přednáška 11 – MZ slajd 11.pptx

cvičení 11: <u>intervaly spolehlivosti a t-test</u> (po probrání předn. 11 a předn. a cvičení č. 10)

<u>Ad př. 1 / předn. 10.</u>

Ve statistickém testu z př. 1 z minulé přednášky (číslo 10) bychom mohli zkonstruovat i interval spolehlivosti

$\alpha = 0{,}05$ ... odpovídá 5%-nímu výskytu chyby 1. druhu

příslušný <u>interval spolehlivosti</u> (confidence interval) je <u>$(1-\alpha)$%-ní</u>, tj. 95%-ní

Naměřená hodnota je $X = 135$, tj. $\frac{135-\mu}{\sigma} \in \langle -1{,}96; 1{,}96 \rangle$ ... interval vypočtený v příkladu 1

když bychom μ nepokládali striktně rovno 120, ale vyjádřili je pomocí mezí daného intervalu (dopočteme $\sigma = \sqrt{N \cdot p(1-p)} = \sqrt{96}$ ... odchylka binomického rozdělení četností),

dostaneme:

$$-1{,}96 \leq \frac{135-\mu}{\sqrt{96}} \leq 1{,}96 \quad / \cdot (-1)$$

$$1{,}96 \geq \frac{\mu-135}{\sqrt{96}} \geq -1{,}96$$

$$135 + 1{,}96 \cdot \sqrt{96} \geq \mu \geq 135 - 1{,}96 \cdot \sqrt{96} \quad \ldots \Rightarrow \mu \in 135 \pm 1{,}96 \cdot \sqrt{96}$$

$$\underline{\mu \in \langle 115{,}796; 154{,}204 \rangle}$$

Tento 95%-ní interval spolehlivosti nám dává lepší informaci, než jen statistický test: říká nejen to, že odhad zájmu by mohl být 120 zákazníků ze 600, ale že zájem o náš výrobek má 116 – 154 zákazníků ze 600. Tento interval je asi cennější informace, pokud ji máme sdílet se šéfem v práci s veřejností, apod.

<u>Ad př. 2 / předn. 10</u>

Test v př. 2 byl jednostranný, analogicky k němu lze zkonstruovat příslušný jednostranný $(1-\alpha)$%-ní interval spolehlivosti takto:

V experimentu jsme naměřili hodnotu 540, a místo abychom automaticky předpokládali $\mu = 500$, necháme μ jako neznámou:

H_0 nezamítáme pro $\frac{540-\mu}{20} \in (-\infty; 1{,}64\rangle$, tedy

$$-\infty \leq \frac{540-\mu}{20} \leq 1{,}64 \quad / \cdot (-1)$$

$$\infty \geq \frac{\mu-540}{20} \geq -1{,}64$$

$$\infty \geq \mu \geq 540 - 20 \cdot 1{,}64 \quad \ldots \Rightarrow \mu \in \langle 540 - 20 \cdot 1{,}64; \infty)$$

$$\underline{\mu \in \langle 507{,}2; \infty)}$$

Tj. z intervalu spolehlivosti vidíme, že průměrný počet bodů se při použití programu v 95% případů zvedne aspoň na 507,2 bodů

(hodnotu 32,8 jsme – v testu přičetli k 500
– v intervalu spolehlivosti odečetli od 540)

Pozn.: v praxi se téměř vždy konstruují oboustranné intervaly spolehlivosti, tj. i v našem příkladu 2 by bylo možné prapsat postup

$$\frac{540-\mu}{20} \in (-1{,}96\,;1{,}96) \ldots$$ u veličiny U jsme na obou stranách „useklí" obsah 0,025, tj. kritický interval je stejný (pro stejné α oboustranného testu) jako v př. 1/předn. 10

$$\mu \in \langle 540-20\cdot 1{,}96\,;540+20\cdot 1{,}96\rangle$$

$$\underline{\mu \in \langle 500{,}8\,;579{,}2\rangle} \ldots$$ oboustranný 95%-ní interval spolehlivosti pro střední hodnotu průměru

Ad př. 3/předn. 10 ... předpokládáme $\sigma^2 = 10\,000$ rozptyl jediné hodnoty měření

$\frac{\sigma^2}{25}$ ... rozptyl průměru 25 hodnot

$\frac{\sigma^2}{30}$ ... rozptyl průměru 30 hodnot

Zde práce jen pro intervaly spolehlivosti obě veličiny $\bar{X}_1, \bar{X}_2$ od sebe oddělíme:
$\mu_{\bar{X}_1}$ (resp. $\mu_{\bar{X}_2}$) předpokládáme neznámé, rozptyl je v pořádku, tj. můžeme vzít

$$\frac{\bar{X}_1-\mu_{\bar{X}_1}}{\sqrt{\frac{\sigma^2}{25}}} \in \langle -1{,}96\,;1{,}96\rangle \ldots$$ spočtené kritické hodnoty z veličiny U s rozdělením No(0;1)

$$\frac{600-\mu_{\bar{X}_1}}{\sqrt{\frac{10\,000}{25}}} \in \langle -1{,}96\,;1{,}96\rangle \qquad \frac{\bar{X}_2-\mu_{\bar{X}_2}}{\sqrt{\frac{\sigma^2}{30}}} \in \langle -1{,}96\,;1{,}96\rangle$$

$$\mu_{\bar{X}_1} \in 600 \pm 1{,}96\cdot 20 \qquad \mu_{\bar{X}_2} \in 575 \pm 1{,}96\cdot 18{,}257$$

$$\underline{\mu_{\bar{X}_1} \in \langle 560{,}8\,;639{,}2\rangle} \qquad \underline{\mu_{\bar{X}_2} \in \langle 539{,}2\,;610{,}8\rangle}$$

toto jsou 95%-ní intervaly spolehlivosti pro průměr každé veličiny zvlášť

- 95%-ní int. spolehlivosti pro $E\bar{X}_1$
- 95%-ní int. spolehlivosti pro $E\bar{X}_2$

Př. 4 (MPSO text, příklad 1.7/str. 37, pdf strana 39). Chceme ověřit statistickým testem hypotézu, že prvorozené děti jsou mnohem nesmělejší než druhorozené. U šesti prvorozených chlapců jsme zjistili počet jejich schůzek s děvčetem za rok: 2, 5, 4, 2, 1, 4. U šesti druhorozených chlapců byl stejným pohlížením zjištěn počet schůzek s děvčetem: 6, 4, 5, 7, 3, 5.

a) proveďte t-test

b) vypočtěte 95%ní interval spolehlivosti pro průměrný počet schůzek s děvčetem v obou případech

---

Řešení: a) Provedeme oboustranný statistický test pro $\alpha = 0{,}05$:

krok 1: H_0: počet schůzek s děvčetem nezávisí na prvorozenosti: $\mu_{\bar{X}_1} = \mu_{\bar{X}_2}$

H_1: — " — závisí — " — $\mu_{\bar{X}_1} \neq \mu_{\bar{X}_2}$

krok 2: kritériem bude rozdíl průměrů $\bar{X}_1 - \bar{X}_2$

krok 3: popíšeme chování veličiny $\bar{X}_1 - \bar{X}_2$ za předpokladu, že platí H_0:

Protože rozptyl je zde neznámý nejasnost/nejistota se zde zvyšuje, rozptyl σ^2 budeme muset odhadnout, tj. naši veličinu $\bar{X}_1 - \bar{X}_2$ nepopíšeme rozdělením normálním, ale t-rozdělením (tzv. Studentovo t-rozdělení), jehož hustota je trochu více do šířky rozložená kolem počátku, tj. pro $\alpha = 0{,}05$ oboustranného testu nevezmeme interval $\langle -1{,}96; 1{,}96 \rangle$ jako interval pro nezamítnutí H_0, nýbrž interval trochu větší

soubor první: měření 2, 5, 4, 2, 1, 4 ... $\Rightarrow$ $\bar{x}_1 = \frac{18}{6} = 3$ $\quad s_1^2 = \frac{6}{5}\left(\frac{2^2+5^2+4^2+2^2+1^2+4^2}{6} - 3^2\right) = 2{,}4$

soubor druhý: měření 6, 4, 5, 7, 3, 5 ... $\Rightarrow$ $\bar{x}_2 = \frac{30}{6} = 5$ $\quad s_2^2 = \frac{6}{5}\left(\frac{6^2+4^2+5^2+7^2+3^2+5^2}{6} - 5^2\right) = 2$

Předpokládejme, že ač jsou střední hodnoty obou veličin jakékoli, vnitřní rozptyl v různých skupinách, daný rozdílností lidí, je stále stejný – proto udělejme průměr těchto rozptylů a získáme ideální odhad rozptylu

(teď napíšu vzorec pro $N_1 \neq N_2$, abyste byli schopni jej správně použít ve svých zápočtových příkladech:

$s_1^2 = 2{,}4$ ... odhad rozptylu jediné hodnoty, s volností $N_1 - 1 = 6 - 1 = 5 = V_1$

$s_2^2 = 2$ ... odhad rozptylu jediné hodnoty s volností $V_2 = N_2 - 1 = 6 - 1 = 5$

$\Rightarrow$ průměrný odhad rozptylu jediné hodnoty je

$$\text{est}\,\sigma^2 = \frac{V_1}{V_1+V_2}\cdot s_1^2 + \frac{V_2}{V_1+V_2}\cdot s_2^2 = \frac{5}{5+5}\cdot 2{,}4 + \frac{5}{5+5}\cdot 2 = 2{,}2$$

... s volností $V_1 + V_2 = 10$ stupňů volnosti!!

$$\text{est}\,\sigma^2_{\bar{X}_1-\bar{X}_2} = \frac{\text{est}\,\sigma^2}{N_1} + \frac{\text{est}\,\sigma^2}{N_2} = \frac{2{,}2}{6} + \frac{2{,}2}{6} = 0{,}7333 \Rightarrow \text{est}\,\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{0{,}7333} \doteq 0{,}8563$$

Celkem $\frac{\bar{X}_1 - \bar{X}_2 - 0}{0{,}8563} \overset{H_0}{\sim} t$ ($\nu = \nu_1 + \nu_2 = 10$ stupňů volnosti)

krok 4: pro zvolené $\alpha = 0{,}05$ najdeme kritický interval (MPSO.pdf skripta, str. 18)

$\alpha = 2Q$ ... oboustranný test ... máme sloupec tabulky
$\nu = 10$ ... volnost ... máme řádek tabulky $\Big\}\ t_h = 2{,}228$

$\Downarrow$

$t_d = -2{,}228$

(jedná se o symetrické hodnoty vzhledem k počátku, liší se jen znaménkem)

krok 5: $\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{est\,\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{3 - 5 - 0}{0{,}8563} = -2{,}336 \notin (-2{,}228; 2{,}228)$,

tedy H_0 zamítáme,
statistický test prokázal, že prvorození se scházejí s dívčaty statisticky významně méně

b) ~~sest~~ sestavme intervaly spolehlivosti průměrného počtu schůzek v obou případech, přitom obě veličiny $\bar{X}_1, \bar{X}_2$ od sebe oddělíme – do obou vzorců ovšem vezmeme $est\,\sigma^2 = 2{,}2$ a volnost $\nu = 10$:
Vyjdeme z „intervalu pro nezamítnutí H_0":

$$\frac{\bar{X}_1 - \mu_1}{\frac{est\,\sigma}{\sqrt{N_1}}} \in \langle -2{,}228; 2{,}228 \rangle \qquad \frac{\bar{X}_2 - \mu_2}{\frac{est\,\sigma}{\sqrt{N_2}}} \in \langle -2{,}228; 2{,}228 \rangle$$

$$\frac{3 - \mu_1}{\sqrt{\frac{2{,}2}{6}}} \in \langle -2{,}228; 2{,}228 \rangle \qquad \frac{5 - \mu_2}{\sqrt{\frac{2{,}2}{6}}} \in \langle -2{,}228; 2{,}228$$

$$\mu_1 \in 3 \pm 2{,}228 \cdot \sqrt{\frac{2{,}2}{6}} \qquad \mu_2 \in 5 \pm 2{,}228 \cdot \sqrt{\frac{2{,}2}{6}}$$

$$\mu_1 \in 3 \pm 1{,}349 \qquad \mu_2 \in 5 \pm 1{,}349$$

$$\underline{\mu_1 \in \langle 1{,}651; 4{,}349 \rangle} \qquad \underline{\mu_2 \in \langle 3{,}651; 6{,}349 \rangle}$$

Je vidět, že oba intervaly spolehlivosti pro dílčí střední hodnoty se překrývají i když H_0 v části (a) byla zamítnuta – zde skutečně neplatí (H_0 zamítnuto $\Leftrightarrow$ dílčí i.s. se nepřekrývají), protože dílčí i.s. jsou konstruovány ~~až~~ po rozdělení obou proměnných,

tj. dílčí intervaly spolehlivosti odpovídají nějakým jiným, dílčím testům, nikoli souhrnnému testu $H_1: \mu_1 \neq \mu_2$.

Př. 5 (MPSO text, př. 1.8/str. 37, pdf strana 39) Chceme a) statistickým testem b) intervalem spolehlivosti ověřit, zda kvalita reakcí člověka je stejná za denního jako za umělého světla. U náhodně vybrané skupiny 10 lidí byly zjištěny výsledky zkoušky za dvou situací

11/5

situace „denní světlo": 9, 2, 7, 12, 14, 10, 6, 7, 12, 10 … x_i

situace „umělé světlo": 7, 2, 4, 13, 13, 7, 4, 6, 8, 9 … y_i

(oba soubory jsou uspořádané, tj. data i-tého respondenta jsou v obou souborech na i-té pozici)

Řešení: jedná se o experiment dvou měření u každého respondenta, tj. užijeme párový test: rozdíly hodnot měření u každého člověka, tj.

$z_i = x_i - y_i$ jsou: 2, 0, 3, −1, 1, 3, 2, 1, 4, 1

$$\bar{z} = \frac{16}{10} = 1,6$$

$$s^2 = \frac{10}{9}\left(\frac{2^2+3^2+(-1)^2+1+3^2+2^2+1+4^2+1}{10} - 1,6^2\right) \doteq 2,2667$$

$\nu = N-1 = 10-1 = 9$ stupňů volnosti

a) statistický test:

krok 1: H_0: odchylky z_i jsou blízké 0, tj. $E\bar{Z} = 0$

H_1: odchylky z_i nejsou blízké 0, tj. $E\bar{Z} \neq 0$

krok 2:

$\bar{Z}$ … průměr odchylek bude kritériem našeho testu

krok 3

$$\frac{\bar{Z} - 0}{\sqrt{\text{est}\,\sigma^2_{\bar{Z}}}} \overset{H_0}{\sim} t(\nu = 10-1 = 9)$$

krok 4: pro $\alpha = 0,05$: najdeme kritické hodnoty t-rozdělení v tabulce:

$\alpha = 2Q = 0,05$ … oboustranný test … určuje sloupec tabulky

$\nu = 9$ … počet stupňů volnosti

$t_K = 2,262$

$t_m = -2,262$

krok 5: rozhodnutí stat. testu:

$$\frac{1,6 - 0}{\sqrt{\frac{2,2667}{10}}} = 3,361 \notin \langle -2,262; 2,262 \rangle \Rightarrow H_0 \text{ zamítneme,}$$

hodnoty x_i jsou statisticky významně vyšší než hodnoty y_i,

tj. 1. situace (počet bodů testu za denního světla) má průměr významně vyšší než 2. situace

b) Vypočtěme ještě interval spolehlivosti pro střední hodnotu veličiny $\bar{Z}$:

Místo 0 do vzorce dosadíme μ:

H_0 nezamítáme pro $\dfrac{1{,}6-\mu_{\bar{Z}}}{\sqrt{\frac{2{,}2667}{10}}} \in \langle -2{,}262; 2{,}262 \rangle$, tedy

$$\mu_{\bar{Z}} \in 1{,}6 \pm 2{,}262 \cdot \sqrt{\frac{2{,}2667}{10}} = \underline{\underline{\langle 0{,}5231; 2{,}677 \rangle}}$$

Tj: skupina „za denního světla" má o 0,523 až 2,677 bodů více průměrně než průměrný jedinec „za umělého osvětlení" na stupnici kvality reakce.

(Pravděpodobně, to se ještě týká vzdělání; čím lepší kvalita reakce, tím je vyšší počet bodů na stupnici)

Po absolvování tohoto týdne byste měli být schopni vyřešit všechny výpočtové příklady kromě příkladů 2, 10, 11, které nám zbývají na příští týden.