11. cvičení z STP

29. dubna - 3. května 2019

Příklady 9.3, 10.1, 10.2 a 10.4.

Příklad 11.1 *Počet neúspěšných zásahů terče předtím, než se střelec trefí, má geometrické rozdělení* Geom(p). *Zaznamenali jsme, že terč byl zasažen*

20 *krát napoprvé*
10 *krát až napodruhé*
7 *krát až napotřetí*
3 *krát až napočtvrté.*

Metodou maximální věrohodnosti (příp. metodou momentů) odhadněte parametr p, představující pravděpodobnost úspěšného zásahu.

Řešení:
Veličina

$$X = \text{“}\textit{počet neúspěšných zásahů, než se trefíme}\text{”}$$

má geometrické rozdělení Geom(p) pro $p \in (0,1)$ a

$$P_p(X = i) = (1-p)^i p, \quad i \in \mathbb{N}_0 \ .$$

Z tabulky

hodnota i veličiny X	0	1	2	3
pozorovaná četnost n_i	20	10	7	3

vidíme, že počet měření je $n = \sum_i n_i = 20 + 10 + 7 + 3 = 40$. Naměřené hodnoty $(x_1, \ldots, x_n)$ se skládají z hodnot $i \in \{0, 1, 2, 3\}$, kde každá se vyskytuje se svojí četností n_i.

Metoda maximální věrohodnosti:
Hledáme hodnotu $p \in (0,1)$, která maximalizuje funkci věrohodnosti

$$L(p) = P_p(X_1 = x_1, \ldots, X_n = x_n) = \prod_{j=1}^{n} P_p(X_j = x_j) =$$

$$= \Big((1-p)^0 p\Big)^{20} \Big((1-p)^1 p\Big)^{10} \Big((1-p)^2 p\Big)^{7} \Big((1-p)^3 p\Big)^{3} =$$

$$= \ (1-p)^{0\cdot 20 + 1\cdot 10 + 2\cdot 7 + 3\cdot 3} \cdot p^{20+10+7+3} \ = \ (1-p)^{33} \cdot p^{40}$$

kde X_j jsou jednotlivé nezávislé veličiny (odpovídající jednotlivým pokusům) a x_j naměřené hodnoty.
Ekvivalentně budeme hledat maximum funkce

$$\ell(p) = \ln\big(L(p)\big) = 33 \cdot \ln(1-p) + 40 \cdot \ln p \ .$$

Z její derivace

$$\ell'(p) = -\frac{33}{1-p} + \frac{40}{p} = \frac{-73p + 40}{(1-p)p}$$

dostáváme řešení

$$\ell'(\hat{p}) = 0 \qquad \Longrightarrow \qquad \hat{p} = \frac{40}{73} \doteq 0.54795$$

které vyhovuje zadání, tj. $\hat{p} \in (0,1)$. Ze znamének derivace je snadno vidět, že v $\hat{p} = \frac{40}{73}$ je skutečně maximum.

Metoda momentů:

Porovnáváme teoretické k-té momenty $E(X^k)$ s jejich odhady $m_k = \frac{1}{n}\sum_{i=1}^n x_i^k$ pro prvních několik $k = 1, 2, \ldots$.

Střední hodnota geometrického rozdělení $X \sim \mathrm{Geom}(p)$ je

$$E(X) = \frac{1-p}{p}$$

a její odhad z realizace je

$$m_1 = \overline{\boldsymbol{x}} = \frac{\sum\limits_{j=1}^{n} x_j}{n} = \frac{\sum\limits_i i \cdot n_i}{\sum\limits_i n_i} = \frac{0 \cdot 20 + 1 \cdot 10 + 2 \cdot 7 + 3 \cdot 3}{20 + 10 + 7 + 3} = \frac{33}{40}\ .$$

Porovnáním dostaneme

$$\frac{1-\hat{p}}{\hat{p}} = E(X) = \overline{x} = \frac{33}{40}$$

což dává opět řešení

$$\hat{p} = \frac{40}{73} \doteq 0.54795$$

jako v předchozí metodě.

Jak je vidět, v případě geometrického rozdělení dostáváme pro jeho parametr p stejné výsledky pro obě metody.

Příklad 11.2 *Letecká společnost prodává letenky a chce co nejvíce utržit. Letadlo má 216 míst, ale ví se, že zhruba 5% lidí se k odletu nedostaví. Jaká je pravděpodobnost, že pokud společnost prodá 220 letenek, nepřesáhne počet cestujících kapacitu letadla?*

Řešení:
Pro veličinu

$$Z = \text{“počet cestujících (z těch, co si koupili letenku), kteří se dostaví k odletu”}$$

nás zajímá $P(Z \leq 216)$.

K řešení použijeme centrální limitní větu. Označme si tedy pro $i = 1, 2, ..., n$, kde $n = 220$, veličiny

$$X_i = \begin{cases} 1 & , \ i\text{-tý cestující se dostaví k odletu,} \\ 0 & , \ i\text{-tý cestující se nedostaví k odletu.} \end{cases}$$

Pokládáme je za nezávislé s alternativním rozdělením $X_i \sim \mathrm{Alt}(p) = \mathrm{Alt}(0.95)$. Protože platí $Z = \sum\limits_{i=1}^{n} X_i$

dostaneme

$$
\begin{aligned}
E(Z) &= n \cdot E(X_1) = n \cdot p = 220 \cdot 0.95 = 209 \\
\mathrm{var}(Z) &= n \cdot \mathrm{var}(X_1) = n \cdot p \cdot (1-p) = 209 \cdot 0.05 = 10.45 \\
&\Rightarrow \quad \sqrt{\mathrm{var}(Z)} = \sqrt{10.45} = 14
\end{aligned}
$$

což v případě rozptylu platí díky nezávislosti veličin.

Podle CLV bude mít veličina $norm(Z) = \frac{Z-E(Z)}{\sqrt{var(Z)}} = \frac{Z-209}{\sqrt{10.45}}$ přibližně rozdělení $N(0,1)$. Můžeme proto psát

$$
\begin{aligned}
P(Z \leq 216) &= P\left(\frac{Z-209}{\sqrt{10.45}} \leq \frac{216-209}{\sqrt{10.45}}\right) = P\left(norm(Z) \leq \frac{7}{\sqrt{10.45}}\right) \doteq \\
&\doteq \Phi\left(\frac{7}{\sqrt{10.45}}\right) \doteq \Phi(2.165) \doteq 0.985 \;.
\end{aligned}
$$

Doplnění: Rozdělení veličiny je zřejmě $Z \sim \mathrm{Bi}(n,p) = \mathrm{Bi}(220,\ 0.95)$ a její hodnoty jsou $Z \in \{0, 1, \ldots, 220\}$. Výpočet můžeme teď, vzhledem k malému počtu sčítanců, udělat i přímo:

$$
P(Z \leq 216) = 1 - P(Z > 216) = 1 - \sum_{i=217}^{220} \binom{220}{i} 0.95^i \cdot 0.05^{220-i} =
$$

$$
= 1 - 0.95^{217}\Big(1750540 \cdot 0.05^3 + 24090 \cdot 0.95 \cdot 0.05^2 + 220 \cdot 0.95^2 \cdot 0.05 + 1 \cdot 0.95^3 \cdot 1\Big) \doteq
$$

$$
\doteq 1 - 0.95^{217} \cdot 286.82 \doteq 1 - 0.0042 = 0.9958 \;.
$$

Poznámky k testování hypotéz:

Chceme otestovat nějakou hypotézu $\mathbf{H}_0$ o rozdělení náhodné veličiny X (tzv. nulovou hypotézu). Je dobré uvědomit si následující věci:

- Obecně máme vždy k dispozici jen konečně mnoho dat. Pokusů (nebo výsledků) máme ale teoreticky neomezeně, takže principiálně nikdy nemůžeme obsáhnout pomocí konečně mnoha dat všechny možnosti (i při házení mincí se dvěma výsledky, rub nebo líc, vzniká nekonečná posloupnost naměřených dat a všech možných posloupností je navíc nespočetně). Hypotézu tak *nemůžeme potvrdit*, ale nanejvýše v určitém smyslu “vyvrátit” (viz dále).

  Určitou analogií by mohla být situace, že máme zjistit, zda nějaká vlastnost $V(n)$ platí pro všechna přirozená čísla $n \in \mathbb{N}$, a my nemůžeme dělat nic jiného než je postupně všechna procházet. Po konečně mnoha pokusech pak buď narazíme na protipříklad n_0 (a pak určitě víme, že vlastnost V neplatí) anebo ne. Ve druhém případě to může být buď proto, že jsme na protipříklad ještě nenarazili anebo proto, že V zkrátka platí. To vede k tomu, že o závěru takovéhoto pokusu má smysl mluvit jen jako o *zamítnutí* nebo *nezamítnutí* (tj. nemůžeme mluvit o potvrzení). Nezamítnout pak znamená, že nemáme (případné) podklady pro zamítnutí.

  Ovšem to byl příklad, kdy jsme u pokusů (pro konkrétní $n \in \mathbb{N}$) měli vždy odpověď typu ano/ne. U náhodných veličin je to zkomplikováno tím, že i netypická data (která se zdají být v rozporu s $\mathbf{H}_0$) mohou (díky náhodě) pocházet z platnosti námi uvažované hypotézy $\mathbf{H}_0$ (byť s malou pravděpodobností).

- Z výše uvedeného vidíme, že to, jestli $\mathbf{H}_0$ (ve statistice) opravdu platí nebo ne, se tudíž *nikdy nedozvíme* (alespoň ne v rámci testování). A na druhou stranu, pokud bychom to odněkud přece jen věděli, bylo by pak nějaké testování v tomto ohledu už zbytečné.

Protože ale nějaký závěr z testování potřebujeme udělat, je pak naším *ROZHODNUTÍM* buď zamítnutí nebo nezamítnutí $\mathbf{H}_0$ a to na základě nějakého zvoleného kritéria a pochopitelně s určitou možnou chybou (přesněji, se dvěma chybami), které s sebou toto rozhodnutí přináší:

$(\mathbf{H}_0)$	zamítáme	nezamítáme
platí :	chyba 1. druhu	správné rozhodnutí
neplatí :	správné rozhodnutí	chyba 2. druhu

Máme tak

$$\text{“nastává chyba 1. druhu”} \quad \overset{(\text{pokud } \mathbf{H}_0 \text{ platí})}{\Longleftrightarrow} \quad \text{zamítáme } \mathbf{H}_0$$

$$\text{“nastává chyba 2. druhu”} \quad \overset{(\text{pokud } \mathbf{H}_0 \text{ NEplatí})}{\Longleftrightarrow} \quad \text{NEzamítáme } \mathbf{H}_0$$

a zdůrazněme, že ty předpoklady napsané nad ekvivalencemi jsou podstatné. Pokud totiž při testování dojdeme k závěru, že máme hypotézu $\mathbf{H}_0$ zamítnout (na základě nějakého kritéria), NEZNAMENÁ to, že jsem udělali chybu 1. druhu! A to proto, že my zkrátka nevíme, jestli $\mathbf{H}_0$ platí nebo ne.

A navíc, z toho, že chyba 1. druhu opravdu nastává, okamžitě plyne, že $\mathbf{H}_0$ pak musí platit. A to, jak už bylo řečeno výše, prostě nevíme a právě proto také děláme to testování.

- Pravděpodobnosti, že nastanou uvedené chyby se pochopitelně snažíme minimalizovat (volbou kritéria pro zamítání), ale protože není možné obě dvě současně snižovat, musíme si vybrat tu chybu, kterou pokládáme za důležitější. Jelikož celé testování probíhá za předpokladu platnosti $\mathbf{H}_0$ (tj. výpočty, posuzování jejich výsledků atd.), tak za větší prohřešek pokládáme to, že bychom se nakonec spletli a zamítli $\mathbf{H}_0$, která by platila (tj. udělali bychom chybu 1. druhu). Kritérium pro zamítnutí proto volíme tak, aby pravděpodobnost chyby 1. druhu byla omezena předem danou hodnotou α (tzv. hladinou významnosti). Hodnota α by tudíž měla být malé číslo (např. 5% nebo 1% apod.)

- Na tomto místě je ještě potřeba zdůraznit a ujasnit si následující věc: Pravděpodobnosti jednotlivých chyb i správných rozhodnutí jsou VŽDY VZTAŽENY vzhledem k platnosti nebo neplatnosti hypotézy $\mathbf{H}_0$. Důležité ovšem je, že nejde o podmíněnou pravděpodobnost, protože NEMÁME nijak URČENU pravděpodobnost, jestli $\mathbf{H}_0$ platí nebo ne (a takové rozdělení by ani obecně nemuselo dávat smysl). Abychom tuto odlišnost (od podmíněné pravděpodobnosti) trochu více zdůraznili, označíme to způsobem uvedeným níže. Přitom α bude pravděpodobnost chyby 1. druhu a β bude pravděpodobnost chyby 2. druhu:

$$\begin{aligned}
P_{(H_0 \text{ platí})}\Big(\text{nastává chyba 1. druhu}\Big) &= P_{(H_0 \text{ platí})}\Big(\text{zamítáme } \mathbf{H}_0\Big) &&= \alpha \\
P_{(H_0 \text{ platí})}\Big(\text{správné rozhodnutí}\Big) &= P_{(H_0 \text{ platí})}\Big(\text{nezamítáme } \mathbf{H}_0\Big) &&= 1-\alpha \\
P_{(H_0 \text{ NEplatí})}\Big(\text{nastává chyba 2. druhu}\Big) &= P_{(H_0 \text{ NEplatí})}\Big(\text{nezamítáme } \mathbf{H}_0\Big) &&= \beta \\
P_{(H_0 \text{ NEplatí})}\Big(\text{správné rozhodnutí}\Big) &= P_{(H_0 \text{ NEplatí})}\Big(\text{zamítáme } \mathbf{H}_0\Big) &&= 1-\beta
\end{aligned}$$

S podmínkou platnosti/neplatnosti hypotézy jsme tady vlastně ve stejné situaci jako v **Příkladu 2.4**, kde byla na výběr dvě pořadí protihráčů v tenise (tj. dva různé modely nebo také dva různé “světy”). Stejný přístup jsme také používali u metody maximální věrohodnosti, kdy značení $P_\lambda(X_1 = x_1, \ldots, X_n = x_n)$ znamenalo, že hodnota pravděpodobnosti závisí na volbě parametru λ (jenž také určoval, který model zrovna používáme). A ani v případě této metody jsme pochopitelně vůbec neuvažovali o tom, že by zmíněný parametr λ měl nějaké rozdělení pravděpodobnosti.

- Poznamenejme ještě že, pokud jsme $\mathbf{H}_0$ zamítli (při testu s hladinou α), pak jsme se buďto strefili (protože opravdu neplatila) a to pak bylo s pravděpodobností $1-\beta$ (viz výše) anebo jsme udělali chybu 1. druhu (ale jen s malou pravděpodobností α). Tyto dvě pravděpodobnosti ovšem obecně doplňkové nejsou.

- Kritérium zamítnutí pro danou hypotézu $\mathbf{H}_0$, daný počet n naměřených hodnot a danou hladinu α je určeno pomocí tzv. *kritického oboru* $W \subseteq \mathbb{R}^n$ (který tudíž závisí na $\mathbf{H}_0$, n a α) a to prostě tak, že

$$\mathbf{H}_0 \text{ zamítáme (na hladině } \alpha) \quad \Leftrightarrow \quad (x_1, \ldots, x_n) \in W$$

  kde $(x_1, \ldots, x_n)$ je náš soubor naměřených hodnot veličiny X (při nezávislých pokusech), který pro test používáme.

  Volbu kritického oboru W pak v daném případě obvykle děláme pomocí určité náhodné veličiny T (tzv. *testovací statistiky*), která má rozdělení, které známe v jistém smyslu "úplně" (např. $N(0,1)$; t-rozdělení, χ^2-rozdělení atd.) na rozdíl od rozdělení původní veličiny X, u kterého známe zpravidla jen typ rozdělení (např. normální), ale konkrétní parametry už nevíme. (Právě tyto parametry jsou obvykle předmětem našich hypotéz.)

  Výše popsaný kritický obor W (který spadá do $\mathbb{R}^n$) je ale spíše teoretická záležitost, protože v praxi si vystačíme s tím, že zamítací kritérium určíme pomocí rozmezí hodnot statistiky T, tedy pomocí nějaké podmnožiny v $\mathbb{R}$ (např. tak, že $T > c$, kde c je nějaká mezní hodnota pro zamítnutí hypotézy $\mathbf{H}_0$).

  Další možností (jak vyjádřit zamítací kritérium) je použití *intervalu spolehlivosti*, který je ovšem nakonec jen jinak přepsaná podmínka pro zvolenou testovací statistiku T.

  Jak je vidět, všechny uvedené možnosti (tj. kritický obor, podmínka pro T, podmínka pro interval spolehlivosti) jsou jen ekvivalentními verzemi stejné zamítací podmínky. Neboli (pro danou hladinu α) máme

$$\text{zamítáme } \mathbf{H}_0 \quad \Leftrightarrow \quad (x_1, \ldots, x_n) \in W \quad \Leftrightarrow \quad \text{platí podmínka pro } T \quad \Leftrightarrow \quad \text{platí podmínka pro int. spolehl.}$$

- Na závěr si (pro lepší představu) ukážeme volbu ekvivalentních zamítacích podmínek na konkrétním příkladu: pro veličinu $X \sim N(\mu, \sigma^2)$; nulovou hypotézu $\mathbf{H}_0 : \mu \leq \mu_0$ a její alternativu $\mathbf{H}_A : \mu > \mu_0$. Máme soubor naměřených hodnot $(x_1, \ldots, x_n)$. Zde volíme testovací statistiku $T = \frac{\overline{\boldsymbol{X}} - \mu_0}{S_{\boldsymbol{X}}} \sqrt{n}$.

  Veličina T je tady také zapsatelná jako $T = h(X_1, \ldots, X_n)$ pro nějakou reálnou funkci $h : \mathbb{R}^n \to \mathbb{R}$ a naše nezávislé veličiny $X_1, \ldots, X_n$ představující jednotlivé pokusy. Konkrétní tvar funkce h teď nebude vypisovat. Pro hladinu α si dále ještě označme $c := t_{1-\alpha;\, n-1}$ jistou hodnotu (tzv. kvantil). Pak tedy máme:

  - podmínka pro T:
$$t := h(x_1, \ldots, x_n) = \frac{\overline{\boldsymbol{x}} - \mu_0}{s_{\boldsymbol{x}}} \sqrt{n} \ > \ c$$
  - kritický obor:
$$W := \{(y_1, \ldots, y_n) \in \mathbb{R}^n \mid h(y_1, \ldots, y_n) \ > \ c\}$$
  - podmínka pro kritický obor:
$$(x_1, \ldots, x_n) \in W$$
  - interval spolehlivosti:
$$\langle \mu_L, \ +\infty) := \Big\langle \ \overline{\boldsymbol{x}} - \frac{s_{\boldsymbol{x}}}{\sqrt{n}} \cdot c \quad , \quad +\infty \ \Big)$$
  - podmínka pro interval spolehlivosti:
$$\mu_0 \notin \langle \mu_L, \ +\infty)$$

Příklad 11.3 *Výrobce tvrdí, že spotřeba jím vyráběného automobilu je* $\mu_0 = 8\ \ell/100\,\mathrm{km}$. *Průměrná spotřeba u* $n = 49$ *uživatelů ale byla* $\overline{\boldsymbol{x}} = 8.4\ \ell/100\,\mathrm{km}$. *Naměřen byl dále výběrový rozptyl* $s^2_{\boldsymbol{x}} = 2.56\ (\ell/100\,\mathrm{km})^2$.

(a) Testujte na hladině 5%, *zda měl výrobce pravdu (tj. zda spotřeba je rovna* $8\ \ell/100\,\mathrm{km}$*).*

(b) Testujte na hladině 5%, *zda je spotřeba* ***nejvýše*** *rovna* $8\ \ell/100\,\mathrm{km}$.

Jak dopadne testovaní těchto hypotéz na hladině 1%*?*

Řešení:
U veličiny

$$X = \text{“}\textit{spotřeba automobilu}\text{”}$$

budeme předpokládat normální rozdělení $N(\mu, \sigma^2)$. Jednotlivá měření X_i, pro $i = 1, \dots, 49$, jsou nezávislá. Oba parametry jsou neznámé a my chceme testovat střední hodnotu μ.

(a) Podle zadání máme na hladině $\alpha = 5\%$ (příp. 1%) otestovat hypotézu o střední hodnotě

$\mathbf{H}_0 : \mu = \mu_0 (= 8)$

proti alternativní hypotéze:

$\mathbf{H}_1 : \mu \neq \mu_0 (= 8)$.

Pomocí testovací statistiky:
Protože hodnotu rozptylu neznáme, provedeme t-test s testovací veličinou (tzv. *statistikou*):

$$T = \frac{\overline{\boldsymbol{X}} - \mu_0}{S_{\boldsymbol{X}}} \sqrt{n}$$

kde

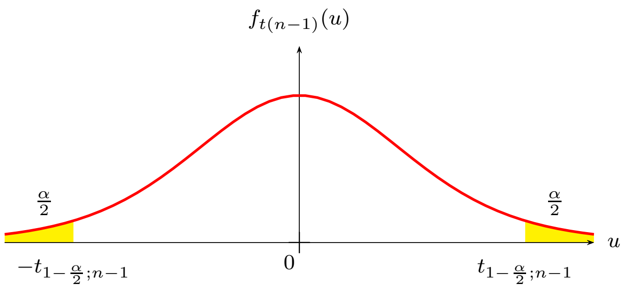
- veličina $\overline{\boldsymbol{X}} = \frac{1}{n} \sum_{i=1}^{n} X_i$ je *výběrový průměr* a
- veličina $S^2_{\boldsymbol{X}} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{\boldsymbol{X}})^2$ je *výběrový rozptyl.*

Kritérium pro **ZAMÍTNUTÍ** H_0 (na hladině α) je tvaru

$$|t| > t_{1-\frac{\alpha}{2};n-1} \quad \Leftrightarrow \quad \text{zamítáme } \mathbf{H}_0 \text{ (na hladině } \alpha) \ .$$

kde t je hodnota T na základě naměřených dat.

Proč má zamítací kritérium uvedený tvar: Za předpokladu platnosti nulové hypotézy, tj. pokud $E(X) = \mu_0$, bude mít statistika T tzv. Studentovo t-rozdělení s $n-1$ stupni volnosti a hustotou $f_{t(n-1)}$ (která má podobný, ale ne stejný, průběh jako u $N(0,1)$):



Očekávané hodnoty takovéto statistiky T by se měly pohybovat blízko nuly. Pokud se příliš odchýlí, bude to důvod k zamítnutí nulové hypotézy. Nebudeme přitom preferovat vychýlení na žádnou ze stran - tj. chybu 1. druhu s pravděpodobností α rozdělíme na poloviny $\frac{\alpha}{2}$ na obě strany. Pak máme

$$P_{(H_0 \text{ platí})}\Big(\text{nastává chyba 1. druhu}\Big) = P_{(H_0 \text{ platí})}\Big(\text{zamítáme } \mathbf{H}_0\Big) =$$
$$= P_{(H_0 \text{ platí})}\Big(\; |T| > t_{1-\frac{\alpha}{2};n-1} \; \Big) = \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$$

Teď už tedy dosadíme konkrétní naměřené hodnoty (které pro jednotlivé veličiny značíme pro odlišení malými písmeny, tj. $\overline{\boldsymbol{x}}$, $s^2_{\boldsymbol{x}}$ a t). Realizace testovací statistiky je

$$t = \frac{\overline{\boldsymbol{x}} - \mu_0}{s_{\boldsymbol{x}}}\sqrt{n} = \frac{8.4 - 8}{\sqrt{2.56}}\sqrt{49} = \frac{0.4}{1.6} \cdot 7 = 1.75 \; .$$

Protože pro $\alpha = 0.05$ je

$$|t| = 1.75 \not> 2.011 \doteq t_{0.975;48} = t_{1-\frac{\alpha}{2};n-1} \; ,$$

nulovou hypotézu **NEZAMÍTÁME** na hladině 5%.

Protože při snížení hladiny se zmenšuje i kritický obor W (je to vidět i na obrázku, kde žlutá plocha bude menší), tak na hladině 1% hypotézu $\mathbf{H}_0$ také **NEZAMÍTÁME**.

(Pro úplnost si ale stejně ještě vyjádříme příslušnou podmínku: $|t| = 1.75 \not> 2.682 \doteq t_{0.995;48}$.)

Obecněji tedy:

snižujeme hladinu chyby 1. druhu (tj. chceme si být více jistí) $\Rightarrow$ musíme tolerovat více "prohřešků" $\Rightarrow$ častěji nezamítáme

Pomocí intervalu spolehlivosti:

Kritérium pro zamítnutí $\mathbf{H}_0$ na hladině α

$$|t| > t_{1-\frac{\alpha}{2};n-1}$$

se dá ekvivalentně přepsat (při vyjádření $t = \frac{\overline{\boldsymbol{x}}-\mu_0}{s_{\boldsymbol{x}}}\sqrt{n}$) jako

$$\mu_0 \notin \left\langle \; \overline{\boldsymbol{x}} - \frac{s_{\boldsymbol{x}}}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2};n-1} \quad , \quad \overline{\boldsymbol{x}} + \frac{s_{\boldsymbol{x}}}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2};n-1} \; \right\rangle \quad =: \langle \mu_L, \; \mu_U \rangle$$

což je hledaný interval spolehlivosti.

Při vyčíslení pro $\alpha = 5\%$, tj. $t_{1-\frac{\alpha}{2};n-1} = t_{0.975;\,48} \doteq 2.011$, tedy dostaneme

$$\langle \mu_L, \; \mu_U \rangle = \left\langle \; 8.4 - \frac{1.6}{\sqrt{49}} \cdot 2.011 \quad , \quad 8.4 + \frac{1.6}{\sqrt{49}} \cdot 2.011 \; \right\rangle = \langle 7.94, \; 8.86 \rangle$$

Protože máme $\mu_0 = 8 \in \langle 7.94, \; 8.86 \rangle = \langle \mu_L, \; \mu_U \rangle$, hypotézu $\mathbf{H}_0$ **NEZAMÍTÁME** na hladině 5%.

(Výsledek musel samozřejmě dopadnout stejně jako při testovací statistice, protože je to ekvivalentní princip.)

- -

(b) V tomto případě budeme na hladině $\alpha = 5\%$ (příp. 1%) testovat hypotézu o střední hodnotě

$\widetilde{\mathbf{H}}_0 : \mu \leq \mu_0 (= 8)$

proti alternativní hypotéze:

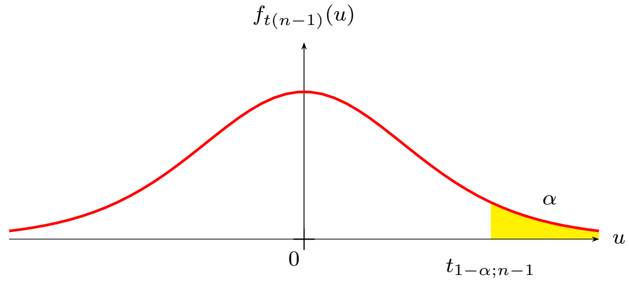
$$\widetilde{\mathbf{H}}_1 : \mu > \mu_0 (= 8) \ .$$

Pomocí testovací statistiky:

Statistika T bude mít stejný tvar jako v předešlém případě (a). Kritérium pro **ZAMÍTNUTÍ** $\widetilde{\mathbf{H}}_0$ (na hladině α) bude ale teď jiné, a sice

$$t > t_{1-\alpha;n-1} \quad \Leftrightarrow \quad \text{zamítáme } \widetilde{\mathbf{H}}_0 \text{ (na hladině } \alpha) \ .$$

Proč má zamítací kritérium uvedený tvar: Za předpokladu platnosti nulové hypotézy, tj. pokud $E(X) \leq \mu_0$, bude mít hustota pro statistiku T svůj vrchol v intervalu $(-\infty, 0\rangle$. Očekávané hodnoty takovéto statistiky T by se měly pohybovat spíše v záporných až nulových hodnotách. Pokud se příliš odchýlí do kladných hodnot, bude to důvod k zamítnutí nulové hypotézy. "Nejhorší" z tohoto hlediska je krajní případ $E(X) = \mu_0$, pro který má T opět Studentovo t-rozdělení s $n-1$ stupni volnosti (viz obrázek). Chyba 1. druhu s pravděpodobností α zde tedy bude soustředěna jen na jedné straně:



Podobně jako předtím máme

$$P_{(H_0 \text{ platí})}\Big(\text{nastává chyba 1. druhu}\Big) = P_{(H_0 \text{ platí})}\Big(\text{zamítáme } \mathbf{H}_0\Big) =$$
$$= P_{(H_0 \text{ platí})}\Big(\ T > t_{1-\alpha;n-1} \ \Big) = \alpha$$

Hodnota statistiky T zůstane stejná jako předtím, tedy $t = 1.75$, a protože pro $\alpha = 0.05$ máme

$$t = 1.75 > 1.677 \doteq t_{0.95;\ 48} = t_{1-\alpha;\ n-1} \ ,$$

hypotézu $\widetilde{\mathbf{H}}_0$ **ZAMÍTNEME**.

(Pozor, jde o jednostranný test, takže kvantil je jiný! Veškerou chybu jsme spotřebovali jen na kladné hodnoty. A toto malé zvětšení, oproti oboustrannému testu, už stačilo na zamítnutí.)

Pro $\alpha = 1\%$ pak máme

$$t = 1.75 \ngtr 2.407 \doteq t_{0.99;\ 48} \ ,$$

takže při této hladině hypotézu $\widetilde{\mathbf{H}}_0$ naopak **NEZAMÍTNEME**.

Pomocí intervalového odhadu:

Kritérium pro zamítnutí $\widetilde{\mathbf{H}}_0$ na hladině α

$$t > t_{1-\alpha;\,n-1}$$

se dá ekvivalentně přepsat (opět při vyjádření $t = \frac{\overline{\boldsymbol{x}}-\mu_0}{s_{\boldsymbol{x}}}\sqrt{n}$) jako

$$\mu_0 \notin \Big\langle \ \overline{\boldsymbol{x}} - \frac{s_{\boldsymbol{x}}}{\sqrt{n}} \cdot t_{1-\alpha;\,n-1} \ , \quad +\infty \ \Big) \quad =: \langle \mu_L, \ +\infty)$$

což je hledaný dolní interval spolehlivosti.

Při vyčíslení pro $\alpha = 5\%$, tj. $t_{1-\alpha;\,n-1} = t_{0.95;\,48} \doteq 1.677$, tedy dostaneme

$$\langle \mu_L,\ +\infty) = \Big\langle\ 8.4 - \frac{1.6}{\sqrt{49}} \cdot 1.677\ ,\ \ +\infty\Big) = \langle 8.017,\ +\infty)$$

Protože máme $\mu_0 = 8 \notin \langle 8.017,\ +\infty) = \langle \mu_L,\ +\infty)$, hypotézu $\mathbf{H}_0$ **ZAMÍTÁME** na hladině 5%. (Výsledek opět dopadne stejně jako při testovací statistice, protože je to ekvivalentní princip.)

Důležitá poznámka: Všimněme si, že jsme došli k těmto (zdánlivě protichůdným výsledkům): na hladině $\alpha = 5\%$ jsme

- hypotézu $\mu = \mu_0$ nezamítli
- hypotézu $\mu \leq \mu_0$ zamítli

přestože nezamítnutý případ je podpřípadem zamítnutého. To vypadá sice jako rozpor, ale ve skutečnosti v každém z případů testujeme hypotézy jiným způsobem. Jak už bylo napsáno výše, chyba se v případě oboustranného testu rozloží symetricky na obě strany, zatímco u jednostranného testu je nahromaděna jen na jednom konci.