Parametrické úlohy o dvou nezávislých náhodných výběrech z normálních rozložení

Motivace: V této situaci je naším úkolem porovnat střední hodnoty či rozptyly dvou normálních rozložení na základě znalosti dvou nezávislých náhodných výběrů pořízených z těchto rozložení. Zpravidla konstruujeme intervaly spolehlivosti pro rozdíl středních hodnot respektive hodnotíme shodu středních hodnot pomocí dvouvýběrového t-testu či dvouvýběrového z-testu a shodu rozptylů pomocí F-testu.

Rozložení statistik odvozených z výběrových průměrů a výběrových rozptylů

Máme dva nezávislé náhodné výběry, první pochází z rozložení $N(\mu_1, \sigma_1^2)$ a má rozsah $n_1 \geq 2$, druhý pochází z rozložení $N(\mu_2, \sigma_2^2)$ a má rozsah $n_2 \geq 2$. Označme M_1, M_2 výběrové průměry, S_1^2, S_2^2 výběrové rozptyly a $S_*^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$ vážený průměr výběrových rozptylů.

Pak platí:

a) Statistiky $M_1 - M_2$ a $S_*^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$ jsou stochasticky nezávislé.

b) $U = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$. (Pivotová statistika U slouží k řešení úloh o $\mu_1 - \mu_2$, když σ_1^2 a σ_2^2 známe.)

c) Nechť $\sigma_1^2 = \sigma_2^2 =: \sigma^2$, pak $K = \frac{(n_1+n_2-2)S_*^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2)$. (Pivotová statistika K slouží k řešení úloh o neznámém rozptylu σ^2.)

d) Jestliže $\sigma_1^2 = \sigma_2^2 =: \sigma^2$, pak $T = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{S_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$. (Pivotová statistika T slouží k řešení úloh o $\mu_1 - \mu_2$, když σ_1^2 a σ_2^2 neznáme, ale víme, že jsou shodné.)

e) $F = \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$. (Pivotová statistika F slouží k řešení úloh o σ_1^2/σ_2^2.)

Intervaly spolehlivosti pro parametrické funkce $\mu_1-\mu_2$, σ_1^2/σ_2^2

Uvedeme přehled vzorců pro meze 100(1-α)% empirických intervalů spolehlivosti pro parametrické funkce $\mu_1 - \mu_2$, σ_1^2/σ_2^2.

a) Interval spolehlivosti pro $\mu_1-\mu_2$, když σ_1^2, σ_2^2 známe (využití pivotové statistiky U)

Oboustranný: $(d, h) = \left(m_1 - m_2 - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\, u_{1-\alpha/2},\ m_1 - m_2 + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\, u_{1-\alpha/2}\right)$

Levostranný: $(d, \infty) = \left(m_1 - m_2 - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\, u_{1-\alpha},\ \infty\right)$

Pravostranný: $(-\infty, h) = \left(-\infty, m_1 - m_2 + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\, u_{1-\alpha}\right)$

b) Interval spolehlivosti pro $\mu_1-\mu_2$, když σ_1^2, σ_2^2 neznáme, ale víme, že jsou shodné (využití pivotové statistiky T)

Oboustranný:

$(d, h) = \left(m_1 - m_2 - S_*\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{1-\alpha/2}(n_1+n_2-2),\ m_1 - m_2 + S_*\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{1-\alpha/2}(n_1+n_2-2)\right)$

Levostranný: $(d, \infty) = \left(m_1 - m_2 - S_*\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{1-\alpha}(n_1+n_2-2),\ \infty\right)$

Pravostranný: $(-\infty, h) = \left(-\infty, m_1 - m_2 + S_*\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{1-\alpha}(n_1+n_2-2)\right)$

c) Interval spolehlivosti pro společný neznámý rozptyl σ^2 (využití pivotové statistiky K)

Oboustranný: $(d, h) = \left(\frac{(n_1+n_2-2)s_*^2}{\chi^2_{1-\alpha/2}(n_1+n_2-2)}, \frac{(n_1+n_2-2)s_*^2}{\chi^2_{\alpha/2}(n_1+n_2-2)} \right)$

Levostranný: $(d, \infty) = \left(\frac{(n_1+n_2-2)s_*^2}{\chi^2_{1-\alpha}(n_1+n_2-2)}, \infty \right)$

Pravostranný: $(-\infty, h) = \left(-\infty, \frac{(n_1+n_2-2)s_*^2}{\chi^2_{\alpha}(n_1+n_2-2)} \right)$

d) Interval spolehlivosti pro podíl rozptylů $\frac{\sigma_1^2}{\sigma_2^2}$ (využití pivotové statistiky F)

Oboustranný: $(d, h) = \left(\frac{s_1^2/s_2^2}{F_{1-\alpha/2}(n_1-1, n_2-1)}, \frac{s_1^2/s_2^2}{F_{\alpha/2}(n_1-1, n_2-1)} \right)$

Levostranný: $(d, \infty) = \left(\frac{s_1^2/s_2^2}{F_{1-\alpha}(n_1-1, n_2-1)}, \infty \right)$

Pravostranný: $(-\infty, h) = \left(-\infty, \frac{s_1^2/s_2^2}{F_{\alpha}(n_1-1, n_2-1)} \right)$

Upozornění: Není-li v bodě (b) splněn předpoklad o shodě rozptylů, lze sestrojit aspoň přibližný 100(1-α)% interval spolehlivosti pro $\mu_1-\mu_2$.

V tomto případě má statistika T přibližně rozložení $t(\nu)$, kde počet stupňů volnosti $\nu = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1-1}+\frac{\left(s_2^2/n_2\right)^2}{n_2-1}}$. Není-li ν celé číslo, použijeme v tabulkách kvantilů Studentova rozložení lineární interpolaci.

Příklad: U 48 náhodně vybraných studentek VŠE v Praze byla zjišťována výška (veličina X, v cm), známka z matematiky v 1. semestru (veličina Y, má varianty výborně, velmi dobře, dobře) a obor studia (veličina Z, má varianty nh – národní hospodářství, inf – informatika).Údaje o výšce studentek oborů národní hospodářství a informatika považujeme za realizace dvou nezávislých náhodných výběrů z rozložení $N(\mu_1, \sigma^2)$ a $N(\mu_2, \sigma^2)$. Sestrojte 95% empirický interval spolehlivosti pro rozdíl středních hodnot $\mu_1 - \mu_2$.

Výpočet pomocí systému STATISTICA:

Statistika – Základní statistiky/tabulky – t-test, nezávislé, dle skupin – OK – Proměnné – Závislé X, Grupovací Z – OK – na záložce Možnosti zaškrtneme Meze spol. pro odhady – ponecháme implicitní hodnotu 95% - Výpočet.

Ve výstupní tabulce nás zajímají poslední dva sloupce:

	t-testy; grupováno: Z: obor stu Skup. 1: nh: narodni hospodar Skup. 2: inf: informatika	
Promě	Int. spol -95,000	Int. spol +95,000
X	-0,450	6,293

Vidíme, že -0,45 cm < $\mu_1 - \mu_2$ < 6,29 cm s pravděpodobností aspoň 0,95.

Jednotlivé typy testů o parametrických funkcích $\mu_1-\mu_2$, σ_1^2/σ_2^2

a) Nechť $X_{11}, \ldots, X_{1n_1}$ je náhodný výběr z rozložení $N(\mu_1, \sigma_1^2)$ a $X_{21}, \ldots, X_{2n_2}$ je na něm nezávislý náhodný výběr z rozložení $N(\mu_2, \sigma_2^2)$, přičemž $n_1 \geq 2$, $n_2 \geq 2$ a σ_1^2, σ_2^2 známe. Nechť c je konstanta.
Test $H_0: \mu_1 - \mu_2 = c$ proti $H_1: \mu_1 - \mu_2 \neq c$ se nazývá dvouvýběrový z-test.

b) Nechť $X_{11}, \ldots, X_{1n_1}$ je náhodný výběr z rozložení $N(\mu_1, \sigma^2)$ a $X_{21}, \ldots, X_{2n_2}$ je na něm nezávislý náhodný výběr rozložení $N(\mu_2, \sigma^2)$, přičemž $n_1 \geq 2$ a $n_2 \geq 2$ a σ^2 neznáme. Nechť c je konstanta.
Test $H_0: \mu_1 - \mu_2 = c$ proti $H_1: \mu_1 - \mu_2 \neq c$ se nazývá dvouvýběrový t-test.

c) Nechť $X_{11}, \ldots, X_{1n_1}$ je náhodný výběr z rozložení $N(\mu_1, \sigma_1^2)$ a $X_{21}, \ldots, X_{2n_2}$ je na něm nezávislý náhodný výběr rozložení $N(\mu_2, \sigma_2^2)$, přičemž $n_1 \geq 2$ a $n_2 \geq 2$. Test $H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$ proti $H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$ se nazývá F-test.

Provedení testů o parametrických funkcích $\mu_1-\mu_2$, σ_1^2/σ_2^2 **pomocí kritického oboru**

a) Provedení dvouvýběrového z-testu

Vypočteme realizaci testového kritéria $t_0 = \dfrac{M_1 - M_2 - c}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$.

Stanovíme kritický obor W.

Pokud $t_0 \in W$, H_0 zamítáme na hladině významnosti α a přijímáme H_1.

Oboustranný test: Testujeme H_0: $\mu_1 - \mu_2 = c$ proti H_1: $\mu_1 - \mu_2 \neq c$. Kritický obor má tvar: $W = (-\infty, -u_{1-\alpha/2}\rangle \cup \langle u_{1-\alpha/2}, \infty)$.

Levostranný test: Testujeme H_0: $\mu_1 - \mu_2 = c$ proti H_1: $\mu_1 - \mu_2 < c$. Kritický obor má tvar: $W = (-\infty, -u_{1-\alpha}\rangle$.

Pravostranný test: Testujeme H_0: $\mu_1 - \mu_2 = c$ proti H_1: $\mu_1 - \mu_2 > c$. Kritický obor má tvar: $W = \langle u_{1-\alpha}, \infty)$.

b) Provedení dvouvýběrového t-testu

Vypočteme realizaci testového kritéria $t_0 = \frac{M_1 - M_2 - c}{S_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$.

Stanovíme kritický obor W.
Pokud $t_0 \in W$, H_0 zamítáme na hladině významnosti α a přijímáme H_1.

Oboustranný test: Testujeme H_0: $\mu_1 - \mu_2 = c$ proti H_1: $\mu_1 - \mu_2 \neq c$. Kritický obor má tvar:
$W = \left(-\infty, -t_{1-\alpha/2}(n_1+n_2-2)\right\rangle \cup \left\langle t_{1-\alpha/2}(n_1+n_2-2), \infty\right)$.

Levostranný test: Testujeme H_0: $\mu_1 - \mu_2 = c$ proti H_1: $\mu_1 - \mu_2 < c$. Kritický obor má tvar: $W = \left(-\infty, -t_{1-\alpha}(n_1+n_2-2)\right\rangle$.

Pravostranný test: Testujeme H_0: $\mu_1 - \mu_2 = c$ proti H_1: $\mu_1 - \mu_2 > c$. Kritický obor má tvar: $W = \left\langle t_{1-\alpha}(n_1+n_2-2), \infty\right)$.

c) Provedení F-testu

Vypočteme realizaci testového kritéria $t_0 = \frac{S_1^2}{S_2^2}$.

Stanovíme kritický obor W.
Pokud $t_0 \in W$, H_0 zamítáme na hladině významnosti α a přijímáme H_1.

Oboustranný test: Testujeme H_0: $\frac{\sigma_1^2}{\sigma_2^2} = 1$ proti H_1: $\frac{\sigma_1^2}{\sigma_2^2} \neq 1$. Kritický obor má tvar:

$$W = \left(0, F_{\alpha/2}(n_1-1, n_2-1)\right\rangle \cup \left\langle F_{1-\alpha/2}(n_1-1, n_2-1), \infty\right).$$

Levostranný test: Testujeme H_0: $\frac{\sigma_1^2}{\sigma_2^2} = 1$ proti H_1: $\frac{\sigma_1^2}{\sigma_2^2} < 1$. Kritický obor má tvar: $W = \left(0, F_{\alpha}(n_1-1, n_2-1)\right\rangle$.

Pravostranný test: Testujeme H_0: $\frac{\sigma_1^2}{\sigma_2^2} = 1$ proti H_1: $\frac{\sigma_1^2}{\sigma_2^2} > 1$. Kritický obor má tvar: $W = \left\langle F_{1-\alpha}(n_1-1, n_2-1), \infty\right)$.

Příklad: V restauraci "U bílého koníčka" měřili ve 20 případech čas obsluhy zákazníka. Výsledky v minutách: 6, 8, 11, 4, 7, 6, 10, 6, 9, 8, 5, 12, 13, 10, 9, 8, 7, 11, 10, 5. V restauraci "Zlatý lev" bylo dané pozorování uskutečněno v 15 případech s těmito výsledky: 9, 11, 10, 7, 6, 4, 8, 13, 5, 15, 8, 5, 6, 8 ,7. Za předpokladu, že uvedené hodnoty pocházejí ze dvou normálních rozložení, na hladině významnosti 0,05 testujte hypotézu, že střední hodnoty doby obsluhy jsou v obou restauracích stejné.

Výpočet pomocí systému STATISTICA:

Otevřeme datový soubor restaurace.sta o dvou proměnných a 35 případech. První proměnnou nazvaná OBSLUHA obsahuje doby obluhy, druhá proměnná ID slouží k rozlišení první a druhé restaurace.

Provedeme dvouvýběrový t-test současně s testem o shodě rozptylů: Statistika – Základní statistiky a tabulky – t-test, nezávislé, dle skupin – OK, Proměnné – Závislé proměnné OBSLUHA, Grupovací proměnná ID – OK.

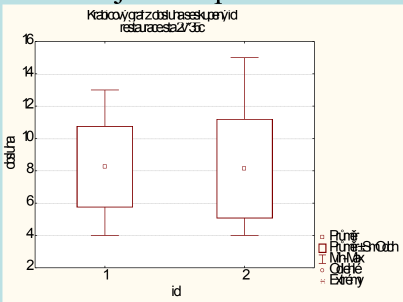
Po kliknutí na tlačítko Souhrn dostaneme tabulku

Proměi	Prům 1	Prům 2	t	sv	p	Poč.p 1	Poč.pl 2	Sm.od 1	Sm.od 2	F-pom Rozpt	p Rozpt
obsluh	8,250	8,133	0,123	3:	0,902	2(	1:	2,510	3,067	1,492	0,410

Testová statistika pro test shody rozptylů se realizuje hodnotou 1,492952 odpovídající p-hodnota je 0,41044, tedy na hladině významnosti 0,05 nezamítáme hypotézu o shodě rozptylů. (Upozornění: v případě zamítnutí hypotézy o shodě rozptylů je zapotřebí v tabulce t-testu pro nezávislé vzorky dle skupin zaškrtnout volbu Test se samostatnými odhady rozptylu.)

Testová statistika pro test shody středních hodnot se realizuje hodnotou 0,12373, počet stupňů volnosti je 33, odpovídající p-hodnota 0,902279, tedy hypotézu o shodě středních hodnot nezamítáme na hladině významnosti 0,05. Znamená to, že se neprokázal rozdíl ve středních hodnotách dob obsluhy v restauracích "U bílého koníčka" a „Zlatý lev“.

Tabulku ještě doplníme krabicovými diagramy. Na záložce Detaily zaškrtneme krabicový graf a vybereme volbu Průměr/SmOdch/Min-Max.



Z grafu je vidět, že průměrná doba obsluhy v první restauraci je nepatrně delší a má menší variabilitu než ve druhé restauraci. Extrémní ani odlehlé hodnoty se zde nevyskytují.

Cohenův koeficient věcného účinku – doplnění významu dvouvýběrového t-testu:

Nechť $X_{11},\dots,X_{1n_1}$ je náhodný výběr z rozložení $N(\mu_1, \sigma^2)$ a $X_{21},\dots,X_{2n_2}$ je na něm nezávislý náhodný výběr rozložení $N(\mu_2, \sigma^2)$, přičemž $n_1 \geq 2$ a $n_2 \geq 2$ a σ^2 neznáme. Nechť c je konstanta.
Testujeme H_0: $\mu_1 - \mu_2 = 0$ proti H_1: $\mu_1 - \mu_2 \neq 0$. Označme m_1, m_2 realizace výběrových průměrů hodnot dané veličiny v těchto dvou skupinách, s_1^2, s_2^2 realizace výběrových rozptylů a $s_*^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$ realizaci váženého průměru výběrových rozptylů.

Cohenův koeficient d vypočteme podle vzorce: $d = \frac{|m_1 - m_2|}{s_*}$.

Tento koeficient slouží k posouzení velikosti rozdílu průměrů, který je standardizován pomocí odmocniny z váženého průměru výběrových rozptylů. Jedná se o tzv. věcnou významnost neboli velikost účinku skupiny na variabilitu hodnot sledované náhodné veličiny. Velikost účinku hodnotíme podle následující tabulky:

Hodnota d	účinek
aspoň 0,8	velký
mezi 0,5 až 0,8	střední
mezi 0,2 až 0,5	malý
pod 0,2	zanedbatelný

(Uvedené hodnoty nemají samozřejmě absolutní platnost, posouzení, jaký účinek považujeme za velký či malý, závisí na kontextu.)

Je zapotřebí si uvědomit, že při dostatečně velkých rozsazích náhodných výběrů i malý rozdíl ve výběrových průměrech způsobí zamítnutí nulové hypotézy na hladině významnosti α, i když z věcného hlediska tak malý rozdíl nemá význam. Naopak, máme-li výběry malých rozsahů, pak i značně velký rozdíl ve výběrových průměrech nemusí vést k zamítnutí nulové hypotézy na hladině významnosti α.

Příklad:

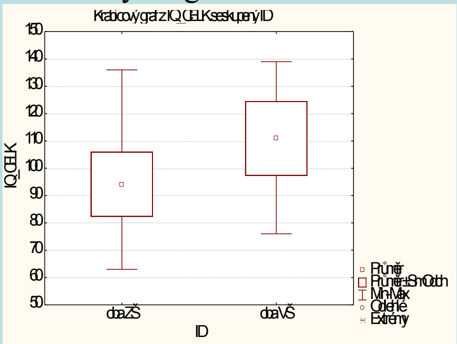
Máme k dispozici údaje o celkovém IQ 856 žáků ZŠ. Zajímáme se jednak o skupinu dětí, jejichž oba rodiče mají pouze základní vzdělání (je jich 296) a jednak o skupinu dětí, jejichž oba rodiče mají vysokoškolské vzdělání (těch je 75). Na hladině významnosti 0,05 budeme testovat hypotézu, že střední hodnota celkového IQ je v obou skupinách stejná a také vypočteme Cohenův koeficient věcného účinku.

Řešení: Provedeme dvouvýběrový t-test:

	t-testy; grupováno:ZŠ a VŠ (IQ) Skup. 1: oba ZŠ Skup. 2: oba VŠ										
Proměi	Průmě oba Z	Průmě oba V	t	sv	p	Poč.p oba Z	Poč.pl oba V	Sm.od oba Z	Sm.od oba V	F-pom Rozpt	p Rozpt
IQ CE	94,13	110,9	-10,6	36	0,000	29	7	11,82	13,60	1,322	0,110

Hypotézu o shodě středních hodnot zamítáme na hladině významnosti 0,05, protože odpovídající p-hodnota je velmi blízká 0 (hypotézu o shodě rozptylů nezamítáme na hladině významnosti 0,05, p-hodnota F-testu je 0,110124, což je větší než 0,05).

Krabicový diagram:



Vidíme, že průměrné celkové IQ dětí v 1. skupině je 94,1, zatímco ve 2. skupině 110,9. Vliv skupiny na variabilitu hodnot celkového IQ posoudíme pomocí Cohenova koeficientu.

	1 n1	2 n2	3 m1	4 m2	5 s1	6 s2	7 d
1	29	7	94,13	110,9	11,82	13,60	1,374

Cohenův koeficient nabývá hodnoty 1,37, tudíž vliv skupiny na variabilitu hodnot celkového IQ lze považovat za velký.

Příklad: Výrobce limonád chtěl zjistit, zda změna technologie výroby se projeví v prodeji limonád. Proto sledoval po 14 náhodně vybraných dnů před zavedením nových limonád tržby v určitém regionu a zjistil, že za den utržil v průměru 39 600 Kč se směrodatnou odchylkou 5 060 Kč. Po zavedení nových limonád prověřil stejným způsobem tržby v 11 náhodně vybraných dnech v témž regionu a zjistil průměrný příjem 41 200 Kč se směrodatnou odchylkou 4 310 Kč. Předpokládejte, že tržby za starý typ limonád se řídí rozložením $N(\mu_1, \sigma^2)$ a tržby za nový typ limonád se řídí rozložením $N(\mu_2, \sigma^2)$.
Na hladině významnosti 0,05 testujte hypotézu H_0: $\mu_1 - \mu_2 = 0$ proti oboustranné alternativě H_1: $\mu_1 - \mu_2 \neq 0$.

Řešení:
Za odhad společného neznámého rozptylu vezmeme vážený průměr výběrových rozptylů:

$$s_*^2 = \frac{13 \cdot 5060^2 + 10 \cdot 4310^2}{23} = 22548165{,}2$$

Realizace testového kritéria:

$$t_0 = \frac{m_1 - m_2 - c}{s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{39600 - 41200}{\sqrt{22548165{,}2}\sqrt{\frac{1}{14} + \frac{1}{11}}} = -0{,}836$$

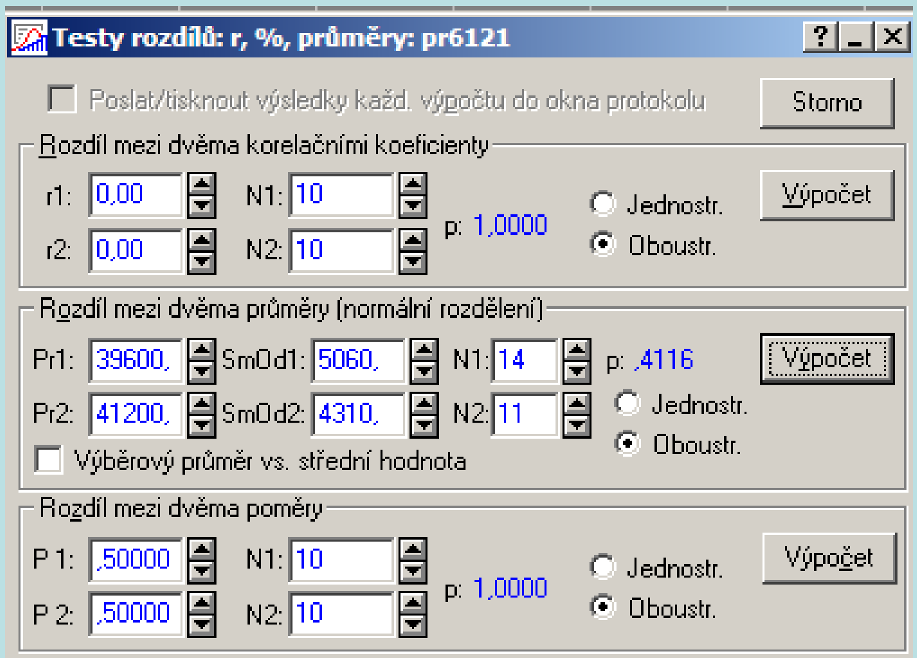
Kritický obor:

$$W = \left(-\infty, -t_{1-\alpha/2}(n_1+n_2-2)\right\rangle \cup \left\langle t_{1-\alpha/2}(n_1+n_2-2), \infty\right) =$$
$$= \left(-\infty, -t_{0{,}975}(23)\right\rangle \cup \left\langle t_{0{,}975}(23), \infty\right) = \left(-\infty, -2{,}0687\right\rangle \cup \left\langle 2{,}0687, \infty\right)$$

Protože testové kritérium se nerealizuje v kritickém oboru, na hladině významnosti 0,05 nelze zamítnout hypotézu o shodě středních hodnot.

Výpočet pomocí systému STATISTICA:

Statistiky – Základní statistiky a tabulky – Testy rozdílů: r, %, průměry – OK – vybereme Rozdíl mezi dvěma průměry (normální rozdělení) – do políčka Pr1 napíšeme 39600, do políčka SmOd1 napíšeme 5600, do políčka N1 napíšeme 14, do políčka Pr2 napíšeme 41200, do políčka SmOd1 napíšeme 4310, do políčka N1 napíšeme 14 - Výpočet. Dostaneme p-hodnotu 0,4116 tedy nezamítáme nulovou hypotézu na hladině významnosti 0,05.



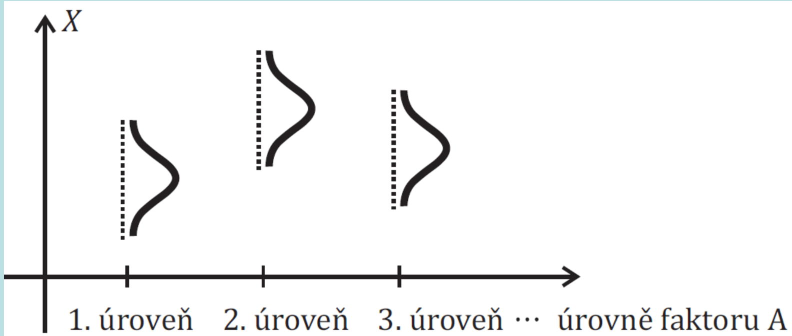
Jelikož p-hodnota je větší než hladina významnosti 0,05, H_0 nezamítáme na hladině významnosti 0,05. Znamená to, že změna technologie výroby se neprojevila ve střední hodnotě tržeb.

Analýza rozptylu jednoduchého třídění

Motivace: Zajímáme se o problém, zda lze určitým faktorem (tj. nominální náhodnou veličinou A) vysvětlit variabilitu pozorovaných hodnot náhodné veličiny X, která je intervalového či poměrového typu. Např. zkoumáme, zda metoda výuky určitého předmětu (faktor A) ovlivňuje počet bodů dosažených studenty v závěrečném testu (náhodná veličina X).
Předpokládáme, že faktor A má $r \geq 3$ úrovní a přitom i-té úrovni odpovídá n_i pozorování $X_{i1}, \ldots, X_{in_i}$, které tvoří náhodný výběr z rozložení $N(\mu_i, \sigma^2)$, $i = 1, \ldots, r$ a jednotlivé náhodné výběry jsou stochasticky nezávislé, tedy $X_{ij} = \mu_i + \varepsilon_{ij}$, kde ε_{ij} jsou stochasticky nezávislé náhodné veličiny s rozložením $N(0, \sigma^2)$, $i = 1, \ldots, r$, $j = 1, \ldots, n_i$.
Výsledky lze zapsat do tabulky

faktor A	výsledky
úroveň 1	$X_{11}, \ldots, X_{1n_1}$
úroveň 2	$X_{21}, \ldots, X_{2n_2}$
…	…
úroveň r	$X_{r1}, \ldots, X_{rn_r}$

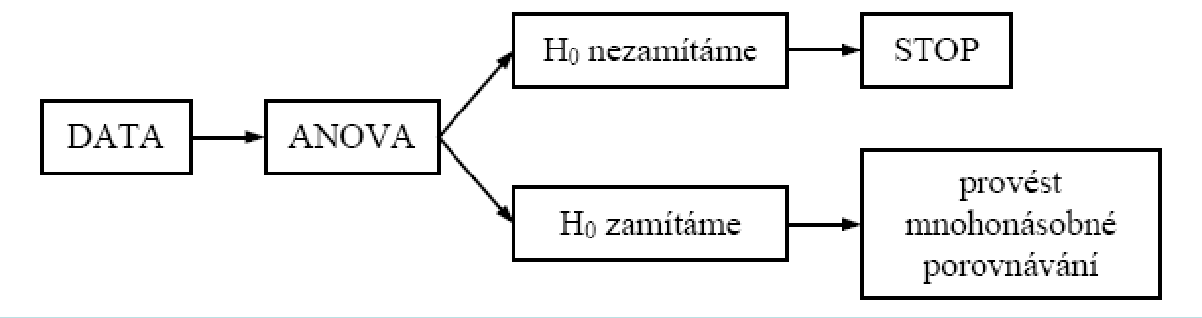
Ilustrace:



Na hladině významnosti α testujeme nulovou hypotézu, která tvrdí, že všechny střední hodnoty jsou stejné, tj. H_0: $\mu_1 = \ldots = \mu_r$ proti alternativní hypotéze H_1, která tvrdí, že aspoň jedna dvojice středních hodnot se liší.

Jedná se tedy o zobecnění dvouvýběrového t-testu a na první pohled se zdá, že stačí utvořit $\binom{r}{2}$ dvojic náhodných výběrů a na každou dvojici aplikovat dvouvýběrový t-test. Hypotézu o shodě všech středních hodnot bychom pak zamítli, pokud aspoň v jednom případě z $\binom{r}{2}$ porovnávání se prokáže odlišnost středních hodnot. Odtud je vidět, že k neoprávněnému zamítnutí nulové hypotézy (tj. k chybě 1. druhu) může dojít s pravděpodobností větší než α. Proto ve 30. letech 20. století vytvořil R. A. Fisher metodu ANOVA (analýza rozptylu, v popsané situaci konkrétně analýza rozptylu jednoduchého třídění), která uvedenou podmínku splňuje.

Pokud na hladině významnosti α zamítneme nulovou hypotézu, zajímá nás, které dvojice středních hodnot se od sebe liší. K řešení tohoto problému slouží metody mnohonásobného porovnávání, např. Schefféo nebo Tukeyova metoda.



Označení:

V analýze rozptylu jednoduchého třídění se používá tzv. tečková notace.

$n = \sum_{i=1}^{r} n_i$ … celkový rozsah všech r výběrů

$X_{i.} = \sum_{j=1}^{n_i} X_{ij}$ … součet hodnot v i-tém výběru

$M_{i.} = \frac{1}{n_i} X_{i.}$ … výběrový průměr v i-tém výběru

$X_{..} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} X_{ij}$ … součet hodnot všech výběrů

$M_{..} = \frac{1}{n} X_{..}$ … celkový průměr všech r výběrů

Zavedeme součty čtverců

$S_T = \sum_{i=1}^{r} \sum_{j=1}^{n_i} \left(X_{ij} - M_{\cdot\cdot}\right)^2$ … celkový součet čtverců (charakterizuje variabilitu jednotlivých pozorování kolem celkového průměru),

počet stupňů volnosti $f_T = n - 1$,

$S_A = \sum_{i=1}^{r} n_i \left(M_{i\cdot} - M_{\cdot\cdot}\right)^2$ … skupinový součet čtverců (charakterizuje variabilitu mezi jednotlivými náhodnými výběry),

počet stupňů volnosti $f_A = r - 1$.

Sčítanec $\left(M_{i\cdot} - M_{\cdot\cdot}\right)$ představuje bodový odhad efektu α_i.

$S_E = \sum_{i=1}^{r} \sum_{j=1}^{n_i} \left(X_{ij} - M_{i\cdot}\right)^2$ … reziduální součet čtverců (charakterizuje variabilitu uvnitř jednotlivých výběrů),

počet stupňů volnosti $f_E = n - r$.

Lze dokázat, že $S_T = S_A + S_E$.

Testování hypotézy o shodě středních hodnot

Náhodné veličiny X_{ij} se řídí modelem

M0: $X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

pro $i = 1, \ldots, r$, $j = 1, \ldots, n_i$, přičemž

ε_{ij} jsou stochasticky nezávislé náhodné veličiny s rozložením $N(0, \sigma^2)$,

μ je společná část střední hodnoty závisle proměnné veličiny,

α_i je efekt faktoru A na úrovni i.

Parametry μ, α_i neznáme.

Požadujeme, aby platila tzv. reparametrizační rovnice: $\sum^{r}$ [illegible].

(Pokud je třídění vyvážené, tj. pokud mají všechny výběry stejný rozsah: $n_1 = n_2 = \ldots = n_r$, pak lze použít zjednodušenou podmínku $\sum^{r}$ [illegible].)

Kdyby nezáleželo na faktoru A, platila by hypotéza $\alpha_1 = \ldots = \alpha_r = 0$ a dostali bychom model
M1: $X_{ij} = \mu + \varepsilon_{ij}$.
Během analýzy rozptylu tedy zkoumáme, zda výběrové průměry $M_1, \ldots, M_r$ se od sebe liší pouze v mezích náhodného kolísání kolem celkového průměru M nebo zda se projevuje vliv faktoru A.

Rozdíl mezi modely M0 a M1 ověřujeme pomocí testové statistiky
$F_A = \frac{S_A/f_A}{S_E/f_E}$, která se řídí rozložením F(r-1,n-r), je-li model M1 správný. Hypotézu o nevýznamnosti faktoru A tedy zamítneme na hladině významnosti α, když platí: $F_A \geq F_{1-\alpha}(r-1,n-r)$.

Výsledky výpočtů zapisujeme do tabulky analýzy rozptylu jednoduchého třídění.

Zdroj variability	součet čtverců	stupně volnosti	podíl	F_A
skupiny	S_A	$f_A = r - 1$	S_A/f_A	$\frac{S_A/f_A}{S_E/f_E}$
reziduální	S_E	$f_E = n - r$	S_E/f_E	-
celkový	S_T	$f_T = n - 1$	-	-

Sílu závislosti náhodné veličiny X na faktoru A můžeme měřit pomocí poměru determinace: $P^2 = \frac{S_A}{S_T}$. Nabývá hodnot z intervalu $\langle 0,1 \rangle$.

Testování hypotézy o shodě rozptylů

Před provedením analýzy rozptylu je zapotřebí ověřit předpoklad o shodě rozptylů v daných r výběrech.

a) Levenův test: Položme $Z_{ij} = |X_{ij} - M_i|$. Označíme

$$M_{Zi} = \frac{1}{n_i}\sum_{j=1}^{n_i} Z_{ij},$$

$$M_Z = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i} Z_{ij},$$

$$S_{ZE} = \sum_{i=1}^{r}\sum_{j=1}^{n_i} (Z_{ij} - M_{Zi})^2,$$

$$S_{ZA} = \sum_{i=1}^{r} n_i (M_{Zi} - M_Z)^2$$

Platí-li hypotéza o shodě rozptylů, pak statistika

$$F_{ZA} = \frac{S_{ZA}/(r-1)}{S_{ZE}/(n-r)} \approx F(r-1, n-r).$$

Hypotézu o shodě rozptylů tedy zamítáme na asymptotické hladině významnosti α, když $F_{ZA} \geq F_{1-\alpha}(r-1, n-r)$.

(Levenův test je vlastně založen na analýze rozptylu absolutních hodnot centrovaných pozorování. Vzhledem k tomu, že náhodné veličiny $X_{ij} - M_i$ nejsou stochasticky nezávislé a absolutní hodnoty těchto veličin nemají normální rozložení, je Levenův test pouze aproximativní.)

b) Brownův – Forsytheův test je modifikací Levenova testu. Modifikace spočívá v tom, že místo výběrového průměru i-tého výběru se při výpočtu veličiny Z_{ij} používá medián i-tého výběru.

Post – hoc metody mnohonásobného porovnávání

Zamítneme-li na hladině významnosti α hypotézu o shodě středních hodnot, chceme zjistit, které dvojice středních hodnot se liší na dané hladině významnosti α, tj. na hladině významnosti α testujeme H_0: $\mu_l = \mu_k$ proti H_1: $\mu_l \neq \mu_k$ pro všechna $l, k = 1, .., r$, $l \neq k$.

a) Mají-li všechny výběry týž rozsah p (říkáme, že třídění je vyvážené), použijeme Tukeyovu metodu.

Testová statistika má tvar $\frac{|M_k - M_l|}{\frac{S_*}{\sqrt{p}}}$. Rovnost středních hodnot μ_k a μ_l zamítneme na hladině významnosti α, když

$\frac{|M_k - M_l|}{\frac{S_*}{\sqrt{p}}} \geq q_{1-\alpha}(r, n-r)$, kde hodnoty $q_{1-\alpha}(r, n-r)$ jsou kvantily studentizovaného rozpětí a najdeme je ve statistických tabulkách. (Studentizované rozpětí je náhodná veličina $Q = \frac{X_{(n)} - X_{(1)}}{S}$.)

Existuje modifikace Tukeyovy metody pro nestejné rozsahy výběrů, nazývá se Tukeyova HSD metoda. V tomto případě má testová statistika tvar $\frac{|M_k - M_l|}{S_*\sqrt{\frac{1}{2}\left(\frac{1}{n_k} + \frac{1}{n_l}\right)}}$. Rovnost středních hodnot μ_k a μ_l zamítneme na hladině významnosti α, když

$$\frac{|M_k - M_l|}{S_*\sqrt{\frac{1}{2}\left(\frac{1}{n_k} + \frac{1}{n_l}\right)}} \geq q_{1-\alpha}(r, n-r).$$

b) Nemají-li všechny výběry stejný rozsah, použijeme Schefféhο metodu: rovnost středních hodnot μ_k a μ_l zamítneme na hladině významnosti α, když

$$|M_k - M_l| \geq S_* \sqrt{(r-1)\left(\frac{1}{n_k}+\frac{1}{n_l}\right)F_{1-\alpha}(r-1, n-r)}.$$

Výhodou Schefféhο testu je, že k jeho provedení nepotřebujeme speciální statistické tabulky s hodnotami kvantilů studentizovaného rozpětí, ale stačí běžné statistické tabulky s kvantily Fisherova – Snedecorova rozložení.

V případě vyváženého třídění, kdy lze aplikovat Tukeyovu i Schefféhο metodu, použijeme tu, která je citlivější. Tukeyova metoda tedy bude výhodnější, když
$q_{1-\alpha}^2(r, n-r) < 2(r-1)F_{1-\alpha}(r-1, n-r)$.

Metody mnohonásobného porovnávání mají obecně menší sílu než ANOVA.
Může nastat situace, kdy při zamítnutí H_0 nenajdeme metodami mnohonásobného porovnávání významný rozdíl u žádné dvojice středních hodnot. K tomu dochází zvláště tehdy, když p-hodnota pro ANOVU je jen o málo nižší než zvolená hladina významnosti. Pak slabší test patřící do skupiny metod mnohonásobného porovnávání nemusí odhalit žádný rozdíl.

Doporučený postup při provádění analýzy rozptylu:

a) Ověření normality daných r náhodných výběrů (grafické metody - NP plot, Q-Q plot, histogram, testy hypotéz o normálním rozložení - Lilieforsova varianta Kolmogorovova – Smirnovova testu nebo Shapirův – Wilkův test).
Doporučuje se kombinace obou způsobů. Závěry učiníme až na základě posouzení obou výsledků.
Obecně lze říci, že analýza rozptylu není příliš citlivá na porušení předpokladu normality, zvláště při větších rozsazích výběrů (nad 20), což je důsledek působení centrální limitní věty. Mírné porušení normality tedy není na závadu, při větším porušení použijeme např. Kruskalův – Wallisův test jako neparametrickou obdobu analýzy rozptylu jednoduchého třídění.

b) Po ověření normality se testuje homogenitu rozptylů, tj. předpoklad, že všechny náhodné výběry pocházejí z normálních rozložení s týmž rozpylem. Graficky ověřujeme shodu rozptylů pomocí krabicových diagramů, kdy sledujeme, zda je šířka krabic stejná. Numericky testujeme homogenitu rozptylů pomocí Levenova testu, Brownova – Forsytheova testu (oba jsou implementovány ve STATISTICE, Brownův – Forsytheův test v MINITABu) či Bartlettova testu (je k dispozici v MINITABu).
Slabé porušení homogenity rozptylů nevadí, při větším se doporučuje mediánový test.

c) Pokud jsou splněny předpoklady normality a homogenity rozptylů, můžeme přistoupit k testování shody středních hodnot. Předtím je samozřejmě vhodné vypočítat průměry a směrodatné odchylky či rozptyly v jednotlivých skupinách.

d) Dojde-li na zvolené hladině významnosti k zamítnutí hypotézy o shodě středních hodnot, zajímá nás, které dvojice středních hodnot se od sebe liší. K řešení tohoto problému slouží post-hoc metody mnohonásobného porovnávání, např. Scheffého nebo Tukeyova metoda.

Příklad: V jisté továrně se měřil čas, který potřeboval každý ze tří dělníků k uskutečnění téhož pracovního úkonu. Čas v minutách:

1. dělník: 3,6 3,8 3,7 3,5
2. dělník: 4,3 3,9 4,2 3,9 4,4 4,7
3. dělník: 4,2 4,5 4,0 4,1 4,5 4,4.

Na hladině významnosti 0,05 testujte hypotézu, že výkony těchto tří dělníků jsou stejné. Zamítnete-li nulovou hypotézu, určete, výkony kterých dělníků se liší na dané hladině významnosti 0,05.

Výpočet pomocí systému STATISTICA:

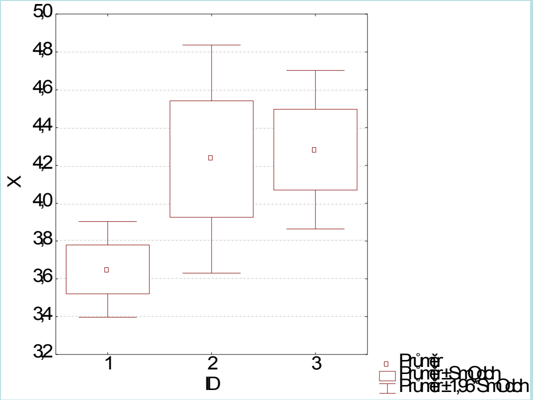
Úloha vede na analýzu rozptylu jednoduchého třídění.
Načteme datový soubor cas_delniku.sta. Proměnná X obsahuje zjištěné časy, proměnná ID nabývá hodnoty 1 pro 1. dělníka, hodnoty 2 pro 2. dělníka a hodnoty 3 pro 3. dělníka.
Statistiky – Základní statistiky/tabulky – Rozklad & jednofakt. ANOVA – Proměnné - Závislé X, Grupovací ID, OK, Kódy pro grupovací proměnné – Vše, OK, Výpočet: Tabulka statistik (zobrazí se průměry, směrodatné odchylky a rozsahy všech tří výběrů).

Rozkladová tabulka popisných stat
N=16 (V seznamu záv. prom. nejso

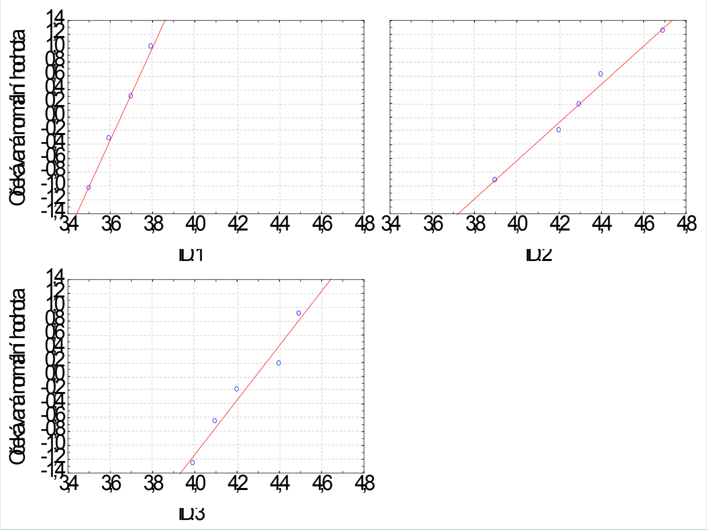
ID	X průmě	X N	X Sm.odc
1	3,650	4	0,129
2	4,233	6	0,307
3	4,283	6	0,213
Vš.sku	4,106	1	0,353

Na uskutečnění daného pracovního úkonu potřebuje nejkratší čas 1. dělník. Podává také nejvyrovnanější výkony – směrodatná odchylka proměnné X je u něj nejmenší. Naopak nejpomalejší je 3. dělník.

Nyní vytvoříme krabicové diagramy: Návrat do Statistiky podle skupin – Kategoriz. krabicový graf (současné zobrazení krabicových diagramů pro všechny tři výběry)



Pomocí N-P plot orientačně posoudíme normalitu všech tří výběrů:
Návrat do Statistiky podle skupin – ANOVA & testy – Kategoriz. norm. pravd. grafy



Ve všech třech případech se tečky jen málo odchylují od přímky, lze soudit, že data pocházejí z normálního rozložení.

Provedení testu o shodě rozptylů:
Návrat do Statistiky podle skupin – Leveneovy testy

	Leveneův test homogenity rozptylů (cas_delniku Označ. efekty jsou význ. na hlad. p < ,05000							
Promě	SC efekt	SV efek	PC efekt	SC chyba	SV chyb	PC chyba	F	p
X	0,042	2	0,021	0,183	1	0,014	1,514	0,256

Testová statistika Levenova testu nabývá hodnoty 1,5142, stupně volnosti čitatele = 2, jmenovatele = 13, odpovídající p-hodnota = 0,256, tedy na hladině významnosti 0,05 se nezamítá hypotézu o shodě rozptylů.

Provedení testu o shodě středních hodnot:
Návrat do Statistiky podle skupin – Analýza rozptylu.

	Analýza rozptylu (cas_delniku.sta) Označ. efekty jsou význ. na hlad. p < ,05000							
Promě	SC efekt	SV efek	PC efekt	SC chyba	SV chyb	PC chyba	F	p
X	1,117	2	0,558	0,751	1	0,057	9,665	0,002

Skupinový součet čtverců S_A = 1,1177, počet stupňů volnosti f_A = 2, reziduální součet čtverců S_E = 0,7517, počet stupňů volnosti f_E = 13, testová statistika $F_A = \frac{S_A/f_A}{S_E/f_E}$ nabývá hodnoty 9,6653, počet stupňů volnosti čitatele = 2, jmenovatele = 13, odpovídající p-hodnota = 0,00268, tedy na hladině významnosti 0,05 se zamítá hypotéza o shodě středních hodnot .

Provedení metody mnohonásobného porovnávání (Schefféo test – viz skripta Základní statistické metody, věta 8.2.2.1.):

Návrat do do Statistiky podle skupin – Post- hoc – Schefféův test.

ID	Scheffého test; Proměnná: X (čas dělníků.s Označ. rozdíly jsou významné r {1} M=3,6	 {2} M=4,2	 {3} M=4,2
1		0,008	0,004
2	0,008		0,937
3	0,004	0,937	

Tabulka obsahuje p-hodnoty pro testování hypotéz o shodě středních hodnot všech dvojic výběrů. Výsledek Scheffého metody ukazuje, že na hladině významnosti 0,05 se liší výkony dělníků (1,2), (1,3) a neliší se (2,3).

Význam předpokladů v analýze rozptylu

a) Nezávislost jednotlivých náhodných výběrů – velmi důležitý předpoklad, musí být splněn, jinak dostaneme nesmyslné výsledky.

b) Normalita – ANOVA není příliš citlivá na porušení normality, zvlášť pokud mají všechny výběry rozsah nad 20 (důsledek centrální limitní věty). Při výraznějším porušení normality se doporučuje Kruskalův – Wallisův test.

c) Shoda rozptylů – mírné porušení nevadí, při větším se doporučuje Kruskalův – Wallisův test. Test shody rozptylů má smysl provádět až po ověření předpokladu normality.