9 Analýza rozptylu: porovnání více průměrů

Mnoho studií má komparativní charakter. Srovnáváme např. platy žen a mužů nebo hodnoty cílové proměnné v kontrolovaném klinickém pokusu. Data v takovém výzkumu se mohou přehledně zobrazit pomocí krabicového grafu nebo sloupkového grafu, také můžeme pro číselné srovnání použít kvantilové charakteristiky rozložení dat ve skupinách nebo průměry a směrodatné odchylky. Při statistickém usuzování se ptáme, zda jsou rozdíly mezi skupinami statisticky významné. Pro porovnání průměrů dvou populací nebo průměrů cílové proměnné sledované v kontrolovaném dvouskupinovém experimentu použijeme *t*-test, který jsme popsali v kapitole 6.2. Studie se však neomezují pouze na dvě skupiny. Potřebujeme tedy metody pro porovnání libovolného počtu průměrů. Takové metody existují a používají se pod shrnujícím názvem **analýza rozptylu**. Je vhodná např. k zodpovězení následujících otázek:

- Liší se ve svém účinku tři metody výuky statistiky na středních školách? Naučí se žáci pomocí některé z těchto metod více než žáci, kteří jsou vyučováni jinými metodami?
- Dožívají se leváci kratšího věku než praváci? Protože ženy žijí déle než muži, je nutné připravit dvoufaktorový plán výzkumu, aby se zohlednil vliv faktoru pohlaví.

Analýzu rozptylu (ANOVA, analysis of variance) rozvinul R. A. Fisher začátkem 20. století. Jedná se o třídu statistických modelů a technik, které lze využít v mnoha výzkumných situacích. Fisherovy statistické práce vznikaly v rámci zemědělského výzkumu v ústavu Rothamsted Experimental Station při vyhodnocování zemědělských pokusů. Před Fisherovým příchodem se zemědělský pokus prováděl tak, že se pole rozdělilo na několik „parcel" a každá parcela byla ošetřena (hnojena) určitým způsobem. Následně se srovnávaly výnosy z jednotlivých parcel. Bylo však zřejmé, že rozdíly ve výnosech mohou být způsobeny jinými faktory než hnojivem, např. rozdílnou kvalitou půdy. Fisher navrhl roz-

dělit pole na bloky a každý blok na parcely (plots). Každá parcela uvnitř bloku pak dostala náhodně přiřazené ošetření. Toto schéma výzkumu (design) se stalo základem pro analýzu, známou dnes pod názvem analýza rozptylu.

Jádro Fisherova postupu tvoří dva principy. První spočívá v klíčové roli, již hraje randomizace (znáhodnění) v experimentování. Tím se zajišťuje eliminace systematického zkreslení efektu posuzovaného ošetření. Randomizace také umožňuje validní odhady náhodné chyby, což má vztah k druhému principu, k možnosti kontrolovaně odhadovat náhodnou chybu opakování ošetření. Opakováním každého ošetření v blocích a uvnitř bloku můžeme v pozorované variabilitě lépe identifikovat variabilitu mezi parcelami uvnitř bloků po „očištění" od variability mezi bloky. Fisher popsal základy analýzy rozptylu v knížce *Statistical methods for research* z roku 1925.

Obecně spočívá základní funkce analýzy rozptylu v posouzení hlavních a interakčních efektů kategoriálních nezávislých proměnných na závisle proměnnou kvantitativního typu. Nezávisle proměnné v ANOVA často nazýváme faktory a jejich hodnoty úrovně nebo kategorie.

Nejjednodušším případem je analýza rozptylu jednoduchého třídění, kdy analyzujeme efekt jednoho faktoru na závisle proměnnou. Jde o přímé zobecnění případu zkoumání rozdílu mezi dvěma nezávislými skupinami pomocí t-testu na případy zkoumání rozdílů mezi více skupinami (odpovídajícím jednotlivým úrovním neboli kategoriím faktoru). Pokud zkoumáme vliv více faktorů, rozlišujeme mezi hlavními efekty a efekty, které jsou způsobeny interakcemi mezi faktory při působení na závisle proměnnou. **Hlavní efekt** je přímý efekt faktoru na závisle proměnnou. **Interakční efekt** je spojený efekt kombinace dvou nebo více faktorů na závisle proměnnou.

Pomocí obecnějších modelů analyzujeme podobné úlohy s kategoriální závisle proměnnou (GLM, general linear model). Také existují postupy, jež do modelu zařazují mezi nezávisle proměnné intervalové proměnné (analýza kovariance, ANCOVA, analysis of covariance) nebo modely, které analyzují simultánně ovlivnění více nezávislých proměnných (MANOVA, multiple analysis of variance).

Základní statistikou v analýze rozptylu je F-testovací statistika rozdílnosti skupinových průměrů, pomocí níž se testuje hypotéza, zda průměry ve skupinách určených kombinacemi faktorů se od sebe liší více než na základě působení náhodného kolísání. Pokud se průměry neliší významně, usuzujeme, že faktory nemají na závisle proměnnou vliv. Jestliže F-test indikuje nějaký systematický vliv, používají se testy simultánního srovnávání pro nalezení kombinací hodnot faktorů, které nejvíce přispívají k systematickým vlivům.

Testovací F-statistika musí zohlednit rozdílnosti ve výběrových průměrech a zároveň přirozenou variabilitu závisle proměnné. V podstatě měří určitým

způsobem velikost rozdílnosti výběrových průměrů a její hodnota závisí jednak na výběrových průměrech v jednotlivých skupinách, dále na velikosti těchto skupin a na rozptylu závisle proměnné uvnitř skupin. Obecně má F-statistika v analýze rozptylu formu:

$$F = \frac{\textit{vážený rozptyl mezi průměry skupin}}{\textit{rozptyl mezi jedinci ve stejné skupině}}$$

Pokud celková rozdílnost měřená F-statistikou překročí určenou kritickou mez, zamítá se nulová hypotéza, že všechny teoretické průměry mají stejnou hodnotu. Analýzou konstrukce F-testu lze ukázat, že podobně jako u jiných testů dojde k zamítnutí hypotézy, pokud jsou rozdíly průměrů relativně veliké. Jestliže zamítneme hypotézu rovnosti průměrů, ještě nám to nic neřekne o rozdílech mezi jednotlivými průměry. Další analýza se musí provést pomocí metod následného zkoumání těchto rozdílů.

Jestliže je plán výzkumu uspořádán vnitroskupinově s opakovaným měřením závisle proměnné u *stejných* objektů, jako je tomu např. u hodnocení měření před pokusem a po pokusu, výpočty při provádění F-testů se odlišují od testů v plánech výzkumu prováděných meziskupinově.

Procedury analýzy rozptylu předpokládají, že závisle proměnná v jednotlivých skupinách má normální rozdělení se stejným rozptylem. Často se také předpokládá, že pro každou skupinu je rozsah výběru měřených jednotek stejný. Tato symetrie zjednodušuje některé výpočty.

9.1 Analýza rozptylu při jednoduchém třídění

Analýza rozptylu při jednoduchém třídění (one-way ANOVA) analyzuje diference průměrů sledované závisle proměnné mezi skupinami, které jsou určeny jednou kategoriální nezávisle proměnnou (faktorem). Zkoumá se, zda skupiny vytvořené tímto klasifikačním faktorem jsou podobné, nebo zda jednotlivé průměry tvoří nějaké identifikovatelné shluky. Jestliže faktor má jenom dvě kategorie (úrovně), úloha je totožná s testováním rovnosti průměru ve dvou nezávislých výběrech pomocí t-testu nebo testování hypotézy, že korelační koeficient mezi závisle proměnnou a binární proměnnou určující příslušnost měření do jedné z obou skupin má nulovou hodnotu.

Popišme přesněji situaci, kterou analyzujeme pomocí této metody. Označme závisle proměnnou X. Provedeme měření na prostých náhodných výběrech objektů z m populací ($j = 1, 2, \ldots, m$). Rozsahy výběrů n_j mohou být různé. Pro

každý výběr j vypočítáme příslušný průměr x_{ij} a rozptyl s_j^2. Předpokládáme, že měření vyhovují modelu

$$x_{ij} = \mu + \alpha_j + e_{ij},$$

kde x_{ij} označuje i-té měření ($i = 1, 2, \ldots, n_j$) v j-tém výběru a μ je společná část průměru. Efekt skupiny α_j způsobuje, že průměry μ_j sledované proměnné v populacích si nemusí být rovny. Měření x_{ij} se liší od průměru ve své skupině o náhodnou odchylku e_{ij}. O této odchylce předpokládáme, že je normálně rozdělená s nulovou střední hodnotou a s rozptylem σ^2, jenž je pro všechny měření stejný. Symbolem n označujeme celkový počet měření, který se rovná součtu všech n_j.

Základní hypotéza, jež nás zajímá, předpokládá, že jsou všechny průměry v jednotlivých populacích stejné, tedy $H_0: \mu_1 = \mu_2 = \cdots = \mu_m$, alternativa je H_1: *Ne všechny μ_j jsou stejné*, nebo v jiném přepisu $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$, s alternativou H_1: *Ne všechny a_j jsou nulové.*

Nulová hypotéza tedy znamená, že faktor neovlivňuje závisle proměnnou X. Při analýze dat zkoumáme, zda vypočtené průměry $\bar{x}_j$ se od sebe liší jen v mezích náhodného kolísání od společného průměru $\bar{x}$, nebo zda je nulová hypotéza porušena.

Konstrukce testovací F-statistiky vychází z rozkladu součtu čtverců odchylek měření od společného průměru $\bar{x}$. Odchylku měření x_{ij} od $\bar{x}$ přepíšeme ve formě rozkladu $x_{ij} - \bar{x} = (x_{ij} - \bar{x}_j) + (\bar{x}_j - \bar{x})$. V tomto vyjádření hodnota $\bar{x}_j - \bar{x}$ představuje odhad parametru α_j, tedy efektu kategorie j. Jestliže umocníme a sečteme obě strany rovnice pro všechna měření, pak po úpravách (při kterých se vyruší všechny ostatní členy vzniklé při umocnění) dostaneme

$$S_T = \sum_j \sum_i \left(x_{ij} - \bar{x}\right)^2 = \sum_j \sum_i \left(x_{ij} - \bar{x}_j\right)^2 + \sum_j n_j \left(\bar{x}_j - \bar{x}\right)^2 = S_e + S_A.$$

Tento výraz říká, že celkový součet čtverců S_T se rovná součtu čtverců odchylek uvnitř výběrů S_e a součtu čtverců rozdílů mezi výběry S_A. K součtům čtverců rozdílů patří stupně volnosti $n - 1$, $n - m$ a $m - 1$. Když jimi vydělíme příslušné součty, dostaneme tzv. průměrné čtverce MS (mean squares):

$$MS_T = S_T/(n-1),$$
$$MS_e = S_e/(n-m),$$
$$MS_A = S_A/(m-1)$$

Platí, že statistika $F = MS_A/MS_e$ má za platnosti nulové hypotézy F-rozdělení se stupni volnosti $(m - 1, n - m)$. Tato F-statistika měří globální odchylku dat od nulové hypotézy. Je tomu tak proto, že obě hodnoty MS_A a MS_e odhadují za

Tab. 9.1 Schéma tabulky analýzy rozptylu

Zdroj variability	S	st.v.	MS	F
faktor A	S_A	$m - 1$	$MS_A = \frac{S_A}{m-1}$	$F = \frac{MS_A}{MS_e}$
reziduální	S_e	$n - m$	$MS_e = \frac{S_e}{n-m}$	
Celková variabilita	S_T	$n - 1$		

platnosti nulové hypotézy společný rozptyl σ^2. Pokud je však porušena nulová hypotéza, hodnota MS_A bude významně větší než MS_e. Velikost mezního poměru určuje kritická mez pro F-rozdělení s příslušnými stupni volnosti. Výpočty se obvykle shrnují tabulkou analýzy rozptylu (tab. 9.1).

PŘÍKLAD 9.1

Analýza rozptylu při jednoduchém třídění

Modelová data v tabulce 9.2 popisují hodnoty kontrolního testu u studentů. Každý sloupec obsahuje data pro skupinu náhodně vybraných studentů ze skupin s danou metodou výuky. Zkoumáme nulovou hypotézu, že průměry hodnot jsou stejné ve všech skupinách. To znamená, že typ výuky (faktor A) nemá vliv na průměrnou hodnotu testu. Pro statistický test volíme hladinu významnosti 0,05. Naše hypotézy jsou:

$$H_0: \mu_1 = \mu_2 = \mu_3$$

H_1: Ne všechny μ jsou stejné.

Tab. 9.2 Příklad dat, u nichž provádíme analýzu rozptylu (jednoduché třídění)

	Metoda výuky		
	1	2	3
	89	104	86
	101	120	98
	87	98	100
	87	110	96
Součet	364	432	380

Průměry ve skupinách mají hodnoty

$$\bar{x}_1 = \frac{364}{4} = 91, \quad \bar{x}_2 = \frac{432}{4} = 108 \text{ a } \bar{x}_3 = \frac{380}{4} = 95.$$

Celkový průměr je průměr ze všech hodnot

$$\bar{x} = \frac{364 + 432 + 380}{12} = 98.$$

Vypočítáme tři součty čtverců.

a) Celkový součet čtverců je totéž co čitatel při počítání výběrového rozptylu ze všech 12 měření:

$$S_T = \sum_j \sum_i (x_{ij} - \bar{x})^2 = \begin{pmatrix} (89-98)^2 + (104-98)^2 + (86-98)^2 \\ + (101-98)^2 + (120-98)^2 + (98-98)^2 \\ + (87-98)^2 + (98-98)^2 + (100-98)^2 \\ + (87-98)^2 + (110-98)^2 + (96-98)^2 \end{pmatrix} = 1148$$

b) Součet čtverců uvnitř skupin má stejný počet členů. Ty jsou však určeny rozdíly mezi měřením a příslušným skupinovým průměrem:

$$S_e = \sum_j \sum_i (x_{ij} - \bar{x}_j)^2 = \begin{pmatrix} (89-91)^2 + (104-108)^2 + (86-95)^2 \\ + (101-91)^2 + (120-108)^2 + (98-95)^2 \\ + (87-91)^2 + (98-108)^2 + (100-95)^2 \\ + (87-91)^2 + (110-108)^2 + (96-95)^2 \end{pmatrix} = 516$$

c) Součet čtverců mezi výběry má také stejný počet členů, ale počítá se pouze se skupinovými průměry a celkovým průměrem:

$$S_A = \sum_j \sum_i (\bar{x}_j - \bar{x})^2 = \begin{pmatrix} (91-98)^2 + (108-98)^2 + (95-98)^2 \\ + (91-98)^2 + (108-98)^2 + (95-98)^2 \\ + (91-98)^2 + (108-98)^2 + (95-98)^2 \\ + (91-98)^2 + (108-98)^2 + (95-98)^2 \end{pmatrix} = 632$$

Protože se v tomto posledním vzorci opakují průměry ve sloupci pro skupinu, lze jej zjednodušit: $S_A = \sum_j n_j (\bar{x}_j - \bar{x})^2 = 4(91-98)^2 + 4(108-98)^2 + 4(95-98)^2 = 632$

Všimněte si, že $S_A + S_e = S_T$, takže výpočet jednoho ze třech součtů je zbytečný. Do schématu tabulky rozptylu (tab. 9.1, s. 341) doplníme příslušné vypočtené hodnoty (tab. 9.3)

Tab. 9.3 Příklad vyplněné tabulky analýzy rozptylu

Zdroj variability	S	$st.v.$	MS	F	$F_{0,05}$
mezi výběry	632	2	316	5,51	$F(2, 9) = 4,26$
reziduální	516	9	57,333		
Celková variabilita	1148	11			

a srovnáme vypočítané F s kritickou hodnotou F-rozdělení s příslušnými stupni volnosti s hladinou významnosti 0,05. Pokud je F větší než kritická mez, nulovou hypotézu zamítáme. Protože je v našem případě testovací statistika F větší než kritická mez, můžeme tvrdit, že máme evidenci pro zamítnutí nulové hypotézy o rovnosti průměrů kontrolního testu ve skupinách.

9.1.1 Ověření předpokladů analýzy rozptylu

Aby jednoduchá analýza rozptylu byla validní, musí být splněny následující předpoklady:

1. Všechna měření musí být vzájemně nezávislá uvnitř skupin i mezi skupinami.
2. Měření v každé skupině jsou normálně rozdělená s průměrem μ_i.
3. Ve všech skupinách mají měření stejný rozptyl kolem průměru.

Ověřování uvedených předpokladů se provádí numericky pomocí různých testů, korelační analýzou a graficky. Především se provádí tzv. „analýza reziduálních hodnot", což vyžaduje:

- výpočet reziduálních hodnot $x_{ij} - \bar{x}_j$ a hodnot $\bar{x}_j$;
- grafické znázornění reziduálních hodnot a jejich absolutních hodnot proti hodnotám faktorů a hodnotám závisle proměnné a zjištění změn, trendů a konfigurací v těchto hodnotách;
- prověření normality rozdělení reziduálních hodnot graficky a statistickým testem.

O výsledku této kontroly má výzkumník informovat čtenáře ve výzkumné zprávě. Jestliže je splněn předpoklad nezávislosti a homogenity rozptylů, grafy reziduálních hodnot neobsahují žádné systematické konfigurace bodů (podrobněji kap. 7.3.2).

Předpoklad rovnosti rozptylů se často těžko ověřuje. Minimálně zkoumáme, zda je splněno, že

$$\frac{\max s_i}{\min s_i} \leq 3,$$

kde s_i jsou směrodatné odchylky měření v jednotlivých skupinách.

Obecně lze říci, že největší vliv na validitu F-testu má předpoklad o statistické nezávislosti všech měření. Vůči poruchám ostatních předpokladů je analýza rozptylu poměrně robustní.

Jestliže analýza reziduálních hodnot odhalí nehomogenitu, porušení normality nebo systematické chyby způsobené špatně sestaveným modelem, pokoušíme se tyto nedostatky ošetřit transformací závisle proměnné. Tento postup může fungovat, jestliže poměr (největší hodnota pozorování)/(nejmenší hodnota pozorování) je veliký. Nejpoužívanější transformace jsou:

- logaritmická transformace, pokud se rozptyl zvyšuje úměrně s průměrem;
- arcussinová transformace pro relativní četnosti;
- druhá odmocnina pro četnosti.

Poslední dvě transformace se používají při analýze dat četnostního typu.

9.1.2 Simultánní porovnávání

F-test v analýze rozptylu je tzv. omnibus test. To znamená, že je zaměřen na porušení globální hypotézy rovnosti průměrů. Neříká nám nic o zdrojích diferencí mezi skupinami. Specifické otázky zodpovídá další následné zkoumání rozdílů průměrů, o němž budeme mluvit v následujících odstavcích. Obecný úvod do problematiky jsme podali v kapitole o teorii statistického testování. Rozlišujeme *porovnávání předem plánované* a *post-hoc porovnávání* (post-hoc znamená po testování předem určených hypotéz), kdy se přistupuje ke zkoumání rozdílů dvojic průměrů bez předem daného záměru.

Často je výhodnější přímo odhadovat velikost rozdílu vhodným intervalem spolehlivosti a ten použít pro testování hypotézy, že mezi průměry není statisticky významný rozdíl. Interval spolehlivosti se počítá pomocí rozdílu průměrů porovnávaných skupin a reziduální směrodatné odchylky s_e, která se spočítá jako odmocnina z průměrných reziduálních čtverců MS_e.

Plánované testy a intervaly spolehlivosti

Pokud předem plánujeme určitá porovnání, použije se modifikace t-testu podle Bonferroniho. Tento přístup jsme objasnili v kapitole 5.5. Při výpočtech použijeme upravenou hladinu významnosti a běžný postup srovnání nezávislých

průměrů pomocí t-testu nebo příslušného intervalu spolehlivosti. Čím více porovnávání děláme, tím musíme učinit hladinu významnosti přísnější. Bonferroniho modifikace spočívá v tom, že při provádění k porovnávání na hladině 0,05 nastavíme hladinu významnosti ve výpočtech na $\alpha' = 0{,}05/k$. Například pro 10 srovnání a při celkové hladině významnosti 0,05 musíme provést každé srovnání na hladině $0{,}05/10 = 0{,}005$.

Statistické testy a výpočty intervalů spolehlivosti podle Bonferroniho potřebují při mnoha srovnáních extrémně malé hladiny významnosti, proto se nepoužívají, jestliže chceme provést všechna srovnání průměrů mezi sebou. Zvyšuje se totiž pravděpodobnost chyby II. druhu a rostou šířky intervalů spolehlivosti.

Při srovnání průměru skupin i a j označíme

$$s' = s_e \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \quad \text{kde} \quad s_e = \sqrt{MS_e}.$$

Pro testy a odhady použijeme intervaly spolehlivosti pro rozdíly průměrů v populacích:

$$\mu_i - \mu_j \in \left(\bar{x}_i - \bar{x}_j - t_{\alpha'/2}s';\ \bar{x}_i - \bar{x}_j + t_{\alpha'/2}s'\right)$$

Přitom hledáme kritické hodnoty t-rozdělení s $n - m$ stupni volnosti, kde n je počet všech pozorování a m je počet skupin.

Obecně se plánované srovnání týká kontrastů. **Kontrast** je váženým součtem průměrů (jejich lineární kombinací), který je dán koeficienty (vektorem vah), jejichž součet je roven nule. Jestliže jsou jenom dvě váhy různé od nuly, jedná se o prostý rozdíl dvou průměrů. Testem kontrastu nebo pomocí intervalu spolehlivosti se přesvědčujeme, zda je zkoumaný kontrast významně různý od nuly. V případě srovnání tří průměrů mohou vektory vah mít např. podobu $(-1; 0{,}5; 0{,}5)$ nebo $(1; -1; 0)$. Tyto dva vektory odpovídají kontrastům $\bar{x}_1 - 0{,}5\bar{x}_2 - 0{,}5\bar{x}_3$ a $\bar{x}_1 - \bar{x}_2$.

Post-hoc testy a intervaly spolehlivosti

Pro zkoumání rozdílů, jež nejsou předem vymezeny na základě teorie, existuje několik post-hoc testů. Každý má trochu jiné vlastnosti. Jako obecný princip volby platí, že máme použít test více konzervativní, jinak vzroste nekontrolovatelně pravděpodobnost chyby I. druhu. Konzervativní test si udržuje za dosti volných předpokladů předpokládanou hladinu významnosti. U konzervativních testů jsou příslušná rozhodnutí zpravidla prováděna spíše na menší hladině významnosti.

Scheffeho test se často používá pro kontrolu chyb I. druhu při post-hoc testování. Uplatníme ho teprve po zamítnutí nulové hypotézy rovnosti průměrů pomocí F-testu. Nechť F je kritická mez příslušného F-testu. Pro Scheffeho test použijeme upravenou mez $F' = (m-1)F$, kde m je počet skupin. Udržení hladiny

pravděpodobnosti chyb I. druhu jde na účet snížení síly testu. Zvyšuje se tedy hladina pravděpodobnosti chyby II. druhu. Nulovou hypotézu zamítáme, jestliže

$$\left|\bar{x}_i - \bar{x}_j\right| > \sqrt{F'}s',$$

kde s' počítáme stejně jako v případě plánovaného srovnávání. Interval spolehlivosti pro rozdíl průměrů má tvar

$$\mu_i - \mu_j \in \left(\bar{x}_i - \bar{x}_j\right) \pm \sqrt{F'}s'.$$

LSD test (least significant difference test) se také nazývá **Fisherův LSD test**. Používá se pro porovnání všech dvojic poté, co byla zamítnuta globální hypotéza o rovnosti všech průměrů. LSD test je nejliberálnější test ze všech post-hoc testů (je velmi pravděpodobné, že zamítne nulovou hypotézu o rovnosti zvolené dvojice průměrů). Použije se stejně jako t-test podle Bonferroniho, ale bez úprav hladiny významnosti. To se týká i určení intervalů spolehlivosti. Interval spolehlivosti vhodný pro test rozdílu průměrů ve skupinách má tvar

$$\mu_i - \mu_j \in \left(\bar{x}_i - \bar{x}_j - t_{\alpha/2}s';\ \bar{x}_i - \bar{x}_j + t_{\alpha/2}s'\right),$$

kde kritickou mez $t_{\alpha/2}$ určujeme na základě t-rozdělení s $n - m$ stupni volnosti.

9.1.3 Velikost účinku

Problematiku účinku jsme nastínili již v kapitole o teorii statistického testování (kap. 5.3.8). V analýze rozptylu se používají dvě míry účinku. Jednoduchou míru představuje poměr variability vysvětlené kategoriemi S_A k celkové variabilitě měřené součtem čtverců odchylek od celkového průměru S_T. Tuto míru někdy označujeme η^2:

$$\eta^2 = \frac{S_A}{S_T}$$

Jiný způsob vyjádření účinkuje výpočet tzv. ω^2:

$$\omega^2 = \frac{S_A - (m-1)MS_e}{S_T + MS_e}$$

Tento koeficient je přesnější, protože uvažuje nevysvětlenou variabilitu (MS_e). Obvykle má menší hodnotu než η^2. Obě míry, η^2 i ω^2, nabývají hodnoty mezi nulou a jedničkou. Používají se při srovnávání síly vlivu faktoru na závisle proměnnou v různých situacích a mezi různými studiemi. V podstatě udávají, kolik variability dat je vysvětleno faktorem. Jestliže se jejich hodnoty blíží k jedné, je efekt veliký. Znamená to také, že faktor vysvětluje značnou část variability dat.

9.1.4 Kruskalův-Wallisův test

Jedná se o podobný test jako F-test pro jednoduchou analýzu rozptylu. Použijeme ho, jestliže nemůžeme vycházet z předpokladu, že měření jsou normálně rozdělená. Nulová hypotéza předpokládá, že měření ve skupinách mají stejné mediány:

$$\mathrm{H}_0: \tilde{\mu}_1 = \tilde{\mu}_2 = \cdots = \tilde{\mu}_m$$

$$\mathrm{H}_1: \text{Alespoň pro jednu dvojici } i, j \text{ platí, že } \tilde{\mu}_i \neq \tilde{\mu}_j$$

Při tomto testu nejdříve uspořádáme všechna n měření podle velikosti. Nahradíme hodnoty jejich pořadími a vypočítáme koeficienty SR_i jako součty pořadí měření ze skupiny i. Dále spočítáme testovací statistiku H, která měří rozdílnost průměru pořadí ve skupinách:

$$H = \left[\frac{12}{n(n+1)}\sum_i\left(\frac{(SR_i)^2}{n_i}\right)\right] - 3(n+1)$$

Pro velká n_i má statistika H za platnosti nulové hypotézy přibližně χ^2 rozdělení. Hypotézu o stejných mediánech zamítáme na (přibližné) hladině významnosti α, jestliže H je větší než příslušná kritická hodnota χ^2 rozdělení o $m - 1$ stupních volnosti.

Při větším počtu shod upravíme hodnotu testovací statistiku H tak, že ji vydělíme hodnotou B, která se spočítá podle vzorce

$$B = 1 - \sum_{i=1}^{r}(t_i^3 - t_i)/(n^3 - n),$$

kde r je počet skupin se shodným měřením, t_i je počet stejných měření ve skupině i, n je počet všech měření.

PŘÍKLAD 9.2

Kruskalův-Wallisův test

Použijeme data z příkladu 9.1 o jednoduché analýze rozptylu (tab. 9.2, s. 341). Předpokládáme, že data nepocházejí z normálního rozdělení. Chceme provést test rovnosti mediánů na hladině významnosti $\alpha = 0{,}05$. Základní výpočty zachycuje tabulka 9.4. Vypočítáme testovací statistiku H:

$$H = \frac{12}{12 \times 13}\left(\frac{18{,}0^2}{4} + \frac{39{,}5^2}{4} + \frac{20{,}5^2}{4}\right) - 3 \times 13 = \frac{1}{13} \times 576{,}125 - 39 = 5{,}3173$$

Tab. 9.4 Příklad postupu při Kruskalově-Wallisově testu rovnosti mediánů – srovnávání metod výuky

	Metoda výuky (původní data)			Metoda výuky (pořadí dat)		
	1	2	3	1	2	3
	89	104	86	4	10	1
	101	120	98	9	12	6,5
	87	98	100	2, 5	6,5	8
	87	110	96	2,5	11	5
SR_i				18,0	39,5	20,5
n_i				4	4	4

V tomto případě je velikost výběru malá. Proto by bylo vhodné vyhledat kritickou hodnotu ve speciální tabulce pro tento test. Přesto použijeme k aproximativnímu posouzení významnosti výsledku tabulku χ^2-rozdělení s $m - 1$ stupni volnosti. Kritická hodnota pro hladinu významnosti $\alpha = 0{,}05$ má hodnotu 5,991. Protože testovací statistika $H = 5{,}317$, nemůžeme hypotézu rovnosti mediánů ve skupinách zamítnout.

Kruskalův-Wallisův simultánní test se používá pro párové srovnání mediánů po aplikaci globálního Kruskalův-Wallisova testu. Při simultánním porovnávání párů skupin opět použijeme Kruskalův-Wallisův test, ale s upravenou hladinou významnosti podle Bonferroniho $\alpha' = \alpha/(m(m-1))$, kterou se kontroluje simultánní hladina významnosti α. Testovací statistika se počítá stejným způsobem s tím, že tentokrát pracujeme jenom se dvěma skupinami.

9.1.5 Jonckheere-Terpstra test

V některých situacích není vhodná obecná alternativní hypotéza, jež se uvažuje v Kruskal-Wallisově testu při porovnávání mediánů populací $\tilde{\mu}_i$. Například při testování efektivity nějakého léku u m skupin pacientů chce výzkumník vědět, zda při rostoucí dávce roste také účinek. Pro takové situace je vhodnější uvažovat alternativní hypotézu v uspořádaném tvaru $H_1 : \tilde{\mu}_1 \leq \tilde{\mu}_2 \leq \cdots \leq \tilde{\mu}_m$ *a aspoň jedna nerovnost je ostrá*. Jestliže srovnáváme jenom dvě hladiny dávky léku, pak v této situaci použijeme test pro dva nezávislé výběry (Wilcoxonův test), ale jednostrannou hypotézu.

Jonckheere a Terpstra vyvinuli neparametrický test pro více nezávislých skupin, který má pro uspořádanou alternativu větší sílu než Kruskal-Wallisův test.

Testovací statistika se vypočte podle formule

$$J = \sum_{u<v} U_{uv},$$

kde U_{uv} vyjadřují, kolik měření X_{au} ze skupiny u je menších než měření X_{bv} ze skupiny v (jestliže jsou si měření rovna, připočteme číslo 0,5). Každé pozorování a ze skupiny u se porovná s každým pozorováním b ze skupiny v:

- Jestliže pozorování z prvního výběru je menší než pozorování z druhého výběru, k U_{uv} se připočte jednička.
- Jestliže pozorování z prvního výběru je větší než pozorování z druhého výběru, k U_{uv} se připočte nula.
- Jestliže pozorování z prvního výběru je stejné jako pozorování z druhého výběru, k U_{uv} se připočte číslo 0,5.

Nulová hypotéza se zamítne, pokud je testovací statistika J větší než kritická hodnota rozdělení statistiky J za předpokladu rovnosti mediánů. Asymptoticky platný test se opírá o z-statistiku

$$z = \frac{J - E(J)}{\sqrt{Var(J)}},$$

kde střední hodnota statistiky J za platnosti nulové hypotézy má průměr $E(J) = (n^2 - \sum n_i^2)/4$ a rozptyl $Var(J) = [n^2(2n+3) - \sum n_i^2(2n_i + 3)]/72$, kde n_i jsou počty pozorování v jednotlivých skupinách. Tato statistika je nepříznivě ovlivňována velkým počtem stejných pozorování. Pak je nutná korekce vzorce (viz Sprent, 2001).

Korekce se počítá zejména tehdy, jestliže jsou měření uspořádána do kontingenční tabulky způsobem, který jsme popsali v kapitole 8.4. Modifikujeme $Var(J)$ podle formule $Var(J) = U_1/d_1 + U_2/d_2 + U_3/d_3$, kde:

$$\begin{aligned}
U_1 &= n(n-1)(2n+5) - \sum_i n_{i.}(n_{i.}-1)(2n_{i.}-5) - \sum_j n_{.j}(n_{.j}-1)(2n_{.j}-5) \\
U_2 &= \left[\sum_i n_{i.}(n_{i.}-1)(n_{i.}-2)\right]\left[\sum_j n_{.j}(n_{.j}-1)(n_{.j}-2)\right] \\
U_3 &= \left[\sum_i n_{i.}(n_{i.}-1)\right]\left[\sum_j n_{.j}(n_{.j}-1)\right] \\
d_1 &= 72, \\
d_2 &= 36n(n-1)(n-2), \\
d_3 &= 8n(n-1)
\end{aligned}$$

PŘÍKLAD 9.3

Jonckheere-Terpstra test

Skupina 24 dětí s vadou sluchu rozdělených do 3 věkových kategorií absolvovala psychologický test základních pojmů. Ptáme se, zda existuje trend v mediánech výsledků? Hrubé skóry testu obsahuje tabulka 9.5. Provedeme test na hladině 0,05. Vypočítáme statistiku J.

$$U_{12} = 6 + 6 + 5 + 2{,}5 + 2{,}5 + 1{,}5 = 23{,}5$$

$$U_{13} = 12 + 12 + 10 + 7 + 7 + 4 = 52$$

$$U_{23} = 10{,}5 + 10 + 9 + 7 + 4 + 2 = 42{,}5$$

$$J = 23{,}5 + 52 + 42{,}5 = 118$$

$$E(J) = 0{,}25 \times (24 \times 24 - 6 \times 6 - 6 \times 6 - 12 \times 12) = 90$$

$$V(J) = [24 \times 24 \times (2 \times 24 + 3) - 6 \times 6 \times (2 \times 6 + 3) - 6 \times 6 \times (2 \times 6 + 3) - 12 \times 12 \times (2 \times 12 + 3)]/72 = 339$$

Například srovnáme skupiny šestiletých a sedmiletých. Pozorování 17 je menší než všechna pozorování v druhé skupině. Totéž platí pro pozorování 20. Pozorování 24 je menší než pět pozorování v druhé skupině. Takto porovnáme všechny dvojice a dostaneme číslo U_{12}.

Testovací statistika z má hodnotu $z = (118 - 90)/\sqrt{339} = 1{,}52$. Jelikož pro hladinu 0,05 použijeme kritickou mez 1,96, lze uzavřít, že nemůžeme potvrdit vzestupný trend v hodnotách výsledků psychologického testu s rostoucím věkem.

Tab. 9.5 Příklad dat, v nichž testujeme uspořádanou alternativní hypotézu (trend v mediánech) – výsledky psychologického vyšetření dětí různého věku

Věk 6	Věk 7	Věk 8	
17	23	22	36
20	25	23	38
24	27	26	38
34	34	32	42
34	38	34	48
38	47	34	50

9.2 Analýza rozptylu dvojného třídění

Při dvojném třídění zkoumáme vliv dvou faktorů na závisle proměnnou. Označme tyto faktory A a B. Pomocí analýzy rozptylu dvojného třídění analyzujeme často tzv. blokové experimenty, při nichž zkoumáme vliv faktoru A, který plánovitě měníme, zatímco faktor B považujeme za vliv rušivý. Vliv rušivého faktoru se snažíme oddělit od vlivu faktoru A. Proto se při provádění takového experimentu nejdříve objekty rozdělí do tzv. bloků podle úrovně faktoru B a uvnitř bloku se objekty náhodně přiřadí k úrovním faktoru A. V tomto odstavci se budeme zabývat modelem, jenž vychází z předpokladu, že data mají normální rozdělení. Pokud tomu tak není, použijeme některou z neparametrických metod podobným způsobem jako u jednoduché analýzy rozptylu.

Označme a počet úrovní faktoru A a b počet úrovní faktoru B. Počet objektů odpovídajících i-té úrovni faktoru A a j-té úrovni faktoru B označme n_{ij}. Nejčastěji se setkáváme s tzv. znáhodněnými úplnými bloky neboli vyváženým tříděním, kdy všechny četnosti n_{ij} jsou shodné ($n_{ij} = c$). Stejně jako při analýze rozptylu jednoduchého třídění považujeme efekty obou faktorů za konstantní. Lze též uvažovat případ, kdy je jeden efekt pevný a druhý náhodný, hovoří se pak o modelu se smíšenými efekty. Abychom mohli testovat interakce mezi faktory, je vhodné, aby n_{ij} byly větší než 1. Pro naměřené hodnoty uvažujeme model:

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

kde μ je společná část průměru závislé proměnné,
α_i je efekt faktoru A na úrovni i ($i = 1, \ldots, a$),
β_j je efekt faktoru B na úrovni j ($j = 1, \ldots, b$),
γ_{ij} je interakce mezi faktorem A na úrovni i a faktorem B na úrovni j.

Poslední člen v rovnici označuje náhodnou chybu, o které předpokládáme, že má nulovou střední hodnotu, normální rozdělení a stejný rozptyl pro všechny hodnoty i a j, přičemž podobně jako v jednoduché analýze rozptylu musí všechna měření představovat statisticky nezávislé náhodné proměnné. Pro každou kombinaci faktorů měříme c objektů ($k = 1, \ldots, c$), platí tedy, že všechna $n_{ij} = c$. Předpokládáme, že hodnota c je větší než 1.

Při analýze zkoumáme tři páry hypotéz:

a) $H_{01}: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$.
H_{11}: Ne všechny efekty α_i jsou nulové.
b) $H_{02}: \beta_1 = \beta_2 = \cdots = \beta_b = 0$.
H_{12}: Ne všechny efekty β_j jsou nulové.
c) H_{03}: Mezi faktory A a B není žádná interakce (všechny $\gamma_{ij} = 0$).
H_{13}: Některé interakce jsou nenulové.

Konstrukce testovací F-statistiky opět vychází z rozkladu součtu čtverců odchylek měření od společného průměru $\bar{x}$. Podobně jako u jednoduché analýzy rozptylu platí rozklad odchylky $(x_{ijk} - \bar{x})$ měření x_{ijk} od celkového průměru $\bar{x}$:

$$(x_{ijk} - \bar{x}) = (x_{ijk} - \bar{x}_{ij.}) + (\bar{x}_{i..} - \bar{x}) + (\bar{x}_{.j.} - \bar{x}) + (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x})$$

Sčítanci $(\bar{x}_{i..} - \bar{x})$, $(\bar{x}_{.j.} - \bar{x})$, $(\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x})$ v tomto výrazu představují bodové odhady efektů α_i, β_j a γ_{ij}.

Jestliže sečteme čtverce obou stran rovnice pro všechna měření, dostaneme po úpravách:

$$S_T = \sum_i \sum_j \sum_k (x_{ijk} - \bar{x})^2 = \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 +$$
$$ac \sum_i (\bar{x}_{i..} - \bar{x})^2 + bc \sum_j (\bar{x}_{.j.} - \bar{x})^2 + c \sum_i \sum_j (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x})^2$$

Symbolicky tento výraz zapisujeme jako součet jednotlivých částí celkové variability:

$$S_T = S_e + S_A + S_B + S_I$$

Součty čtverců S_A a S_B zachycují hlavní efekty faktorů, součet S_I jejich interakce. Součet čtverců S_e hodnotí variabilitu měření uvnitř skupin a slouží pro odhad společného rozptylu náhodné chyby ε_{ijk} v modelu analýzy rozptylu. Tabulka analýzy rozptylu (tab. 9.6) obsahuje všechny potřebné informace pro získání testovacích statistik. F-statistiky počítáme podobně jako při jednoduchém třídění. Stupně volnosti pro jednotlivé F-statistiky lze jednoduše zjistit z třetího sloupce tabulky se stupni volnosti jednotlivých součtů čtverců.

Abychom mohli oprávněně prohlásit, že faktor A ovlivňuje cílovou proměnnou X, musí nutně platit nulová hypotéza H_{03}. Ta zaručuje, že efekt faktoru A na dané úrovni je stejný pro všechny úrovně faktoru B. Tuto hypotézu tedy zkoumáme jako první. Hlubší výklad této problematiky nalezne čtenář např. v práci

Tab. 9.6 Tabulka analýzy rozptylu u dvojného třídění

Zdroj variability	S	st.v.	MS	F	H_0
faktor A	S_A	$a - 1$	MS_A	MS_A/MS_e	H_{01}
faktor B	S_B	$b - 1$	MS_B	MS_B/MS_e	H_{02}
interakce	S_I	$(a-1)(b-1)$	MS_I	MS_I/MS_e	H_{03}
reziduální	S_e	$ab(c-1)$	MS_e		
Celkový rozptyl	S_T	$abc - 1$			

Tab. 9.7 Příklad analýzy rozptylu při dvojném třídění

	Chlapci (hladina B_1)	Dívky (hladina B_2)
Metoda výuky 1 (hladina A_1)	89 101	87 87
Metoda výuky 2 (hladina A_2)	120 110	98 104
Metoda výuky 3 (hladina A_3)	100 98	86 96

Dvojice hodnot v každé buňce odpovídá dvěma měřením za každé kombinace podmínek.

Tab. 9.8 Příklad výpočtu průměrů pro analýzu rozptylu při dvojném třídění

	Chlapci (faktor B_1)	Dívky (faktor B_2)	Řádkové průměry
Metoda výuky 1 (faktor A_1)	$\bar{x}_{11.} = 95$	$\bar{x}_{12.} = 87$	$\bar{x}_{1..} = 91$
Metoda výuky 2 (faktor A_2)	$\bar{x}_{21.} = 115$	$\bar{x}_{22.} = 101$	$\bar{x}_{2..} = 108$
Metoda výuky 3 (faktor A_3)	$\bar{x}_{31.} = 99$	$\bar{x}_{32.} = 91$	$\bar{x}_{3..} = 95$
Sloupcové průměry	$\bar{x}_{.1} = 103$	$\bar{x}_{.2} = 93$	$\bar{x} = 98$

(Zvára, 1997). Algoritmy výpočtů jsou realizovány ve všech běžných statistických systémech.

Nyní uvedeme pro ilustraci rozbor problému, který vychází z dat příkladu 9.1, resp. 9.2 z odstavce o jednoduché analýze rozptylu. Studovali jsme, zda existuje rozdíl mezi výsledky studentů vyučovaných různými metodami. Teď studenty rozdělíme na chlapce a dívky, abychom zohlednili efekt pohlaví. Tabulka 9.7 ukazuje, jak jsou jednotlivá měření klasifikována podle nového kritéria. Tento problém má tři úrovně faktoru A ($a = 3$) a dvě úrovně faktoru B ($b = 2$). Pro každou kombinaci faktorů se provedla dvě měření ($c = 2$). Můžeme spočítat průměry pro jednotlivé kombinace a celkové průměry pro jednotlivé úrovně faktorů (tab. 9.8). Pomocí F-testů analýzy rozptylu máme přezkoušet tyto hypotézy:

a) H_{01}: Jsou efekty metod výuky stejné?
(Testujeme pomocí $F = MS_A/MS_e$.)

b) H_{02}: Je rozdíl mezi chlapci a dívkami?
(Testujeme pomocí $F = MS_B/MS_e$.)

c) H_{03}: Je efektivita metod výuky stejná pro chlapce i dívky?
(Testujeme pomocí $F = MS_I/MS_e$.)

Tab. 9.9 Příklad tabulky analýzy rozptylu pro dvojné třídění

Zdroj variability	*S*	*st.v.*	*MS*	*F*	Kritická hodnota
faktor *A*	632	2	316	9,88	$F_{0,05}(2,6) = 5,14$
faktor *B*	300	1	300	9,36	$F_{0,05}(1,6) = 5,99$
interakce *A* × *B*	24	2	12	0,38	$F_{0,05}(2,6) = 5,14$
reziduální	192	6	32		
Celkový S_T	1148	11			

Nejdříve zkoumáme třetí hypotézu; jestliže ta platí, můžeme pokračovat v testování první a druhé hypotézy. Pokud je zamítneme, lze použít testy pro simultánní testování rozdílností průměrů pro jednotlivé kombinace faktorů. Tabulka 9.9 je tabulkou analýzy rozptylu pro naše data. Srovnání *F*-statistik s kritickými hodnotami indikuje, že jak faktor *A*, tak faktor *B* ovlivňují závisle proměnnou, přičemž nemusíme uvažovat interakce obou faktorů mezi sebou.

Interakce mezi faktory

Interpretace výsledků analýzy rozptylu pro dvojné třídění závisí silně na přítomnosti interakcí mezi faktory. Interakce jsou jediným podstatným problémem při zobecnění postupu jednoduché analýzy rozptylu pro použití při hodnocení působení více faktorů. Proto se jim budeme věnovat podrobněji (viz také Zvára, 1997).

Analýza interakcí má tři části: 1. odhalení jejich existence; 2. odhad jejich velikosti; 3. provedení jejich interpretace. V tomto odstavci se budeme krátce zabývat třetí komponentou. První dvě lze vyřešit s odkazem na tabulku analýzy rozptylu a analýzou odhadů efektů, které jsme uvedli pro základní situace analýzy rozptylu. Roli přítomnosti interakcí demonstrujeme pro plán dvojného třídění typu 2 × 2. Půjde o modelová data odlišná od předchozího odstavce.

Vycházíme z tabulky průměrů získaných při výzkumném experimentu pro kombinace faktorů POHLAVÍ (chlapci, dívky) a METODA výuky (metoda 1, metoda 2), když hodnotíme působení vztahu obou faktorů na dosažené výsledky (tab. 9.10). Je zřejmé, že faktor METODA nepůsobí přímo na průměrný výsledek hodnocení žáků. Abychom vypočítali odhady čistých efektů faktorů, musíme nejdříve odečíst celkový průměr od průměrů pro úrovně jednotlivých faktorů. Celkový průměr také odečteme od průměrů uvnitř tabulky, abychom získali reziduální hodnoty, které popisují celkové působení faktorů a jejich kombinací relativně vzhledem k celkovému průměru. Získáme tak tabulku 9.11.

Tab. 9.10 Tabulka průměrů modelových dat pro analýzu a interpretaci interakcí mezi faktory

	Chlapci	Dívky	Celkově
Metoda 1	50	40	45
Metoda 2	30	60	45
Celkově	40	50	$\bar{x} = 45$

Tab. 9.11 Data z tabulky 9.10 upravená odečtením průměru

	Chlapci	Dívky	Řádkový efekt
Metoda 1	5	−5	0
Metoda 2	−15	15	0
Sloupcový efekt	−5	5	$\bar{x} = 45$

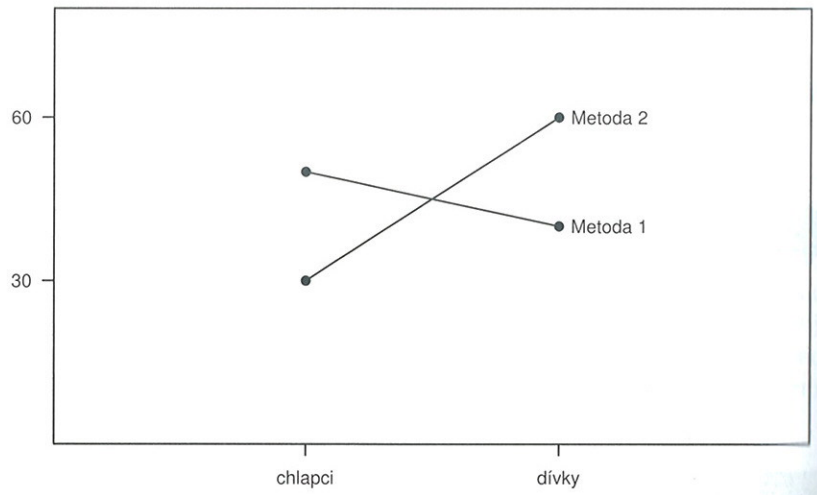
Tab. 9.12 Data z tabulky 9.11 upravená odečtením odhadů faktorů

	Chlapci	Dívky	Řádkový efekt
Metoda 1	10	−10	0
Metoda 2	−10	10	0
Sloupcový efekt	−5	5	$\bar{x} = 45$

Kdyby neexistovaly interakční efekty, pak bychom tabulku čistých efektů mohli rekonstruovat pomocí součtů příslušných řádkových a sloupcových efektů. Například pro políčko [1; 1] = 5 spočítáme 0+(−5) = −5. Obě hodnoty jsou však různé. Tabulku vhodnou pro analýzu interakcí spočítáme, když odhady efektů faktorů odečteme od hodnot uvnitř tabulky – např. políčko [1; 1] = 5−(−5) = 10. Těmito výpočty získáme tabulku 9.12. Vidíme, že interakce svým vlivem značně převyšují čistý vliv jednotlivých faktorů.

Rozklad hodnot původních průměrů do jednotlivých komponent ukazuje, jak se skládá průměr ve skupině z hlavních a interakčních efektů. Ilustruje se tak dobře základní model analýzy dvojného třídění daný rovnicí $x_{ijk} = \mu+\alpha_i+\beta_j+\gamma_{ij}$. Například pro průměr ve skupině dívek s intervencí 1 (metoda výuky 1) má rozklad podobu 40 = 45+5+0−10. Působení interakcí se na grafu průměrů projeví typickým křížem (obr. 9.1). Pokud by v datech žádné interakce neexistovaly, tedy

Obr. 9.1 Působení interakce v tabulce průměrů při dvojném třídění 2 × 2



Tab. 9.13 Tvar tabulky průměrů v případě stejných čistých efektů obou faktorů

	Chlapci	Dívky	Celkově
Metoda 1	40	50	45
Metoda 2	40	50	45
Celkově	40	50	$\bar{x}$ = 45

$\gamma_{ij} = 0$, graf by obsahoval rovnoběžné úsečky. Při stejných velikostech čistých efektů faktorů POHLAVÍ a METODA bychom dostali tabulku 9.13.

Poznamenáváme, že zjištění interakcí nemá vést k závěru, že hlavní efekty faktorů není nutné uvažovat. Doporučujeme, aby výzkumník v rámci analýzy rozptylu analyzoval data komplexním způsobem (také pomocí grafického znázornění), aby tak porozuměl všem závislostem, které mohou nastat. Analýza má zahrnout interpretaci odhadu hlavních efektů a efektů způsobených interakcemi.

9.3 Analýza rozptylu s opakováním měření

Ve výzkumu často potřebujeme rozhodnout, zda je věcně a statisticky významný efekt nějaké intervence (medicínské, tréninkové, zátěžové), přičemž pokus je uspořádán tak, že jednotky pozorování měříme opakovaně v předem daných časových okamžicích. Podobná situace nastane, jestliže měříme změnu názoru vybrané skupiny učitelů na strategii problémového učení během školního roku, nebo jestliže posuzuje každého jedince více posuzovatelů. Celý problém probereme podrobněji, protože se poměrně často vyskytuje. Popisované schéma se také použije pro blokový plán experimentu (s. 64).

Máme-li pouze dva časové okamžiky měření (obvykle před a po intervenci), hodnotíme data pomocí analýzy kovariance, pomocí párového testu t-testu nebo některé verze neparametrického testu (např. znaménkovým testem). Pokud data získáváme ve více časových okamžicích a chceme se přesvědčit, že se sledovaná proměnna mění, je vhodné výsledky hodnotit pomocí některé varianty analýzy rozptylu s opakováním měření. Tento přístup se také používá pro plány výzkumu, kdy kromě faktoru, jenž určuje opakování měření u jednotlivých objektů (např. čas), existují navíc faktory, které rozdělují objekty do několika nezávislých skupin. Například opakujeme stejným způsobem měření u kontrolní a experimentální skupiny jedinců.

Nejdříve si uvedeme jednoduchý modelový příklad dat, jež zkoumáme pomocí analýzy rozptylu s opakováním měření bez rozdělení do skupin.

PŘÍKLAD 9.4

Analýza rozptylu s opakováním měření

Fyziolog provádí měření VO_2 na pěti jedincích při zátěži na „běhacím koberci" pomocí zátěžového testu GXT. Po 3 minutách chůze pro rozehřátí se zátěž zvyšuje každé 2 minuty

Tab. 9.14 Příklad analýzy rozptylu s opakováním měření – data pro pět jedinců změřená při zátěžovém testu

Jedinec	2 minuty	4 minuty	6 minut	8 minut	10 minut	Průměry
1	21	19	27	31	35	26,6
2	18	17	21	28	31	23
3	19	25	29	36	41	30
4	12	17	21	27	32	21,8
5	15	20	27	31	38	26,2
Průměry	16,8	19,6	25	30,6	35,4	25,48

až do 10 minut. Provede se 5 opakovaných měření VO_2 na konci každého dvouminutového intervalu. Získají se data, která zobrazuje tabulka 9.14, v níž jsou uvedeny také řádkové a sloupcové průměry.

Průměr z dat vypočtený pro řádek i, resp. sloupec j je označen symbolem $\bar{x}_{i.}$, resp. $\bar{x}_{.j}$. Obecné schéma zápisu dat je v tabulce 9.15. Předpokládáme, že měření vyhovují modelu

$$x_{ij} = \mu + \alpha_j + \beta_i + \gamma_{ij} + e_{ij},$$

kde μ je společná část průměru závislé proměnné, α_j je efekt ošetření j ($j = 1, 2, \ldots, s$), β_i je efekt osoby i ($i = 1, 2, \ldots, r$), γ_{ij} je interakce mezi ošetřením j a osobou i. O náhodné chybě předpokládáme, že má nulovou střední hodnotu, stejný rozptyl a stejnou kovarianci pro měření dané osoby. Měření jednotlivých osob (mezi bloky) jsou na sobě nezávislá. V tomto modelu se především zajímáme o efekty ošetření. Zkoumané hypotézy mají tvar:

H_{01}: Všechny řádkové průměry jsou stejné (všechny $\alpha_i = 0$).

H_{11}: Ne všechny řádkové průměry jsou stejné.

Podobně jako v jednoduché analýze rozptylu počítáme součty čtverců odchylek a průměrné součty čtverců upravené pro tuto situaci (tab. 9.16). Použité součty se počítají podle následujících vzorců:

$$S_A = \sum_j \sum_i \left(\bar{x}_{.j} - \bar{x}\right)^2 \quad S_B = \sum_j \sum_i (\bar{x}_{i.} - \bar{x})^2 \quad S_T = \sum_j \sum_i \left(x_{ij} - \bar{x}\right)^2$$

Tab. 9.15 Schéma dat pro analýzu rozptylu jedné skupiny s opakováním měření

Jedinci	Ošetření				Průměry
	1	2	⋯	s	
1	x_{11}	x_{12}	⋯	x_{1s}	$\bar{x}_{1.}$
2	x_{21}	x_{22}	⋯	x_{2s}	$\bar{x}_{2.}$
⋮	⋮	⋮	⋱	⋮	⋮
r	x_{r1}	x_{r2}	⋯	x_{rs}	$\bar{x}_{r.}$
Průměry	$\bar{x}_{.1}$	$\bar{x}_{.2}$	⋯	$\bar{x}_{.s}$	$\bar{x}$

Tab. 9.16 Tabulka analýzy rozptylu s opakováním měření

Zdroj variability	S	$st.v.$	MS	F
ošetření (A)	S_A	$s - 1$	$MS_A = \frac{S_A}{s-1}$	$F = \frac{MS_A}{MS_e}$
jedinci (B)	S_B	$r - 1$	$MS_B = \frac{S_B}{r-1}$	$F = \frac{MS_B}{MS_e}$
reziduální	$S_e = S_T - S_A - S_B$	$(r-1)(s-1)$	$MS_e = \frac{S_e}{(r-1)(s-1)}$	
Celková variabilita	S_T	$rs - 1$		

Řešení příkladu 9.4 (s. 357)

Pro náš příklad z tabulky 9.14 (s. 357) získáme hodnoty uvedené v tabulce 9.17. Kritická F-hodnota pro hladinu významnost 0,05 a pro stupně volnosti (4; 16) je 3,01. Protože F-statistika v modelovém příkladu (98,517) je větší než kritická hodnota, přikloníme se k alternativní hypotéze, že došlo ke změně průměrů.

Tab. 9.17 Příklad tabulky analýzy rozptylu pro data ze zátěžového testu

Zdroj variability	S	$st.v.$	MS	F
ošetření (A)	220,640	4	281,760	98,517 $st.v.$ = (4; 16)
jedinci (B)	1127,040	4	281,760	
reziduální	45,760	16	2,860	
Celková variabilita	1393,440	24	58,060	

Jedním z předpokladů této statistické procedury je shoda rozptylů a shoda kovariancí v kovarianční matici, která je vytvořena z tabulky dat pro proměnné určené časovým okamžikem měření (tab. 9.18). Tento předpoklad se nazývá **podmínka sféricity** kovarianční matice. Lepší statistické programy nabízejí provedení testu tohoto předpokladu. Pokud není splněn, roste velikost chyby I. druhu. Proto se v takovém případě upravují stupně volnosti pro použitý F-test pomocí Greenhouseovy-Geisserovy korekce, aby bylo dosaženo deklarované hladiny významnosti.

Tab. 9.18 Příklad matice kovariancí mezi výsledky měření v různých okamžicích zátěžového testu

	Čas 1	Čas 2	Čas 3	Čas 4	Čas 5
Čas 1	10				
Čas 2	3,6	8,64			
Čas 3	5,6	8,4	11,2		
Čas 4	5,4	9,04	9,6	9,84	
Čas 5	3,4	10,36	11,6	10,96	13,84

Alternativní postup k analýze opakovaných měření se opírá o vícerozměrnou analýzu rozptylu (MANOVA), v níž není sféricita problémem. Tato metoda ale vyžaduje větší počet pozorování a má menší sílu než jednorozměrná analýza rozptylu pro tento případ.

9.3.1 Friedmanův test

Data získaná opakovaným měřením skupiny jedinců lze také analyzovat pomocí neparametrického Friedmanova testu, který vychází z hodnocení pořadových hodnot měření. Na rozdíl od testů v analýze rozptylu se nepředpokládá, že měření mají normální rozdělení. Zkoumáme nulovou hypotézu, že mediány souboru opakovaných měření jsou stejné:

$$\mathrm{H}_0: \tilde{\mu}_1 = \tilde{\mu}_2 = \cdots = \tilde{\mu}_m$$

$$\mathrm{H}_1: \text{Alespoň pro jednu dvojici } i, j \text{ platí, že } \tilde{\mu}_i \neq \tilde{\mu}_j\,.$$

Pro testování této hypotézy použijeme pořadové hodnoty, přičemž pořadí určujeme odděleně pro každý měřený objekt (pro každý řádek pozorování). Testovací statistika má tvar

$$F_r = \left[\frac{12}{nm\,(m+1)} \sum_i (SR_i)^2\right] - 3n\,(m+1)\,,$$

kde součet se provádí přes všechny časové okamžiky. Parametr m je počet opakování a n je počet měřených objektů. Součty SR_i označují součet pořadových hodnot pro časový okamžik i. Statistika F_r má za platnosti nulové hypotézy přibližně χ^2 rozdělení o $m-1$ stupních volnosti.

Tab. 9.19 Příklad analýzy rozptylu s opakováním měření – data pro pět jedinců změřená při zátěžovém testu a přiřazená pořadí v závorkách

Jedinec	2 minuty	4 minuty	6 minut	8 minut	10 minut
1	20 (2)	19 (1)	27 (3)	31 (4)	35 (5)
2	18 (2)	17 (1)	21 (3)	28 (4)	31 (5)
3	19 (1)	25 (2)	29 (3)	36 (4)	41 (5)
4	12 (1)	17 (2)	21 (3)	27 (4)	32 (5)
5	15 (1)	20 (2)	27 (3)	31 (4)	38 (5)
Součet pořadí	7	8	15	20	25

Při větším počtu shod uvnitř bloku (řádku) je nutné provést korekci testovací statistiky tím, že ji násobíme korekčním členem B

$$B = n/\left[n - \frac{1}{m^3 - m} \sum_{i=1}^{n} \sum_{j=1}^{r_i} \left(t_{ij}^3 - t_{ij}\right)\right],$$

kde r_i je počet shodných skupin uvnitř řádku i a t_{ij} jsou postupně počty shodných údajů v i-tém řádku (blok).

Užijeme příklad 9.4 o zátěžovém testu z předchozího odstavce. Pro naši tabulku 9.19 mají součty SR_i hodnotu 7, 8, 15, 20 a 25. Dosadíme jejich mocniny a ostatní parametry do vzorce a dostaneme

$$F_r = \left[\frac{12}{25 \times 6}(49 + 64 + 225 + 400 + 625)\right] - 15 \times 6 = 19{,}4.$$

Pro 5 stupňů volnosti na hladině významnost 0,05 je kritická hodnota χ^2 rozdělení 11,07, což vede k závěru, že Friedmanův test indikuje pro naše data zamítnutí shody mediánů měření v jednotlivých časových okamžicích.

Výhoda Friedmanova testu spočívá v tom, že jeho provedení je rychlé, test má neparametrický charakter, výsledek je také rezistentní. To jsou obecné vlastnosti neparametrických testů, pokud analyzujeme relativně malé výběry. Popsaný test lze také použít pro blokově uspořádaný experiment (s. 64).

Jestliže zamítneme nulovou hypotézu, můžeme se ptát, mezi kterými dvěma časovými okamžiky jsou změny mediánů statisticky významné. K tomu lze již použít jednoduchý Wilcoxonův test pro dva závislé výběry, ovšem s upravenou hladinou významnosti podle Bonferroniho $\alpha' = \alpha/k$, kde je k počet provedených srovnání.

V určitých situacích není vhodná obecná alternativní hypotéza, jež se zkoumá ve Friedmanově testu. Například při testování efektivity nějakého léku chce výzkumník vědět, zda u stejné skupiny jedinců při rostoucí dávce roste také účinek. Pro tuto situaci hledáme silnější test, který reaguje na popsané porušení nulové hypotézy. Uspořádanou alternativní hypotézu o mediánech podobně jako Jonckheere-Terpstra test uvažuje test podle Page. Zkoumají se jím hypotézy:

$$H_0: \tilde{\mu}_1 = \tilde{\mu}_2 = \cdots = \tilde{\mu}_m$$

$$H_1: \tilde{\mu}_1 \leq \tilde{\mu}_2 \leq \ldots \leq \tilde{\mu}_m \quad \text{a aspoň jedna nerovnost je ostrá.}$$

Testovací statistika se počítá podle formule

$$L = \sum_{j=1}^{m} jSR_j = SR_1 + 2SR_2 + \cdots + mSR_m,$$

kde SR_i jsou součty pořadí vypočítané stejným způsobem jako u Friedmanova testu. Nulová hypotéza se zamítne, pokud testovací statistika je větší než kritická mez, která závisí na m, k a α, kde k je počet jedinců. Asymptoticky platný test se opírá o z-statistiku

$$z = \frac{L - E(L)}{\sqrt{Var(L)}},$$

kde střední hodnota statistiky L za platnosti nulové hypotézy má průměr $E(L) = mk(1 + m)^2/4$ a rozptyl $Var(L) = k(m^3 - m)^2/[144(m - 1)]$.

Další aproximace a tabulky s přesnými kritickými hodnotami uvádějí Hollander a Wolfe (1999).

9.3.2 Vnitrotřídní koeficient korelace

Reliabilita měření znamená opakovatelnost měření neboli konzistenci/stabilitu opakovaných měření. Podrobně jsme se jí zabývali v kapitole 2.3.2. Často se pro vyhodnocení reliability používá Pearsonův koeficient korelace. Ten však není citlivý ke změně průměrů, což znamená, že neodhalí systematickou chybu. Korelační koeficient také může být aplikován pouze na měření provedená pouze ve dvou časových okamžicích. Jestliže chceme posoudit reliabilitu měření u více sérií dat, musíme použít odlišnou míru kvality. Jedním z přístupů pro zohlednění těchto námitek je použití korelačního koeficientu uvnitř tříd (intraclass correlation coeficient, *ICC*, *R*), který je citlivý ke změně průměru a dovoluje zařadit do hodnocení více řad měření stejné proměnné. Koeficient *ICC* je navržen tak, že nabývá hodnot v intervalu od 0 do 1, přičemž hodnoty blízko číslu 1 znamenají dobrou konzistenci měření. Používá se také pro hodnocení souhlasu hodnocení pozorovatelů daných objektů (inter-rater reliability). Jeho základní definice má tvar

$$ICC = \frac{\sigma^2(u)}{\sigma^2(m) + \sigma^2(u)},$$

kde $\sigma^2(m)$ znamená složku variability měření mezi jedinci a $\sigma^2(u)$ průměrnou intrasubjektovou variabilitu měření. Výraz $\sigma^2(m) + \sigma^2(u)$ znamená celkovou variabilitu měření. Obě složky variability málokdy známe, proto je musíme odhadovat pomocí dat. Tím se výpočty stávají méně přehledné. Ukážeme je pomocí příkladu. Nejdříve uvedeme základní data a výpočet analýzy rozptylu.

PŘÍKLAD 9.6

Vnitrotřídní koeficient korelace

Pokus se týkal problému, jenž byl popsán v příkladu 9.4. Analyzovala se data získaná od pěti jedinců, kteří však byli na rozdíl od předchozího případu měřeni pětkrát po dvou minutách *bez záběže*, aby se pro tento případ zjistila opakovatelnost měření (srovnej Vincent, 1995, s. 182). Data jsou uspořádána do tabulky 9.20. Metodou analýzy rozptylu s opakováním se získala tabulka 9.21. Výsledky indikují velké rozdíly průměrů mezi jedinci ($p < 0{,}01$) i statisticky významnou změnu v čase ($p < 0{,}05$), která však je na hranici významnosti.

Tab. 9.20 Příklad dat, u nichž analyzujeme stabilitu

	Čas				
	2. minuty	4. minuty	6. minut	8. minut	10. minut
1. subjekt	20	21	19	21	24
2. subjekt	17	18	19	20	19
3. subjekt	19	18	20	21	20
4. subjekt	12	13	13	14	12
5. subjekt	15	16	15	16	17

Tab. 9.21 Příklad vyplněné tabulky analýzy rozptylu s opakováním měření

Zdroj variability	*st.v.*	*S*	*MS*	*F*	*p*
ošetření (*A*)	4	12,960	3,240	3,042	0,048
jedinci (*B*)	4	214,160	53,54	53,34	0,005
reziduální	16	17,040	1,065		
Celková variabilita	24	244,160	10,173		

Neopravený vnitrotřídní koeficient korelace R_1 hodnotící přesnost průměru měření pro jedince se spočítá podle vzorce

$$R_1 = \frac{MS_B - MS_{O+e}}{MS_B},$$

přičemž

$$MS_{O+e} = \frac{S_A + S_e}{r(s-1)}$$

je průměrná čtvercová odchylka reprezentující změny průměrů v čase a zbytkovou chybu (r je počet osob a s je počet měření v čase). Pokud chceme zachytit konzistenci jednoho měření, použijeme výpočet:

$$R_1' = \frac{MS_B - MS_{O+e}}{MS_B + (s-1)MS_{O+e}}$$

V našem případě má R_1 hodnotu

$$R_1 = \frac{53{,}54 - 1{,}5}{53{,}54} = 0{,}972,$$

protože $MS_{O+e} = (17{,}04 + 12{,}96)/20 = 1{,}5$. Hodnota R_1 měří konzistenci průměru jako odhadu stacionární hodnoty jedince. Konzistenci jednoho měření, pokud uvažujeme i změnu průměru v čase, odhadneme hodnotou

$$R_1' = \frac{53{,}54 - 1{,}5}{53{,}54 + 4 \times 1{,}5} = 0{,}874.$$

Pokud jsou hodnoty tohoto koeficientu vyšší než 0,90, považuje se konzistence měření za velmi dobrou. V behaviorálních vědách se hodnoty R_1 mezi 0,70 a 0,90 považují ještě za přijatelné.

Jestliže chceme eliminovat variabilitu, která připadá na vrub změn průměrů v čase, lze provést opravu koeficientu R_1 podle Cronbacha, jenž navrhl vnitrotřídní koeficient R_2 počítat podle vzorce

$$R_2 = \frac{MS_B - MS_e}{MS_B}.$$

V našem příkladu, když dosadíme do tohoto vzorce odpovídající hodnoty, dostaneme

$$R_2 = \frac{53{,}54 - 1{,}065}{53{,}54} = 0{,}980.$$

Jestliže se průměr v čase moc nemění, koeficient R_2 se bude lišit od R_1 jen velmi málo. Opět se jedná o koeficient hodnotící nepřesnost odhadu průměru jedince.

Vnitrotřídní korelační koeficient pro hodnocení konzistence jednoho měření je dán vzorcem

$$R_2' = \frac{MS_B - MS_e}{MS_B + (s-1)MS_e}.$$

Konzistenci jednoho měření v našem příkladu – po eliminaci změn průměru v čase – odhadneme hodnotou

$$R_2' = \frac{53{,}54 - 1{,}065}{53{,}54 + 4 \times 1{,}065} = 0{,}909.$$

Doporučuje se používat koeficienty R_1 pro intervalová nebo poměrová data a koeficienty R_2 pro ordinální data, u nichž nemá smysl brát v úvahu změny v průměru. Jestliže data centrovaná průměry jedinců nevykazují velkou variabilitu a změny v čase jsou malé, koeficient R_1 bude mít vyšší hodnotu, než když centrovaná měření budou mít velkou variabilitu nebo změny v čase budou veliké.

Vnitrotřídní korelační koeficient je relativní míra opakovatelnosti. Nevýhodou těchto koeficientů je skutečnost, že stejná metoda může být posuzována jako spolehlivá nebo nespolehlivá v závislosti na variabilitě mezi jedinci $\sigma^2(m)$. Někdy je proto užitečnější posuzovat opakovatelnost nebo souhlas mezi posuzovateli pomocí absolutních měr souhlasu. Speciálně se tento přístup doporučuje, jestliže měříme na Likertově škále.

V literatuře se označuje koeficient R_2' rovněž symbolem $ICC(3,1)$ a koeficient R_1' symbolem $ICC(1,1)$, koeficienty R_2, resp. R_1 symboly $ICC(3,k)$, resp. $ICC(1,k)$, (ICC – intraclass correlation coefficient). Koeficienty typu $ICC(2,1)$ a $ICC(2,k)$ popisují Shrout a Fleiss (1979), kteří uvádějí celkem 6 různých typů intratřídních korelačních koeficientů a podrobně diskutují podmínky jejich uplatnění. Problém prezentují na reliabilitě hodnocení objektů pomocí hodnotitelů. Rozlišují tři typy studií:

1. hodnotitelé jsou pro každý objekt vybráni náhodně z větší skupiny hodnotitelů – $ICC(1)$;
2. stejní hodnotitelé hodnotí každý objekt; jsou náhodně vybráni z větší skupiny – $ICC(2)$;
3. stejní hodnotitelé hodnotí každý objekt; hodnotí se pouze tito hodnotitelé – $ICC(3)$.

Dále se pro každý případ uvažují zvlášť dvě možnosti: odhaduje se reliabilita jednoho hodnocení $ICC(*,1)$; odhaduje se reliabilita hodnocení tvořeného průměrem hodnocení k hodnotitelů $ICC(*,k)$.

9.3.3 Konkordance

Uvažujeme situaci, že n objektů je uspořádáno k způsoby. Máme posoudit, zda existuje vzájemná shoda uspořádání. Kendall navrhl pro tento účel **koeficient konkordance** W. S úlohou řazení se setkáváme často při hodnocení ve sportovních soutěžích. Uvedeme modelový příklad z této oblasti.

Tab. 9.22 Příklad dat, u nichž analyzujeme shodu uspořádání

Sportovec	Pořadí			Součet pořadí
	rozhodčí 1	rozhodčí 2	rozhodčí 3	
A	1	1	6	8
B	6	5	3	14
C	3	6	2	11
D	2	4	5	11
E	5	2	4	11
F	4	3	1	8
Součet				63

Tabulka 9.22 obsahuje pořadí sportovců navržená třemi rozhodčími. Obecně předpokládáme, že tabulka má k sloupců a je řazeno n objektů, každý má jedno z pořadí 1 až n. Vypočítáme součet pořadí SR_i v každém řádku i a hodnotu $S = \sum(SR_i)^2 - n(\overline{SR})^2$, kde $\overline{SR} = (n+1)k/2$ je průměr součtu pořadí v řádcích. Kendallův koeficient korelace W, který měří míru shody rozhodčích, se vypočte podle vzorce

$$W = \frac{S}{(1/12)\,k^2(n^3 - n)},$$

který nabývá hodnoty mezi 0 a 1 (hodnota 1 odpovídá dokonalé shodě – v každém řádku jsou stejná čísla).

PŘÍKLAD 9.7

Výpočet koeficientu konkordancí

Vypočítáme koeficient konkordance pro data z tabulky 9.22. Součet pořadí tabulka již obsahuje.

$$\overline{SR} = \frac{63}{6} = 10{,}5 = \frac{(n+1)\,k}{2}$$

$$S = \sum (SR_i)^2 - n\left(\overline{SR}\right)^2 =$$

$$(8)^2 + (14)^2 + (11)^2 + (11)^2 + (11)^2 + (8)^2 - 6(10{,}5)^2 = 25{,}5$$

$$W = \frac{S}{1/12\,k^2(n^3 - n)} = \frac{25{,}5}{(1/12)\,(3)^2(6^3 - 6)} = 0{,}162$$

Z dat plyne, že konkordance mezi hodnotiteli je nepříliš vysoká.

Jestliže chceme přezkoušet, zda se návrh pořadí liší od náhodné konfigurace pořadí (mezi rozhodčími nepanuje žádná shoda), což je v této souvislosti málo pravděpodobné, můžeme hodnotu $k(k-1)W$ srovnat s kritickou hodnotu χ^2-rozdělení o $2(n-1)$ stupních volnosti. Tento test je ekvivalentní aplikaci Friedmanova asymptotického testu rovnosti průměrných pořadí v jednotlivých řádcích tabulky. Platí vztah mezi W a Friedmanovou testovací statistikou $W = F_r/[(k(n-1)]$. Také existuje zajímavý vztah mezi koeficientem W a Spearmanovým koeficientem korelace r_s. Průměrná hodnota $\bar{r}_s$ těchto korelačních koeficientů mezi všemi různými páry řazení se vypočte pomocí W podle vzorce $\bar{r}_s = (kW - 1)/(k - 1)$.

9.4 Náhodný výběr, randomizace a analýza rozptylu

Než uzavřeme tuto kapitolu, upozorníme na jev, který má značnou důležitost pro pochopení některých souvislostí při používání statistiky. Ve značném procentu studií nemají použité výběry, jež se hodnotí analýzou rozptylu nebo pomocí testů pro dva nezávislé výběry, charakter náhodného výběru. Při hodnocení experimentu zpracováváme častěji, než si myslíme, data z výběrů, které byly získány na základě dostupnosti, ne náhodně. Pak je však standardní použití F-testu nebo t-testu značně pochybné, protože rozdělení těchto testovacích statistik bylo odvozeno pomocí předpokladu existence náhodného výběru z jednotlivých subpopulací. Hladiny kritické meze nebo dopočtené p-hodnoty tedy nemohou odpovídat správným hodnotám.

Lubrook a Dudley (1998) ukazují na základě rešerše 252 prospektivních a komparativních studií z medicínských časopisů, že pouze 4 % z nich dodržely zásady náhodného výběru z definované populace pro vytvoření experimentální a kontrolní skupiny. Navíc pracovaly všechny studie tvořící tato 4 % ne s lidmi, ale se zvířaty. Standardní učebnice statistiky se tomuto metodologickému problému dostatečně nevěnují. Ani náš text nepodává podrobný návrh, jakým způsobem taková data analyzovat pomocí inferenčních metod. Učiníme jenom několik poznámek (podrobněji Edgington, 1995).

Vhodný přístup vychází z požadavku, že při experimentu jsme použili alespoň randomizované přiřazení jedinců do skupin. Randomizace umožňuje aplikovat tzv. permutační testy (Rao, 1978, s. 546), které jsou vhodné pro data z takového plánu výzkumu. R. A. Fisher, autor metod analýzy rozptylu, považoval znáhodněné přiřazení do skupin za základ experimentální práce a statistické inference. Byl to také on, kdo se zabýval jako první permutačními testy. Později však Fisher ztratil zájem o tento způsob statistického usuzování pro jeho značnou výpočetní náročnost a obrátil svoji pozornost k odvození F-rozdělení testovací statistiky na principu náhodného výběru. Pomocí hypotetického příkladu ukážeme princip

permutačního testu a zároveň výpočetní obtíže, se kterými se Fisher ve své době nemohl vyrovnat.

■ ■ ■

Dvanáct mužů bylo získáno k pokusu zjišťujícímu, zda strava bohatá na ryby snižuje koncentraci cholesterolu v krvi. Tato skupina byla získána ne jako náhodný výběr, nýbrž na základě dostupnosti. Sedm z nich bylo náhodně přiřazeno do skupiny s „rybí" dietou a ostatní měli dietu převážně s červeným masem. Na konci roku byl u celé skupiny změřen cholesterol a získaly se následující výsledky:

- „Rybí" dieta: 5,42; 5,86; 6,16; 6,55; 6,80; 7,00; 7,11
- „Masová" dieta: 6,51; 7,56; 7,61; 7,84; 11,50

Jak se provádí permutační test? Vycházíme z toho, že pokud neexistuje rozdíl mezi působením obou diet, můžeme získat se stejnou pravděpodobností konfiguraci čísel, v níž některá čísla, která byla v první („rybí") skupině, se ocitají ve druhé „masové" skupině, a naopak. Jedné takto vytvořené konfiguraci dat se říká permutace. Počet čísel ve skupinách musí zůstat zachován. Ukážeme dvě takové permutace jako příklad. Vyměněná čísla označujeme *:

1. „Rybí" dieta: 5,42; 6,16; 6,55; 6,80; 7,00; 7,11; 7,56*
   „Masová" dieta: 5,86*; 6,51; 7,61; 7,84; 11,50
2. „Rybí" dieta: 5,42; 5,86; 6,16; 6,55; 7,61*; 7,84*; 11,50*
   „Masová" dieta: 6,51; 6,80*; 7,00*; 7,11*; 7,56
3. ...

Takto generovaných skupin dat (permutací) lze vytvořit 792.

Při provádění permutačního testu se postupuje následovně:

1. Vybere se testovací statistika S, která je citlivá k odchylce od nulové hypotézy, o niž jde (např. rozdíl průměrů nebo t-statika pro dva nezávislé výběry).
2. Vypočítá se S' pro původní množinu dat (t-statistika = 1,79 pro náš příklad).
3. Rozdělení statistiky S se získá provedením všech možných permutací, tedy výměn hodnot mezi oběma množinami dat („rybí" a „masovou"). Všechny hodnoty S určují referenční rozdělení pro posouzení hodnoty S'.
4. Spočítá se pro S', jakou dosahuje p-hodnotu vzhledem k tato získané distribuci.
5. Podle této p-hodnoty se zamítne nebo ponechá v platnosti nulová hypotéza.

Potíž je v tom, že abychom nalezli S pro všechny konfigurace dat, musíme záměnu provést v našem příkladu 792krát. Při rostoucím rozsahu výběru toto číslo roste

velmi rychle k ohromným hodnotám. Pro náš příklad Lodbrook a Dudley získali pro statistiku S počítanou jako běžná t-statistika p-hodnotu 7/792 = 0,088. Číslo 7 udává, kolik vypočtených t-hodnot z celkem 792 výpočtů bylo větších než číslo 1,79 (t-statistika pro základní konfiguraci hodnot). To indikuje, že vzhledem k distribuční funkci získané permutacemi je hodnota 1,79 extrémní hodnotou. Proto po tomto experimentu mohou výzkumníci tvrdit, že „rybí" dieta měla statisticky významný efekt při snižování koncentrace cholesterolu ve srovnání s „masovou" dietou (na hladině významnosti permutačního testu 0,088). Tento závěr však platí pouze pro populaci, ke které měl výzkumný tým přístup. Ještě přesněji: platí pouze pro populaci čísel, pro něž se test prováděl. Někdy říkáme, že tímto postupem provádíme lokálně platnou inferenci.

■ ■ ■

Permutační testy a randomizační testy (kdy se permutace provádějí náhodně do určitého počtu, aby se omezila výpočetní náročnost) jsou neparametrické testy, jež nevyžadují žádné předpoklady o populacích ani o náhodnosti výběru. Volnost, s kterou tyto testy můžeme provádět, je užitečná i v behaviorálních vědách, kde získání náhodného výběru pro experimenty představuje podobné obtíže jako v medicínském výzkumu.

Poznamenejme, že neparametrické testy založené na znaménku nebo pořadí jsou v podstatě permutační testy, a tudíž nevyžadují náhodný výběr. Permutační testy s původními hodnotami měření mají tu výhodu, že je můžeme aplikovat bez nutnosti provádět transformaci dat na pořadí nebo znaménka.

Ačkoli pro nenáhodné výběry nemohou být standardní F-testy v analýze rozptylu přesnými testy, bylo naštěstí ukázáno, že s rostoucí velikostí výběrů jsou používané kritické meze založené na F-rozdělení nebo t-rozdělení nebo standardně vypočítané p-hodnoty asymptoticky platné i pro takto provedené testy, pokud jsem použili randomizované přiřazení do skupin. Ani tato okolnost nás však nezbavuje nutnosti posoudit externí validitu závěru.

Souhrn

Popsali jsme základní modely analýzy rozptylu. Pomocí těchto modelů zkoumáme závislost spojitě rozdělené proměnné na ovlivňujících faktorech – kategoriálních nezávislých proměnných. Jednoduché třídění reprezentuje úlohu s jedním faktorem. Srovnání dvou skupin v analýze rozptylu je shodné s použitím t-testu průměrů ve dvou skupinách. F-test v analýze rozptylu testuje globální hypotézu, že ve všech skupinách jsou průměry stejné. Pokud tento test indikuje rozdílnosti, je nutné se zabývat vyhledáním zdrojů poruchy nulové hypotézy.

Jednoduchá analýza rozptylu předpokládá nezávislé náhodné výběry z populací, normální rozdělení závisle proměnné a stejnou směrodatnou odchylku rozdělení. V praxi je analýza rozptylu relativně robustní k poruše předpokladu normality, zvláště pokud jsou výběry veliké. Přesto je nutné před použitím statistických testů analýzy rozptylu zkoumat rozdělení dat a vyhledávat zdroje zešikmenosti rozdělení dat a odlehlé hodnoty.

Analýza rozptylu dvojného třídění uvažuje dva faktory jako dvě nezávislé kategoriální proměnné. Také jsme se zabývali jednoduchým plánem s opakováním měření. Probrali jsme jenom několik příkladů z mnoha. K celé problematice se vrátíme v následující kapitole, protože obecný model mnohonásobné lineární regrese v sobě zahrnuje všechny modely analýzy rozptylu. Výhoda spočívá v tom, že pro všechny tyto modely závislostí použijeme jednotnou statistickou metodologii.

Pro základní situace jsme uvedli také neparametrické přístupy k testování hypotéz. Testem Kruskala a Wallise hodnotíme homogenitu mediánů. Jestliže máme uspořádanou alternativní hypotézu, použijeme test Jonckheere a Terpstra. V případě jednoduchého pokusu s opakováním měření použijeme test podle Friedmana. Pro uspořádanou alternativu bychom měli použít v tomto případě test podle Pagea.

Podrobnosti o analýze rozptylu obsahují práce Havránek (1993), Reif (2000), Zvára (1997).

10 Mnohonásobná lineární regrese

V kapitole 7 jsme se zabývali situací, v níž bylo zapotřebí analyzovat vztah mezi jednou závisle a jednou nezávisle proměnnou. V této kapitole rozšíříme tento přístup, abychom mohli analyzovat komplexnější data s více nezávislými proměnnými. Popisné prostředky, které jsme poznali – bodové dvojrozměrné grafy, grafy reziduálních hodnot, korelace a regresní koeficienty, metoda nejmenších čtverců – se používají i v tomto případě. Jsou předpokladem pro získání hodnověrného modelu dat a oprávněnou aplikaci statistických testů významnosti a intervalů spolehlivosti.

Příklady otázek, které mohou být zodpovězeny pomocí mnohonásobné regrese:

- Jak závisí hodnota výsledku znalostního testu na demografických charakteristikách žáka?
- Jak závisí úspěšnost hokejového mužstva na průměrném BMI indexu, přesnosti přihrávek a úspěšnosti střelby z trestného nájezdu?
- Jak závisí celková osmolalita roztoku na koncentracích různých osmoticky aktivních látek?

Mnohonásobná regrese je prostředkem pro zkoumání statistické závislosti pomocí modelu, jenž zahrnuje jednu závislou proměnnou a několik nezávislých proměnných. Data získáme tak, že u prvků výběru zjistíme hodnoty všech uvažovaných proměnných. Rozlišujeme tři druhy úloh, pro jejichž řešení je vhodné aplikovat mnohonásobnou regresní analýzu:

1. Chceme poznat efekt, který má na cílovou proměnnou Y souhrn změn ovlivňujících parametrů $X_1, X_2, \ldots, X_k$.
2. Chceme predikovat hodnotu závisle proměnné Y pro budoucí hodnoty proměnných $X_1, X_2, \ldots, X_k$.
3. V rámci explorační statistické analýzy chceme vyhledat statistické vztahy mezi závisle proměnnou a několika nezávisle proměnnými.