí i pro rozdělení podílů centrální limitní věta
- rozdělení výběrových podílů je normální rozdělení, s **průměrnou četností π** a směrodatnou odchylkou (výběrovou chybou)

$$SE = \sqrt{\frac{\pi(1-\pi)}{n}}$$

Příklad 4

- chceme zjistit, jaká je podpora odsunu hlavního nádraží v Brně
 - náhodný výběr z populace brněnských občanů starších 18 let
 $n=1000$ osob
 - 585 osob se vyjádřilo pro
 - $p=0,585$
 - odhadněte s 95% spolehlivostí podporu odsunu nádraží
-

Odhady podílů

- interval spolehlivosti pro podíly se spočítá podobně jako pro průměry:

$$p \pm z_{1-\alpha/2} \sigma_p$$

Odhady podílů

- ❑ nemůžeme však spočítat výběrovou chybu, protože neznáme π
 - ❑ v tomto případě je však možné dosadit místo toho p a přitom použít normální rozdělení (pokud je $n > 30$)
 - ❑ pokud je $n < 30$, pak dosadíme místo π hodnotu 0,5
-

Příklad 4

□ $p=0,585$

□ $z=1,96$

□ $SE(p) = \sqrt{[0,585(1-0,585)/1000]}$
 $=0,156$

interval spolehlivosti

$$0.585 \pm 1.96(0.0156)$$

$$0.585 \pm 0,0305$$

--- přesnost odhadu je $\pm 3\%$

Příklad 4

- s 95% pravděpodobností je podíl osob souhlasících se odsunem nádraží **mezi 55.4% a 61.6%**
 - tj. máme 95% pravděpodobnost, že kdyby se v době průzkumu hlasovalo, bude většina pro
-

Odhady podílů

vztah mezi velikostí vzorku a přesností odhadu

- $n=100$ $\pm 10\%$
 - $n=200$ $\pm 7\%$
 - $n=400$ $\pm 5\%$
 - $n=1000$ $\pm 3\%$
 - $n=2400$ $\pm 2\%$
 - $n=9600$ $\pm 1\%$
-

Odhady podílů

- ❑ požadovaná velikost vzorku roste mnohem rychleji než spolehlivost odhadu (pro zdojnásobení spolehlivosti je nutné asi čtyřnásobně zvětšit vzorek)
 - ❑ důležité při plánování výzkumu – jakou přesnost potřebujeme? jaké budou náklady?
 - ❑ podobný vztah platí pro odhad průměrů
-

Kontrolní otázky

- 2 typy odhadů
 - na čem závisí šířka intervalu spolehlivosti? (*není nutno znát zpaměti vzorce, ale je třeba chápat princip výpočtu*)
 - vztah velikosti výběru a spolehlivosti odhadu
-