Kapitola 5.: Analýza rozptylu jednoduchého třídění

Cíl kapitoly

Po prostudování této kapitoly budete umět
- hodnotit vliv faktoru o $r \geq 3$ úrovních na variabilitu hodnot sledované náhodné veličiny
- sestrojit tabulku analýzy rozptylu
- identifikovat dvojice náhodných výběrů, které se významně liší střední hodnotou
- provést test shody rozptylů
- znázornit rozložení dat v daných $r \geq 3$ náhodných výběrech graficky pomocí kategorizovaných krabicových diagramů

Časová zátěž

Na prostudování této kapitoly a splnění úkolů s ní spojených budete potřebovat asi 5 hodin studia.

5.1. 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 i-té úrovni odpovídá n_i výsledků $X_{i1},\ldots,X_{in_i}$, které tvoří náhodný výběr z rozložení $N(\mu_i, \sigma^2)$, i = 1, ..., 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	$X_{r1},\ldots,X_{rn_r}$

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 nejvýše 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.

5.2. Označení

V analýze rozptylu jednoduchého třídění se používá následující označení.

$n = \sum_{i=1}^{r} n_i$ … 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

5.3. 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:

$\sum_{i=1}^{r} n_i \alpha_i = 0$.

Zavedeme součty čtverců

$S_T = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{ij} - M_{..})^2$ … celkový součet čtverců (charakterizuje variabilitu jednotlivých pozorování kolem celkového průměru), má počet stupňů volnosti $f_T = n - 1$,

$S_A = \sum_{i=1}^{r} n_i (M_{i.} - M_{..})^2$ … skupinový součet čtverců (charakterizuje variabilitu mezi jednotlivými náhodnými výběry), má počet stupňů volnosti $f_A = r - 1$,

$S_E = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{ij} - M_{i.})^2$ … reziduální součet čtverců (charakterizuje variabilitu uvnitř jednotlivých výběrů), má počet stupňů volnosti $f_E = n - r$.

Lze dokázat, že $S_T = S_A + S_E$.
Sčítanec $(M_{i.} - M_{..})$ představuje bodový odhad efektu α_i.

Kdyby nezáleželo na faktoru A, platila by hypotéza $\alpha_1 = \ldots = \alpha_r = 0$ a dostali bychom model
M1: $X_{ij} = \mu + \varepsilon_{ij}$.
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	$f_T = n-1$	-	-

5.4. Testy shody rozptylů

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

5.4.1. Levenův test

Položme $Z_{ij} = \left|X_{ij} - M_{i.}\right|$. Označíme

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

Platí-li hypotéza o shodě rozptylů, pak statistika $F_{ZA} = \frac{S_{ZA}/(r-1)}{S_{ZE}/(n-r)}$ má asymptoticky rozložení F(r-1,n-r). H_0 tedy zamítáme na asymptotické hladině významnosti α, když $F_{ZA} \geq F_{1-\alpha}(r-1,n-r)$. Často užívanou modifikací Levenova testu je Brownův – Forsytheův test. Modifikace spočívá v tom, že místo výběrového průměru i-tého výběru se při výpočtu veličiny Z_{ij} používá medián i-tého výběru. Simulační studie ukazují, že Brownův – Forsytheův test lze použít i při porušení předpokladu normality.

5.4.2. Bartlettův test

Platí-li hypotéza o shodě rozptylů, pak statistika

$$B = \frac{1}{C}\left[(n-r)\ln S_*^2 - \sum_{i=1}^{r}(n_i - 1)\ln S_i^2\right]$$ má asymptoticky rozložení χ^2(r-1), kde

$$C = 1 + \frac{1}{3(r-1)}\left(\sum_{i=1}^{r}\frac{1}{n_i - 1} - \frac{1}{n-r}\right), S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(X_{ij} - M_{i.}\right)^2$$ je výběrový rozptyl i-tého výběru, $S_*^2 = \frac{\sum_{i=1}^{r}(n_i - 1)S_i^2}{n-r} = \frac{S_E}{n-r}$ je vážený průměr výběrových rozptylů.

H_0 zamítáme na asymptotické hladině významnosti α, když $B \geq \chi^2_{1-\alpha}(r-1,n-r)$. Bartlettův test lze použít, pokud rozsahy všech výběrů jsou aspoň 7. Je velmi citlivý na porušení předpokladu normality.

5.5. 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.

5.5.1. Tukeyova metoda

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

$\left|M_{k.} - M_{l.}\right| \geq q_{1-\alpha}(r, n-r)\frac{S_*}{\sqrt{p}}$, kde hodnoty $q_{1-\alpha}$(r,n-r) jsou kvantily studentizovaného rozpětí a najdeme je ve statistických tabulkách.

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

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

Metody mnohonásobného porovnávání mají obecně menší sílu než ANOVA. Může proto 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.

5.6. 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, $M_{3.}$ = 1,4, $M_{4.}$ = 1,1, $M_{..}$ = 1,14, S_E = 0,3, S_A = 0,816, S_T = 1,116, F_A = 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

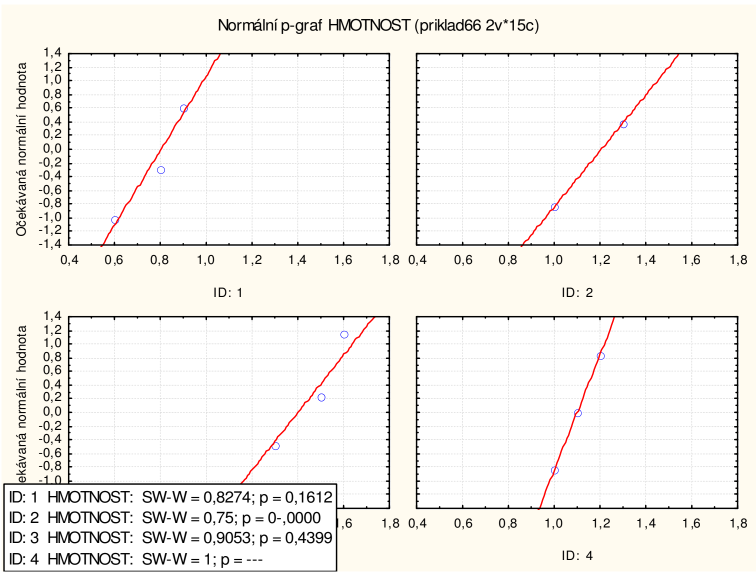
Zdroj variability	Součet čtverců	Stupně volnosti	podíl	F_A
skupiny	S_A = 0,816	f_A = 3	S_A/3 = 0,272	$\frac{S_A/f_A}{S_E/f_E} = 9{,}97$
reziduální	S_E = 0,3	f_E = 11	S_E/11 = 0,02727	-
celkový	S_T = 1,116	f_T = 14	-	-

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

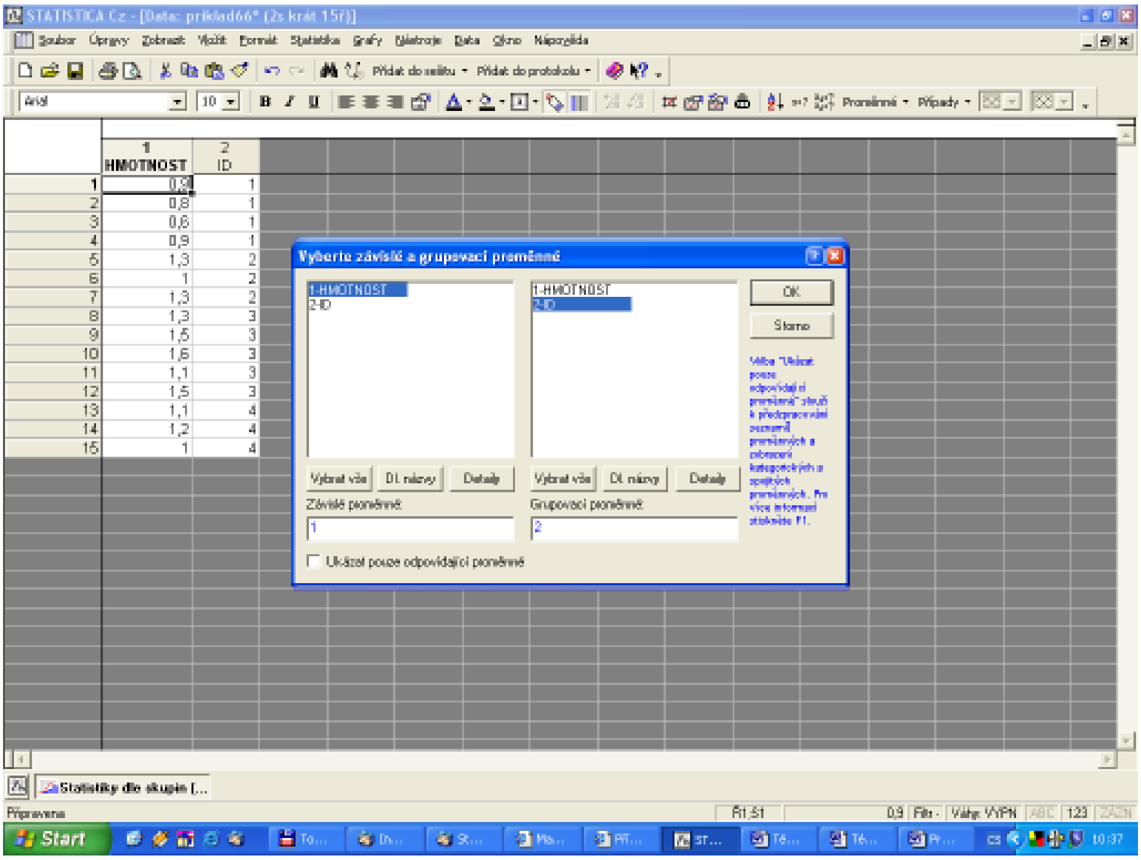
Srovnávané odrůdy	Rozdíly $\lvert M_{k.} - M_{l.} \rvert$	Pravá strana vzorce
A, B	0,4	0,41
A, C	0,67	0,36
A, D	0,3	0,41
B, C	0,2	0,40
B, D	0,1	0,44
C, D	0,3	0,40

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

Řešení pomocí systému STATISTICA:
Otevřeme nový datový soubor o dvou proměnných a 15 případech. První proměnnou nazveme HMOTNOST, druhou ID. Do proměnné HMOTNOST zapíšeme zjištěné hmotnosti, do proměnné ID, která slouží k rozlišení odrůd, zapíšeme 4 krát jedničku, 3 krát dvojku, 5 krát trojku a 3 krát čtyřku.
Pomocí NP-grafu a S-W testu ověříme normalitu dat v daných čtyřech skupinách. Grafy – 2D Grafy – Normální pravděpodobnostní grafy – zaškrtneme S-W test, Proměnné HMOTNOST, OK, Kategorizovaný – Kategorie X, zaškrtneme Zapnuto, Změnit proměnnou – ID, OK. Dostaneme graf



Vidíme, že ve všech čtyřech případech jsou odchylky teček od přímky jenom malé a data tedy lze považovat za realizace náhodných výběrů z normálních rozložení.
Nyní budeme na hladině významnosti 0,05 testovat hypotézu o shodě rozptylů: Statistika – Základní statistiky a tabulky – Rozklad & jednofakt. ANOVA – OK, Proměnné – Závislé proměnné HMOTNOST, Grupovací proměnná ID – OK.



Na záložce ANOVA & testy vybereme Leveneovy testy. Ve výstupu dostaneme tabulku

	Leveneův test homogenity rozptylů (priklad66) Označ. efekty jsou význ. na hlad. p < ,05000							
Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
HMOTNOST	0,018667	3	0,006222	0,065333	11	0,005939	1,047619	0,410027

Testová statistika Levenova testu se realizuje hodnotou 1,047619, počet stupňů volnosti čitatele je 3, jmenovatele 11, odpovídající p-hodnota je 0,410027, tedy na hladině významnosti 0,05 nezamítáme hypotézu o shodě rozptylů.
Dále budeme na hladině významnosti 0,05 testovat hypotézu o shodě středních hodnot. Na záložce ANOVA & testy vybereme Analýza rozptylu. Ve výstupu dostaneme tabulku

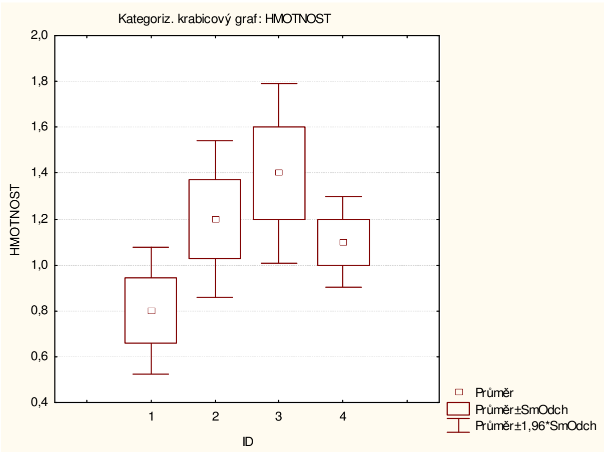
	Analýza rozptylu (priklad66) Označ. efekty jsou význ. na hlad. p < ,05000							
Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
HMOTNOST	0,816000	3	0,272000	0,300000	11	0,027273	9,973333	0,001805

Testová statistika F_A se realizuje hodnotou 9,97333, počet stupňů volnosti čitatele je 3, jmenovatele 11, odpovídající p-hodnota je 0,001805, tedy na hladině významnosti 0,05 zamítáme hypotézu o shodě středních hodnot.

Vytvoříme ještě tabulku s hodnotami výběrových průměrů a výběrových směrodatných odchylek tak, že na záložce Popisné statistiky zvolíme Výpočet: Tabulka statistik.

Rozkladová tabulka popisných statistik (priklad66) N=15 (V seznamu záv. prom. nejsou ChD)			
ID	HMOTNOST průměr	HMOTNOST N	HMOTNOST Sm.odch.
1	0,800000	4	0,141421
2	1,200000	3	0,173205
3	1,400000	5	0,200000
4	1,100000	3	0,100000
Vš.skup.	1,140000	15	0,282337

Rovněž sestrojíme krabicové diagramy tak, že na záložce Popisné statistiky zvolíme Kategoriz. krabicový graf. Vybereme typ Průměr/SmOdch/1.96SmOdch.



Vidíme, že nejnižší průměrnou hmotnost má odrůda A, nejnižší variabilitu hmotnosti vykazuje odrůda D.
Abychom zjistili, které dvojice odrůd se liší na hladině významnosti 0,05, na záložce Post-hoc vybereme Schefféův test.

		Scheffeho test; proměn.:HMOTNOST (priklad66) Označ. rozdíly jsou významné na hlad. p < ,05000			
ID		{1} M=,80000	{2} M=1,2000	{3} M=1,4000	{4} M=1,1000
1	{1}		0,059165	0,001950	0,190463
2	{2}	0,059165		0,464537	0,905502
3	{3}	0,001950	0,464537		0,163499
4	{4}	0,190463	0,905502	0,163499	

V tabulce jsou uvedeny p-hodnoty pro testování hypotéz o shodě dvojic středních hodnot. Pouze jediná z těchto p-hodnot je menší nebo rovna 0,05, tedy na hladině významnosti 0,05 se liší odrůdy A a C.

5.7. 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.

Shrnutí

Analýza rozptylu jednoduchého třídění slouží k hodnocení vlivu faktoru o $r \geq 3$ úrovních na variabilitu hodnot sledované náhodné veličiny s normálním rozložením. Test hypotézy o shodě středních hodnot odvodil *R. A. Fisher*. Výpočty spojené s testováním této hypotézy se zapisují do *tabulky ANOVA*. Dojde-li na dané hladině významnosti α k zamítnutí nulové hypotézy, použijeme některou z *metod mnohonásobného porovnávání* (např. *Scheffého* nebo *Tukeyovu metodu*), abychom identifikovali dvojice náhodných výběrů, které přispěly k zamítnutí nulové hypotézy. ANOVA předpokládá shodu rozptylů. Hypotézu o shodě rozptylů můžeme testovat pomocí *Bartlettova testu*, *Levenova testu* nebo *Brownova – Forsytheova testu*. Vlastnosti rozložení dat v daných $r \geq 3$ náhodných výběrech graficky znázorňujeme pomocí *kategorizovaných krabicových diagramů* typu průměr – směrodatná odchylka – 1,96 směrodatná odchylka.

Kontrolní otázky

1. Jaký problém řeší analýza rozptylu jednoduchého třídění?
2. Jak je definován celkový, skupinový a reziduální součet čtverců a co tyto součty čtverců charakterizují?
3. Popište způsob testování hypotézy o shodě středních hodnot.
4. Jak se testuje hypotéza o shodě rozptylů?
5. Které metody mnohonásobného porovnání se používají v analýze rozptylu jednoduchého třídění?
6. Pojednejte o významu předpokladů v analýze rozptylu jednoduchého třídění.

Autokorekční test

1. Z následujících tří možností vyberte tu správnou:
Analýza rozptylu jednoduchého třídění slouží k vyhodnocení dat, které vznikly:
a) při párovém uspořádání pokusu
b) při blokovém uspořádání pokusu
c) při mnohovýběrovém uspořádání pokusu.

2. Která z následujících tvrzení jsou pravdivá?
Analýza rozptylu jednoduchého třídění vyžaduje, aby
a) jednotlivé náhodné výběry byly stochasticky nezávislé
b) jednotlivé náhodné výběry pocházely z binomického rozložení
c) jednotlivé náhodné výběry měly stejný rozptyl.

3. Která z následujících tvrzení jsou pravdivá?
a) Celkový součet čtverců charakterizuje variabilitu jednotlivých pozorování kolem celkového průměru.
b) Skupinový součet čtverců charakterizuje variabilitu jednotlivých pozorování kolem skupinových průměrů.
c) Reziduální součet čtverců charakterizuje variabilitu skupinových průměrů kolem celkového průměru.

4. Nulovou hypotézu o shodě středních hodnot zamítáme na hladině významnosti α, když testové kritérium F_A se realizuje v kritickém oboru

a) $W = (0, F_{\alpha}(r-1, n-r))$

b) $W = (0, F_{1-\alpha}(r-1, n-r))$

c) $W = (F_{1-\alpha}(r-1, n-r), \infty)$

5. Pokud zamítneme hypotézu o shodě středních hodnot a všechny výběry mají stejný rozsah, pak pro zjištění, které dvojice středních hodnot se liší na zvolené hladině významnosti, použijeme

a) Tukeyovu metodu mnohonásobného porovnávání

b) Bartlettův test

c) Levenův test

Správné odpovědi: 1c)   2a), c)   3a)   4c)   5a)

Příklady

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.

Výsledek:

$M_{1.} = 10{,}17$, $M_{2.} = 12$, $M_{3.} = 16{,}83$, $M_{..} = 13$, $S_E = 27{,}7$, $S_A = 142{,}3$, $S_T = 170$, $F_A = 38{,}58$, $F_{0,95}(2, 15) = 3{,}6823$, H_0 tedy zamítáme na hladině významnosti 0,05.

Výsledky zapíšeme do tabulky ANOVA

Zdroj variability	Součet čtverců	Stupně volnosti	podíl	F_A
skupiny	$S_A = 142{,}3$	$f_A = 2$	$S_A/f_A = 71{,}17$	$\frac{S_A/f_A}{S_E/f_E} = 38{,}58$
reziduální	$S_E = 27{,}7$	$f_E = 15$	$S_E/f_E = 1{,}84$	-
celkový	$S_T = 170$	$f_T = 17$	-	-

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	2,03
1, 3	6,67*	2,03
2, 3	4,83*	2,03

Pravá strana: $q_{1-\alpha}(r, n-r)\frac{S_*}{\sqrt{p}} = q_{0,95}(3,15)\frac{\sqrt{1{,}84}}{\sqrt{6}} = 4{,}83\frac{\sqrt{1{,}84}}{\sqrt{6}} = 2{,}03$, kde $S_*^{\,2} = \frac{S_E}{n-r} = 1{,}84$

Na hladině významnosti 0,05 se liší tržby prodavačů 1, 3 a 2, 3.

2. Je dáno pět nezávislých náhodných výběrů o rozsazích 5, 7, 6, 8, 5, přičemž i-tý výběr pochází z rozložení $N(\mu_i,\sigma^2)$, i = 1, ..., 5. Byl vypočten celkový součet čtverců $S_T = 15$ a reziduální součet čtverců $S_E = 3$. Na hladině významnosti 0,05 testujte hypotézu o shodě středních hodnot.

Výsledek:

$n = 5 + 7 + 6 + 8 + 5 = 31$, $r = 5$, $S_A = S_T - S_E = 15 - 3 = 12$

$$F = \frac{S_A/(r-1)}{S_E/(n-r)} = \frac{12/4}{3/26} = 26, F_{0,95}(4,26) = 2,9752$$

Protože $F \geq F_{0,95}(4, 26)$, H_0 zamítáme na hladině významnosti 0,05.

3. Je dána neúplná tabulka ANOVA. Místo otazníků doplňte chybějící čísla.

zdroj variability	součet čtverců	stupně volnosti	podíl	F
skupiny	?	2	?	?
reziduální	16,033	?	?	-
celkový	17,301	35	-	-

Výsledek:

zdroj variability	součet čtverců	stupně volnosti	podíl	F
skupiny	**1,268**	2	**0,634**	**1,304**
reziduální	16,033	**33**	**0,486**	-
celkový	17,301	35	-	-

4. 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 nulové hypotézy zjistěte, které výběry se liší na hladině významnosti 0,05.

metoda	počet bodů							
tradiční	76,2	48,3	85,1	63	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		

Výsledek:

Všech pět náhodných výběrů má rozložení blízké normálnímu rozložení. Levenův test shody rozptylů má testové kritérium 0,863, počet stupňů volnosti je 4 a 26, odpovídající p-hodnota je 0,4991, tedy na hladině významnosti 0,05 hypotézu o shodě rozptylů nezamítáme. Analýza rozptylu má testové kritérium 1,6236, počet stupňů volnosti je 4 a 26, odpovídající p-hodnota je 0,1983, tedy na hladině významnosti 0,05 hypotézu o shodě středních hodnot nezamítáme. Znamená to, že na hladině významnosti 0,05 se neprokázaly odlišnosti ve znalostech studentů.

5. 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	doba čekání						
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.
Výsledek:
Průměrné časy cestování pro tři způsoby dopravy jsou 37 min, 30 min, 36 min.
Všechny tři náhodné výběry mají rozložení blízké normálnímu rozložení. Levenův test shody rozptylů má testové kritérium 0,1054, počet stupňů volnosti je 2 a 15, odpovídající p-hodnota je 0,9007, tedy na hladině významnosti 0,05 hypotézu o shodě rozptylů nezamítáme. Analýza rozptylu má testové kritérium 6,7151, počet stupňů volnosti je 2 a 15, odpovídající p-hodnota je 0,0083, tedy na hladině významnosti 0,05 hypotézu o shodě středních hodnot zamítáme. Scheffého metoda mnohonásobného porovnávání prokázala na hladině významnosti 0,05 rozdíl mezi způsoby A a B a mezi způsoby B a C.