Intervalové odhady

- měli jsme: bodové odhady (odhadem charakteristiky je číslo)
- nevyjadřuje nic o přesnosti odhadu

Intervalový odhad parametru θ

- konstruujeme z pozorovaných dat tak, aby pokrýval neznámou hodnotu θ s předepsanou pravděpodobností (např. 95 %)
- interval s náhodnými mezemi, který překryje θ s předepsanou pravděpodobností
- např. 95% interval spolehlivosti, interval na hladině 99% apod.
- též *konfidenční interval* nebo intervalový odhad

Interval spolehlivosti pro střední hodnotu v $\mathrm{N}(\mu, \sigma^2)$

Situace: $X_1, \ldots, X_n$ náhodný výběr z $\mathrm{N}(\mu, \sigma^2)$, kde $\sigma^2 > 0$ známe

- víme $\overline{X} \sim \mathrm{N}(\mu, \sigma^2/n)$ a proto (už jsme viděli dříve)

$$0.95 = \mathrm{P}\left(\sqrt{n}\frac{|\overline{X} - \mu|}{\sigma} < u_{0.975}\right)$$

- po úpravě

$$\mathrm{P}\left(|\overline{X} - \mu| < 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

a tedy

$$\mathrm{P}\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

- dostali jsme 95% interval spolehlivosti pro μ

Interpretace intervalu spolehlivosti

- 95% interval spolehlivosti překryje s pravděpodobností 95 % skutečnou hodnotu μ
- kdybych postup prováděli opakovaně, tak cca v 95 % případů interval pokryje skutečnou hodnotu μ, ve zbylých 5 % bude skutečné μ mimo

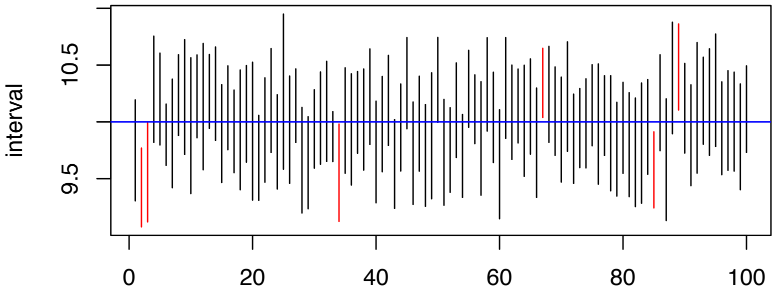
Obecně, interval spolehlivosti pro μ na hladině $1 - \alpha$:

$$\left(\overline{X} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

pokryje skutečnou hodnotu μ s pstí $1 - \alpha$

Interval spolehlivosti – ilustrace

- 100 výběrů z $N(10, 1)$ o rozsahu $n = 20$
- v každém výběru spočten 95 % interval spolehlivosti pro μ
- skutečná hodnota $\mu = 10$ není překryta v 6 případech



Interval spolehlivosti pro střední hodnotu při neznámém σ

Situace: $X_1, \ldots, X_n$ náhodný výběr z $\mathrm{N}(\mu, \sigma^2)$, kde $\sigma^2 > 0$ neznáme

- neznámé σ nahradíme odhadem S_n
- kvantily normálního rozdělení musíme nahradit kvantily Studentova t-rozdělení
- dostaneme

$$\mathrm{P}\left(\overline{X} - \frac{S_n}{\sqrt{n}} t_{n-1}(1 - \frac{\alpha}{2}) < \mu < \overline{X} + \frac{S_n}{\sqrt{n}} t_{n-1}(1 - \frac{\alpha}{2})\right) = 1 - \alpha$$

- interval s náhodnými mezemi, který pokryje skutečnou hodnotu μ s pstí $1 - \alpha$

Příklad - pivo (viz minule)

Bylo zakoupeno 10 piv a jejich objem byl (v litrech):

0.510, 0.462, 0.491, 0.466, 0.461, 0.503, 0.495, 0.488, 0.512, 0.505.

Předpokládali jsme, že data pochází z $N(\mu, \sigma^2)$.

- měli jsme $n = 10$,

$$\overline{X} = 0.4893, \quad S_n = 0.0197, \quad t_9(0.975) = 2.262$$

- 95% interval spolehlivosti pro střední hodnotu natočeného objemu piva:

$$(0.475,\ 0.503)$$

- 99% interval spolehlivosti (využijeme $t_9(0.995) = 3.250$):

$$(0.469,\ 0.510)$$

Vlastnosti intervalu spolehlivosti

Délka intervalu spolehlivosti pro střední hodnotu je rovna

$$2t_{n-1}\big(1-\frac{\alpha}{2}\big)\frac{S_n}{\sqrt{n}}.$$

Závisí tedy na pravděpodobnosti pokrytí α, počtu pozorování n a rozptylu pozorování σ^2 (skrze jeho odhad S_n^2):

- vyšší je požadovaná pravděpodobnost pokrytí $\rightsquigarrow$ delší interval
- více pozorování $\rightsquigarrow$ kratší interval
- větší rozptyl pozorování $\rightsquigarrow$ delší interval

Poznámka

Lze uvažovat i *jednostranné* intervaly spolehlivosti

- např. z rovnosti

$$0.95 = \mathrm{P}\left(\sqrt{n}\frac{\overline{X} - \mu}{S_n} < t_{n-1}(0.95)\right)$$

  dostaneme po úpravách 95% *levostranný* interval spolehlivosti pro μ

$$\left(\overline{X} - \frac{S_n}{\sqrt{n}} t_{n-1}(0.95), \infty\right)$$

- podobně *pravostranný* 95 % interval spolehlivosti pro μ je

$$\left(-\infty, \overline{X} + \frac{S_n}{\sqrt{n}} t_{n-1}(0.95)\right)$$

Předpoklad normality

- ověřuje se stejně jako t-testu
- je-li n dost velké, lze uvedené intervaly použít i při porušení normality
  - interpretace: *asymptotické* intervaly spolehlivosti $\leftrightsquigarrow$ intervalové odhady se spolehlivostí, která se blíží k $1-\alpha$ pro $n \to \infty$

Souvislost mezi testy a intervaly spolehlivosti

- oboustranný interval spolehlivosti pro μ

$$\left(\overline{X} - \frac{S_n}{\sqrt{n}} t_{n-1}(1 - \frac{\alpha}{2}), \overline{X} + \frac{S_n}{\sqrt{n}} t_{n-1}(1 - \frac{\alpha}{2})\right)$$

- μ_0 patří do intervalu spolehlivosti $\Leftrightarrow$ platí

$$|\overline{X} - \mu_0| < \frac{S_n}{\sqrt{n}} t_{n-1}(1 - \alpha/2)$$

- tj. μ_0 patří do intervalu spolehlivosti $\Leftrightarrow$ nezamítáme $H_0 : \mu = \mu_0$ proti $H_1 : \mu \neq \mu_0$
- interval spolehlivosti obsahuje takové hodnoty μ_0, pro které bychom nezamítli $H_0 : \mu = \mu_0$ $\rightsquigarrow$ intervaly spolehlivosti lze použít pro testování hypotéz
- podobná souvislost mezi jednostrannými intervaly spolehlivosti a jednostrannými alternativami H_1

Příklad – pivo

- 95% interval spolehlivosti pro střední hodnotu natočeného objemu piva byl
$$(0.475,\ 0.503)$$
$\rightsquigarrow$ nezamítáme $H_0 : \mu = 0.5$ proti $H_1 \neq 0.5$ na hladině 5%
- 95% pravostranný interval spolehlivosti
$$(-\infty, 0.501)$$
$\rightsquigarrow$ nezamítáme $H_0 : \mu = 0.5$ proti $H_1 < 0.5$ na hladině 5%

Poznámka

Intervalový odhad

- interval spolehlivosti se počítá i pro jiné parametry než μ
- lze uvažovat interval spolehlivosti pro pravděpodobnost, rozptyl, rozdíl středních hodnot dvou výběrů ...
- vždy je to interval, který s požadovanou pravděpodobností překryje skutečnou hodnotu odhadovaného parametru
- úzká souvislost s příslušným testem

Párový problém

- na každém subjektu měříme dvě veličiny
- otázka: Mají tyto dvě veličiny stejnou střední hodnotu? Neboli, jsou co do polohy stejné?

Příklady:

- Věk rodičů: Jsou otcové starší než matky?
- Účinnost redukční diety: Je hmotnost po dietě nižší než před ní?
- Výška rodičů a dětí: Jsou synové vyšší než jejich otcové?
- Úspěšnost reklamní kampaně: Je prodejnost výrobku vyšší po kampani než před ní?
- Jsou dvojčata stejně inteligentní?
- ...

Matematický zápis

- párová pozorování $(X_1, Y_1), \ldots, (X_n, Y_n)$
  nezávislé dvojice náhodných veličin
  náhodný výběr z dvourozměrného rozdělení
- X_i a Y_i měřeny na stejném subjektu i
- příklady: věk matky a věk otce, ...
- $\mu_X = \mathrm{E}X_i$, $\mu_Y = \mathrm{E}Y_i \rightsquigarrow$ chceme otestovat hypotézu

$$H_0 : \mu_X = \mu_Y \text{ proti } H_1 : \mu_X \neq \mu_Y.$$

(příp. proti jednostranným H_1)

Párový t-test

Idea:

- zavedeme $Z_i = X_i - Y_i$ rozdíly
  (např. rozdíl věku rodičů)
- předpoklad $Z_1, \dots, Z_n$ stejné rozdělení $\leftrightsquigarrow$ normální
- zjevně $\mu_Z = \mu_X - \mu_Y$, a proto

$$H_0 : \mu_X = \mu_Y \text{ platí} \quad \Leftrightarrow \quad \text{platí } \mu_Z = 0$$

- střední hodnota X_i a Y_i je stejná $\Leftrightarrow$ X_i kolísají kolem nuly $\rightsquigarrow$ úloha převedena na jednovýběrový test

Párový t-test

- definujeme $Z_i = X_i - Y_i$, $i = 1, \dots, n$
- předpoklad: $Z_1, \dots, Z_n$ náhodný výběr z $\mathrm{N}(\mu_Z, \sigma^2)$
- hypotézy

$$H_0 : \mu_Z = 0 \quad \text{proti} \quad H_1 : \mu_Z \neq 0$$

- jednovýběrový t-test: spočteme $\overline{Z}$ odhad μ_Z, S^2 odhad σ^2 $\rightsquigarrow$ testová statistika

$$T_n = \sqrt{n}\frac{\overline{Z}}{S} = \sqrt{n}\frac{\overline{X} - \overline{Y}}{S}$$

H_0 zamítáme, pokud

$$|T_n| > t_{n-1}(1 - \alpha/2)$$

Další varianty testu

$H_0 : \mu_Z = 0$ proti $H_1 : \mu_Z > 0$

- zamítáme H_0, pokud

$$T_n > t_{n-1}(1-\alpha)$$

$H_0 : \mu_Z = 0$ proti $H_1 : \mu_Z < 0$

- zamítáme H_0, pokud

$$T_n < -t_{n-1}(1-\alpha)$$

Obecnější hypotézy:

- lze testovat obecněji $H_0 : \mu_X - \mu_Y = \delta$
- testová statistika: $T_n = \sqrt{n}\,\frac{\overline{Z}_n - \delta}{S}$

Předpoklad normality

Porušení předpokladů:

- test dodržuje požadovanou hladinu α, pokud
  - Z_i mají normální rozdělení, nebo
  - počet pozorovaných dvojic n je dost velký ($n > 50$)
- jestliže normalitu nelze předpokládat
  - je-li n dost velké $\rightsquigarrow$ lze párový t-test
  - je-li n malé $\rightsquigarrow$ párový test může dávat nesprávné výsledky nutné použít jiný postup (Wilcoxonův párový test)

Příklad — věk otce vs. věk matky

Otázka: Jsou otcové studentů starší než matky studentů?

- $n = 256$ studentů z let $2006-2011$ $\rightsquigarrow$ sledujeme věk otce a věk matky
- X - věk otce, Y - věk matky, $Z = X - Y$ rozdíl věků
- test $H_0 : \mu_Z = 0$ proti $H_1 : \mu_Z > 0$ na hladině $\alpha = 0.05$
- vypočteme $\overline{X} = 48.88$, $\overline{Y} = 46.60$, $\overline{Z} = 2.28$, $S = 4.12$ testová statistika

$$T_n = \sqrt{256}\,\frac{2.28}{4.12} = 8.85$$

- kritická hodnota $t_{255}(0.95) = 1.65$

Příklad — věk otce vs. věk matky

- $T_n = 8,85 > t_{255}(0.95) = 1.65 \Rightarrow$ zamítáme hypotézu $H_0 : \mu_X = \mu_Y$ ve prospěch $H_1 : \mu_X > \mu_Y$
- p-hodnota $< 10^{-16}$
- Závěr: Prokázali jsme, že střední věk otců je statisticky významně vyšší než střední věk matek

Ověření předpokladu normality:

- graficky— histogram, QQ graf
- Shapirův-Wilkův test: p-hodnota $6 \cdot 10^{-14}$
- normalitu dat nelze předpokládat; nicméně n dostatečně vysoké $\rightsquigarrow$ párový t-test lze použít

Příklad – Věk otce vs. věk matky

95 % intervalový odhad rozdílu věku rodičů:

- obecný vzorec

$$\left(\overline{Z} - \frac{S}{\sqrt{n}} t_{n-1}(1-\alpha/2), \overline{Z} + \frac{S}{\sqrt{n}} t_{n-1}(1-\alpha/2)\right)$$

  dosadíme:

$$(1.771, 2.784)$$

  interval, který s pravděpodobností 95 % pokryje skutečný rozdíl středních hodnot věku rodičů

- levostranný 95% interval spolehlivosti

$$(1.855, \infty)$$

  0 zde neleží $\rightsquigarrow$ výsledek testu

Dvouvýběrový problém

- jedna veličina měřená ve dvou nezávislých skupinách
- m nezávislých pozorování X_i a n nezávislých pozorování Y_j navzájem nezávislé
- zajímá nás porovnání jejich středních hodnot

Příklad:

- výška mužů a žen $\leftrightsquigarrow$ jsou muži vyšší než ženy? (je v jejich průměrné výšce systematický rozdíl?)
- plat mužů a žen $\leftrightsquigarrow$ je plat mužů stejný jako plat žen? (je v platech mužů a žen rozdíl, který se projevuje ve střední hodnotě?)
- liší se výše cholesterolu u kuřáků a nekuřáků?

Matematický zápis

Model:

- dva nezávislé náhodné výběry
  $X_1, \ldots, X_m$ z normálního rozdělení $\mathrm{N}(\mu_X, \sigma_X^2)$
  $Y_1, \ldots, Y_n$ z normálního rozdělení $\mathrm{N}(\mu_Y, \sigma_Y^2)$
- předpoklad: shodné rozptyly $\sigma_X^2 = \sigma_Y^2$

Chceme otestovat

$$H_0 : \mu_X = \mu_Y \text{ proti } H_1 : \mu_X \neq \mu_Y$$

(resp. proti jednostranným alternativám)

Test: dvouvýběrový t-test

Dvouvýběrový t-test: odvození

Idea:

- porovnáme průměry $\overline{X}$ a $\overline{Y}$
  velký rozdíl $\rightsquigarrow$ zamítnutí hypotézy H_0
- je třeba brát v úvahu také rozsahy výběrů a rozptyl

Testová statistika:

$$T = \frac{\overline{X} - \overline{Y}}{\text{S.E.}(\overline{X} - \overline{Y})} = \sqrt{\frac{mn}{m+n}}\,\frac{\overline{X}_m - \overline{Y}_n}{S},$$

kde S je společný odhad rozptylu σ^2 spočítaný z obou výběrů

$$S^2 = \frac{1}{m+n-2}\left[(m-1)S_X^2 + (n-1)S_Y^2\right]$$

a S.E.() značí odhad směrodatné odchylky

Dvouvýběrový t-test: odvození

Společný odhad rozptylu:

- umíme odhadnout σ^2 z každého výběru zvlášť pomocí výběrových rozptylů

$$S_X^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \overline{X}_m)^2$$

$$S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2$$

- vezmeme vážený průměr

$$S_{m,n}^2 = \frac{1}{m+n-2}\left[(m-1)S_X^2 + (n-1)S_Y^2\right]$$

Rozdělení testové statistiky

Pak za $H_0 : \mu_X = \mu_Y$ má testová statistika

$$T = \sqrt{\frac{mn}{m+n}} \frac{\overline{X}_m - \overline{Y}_n}{S},$$

t_{m+n-2} rozdělení, tj. t-rozdělení s $m + n - 2$ stupni volnosti.

$H_0 : \mu_X = \mu_Y$ zamítáme ve prospěch $H_1 : \mu_X \neq \mu_Y$, pokud

$$|T| > t_{m+n-2}\left(1 - \frac{\alpha}{2}\right)$$

zamítáme-li H_0, říkáme, že rozdíl ve výběrových průměrech je statisticky významný

Dvouvýběrový t-test:

$H_0 : \mu_X = \mu_Y$ zamítáme ve prospěch alternativy

- $H_1 : \mu_X > \mu_Y$, pokud $T > t_{m+n-2}\big(1-\alpha\big)$
- $H_1 : \mu_X < \mu_Y$, pokud $T < -t_{m+n-2}\big(1-\alpha\big)$

Poznámka

- lze obecnější hypotéza $H_0 : \mu_X - \mu_Y = \delta \rightsquigarrow$ testová statistika

$$T = \sqrt{\frac{mn}{m+n}}\,\frac{\overline{X}_m - \overline{Y}_n - \delta}{S}$$

Ověření předpokladů

Normalita

- ověření normality pro každý výběr zvlášť
- pro velká n, m porušení normality velmi nevadí

Shoda rozptylů

- S_X^2 a S_Y^2 podobné
- F-test shody rozptylů $\rightsquigarrow H_0 : \sigma_X^2 = \sigma_Y^2$ proti $H_1 : \sigma_X^2 \neq \sigma_Y^2$
- pochyby o shodě $\rightsquigarrow$ Welchův test (modifikace t-testu pro nestejné rozptyly)

Welchův t-test:

- Model: nezávislé výběry $X_1, \ldots, X_m$ z $\mathrm{N}(\mu_X, \sigma_X^2)$ a $Y_1, \ldots, Y_n$ z $\mathrm{N}(\mu_Y, \sigma_Y^2)$
- testová statistika
$$T = \frac{\overline{X}_m - \overline{Y}_n}{\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}}$$
(jiný jmenovatel $\rightsquigarrow$ jiný odhad S.E.($\overline{X} - \overline{Y}$))
- za nulové hypotézy má T přibližně t-rozdělení s ν stupni volnosti, kde ν je (necelé číslo), které se počítá z S_X^2/m a S_Y^2/n
- je-li rozptyl ve výběrech shodný, je vhodnější použít standardní dvouvýběrový t-test

Příklad – plat

Problém: Je plat mužů vyšší než plat žen?

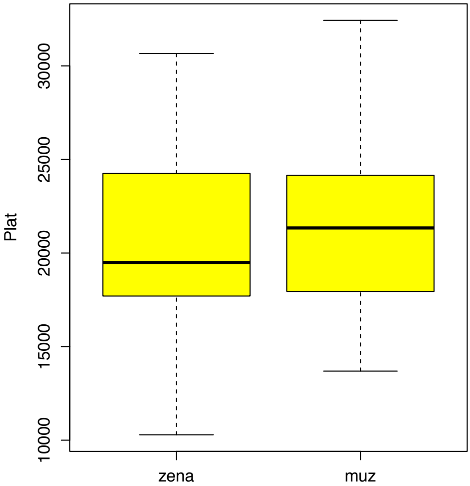
- 100 náhodně vybraných zaměstnanců ⇝ měsíční plat v Kč
- 35 žen a 65 mužů
- X – plat žen, Y – plat mužů

	rozsah	průměr	směr. odchylka
ženy	35	20 686	5 180
muži	65	21 364	4 334

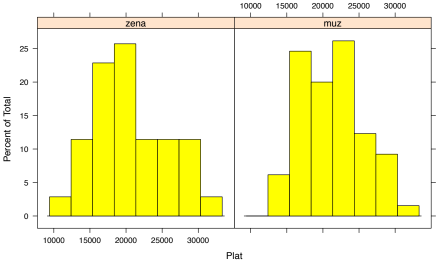
Předpoklady:

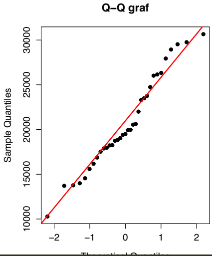
- normalita muži ⇝ p-hodnota 0.134
- normalita ženy ⇝ p-hodnota 0.310
- test shody rozptylů ⇝ p-hodnota 0.218

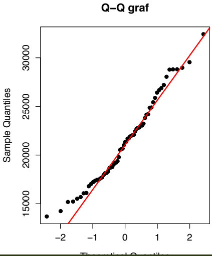
Příklad – grafické znázornění



Příklad – předpoklady







Příklad – řešení

- $H_0 : \mu_X = \mu_Y$ proti $H_1 : \mu_X < \mu_Y$
- společný odhad rozptylu

$$S^2 = \frac{35-1}{35+65-1}5180^2 + \frac{65-1}{35+65-1}4334^2 = 21\,574\,090$$

- testová statistika

$$T = \sqrt{\frac{35 \cdot 65}{100}} \cdot \frac{20686 - 21364}{\sqrt{21\,574\,090}} = -0.697$$

- kritická hodnota $-t_{98}(0.95) = -1.661$
- na základě našich dat nelze zamítnout H_0

Příklad – řešení

Řešení v programu R:

```
> t.test(zeny,muzi,var.equal=T,alternative=‘‘less‘‘)
                     Two Sample t-test
data: zeny and muzi
t = -0.6971, df = 98, p-value = 0.2437
alternative hypothesis: true difference in means is less
than 0
95 percent confidence interval:
-Inf 938.2113
sample estimates:
mean of x mean of y
20685.51 21364.37
```

Shrnutí

Testy o střední hodnotě

1. jeden výběr
   - jednovýběrový t-test
   - normalita
     (není nezbytné při dostatečně velkém rozsahu výběru)
2. párová pozorování
   - párový t-test
   - normalita rozdílu
     (není nezbytné při dostatečně velkém rozsahu výběru)
3. dva nezávislé výběry
   - dvouvýběrový t-test
   - nezávislost
   - normalita
     (není nezbytné při dostatečně velkém rozsahu výběru)
   - shoda rozptylů
     (neplatí-li, lze použít Welshův test)

Porušení normality

Jestliže nelze normalitu předpokládat a rozsah výběru je malý

- nutné použít jiné testy, které předpoklad normality nepotřebují
- neparametrické testy
- založeny na pořadí $\rightsquigarrow$ pořadové testy

Uvedeme si

- jednovýběrový Wilcoxonův test
- dvouvýběrový Wilcoxonův test