

ADDS cvičení 7

©Pavlína Kuráňová

Příklad 1.

řešíme pomocí kontingenční tabulky

- Analyzujte závislost věku obyvatel na místě kde nejčastěji tráví dovolenou.
(dotazník dovolená, sloupce Jaký je Váš věk a Kde nejčastěji trávíte dovolenou)

Příklad 1.

řešíme pomocí kontingenční tabulky

- ▶ Analyzujte závislost věku obyvatel a místě kde nejčastěji tráví dovolenou.
(dotazník dovolená, sloupce Jaký je Váš věk a Kde nejčastěji cestujete)

Řešení:

1) Překódujeme proměnnou věk do skupin...

(respondenti 15-59 let; do 25-ti; 25 - 35, 35 -45; nad 45 let)

Transform - Visual Binning - Jaký je Váš věk -

Vypisujeme hodnoty Value a popis Label

Vizuální kontrola, přes tabulku četností Analyze - Descriptive Statistics -
Frequencies

Příklad 1.

řešíme pomocí kontingenční tabulky

- Analyzujte závislost věku obyvatel a místě kde nejčastěji tráví dovolenou. (dotazník dovolená, sloupce Jaký je Váš věk a Kde nejčastěji cestujete)

Řešení:

Jaký je Váš věk? (Binned) * Kde nejčastěji trávíte dovolenou? Crosstabulation

Count

		Kde nejčastěji trávíte dovolenou?		Total
		v ČR (včetně chalupaření)	v zahraničí	
Jaký je Váš věk? (Binned)	do 25-ti	10	25	35
	25 - 35 let	17	24	41
	35 - 45 let	5	9	14
	nad 45 let	7	5	12
Total		39	63	102

Analyze - Descriptive Statistics - CrossTabs

Příklad 1.

řešíme pomocí kontingenční tabulky

- Analyzujte závislost věku obyvatel a místě kde nejčastěji tráví dovolenou. (dotazník dovolená, sloupce Jaký je Váš věk a Kde nejčastěji cestujete)

Jaký je Váš věk? (Binned) * Kde nejčastěji trávíte dovolenou? Crosstabulation

Řešení:

		Kde nejčastěji trávíte dovolenou?				Total	
		v ČR (včetně chalupaření)		v zahraničí			
		Count	Expected Count	Count	Expected Count	Count	Expected Count
Jaký je Váš věk? (Binned)	do 25-ti	10	13,4	25	21,6	35	35,0
	25 - 35 let	17	15,7	24	25,3	41	41,0
	35 - 45 let	5	5,4	9	8,6	14	14,0
	nad 45 let	7	4,6	5	7,4	12	12,0
Total		39	39,0	63	63,0	102	102,0

Analyze - Descriptive Statistics - CrossTabs -

položka Cells - Observed i Expected a úprava tabulky přes „pivotování“

Příklad 1.

řešíme pomocí kontingenční tabulky

- ▶ Analyzujte závislost věku obyvatel a místě kde nejčastěji tráví dovolenou.
(dotazník dovolená, sloupce Jaký je Váš věk a Kde nejčastěji cestujete)

Řešení:

- 2) Jaké hypotézy testujeme v kontingenční tabulce? Jaké testy k tomu používáme?
S jakými předpoklady?

Příklad 1.

řešíme pomocí kontingenční tabulky

- ▶ Analyzujte závislost věku obyvatel a místě kde nejčastěji tráví dovolenou.
(dotazník dovolená, sloupce Jaký je Váš věk a Kde nejčastěji cestujete)

Řešení:

2) Jaké hypotézy testujeme v kontingenční tabulce? Jaké testy k tomu používáme?
S jakými předpoklady?

- ▶ hypotéza o shodnosti struktury (homogenity),
 - ▶ hypotéza o nezávislosti,
 - ▶ hypotéza o symetrii.
- ▶ Všechny hypotézy se matematicky formulují pomocí očekávaných četností, testuje se pomocí statistiky χ^2 nebo pomocí věrohodnostního poměru.

Příklad 1.

řešíme pomocí kontingenční tabulky

- ▶ Analyzujte závislost věku obyvatel a místě kde nejčastěji tráví dovolenou.
(dotazník dovolená, sloupce Jaký je Váš věk a Kde nejčastěji cestujete)

Řešení:

2) Jaké hypotézy testujeme v kontingenční tabulce? Jaké testy k tomu používáme?
S jakými předpoklady?

hypotéza o shodnosti struktury (homogeneity), výpočet pomocí statistiky χ^2

Předpoklady:

Příklad 1.

řešíme pomocí kontingenční tabulky

- ▶ Analyzujte závislost věku obyvatel a místě kde nejčastěji tráví dovolenou.
(dotazník dovolená, sloupce Jaký je Váš věk a Kde nejčastěji cestujete)

Řešení:

- 2) Jaké hypotézy testujeme v kontingenční tabulce? Jaké testy k tomu používáme?
S jakými předpoklady?

hypotéza o shodnosti struktury (homogeneity), výpočet pomocí statistiky χ^2

Předpoklady:

- ▶ Alespoň 80% očekávaných četností musí být větších než 5.
- ▶ Všechny očekávané četnosti musí být větší než 1.

Příklad 1.

řešíme pomocí kontingenční tabulky

- Analyzujte závislost věku obyvatel a místě kde nejčastěji tráví dovolenou.
(dotazník dovolená, sloupce Jaký je Váš věk a Kde nejčastěji cestujete)

Řešení:

Jaký je Váš věk? (Binned) * Kde nejčastěji trávíte dovolenou? Crosstabulation

		Kde nejčastěji trávíte dovolenou?				Total	
		v ČR (včetně chalupaření)		v zahraničí			
		Count	Expected Count	Count	Expected Count	Count	Expected Count
Jaký je Váš věk? (Binned)	do 25-ti	10	13,4	25	21,6	35	35,0
	25 - 35 let	17	15,7	24	25,3	41	41,0
	35 - 45 let	5	5,4	9	8,6	14	14,0
	nad 45 let	7	4,6	5	7,4	12	12,0
Total		39	39,0	63	63,0	102	102,0

Příklad 1.

řešíme pomocí kontingenční tabulky

- Analyzujte závislost věku obyvatel a místě kde nejčastěji tráví dovolenou. (dotazník stejný dovolená, sloupce Jaký je Váš věk a Kde nejčastěji cestujete)

Řešení:

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3,655 ^a	3	,301
Likelihood Ratio	3,636	3	,304
N of Valid Cases	102		

a. 1 cells (12,5%) have expected count less than 5. The minimum expected count is 4,59.

Analyze - Descriptive Statistics - CrossTabs - položka Statistics - Chi-Square

Můžeme pokračovat? Nějaký nápad na řešení?

Příklad 1.

řešíme pomocí kontingenční tabulky

- Analyzujte závislost věku obyvatel a místě kde nejčastěji tráví dovolenou. (dotazník stejný dovolená, sloupce Jaký je Váš věk a Kde nejčastěji cestujete)

Závěr: Co jsem analyzoval jaká byla hypotéza a jaká alternativa?

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3,655 ^a	3	,301
Likelihood Ratio	3,636	3	,304
N of Valid Cases	102		

a. 1 cells (12,5%) have expected count less than 5. The minimum expected count is 4,59.

Příklad 1.

řešíme pomocí kontingenční tabulky

- Analyzujte závislost věku obyvatel a místě kde nejčastěji tráví dovolenou.
(dotazník stejný dovolená, sloupce Jaký je Váš věk a Kde nejčastěji cestujete)

Závěr: Co jsem analyzoval jaká byla hypotéza a jaká alternativa?

H_0 : Proměnné věk a místo kam nejraději cestujete jsou nezávislé.

H_A : Proměnné věk a místo kam nejraději cestujete jsou závislé..

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3,655 ^a	3	,301
Likelihood Ratio	3,636	3	,304
N of Valid Cases	102		

a. 1 cells (12,5%) have expected count less than 5. The minimum expected count is 4,59.

Příklad 1.

řešíme pomocí kontingenční tabulky

- Analyzujte závislost věku obyvatel a místě kde nejčastěji tráví dovolenou. (dotazník stejný dovolená, sloupce Jaký je Váš věk a Kde nejčastěji cestujete)

Závěr: Co jsem analyzoval jaká byla hypotéza a jaká alternativa?

H_0 : Proměnné věk a místo kam nejraději cestujete jsou nezávislé.

H_A : Proměnné věk a místo kam nejraději cestujete jsou závislé..

Nezamítáme nulovou hypotézu o nezávislosti proměnných věk a místem, kam respondenti nejraději cestují.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3,655 ^a	3	,301
Likelihood Ratio	3,636	3	,304
N of Valid Cases	102		

a. 1 cells (12,5%) have expected count less than 5. The minimum expected count is 4,59.

Příklad 1.

řešíme pomocí kontingenční tabulky

- ▶ Analyzujte závislost věku obyvatel a místě kde nejčastěji tráví dovolenou. (dotazník stejný dovolená, sloupce Jaký je Váš věk a Kde nejčastěji cestujete)

Další testy nezávislosti proměnných:

- Pearsonův kontingenční koeficient
- Koeficient φ
- Cramérovo V

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,189	,301
	Cramer's V	,189	,301
	Contingency Coefficient	,186	,301
N of Valid Cases		102	

Analyze - Descriptive Statistics - CrossTabs - položka Statistics - Chi-Square - Contingency Coefficient, Phi and Cramer's V

Příklad 1.

řešíme pomocí kontingenční tabulky

- ▶ Analyzujte závislost věku obyvatel a místě kde nejčastěji tráví dovolenou.
(dotazník stejný dovolená, sloupce Jaký je Váš věk a Kde nejčastěji cestujete)

Další testy nezávislosti proměnných:

- Pearsonův kontingenční koeficient
- Koeficient φ
- Cramérovo V

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,189	,301
	Cramer's V	,189	,301
	Contingency Coefficient	,186	,301
N of Valid Cases		102	

Na základě výsledků vidíme že nemůžeme zamítnout nulovou hypotézu o nezávislosti proměnných.

Příklad 1. navíc adjustovaná rezidua?

řešíme pomocí kontingenční tabulky

Adjustovaná residua ???

Příklad 1. navíc adjustovaná rezidua?

řešíme pomocí kontingenční tabulky

Adjustovaná residua ???

- ▶ Residuum v daném políčku tabulky (= Pozorovaná (observed) minus Očekávaná (expected) hodnota) dělené odhadem vlastní standardní chyby.
Standardizovaný residuál je vyjádřen v jednotkách směrodatné odchylky nad nebo pod průměrem.

$$e_{ij}^A = \frac{n_{ij} - o_{ij}}{\sqrt{o_{ij}(1 - p_{i+})(1 - p_{+j})}}$$

Znaménkové schéma → jednoduchá vizualizace, kde

- ▶ $\text{abs}(z) \geq 3.29$ nahradíme +++ resp. ---
- ▶ $\text{abs}(z) \geq 2.58$ nahradíme ++ resp. --
- ▶ $\text{abs}(z) \geq 1.96$ nahradíme + resp. -

Příklad 1. navíc adjustovaná rezidua?

řešíme pomocí kontingenční tabulky

Adjustovaná residua

Znaménkové schéma → jednoduchá vizualizace,

- ▶ $abs(z) \geq 3.29$ nahradíme +++ resp. ---
- ▶ $abs(z) \geq 2.58$ nahradíme ++ resp. --
- ▶ $abs(z) \geq 1.96$ nahradíme + resp. -

Analyze - Descriptive Statistics - CrossTabs -

položka Cells - Adjusted Standardized

- ▶ Z tabulky je patrné že doplníme jen + nebo jen - . Mezi proměnnými opravdu neexistuje významná závislost.

Jaký je Váš věk? (Binned) ^ Kde nejčastěji trávíte dovolenou? Crosstabulation

		Kde nejčastěji trávíte dovolenou?		Total	
		v ČR (včetně chalupaření)	v zahraničí		
Jaký je Váš věk? (Binned)	do 25-ti	Count	10	25	35
		Expected Count	13,4	21,6	35,0
		Adjusted Residual	-1,5	1,5	
25 - 35 let		Count	17	24	41
		Expected Count	15,7	25,3	41,0
		Adjusted Residual	,6	-,6	
35 - 45 let		Count	5	9	14
		Expected Count	5,4	8,6	14,0
		Adjusted Residual	-,2	,2	
nad 45 let		Count	7	5	12
		Expected Count	4,6	7,4	12,0
		Adjusted Residual	1,5	-,5	
Total		Count	39	63	102
		Expected Count	39,0	63,0	102,0

Příklad 2.

řešíme pomocí kontingenční tabulky

- Analyzujte závislost proměnných pohlaví a počtu dětí. Viz příklad z předchozí hodiny. (dotazník stejný dovolená, sloupce Jste a Máte děti)

Řešení:

Formulace H_0 a H_A

H_0 : Mezi pohlavím a počtem dětí neexistuje závislost.

H_A : Proměnné pohlaví a počet dětí jsou závislé.

Příklad 2.

řešíme pomocí kontingenční tabulky

- Analyzujte závislost proměnných pohlaví a počtu dětí. Viz příklad z předchozí hodiny. (dotazník stejný dovolená, sloupce Jste a Máte děti)

Jste? * Máte děti? Crosstabulation

			Máte děti?			Total
			ano - 1	ano - 2-3	nemám	
Jste?	muž	Count	4	14	27	45
		Expected Count	5,7	7,5	31,8	45,0
		Adjusted Residual	-1,0	3,5	-2,1	
	žena	Count	9	3	45	57
		Expected Count	7,3	9,5	40,2	57,0
		Adjusted Residual	1,0	-3,5	2,1	
Total		Count	13	17	72	102
		Expected Count	13,0	17,0	72,0	102,0

Analyze - Descriptive Statistics - CrossTabs - položka Cells - Expected, Observed, Adjusted Standardized

Příklad 2.

řešíme pomocí kontingenční tabulky

- Analyzujte závislost proměnných pohlaví a počtu dětí. Viz příklad z předchozí hodiny. (dotazník stejný dovolená, sloupce Jste a Máte děti)

Jste? * Máte děti? Crosstabulation

			Máte děti?			Total
			ano - 1	ano - 2-3	nemám	
Jste?	muž	Count	4	14	27	45
		Expected Count	5,7	7,5	31,8	45,0
		Adjusted Residual	-1,0	3,5	-2,1	
žena	Count	9	3	45	57	
	Expected Count	7,3	9,5	40,2	57,0	
	Adjusted Residual	1,0	-3,5	2,1		
Total	Count	13	17	72	102	
	Expected Count	13,0	17,0	72,0	102,0	

Všechny Expected Count jsou větší než 5, mohou pokračovat v testování.

Příklad 2.

řešíme pomocí kontingenční tabulky

- Analyzujte závislost proměnných pohlaví a počtu dětí. Viz příklad z předchozí hodiny. (dotazník stejný dovolená, sloupce Jste a Máte děti)

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	12,299 ^a	2	,002
Likelihood Ratio	12,830	2	,002
N of Valid Cases	102		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 5,74.

Na hladině významnosti 1% i 5% zamítám nulovou hypotézu. Mezi proměnnými existuje významná závislost.

Příklad 2.

řešíme pomocí kontingenční tabulky

- Analyzujte závislost proměnných pohlaví a počtu dětí. Viz příklad z předchozí hodiny.

Míry závislosti založené na Pearsonově chí-kvadrát

		Value	Approx. Sig.
Nominal by Nominal	Phi	,347	,002
	Cramer's V	,347	,002
	Contingency Coefficient	,328	,002
N of Valid Cases		102	

Na hladině významnosti 1% i 5% zamítám nulovou hypotézu. Mezi proměnnými existuje významná závislost.

Příklad 2.

řešíme pomocí kontingenční tabulky

- ▶ Analyzujte závislost proměnných pohlaví a počtu dětí. Viz příklad z předchozí hodiny.

Míry závislosti založené na Pearsonově chí-kvadrát

Doporučení:

- ▶ tabulky s alespoň jednou dichotomickou proměnnou - Crameruv koeficient;
- ▶ jinak korigovaný Pearsonuv koeficient;
- ▶ častěji se koeficienty kontingence používají pro porovnání intenzity závislosti.

Příklad 2. navíc adjustovaná rezidua?

řešíme pomocí kontingenční tabulky

Adjustovaná residua

Znaménkové schéma → jednoduchá vizualizace,

- ▶ $abs(z) \geq 3.29$ nahradíme +++ resp. ---
- ▶ $abs(z) \geq 2.58$ nahradíme ++ resp. --
- ▶ $abs(z) \geq 1.96$ nahradíme + resp. -

Analyze - Descriptive Statistics - CrossTabs -
položka Cells - Adjusted Standardized

Jste? * Máte děti? Crosstabulation

		Máte děti?			Total	
		ano - 1	ano - 2-3	nemám		
Jste?	muž	Count	4	14	27	45
		Expected Count	5,7	7,5	31,8	45,0
		Adjusted Residual	-1,0	3,5	-2,1	
	žena	Count	9	3	45	57
		Expected Count	7,3	9,5	40,2	57,0
		Adjusted Residual	1,0	-3,5	2,1	
Total		Count	13	17	72	102
		Expected Count	13,0	17,0	72,0	102,0

- Z tabulky je patrné že mezi proměnnými opravdu existuje významná závislost.

Příklad 2.

řešíme pomocí kontingenční tabulky

- ▶ Analyzujte závislost proměnných pohlaví a počtu dětí.

Jaké další míry závislosti máme?

- ▶ Asymetrické míry:

lambda

Goddman - kruskalovo tau

Koeficient neurčitosti

Příklad 2.

řešíme pomocí kontingenční tabulky

- Analyzujte závislost proměnných pohlaví a počtu dětí.

Directional Measures

			Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Nominal by Nominal	Lambda	Symmetric	,147	,046	2,766	,006
		Jste? Dependent	,244	,080	2,766	,006
		Máte děti? Dependent	,000	,000	. ^c	. ^c
Goodman and Kruskal tau	Jste? Dependent	,121	,056		,002 ^d	
	Máte děti? Dependent	,058	,034		,003 ^d	
Uncertainty Coefficient	Symmetric	Jste? Dependent	,084	,043	1,913	,002 ^e
		Jste? Dependent	,092	,048	1,913	,002 ^e
		Máte děti? Dependent	,078	,040	1,913	,002 ^e

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Cannot be computed because the asymptotic standard error equals zero.

d. Based on chi-square approximation

e. Likelihood ratio chi-square probability.

- **Analyze - Descriptive Statistics - CrossTabs - položka Statistics - Lambda, Uncertainty Coefficient**

Příklad 2.

řešíme pomocí kontingenční tabulky

- ▶ Analyzujte závislost proměnných pohlaví a počtu dětí.

Formulace H_0 a H_A

H_0 : Mezi pohlavím a počtem dětí existuje jednostranná závislost.

$H_A: \overline{H_0}$

- ▶ Výsledek?
- ▶ Zamítám H_0 ve prospěch alternativy nebo nezamítáme H_0