MUNI
ECON

Analýza rozptylu

Povinná literatura: Mann (2016), Kapitola 12

Z čeho studovat desátou lekci?

Povinná literatura: Mann (2016), kap. 12

Příprava na zkoušku: Mann (2016), kap. 12

Proč se zabývat Analýzou rozptylu (ANOVA)?

- **Odhalování skrytých rozdílů:** ANOVA nám umožňuje zjistit, zda mezi více skupinami existují významné rozdíly, aniž bychom museli provádět opakované párové testy.
- **Praktické využití:**
  - Posuzování efektivity různých výukových metod.
  - Testování produktivity v pracovním prostředí.
  - Porovnávání efektů různých léků nebo terapií.
- **Jednoduchý princip:** Pomocí F-testu porovnáváme rozptyl mezi skupinami a uvnitř skupin.
- **Univerzálnost:** Používá se v mnoha oblastech – od vzdělávání, přes ekonomii až po medicínu.

Obsah

Vlastnosti *F* rozdělení

1. Rozdělení *F* je **spojité** a má kladnou šikmost (je **zešikmené doprava**).
2. Rozdělení *F* má **dva stupně volnosti**: **df** pro čitatel a **df** pro jmenovatel.

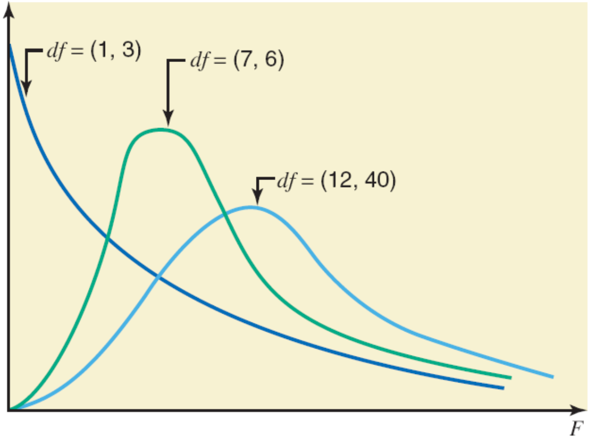
$$df = (8,\ 14)$$

First number denotes the *df* for the numerator

Second number denotes the *df* for the denominator

3. Hodnoty rozdělení *F*, označované jako **F**, jsou **vždy nezáporné**.

Tvar tří křivek *F* rozdělení



Příklad 1

Zadání: Najděte hodnotu *F* pro 8 stupňů volnosti v čitateli, 14 stupňů volnosti ve jmenovateli a pravděpodobnost 0.05 v pravém konci F rozdělení.

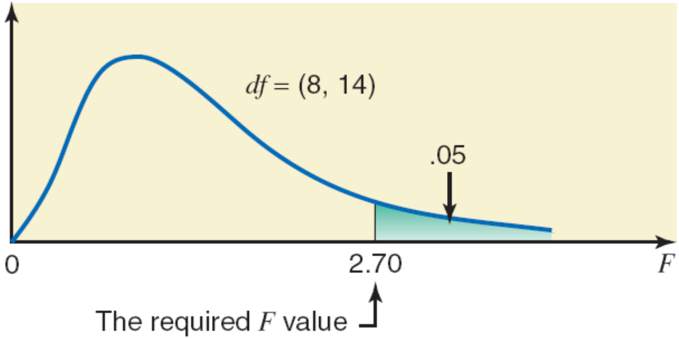
Řešení:

Degrees of Freedom for the Denominator		Degrees of Freedom for the Numerator					
		1	**2**	**. . .**	**8**	**. . .**	**100**
	1	161.5	199.5	. . .	238.9	. . .	253.0
	2	18.51	19.00	. . .	19.37	. . .	19.49
	.	. . .	. . .	. . .	. . .	. . .	. . .
	.	. . .	. . .	. . .	. . .	. . .	. . .
	14	4.60	3.74	. . .	**2.70**	. . .	2.19
	.	. . .	. . .	. . .	. . .	. . .	. . .
	.	. . .	. . .	. . .	. . .	. . .	. . .
	100	3.94	3.09	. . .	2.03	. . .	1.39

The *F* value for 8 *df* for the numerator, 14 *df* for the denominator, and .05 area in the right tail

Příklad 1: Řešení

Řešení graficky:



Obsah

Jednofaktorová ANOVA

ANOVA (analýza rozptylu) je statistická metoda používaná k testování nulové hypotézy, že střední hodnoty tří nebo více populací jsou si rovny.

Předpoklady pro jednofaktorovou ANOVA

1. Populace, ze kterých jsou vzorky vybírány, mají přibližně normální rozdělení.
2. Populace, ze kterých jsou vzorky vybírány, mají stejný rozptyl (nebo směrodatnou odchylku).
3. Vzorky odebrané z různých populací jsou náhodné a nezávislé.

Testovací statistika F pro jednofaktorovou ANOVA

Hodnota **testovací statistiky** F pro jednofaktorovou ANOVA se vypočítá jako:

$$F = \frac{\text{Rozptyl mezi vzorky}}{\text{Rozptyl uvnitř vzorků}} \quad \text{nebo} \quad F = \frac{\text{MSB}}{\text{MSW}}$$

Poznámka: Výpočet hodnot MSB (Mean Square Between) a MSW (Mean Square Within) je vysvětlen v Příkladu 2.

Příklad 2

Zadání: Patnáct žáků čtvrté třídy bylo náhodně rozděleno do tří skupin, aby vyzkoušeli tři různé metody výuky aritmetiky. Na konci semestru všichni studenti absolvovali stejný test. Tabulka uvádí výsledky testů studentů ve třech skupinách.

Method I	Method II	Method III
48	55	84
73	85	68
51	70	95
65	69	74
87	90	67

Vypočítejte hodnotu testovací statistiky F. Předpokládejte, že všechny požadované předpoklady uvedené na slidu 10 jsou splněny.

Příklad 2: Řešení

Označme

x = skóre studenta

k = počet různých vzorků (nebo treatmentů)

n_i = velikost i-tého vzorku

T_i = součet hodnot v i-tém vzorku

n = počet hodnot ve všech vzorcích

$$n = n_1 + n_2 + n_3 + \dots$$

Σx = součet všech hodnot ve všech vzorcích

$$\Sigma x = T_1 + T_2 + T_3 + \dots$$

Σx^2 = součet druhých mocnin všech hodnot ve všech vzorcích

Příklad 2: Řešení

Pro výpočet MSB a MSW nejprve vypočítáme:

- Součet čtverců mezi vzorky (between-samples sum of squares), označený jako SSB,
- Součet čtverců uvnitř vzorků (within-samples sum of squares), označený jako SSW.

Součet SSB a SSW se nazývá celkový součet čtverců (total sum of squares) a je označen jako SST, tedy:

$$\text{SST} = \text{SSB} + \text{SSW}$$

Hodnoty SSB a SSW se počítají pomocí následujících vzorců.

Příklad 2: Řešení

Součet čtverců mezi vzorky (between-samples sum of squares), označený jako **SSB**, se vypočítá jako:

$$\mathrm{SSB} = \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \dots\right) - \frac{(\Sigma x)^2}{n}$$

Součet čtverců uvnitř vzorků (within-samples sum of squares), označený jako **SSW**, se vypočítá jako:

$$\mathrm{SSW} = \Sigma x^2 - \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \dots\right)$$

Příklad 2: Řešení

Method I	Method II	Method III
48	55	84
73	85	68
51	70	95
65	69	74
87	90	67
$T_1 = 324$	$T_2 = 369$	$T_3 = 388$
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$

Výpočty dodatečných potřebných hodnot

$$\Sigma x = T_1 + T_2 + T_3 = 324 + 369 + 388 = 1081$$

$$n = n_1 + n_2 + n_3 = 5 + 5 + 5 = 15$$

$$\Sigma x^2 = (48)^2 + (73)^2 + (51)^2 + (65)^2 + (87)^2 + (55)^2 + (85)^2 + (70)^2 + (69)^2 + (90)^2 + (84)^2 + (68)^2 + (95)^2 + (74)^2 + (67)^2 = 80\,709$$

Příklad 2: Řešení

Výpočet součtů čtverců:

$$\mathrm{SSB} = \frac{(324)^2}{5} + \frac{(369)^2}{5} + \frac{(388)^2}{5} - \frac{(1081)^2}{15} = 432.1333$$

$$\mathrm{SSW} = 80\,709 - \left(\frac{(324)^2}{5} + \frac{(369)^2}{5} + \frac{(388)^2}{5}\right) = 2372.80$$

$$\mathrm{SST} = \mathrm{SSB} + \mathrm{SSW} = 432.1333 + 2372.8000 = 2804.9333$$

Příklad 2: Řešení

Výpočet MSB a MSW:

$$\text{MSB} = \frac{\text{SSB}}{k-1} \quad \text{a} \quad \text{MSW} = \frac{\text{SSW}}{n-k}$$

kde $k-1$ a $n-k$ představují stupně volnosti (df) pro čitatele a jmenovatele ve F-rozdělení. Pamatujte, že k je počet různých vzorků.

Konkrétní realizace a F statistika:

$$\text{MSB} = \frac{\text{SSB}}{k-1} = \frac{432.1333}{3-1} = 216.0667$$

$$\text{MSW} = \frac{\text{SSW}}{n-k} = \frac{2372.8000}{15-3} = 197.7333$$

$$F = \frac{\text{MSB}}{\text{MSW}} = \frac{216.0667}{197.7333} = 1.09$$

Příklad 2: Řešení

Vše lze zapsat do přehledné tabulky:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Value of the Test Statistic
Between	$k-1$	SSB	MSB	$F = \dfrac{\text{MSB}}{\text{MSW}}$
Within	$n-k$	SSW	MSW	
Total	$n-1$	SST		

A to i s konkrétními realizacemi:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Value of the Test Statistic
Between	2	432.1333	216.0667	$F = \dfrac{216.0667}{197.7333} = 1.09$
Within	12	2372.8000	197.7333	
Total	14	2804.9333		

Příklad 3

Zadání: Zvažte znovu Příklad 2 týkající se výsledků 15 žáků čtvrté třídy, kteří byli náhodně rozděleni do tří skupin za účelem experimentování se třemi různými metodami výuky aritmetiky.

Můžeme na hladině významnosti 1 % zamítnout nulovou hypotézu, že průměrné skóre z aritmetiky všech žáků čtvrté třídy, vyučovaných těmito třemi metodami, je stejné?

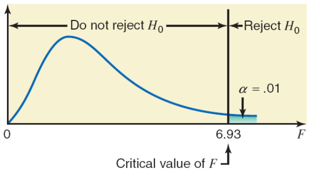
Předpokládejte, že všechny požadavky potřebné pro použití jednofaktorové ANOVA jsou splněny.

Příklad 3: Řešení

Krok 1 $H_0 : \mu_1 = \mu_2 = \mu_3$ (průměrné skóre tří skupin je stejné)
H_1 : Alespoň jeden průměr je rozdílný

Krok 2 Protože porovnáváme průměry tří normálně rozdělených populací a všechny předpoklady potřebné pro použití procedury ANOVA jsou splněny, použijeme F-rozdělení pro tento test.

Krok 3



Jednofaktorový ANOVA test je vždy jednostranný (pravostranný). Jelikož $\alpha = 0.01$, pak plocha v pravém chvostu je 0.01.

Stupně volnosti pro:
čitatel: df $= k - 1 = 3 - 1 = 2$
jmenovatel: df $= n - k = 15 - 3 = 12$

Kritická hodnota $F = 6.93$

Příklad 3: Řešení

Krok 4 Hodnota testové statistiky je vypočítána jako $F = 1.09$. Viz Excel.

Krok 5 Hodnota testové statistiky $F = 1.09$ je menší než kritická hodnota $F = 6.93$. Spadá do oblasti nezamítnutí. Proto nezamítáme nulovou hypotézu.

Vyvozujeme, že průměry tří populací jsou stejné. Jinými slovy, tři různé metody výuky aritmetiky nemají vliv na průměrné skóre studentů. Rozdíly v průměrných skórech našich tří vzorků mohou být způsobeny pouze chybou výběru vzorků.

Příklad 4

Zadání: Čas od času, aniž by o tom zaměstnanci věděli, výzkumné oddělení banky Post Bank sleduje různé zaměstnance kvůli jejich pracovní produktivitě. Nedávno chtělo toto oddělení zjistit, zda čtyři pokladní na jedné z poboček této banky obslouží v průměru stejný počet zákazníků za hodinu. Výzkumný manažer pozoroval každého ze čtyř pokladních po určitý počet hodin.

Tabulka na následujícím slidu uvádí počet zákazníků, které čtyři pokladní obsloužili během jednotlivých pozorovaných hodin.

Na hladině významnosti 5 % otestujte nulovou hypotézu, že průměrný počet zákazníků obsloužených za hodinu každým ze čtyř pokladních je stejný. Předpokládejte, že jsou splněny všechny požadavky potřebné pro použití jednofaktorové analýzy rozptylu (ANOVA).

Příklad 4

Teller A	Teller B	Teller C	Teller D
19	14	11	24
21	16	14	19
26	14	21	21
24	13	13	26
18	17	16	20
	13	18	

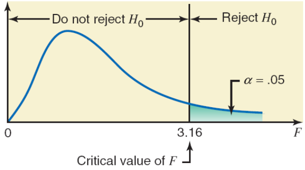
Příklad 4: Řešení

Krok 1 $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ (Průměrný počet zákazníků obsloužených za hodinu u každého ze čtyř pokladních je stejný.)

H_1 : Alespoň jeden průměr je rozdílný.

Krok 2 Protože porovnáváme průměry tří normálně rozdělených populací a všechny předpoklady potřebné pro použití procedury ANOVA jsou splněny, použijeme F-rozdělení pro tento test.

Krok 3



Jednofaktorový ANOVA test je vždy jednostranný (pravostranný). Jelikož $\alpha = 0.05$, pak plocha v pravém chvostu je 0.05.

Stupně volnosti pro:
čitatel: df $= k - 1 = 4 - 1 = 3$
jmenovatel: df $= n - k = 22 - 4 = 18$

Kritická hodnota $F = 3.16$

Příklad 4: Řešení

Krok 4 Hodnota testové statistiky je postupně vypočítána jako:

Teller A	Teller B	Teller C	Teller D
19	14	11	24
21	16	14	19
26	14	21	21
24	13	13	26
18	17	16	20
	13	18	
$T_1 = 108$	$T_2 = 87$	$T_3 = 93$	$T_4 = 110$
$n_1 = 5$	$n_2 = 6$	$n_3 = 6$	$n_4 = 5$

Příklad 4: Řešení

$$\Sigma x = T_1 + T_2 + T_3 + T_4 = 108 + 87 + 93 + 110 = 398$$
$$n = n_1 + n_2 + n_3 + n_4 = 5 + 6 + 6 + 5 = 22$$
$$\Sigma x^2 = (19)^2 + (21)^2 + \cdots + (26)^2 + (20)^2 = 7614$$

$$SSB = \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \frac{T_4^2}{n_4}\right) - \frac{(\Sigma x)^2}{n}$$
$$= \left(\frac{(108)^2}{5} + \frac{(87)^2}{6} + \frac{(93)^2}{6} + \frac{(110)^2}{5}\right) - \frac{(398)^2}{22} = 255.6182$$
$$SSW = \Sigma x^2 - \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \frac{T_4^2}{n_4}\right)$$
$$= 7614 - \left(\frac{(108)^2}{5} + \frac{(87)^2}{6} + \frac{(93)^2}{6} + \frac{(110)^2}{5}\right) = 158.2000$$

$$MSB = \frac{SSB}{k-1} = \frac{255.6182}{4-1} = 85.2061 \qquad MSW = \frac{SSW}{n-k} = \frac{158.2000}{22-4} = 8.7889$$
$$F = \frac{MSB}{MSW} = \frac{85.2061}{8.7889} = 9.69$$

Příklad 4: Řešení

Krok 4 Hodnota testové statistiky je vypočítána jako $F = 9.69$.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Value of the Test Statistic
Between	3	255.6182	85.2061	$F = \frac{85.2061}{8.7889} = 9.69$
Within	18	158.2000	8.7889	
Total	21	413.8182		

Příklad 4: Řešení

Krok 5 Hodnota testové statistiky $F = 9.69$ je menší než kritická hodnota $F = 3.16$. Spadá do oblasti zamítnutí. Proto zamítáme nulovou hypotézu.

Závěrem je, že průměrný počet zákazníků obsloužených za hodinu každým ze čtyř pokladníků není stejný.

Shrnutí přednášky:

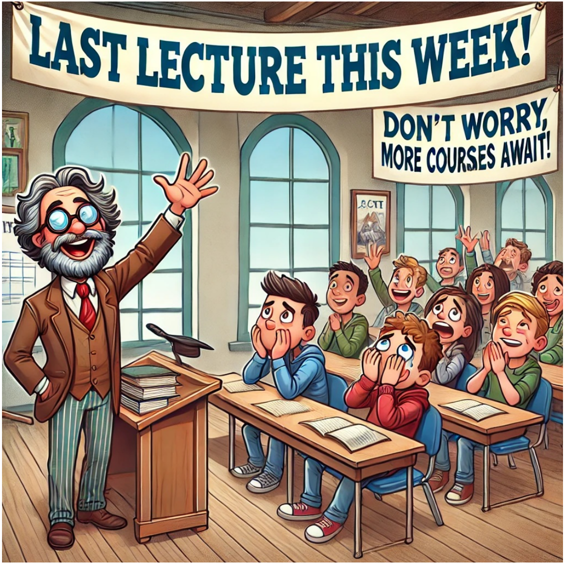
F rozdělení

Jednofaktorová ANOVA

Z čeho studovat na zkoušku?

Učebnice	Mann (2016), Kapitola 1-12
Přednášky	Mann (2016), Kapitola 1-12
Leaflety	Materiály ke cvičení 1-12
Koncepty a procedury	Materiály ke cvičení 2-12
Sbírky úloh	Materiály ke cvičení 2-12

Děkuji za pozornost!



Nabídka navazujících kurzů

Bakalářské kurzy

ZAEK	Základy ekonometrie
STAF	Statistics for finance
CARA	Časové řady

Magisterské kurzy

AVED	Analýza a vizualizace ekonomických dat
APIS	Aplikované identifikační strategie
AIIF	AI in Finance
APFE	Applied financial econometrics
BAAN	Bayesiánská analýza
EKON	Ekonometrie
VSM	Vícerozměrné statistické metody
...	a další :)