

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 „očistění“ 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 rozdílů 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:

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

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 velké. 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 rozdíků.

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 ($j = 1, 2, \dots, m$). 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 ($i = 1, 2, \dots, n_j$) 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 $H_0: \mu_1 = \mu_2 = \dots = \mu_m$, alternativa je H_1 : *Ne všechny μ_j jsou stejné*, nebo v jiném prepisu $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_m = 0$, s alternativou H_1 : *Ne všechny α_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

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

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:

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

$$H_1: \text{Ne všechny } \mu \text{ jsou stejné.}$$

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 číselník 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 měřeními 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 zkouáme, 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 rozdílů vhodným intervalem spolehlivosti a ten použít pro testování hypotézy, že mezi průměry není statisticky významný rozdíl. Interval spolehlivosti se počítá pomocí rozdílů 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 označme

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

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

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

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