ŠTATISTICKÉ SPRACOVANIE DÁT

prof. Ing. Dušan KUDELAS, PhD

Testy hypotéz o zhode dvoch stredných hodnôt nezávislých výberových súborov t-testy

Nech štatistický znak X_1 má v prvom základnom súbore približne normálne rozdelenie ….$N(\mu_1, \sigma_1^2)$

Štatistický znak X_2 má v druhom základnom súbore tiež približne normálne rozdelenie ….$N(\mu_2, \sigma_2^2)$

Predpokladajme, že odhadované stredné hodnoty μ_1 a μ_2 sú zhodné, t.j.

testujeme $$H_0 : \mu_1 = \mu_2$$

oproti alternatívnej hypotéze pri obostrannom teste $$H_1 : \mu_1 \neq \mu_2$$

Testy zhody stredných hodnôt **t – test**

- Studentov t – test umožňuje testovanie hypotézy H_0: $\mu_1 = \mu_2$ proti alternatívnej H_1: $\mu_1 \neq \mu_2$
- Klasický Studentov t – test pre **zhodné rozptyly**, $\nu = n_1 + n_2 - 2$

$$t = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{(n_1 - 1){s_1}^2 + (n_2 - 1){s_2}^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

- Studentov t – test pre **rôzne rozptyly**

$$t = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} \qquad \nu = \frac{\left(\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}\right)^2}{\frac{{s_1}^4}{{n_1}^2(n_1 - 1)} + \frac{{s_2}^4}{{n_2}^2(n_2 - 1)}}$$

Testy zhody rozptylov: **F – test**, χ^2 test

- Umožňuje overiť hypotézy $\frac{\sigma_1^{\;2}}{\sigma_2^{\;2}}$ = 1 t.j. $\sigma_1^{\;2} = \sigma_2^{\;2}$
- Testovacie kritérium F = $\frac{s_1^{\;2}}{s_2^{\;2}}$ pričom $s_1^{\;2} > s_2^{\;2}$
- Kritická hodnota sa určí z tabuliek Fisherovho rozdelenia, pre zvolenú hladinu významnosti α a počet stupňov voľnosti $\nu_1 = n_1 - 1$ a $\nu_2 = n_2 - 1$

$$\sigma^{\;2} = s^2$$

$$\chi^2 = \frac{[n-1].s^2}{\sigma^2}$$

Pr

- Treba overiť efektívnosť nového spôsobu výučby vybraných tematických celkov matematiky. Preto výučba matematiky bola realizovaná v jednej triede experimentálnym spôsobom. Z tejto triedy boli náhodne vybraní žiaci, ktorí budú predstavovať experimentálnu skupinu. Žiaci z experimentálnej skupiny dosiahli v záverečnom teste, ktorý písali po realizácii experimentu, nasledujúce počty bodov: 22, 34, 26, 31, 26, 35, 25, 38, 36, 22, 23, 32. V ostatných triedach prebiehala výučba matematiky štandardným spôsobom. Z týchto tried boli náhodne vybraní žiaci, ktorí tvoria kontrolnú skupinu. Dosiahli v záverečnom teste nasledujúce počty bodov: 24, 33, 23, 20, 26, 32, 35, 21, 25.

$$s_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2 = 33{,}79$$

$$s_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n}(x_{2i} - \bar{x}_2)^2 = 29{,}78$$

$$\bar{x}_1 = \frac{1}{n_{1i}}\sum_{i=1}^{n} x_{1i} = 29{,}14$$

$$\bar{x}_2 = \frac{1}{n_{2i}}\sum_{i=1}^{n} x_{2i} = 26{,}56$$

$$F = \frac{s_1^2}{s_2^2} = 1{,}135$$

ν_2 \ ν_1	10	12
1	241´,88	243,91
2	19,396	19,413
3	8,786	8,745
4	5,964	5,912
5	4,735	4,678
6	4,060	4,000
7	3,637	3,575
8	3,347	3,284

TK < KH=> H_0

TK: t

$$t = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{(n_1-1)s_1^{\ 2} + (n_2-1)s_2^{\ 2}}}\sqrt{\frac{n_1 n_2 (n_1+n_2-2)}{n_1+n_2}} = \frac{\left|29{,}17 - 26{,}56\right|}{\sqrt{(12-1)33{,}79 + (9-1)29{,}78}}\sqrt{\frac{12.9(12+9-2)}{12+9}} = 1{,}05$$

KH: t = 2,093

	P		
ν	0,90	0,95	0,975
1	3,078	6,314	12,706
2	1,886	2,920	4,303
3	1,638	2,353	3,182
4	1,553	2,132	2,776
5	1,476	2,015	2,571
6	1,440	1,943	2,447
7	1,415	1,895	2,365
8	1,397	1,860	2,306
9	1,383	1,833	2,262
10	1,372	1,812	2,228
11	1,363	1,796	2,201
12	1,356	1,782	2,179
13	1,350	1,771	2,160
14	1,345	1,761	2,145
15	1,341	1,753	2,131
16	1,337	1,746	2,120
17	1,333	1,740	2,110
18	1,330	1,734	2,101
19	1,328	1,729	2,093

TK < KH=> H_0

Pr.

- Psychológovia vypracovali určitý postup na upokojenie žiakov (rozprávanie s hudobnými efektmi), ktorý by sa mal prejaviť znížením krvného tlaku. Na overenie účinnosti tohto postupu bol meraný krvný tlak u žiakov dvoch skupín; v prvej skupine bez aplikácie tohto postupu a v druhej skupine s aplikáciou tohto postupu. Výsledky

(v mm Hg) sú takéto: 1. skupina: 105, 107, 110, 117, 124, 153, 137, 174, 109, 119, 143, 162, 91, 146, 109;

- 2. skupina: 92, 96, 104, 119, 106, 100, 93, 90, 98, 109, 106, 91, 88, 94

$$\bar{x}_1 = \frac{1}{n_{1i}} \sum_{i=1}^{n} x_{1i} = 127{,}1$$

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2 = 579{,}8$$

$$\bar{x}_2 = \frac{1}{n_{2i}} \sum_{i=1}^{n} x_{2i} = 99$$

$$s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n} (x_{2i} - \bar{x}_2)^2 = 77{,}7$$

$$F = \frac{s_1^2}{s_2^2} = 7{,}5$$

TK > KH=> H_1

ν_2 \ ν_1	10	12	15
13	2,671	2,604	2,533

$$t' = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{127{,}1 - 99}{\sqrt{\frac{579{,}8}{15} + \frac{77{,}7}{14}}} = \frac{28{,}1}{6{,}65} = 4{,}23$$

$$\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} = \frac{\left(\frac{579{,}8}{15} + \frac{77{,}7}{14}\right)^2}{\frac{\left(\frac{579{,}8}{15}\right)^2}{14} + \frac{\left(\frac{77{,}7}{14}\right)^2}{13}} \doteq 18.$$

	P		
ν	0,90	0,95	0,975
1	3,078	6,314	12,706
2	1,886	2,920	4,303
3	1,638	2,353	3,182
4	1,553	2,132	2,776
5	1,476	2,015	2,571
6	1,440	1,943	2,447
7	1,415	1,895	2,365
8	1,397	1,860	2,306
9	1,383	1,833	2,262
10	1,372	1,812	2,228
11	1,363	1,796	2,201
12	1,356	1,782	2,179
13	1,350	1,771	2,160
14	1,345	1,761	2,145
15	1,341	1,753	2,131
16	1,337	1,746	2,120
17	1,333	1,740	2,110
18	1,330	1,734	2,101

TK > KH=> H_1

- Na výrobnej linke pre balenie údenín bolo preväžených 20 náhodne vybraných balíčkov s týmito výsledkami v gramoch: 251,5; 258,1; 253,4; 256; 245; 251,1; 255,7; 255,5; 249,2; 254,6; 251,9; 248,3; 255,7; 256,2; 250,5; 251,7; 254; 259,5; 257; 252,8. Je potrebné zistiť, či hmotnosti balíčkov vyhovujú norme, podľa ktorejpriemerná hodnota hmotnosti má byť 250 g a kolísavosť hmotnosti vyjadrená rozptylom má byť menšia alebo rovná 4.

$$t = \frac{\bar{x} - \mu_0}{s} \cdot \sqrt{n} = \frac{253{,}385 - 250}{3{,}559} \cdot \sqrt{20} = 4{,}253.$$

$$t_{0,05}(19) = 2{,}093$$

$$H_0: \sigma^2 = 4 \qquad \text{oproti} \qquad H_1: \sigma^2 > 4$$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{19 \cdot 12{,}667}{4} = 60{,}168$$

$$\chi^2_{0,05}(19) = 30{,}14$$

Test hypotézy o zhode dvoch stredných hodnôt závislých súborov - Párový t-test

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \qquad t = \frac{\bar{d}}{S}\cdot\sqrt{n} \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2$$

Je potrebné overiť, či sa u osobného automobilu určitej značky pri správnej geometrii kolies ojazdia obe pneumatiky rovnako rýchlo. Preto bolo náhodne vybraných 6 nových áut a po určitej dobe bolo zistené, o koľko milimetrov sa ojazdili ich pravé a ľavé predné pneumatiky. Výsledky zisťovania sú uvedené v nasledujúcej tabuľke

auto č. i	1	2	3	4	5	6
x_i	1,8	1,0	2,2	0,9	1,5	1,6
y_i	1,5	1,1	2,0	1,1	1,4	1,4

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = 0{,}0833; \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2 = 0{,}0377; \quad s = 0{,}194$$

$$t = \frac{\bar{d}}{s}\cdot\sqrt{n} = \frac{0{,}0833}{0{,}194}\cdot\sqrt{6} = 1{,}0512$$

	P		
ν	0,90	0,95	0,975
1	3,078	6,314	12,706
2	1,886	2,920	4,303
3	1,638	2,353	3,182
4	1,553	2,132	2,776
5	1,476	2,015	2,571