Matematická statistika

Šárka Hudecová

Katedra pravděpodobnosti a matematické statistiky
Matematicko-fyzikální fakulta Univerzity Karlovy

letní semestr 2012

Testování hypotéz – motivační příklad

Matematická statistika
Testování hypotéz
Test o střední hodnotě normálního rozdělení při známém σ^2
Jednovýběrový t-test
Interval spolehlivosti pro střední hodnotu

Příklad (Platová diskriminace)

- firma provedla šetření s cílem zmapovat platové podmínky a spokojenost zaměstnanců
- jednou z otázek bylo zjistit, zda dochází k platové diskriminaci žen

- do studie zahrnuto 100 náhodně vybraných zaměstnanců, z toho 35 žen a 65 mužů
  - měsíční plat žen $\rightsquigarrow$ $\overline{X}_{35} = 20\,685.5$ Kč, $S_X = 5\,179.5$ Kč
  - měsíční plat mužů $\rightsquigarrow$ $\overline{Y}_{65} = 21\,364.4$ Kč, $S_Y = 4\,334,0$ Kč
- Je rozdíl $\overline{Y} - \overline{X} = 678.9 > 0$ Kč dostatečně průkazný na to, abychom mohli tvrdit, že muži mají (v dané firmě) obecně vyšší platy než ženy?

Testování hypotéz – motovační příklad

Matematická statistika
Testování hypotéz
Test o střední hodnotě normálního rozdělení při známém σ^2
Jednovýběrový t-test
Interval spolehlivosti pro střední hodnotu

Otázka: Mají muži vyšší příjem než ženy?

- přesnější formulace $\rightsquigarrow$ zajímá nás zřejmě porovnání středních hodnot platů mužů a žen (nikoliv, zda libovolný muž vydělává více než libovolná žena)
- je střední hodnota platů vyšší u mužů než u žen?
- víme: střední hodnoty odhadujeme výběrovými průměry $\rightsquigarrow$ je rozumné rozhodnout na základě porovnání $\overline{X}$ a $\overline{Y}$
  - jiný náhodný výběr by zahrnul jiných 100 zaměstnanců $\rightsquigarrow$ dostali bychom odlišné výběrové průměry $\overline{X}$ a $\overline{Y}$
  - výběrové průměry $\overline{X}$ a $\overline{Y}$ jsou náhodné veličiny
  - jejich hodnoty neodpovídají středním hodnotám $\mathrm{E}\,X$ a $\mathrm{E}\,Y$ přesně
- vyšlo nám $\overline{Y} - \overline{X} = 678.9 > 0$ Kč $\leftrightsquigarrow$ opravňuje nás to tvrdit, že $\mathrm{E}\,Y - \mathrm{E}\,X > 0$?

Testování hypotéz – motovační příklad

Matematická statistika
Testování hypotéz
Test o střední hodnotě normálního rozdělení při známém σ^2
Jednovýběrový t-test
Interval spolehlivosti pro střední hodnotu

Otázka: Kdy se výběrové průměry liší „dostatečně“?

- když vezmeme dva výběry z téhož rozdělení $\rightsquigarrow$ výběrové průměry se budou lišit (byť jen málo), i když jsou střední hodnoty stejné
- příklad s platy: otázka je, zda je rozdíl v průměrech jen vlivem „náhody“ nebo se skutečně liší střední hodnoty
- pokud oba výběrové průměry odhadují tutéž střední hodnotu $\rightsquigarrow$ průměry by se neměly lišit velmi
- je potřeba zohlednit:
  - $\rightsquigarrow$ počet pozorování (více dat $\leftrightsquigarrow$ větší přesnost odhadů)
  - $\rightsquigarrow$ variabilitu (vysoký rozptyl = větší nejistota)
- vezmeme v úvahu rozdělení $\overline{Y} - \overline{X}$ a z něho zjistíme, jaké hodnoty jsou již „extrémní a málo pravděpodobné“ $\leftrightsquigarrow$ statistické testy

Testování hypotéz

Testování hypotéz

= vyhodnocování pravdivostní hodnoty výroků na základě náhodného výběru
(tj. ověřování platnosti nějakého výroku)

– provádíme pomocí statistických testů

Hypotéza

= výrok, o jehož pravdivosti chceme rozhodnout

- **nulová hypotéza** H_0
  - tvrzení o populaci, o jehož platnosti rozhodujeme
  - (není rozdíl, nezávisí, neliší se, ...)
- **alternativní hypotéza** H_1:
  - alternativa (doplňující možnost) k H_0
  - často tvrzení, které chceme prokázat

Statistický test

Statistický test = rozhodovací pravidlo, na jehož základě zamítáme nebo nezamítáme H_0

- *testová statistika* $T_n = T_n(X_1, \dots, X_n)$ = náhodná veličina, která je funkcí pozorování $X_1, \dots, X_n$
- *kritický obor* C = možné výsledky pokusu, kdy H_0 zamítáme

Rozhodovací pravidlo:

- pokud $T_n \in C \rightsquigarrow$ zamítáme hypotézu H_0 ve prospěch alternativy H_1
(*naše data svědčí proti* H_0 *a prokazujeme tak platnost* H_1)
- pokud $T_n \notin C \rightsquigarrow$ hypotézu H_0 nezamítáme
(*na základě našich dat nelze zamítnout* H_0, *naše data nejsou v rozporu s* H_0)

Pozor: nesymetrie mezi H_0 a H_1

Chyba I. a II. druhu

- rozhodujeme na základě náhodného výběru $\rightsquigarrow$ nemůžeme testovanou otázku zodpovědět s absolutní jistotou
- můžeme se dopustit chyby $\rightsquigarrow$ tyto chyby se budeme snažit omezit (resp. kontrolovat jejich pravděpodobnosti)

	H_0 zamítáme	H_0 nezamítáme
H_0 platí	chyba 1. druhu	OK
H_0 neplatí	OK	chyba 2. druhu

Označíme:

$$\alpha = \mathrm{P}(\text{chyba 1. druhu}) = \mathrm{P}(\text{zamítáme } H_0 |\ H_0 \text{ platí})$$
$$\beta = \mathrm{P}(\text{chyba 2. druhu}) = \mathrm{P}(\text{nezamítáme } H_0 |\ H_0 \text{ neplatí})$$

Přirozený požadavek: $\alpha, \beta \to \min \leftrightsquigarrow$ bohužel nelze současně

Chyba I. a II. druhu

- zvolíme **hladinu testu** α (zpravidla $\alpha = 0.05$)
  - maximální dovolená pst chyby 1. druhu
  - maximální pst falešného prokázání vědecké hypotézy
  - chyba 1.druhu je závažnější (falešně něco prokazujeme)
  - volíme před pokusem, nezávisle na jeho výsledku
- pro dané α chceme minimální $\beta \leftrightsquigarrow$ maximální $1 - \beta$
- **síla testu** $1 - \beta$
  - pst zamítnutí neplatné H_0
  - pst, s jakou prokážeme platnou vědeckou hypotézu H_1
  - nemáme pod kontrolou (závisí na tom, co opravdu platí)
  - můžeme ovlivnit volbou statistického testu, počtem pozorování, ...

$\rightsquigarrow$ α máme plně pod kontrolou, o β toho moc nevíme

Dosažená hladina testu

Dosažená hladina testu p-hodnota (angl. p-value)

- pravděpodobnost, že dostaneme výsledek, který stejně nebo ještě méně podporuje H_0, jestliže H_0 platí
- nejmenší hladina α, na které lze ještě H_0 zamítnout
- „stupeň důvěry" v platnost H_0
- výsledek provedení statistického testu pomocí softwaru

Pravidlo:

- je-li $p \leq \alpha \rightsquigarrow$ zamítáme H_0
- je-li $p > \alpha \rightsquigarrow$ nezamítáme H_0

(Zapamatovat!)

Testování hypotéz – shrnutí

Matematická statistika

Testování hypotéz

Test o střední hodnotě normálního rozdělení při známém σ^2

Jednovýběrový t-test

Interval spolehlivosti pro střední hodnotu

Při testování můžeme

- dojít ke správnému rozhodnutí
- udělat chybu 1. druhu (zamítnout platnou H_0) anebo 2. druhu (nezamítnout neplatnou H_0)

Postup:

- stanovíme dostatečně nízkou pravděpodobnost chyby 1. druhu α (hladinu testu)
  (chyba 1. druhu je závažnější $\rightsquigarrow$ její pst máme pod kotrolou)
- pravděpodobnost chyby 2. druhu β závisí na okolnostech (volba testu, počet pozorování, variabilita, ...) a její hodnotu v praxi neznáme

Proto: H_0 a H_1 nejsou posuzovány symetricky

Nesymetrie H_0 a H_1

Matematická statistika

Testování hypotéz

Test o střední hodnotě normálního rozdělení při známém σ^2

Jednovýběrový t-test

Interval spolehlivosti pro střední hodnotu

H_0 a H_1 nejsou posuzovány symetricky:

- H_0 považujeme a priori za platnou a zamítáme ji jen tehdy, pokud k tomu máme dostatečně silné důvody
- pokud jsme zamítli $H_0 \rightsquigarrow$ můžeme tvrdit, že data svědčí o tom, že H_0 neplatí (a prokazujeme platnost H_1)
- pokud jsme H_0 nezamítli $\rightsquigarrow$ pak
  - buď H_0 opravdu platí
  - anebo H_0 neplatí, ale data neposkytují dostatečné „důkazy" k jejímu zamítnutí (malá síla testu)

  $\rightsquigarrow$ nutné volit opatrné formulace závěrů (*hypotézu H_0 nelze na základě našich dat zamítnout apod.)*

Závěr

Hypotézu H_0 nemůžeme prokázat, ale pouze vyvrátit

Filozofie testování hypotéz

Matematická statistika

Testování hypotéz

Test o střední hodnotě normálního rozdělení při známém σ^2

Jednovýběrový t-test

Interval spolehlivosti pro střední hodnotu

soudní proces ctící princip presumpce neviny.

H_0: „Obžalovaný je nevinen".
H_1: „Obžalovaný je vinen".

- zamítneme-li $H_0 \rightsquigarrow$ odsoudíme obžalovaného k trestu
- nezamítneme-li $H_0 \rightsquigarrow$ obžalovaný je propuštěn
- data $\leftrightsquigarrow$ důkazy svědčící pro nebo proti vině
- testová statistika $\leftrightsquigarrow$ soudce
- kritický obor $\leftrightsquigarrow$ důkazní břemeno = množství důkazů nutné k odsouzení
- hladina testu $\alpha \leftrightsquigarrow$ pst odsouzení nevinného
- síla testu $1-\beta \leftrightsquigarrow$ pst odsouzení skutečného pachatele

Filozofie testování hypotéz

Hladinu α zvolíme malou (např. 5 %)

- chráníme nevinné před odsouzením
- čím menší α $\rightsquigarrow$ vyšší důkazní břemeno $\rightsquigarrow$ řada viníků propuštěna pro nedostatek důkazů

Výsledek procesu:

- odsouzení obžalovaného $\rightsquigarrow$ lze tvrdit, že je vinen
- propuštění obžalovaného $\rightsquigarrow$ buď opravdu nevinen, anebo vinen, ale soud neměl dostatečné důkazy k jeho odsouzení

K odsouzení viníka může dopomoci (bez porušení presumpce neviny):

- schopný soudce (vhodná testová statistika),
- dodatečné množství důkazů (více dat)
- málo chyb a omylů v důkazním materiálu (malý rozptyl)

Statistické testy

Matematická statistika
Testování hypotéz
Test o střední hodnotě normálního rozdělení při známém σ^2
Jednovýběrový t-test
Interval spolehlivosti pro střední hodnotu

- statistických testů je obrovské množství
- uvedeme si jen několik vybraných základních
- podrobné odvození testové statistiky a kritického oboru jen u některých
- v praxi důležité:
  - výběr vhodného testu pro daný problém
  - provedení testu
  - ověření předpokladů testu (např. normalita dat, ...)
  - správná interpretace výsledku

Test o střední hodnotě v normálním rozdělení

Matematická statistika
Testování hypotéz
Test o střední hodnotě normálního rozdělení při známém σ^2
Jednovýběrový t-test
Interval spolehlivosti pro střední hodnotu

Situace: $X_1, \dots, X_n$ náhodný výběr z normálního rozdělení $\mathrm{N}(\mu, \sigma^2)$, kde $\sigma > 0$ známe; chceme testovat

$$H_0 : \mu = \mu_0,$$

kde μ_0 je nějaká známá (předepsaná) hodnota, proti alternativě

$$H_1 : \mu > \mu_0.$$

Chceme: statistický test $\leftrightsquigarrow$ testová statistka a kritický obor

Jiné možné alternativy:

- $H_1 : \mu \neq \mu_0$ tzv. oboustranná alternativa
- $H_1 : \mu < \mu_0$ tzv. jednostranná alternativa

(volba alternativy dle zadání úlohy – vyplývá z její formulace)

Odvození testu

Matematická statistika
Testování hypotéz
Test o střední hodnotě normálního rozdělení při známém σ^2
Jednovýběrový t-test
Interval spolehlivosti pro střední hodnotu

Víme:

- střední hodnotu μ odhadujeme pomocí výběrového průměru $\overline{X}_n$
- rozdělení $\overline{X}_n$ je $\mathrm{N}(\mu, \frac{\sigma^2}{n})$, a proto

$$\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \sim \mathrm{N}(0, 1)$$

Proti H_0 a ve prospěch H_1 svědčí hodnoty $\overline{X}_n >> \mu_0$

- zvolíme hladinu testu α
- budeme hledat „mezní hodnotu“ K tak, že pro $\overline{X}_n > K$ už zamítneme H_0
- musí platit

$$\mathrm{P}(\text{ chyba 1.druhu }) = \mathrm{P}(\overline{X}_n > K | H_0 \text{ platí}) = \alpha$$

Odvození testu – pokrač.

Postupně upravíme:

$$\alpha = \mathrm{P}(\overline{X}_n > K | H_0 \text{ platí}) = \mathrm{P}(\overline{X}_n > K | \mu = \mu_0)$$
$$= \mathrm{P}(\underbrace{\sqrt{n}\frac{\overline{X} - \mu_0}{\sigma}}_{T_n \sim \mathrm{N}(0,1)} > \underbrace{\sqrt{n}\frac{K - \mu_0}{\sigma}}_{=C} | \mu = \mu_0) = 1 - \Phi(C)$$

Odtud

$$C = \Phi^{-1}(1-\alpha) = z_{1-\alpha}, \quad K = \dots$$

(Připomenutí: $z_{1-\alpha}$ je $1-\alpha$ kvantil normálního rozdělení)

Test o střední hodnotě normálního rozdělení při známém σ^2

Odvodili jsme test:

- testová statistika

$$T_n = \sqrt{n}\frac{\overline{X} - \mu_0}{\sigma}$$

- kritický obor $C = (z_{1-\alpha}, \infty)$
- je-li $T_n > z_{1-\alpha} \rightsquigarrow$ zamítáme H_0
- je-li $T_n \leq z_{1-\alpha} \rightsquigarrow$ nezamítáme H_0

Podobným postupem bychom odvodili test H_0 proti

- $H_1 : \mu < \mu_0 \rightsquigarrow$ kritický obor $T_n < -z_{1-\alpha}$
- $H_1 : \mu \neq \mu_0 \rightsquigarrow$ kritický obor $|T_n| > z_{1-\frac{\alpha}{2}}$

Zvolíme-li hladinu $\alpha = 0.05$ dostaneme $z_{1-\alpha} = z_{0.95} = 1.64$ a $z_{1-\alpha/2} = z_{0.975} = 1.96$.

⋆ Obecný postup

Testová statistika T_n

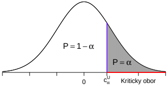
- musí být citlivá na parametr, který testujeme
- musíme znát její rozdělení za hypotézy H_0
- nesmí záviset na neznámých konstantách

Kritický obor C

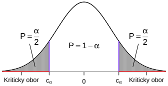
- musí splňovat $\alpha = \mathrm{P}(T_n \in C | H_0)$
- někdy se spokojíme s $\mathrm{P}(T_n \in C | H_0) \to \alpha$ pro $n \to \infty$ (tzv. asymptotické testy, testy na asymptotické hladině α)
- většinou tvaru $|T_n| > c_\alpha$ nebo $T_n > c_\alpha^U$ nebo $T_n < c_\alpha^L$
- $c_\alpha, c_\alpha^U, c_\alpha^L$ *kritické hodnoty* $\leftrightsquigarrow$ většinou spočteny jako vhodné kvantily rozdělení testové statistiky T_n

⋆ Nalezení kritického oboru

Kritická hodnotapro zamítání H_0 pro příliš velké hodnoty T_n:



Kritické hodnoty pro zamítání H_0 pro příliš malé anebo příliš velké hodnoty T_n:



Příklad

Příklad (Výška chlapců)

- v roce 1951 rozsáhlé měření výšky desetiletých chlapců $\rightsquigarrow$ populační parametry $\mu = 136.1$ cm, $\sigma = 6.4$ cm
- v roce 1961 provedeno měření na náhodném výběru 15 chlapců $\rightsquigarrow$ máme rozhodnout, zda se výška desetiletých chlapců zvýšila (tj. zda je nová generace vyšší)

- hodnoty zjištěné v roce 1961 $\rightsquigarrow$ $\overline{X}_{15} = 139.13$ cm
- zvolíme testovací hladinu $\alpha = 0.05$
- hypotézy: $H_0 : \mu = 136.1$ proti $H_1 : \mu > 136.1$
- hodnota testové statistiky

$$T = \sqrt{15}\,\frac{139.13 - 136.1}{6.4} = 1.836$$

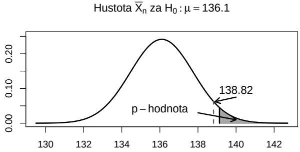
Příklad – pokrač.

- testová statistika $T_n = 1.836$
- kritická hodnota $z_{1-\alpha} = 1.645$
- máme $1.836 > 1.645$ $\rightsquigarrow$ zamítáme H_0 na hladině významnosti 0.05 ve prospěch H_1
- na 5% hladině jsme prokázali, že je nová generace vyšší
- v případě, že nová generace není vyšší (tj. platí H_0) $\rightsquigarrow$ riskovali jsme jen 5% pravděpodobnost, že budeme nesprávně tvrdit, že je vyšší

Příklad – pokrač.

Dosažená hladina testu (p-hodnota)

- pst za platnosti H_0, že dostaneme výsledek ještě „extrémněji" svědčící proti H_0 a ve prospěch H_1
- $\mathrm{P}(\sqrt{n}\frac{\overline{X}_n - \mu_0}{\sigma} > 1.836) = \mathrm{P}(\overline{X}_n > 139.13) = 0.033$
- p-hodnota je rovna 0.033 (je tedy nižší než $\alpha = 0.05$)



Jednovýběrový t-test – úvod

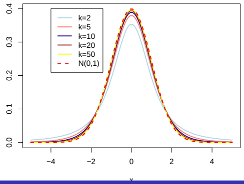
- dosud jsme předpokládali, že σ^2 je známé
- v praxi většinou σ^2 neznámé $\rightsquigarrow$ musíme jej odhadnout pomocí S_n^2
(připomenutí: $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$)
- testová statistika (místo σ použijeme S_n)

$$T_n = \sqrt{n}\,\frac{\overline{X}_n - \mu_0}{S_n}$$

- rozdělení T_n za H_0 není normální $\mathrm{N}(0,1)$, ale nazývá se t-rozdělení s $n-1$ stupni volnosti, značíme jej t_{n-1} rozdělení

t-rozdělení

- hustota rozdělení t_k má pro velké k tvar velmi podobný normovanému normálnímu rozdělení
- pro malé k má rozptyl větší než 1 a kritické hodnoty jsou vzdálenější od 0



Kvantily t-rozdělení

- α kvantil budeme značit $t_k(\alpha)$

Tabulka: Vybrané kvantily rozdělení t_k.

	α				
k	0.9	0.95	0.975	0.99	0.995
5	1.476	2.015	2.571	3.365	4.032
10	1.372	1.812	2.228	2.764	3.169
20	1.325	1.725	2.086	2.528	2.845
50	1.299	1.676	2.009	2.403	2.678
100	1.290	1.660	1.984	2.364	2.626
N(0,1)	1.282	1.645	1.960	2.326	2.576

Jednovýběrový t-test

Situace: $X_1, \dots, X_n$ náhodný výběr z rozdělení $N(\mu, \sigma^2)$, chceme otestovat hypotézu $H_0 : \mu = \mu_0$ proti (oboustranné) alternativě $H_1 : \mu \neq \mu_0$.

Testová statistika:

$$T_n = \sqrt{n}\frac{\overline{X}_n - \mu_0}{S_n}$$

Kritický obor

$$|T_n| > t_{n-1}(1 - \frac{\alpha}{2}) \implies \text{zamítáme } H_0,$$

kde $t_{n-1}(1 - \frac{\alpha}{2})$ je $(1 - \frac{\alpha}{2})$-kvantil rozdělení t_{n-1}

Tento test nazýváme *jednovýběrový t-test* na hladině α.

Další možné alternativy

Lze uvažovat také jednostranné alternativy:

- $H_1 : \mu > \mu_0 \rightsquigarrow$ kritický obor $T_n > t_{n-1}(1 - \alpha)$
- $H_1 : \mu < \mu_0 \rightsquigarrow$ kritický obor $T_n < -t_{n-1}(1 - \alpha)$

Poznámky

- t-rozdělení se někdy také nazývá *Studentovo* t-rozdělení
- William Sealy Gosset (1876–1937) $\leftrightsquigarrow$ chemik pracující v pivovaru Guinness
- 1908 odvození jednovýběrového t-testu (pseudonym Student)

Příklad

Uvažujme příklad s výškou chlapců pro neznámé σ^2

- opět testujeme $H_0 : \mu = 136.1$ proti $H_1 : \mu > 136.1$ na hladině $\alpha = 0.05$
- kritická hodnota $t_{14}(0.95) = 1.761$
- máme $\overline{X} = 139.133$, $S^2 = 6.556^2$, odtud

$$T_n = \sqrt{15}\,\frac{139.133 - 136.1}{6.556} = 1.792$$

  a tudíž $1.792 > 1.761 \rightsquigarrow$ zamítáme H_0 ve prospěch H_1 na hladině 5 %
- p-hodnota je rovna $\mathrm{P}(T > 1.792) = 0.047 < 0.05$
- na 5% hladině významnosti jsme prokázali, že nová generace desetiletých chlapců je statisticky významně vyšší

Příklad — pokrač.

- kdybychom dopředi neměli určenou jednostrannou alternativu $\rightsquigarrow$ museli bychom brát $H_1 : \mu \neq 136.1$
- hodnotu $|T_n| = 1.792$ bychom porovnávali s $t_{14}(0.975) = 2.145$
- zde $|T_n| = |1.792| < 2.145 \rightsquigarrow H_0$ na hladině výynamnosti 5 % nelze zamítnout
- p-hodnota 0.0948
- interpretace: *nelze prokázat, že by se výška nové generace statisticky významně změnila*

výsledek testu tedy zjevně závisí, proti které alternativě H_1 testujeme

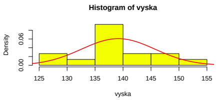
Ověření předpokladu o normálním rozdělení

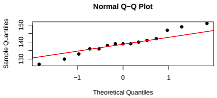
Jak ověříme, že data pocházejí skutečně z normálního rozdělení

- graficky
  - porovnáme histogram s hustotou normálního rozdělení
  - Q-Q graf (diagram normality) $\rightsquigarrow$ porovnává výběrové kvantily s teoretickými kvantily normálního rozdělení
- statistickým testem
  - např. Shaprův-Wilkův test
  - H_0 : výběr pochází z normálního rozdělení
    H_1 : výběr nepochází z normálního rozdělení
  - `shapiro.test`: p-hodnota $\rightsquigarrow$ zamítáme nebo nezamítáme normalitu $\rightsquigarrow$ lze nebo nelze použít jednovýběrový t-test

Příklad – výška chlapců

- Shapirův-Wilkův test: p-hodnota $0.8 > 0.05 \rightsquigarrow$ normalitu nelze zamítnout (a je proto možné použít jednovýběrový t-test)





Porušení normality

Situace: $X_1, \dots, X_n$ výběr z (jiného než normálního) rozdělení se střední hodnotou μ a konečným rozptylem, chceme testovat $H_0: \mu = \mu_0$ proti alternativě $H_1: \mu \neq \mu_0$

Jednovýběrový *t*-test

- předpoklad normality není splněn $\rightsquigarrow$ skutečná hladina testu není rovna předepsanému α (a můžeme se tedy dopouštět daleko větší chyby)
- použijeme-li $\rightsquigarrow$ hladinu testu se blíží k α pro $n \to \infty$ (zdůvodnění: centrální limitní věta)
- při velkém počtu pozorování n je hladina testu $\approx \alpha$

Závěr: Jednovýběrový t-test je možné použít k testování střední hodnoty nenormálních dat, pokud je pozorování dostatečně mnoho (řekněme alespoň 50).

Porušení normality

Co dělat v případě, že normální rozdělení nelze pro naše data uvažovat?

- použijeme *t*-test
  - výsledek interpretujeme „asymptoticky" (ve smyslu, že test je proveden na hladině, která je asymptoticky rovna α)
  - nutné mít dostatečný počet pozorování
- použijeme jiný (např. neparametrický) test

Interval spolehlivosti

Situace: $X_1, \dots, X_n$ náhodný výběr z rozdělení se střední hodnotou μ, zajímá nás její odhad

- $\overline{X}_n$ je dobrý odhad střední hodnoty
- jediné číslo $\leftrightsquigarrow$ *bodový odhad*
- nedává nám představu o přesnosti tohoto odhadu
- chtěli bychom najít *interval*, který nám řekne, v jakém rozmezí se μ může nacházet
- zkonstruujeme tzv. *intervalový odhad*, též *interval spolehlivosti* pro μ nebo *konfidenční interval* pro μ
- interval spočítáme z pozorovaných dat tak, aby pokrýval neznámou hodnotu μ s předepsanou pravděpodobností, např. 95 %

Interval spolehlivosti pro střední hodnotu při známém σ^2

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

- víme $\overline{X} \sim \mathrm{N}(\mu, \sigma^2/n)$ a proto (už jsme viděli dříve)
$$0.95 = \mathrm{P}\left(\sqrt{n}\frac{|\overline{X} - \mu|}{\sigma} < z_{0.975}\right)$$
- 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

- $\overline{X}$ bodový odhad μ
- intervalový odhad $\leftrightsquigarrow$ interval s náhodnými mezemi
- 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
- obecněji: interval spolehlivosti pro μ na hladině $1-\alpha$:

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

  pokryje skutečnou hodnotu μ s pstí $1-\alpha$
- pozor na špatnou interpretaci!

Matematická statistika
Testování hypotéz
Test o střední hodnotě normálního rozdělení při známém σ^2
Jednovýběrový t-test
Interval spolehlivosti pro střední hodnotu

Příklad – výška chlapců

- měli jsme $n = 15$, $\bar{X} = 139.13$, $\sigma = 6.4$
- 95% interval spolehlivosti pro střední výšku všech desetiletých chlapců

$$\left(139.13 - \frac{6.4}{\sqrt{15}}\cdot 1.96, 139.13 + \frac{6.4}{\sqrt{15}}\cdot 1.96\right) = (135.9; 142.3)$$

- průměr výšek všech desetiletých chlapců leží s pstí 95 % v rozmezí 135.9 cm až 142.3 cm
- příklady nesprávné interpretace:
  - *95% desetiletých chlapců má výšku v rozmezí 135.9 cm až 142.3 cm*
  - *výběrový průměr výšek leží s pravděpodobností 95 % v intervalu 135.9 cm až 142.3 cm*

Matematická statistika
Testování hypotéz
Test o střední hodnotě normálního rozdělení při známém σ^2
Jednovýběrový t-test
Interval spolehlivosti pro střední hodnotu

Interval spolehlivosti pro střední hodnotu při neznámém σ

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

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

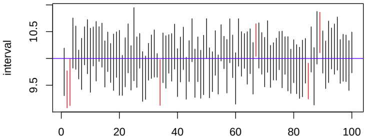
$$\mathrm{P}\left(\overline{X} - \frac{S_n}{\sqrt{n}} t_{n-1}(1-\frac{\alpha}{2}) < \mu < \overline{X} + \frac{S_n}{\sqrt{n}} t_{n-1}(1-\frac{\alpha}{2})\right) = 1-\alpha$$

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

Matematická statistika
Testování hypotéz
Test o střední hodnotě normálního rozdělení při známém σ^2
Jednovýběrový t-test
Interval spolehlivosti pro střední hodnotu

Interval spolehlivosti – ilustrace

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



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

Příklad

Výška chlapců (σ^2 neznámé)

- měli jsme $\bar{X} = 139.133$, $S_n = 6.556$
- 95% interval spolehlivosti pro střední hodnotu výšky chlapců

$$\left(139.133 - \frac{6.556}{\sqrt{15}} \cdot 2.145, 139.133 + \frac{6.556}{\sqrt{15}} \cdot 2.145\right)$$

$= (135.50; 142.76)$

- 99% interval spolehlivosti $\rightsquigarrow$ větší spolehlivost $\leftrightsquigarrow$ širší interval $\rightsquigarrow$ vyjde $(134.09; 144.17)$

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

Poznámka II.

- 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

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

- lze uvažovat *asymptotické* intervaly spolehlivosti = 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– výška chlapců (σ^2 neznámé)

Matematická statistika

Testování hypotéz

Test o střední hodnotě normálního rozdělení při známém σ^2

Jednovýběrový t-test

Interval spolehlivosti pro střední hodnotu

Oboustranný interval

- 95% interval spolehlivosti pro μ je $(135.5, 142.8)$
- s pravděpodobností 0.95 je skutečná hodnota μ v tomto intervalu $\rightsquigarrow$ jen 5 % riziko, že leží mimo něj
- 136.1 leží v tomto intervalu $\rightsquigarrow$ nelze zamítnout $H_0 : \mu = 136.1$ proti $H_1 : \mu \neq 136.1$ na hladině 5 %

Jednostranné intervaly

- 95% levostranný interval spolehlivosti:

$$\left(139.133 - \frac{6.556}{\sqrt{15}} \cdot 1.761, \infty\right) = (136.15, \infty)$$

- 136.1 neleží v tomto intervalu $\rightsquigarrow$ zamítáme $H_0 : \mu = 136.1$ proti $H_1 : \mu > 136.1$ na hladině 5 %
- 95 % pravostranný interval vyjde $(-\infty, 142.11)$ $\leftrightsquigarrow$ test $H_0 : \mu = 136.1$ proti $H_1 : \mu < 136.1$