

Téma 2 – Příklad 1

Zadání příkladu

Sledujeme výzkum, který se zabývá vlivem 4 typů učebních metod angličtiny na výsledky studentů, posuzované podle výsledků dosažených v závěrečném testu (v bodech). Údaje jsou v následující tabulce. Data byla získána náhodným výběrem, rozsah souboru je 20 studentů. Předpokládáme, že výkon studentů má normální rozdělení. Pomocí analýzy rozptylu posuďte závislost proměnné y (výsledky v bodech) na metodě výuky (x). Posuďte sílu případné závislosti.

Typ učební metody (x_i) $i = 1,2,3,4$	Počet bodů v závěrečném testu (y_{ij}) $j = 1,2,3,4,5$	Četnost (n_i)	Podmíněné průměry $\bar{y}_i = \frac{\sum y_{ij}}{n_i}$	Podmíněné rozptyly $s_i^2 = \frac{\sum (y_{ij} - \bar{y}_i)^2}{n_i}$
1. typ	71 83 65 77 84	5	76	52
2. typ	60 51 54 80 55	5	60	108,4
3. typ	55 55 62 65 63	5	60	17,6
4. typ	68 73 67 59 53	5	64	50,4
Suma		20	X	X

Vypracování příkladu

Testujeme hypotézu o neúčinnosti faktoru x na proměnnou y . Postup je obdobný jako u jiného statistického testování hypotéz a probíhá v několika krocích.

1. $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ (obecně y na x nezávisí) – rozdělení proměnné y mají na různých úrovních faktoru x stejné střední hodnoty μ

$H_1: \text{non } H_0$

2. Testové kritérium

$$F = \frac{\frac{S_{ym}}{k-1}}{\frac{S_{yv}}{n-k}} = \frac{\frac{860}{3}}{\frac{1142}{16}} = 4,016$$

3. $W \equiv (F, F \geq F_{1-\alpha}(k-1; n-k))$

$W \equiv (F, F \geq F_{0,95}(3; 16))$

$W \equiv (F, F \geq 3,2391)$

4. $4,0160 \geq 3,2391$

Testové kritérium je prvkem kritického oboru. Nulovou hypotézu zamítáme, přijímáme hypotézu alternativní.

Na základě analýzy rozptylu tedy můžeme říci, že na 5% hladině významnosti **existuje statisticky významná závislost** výsledků studentů angličtiny (posuzovaných pomocí počtu bodů v závěrečném testu) na zvolené metodě výuky.

Sílu závislosti posoudíme pomocí poměru determinace. Jde o poměr meziskupinové variability na celkové variabilitě proměnné y . Výsledek je možné vyjádřit v procentech, jelikož jde o poměr části z celku, potenciálně tudíž vychází v intervalu od 0 do 1.

$$P^2 = \frac{S_{ym}}{S_y} = \frac{860}{2002} = 42,96\%$$

Řešení v SGP

Důležité je správné zadání dat do datového listu. Každá proměnná musí mít formu datového vektoru. Závislou proměnnou jsou body studentů z testu (*Test*), faktorem je typ učební metody (*Ucební_metoda*). Je třeba, aby závislá a nezávislá proměnná byly párově správně propojeny.

	Ucebni_metoda	Test
		body
1	1	71
2	1	83
3	1	65
4	1	77
5	1	84
6	2	60
7	2	51
8	2	54
9	2	80
10	2	55
11	3	55
12	3	55
13	3	62
14	3	65
15	3	63
16	4	68
17	4	73
18	4	67
19	4	59
20	4	53
21		
22		
23		

Procedura v SGP: Compare - Analysis of Variance - One-Way ANOVA (ANOVA Table, Variance check)

Při vstupním dialogu zadáme *Dependent Variable* jako proměnnou *Test*; *Factor* je *Ucebni_metoda*.

V programu Statgraphics můžeme ověřit jeden z předpokladů analýzy rozptylu – shodu skupinových rozptylů. Program nabízí čtyři různé testy (kromě nám známého Bartlettova testu jde o testy Leveneho, Cochran a Hartleye). Výsledek testu vidíme ze stručné tabulky.

Variance Check

	Test	P-Value
Bartlett's	2,73239	0,434747

Comparison	Sigma1	Sigma2	F-Ratio	P-Value
1 / 2	8,06226	11,6404	0,479705	0,4943
1 / 3	8,06226	4,69042	2,95455	0,3190
1 / 4	8,06226	7,93725	1,03175	0,9766
2 / 3	11,6404	4,69042	6,15909	0,1062
2 / 4	11,6404	7,93725	2,15079	0,4765
3 / 4	4,69042	7,93725	0,349206	0,3326

Položka *Test* zobrazuje hodnotu testové statistiky. Hodnota *P-Value* ukazuje maximální možnou hladinu významnosti, na které nezamítáme nulovou hypotézu. V tomto případě tedy na 5% hladině významnosti nezamítáme nulovou hypotézu o shodě skupinových rozptylů. Předpoklad byl naplněn.

Doplňková tabulka zobrazuje porovnání a testy pro jednotlivé dvojice rozptylů (směrodatných odchylek).

Hlavní výstup je zobrazen v okně ANOVA Table.

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	860,0	3	286,667	4,02	0,0262
Within groups	1142,0	16	71,375		
Total (Corr.)	2002,0	19			

Sloupec *Source* zobrazuje rozklad variability na část meziskupinovou (Between groups) a vnitroskupinovou (Within groups). Příslušné sumy čtverců jsou ve sloupci *Sum of Squares*. Hodnota testové statistiky je označena jako *F-Ratio*. *P-Value* v tomto případě nasvědčuje tomu, že na 5% hladině významnosti zamítáme nulovou hypotézu a přijímáme hypotézu alternativní. Výpočet poměru determinace je třeba udělat jako podíl meziskupinové variability na celkové variabilitě (SGP nemá zvláštní funkcionalitu pro tento ukazatel).

Stejně jako v ostatních případech používá Statgraphics k vyhodnocení testu hypotézy ukazatel *P-Value* (v jiných programech např. *Significance Level* apod.), což je maximální možná hodnota hladiny významnosti, na které ještě nezamítáme nulovou hypotézu. Není tudíž nutné určovat kritický obor pro námi zvolenou hladinu významnosti.

Interpretace

Na 5% hladině významnosti bylo prokázán statisticky významný vliv použité učební metody na výsledky studentů angličtiny posuzované výkonem v závěrečném testu. Výsledky studentů jsou vysvětleny faktorem učební metody asi z 42,96 % (42,96 % celkové variability proměnné *y* je vysvětleno pomocí faktoru *x*).

Řešení v MS Excel

Dle verze programu se liší i možnosti zpracování analýzy rozptylu. Nejblíže výstupu ze Statgraphicsu je použití nástroje *Analýza dat*. Po kliknutí na toto tlačítko lze vybrat ANOVA: jednofaktorová. V následujícím dialogu určíte pole se vstupními daty. Lze je mít uspořádána ve sloupcích, nebo řádcích. Pozor ovšem, že každá kategorie musí mít samostatný sloupec/řádek (tím se liší zadávání od SGP). Lze také definovat, jestli vstup obsahuje hlavičky kategorií a určit hladinu významnosti pro testování. Nezapomeňte také určit pole pro výstup procedury (případně nový list).

Výstup je obdobný ANOVA Table ze Statgraphicsu. Variabilita je rozložena na meziskupinovou a vnitroskupinovou. Kromě *P-Value* je zobrazen i kvantil kritického oboru pro zadanou hladinu významnosti. V tabulce *Summary* jsou podmíněné charakteristiky pro jednotlivé kategorie (obdoba *Subset Analysis* v SGP).