

## Nové možnosti rozvoje vzdělávání na Technické univerzitě v Liberci

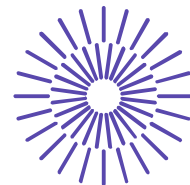
Specifický cíl A3: Tvorba nových profesně zaměřených studijních programů

NPO\_TUL\_MSMT-16598/2022



## Téma 8: Příklad 1 – Test nezávislosti kategoriálních dat

Ing. Vladimíra Hovorková Valentová, Ph.D.



### Zadání příkladu:

Zásilky ve firmě se dopravují do distribučních center po čtyřech trasách (A, B, C a D). Termín dodání zásilky se předem avizuje. Posuďte na hladině významnosti 5 %, zda existuje závislost mezi trasou dodání zásilky a včasností dodání zásilky v avizovaném termínu. Případně změřte sílu dané závislosti pomocí vhodné charakteristiky. Bylo náhodně vybráno 470 zásilek a zjištěny tyto údaje:

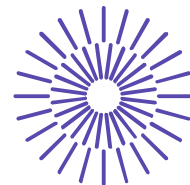
Trasa ( $a_i$ )	Dodání v avizovaném termínu ( $b_j$ )	
	ANO	NE
Trasa A	56	10
Trasa B	112	13
Trasa C	118	45
Trasa D	84	32

### Řešení příkladu:

Vzhledem k tomu, že obě proměnné jsou slovní – nominální, je možné ke zkoumání dané závislosti použít chí-kvadrát test o nezávislosti v kontingenční tabulce (nazývaný také test nezávislosti kategoriálních znaků). Tento test vyžaduje dostatečně velký rozsah souboru – takový, aby teoretické četnosti ve všech políčkách kontingenční tabulky byly větší nebo rovny 5. Tento předpoklad ověříme v průběhu testování v programu SPSS. Nejprve tedy formulujeme hypotézy:

$H_0$ : Mezi včasností dodání zásilky a zvolenou trasou doručení není závislost

$H_1$ : non  $H_0$



Do programu SPSS vložíme data následujícím způsobem:

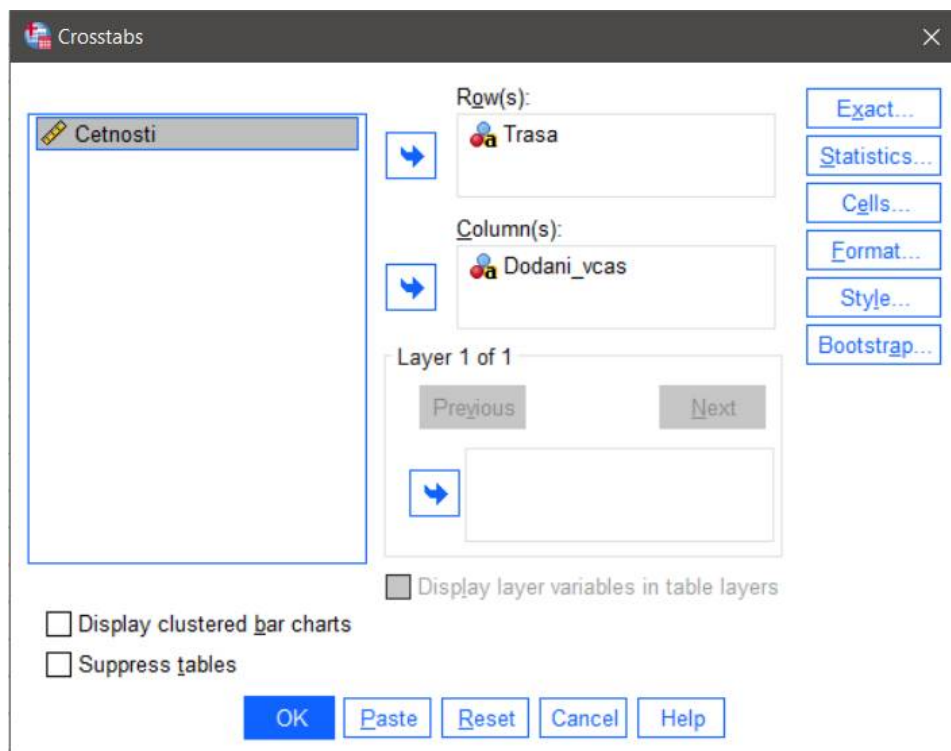
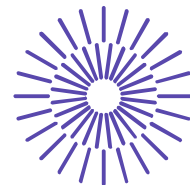
	Trasa	Dodani_vcas	Cetnosti
1	A	ANO	56
2	A	NE	10
3	B	ANO	112
4	B	NE	13
5	C	ANO	118
6	C	NE	45
7	D	ANO	84
8	D	NE	32

V této podobě zatím pracujeme se třemi různými proměnnými, které chceme „svázat“ do tabulky. Následně proto zvolíme posloupnost procedur: **Data – Weight Cases**. Tam jako proměnnou, kterou budeme vážit jednotlivé dvojice případů, označíme proměnnou Cetnosti a potvrdíme OK:

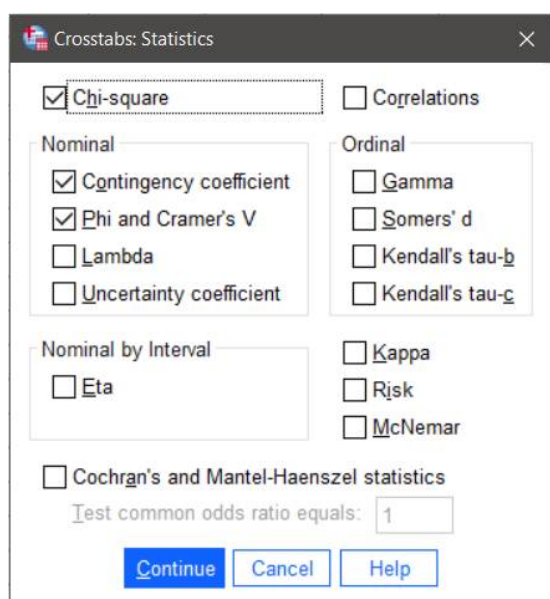


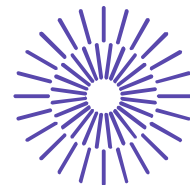
Nyní už trojice samostatných proměnných vystupuje jako kontingenční tabulka.

Pokračujeme posloupností procedur: **Analyze – Descriptive Statistics – Crosstabs ...** Zde vyplníme vstupní panel následovně:



Kromě zadání obou proměnných je potřeba spustit tlačítkovou volbu *Statistics*, kterou vyplníme následujícím způsobem:





Vysvětlení:

- Chi-square – provede se test chí-kvadrát o nezávislosti.
- Contingency Coefficient (nepovinné) – vypočte se hodnota Pearsonova koeficientu kontingence.
- Phi and Cramer's V (doporučeno) – vypočte se buď hodnota koeficientu asociace (Phi) pro alternativní data nebo hodnota Cramérova koeficientu kontingence pro ostatní případy.

Pokud si chceme ověřit hodnoty teoretických četností v jednotlivých políčkách kontingenční tabulky, zvolíme ještě tlačítko *Cells* a zaškrtneme výpočet teoretických četností – *Expected*:

Crosstabs: Cell Display

Counts

Observed

Expected

Hide small counts

Less than 5

z-test

Compare column proportions

Adjust p-values (Bonferroni method)

Percentages

Row

Column

Total

Create APA style table

Residuals

Unstandardized

Standardized

Adjusted standardized

Noninteger Weights

Round cell counts

Round case weights

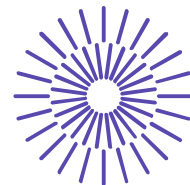
Truncate cell counts

Truncate case weights

No adjustments

Continue Cancel Help

Po potvrzení klávesou Continue a následně OK dostaneme výstup obsahující 3 důležité tabulky:



### Trasa \* Dodani\_vcas Crosstabulation

		Dodani_vcas		Total	
		ANO	NE		
Trasa	A	Count	56	10	66
		Expected Count	52,0	14,0	66,0
B	Count	112	13	125	
	Expected Count	98,4	26,6	125,0	
C	Count	118	45	163	
	Expected Count	128,3	34,7	163,0	
D	Count	84	32	116	
	Expected Count	91,3	24,7	116,0	
Total	Count	370	100	470	
	Expected Count	370,0	100,0	470,0	

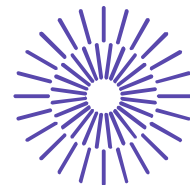
V této tabulce je možné si ověřit, zda je splněn požadavek na rozsah výběru. Jak je vidět ve všech řádcích označených jako Expected Count, všechny teoretické četnosti jsou větší než 5. Požadavek na dostatečně velký rozsah výběru je tedy splněn.

Další tabulka obsahuje výsledek chí-kvadrát testu o nezávislosti:

### Chi-Square Tests

	Value	df	Asymptotic Significance (2- sided)
Pearson Chi-Square	16,964 <sup>a</sup>	3	<,001
Likelihood Ratio	18,223	3	<,001
N of Valid Cases	470		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 14,04.



Relevantní pro naše data jsou údaje v řádku označeném jako Pearson Chi-Square. Hodnota testového kritéria G je 16,964 a vypočtená hladina významnosti Sig. je menší než 0,001 (konkrétně je 0,000719 – uvidíme ji, když 2x poklikáme na tabulku a pak na údaj u Sig.). Vzhledem k tomu, že Sig. je menší než  $\alpha$ , zamítáme  $H_0$  a přijímáme  $H_1$ . Na hladině významnosti 5 % jsme tedy prokázali, že mezi včasností dodání zásilky a zvolenou trasou doručení je závislost. Zároveň si všimněme, že pod tabulkou u poznámky a. je uvedeno, že 0 buněk má očekávané četnosti menší než 5 a že nejmenší hodnota očekávaných (teoretických) četností je 14,04 – i zde tedy najdeme ověření dostatečné velikosti výběru.

Ve třetí tabulce pak najdeme hodnoty měř těsnosti závislosti kategoriálních dat:

		Value	Approximate Significance
Nominal by Nominal	Phi	,190	<,001
	Cramer's V	,190	<,001
	Contingency Coefficient	,187	<,001
N of Valid Cases		470	

Hodnota Cramérova koeficientu kontingence je 0,190, což znamená, že závislost mezi včasností dodání zásilky a zvolenou trasou doručení je velmi slabá. Podobně bychom mohli interpretovat i hodnotu Pearsonova koeficientu kontingence, která je 0,187.