TECHNICKÁ UNIVERZITA V LIBERCI

Nové možnosti rozvoje vzdělávání na Technické univerzitě v Liberci

Specifický cíl A3: Tvorba nových profesně zaměřených studijních programů

NPO_TUL_MSMT-16598/2022

Téma 7: Příklad 1 – Analýza rozptylu

Ing. Vladimíra Hovorková Valentová, Ph.D.

Financováno Evropskou unií NextGenerationEU

Národní plán obnovy

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Zadání příkladu:

Zásilky ve firmě se balí ve 3 střediscích (v přibližně stejném počtu). Jsou evidovány vady zásilek za každé středisko. Posuďte na hladině významnosti 5 %, zda měsíční počet vad zásilek závisí na středisku, ve kterém byla zásilka balena. Případně změřte sílu dané závislosti pomocí vhodné charakteristiky. Předpokládáme, že měsíční počet vad se řídí normálním rozdělením. Bylo náhodně vybráno 12 měsíců za poslední 3 roky a zjištěny tyto údaje:

Středisko (x_i)	**Počet vad za měsíc (y_{ij})**	**n_i**
Středisko 1	6, 10, 10, 8	4
Středisko 2	12, 10, 9, 10	4
Středisko 3	18, 15, 18, 20	4
Celkem		**12**

Řešení příkladu:

Vzhledem k tomu, že počet vad za měsíc (závisle proměnná) je číselná proměnná (metrická) a středisko (nezávisle proměnná) je slovní proměnná (nominální), lze uvažovat o použití analýzy rozptylu pro řešení dané otázky, zda středisko ovlivňuje počet vad na balených zásilkách. Nejprve je ale potřeba ověřit vstupní podmínky, které použití analýzy rozptylu vyžaduje.

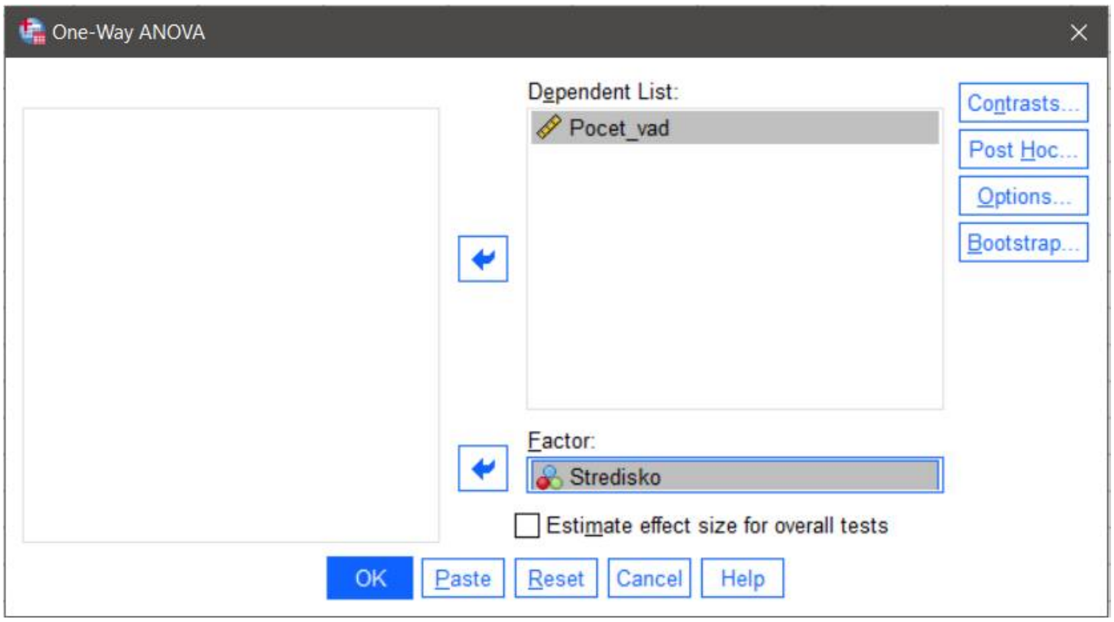
Ověření podmínek ANOVA.

1) Závisle proměnná je číselná a alespoň ordinální – splněno, závisle proměnná je metrická.
2) Výběr pochází z normálního rozdělení – splněno, uvedeno v zadání, že počet vad se řídí normálním rozdělením (jinak je možné ověřit např. pomocí Shapirova-Wilkova testu normality).
3) Výběry jsou nezávislé – mezi pozorováními není žádný vztah, nepracujeme s párovými hodnotami.
4) Počet pozorování je větší než počet skupin, tj. n>k – splněno, protože 12>3.
5) Homoskedasticita (shoda rozptylů ve všech skupinách) – ověříme pomocí Leveneova testu v programu SPSS:

Nejprve stanovíme hypotézy (viz níže u komentáře k výstupu z SPSS) a do programu SPSS zadáme data následujícím způsobem:

	Stredisko	Pocet_vad
1	1	6
2	1	10
3	1	10
4	1	8
5	2	12
6	2	10
7	2	9
8	2	10
9	3	18
10	3	15
11	3	18
12	3	20

Potom zvolíme tuto posloupnost procedur: Analyze – Compare Means – One-Way ANOVA. Vstupní pole procedury vyplníme následovně:



Kromě zadání obou proměnných je potřeba spustit tlačítkovou volbu *Options* (odshora třetí tlačítko vpravo), kde vybereme položku **Homogeneity of variance test**. Po potvrzení tlačítky Continue a následně OK, dostaneme výstup, ve kterém se nejprve zaměříme na Leveneův test homoskedasticity:

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Pocet_vad	Based on Mean	,444	2	9	,655
	Based on Median	,420	2	9	,669
	Based on Median and with adjusted df	,420	2	7,545	,672
	Based on trimmed mean	,445	2	9	,654

Výstup relevantní pro naše potřeby je podsvícen žlutě. Postup Leveneova testu pomocí SPSS:

$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$

$H_1: non\ H_0$

$F = 0{,}444$ (Leveneova testovací statistika)

Sig. (0,655) $> \alpha\ (0{,}05)$, proto nezamítáme H_0 a nepřijímáme H_1.

Na hladině významnosti 5 % nezamítáme předpoklad o tom, že rozptyly jsou ve všech skupinách stejné – předpoklad homoskedasticity je tedy splněn.

Nyní po ověření vstupních předpokladů pro použití analýzy rozptylu pro zkoumání dané závislosti můžeme provést **test nezávislosti**:

H_0: měsíční počet vad na zásilkách není závislý na středisku, kde byla zásilka balena

(NEBO $H_0: \mu_1 = \mu_2 = \mu_3$)

$H_1: non\ H_0$

Financováno Evropskou unií NextGenerationEU

Národní plán obnovy

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Relevantní pro tento test je následující výstup z SPSS:

ANOVA

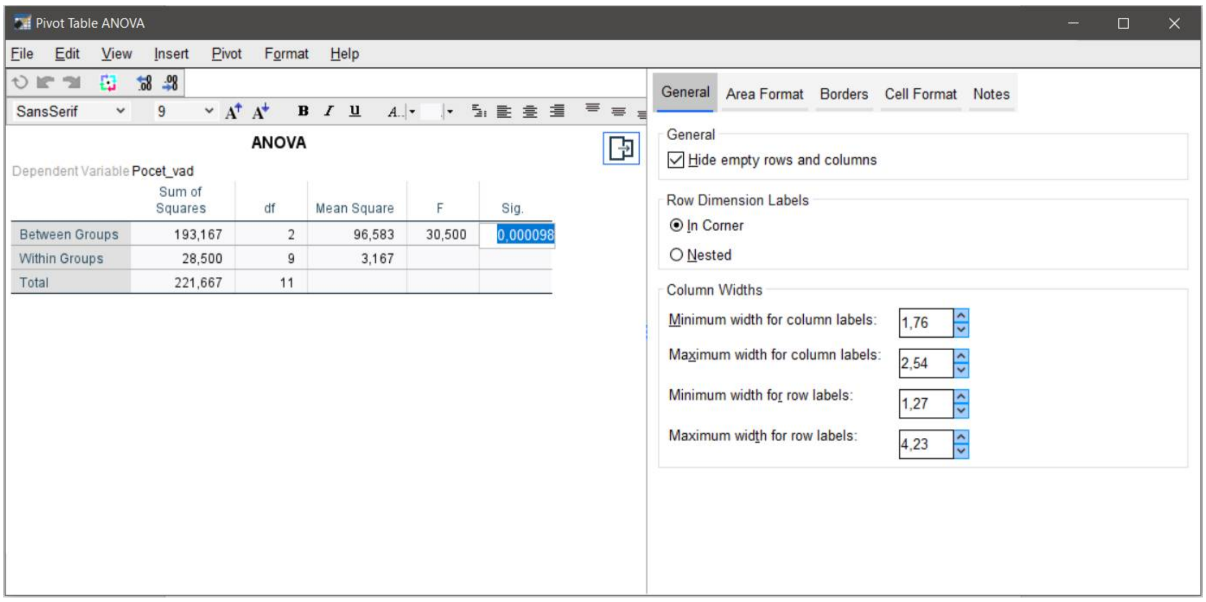
Pocet_vad

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	193,167	2	96,583	30,500	<,001
Within Groups	28,500	9	3,167		
Total	221,667	11			

Obsah tabulky tvoří jednotlivé položky, ze kterých vypočítáme hodnotu testového kritéria F (barevně přiřazeny hodnoty z tabulky hodnotám v testovém kritériu F), tj.:

$$F = \frac{\frac{Sym}{k-1}}{\frac{Syv}{n-k}} = \frac{\frac{193{,}167}{2}}{\frac{28{,}500}{9}} = \frac{96{,}583}{3{,}167} = 30{,}500$$

Výsledek testu rozhodneme podle vypočtené hladiny významnosti, zde označené jako Sig. Pokud program uvádí, že Sig. < 0,001, je vypočtená hladina významnosti velmi malá. Pokud bychom chtěli znát její skutečnou hodnotu, je potřeba na tabulku s výstupem 2x rychle kliknout a následně 2x kliknout na hodnotu Sig. v tabulce:



Financováno Evropskou unií NextGenerationEU
Národní plán obnovy
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Vidíme, že Sig. = 0,000098. To je méně, než hladina významnosti, na které testujeme (0,05).

Sig. $< \alpha$, proto zamítáme H_0 a přijímáme H_1.

Na hladině významnosti 5 % jsme prokázali, že měsíční počet vad na zásilkách je závislý na středisku, kde byla zásilka balena.

Pokud byla závislost sledované proměnné na faktoru x prokázána, změříme sílu této závislosti. Vhodnou charakteristikou je poměr determinace, kdy hodnoty pro jeho výpočet získáme v tabulce nazvané ANOVA v SPSS:

ANOVA

Pocet_vad

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	193,167	2	96,583	30,500	<,001
Within Groups	28,500	9	3,167		
Total	221,667	11			

$$P^2 = \frac{Sym}{Sy} = \frac{193{,}167}{221{,}667} = 0{,}871$$

Závislost počtu vad na zásilkách na středisku je dosti vysoká. 87,1 % z celkové variability počtu vad lze vysvětlit pomocí střediska, kde byla zásilka balena.

Financováno Evropskou unií NextGenerationEU
Národní plán obnovy
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY