

Test dobré shody

Chí-kvadrát test dobré shody (1)

Testuje hypotézu, že rozložení četností kategorizované proměnné odpovídá předpokládanému rozdělení (tj. výběr pochází z daného rozdělení)

- *Jsou výzkumná data reprezentativní?*
- *Odpovídá rozložení vzdělanosti populačním hodnotám?*
- *Pokrývá trh výrobku proporcionálně jednotlivé věkové kategorie?*

- **testová statistika**

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad , \text{ kde}$$

E_i ...očekávané četnosti kategorií (Expected counts)

O_i... pozorované četnosti za platnosti nulové hypotézy (Observed counts)

- **stupně volnosti (df):** počet kategorií - 1

Chí-kvadrát test dobré shody (2)

H_0 : datový soubor představuje výběr z populace, kde je zastoupení mužů 48,76 % a žen 51,24 %

	Observed N	Expected N	Residual
muž	257	245,8	11,2
žena	247	258,2	-11,2
Total	504		

Test Statistics

	Pohlaví
Chi-Square	1,005 ^a
df	1
Asymp. Sig.	,316

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 245,8.

Závěr: nezamítáme H_0

- *Je datový soubor reprezentativní vzhledem k pohlaví?*
- *Mohou být odchylky od předpokládané struktury způsobené náhodou?*

Předpoklady testu:

- alespoň 30 případů
- všechny očekávané četnosti jsou větší než 1
- nejméně 80 % očekávaných četností je větší než 5

- Ovlivňuje kouření matky během těhotenství porodní hmotnost dítěte?

Kouření během těhotenství * Porodní hmotnost Crosstabulation

			Porodní hmotnost		Total
			normální (>= 2500g)	nízká (<2500g)	
Count	Kouření během těhotenství	ne	86	29	115
		ano	44	30	74
	Total		130	59	189
% within těhotenství	Kouření během těhotenství	ne	74,8%	25,2%	100,0%
		ano	59,5%	40,5%	100,0%
	Total		68,8%	31,2%	100,0%

Test nezávislosti řádkové a sloupcové proměnné

Pearsonův test chí-kvadrát

H_0 : řádková a sloupcová proměnná v kontingenční tabulce jsou nezávislé

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4,924 ^a	1	,026		
Continuity Correction ^b	4,236	1	,040		
Likelihood Ratio	4,867	1	,027		
Fisher's Exact Test				,036	,020
Linear-by-Linear Association	4,898	1	,027		
N of Valid Cases	189				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 23,10.

b. Computed only for a 2x2 table

Test je založený na komparaci pozorovaných (O_{rs}) a očekávaných (E_{rs}) a četností :

- $\chi^2 = \sum (E_{rs} - O_{rs})^2 / E_{rs}$
- $df = (R-1) * (S-1)$, kde
 R, S je počet kategorií řádkové resp. sloupcové proměnné

Předpoklady:

- alespoň 30 případů
- všechny očekávané četnosti jsou větší než 1
- nejméně 80% buněk má očekávané četnosti větší než 5

- pro zjišťování u kterých kategorií nastal významný rozdíl lze užít **adjustovaná standardizovaná rezidua**
- porovnávají se s kvantily standardizovaného normálního rozložení pro zvolenou hladinu spolehlivosti
- výsledek lze graficky znázornit znaménkovým schématem:

Kouření během těhotenství * Porodní hmotnost Crosstabulation

			Porodní hmotnost		Total
			normální (>= 2500g)	nízká (<2500g)	
Count	Kouření během těhotenství	ne	86	29	115
		ano	44	30	74
	Total		130	59	189
% within těhotenství	Kouření během těhotenství	ne	74,8%	25,2%	100,0%
		ano	59,5%	40,5%	100,0%
	Total		68,8%	31,2%	100,0%
Sign Scheme	Kouření během těhotenství	ne	+	-	
		ano	-	+	

- Tabulka s adjustovanými rezidui je upravena pomocí skriptu *Znaménkové schéma* (volně k dispozici na stránkách www.acrea.cz)
- Skript porovnává adjustovaná rezidua s 95%, 99% a 99,9% kvantily standardizovaného normálního rozdělení (tj. zaokrouhleně $\pm 1,96$ / $\pm 2,58$ / $\pm 3,29$)