

Odhady

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

Odhady

- jedna z aplikace postupů statistického usuzování (inference)
 - zobecňujeme z náhodného výběru na populaci
 - vztah mezi výběrovou statistikou a parametrem populace popisují výběrová rozdělení statistik (např. rozdělení výběrových průměrů) = teoretická pravděpodobnostní rozdělení
 - dva druhy výběrových chyb: systematická chyba, náhodná chyba
-

Odhady

- dva druhy výběrových chyb: systematická chyba, náhodná chyba
 - systematickou výběrovou chybu redukuje náhodným výběrem
 - náhodná výběrová chyba = směrodatná odchylka rozdělení výběrové statistiky (tzv. směrodatná chyba)
 - její velikost závisí na velikosti výběru a na variabilitě znaku v populaci
-

Odhady

- v příkladech v předchozí přednášce 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**: jediná hodnota - použijeme např. 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 náhodná výběrová chyba)
 - **variabilita hodnot v populaci** (čím vyšší, tím vyšší i výběrová chyba)
-

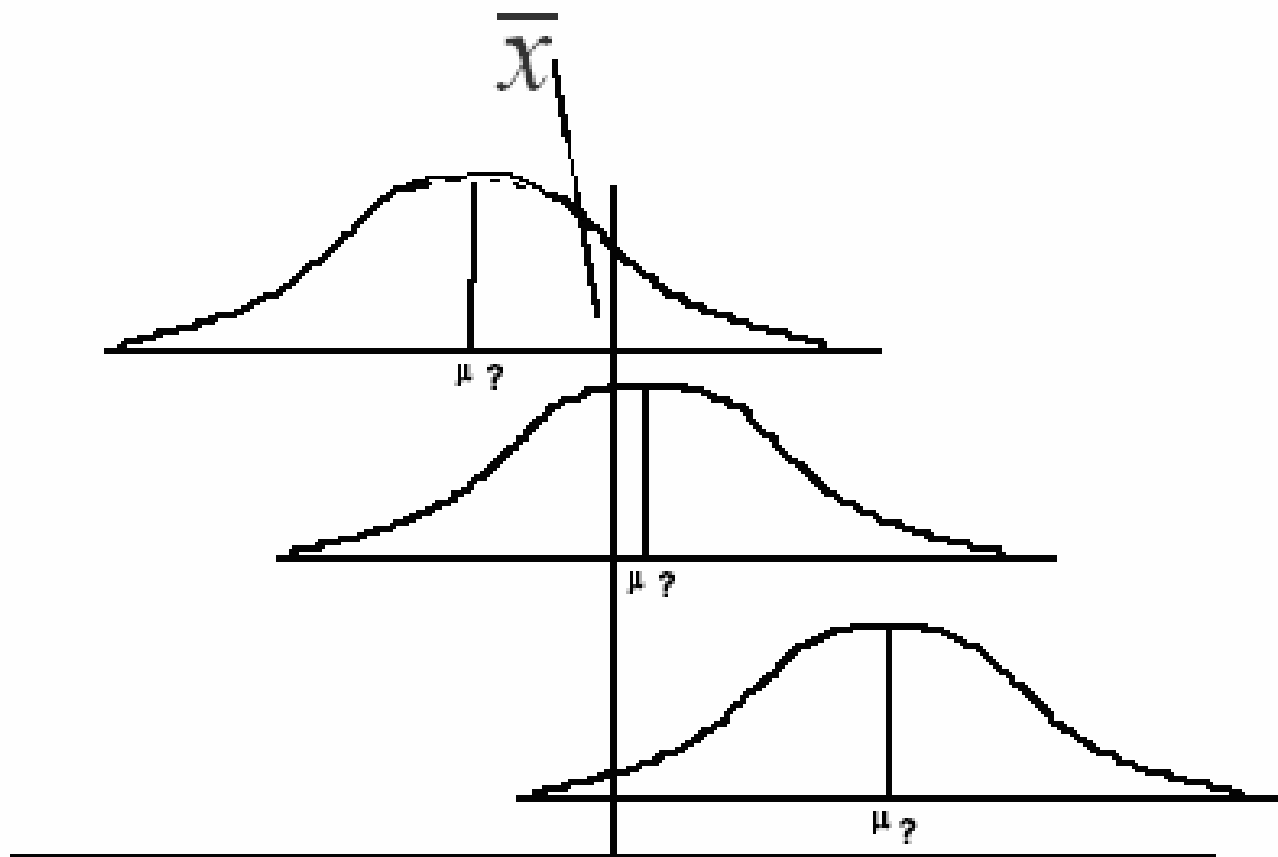
Intervalový odhad

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

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



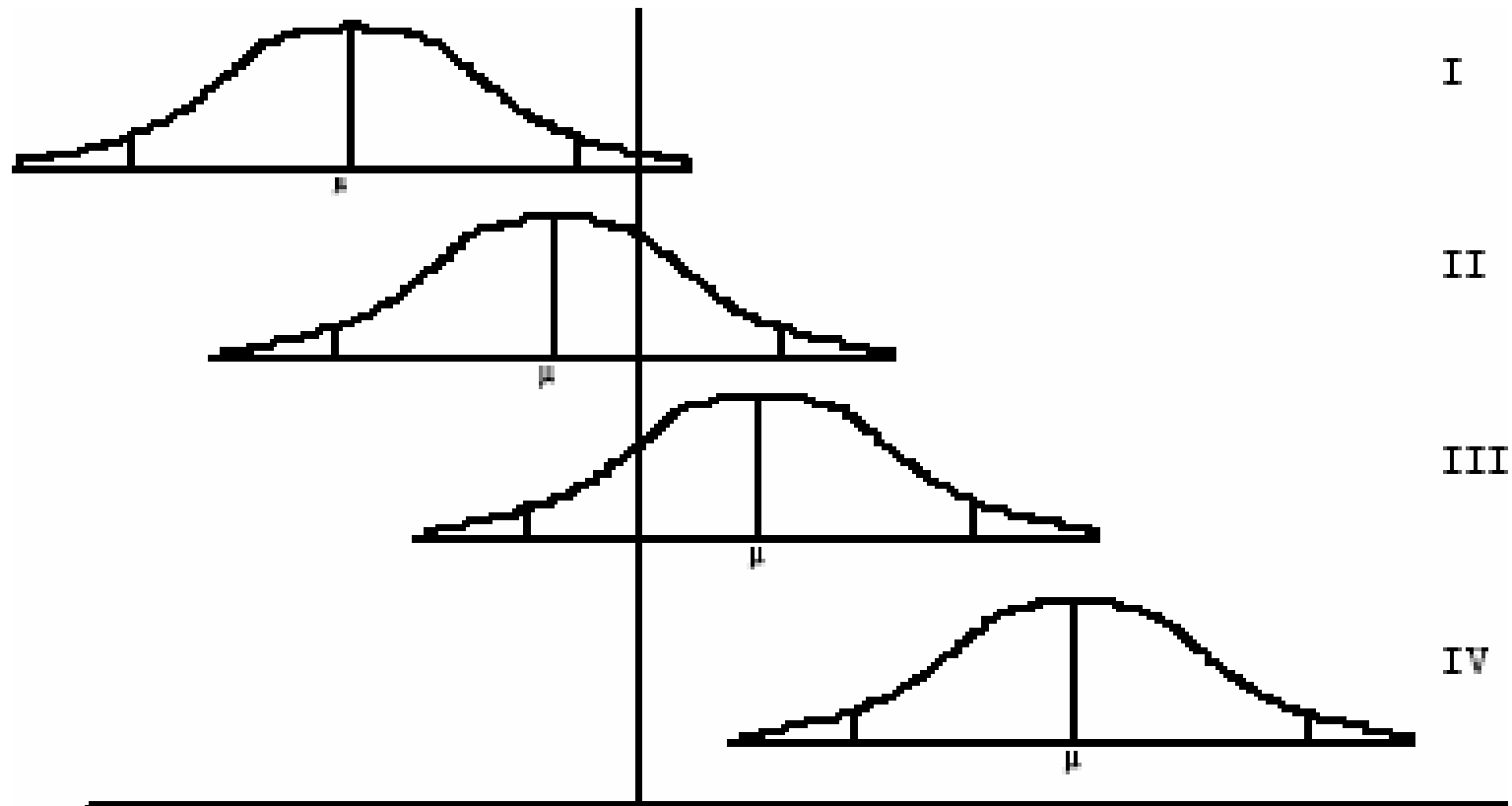
Intervalový odhad

- 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 (resp. ve kterém se při opakovaném provádění výběru ocitne náš odhad v určitém % případů)
-

Intervalové odhady

- konstrukce intervalu spolehlivosti pro průměr:
 - výběrový průměr je bodovým odhadem průměru populace – tvoří střed intervalu
 - směrem nalevo od něj je dolní hranice intervalu, napravo horní hranice
-

Intervalové odhady

- hranice intervalu (a tím jeho délka, tj. spolehlivost odhadu) jsou dány velikostí výběrové chyby a požadovanou úrovní spolehlivosti
 - čím větší výběrová chyba, tím delší interval
 - čím vyšší požadovaná spolehlivost odhadu, tím delší interval
-

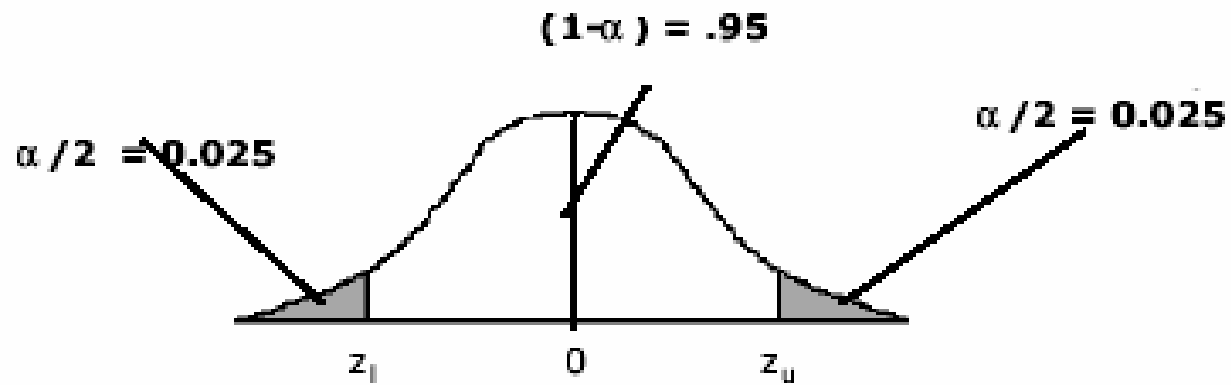
Interval spolehlivosti

- nejprve je třeba si **stanovit pravděpodobnost** – tj. úroveň přesnosti intervalu;
 - obvyklá je např. **95%** (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



Tabulka z-rozdělení

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
...										
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857

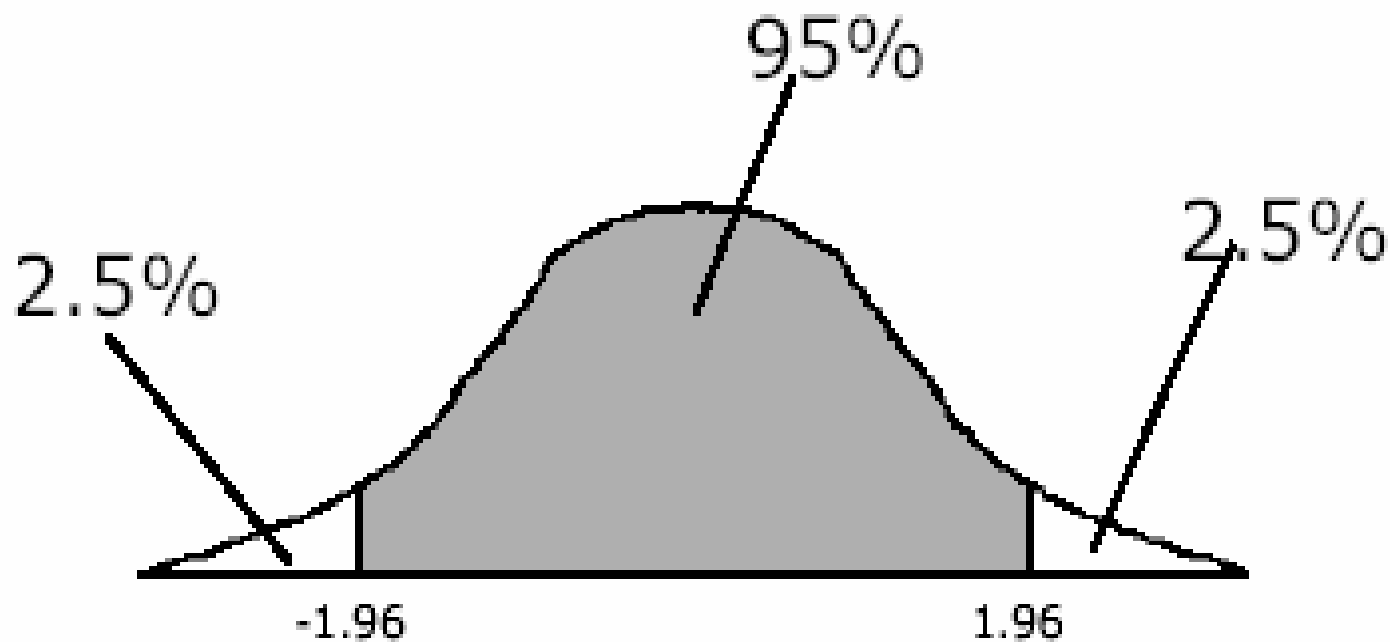
Interval spolehlivosti

□ tomu odpovídají hodnoty

$$z = -1,96$$

$$z = 1,96$$

Interval spolehlivosti



Interval spolehlivosti

- interval spolehlivosti pro průměr tedy leží kolem výběrového průměru v rozmezí
 - 1.96 směrodatné odchylky rozdělení až
 - +1.96 směrodatné odchylky od průměru
 - směrodatná odchylka rozdělení výběrových průměrů = výběrová chyba (směrodatná chyba)
-

Interval spolehlivosti - výpočet

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

Interval spolehlivosti

- interpretace intervalu spolehlivosti:
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

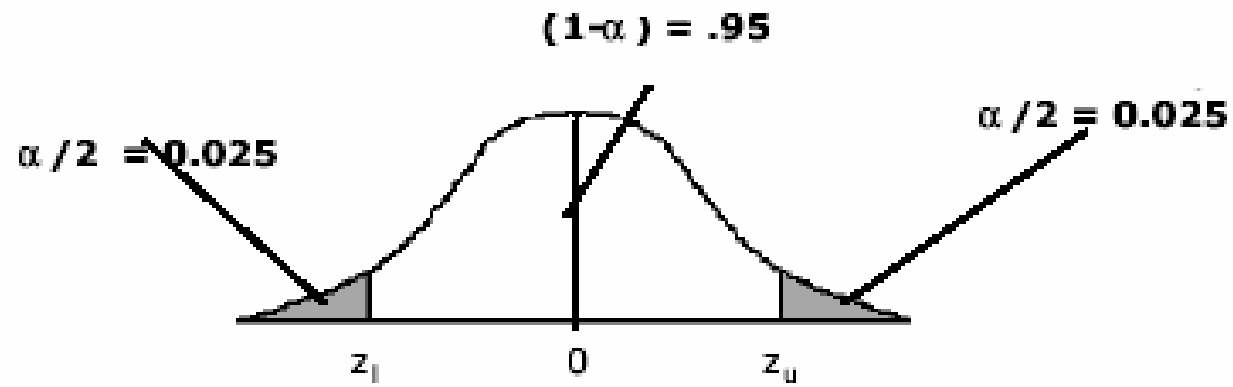
- náhodný výběr 36 dětí, průměrné IQ vzorku = 96
 - na základě tohoto zjištění odhadněte průměrné IQ populace, ze které děti pocházejí (sestavte 95% interval spolehlivosti)
-

Příklad

□ 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



Tabulka z-rozdělení

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
...										
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857

Příklad

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

Příklad

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

Příklad

□ 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

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

Příklad

□ 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

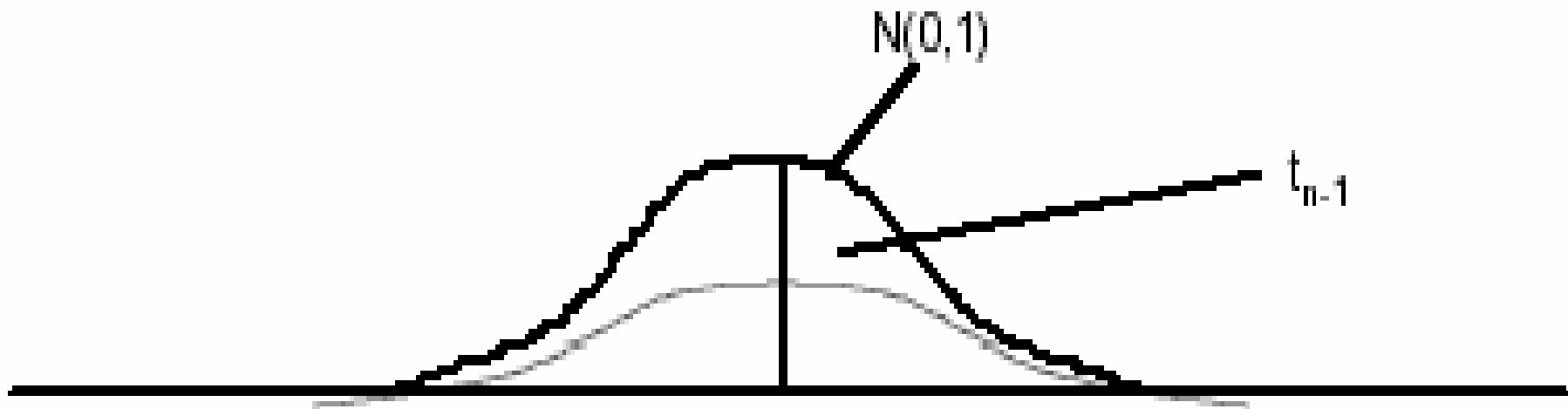
- pokud **za σ nahradíme s** (směr. odchylku výběrového průměru), pak musíme při konstrukci intervalu spolehlivosti místo z rozdělení použít místo z-rozdělení tzv. **Studentovo t rozdělení**
-

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 náhodný výběr z ní, nezajímá nás tolik, jaký je podíl kategorií proměnné ve výběru (četnost **p**)
 - ale spíše jaký je skutečný podíl v populaci – četnost **π**
-

Odhady podílů

□ také zde platí obecný vzorec pro interval spolehlivosti

= bodový odhad $-/+$ koeficient spolehlivosti * výběrová chyba

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 (pak je hodnota výběrové chyby nejvyšší možná)
-

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%**
-

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 vzorce, jen chápat princip výpočtu*)
 - vztah velikosti výběru a spolehlivosti odhadu
-