8. Parametrické úlohy o více nezávislých náhodných výběrech

8.1. Více nezávislých náhodných výběrů z normálních rozložení

8.1.1. Motivace: Zajímáme se o problém, zda lze určitým faktorem (tj. nominální náhodnou veličinou Y, která má varianty $y_{[1]}, \ldots, y_{[r]}$, $r \geq 3$, vysvětlit variabilitu pozorovaných hodnot náhodné veličiny X, která je intervalového či poměrového typu. Např. zkoumáme, zda počet dětí (náhodná veličina Y) ovlivňuje průměrné roční výdaje rodiny na průmyslové zboží (náhodná veličina X).

Provede se tedy náhodný výběr $X_{11}, \ldots, X_{n_1 1}$ odpovídající variantě $y_{[1]}$ atd. až náhodný výběr $X_{1r}, \ldots, X_{n_r r}$ odpovídající variantě $y_{[r]}$. Přitom předpokládáme, že $X_{1j}, \ldots, X_{n_j j}$ se řídí rozložením $N(\mu_j, \sigma^2)$, $j = 1, \ldots, r$ a že jednotlivé náhodné výběry jsou stochasticky nezávislé.

Na hladině významnosti α testujeme nulovou hypotézu, která tvrdí, že všechny střední hodnoty jsou stejné proti alternativní hypotéze, 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. Tento postup však nelze použít, neboť nezaručuje splnění podmínky, že pravděpodobnost chyby 1. druhu je 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. Scheffého nebo Tukeyova metoda.

8.1.2. Vlastnosti statistik odvozených z výběrových průměrů a výběrových rozptylů

Nechť $X_{1j}, \ldots, X_{n_j j} \sim N(\mu_j, \sigma^2)$, $j = 1, 2, \ldots, r$ jsou nezávislé náhodné výběry. Označme

$n = \sum_{j=1}^{r} n_j$ celkový rozsah všech r výběrů, $M_j = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij}$ výběrový průměr a

$S_j^2 = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}(X_{ij} - M_j)^2$ výběrový rozptyl j-tého výběru, $j = 1, \ldots, r$. Dále označme

$S_*^2 = \frac{\sum_{j=1}^{r}(n_j - 1)S_j^2}{n - r}$ vážený průměr výběrových rozptylů a $M = \frac{1}{n}\sum_{j=1}^{r}\sum_{i=1}^{n_j} X_{ij} = \frac{1}{n}\sum_{j=1}^{r} n_j M_j$ celkový průměr všech r výběrů.

Součet $S_E = \sum_{j=1}^{r}(n_j - 1)S_j^2 = \sum_{j=1}^{r}\sum_{i=1}^{n_j}(X_{ij} - M_j)^2$ se nazývá reziduální součet čtverců a charakterizuje variabilitu uvnitř jednotlivých výběrů.

Součet $S_A = \sum_{j=1}^{r} n_j (M_j - M)^2$ se nazývá skupinový součet čtverců a charakterizuje variabilitu mezi jednotlivými výběry.

Součet $S_T = \sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(X_{ij} - M\right)^2$ se nazývá celkový součet čtverců a charakterizuje variabilitu jednotlivých pozorování kolem celkového průměru.

Platí: $S_T = S_E + S_A$.

Důkaz:

$$S_T = \sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(X_{ij} - M\right)^2 = \sum_{j=1}^{r}\sum_{i=1}^{n_j}\left[\left(X_{ij} - M_j\right) + \left(M_j - M\right)\right]^2 = \sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(X_{ij} - M_j\right)^2 +$$

$$+ 2\sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(X_{ij} - M_j\right)\left(M_j - M\right) + \sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(M_j - M\right)^2 =$$

$$= S_E + 2\sum_{j=1}^{r}\left(M_j - M\right)\sum_{i=1}^{n_j}\left(X_{ij} - M_j\right) + \sum_{j=1}^{r} n_j\left(M_j - M\right)^2 =$$

$$= S_E + 2\sum_{j=1}^{r}\left(M_j - M\right)\left(n_j M_j - n_j M_j\right) + S_A = S_E + S_A$$

8.1.3. Test hypotézy o shodě středních hodnot

Na hladině významnosti α testujeme H_0: $\mu_1 = \ldots = \mu_r$ proti H_1: aspoň jedna dvojice středních hodnot se liší. Platí-li H_0, pak

a) nestranným odhadem společné střední hodnoty μ je celkový průměr $M = \frac{1}{n}\sum_{j=1}^{r} n_j M_j$

b) nestranným odhadem rozptylu σ^2 je statistika $\frac{S_A}{r-1}$

c) statistika $\frac{S_A}{\sigma^2} \sim \chi^2(r\text{-}1)$

d) statistika $F = \frac{S_A/(r-1)}{S_E/(n-r)} \sim F(r\text{-}1,\ n\text{-}r)$.

H_0 tedy zamítáme na hladině významnosti α, když $F \geq F_{1-\alpha}(r\text{-}1,\ n\text{-}r)$.

Důkaz:

Nechť platí H_0.

ad a) Označme μ společnou střední hodnotu. Počítáme

$$E(M) = E\left(\frac{1}{n}\sum_{j=1}^{r} n_j M_j\right) = \frac{1}{n}\sum_{j=1}^{r} n_j E(M_j) = \frac{1}{n}\sum_{j=1}^{r} n_j \mu = \frac{1}{n} n\mu = \mu \text{. Tedy M je nestranný odhad } \mu.$$

ad b)

$$E\left(\frac{S_A}{r-1}\right)=\frac{1}{r-1}E\left(\sum_{j=1}^{r}n_j(M_j-M)^2\right)=\frac{1}{r-1}\sum_{j=1}^{r}n_jE\left([(M_j-\mu)-(M-\mu)]^2\right)=$$

$$=\frac{1}{r-1}\sum_{j=1}^{r}n_jE\left((M_j-\mu)^2-2(M_j-\mu)(M-\mu)+(M-\mu)^2\right)=$$

$$=\frac{1}{r-1}\left[\sum_{j=1}^{r}n_jD(M_j)-2E\left((M-\mu)\left(\sum_{j=1}^{r}n_jM_j-\sum_{j=1}^{r}n_j\mu\right)\right)+\sum_{j=1}^{r}n_jE\left((M-\mu)^2\right)\right]=$$

$$=\frac{1}{r-1}\left[\sum_{j=1}^{r}n_j\frac{\sigma^2}{n_j}-2nE\left((M-\mu)^2\right)+nD(M)\right]=\frac{1}{r-1}\left(r\sigma^2-nD(M)\right)=\frac{1}{r-1}\left(r\sigma^2-n\frac{\sigma^2}{n}\right)=\sigma^2$$

Statistika $\frac{S_A}{r-1}$ je tedy nestranným odhadem σ^2.

ad c)

$$\frac{S_A}{\sigma^2}=\sum_{j=1}^{r}n_j\frac{(M_j-M)^2}{\sigma^2}=\sum_{j=1}^{r}n_j\left(\frac{M_j-\mu}{\sigma}-\frac{M-\mu}{\sigma}\right)^2=$$

$$=\sum_{j=1}^{r}n_j\left(\frac{M_j-\mu}{\sigma}\right)^2-2\sum_{j=1}^{r}n_j\frac{M_j-\mu}{\sigma}\cdot\frac{M-\mu}{\sigma}+\sum_{j=1}^{r}n_j\left(\frac{M-\mu}{\sigma}\right)^2=$$

$$=\sum_{j=1}^{r}\left(\frac{M_j-\mu}{\frac{\sigma}{\sqrt{n_j}}}\right)^2-2\frac{M-\mu}{\sigma}\left(\sum_{j=1}^{r}n_jM_j-\sum_{j=1}^{r}n_j\mu\right)+n\left(\frac{M-\mu}{\sigma}\right)^2=$$

$$=\sum_{j=1}^{r}\left(\frac{M_j-\mu}{\frac{\sigma}{\sqrt{n_j}}}\right)^2-2n\left(\frac{M-\mu}{\sigma}\right)^2+n\left(\frac{M-\mu}{\sigma}\right)^2=\sum_{j=1}^{r}\left(\frac{M_j-\mu}{\frac{\sigma}{\sqrt{n_j}}}\right)^2-\left(\frac{M-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2$$

Náhodná veličina, která vznikne standardizací M_j či M, má standardizované normální rozložení (M_j resp. M je totiž lineární kombinací náhodných veličin s normálním rozložením). Součet kvadrátů stochasticky nezávislých náhodných veličin s rozložením N(0,1) má rozložení χ^2 s počtem stupňů volnosti rovným počtu sčítanců. V našem případě je sčítanců r-1, dostáváme tedy veličinu s rozložením χ^2(r-1).

ad d) Zobecněním 7.1.1. c) dostaneme, že statistika $\frac{(n-r)S_*^2}{\sigma^2}=\frac{\sum_{j=1}^{r}(n_j-1)S_j^2}{\sigma^2}=\frac{S_E}{\sigma^2}$ ~ χ^2(n-r). Platí-li H_0, má statistika $\frac{S_A}{\sigma^2}$ ~ χ^2(r-1) (viz bod b) a přitom S_E a S_A jsou stochasticky nezávislé. Tedy $F=\frac{\frac{S_A/(r-1)}{\sigma^2}}{\frac{S_E/(n-r)}{\sigma^2}}=\frac{S_A/(r-1)}{S_E/(n-r)}$ ~ F(r-1, n-r), pokud H_0 platí. Když H_0 platí,

bude podíl $F = \dfrac{S_A/(r-1)}{S_E/(n-r)}$ blízký 1, protože $S_A/(r-1)$ je nestranný odhad σ^2 a $S_E/(n-r)$ je nestranný odhad σ^2 (bez ohledu na platnost H_0). Není-li H_0 pravdivá, čitatel vzroste, protože rozdílnost středních hodnot se projeví větší variabilitou výběrových průměrů kolem celkového průměru. Jmenovatel ovšem zůstane týž, proto testové kritérium F bude nabývat hodnot větších než 1. Jelikož hladina významnosti je α, bude mít kritický obor tvar:
$W = \langle F_{1-\alpha}(r-1, n-r), \infty)$.

8.1.4. Tabulka jednofaktorové analýzy rozptylu

Bývá zvykem zapisovat výsledky do tabulky ANOVA.

Zdroj variability	Součet čtverců	Stupně volnosti	podíl	F
skupiny	S_A	r-1	$S_A/(r-1)$	$\dfrac{S_A/(r-1)}{S_E/(n-r)}$
reziduální	S_E	n-r	$S_E/(n-r)$	-
celkový	S_T	n-1	-	-

8.1.5. Bartlettův a Levenův test shody rozptylů

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

Levenův test: Položme $Z_{ij} = \left|X_{ij} - M_j\right|$. Označíme

$$M_{Zj} = \frac{1}{n_j}\sum_{i=1}^{n_j} Z_{ij},\ M_Z = \frac{1}{n}\sum_{j=1}^{r}\sum_{i=1}^{n_j} Z_{ij},\ S_{ZE} = \sum_{j=1}^{r}\sum_{i=1}^{n_j}\left(Z_{ij} - M_{Zj}\right)^2,\ S_{ZA} = \sum_{j=1}^{r} n_j\left(M_{Zj} - M_Z\right)^2.$$

Platí-li hypotéza o shodě rozptylů, pak statistika $F_Z = \dfrac{S_{ZA}/(r-1)}{S_{ZE}/(n-r)} \sim F(r-1, n-r)$. H_0 tedy zamítáme na hladině významnosti α, když $F_Z \geq F_{1-\alpha}(r-1, n-r)$.

Bartlettův test: Platí-li hypotéza o shodě rozptylů, pak statistika

$$B = \frac{1}{C}\left[(n-r)\ln S_*^2 - \sum_{j=1}^{r}(n_j - 1)\ln S_j^2\right] \sim \chi^2(r-1),\ \text{kde}\ C = 1 + \frac{1}{3(r-1)}\left(\sum_{j=1}^{r}\frac{1}{n_j - 1} - \frac{1}{n-r}\right).$$

H_0 zamítáme na hladině významnosti α, když $B \geq \chi^2_{1-\alpha}(r-1, n-r)$.

8.1.6. 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. Mají-li všechny výběry týž rozsah p (říkáme, že třídění je vyvážené), použijeme Tukeyovu metodu: rovnost středních hodnot μ_k a μ_l zamítneme na hladině významnosti α, když

$\left|M_k - M_l\right| \geq q_{1-\alpha}(r, n-r)\dfrac{S_*}{\sqrt{p}}$, kde hodnoty $q_{1-\alpha}(r, n-r)$ najdeme ve statistických tabulkách.

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

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

Pozor, může nastat situace, kdy při zamítnutí H_0 nenajdeme významný rozdíl u žádné dvojice středních hodnot. Pak je významně rozdílná některá složitější kombinace středních hodnot.

8.1.6. Příklad

U čtyř odrůd brambor (označených symboly A, B, C, D) se zjišťovala celková hmotnost brambor vyrostlých vždy z jednoho trsu. Výsledky (v kg):

odrůda	hmotnost
A	0,9 0,8 0,6 0,9
B	1,3 1,0 1,3
C	1,3 1,5 1,6 1,1 1,5
D	1,1 1,2 1,0

Na hladině významnosti 0,05 testujte hypotézu, že střední hodnota hmotnosti trsu brambor nezávisí na odrůdě. Zamítnete-li nulovou hypotézu, zjistěte, které dvojice odrůd se liší na hladině významnosti 0,05.

Řešení

Data považujeme za realizace čtyř nezávislých náhodných výběrů ze čtyř normálních rozložení se stejným rozptylem. Testujeme hypotézu, že všechny čtyři střední hodnoty jsou stejné.

$M_1 = 0{,}8$, $M_2 = 1{,}2$, $M_3 = 1{,}1$, $M = 1{,}14$, $S_E = 0{,}3$, $S_A = 0{,}816$, $S_T = 1{,}116$, $F = 9{,}97$, $F_{0,95}(3,11) = 3{,}59$. Protože testová statistika se realizuje v kritickém oboru, H_0 zamítáme na hladině významnosti 0,05.

Výsledky zapíšeme do tabulky ANOVA

Zdroj variability	Součet čtverců	Stupně volnosti	podíl	F
skupiny	$S_A = 0{,}816$	3	$S_A/3 = 0{,}272$	$\dfrac{S_A/(r-1)}{S_E/(n-r)} =$ 9,97
reziduální	$S_E = 0{,}3$	11	$S_E/11 = 0{,}02727$	-
celkový	$S_T = 1{,}116$	14	-	-

Nyní pomocí Scheffého metody zjistíme, které dvojice odrůd se liší na hladině významnosti 0,05.

Srovnávané odrůdy	Rozdíly $\lvert M_k - M_l \rvert$	Pravá strana vzorce
A, B	0,4	0,41
A, C	0,67	0,36
A, D	0,3	0,41
B, C	0,2	0,40
B, D	0,1	0,44
C, D	0,3	0,40

Na hladině významnosti 0,05 se liší odrůdy A a C.

8.1.7. 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.

8.2. Více nezávislých náhodných výběrů z alternativních rozložení

8.2.1. Test homogenity binomických rozložení

Nechť $X_{1j},\dots,X_{n_j j} \sim A(\vartheta_j)$, j = 1, 2, ..., r jsou nezávislé náhodné výběry. Testujeme hypotézu H_0: $\vartheta_1 = \dots = \vartheta_r$ proti alternativní hypotéze H_1: aspoň jedna dvojice parametrů je různá. Označme $n = \sum_{j=1}^{r} n_j$ celkový rozsah všech r výběrů a M_* vážený průměr výběrových průměrů. Jako testové kritérium slouží statistika $Q = \frac{1}{M_*(1-M_*)}\sum_{j=1}^{r} n_j (M_j - M_*)^2$, která v případě platnosti nulové hypotézy má asymptoticky rozložení $\chi^2(r-1)$. H_0 tedy zamítáme na asymptotické hladině významnosti α, když $Q \geq \chi^2_{1-\alpha}(r-1)$.

Test lze použít, pokud $n_j M_* > 5$ pro všechna j = 1, ..., r.

Statistiku Q lze snadno upravit do Brandtova – Snedecorova výpočetního tvaru

$$Q = \frac{1}{M_*(1-M_*)}\sum_{j=1}^{r} n_j M_j^2 - n\frac{M_*}{1-M_*}.$$

Důkaz: Jak bylo ukázáno v 6.3.1., statistika $U_j = \frac{M_j - \vartheta_j}{\sqrt{\frac{\vartheta_j(1-\vartheta_j)}{n_j}}} \approx N(0,1)$. Nechť platí H_0.

Označme ϑ společnou hodnotu všech parametrů ϑ_j, j = 1, 2, ..., r. Pak statistika

$U_j = \frac{M_j - \vartheta}{\sqrt{\frac{\vartheta(1-\vartheta)}{n_j}}} \approx N(0,1)$ a $U_j^2 = \frac{n_j (M_j - \vartheta)^2}{\vartheta(1-\vartheta)} \approx \chi^2(1)$. Lze ukázat, že statistika

$Q = \sum_{j=1}^{r} U_j^2 = \frac{1}{\vartheta(1-\vartheta)}\sum_{j=1}^{r} n_j (M_j - \vartheta)^2 \approx \chi^2(r-1)$. Parametr ϑ však neznáme, nahradíme ho váženým průměrem výběrových průměrů $M_* = \frac{\sum_{j=1}^{r} n_j M_j}{n}$ a dostaneme

$Q = \frac{1}{M_*(1-M_*)}\sum_{j=1}^{r} n_j (M_j - M_*)^2 \approx \chi^2(r-1)$. Kritický obor tedy bude $W = \langle \chi^2_{1-\alpha}(r-1), \infty)$.

8.2.2. Test homogenity binomických rozložení založený na arkussinusové transformaci

Není-li splněna podmínka $n_j M_* > 5$ pro všechna j = 1, ..., r, doporučuje se následující postup: označme $Y_j = \sum_{i=1}^{n_j} X_{ij}$, $A_j = \arcsin \sqrt{Y_j}$, j = 1, ..., r, $B = \frac{1}{n}\sum_{j=1}^{r} n_j A_j$. Pak statistika

$Q = n\sum_{j=1}^{r} n_j (A_j - B)^2 \approx \chi^2(r-1)$. H_0 tedy zamítáme na asymptotické hladině významnosti α, když $Q \geq \chi^2_{1-\alpha}(r-1)$.

8.2.3. Mnohonásobné porovnávání

Zamítneme-li nulovou hypotézu na asymptotické hladině významnosti α, chceme zjistit, které dvojice parametrů ϑ_k, ϑ_l se liší. Platí-li nerovnost

$|A_k - A_l| \geq \sqrt{\frac{1}{8}\left(\frac{1}{n_k} + \frac{1}{n_l}\right)} \cdot q_{1-\alpha}(r, \infty)$, pak na hladině významnosti α zamítáme hypotézu o shodě parametrů ϑ_k, ϑ_l. (Hodnoty $q_{1-\alpha}(r, \infty)$ najdeme v tabulkách.)

8.2.4. Příklad

Na gymnázium bylo přijato 142 studentů. Ti byli náhodně rozděleni do čtyř tříd A, B, C, D. V každé třídě byla matematika vyučována jinou metodou. Na konci školního roku psali všichni studenti stejnou písemnou práci a byl zaznamenán počet těch studentů, kteří vyřešili všechny zadané úkoly.

Třída	A	B	C	D
Počet studentů	35	36	37	34
Počet úspěšných studentů	5	8	17	15

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že rozdíly mezi třídami jsou způsobeny pouze náhodnými vlivy.

Řešení

Máme čtyři nezávislé náhodné výběry, j-tý pochází z rozložení $A(\vartheta_j)$, j = 1, 2, 3, 4. Testujeme hypotézu H_0: $\vartheta_1 = \vartheta_2 = \vartheta_3 = \vartheta_4$. Přitom $n_1 = 35$, $n_2 = 36$, $n_3 = 37$, $n_4 = 34$, $m_1 = 5/35$, $m_2 = 8/36$, $m_3 = 17/37$, $m_4 = 15/34$, $m_* = (5+8+17+15)/142 = 45/142$. Testová statistika Q = 12,288, $\chi^2_{0,95}(3) = 7,81$. Protože testové kritérium se realizuje v kritickém oboru, H_0 zamítáme na asymptotické hladině významnosti 0,05.

Protože v jednom případě není splněna podmínka $n_j m_* > 5$, použijeme ještě test založený na arkussinusové transformaci.

$A_1 = 0,388$, $A_2 = 0,491$, $A_3 = 0,745$, $A_4 = 0,768$, $B = 0,588$. Pak Q = 13,227 a H_0 opět zamítáme na asymptotické hladině významnosti 0,05.

Nyní metodou mnohonásobného porovnávání zjistíme, které dvojice parametrů se od sebe liší na hladině významnosti 0,05.

Srovnávané třídy	Rozdíly $\lvert A_k - A_l \rvert$	Pravá strana vzorce
A, B	0,103	0,30
A, C	0,357	0,30
A, D	0,339	0,31
B, C	0,254	0,30
B, D	0,254	0,31
C, D	0,018	0,30

Na hladině významnosti 0,05 se liší třídy A, C a A, D.

Příklady k 8. kapitole

Příklad 1.: Jsou známy měsíční tržby (v tisících Kč) tří prodavačů za dobu půl roku.
1. prodavač: 12 10 9 10 11 9
2. prodavač: 10 12 11 12 14 13
3. prodavač: 19 18 16 16 17 15
Na hladině významnosti 0,05 testujte hypotézu, že střední hodnoty tržeb všech tří prodavačů jsou stejné. Pokud zamítneme nulovou hypotézu, zjistěte, tržby kterých dvou prodavačů se liší na hladině významnosti 0,05.

Řešení
$m_1 = 10{,}17$, $m_2 = 12$, $m_3 = 16{,}83$, $m = 13$, $s_1^2 = 1{,}37$, $s_2^2 = 2$, $s_3^2 = 2{,}17$, $S_E = 27{,}7$, $S_A = 142{,}3$, $S_T = 170$, $F = 38{,}58$, $F_{0,975}(2{,}015) = 3{,}6823$, H_0 tedy zamítáme na hladině významnosti 0,05.
Výsledky zapíšeme do tabulky ANOVA

Zdroj variability	Součet čtverců	Stupně volnosti	podíl	F
skupiny	$S_A = 142{,}3$	2	$S_A/2 = 71{,}17$	$\dfrac{S_A/(r-1)}{S_E/(n-r)} = 38{,}58$
reziduální	$S_E = 27{,}7$	15	$S_E/15 = 1{,}84$	-
celkový	$S_T = 170$	17	-	-

Nyní pomocí Tukeyovy metody zjistíme, které dvojice prodavačů se liší na hladině významnosti 0,05.

Srovnávaní prodavači	Rozdíly $\lvert M_k - M_l \rvert$	Pravá strana vzorce
1, 2	1,83	2,68
1, 3	6,67*	2,68
2, 3	4,83*	2,68

Pravá strana: $q_{1-\alpha}(r, n-r)\dfrac{s_*}{\sqrt{p}} = q_{0,95}(3{,}15)\dfrac{\sqrt{1{,}84}}{\sqrt{6}} = 4{,}83\dfrac{\sqrt{1{,}84}}{\sqrt{6}} = 2{,}68$, kde $s_*^2 = \dfrac{S_E}{n-r} = 1{,}84$
Na hladině významnosti 0,05 se liší tržby prodavačů 1, 3 a 2, 3.

Příklad 2.: 99 náhodně vybraných matek bylo dotázáno, zda jejich kojenec dostává dudlík. Zjišťoval se též nejvyšší stupeň dosaženého vzdělání matky.

Vzdělání matky	Počet matek	Počet dětí s dudlíkem
Základní	39	27
Středoškolské	47	34
Vysokoškolské	18	15

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že podíly dětí s dudlíkem nezávisí na vzdělání matky.

Řešení
$n_1 = 39$, $n_2 = 47$, $n_3 = 18$, $m_1 = 27/39$, $m_2 = 34/47$, $m_3 = 15/18$, $m_* = (27+34+15)/99 = 76/99$.
Testová statistika $Q = 1{,}085$, kvantil $\chi^2_{0,95}(2) = 5{,}992$. Protože testové kritérium se nerealizuje v kritickém oboru, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

Práce se systémem STATISTICA

Téma: Parametrické úlohy o více nezávislých náhodných výběrech (analýza rozptylu jednoduchého třídění)

Vzorový 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.
Návod: Vytvořte datový soubor se dvěma proměnnými (X a ID) a 16 případy. Do 1. sloupce napište změřené časy, do 2. sloupce dejte čtyřikrát jedničku, šestkrát dvojku a šestkrát trojku. Statistics - Basic Statistics and Tables - Breakdown & one-way ANOVA - Variables Dependent X, Grouping ID, OK, Codes for grouping variables – All, OK, Quick – Summary: Table of statistics (zobrazí se průměry, směrodatné odchylky a rozsahy všech tří výběrů) – návrat do Statistics by Groups – Categorized box & whisker plot (současné zobrazení krabicových diagramů pro všechny tři výběry – změna typu krabicového diagramu se provede po dvojnásobném kliknutí myší na graf v menu Plot:Box/Whisker) – návrat do Statistics by Groups – ANOVA & tests – Categorized normal prob. plots (vizuální posouzení normality všech výběrů) – návrat do Statistics by Groups – Levene tests (testování homogenity rozptylů všech tří výběrů - p-hodnota = 0,256, tedy na hladině významnosti 0,05 se nezamítá hypotézu o shodě rozptylů) – návrat do Statistics by Groups – Analysis of Variance (provední analýzy rozptylu). Ve výstupní tabulce je použito následující označení: SS Effect … skupinový součet čtverců S_A, MS Effect … $S_A/(r-1)$, SS Error … reziduální součet čtverců S_E, MS Error … $S_E/(n-r)$. Protože p-hodnota = 0,00268, zamítá se na hladině významnosti 0,05 hypotézu o shodě středních hodnot. Návrat do Statistics by Groups – Post- hoc – Scheffé test. Výsledek Scheffého metody ukazuje, že na hladině významnosti 0,05 se liší výkony dělníků (1,2), (1,3) a neliší se (2,3).

Příklady k samostatnému řešení:
1. Studenti byli vyučováni předmětu za využití pěti pedagogických metod: tradiční způsob, programová výuka, audiotechnika, audiovizuální technika a vizuální technika. Z každé skupiny byl vybrán náhodný vzorek studentů a všichni byli podrobeni témuž písemnému testu. Na hladině významnosti 0,05 testujte hypotézu, že znalosti všech studentů jsou stejné a nezávisí na použité pedagogické metodě. V případě zamítnutí hypotézy zjistěte, které výběry se liší na hladině významnosti 0,05.

metoda	počet bodů
tradiční	76,2 48,3 85,1 63,7 91,6 87,2
programová	85,2 74,3 76,5 80,3 67,4 67,9 72,1 60,4
audio	67,3 60,1 55,4 72,3 40
audiovizuální	75,8 81,6 90,3 78 67,8 57,6
vizuální	50,5 70,2 88,8 67,1 77,7 73,9

2. Pan Novák může cestovat z místa bydliště do místa pracoviště třemi různými způsoby: tramvají (způsob A), autobusem (způsob B) a metrem s následným přestupem na tramvaj (způsob C). Máme k dispozici jeho naměřené časy cestování do práce v době ranní špičky (včetně čekání na příslušný spoj) v minutách.
Způsob A: 32, 39, 42, 37, 34, 38, B: 30, 34, 28, 26, 32, C: 40, 37, 31, 39, 38, 33, 34

Pro všechny tři způsoby dopravy vypočtěte průměrné časy cestování. Na hladině významnosti 0,05 testujte hypotézu, že doba cestování do práce nezávisí na způsobu dopravy. V případě zamítnutí nulové hypotézy zjistěte, které způsoby dopravy do práce se od sebe liší na hladině významnosti 0,05.