vodní hospodářství®

www.vodnihospodarstvi.cz

ročník 57
4
2007

ČOV Pivovar Černá Hora, IC reaktor

ČOV VELVETA a.s. Varnsdorf

ČOV Danzer Bohemia - Dýhárna, montáž IC reaktoru

Vracíme vodě život

1997-2007 EWA 10 let členství

Vodní hospodářství
je mediálním partnerem
výstavy VODOVODY-KANALIZACE
a veletrhu ENVIBRNO
29. - 31. května 2007 Výstaviště Brno
Časopis bude k dispozici všem vystavovatelům
a návštěvníkům obou akcí.
Najdete nás na společném stánku s AČE ČR č. 30 v pavilonu D.

vodní hospodářství · AČE · VODOVODY - KANALIZACE 2007 · ENVIBRNO

PŘÍLOHA VODAŘ

Statistické zpracování vodohospodářských dat

4. Testování správnosti a shodnosti v kontrolní laboratoři

Milan Meloun

Klíčová slova
test správnosti - test shodnosti - párový test - Hornův postup - malé výběry – heteroskedasticita - nulová hypotéza - alternativní hypotéza - Studentův t-test - Fischer-Snedecorův F-test

Souhrn

Pro testování hypotéz o parametrech základního souboru na základě jednoho výběru jsou odvozeny testovací statistiky ze vztahů pro intervaly spolehlivosti. Jednodušší způsob spočívá v přímém užití 100(1 - α)%ního intervalu spolehlivosti: padne-li zadaná hodnota Θ_0 parametru Θ do tohoto intervalu spolehlivosti, nezamítá se nulová hypotéza H_0: $\Theta = \Theta_0$ a odhad Θ_0 je správný. Padne-li Θ_0 mimo tento interval spolehlivosti, zamítá se nulová hypotéza H_0 a odhad Θ_0 není správný. Při testování hypotéz o dvou základních souborech, které jsou vzájemně nezávislé a jejichž rozdělení je přitom normální, $x_i \sim N(\mu_x, \sigma_x^2)$ a $y_j \sim N(\mu_y, \sigma_y^2)$, charakterizovaných dvěma výběry $\{x_i\}$, $i = 1, \ldots, n_1$, a $\{y_j\}$, $j = 1, \ldots, n_2$, se nejdříve ověří shoda rozptylů testováním nulové hypotézy H_0: $\sigma_x^2 = \sigma_y^2$ proti alternativě H_A: $\sigma_x^2 \neq \sigma_y^2$ Fisherovým-Snedecorovým *F*-testem. Protože *F*-test je značně citlivý na odchylky od normality, bývá výhodnější použít Jackknifův test F_J nebo robustní testy polohy T_4, resp. T_5 pro nulovou hypotézu H_0: $\mu_x = \mu_y$ proti alternativě H_A: $\mu_x \neq \mu_y$.. Kritériem klasického Studentova *t*-testu je T_1 test, resp. T_2 test, který je robustní vůči odchylkám od heteroskedasticity, zejména pokud jsou velikosti výběrů přibližně shodné. Pro případ, že se výběry liší v šikmostech od normálního rozdělení, je vhodné užití testační statistiky modifikovaného *t*-testu T_3. Vedle testu shodnosti je výhodné použití párového testu u párových dat. U malých výběrů $4 < n < 20$ je vhodné dát přednost Hornovu postupu.

◆

1 Úvod

Statistická hypotéza je předpoklad o rozdělení pravděpodobnosti jedné nebo více náhodných veličin. Předpoklad se týká parametrů rozdělení náhodné veličiny v základním souboru nebo se může vztahovat pouze k zákonu rozdělení náhodné veličiny. *Test statistické hypotézy* je pravidlo, které na základě výsledků zjištěných z náhodného výběru objektivně předepisuje rozhodnutí, má-li být ověřovaná hypotéza zamítnuta či nikoliv. Při testování statistické hypotézy se rozlišuje testovaná *nulová hypotéza* H_0 a *alternativní hypotéza* H_A. O nulové hypotéze má test rozhodnout, zda se zamítne či nikoliv. Alternativní hypotéza je ta, kterou přijímáme, zamítneme-li hypotézu nulovou. Celý postup testování je vlastně zamítání alternativních hypotéz. Testování je předmětem řady monografií [1-19]. K testování nulové hypotézy se sestrojuje určitá testovací statistika. Padne-li tato statistika do *oboru přijetí*, nulová hypotéza se nezamítá. Padne-li však do *kritického oboru*, je nulová hypotéza zamítnuta. Pravděpodobnost padnutí testovací charakteristiky do kritického oboru se nazývá *hladina významnosti* α. Vyjadřuje se v %, jako 100α%ní hladina významnosti, např. 5%ní nebo 1%ní. Kritický obor je možno vymezit oboustranný nebo jednostranný. Oboustranný se vymezuje tehdy, neexistuje-li důvod, proč by testovací statistika měla mít buď jen kladné, nebo jen záporné znaménko. Hladina významnosti α je pak rozložena na dvě stejné části o velikosti $\alpha/2$.

2 Obecný postup testování statistických hypotéz

Postup testování statistických hypotéz lze vyjádřit těmito kroky [20-25]: (1) Formulace nulové H_0 a alternativní hypotézy H_A, (2) Volba hladiny významnosti α, (3) Volba testové statistiky, např. *t*. (4) Určení kritického oboru testové charakteristiky, např. $t_{1-\alpha/2}(n-1)$, (5) Vyčíslení testové statistiky a jejích kvantilů, (6) Rozhodnutí, zda (a) zamítnout hypotézu H_0 a přijmout H_A, jestliže testové statistika padne do kritického oboru, (b) nezamítnout hypotézu H_0, jestliže testová statistika nepadne do kritického oboru.

3 Hornův postup malých výběrů, $4 \leq n \leq 20$

Postup obsahuje tyto kroky: 1. krok: Postup je založen na pořádkových statistikách, $x_{(i)}$. 2. krok: Hloubka pivotu se vyčíslí dle vztahu $H = (\text{int}((n + 1)/2))/2$ nebo $H = (\text{int}((n + 1)/2) + 1)/2$ podle toho, které číslo vyjde celé a dolní pivot je potom $x_D = x_{(H)}$ a horní pivot $x_H = x_{(n+1-H)}$. 3. krok: Odhadem parametru polohy je *pivotová polosuma* $P_L = (x_D + x_H)/2$ a odhadem parametru rozptýlení je *pivotové rozpětí* $R_L = x_H - x_D$. 4. krok: Náhodná veličina k testování $T_L = \frac{P_L}{R_L} = \frac{x_D + x_H}{2(x_H - x_D)}$ má přibližně symetrické rozdělení, jehož vybrané kvantily $t_{L,0.975}(n)$ jsou uvedeny v **tabulce 1.** v cit. [25]. 5. krok: 95% *interval spolehlivosti střední hodnoty* se vypočte dle vztahu $P_L - R_L\ t_{L,0.975}(n) \leq \mu \leq P_L + R_L\ t_{L,0.975}(n)$.

n	1 - α =	0.9	0.95	0.975	0.99	0.995
4		0.477	0.555	0.738	1.040	1.331
5		0.869	1.370	2.094	3.715	5.805
6		0.531	0.759	1.035	1.505	1.968
7		0.451	0.550	0.720	0.978	1.211
8		0.393	0.469	0.564	0.741	0.890
9		0.484	0.688	0.915	1.265	1.575
10		0.400	0.523	0.668	0.878	1.051
11		0.363	0.452	0.545	0.714	0.859
12		0.344	0.423	0.483	0.593	0.697
13		0.389	0.497	0.608	0.792	0.945
14		0.348	0.437	0.525	0.661	0.776
15		0.318	0.399	0.466	0.586	0.685
16		0.299	0.374	0.435	0.507	0.591
17		0.331	0.421	0.502	0.637	0.774
18		0.300	0.380	0.451	0.555	0.650
19		0.288	0.361	0.423	0.502	0.575
20		0.266	0.337	0.397	0.464	0.519

4 Testy správnosti přímých měření

Testy správnosti vycházejí ze známého tvaru rozdělení pravděpodobností základního souboru, z něhož pochází náhodný výběr. Standardním předpokladem je, že ze základního souboru s rozdělením $N(\mu, \sigma^2)$ byl proveden náhodný výběr rozsahu *n*. Z dat byl vypočten výběrový průměr $\bar{x}$ a směrodatná odchylka *s*. Základní hypotéza H_0 testu správnosti se týká rovnosti neznámé střední hodnoty μ dané známé konstantě μ_0. Vhodná testová statistika je náhodná veličina $t = \frac{\bar{x} - \mu_0}{s}\sqrt{n}$. Analogicky se postupuje i u testu rovnosti rozptylu σ^2 známé hodnotě σ^2_0, kde je nulová hypotéza H_0: $\sigma^2 = \sigma_0^2$ a testová statistika je $\chi^2 = \frac{(n-1)\,s^2}{\sigma_0^2}$. Hraniční body kritického oboru představují 100α%ní kvantily známých rozdělení. Místo formálního testování, zda jsou tyto kvantily větší než testové statistiky, je možné přímo vyčíslit velikost pravděpodobnosti (1 - α) a u oboustranného testu (1 - $\alpha/2$). V řadě statistických programů jsou vyčíslené hodnoty pravděpodobnosti (1 - α) rovněž výsledkem testu. Je třeba věnovat pozornost také volbě hladiny významnosti α a objasnění výsledku testování podle následujících pravidel: 1. Pokud není nulová hypotéza H_0 zamítnuta na hladině významnosti $\alpha = 0.05$, považuje se rozdíl mezi zadanou hodnotou θ_0 a odhadem $\hat{\Theta}$ parametru θ za nevýznamný. 2. Pokud je nulová hypotéza H_0 zamítnuta i na hladině významnosti $\alpha = 0.01$, považuje se rozdíl mezi zadanou hodnotou θ_0 a odhadem $\hat{\Theta}$ za statisticky významný. 3. Pokud je nulová hypotéza H_0 zamítnuta na hladině $\alpha = 0.05$, ale není zamítnuta na hladině $\alpha = 0.01$, uvádí se, že test neposkytl pro daný rozsah výběru měření dostatečné informace k rozhodnutí. Testy významnosti úzce souvisejí s intervaly spolehlivosti. *Test správnosti* lze také vyjádřit formulací, že pokud 100(1 - α) %ní interval spolehlivosti parametru Θ obsahuje zadanou hodnotu Θ_0, nelze na hladině významnosti α zamítnout hypotézu $H_0 : \Theta = \Theta_0$. U malých výběrů je vhodnější použít Hornův postup.

5 Postup testů shodnosti parametrů dvou souborů

Porovnání dvou souborů na základě náhodných výběrů $\{x_i\}$, $i = 1, \ldots, n_1$, a $\{y_j\}$, $j = 1, \ldots, n_2$, patří k častým úlohám v přírodních i technických vědách, a to při (a) porovnání výsledků z různých měřících metod nebo laboratoří, (b) ověřování nutnosti dělení heterogenních výběrů do homogenních podskupin, (c) hodnocení rozdílu mezi rozličnými materiály a přístroji. Někdy lze tuto úlohu převést na testování jednoho výběru. To je případ *párového testu*, kdy mezi prvky obou výběrů existuje jistá logická vazba. Představují-li prvky x_i vlastnosti před úpravou materiálu

a prvky y_i po úpravě materiálu ***těchže*** vzorků ($n_1 = n_2$), lze utvořit kombinovaný jednorozměrný výběr, $D_i = x_i - y_i$, pro který lze užít klasickou statistickou analýzu. Pokud se odhad střední hodnoty $\overline{D}$ významně neliší od nuly, znamená to, že $\mu_x = \mu_y$ a efekt zpracování materiálu není pro sledovanou vlastnost statisticky významný. V obecnějším případě dvou výběrů lze zjistit, zda pocházejí ze stejného rozdělení pravděpodobnosti a zda se neliší v parametrech polohy a rozptýlení.

Před vlastní statistickou analýzou je výhodné prozkoumat nejprve metodami průzkumové analýzy rozdíly ve statistickém chování obou výběrů. Pro každý výběr se konstruuje krabicový graf a vizuálně jsou porovnávány rozdíly mezi relativním rozptýlením (délky krabicového grafu) i parametry polohy (mediánové čáry). K ověření předpokladu shody rozdělení obou výběrů se užívá *empirický graf Q-Q*. U výběrů stejného rozsahu $n_1 = n_2$ se na osu y vynášejí pořádkové statistiky $y_{(i)}$ a na osu *x* se vynášejí pořádkové statistiky $x_{(i)}$. Při shodě obou rozdělení by měly body $\{y_{(i)}, x_{(i)}\}$ ideálně ležet na přímce $y = x$.

Pokud však směrnice přímky není rovna jedné, liší se obě rozdělení o jistý násobek, úměrný velikosti směrnice. Je-li navíc úsek na ose *y* nenulový, udává jeho velikost posun středních hodnot obou výběrů. Vyjde-li tedy v empirickém grafu *Q-Q* regresní přímka $y_{(i)} = k\ x_{(i)} + q$, znamená to, že střední hodnota výběru $y_1, \ldots, y_n$ je $y = k\,x + q$ a pro rozptyl platí $s^2_y = k^2\, s^2_x$. Nelineární průběh empirického grafu *Q-Q* pak indikuje rozdíly v typu rozdělení obou výběrů.

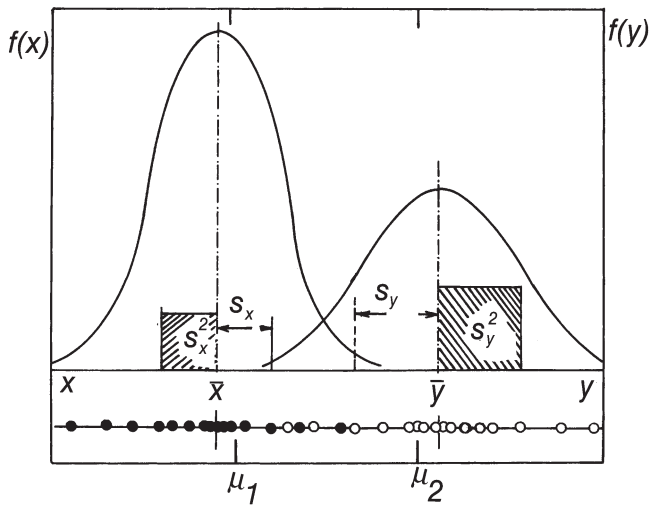
Pro praktické účely postačuje obvykle empirický graf *Q-Q* nebo krabicové grafy [20, 25]. Pokud se konstruují vrubové krabicové grafy, je možné orientačně zhodnotit význam rozdílů mezi parametry polohy, mediány. Jestliže se vruby nepřekrývají, jsou oba výběry co do parametru polohy významně odlišné [20-25].

Formálnější postup analýzy rozdílů mezi parametry polohy a rozptýlení dvou výběrů je založen na testech významnosti. Klasické testy vycházejí z předpokladů: (a) výběry $\{x_i\}$, $i = 1, \ldots, n_1$, a $\{y_j\}$, $j = 1, \ldots, n_2$ jsou vzájemně nezávislé; (b) rozdělení obou výběrů je normální, $x_i \sim N(\mu_x, \sigma^2_x)$ a $y_j \sim N(\mu_y, \sigma^2_y)$. K testování hypotézy o shodě středních hodnot nebo rozptylů existuje řada různých metod. Některé jsou použitelné i v případech, kdy jsou tyto předpoklady narušeny. Obecný postup testu shodnosti středních hodnot dvou souborů obsahuje tyto kroky:

1. *Ověření normálního rozdělení obou souborů:* testy a statistické diagnostiky k ověření předpokladů o výběru.
2. *Test shody rozptylů:*
   a) Klasický Fisherův-Snedecorovův *F*-test,
   b) Modifikovaný Fisherův-Snedecorův *F*-test,
   c) Robustní Jackknife test F_J.
3. *Test shody středních hodnot dvou souborů:*
   a) Klasický Studentův *t*-test T_1 pro homoskedasticitu,
   b) Klasický Studentův *t*-test T_2 pro heteroskedasticitu,
   c) Modifikovaný Studentův *t*-test T_3 pro výběry, odchýlené od normálního rozdělení.
   d) Robustní Jackknife test polohy T_4 pro homoskedasticitu,
   e) Robustní Jackknife test polohy T_5 pro heteroskedasticitu.

6 Testy shody rozptylů

Klasický *F-test* umožňuje ověření nulové hypotézy H_0: $\sigma^2_x = \sigma^2_y$ proti alternativě H_A: $\sigma^2_x \neq \sigma^2_y$. Vychází se z předpokladu, že oba výběry jsou nezávislé a pocházejí z normálního rozdělení, **obr. 1.**



Obr. 1. Znázornění testu shodnosti výsledků při nestejných rozptylech

Testovací kritérium má tvar $F = \max\left(\frac{s_x^2}{s_y^2}, \frac{s_y^2}{s_x^2}\right)$. Platí-li hypotéza H_0 a $s^2_x > s^2_y$, má *F* kritérium *F*-rozdělení s $v_1 = (n_1 - 1)$ a $v_2 = (n_2 - 1)$ stupni volnosti. V opačném případě se pořadí stupňů volnosti zamění. Tento klasický test je značně citlivý na předpoklad normality. Mají-li obě výběrová rozdělení jinou špičatost než odpovídá normálnímu, je třeba užít kvantil $F_{1-\alpha/2}(v_1, v_2)$ se stupni volnosti v_1 a v_2, vyčíslenými podle vztahů

$$\nu_1 = \frac{n_1 - 1}{1 + \frac{\hat{g}_{2c}}{2}} \text{ a } \nu_2 = \frac{n_2 - 1}{1 + \frac{\hat{g}_{2c}}{2}},$$

kde
$$\hat{g}_{2c} = \frac{2\ (n_1 + n_2)\left[\sum_{i=1}^{n_1}(x_i - \overline{x})^4 + \sum_{i=1}^{n_2}(y_i - \overline{y})^4\right]}{\left[\sum_{i=1}^{n_1}(x_i - \overline{x})^2 + \sum_{i=1}^{n_2}(y_i - \overline{y})^2\right]^2} - 3$$

Z dalších testů shodnosti rozptylů i pro více výběrů se jeví spolehlivým robustní Jackknife test. Testovací kritérium má tvar

$$F_J = \frac{n_1\ (\overline{z}_1 - \overline{z})^2 + n_2\ (\overline{z}_2 - \overline{z})^2}{\dfrac{\sum_{i=1}^{n_1}(z_{1i} - \overline{z}_1)^2 + \sum_{i=1}^{n_2}(z_{2i} - \overline{z}_2)^2}{n_1 + n_2 - 2}},$$

kde
$$\overline{z} = \frac{n_1\ \overline{z}_1 + n_2\ \overline{z}_2}{n_1 + n_2}$$

a
$$z_j = \frac{\sum_{i=1}^{n_j} z_{ji}}{n_j}, \quad j = 1, 2.$$

Veličiny z_{1i} se počítají podle vztahu:

$$z_{1i} = n_1 \ln s_x^2 - (n_1 - 1) \ln s_{1(i)}^2,$$

kde
$$s_{1(i)}^2 = \frac{1}{n_1 - 2}\sum_{j \neq i}^{n_1}(x_j - \overline{x}_{(i)})^2.$$

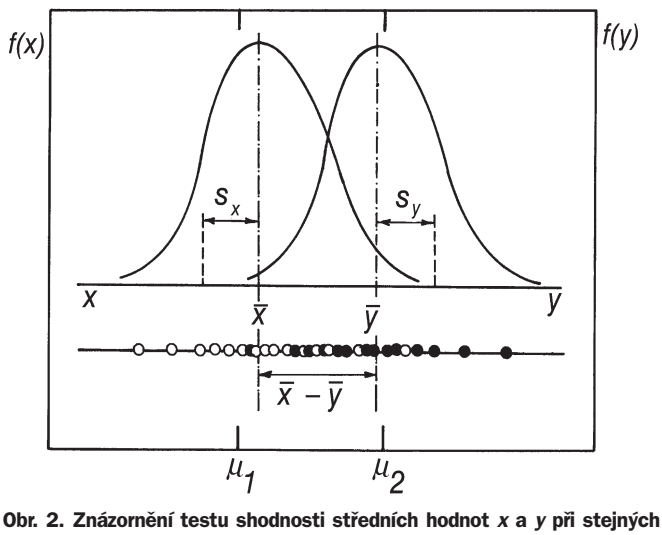
Ve vztahu se vyskytuje průměr s vynechanou i-tou hodnotou, pro který platí

$$\overline{x}_{(i)} = \frac{1}{n_1 - 1}\sum_{j \neq i}^{n_1} x_j.$$

Při výpočtu z_{2i} se ve výše uvedených vztazích dosazují hodnoty $\{y_j\}$, j = 1, ..., n_2, rozptyl s^2_y a rozsah výběru n_2. Platí-li nulová hypotéza H_0, má testovací kritérium F_J přibližně F-rozdělení s $v_1 = 2$, $v_2 = n_1 + n_2 - 2$ stupni volnosti. Vyjde-li, že $F_J > F_{1-\alpha/2}(v_1, v_2)$, je nutné zamítnout hypotézu H_0 o shodnosti obou výběrových rozptylů na hladině významnosti. Testy shody rozptylů se používají k rozhodování, zda lze při testování shody výběrových parametrů vycházet z předpokladu $\sigma^2_x = \sigma^2_y$ či nikoliv.

7 Testy shody středních hodnot („testy shodnosti“)

Studentův t-test umožňuje testování hypotézy H_0: $\mu_x = \mu_y$ proti alternativní H_A: $\mu_x \neq \mu_y$ při splnění obou uvedených předpokladů o výběrech, **obr. 2.**



Obr. 2. Znázornění testu shodnosti středních hodnot *x* a *y* při stejných rozptylech

1. Je-li $\sigma_x^2 = \sigma_y^2$, má testovací kritérium tvar

$$T_1 = \frac{|\bar{x} - \bar{y}|}{\sqrt{(n_1 - 1)\ s_x^2 + (n_2 - 1)\ s_y^2}} \sqrt{\frac{n_1\ n_2\ (n_1 + n_2 - 2)}{n_1 + n_2}}.$$

V případě platnosti hypotézy H_0 má tato testová statistika Studentovo rozdělení s $\nu = n_1 + n_2 - 2$ stupni volnosti. Platí-li, že $T_1 > t_{1-\alpha/2}(\nu)$, je hypotéza H_0 o shodě středních hodnot na hladině významnosti α zamítnuta.

2. Je-li $\sigma_x^2 \neq \sigma_y^2$, má testovací kritérium tvar $T_2 = \dfrac{|\bar{x} - \bar{y}|}{\sqrt{\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}}}$.

Platí-li hypotéza H_0, má tato testová statistika Studentovo rozdělení s „ekvivalentními" stupni volnosti

$$\nu = \frac{\left[\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}\right]^2}{\dfrac{s_x^4}{n_1^2\ (n_1 - 1)} + \dfrac{s_y^4}{n_2^2\ (n_2 - 1)}}.$$

Platí-li, že $T_2 > t_{1-\alpha/2}(\nu)$, je hypotéza H_0 o shodě středních hodnot na hladině významnosti α zamítnuta. Určitým problémem při testování je neznalost hodnot obou rozptylů σ_x^2 a σ_y^2. Bylo však zjištěno, že v případě stejných rozsahů obou výběrů $n_1 = n_2$, větších než 8, lze použít testovací kritérium T_1, i když $\sigma_x^2 \neq \sigma_y^2$, cit. [9]. Jestliže se rozsahy výběrů výrazně liší, $n_1 \neq n_2$, lze užít kritéria T_1 jenom tehdy, když je poměr výběrových rozptylů blízký jedné. Je-li $n_2 > n_1$ a navíc n_1 i n_2 jsou dostatečně veliké, je možné užít kritérium T_1. Pro $\alpha = 0.05$ pak platí nerovnost dle [10] jak

$$0.82 \leq \frac{\dfrac{n_2}{n_1}\dfrac{s_x^2}{s_y^2} + 1}{\dfrac{s_x^2}{s_y^2}\dfrac{n_2}{n_1}} \leq 1.17$$

uvádí , kde $s_x^2 > s_y^2$. Testovací kritérium T_1 není robustní vůči heteroskedasticitě, tj. případu, kdy data jsou ve výběrech měřena s různou přesností. V této situaci je správnější užít testovacího kritéria T_2, které je vůči heteroskedasticitě robustnější.

Pokud se oba výběry odchylují od normality, je vhodné použít modifikované testovací kritérium

$$T_3 = \frac{|\bar{x} - \bar{y}| + C + D\ (\bar{x} - \bar{y})^2}{\sqrt{\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}}},$$

kde

$$C = \frac{1}{6}\frac{\dfrac{\hat{g}_{1x}}{n_1^2}\dfrac{s_x^3}{\sqrt{n_1}} - \dfrac{\hat{g}_{1y}}{n_2^2}\dfrac{s_y^3}{\sqrt{n_2}}}{\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}}$$

a

$$D = \frac{1}{3}\frac{\dfrac{\hat{g}_{1x}}{n_1^2}\dfrac{s_x^3}{\sqrt{n_1}} - \dfrac{\hat{g}_{1y}}{n_2^2}\dfrac{s_y^3}{\sqrt{n_2}}}{\left[\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}\right]^2}.$$

V těchto vztazích jsou $\hat{g}_{1x}$ a $\hat{g}_{1y}$ výběrové šikmosti. Aby bylo možné užít kvantilů Studentova rozdělení pro předepsanou hladinu významnosti α, je třeba přeformulovat testovací kritérium T_3 do tvaru $T_3 = T_2 + B_x - B_y$, kde

$$B_x = \frac{\dfrac{\hat{g}_{1x}\, s_x^3}{6\ n_1^2\ \sqrt{n_1}\left[\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}\right]} + \dfrac{\hat{g}_{1x}\ s_x^2\ (\bar{x} - \bar{y})^2}{3\ n_2^2\ \sqrt{n_2}\left[\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}\right]}}{\sqrt{\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}}}$$

a B_y se vyčíslí analogicky, pouze šikmost $\hat{g}_{1x}$ se nahradí hodnotou $\hat{g}_{1y}$, rozptyl σ_x^2 hodnotou σ_y^2 a rozsah n_1 hodnotou n_2. Za předpokladu platnosti hypotézy H_0 má testovací kritérium T_3 Studentovo rozdělení s počtem stupňů volnosti ν. Test založený na kritériu T_3 je robustní vůči zešikmení výběrových rozdělení i vůči heteroskedasticitě a není u něho požadována ani shoda rozptylů, $\sigma_x^2 \neq \sigma_y^2$. Vůči odchylkám rozdělení od normality ve špičatosti jsou uvedené *t*-testy T_1, T_2 a T_3 dostatečně robustní. Je možné použít i korekcí na špičatost [10], které však nepřinášejí výrazné zlepšení.

Z ostatních typů testů shody středních hodnot uvedeme ***test Brownův a Forsythův***, který je vhodný pro obecnější testování shody *k*-tice středních hodnot, je-li definována nulová hypotéza H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$. Test vychází z předpokladu *k*-tice výběrů pocházejících z normálního rozdělení. Velikosti výběrů jsou n_i, $i = 1, \ldots, k$. Jsou určeny také průměry $\bar{x}_i$ a rozptyly s_i^2, $i = 1, \ldots, k$. Testovací kritérium je formulováno vztahem:

$$F = \frac{\sum_{i=1}^{k} n_i (\bar{x}_i - \bar{X})^2}{\sum_{i=1}^{k}\left(1 - \dfrac{n_i}{\bar{n}}\right) s_i^2},$$

kde $\bar{X} = \dfrac{1}{\bar{n}}\sum_{i=1}^{k} n_i\ \bar{x}_i$ a $\bar{n} = \sum_{i=1}^{k} n_i$.

Veličina F má za předpokladu platnosti hypotézy H_0 rozdělení F s $(k - 1)$ a ν stupni volnosti. Platí-li, že $F > F_{1-\alpha/2}(k - 1, \nu)$, hypotéza H_0 se zamítá. Počet stupňů volnosti ν se určí ze vztahu

$$\nu = \frac{1}{\sum_{i=1}^{k}\dfrac{o_i^2}{n_i - 1}},$$

kde

$$o_i = \frac{\left(1 - \dfrac{n_i}{\bar{n}}\right) s_i^2}{\sum_{i=1}^{k}\left(1 - \dfrac{n_i}{\bar{n}}\right) s_i^2}.$$

K testování shody dvou středních hodnot stačí dosadit $k = 2$, $\bar{x}_1 = \bar{x}$, $\bar{x}_2 = \bar{y}$, $s_1^2 = s_x^2$, $s_2^2 = s_y^2$.

8 Ilustrativní úlohy

Úloha 1 *Analýza malého výběru při testu správnosti obsahu síranu v pitné vodě* (**E3.10** v [25])

Ve vzorku povrchové vody, který byl 7krát vedle sebe naředěn v poměru 1:10, byly stanovovány sírany metodou RQ-flex (Merck). Proveďte analýzu využitím Hornova postupu. Deklarovaný obsah má být 50 mg. l^{-1}. Obsah síranů v povrchové vodě [mg. l^{-1}]: 50.0, 50.4, 55.0, 55.0, 55.5, 56.0, 56.5.

Řešení: (a) Hornův postup pivotů pro malé výběry ($4 < n < 20$) vede k tomuto výpočetnímu schématu:

1. Pořádkové statistiky: nejprve se seřadí prvky od nejmenší do největší hodnoty a zapíšou se do následující **tabulky:**

i	1	2	3	4	5	6	7
$x_{(i)}$	50.0	50.4	55.0	55.0	55.5	56.0	56.5

2. Hloubka pivotu: vyčíslí se hloubka pivotu pro lichou či sudou hodnotu počtu prvků n. Vlevo je zde uveden vzorec a vpravo je pak dosazení do tohoto vzorce: n = 7, liché

$$H = \text{int}\,\frac{\dfrac{n+1}{2}}{2}$$

int(2) = 2

3. Pivoty: vyčíslí se pivoty, a to dolní pivot

$x_D = x_{(H)}$ $x_{(2)} = 50.4$

a horní pivot

$x_H = x_{(n+1-H)}$ $x_{(6)} = 56.0$

4. Pivotová polosuma bude $P_L = \dfrac{x_D + x_H}{2}$ $P_L =$ 53.20

5. Pivotové rozpětí bude $R_L = x_H - x_D$ $R_L =$ 5.60

6. 95% interval spolehlivosti střední hodnoty μ: $t_{L,1-\alpha/2} = 0.720$

$P_L - R_L\ t_{L,1-\alpha/2}(n) \leq \mu \leq P_L + R_L\ t_{L,1-\alpha/2}(n)$

$53.2 - 5.6 \times 0.720 \leq \mu \leq 53.2 + 5.6 \times 0.720$

$49.17 \leq \mu \leq 57.23.$

7. Závěr: Bodový odhad míry polohy je 53.20, míry rozptýlení 5.60 a intervalový odhad míry polohy je $49.17 \leq \mu \leq 57.23$.

Úloha 2 *Test správnosti dodržení přípustné koncentrace fluoridů v pitné vodě* (**E3.19** v [25])

Předpis normy ČSN 757111-Pitná voda připouští v pitné vodě maximální koncentraci fluoridů 1.5 mg. l^{-1}. U výběru 33 dat je na hladině významnosti $\alpha = 0.05$ třeba rozhodnout, byl-li dodržen maximálně přípustný odhad střední hodnoty koncentrace fluoridů ve vodě. Je třeba rozhodnout, který odhad střední hodnoty lze použít jako statisticky nejlepší. K tomu se nejprve vyšetří, zda je rozdělení symetrické a bez odlehlých hodnot. Koncentrace fluoridů [mg. l^{-1}] v pitné vodě: 1.50, 1.54, 1.43, 1.44, 1.46, 1.44, 1.48, 1.46, 1.51, 1.42, 1.91, 1.55, 1.49, 1.55, 1.45, 1.53, 1.49, 1.50, 1.52, 1.54, 1.55, 1.46, 1.47, 1.44, 1.53, 1.52, 1.52, 1.48, 1.46, 1.58, 1.51, 1.46, 1.47.

Řešení: Z řady grafických diagnostik exploratorní analýzy dat byla zjištěna asymetrie rozdělení a bylo detekováno spíše lognormální rozdělení než normální za přítomnosti jednoho vybočujícího bodu. Z vyšetření základních předpokladů o výběru vyplývá, že data nejsou homogenní, protože výběr obsahuje 1 silně odlehlou hodnotu 1.91 mg.l^{-1}, která leží mimo vnitřní meze povoleného intervalu homogenity 1.33 a 1.65 mg.l^{-1}. Normalita je Jarque-Beraovým testem zamítnuta, protože vypočtená hladina významnosti $P = 0.00012$ je menší než zvolená $\alpha = 0.05$. Von Neumanův test prokázal, že data jsou nezávislá, protože vyčíslená hladina významnosti $P = 0.403$ je větší než zvolená $\alpha = 0.05$ a dále jsou také bez autokorelačního trendu, protože u testu vyčíslené $P = 0.515$ je větší než $\alpha = 0.05$. Protože exploratorní analýza odhalila asymetrické rozdělení (lognormální), nelze použít odhady klasických parametrů ($\bar{x} = 1.505$, $s = 0.083$, $\hat{g}_1 = 3.53$ a $\hat{g}_2 = 18.02$, které vedou k chybnému 95.0%nímu intervalovému odhadu $L_D = 1.475$ a $L_H = 1.534$). Je proto třeba použít robustní odhady nebo nejlépe data transformovat. *Robustní odhady parametrů* vedou k hodnotám mediánu $\tilde{x}_{0.5} = 1.49$, jeho směrodatné odchylky $s(\tilde{x}_{0.5}) = 0.015$ a 95.0% intervalovému odhadu $L_D = 1.459$ a $L_H = 1.521$. Prostá mocninná transformace vede k bodovému *odhadu retransformovaného průměru* $\bar{x}_R = 1.489$ a 95.0% intervalovému odhadu $L_D = 1.45$ a $L_H = 1.55$, který obsahuje předepsanou hodnotu 1.50 mg.l^{-1}.

Závěr: Z intervalových odhadů (robustních a retransformovaných) vyplývá, že normou požadovaná koncentrace fluoridů v pitné vodě 1.50 mg.l^{-1} leží v nalezeném rozmezí L_D a L_D a naměřená data jsou proto v souladu s normou.

Úloha 3 *Test shodnosti středních hodnot BSK$_5$ dvěma metodami* (**E3.36** v [25])

Koncentrace BSK$_5$ v odpadní vodě je stanovována dvěma metodami, standardní zřeďovací metodou a metodou pomocí biosenzoru. Oběma metodami byla provedeno 14 stanovení. Je třeba testovat shodnost výsledků z obou metod. Hodnoty koncentrace kyslíku [mg. l^{-1}] jsou následující: *E336a* standardní zřeďovací metodou 17.0, 14.0, 17.0, 17.0, 17.0, 14.0, 14.0, 14.0, 17.0, 17.0, 17.0, 14.0, 14.0, 17.0. *E336b* biosenzorem 14.6, 15.3, 17.8, 14.5, 13.8, 16.2, 17.6, 14.1, 14.7, 15.2, 17.5, 14.3, 13.9, 17.1.

Řešení: Z ověření základních předpokladů pro oba jednotlivé výběry vyplývá, že data jsou nezávislá, homogenní bez odlehlých bodů, test normality u obou výběrů prokázal Gaussovo rozdělení. Test shodnosti u porovnání koncentrace BSK$_5$ dvou výběrů odpadní vody (ADSTAT) vede k výsledkům: po exploratorní analýze dat a ověření normality u obou výběrů jsou vyčísleny klasické odhady parametrů polohy, rozptýlení a tvaru: průměr $\bar{x}$(A) = 15.71, $\bar{x}$(B) = 15.47, rozptyl s^2(A) = 2.37, s^2(B) = 2.17, šikmost $\hat{g}_1$(A) =-0.29, $\hat{g}_1$(B) = 0.51, špičatost $\hat{g}_2$(A) = 1.08, $\hat{g}_2$(B) = 1.68. Následuje test homogenity rozptylu Fisherovým-Snedecorovým testem nulové hypotézy H$_0$: $s_1^2 = s_2^2$ pro počet stupňů volnosti $n_1 - 1 = 13$ a $n_2 - 1 = 13$. Tabulkový kvantil $F_{1-\alpha}(n_1 - 1, n_2 - 1) = 2.637$ a experimentální hodnota činí $F_{exp} = 1.095$ vede k vyčíslené hladině významnosti $P = 0.473$ přijetí nulové hypotézy shodných rozptylů. Pak následuje vlastní test shody průměrů nulové hypotézy H$_0$: $\mu_1 = \mu_2$, ke kterému se užije Studentův *t*-test pro shodné rozptyly. Tabulkový kvantil $t_{1-\alpha/2}(\nu) = 2.055$ a experimentální hodnota Studentovy testační statistiky $T_2 = 0.426$, což je menší a proto je nulová hypotéza shodných středních hodnot přijata.

Závěr: Na hladině významnosti $\alpha = 0.05$ potvrzuje oboustranný klasický test shodu středních hodnot obou výběrů při shodnosti obou rozptylů. Obě metody vedou proto ke stejným výsledkům a se stejnou variabilitou.

Úloha 4 *Párový test obsahu železa ve vodě určeného ve dvou laboratořích* (**E3.30 v** [25])

Koncentrace železa v pitné vodě je sledována ve dvou laboratořích A a B. Vzorkování proběhlo v jeden den, voda byla nabírána z různých vodovodů. Je třeba vyšetřit, zda je možné považovat stanovení v obou laboratořích za shodná. Obsah železa [mg. l^{-1}] ve vodě dvěma laboratořemi činil: *E330a*: 0.05, 0.06, 0.06, 0.16, 0.13, 0.04, 0.24, 0.06, 0.95, 1.20, 0.11, 0.18, 0.32 a *E330b*: 0.06, 0.08, 0.07, 0.16, 0.15, 0.06, 0.26, 0.07, 0.79, 1.15, 0.26, 0.07, 0.30.

Řešení: Párový test obsahu železa ve vodě určeného ve dvou laboratořích (ADSTAT) vede k výsledkům: Průměrný rozdíl $\bar{D} = 0.006153$, rozptyl $s_D^2 = 0.08903$, počet stupňů volnosti $n - 1 = 12$, tabulkový kvantil $t_{1-\alpha/2}(n-1) = 2.179$, $t_{exp} = 0.249$ vedou k výsledku, že průměry se považují za shodné, H$_0$ je přijata na hladině významnosti 0.807.

Závěr: Párový test přijal hypotézu o shodě koncentraci železa v pitné vodě, stanovené ve dvou laboratořích.

9 Závěr

Studentův *t*-test správnosti analytického výsledku je základním testem s analogií vůči intervalu spolehlivosti. Nachází-li se totiž hodnota μ_0 (tj. Apravda@, správná hodnota, norma, standard) v intervalu spolehlivosti [L_D; L_H], je stanovení správné. Exploratorní analýza předurčí volbu, zda k testu správnosti využijeme intervalový odhad aritmetického průměru, retransformovaného průměru, mediánu nebo pivotové polosumy [20-25]. Interaktivní statistická analýza vhodným software umožnuje snadno a jednoznačně vyšetřit správnost výsledku měření. Při testování hypotéz o dvou základních souborech, charakterizovaných dvěma výběry, se nejdříve ověří shoda rozptylů Fisherovým-Snedecorovým *F*-testem. Protože *F*-test je značně citlivý na odchylky od normality, je výhodnější použít robustní Jackknifův test shodných rozptylů F_J. Klasický Studentův *t*-test T_1 pro výběry shodných rozptylů, resp. T_2 pro výběry neshodných rozptylů je poměrně robustní vůči odchylkám od normality, zejména pokud jsou velikosti výběrů přibližně shodné. Pro případ, že se výběry liší v šikmostech od normálního rozdělení, je vhodné užití testační statistiky modifikovaného Studentova *t*-testu T_3, určeného především pro případy, kdy jeden z výběrů není Gaussova rozdělení.

Poděkování
Článek vznikl za finační podpory vědeckého záměru Ministerstva školství mládeže a tělovýchovy č. MSM253100002.

Literatura

[1] Hensgaard. D.: *Commun. Statist.* **B8**, 359 (1979).
[2] Tukey J. W., McLaughlin: *Sankya* **125**, 331 (1963).
[3] Johnson N. L., Kotz S.: *Continuous Univariate Distributions*, Mifflin 1970.
[4] Hogg R. V.: *J. Amer. Statist. Assoc.* **69**, 909 (1964).
[5] Du Mond Ch., Lenth R. V.: *Technometrics* **29**, 211 (1987).
[6] Blackman N. M., Machol R. E.: *IEEE Trans. on Inform. Theory* **IT-33**, 373 (1987).
[7] Horn J.: *J. Amer. Statist. Assoc.* **78**, 930 (1983).
[8] Efron B.: *Canad. J. Statist.* **9**, 139 (1981).
[9] Posten H. O., Yeh H. C., Owen D. B.: *Commun. Statist.* **A11**, 109 (1982).
[10] Cressie N. A. C., Whitford H. J.: *Biom. J.* **28**, 131 (1986).
[11] Yuen K., Dixon W. J.: *Biometrika* **60**, 369 (1973).
[12] Owen D. B.: *Handbook of Statistical Tables*, Addison Wesley Publ. Reading 1963.
[13] Green J. R., Margerison D.: *Statistical Treatment of Experimental Data*, Elsevier, Amsterdam 1978.
[14] Miller J. C. a Miller J. N.: *Statistics for Analytical Chemistry*, Ellis Horwood, Chichester, 1984.
[15] Himmelblau D. M.: *Process Analysis by Statistical Methods*, Wiley New York 1969.
[16] Liteanu C., Rica I.: Statistical Theory and Methodology of Trace Analysis, Ellis Horwood, Chichester 1980.
[17] Anderson R. L.: Practical Statistics for Analytical Chemists, van Nostrand Reinhold Comp., New York 1987.
[18] Eason G. a kol.: Mathematics and Statistics for the Bio-Sciences, Ellis Horwood, Chichester 1980.
[19] Stoodly K.: Applied and Computation Statistics, Ellis Horwood, Chichester, 1984.
[20] Meloun M., Militký J.: Statistické zpracování experimentálních dat, Plus Praha 1994, East Publishing Praha 1998, Academia Praha 2004.
[21] Meloun M., Militký J.: Statistické zpracování experimentálních dat - Sbírka úloh s disketou,Univerzita Pardubice 1996.
[22] Meloun M., Militký J., Forina M.: Chemometrics for Analytical Chemistry, Volume 1. PC-Aided Statistical Data Analysis, Ellis Horwood Chichester, 1992.
[23] Meloun M., Militký J., Forina M.: Chemometrics for Analytical Chemistry, Volume 2. PC-Aided Regression and Related Methods, Ellis Horwood Chichester, 1994.
[24] ADSTAT, TriloByte Statistical Software s. r. o., Pardubice 1990.
[25] Meloun M., Militký J.: Kompendium statistického zpracování dat, Academia Praha 2002 (1. vydání), 2006 (2. rozšířené a doplněné vydání).

Prof. RNDr. Milan Meloun, DrSc.
Katedra analytické chemie
Chemickotechnologická fakulta

NÁZEV V ANGLIČTINĚ DO KOREKTUR

Key Words
accuracy test - equality test - pair test - Horn procedure of small symplex – heteroscedasticity - null hypothesis - alternative hypothesis - Student t-test - Fischer-Snedecor F-test

In many applications of statistics we are interested in making inferences about population characteristics on the basis of observations made on a random sample of items from population. The characteristics of interest may often be expressed in terms of popula-

tion parameters, such as the population mean or variance. In other situations we may wish to make inferences about the diference between two or more populations, such as the difference between two population means. A statistical hypothesis is a statement about the population distribution of some random variable. Hypothesis testing consists of comparing some statistical measures called test criteria deduced from a data sample with the values of these criteria taken on the assumption that a given hypothesis is correct. This article brings several frequently used tests, such as the accuracy test, the equality test, the pair test and Horn procedure of small samples with demonstration on typical problems.