

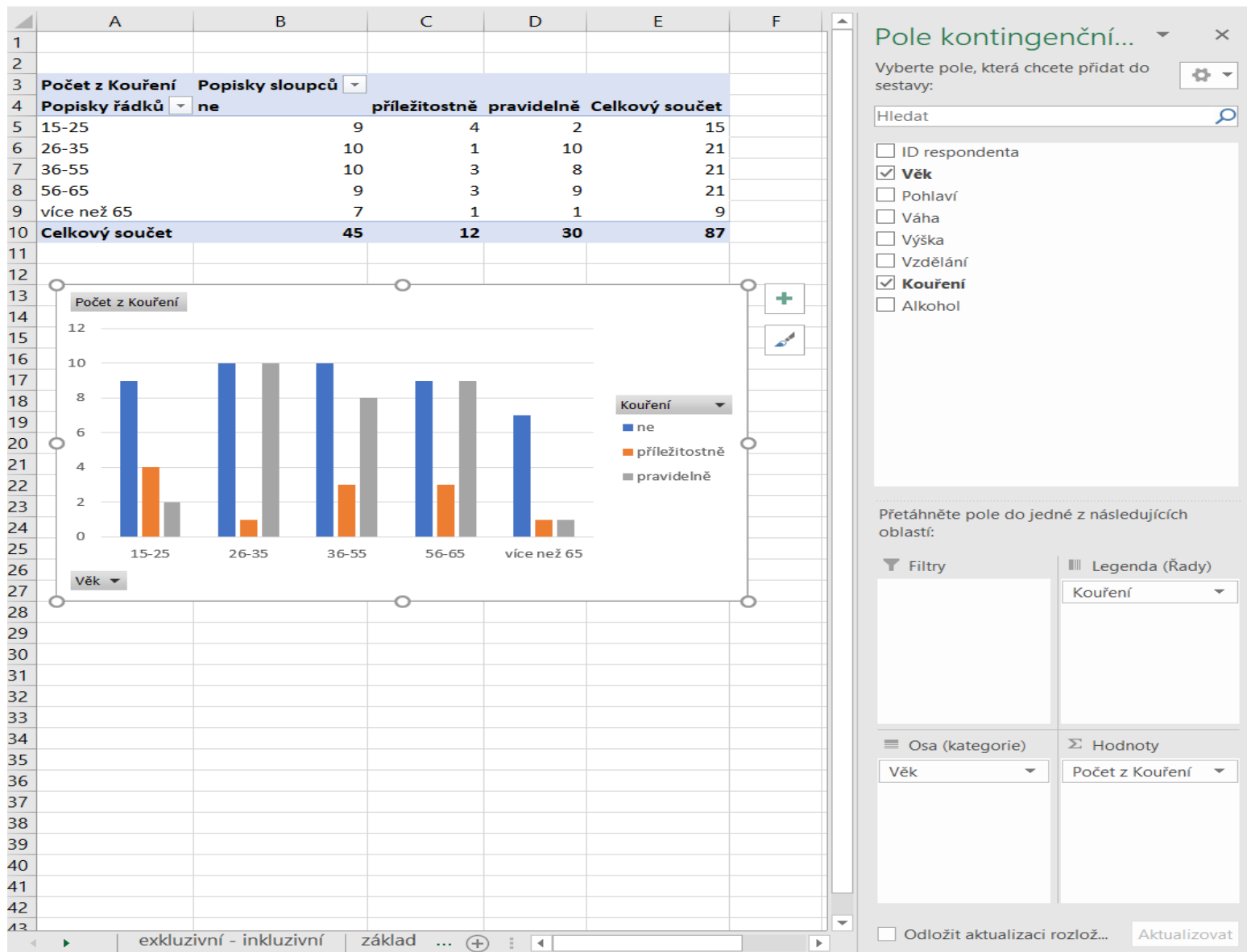
Test  $\chi^2$  (chí kvadrát)

# Test $\chi^2$ (chí kvadrát)

- Máme k dispozici náhodný výběr rozsahu  $n$  rozdělený do dvou znaků (znak 1, znak 2)
- Úkolem testu je rozhodnout, zda jsou znaky na sobě závislé nebo nezávislé (zda znak 1 má vliv na znak 2)
- Znak 1
- Znak 2



# Kontingenční tabulka



# Hypotézy

- Nulová hypotéza: znaky 1 a 2 jsou nezávislé
- Alternativní hypotéza: mezi znaky 1 a 2 existuje závislost

# Postup výpočtu

- Sestaví se tabulka skutečných (naměřených) relativních četností
  - Sestaví se tabulka očekávaných četností
  - Podmínky pro použití testu nezávislosti v kontingenční tabulce:
    - nejvíce 20 % teoretických četností může být menších než 5
    - žádná teoretická četnost nesmí být menší než 1
- Pro tabulku 2x2:
- $n > 40$
  - pokud  $20 < n < 40$ , pak je nutná úprava testového kritéria pomocí Yatesovy korekce
  - pokud  $n < 20$ , pak použijeme Fisherův test
- Vypočte se testové kritérium (dosazení do vzorce – výsledek hodnota)
  - Testové kritérium se srovná s kritickou hodnotou (tabulková hodnota, je potřeba zohlednit počet stupňů volnosti)
  - Je-li testové kritérium  $<$  kritická hodnota, potom nezamítáme hypotézu o nezávislosti a nezávislost lze předpokládat

# Kritické hodnoty testového kritéria chí-kvadrát

Stupně volnosti	Hladina významnosti	
	0,05	0,01
1	3,841	6,635
2	5,991	9,21
3	7,815	11,341
4	9,483	13,277
5	11,070	15,086
6	12,592	16,812
7	14,067	18,475
8	15,507	20,09
9	16,919	21,666
10	18,307	23,209
11	19,675	24,725
12	21,026	26,217
13	22,362	27,688
14	23,685	29,141
15	24,996	30,578
16	26,296	32
17	27,587	33,409
18	28,868	34,805
19	30,144	36,191
20	31,410	37,566

# Příklad 1

- 2 vybrané otázky z dotazníku (2 znaky). Souvisí spolu dosažené vzdělání a péče o zrak?

1. Domníváte se, že se dostatečně pečujete o svůj zrak?

- a) ano
- b) ne
- c) někdy

18. Nejvyšší dosažené vzdělání

- a) základní
- b) středoškolské
- c) vyšší odborné
- d) vysokoškolské



Sestaví se tabulka skutečných (relativních) četností

	základní	SŠ	VOŠ	VŠ	celkem
ano	33	132	28	69	262
někdy	6	74	0	70	150
ne	11	128	6	103	248
celkem	50	334	34	242	660

- <http://www.milankabrt.cz/testNezavislosti/index.php>

# Aplikovaná statistika

Test chí-kvadrát nezávislosti v kontingenční tabulce

- » o aplikaci
- » aplikace

**Aplikovaná statistika**

Test chí-kvadrát nezávislosti v kontingenční tabulce

**Test chí-kvadrát nezávislosti v kontingenční tabulce**

**vank**

Tato webová aplikace byla vyvíjena na jaře roku 2011, kdy vznikla jako výsledek semestrálního projektu z předmětu "Aplikovaná statistika" v rámci studia na [Univerzitě Jihočeské, Katedře oboru "Informační management"](#).

Vytvořil Milan Kabrt

[Spustit aplikaci](#)

**Test chí-kvadrát nezávislosti v kontingenční tabulce**

Máme v dispozici nějaký soubor rozdělený do dvou statistických znaků: znak1, znak2, které nám tvoří tabulku, přičemž znak1 z nás známkou je rozdělen do 3 skupin, resp. 3 skupin čísel. Sledem toho je vzorová tabulka (pro více zkusů na soubě závislosti) rozdělena. Test chí-kvadrát nezávislosti (nebo také test nezávislosti) je test, který nám umožňuje zjistit, zda existuje nějaká závislost mezi dvěma znaky (nebo také mezi dvěma znaky) v tabulce (nebo také mezi dvěma znaky).

	znak1 - 1.skupina	znak1 - 2.skupina	celkem
znak2 - 1.skupina	skutečné četnosti	skutečné četnosti	$n_{1.}$
znak2 - 2.skupina	skutečné četnosti	skutečné četnosti	$n_{2.}$
znak2 - 3.skupina	skutečné četnosti	skutečné četnosti	$n_{3.}$
...	...	...	$n_{.j}$
celkem	$n_{.1}$	$n_{.2}$	$n_{.j}$

$$N_{1j} = \sum_{i=1}^3 n_{ij} \quad N_{2j} = \sum_{i=1}^3 n_{ij}$$

V zobrazení aplikace máme doplněnou stručně (nenáhraně) nastavení testu (četnosti, náhodující tabulka a další ukazatele závislosti (četnosti, závislosti, závislosti, závislosti).

**Aplikovaná statistika**

Test chí-kvadrát nezávislosti v kontingenční tabulce

**Zadáno parametry**

Počet skupin znak1:

Počet skupin znak2:

Hladina významnosti alpha:

[Pokačte](#)

© CREATED BY: Milan Kabrt | www.milankabrt.cz

# Aplikovaná statistika

» o aplikaci

» aplikace

## Test chí-kvadrát nezávislosti v kontingenční tabulce

základní parametry

relativní četnosti

výsledek

1

Počet skupin **znaku 1**: ?

Počet skupin **znaku 2**: ?

Hladina významnosti  **$\alpha$** : ?

Pokračovat

# Aplikovaná statistika

» o aplikaci

» aplikace

## Test chí-kvadrát nezávislosti v kontingenční tabulce

základní parametry

relativní četnosti

výsledek


1

Počet skupin znaku 1:

Počet skupin znaku 2:

Hladina významnosti  $\alpha$ :

2

Zadejte do tabulky naměřené relativní četnosti: 

	znak1 - 1. sk.	znak1 - 2. sk.	znak1 - 3. sk.	znak1 - 4. sk.
znak2 - 1. sk.	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
znak2 - 2. sk.	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
znak2 - 3. sk.	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

zpět

pokračovat

# Aplikovaná statistika

» o aplikaci  
» aplikace

## Test chí-kvadrát nezávislosti v kontingenční tabulce

základní parametry    relativní četnosti    výsledek

1

Počet skupin znaku 1:

Počet skupin znaku 2:

Hladina významnosti  $\alpha$ :

2 Zadejte do tabulky naměřené relativní četnosti: ?

	znak1 - 1. sk.	znak1 - 2. sk.	znak1 - 3. sk.	znak1 - 4. sk.
znak2 - 1. sk.	<input type="text" value="33"/>	<input type="text" value="132"/>	<input type="text" value="28"/>	<input type="text" value="69"/>
znak2 - 2. sk.	<input type="text" value="6"/>	<input type="text" value="74"/>	<input type="text" value="0"/>	<input type="text" value="70"/>
znak2 - 3. sk.	<input type="text" value="11"/>	<input type="text" value="128"/>	<input type="text" value="6"/>	<input type="text" value="103"/>

zpět    pokračovat

základní parametry

relativní četnosti

výsledek

### Výsledek testu

Počet skupin znaku 1:	4
Počet skupin znaku 2:	3
Hladina významnosti $\alpha$ :	5 %

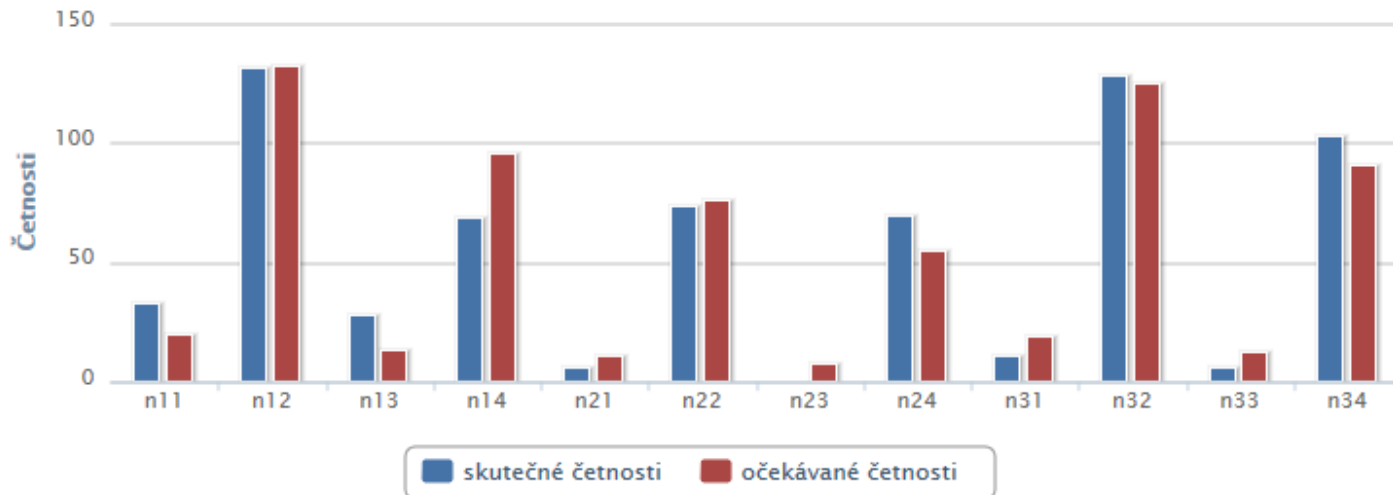
#### Skutečné četnosti

	znak1 - 1. sk.	znak1 - 2. sk.	znak1 - 3. sk.	znak1 - 4. sk.	$n_{j\cdot}$
znak2 - 1. sk.	33	132	28	69	262
znak2 - 2. sk.	6	74	0	70	150
znak2 - 3. sk.	11	128	6	103	248
$n_{\cdot j}$	50	334	34	242	660

#### Očekávané četnosti

	znak1 - 1. sk.	znak1 - 2. sk.	znak1 - 3. sk.	znak1 - 4. sk.	$n_{j\cdot}$
znak2 - 1. sk.	19.85	132.59	13.5	96.07	262
znak2 - 2. sk.	11.36	75.91	7.73	55	150
znak2 - 3. sk.	18.79	125.5	12.78	90.93	248
$n_{\cdot j}$	50	334	34	242	660

## skutečné a očekávané četnosti



Highcharts.com

**testové kritérium:**

$$G = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}}$$

Po dosazení do vzorce vychází testové kritérium:  
 $G = 54.792$

**Kritická hodnota:**

Kritická hodnota:  
 $\chi_{(1-\alpha); ar} = 12.592$

**Rozhodnutí:**

**Na hladině významnosti 5 % nulovou hypotézu ( $H_0$ ) o nezávislosti jednotlivých znaků zamítáme a přijímáme hypotézu  $H_1$ , která nám říká, že zde určitá závislost existuje.**

# Bakalářská práce

- Z našeho dotazníku vyplynulo, že péče o zrak a vzdělání jsou na sobě závislé veličiny. Závislost jsme ověřili pomocí kontingenčních tabulek a použili jsme test chí kvadrát. Kritická mez pro hladinu významnosti byla zvolena 0,05. Vypočtená hodnota testového kritéria je 54,792. Počet stupňů volnosti je 6, kritická hodnota pro 6 stupňů volnosti je 12,592. Protože kritická hodnota je menší než vypočtená hodnota, z provedeného testu vyplývá, že veličiny jsou na sobě závislé.
- Korigovaný koeficient kontingence pomocí Pearsona
- Cramerův koeficient
- Korelace mezi hodnotami je podle ... středně silná.



# Příklad 2

- 2 vybrané otázky z dotazníku (2 znaky). Souvisí spolu pohlaví a péče o zrak?

1. Domníváte se, že se dostatečně pečujete o svůj zrak?

- a) ano
- b) ne
- c) někdy

16. Pohlaví

- a) žena
- b) muž

# Výsledek testu

Počet skupin znaku 1:	2
Počet skupin znaku 2:	3
Hladina významnosti $\alpha$ :	5 %

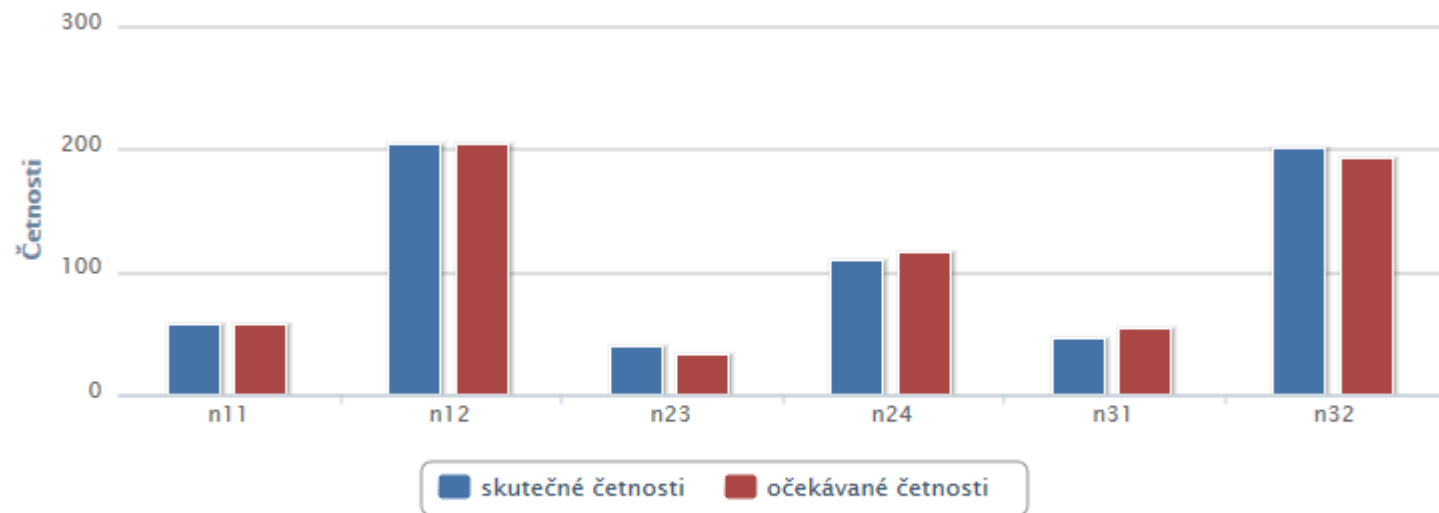
## Skutečné četnosti

	znak1 - 1. sk.	znak1 - 2. sk.	$n_{j\cdot}$
znak2 - 1. sk.	58	204	262
znak2 - 2. sk.	40	110	150
znak2 - 3. sk.	47	201	248
$n_{\cdot j}$	145	515	660

## Očekávané četnosti

	znak1 - 1. sk.	znak1 - 2. sk.	$n_{j\cdot}$
znak2 - 1. sk.	57.56	204.44	262
znak2 - 2. sk.	32.95	117.05	150
znak2 - 3. sk.	54.48	193.52	248
$n_{\cdot j}$	145	515	660

## skutečné a očekávané četnosti



Highcharts.com

**testové kritérium:**

$$G = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}}$$

Po dosazení do vzorce vychází testové kritérium:

$G = 3.253$

**Kritická hodnota:**

Kritická hodnota:

$\chi_{(1-\alpha); dr} = 5.991$

**Rozhodnutí:**

**Na hladině významnosti 5 % nulovou hypotézu ( $H_0$ ) o nezávislosti jednotlivých znaků nezamítáme.**

# Bakalářská práce

- Z našeho dotazníku vyplynulo, že péče o zrak a pohlaví na sobě nezávisí. Závislost jsme ověřili pomocí kontingenčních tabulek a použili jsme test chí kvadrát. Kritická mez pro hladinu významnosti byla zvolena 0,05. Vypočtená hodnota testového kritéria je 3,253. Počet stupňů volnosti je 4, kritická hodnota pro 4 stupně volnosti je 5,991. Protože kritická hodnota je větší než vypočtená hodnota, z provedeného testu vyplývá, že veličiny jsou na sobě nezávislé.

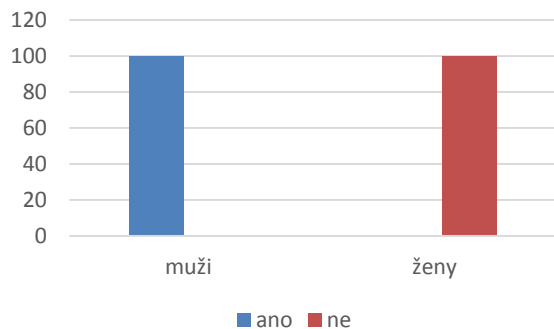
# Závislost

## Skutečné četnosti

	ano	ne	celkem
muži	100	0	100
ženy	0	100	100
celkem	100	100	200

## Očekávané četnosti

	ano	ne	celkem
muži	50	50	100
ženy	50	50	100
celkem	100	100	200



Testová statistika 200  
Kritická hodnota 3,841  
Pearsonův koeficient 1  
Cramerův koeficient 1

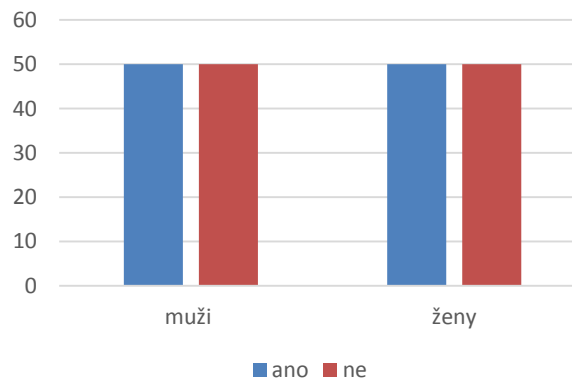
# Nezávislost

## Skutečné četnosti

	ano	ne	celkem
muži	50	50	100
ženy	50	50	100
celkem	100	100	200

## Očekávané četnosti

	ano	ne	celkem
muži	50	50	100
ženy	50	50	100
celkem	100	100	200



Testová statistika 0  
Kritická hodnota 3,841  
Pearsonův koeficient 0  
Cramerův koeficient 0

	ano	ne	celkem
muži	50	50	=SUMA(B2:C2)
ženy	50	50	=SUMA(B3:C3)
celkem	=SUMA(B2:B3)	=SUMA(C2:C3)	=SUMA(B4:C4)
	ano	ne	celkem
muži	= $\$D2 * B\$4 / \$D\$4$	= $\$D2 * C\$4 / \$D\$4$	=SUMA(B7:C7)
ženy	= $\$D3 * B\$4 / \$D\$4$	= $\$D3 * C\$4 / \$D\$4$	=SUMA(B8:C8)
celkem	=SUMA(B7:B8)	=SUMA(C7:C8)	=SUMA(B9:C9)
testová statistika		= $\text{POWER}(B2-B7;2)/B7$	= $\text{POWER}(C2-C7;2)/C7$
		= $\text{POWER}(B3-B8;2)/B8$	= $\text{POWER}(C3-C8;2)/C8$
		=C11+D11+C12+D12	
kritická hodnota		3,841	
Cramerův koeficient		= $\text{ODMOCNINA}(C14/D9*(2-1))$	
Pearsonův koeficient		= $\text{ODMOCNINA}((C14/(C14+D4))/((2-1)/2))$	

# Korigovaný koeficient kontingence pomocí Pearsona

$$C_{kor} = \frac{\sqrt{\frac{\chi^2}{\chi^2 + n}}}{\sqrt{\frac{m-1}{m}}}$$

kde  $\chi^2$  je hodnota testového kritéria

$n$  je rozsah souboru

$m$  je počet řádků nebo počet sloupců v kontingenční tabulce (je-li větší počet řádků, je  $m$  počet řádků; je-li větší počet sloupců, je  $m$  počet sloupců)

# Cramerův koeficient

$$V = \sqrt{\frac{\chi^2}{n(m-1)}}$$

$\chi^2$  je hodnota testového kritéria

$n$  je rozsah souboru

$m$  je počet řádků nebo počet sloupců

v kontingenční tabulce (je-li větší počet řádků, je  $m$  počet řádků; je-li větší počet sloupců, je  $m$  počet sloupců)



# $p$ hodnota - Excel

## **CHISQ.TEST (funkce)**

Vrátí test nezávislosti. Funkce CHISQ.TEST vrátí hodnotu rozdělení chí-kvadrát ( $\chi^2$ ) pro dané testové kritérium a příslušné stupně volnosti. Pomocí testů  $\chi^2$  můžete určit, zda experiment potvrzuje předpokládané výsledky.

## **Syntaxe**

CHISQ.TEST(aktuální,očekávané)

$p < 0,05 \rightarrow$  nezávislost zamítáme, určitá závislost existuje

# $p$ -hodnota

$p$ -hodnota je nejmenší hladina významnosti, při které ještě zamítneme nulovou hypotézu

$p$ -hodnota je pravděpodobnost, že při platnosti nulové hypotézy nabývá testová statistika své stávající hodnoty anebo hodnot ještě extrémnějších (nepříznivějších vůči nulové hypotéze).