


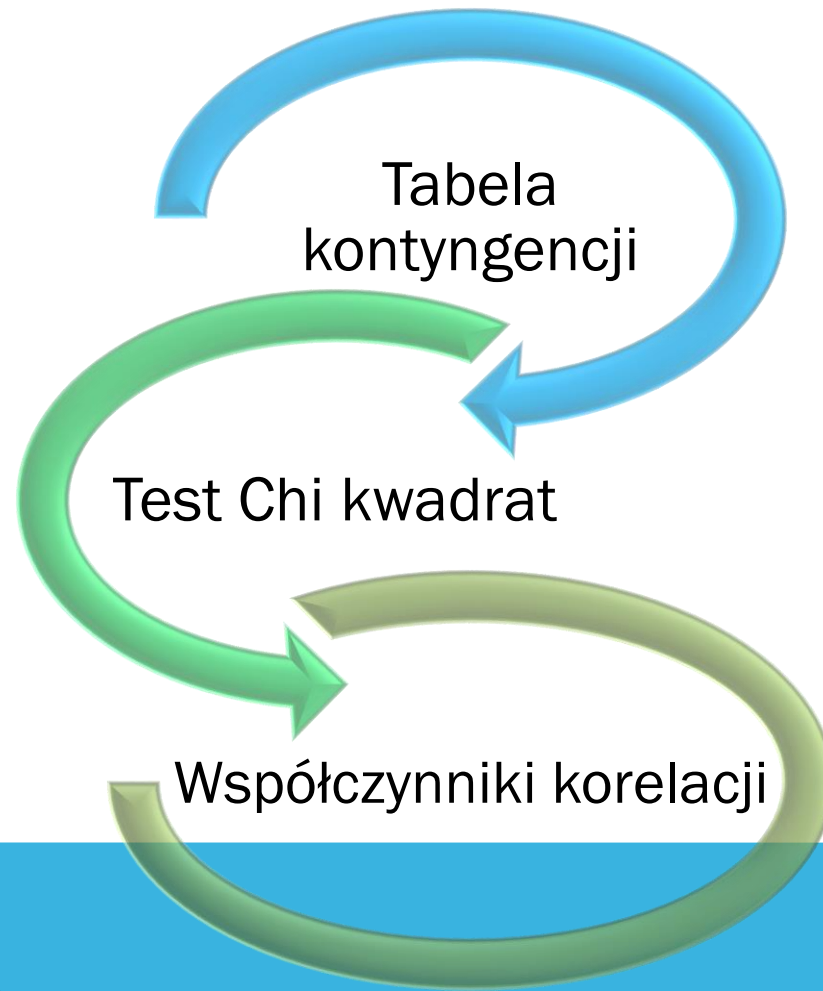
# ANALIZA ZALEŻNOŚCI POMIĘDZY ZMIENNYMI JAKOŚCIOWYMI

Karolina Tura-Gawron

# PLAN

- I. Tabele kontyngencji
  - II. Test chi kwadrat
  - III. Współczynniki
  - IV. Statistica
- 

# ANALIZA ZALEŻNOŚCI POMIĘDZY ZMIENNYMI JAKOŚCIOWYMI



# TABELA KONTYNGENCJI

$X \backslash Y$	$Y_1$	$Y_2$	...	...	$Y_p$	
$X_1$	$n_{11}$	$n_{12}$	...	...	$n_{1p}$	$\sum_{j=1}^p n_{1j}$
$X_2$	$n_{21}$	$n_{22}$	...	...	$n_{2p}$	$\sum_{j=1}^p n_{2j}$
...	...	...	...	...	...	
...	...	...	...	...	...	
$X_k$	$n_{k1}$	$n_{k2}$	...	...	$n_{kp}$	$\sum_{j=1}^p n_{kj}$
	$\sum_{i=1}^k n_{i1}$	$\sum_{i=1}^k n_{i2}$			$\sum_{i=1}^k n_{ip}$	$\sum_{i=1}^k \sum_{j=1}^p n_{ij} = N$

# TEST CHI KWADRAT

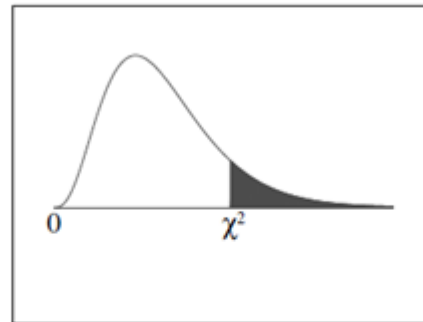
$H_0$ : Zmienne X i Y są niezależne

$H_1$ : Zmienne X i Y są zależne

Obszar odrzucenia:  $\chi^2 > \chi^2_\alpha$

Gdzie  $\chi^2_\alpha \sim (k-1)(p-1)$  df

Chi-Square Distribution Table



The shaded area is equal to  $\alpha$  for  $\chi^2 = \chi^2_\alpha$ .

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955

# TEST CHI KWADRAT

Porównanie licznosci obserwowanych i oczekiwanych

$$E_{ij} = \frac{(\text{row } i \text{ total}) * (\text{column } j \text{ total})}{\text{total sum}} = \frac{\sum_{j=1}^p n_{ij} \sum_{i=1}^k n_{ij}}{\sum_{i=1}^k \sum_{j=1}^p n_{ij}}$$

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \sum_{i=1}^k \sum_{j=1}^p \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

O - licznosci obserwowane

E - licznosci oczekiwane

# TEST CHI KWADRAT- TABELA KONTYNGENCJI 2 X 2

a	b
c	d

$$\chi^2 = \frac{(ad - bc)^2 N}{(a + b)(c + d)(a + c)(b + d)}$$

# POPRAWKA YATES'A

20 < N < 40, E < 5

$$\chi^2 = \sum \frac{(|O - E| - 0.5)^2}{E}$$

a	b
c	d

$$\chi^2 = \frac{\left(|ad - bc| - \frac{N}{2}\right)^2 N}{(a + b)(c + d)(a + c)(b + d)}$$



# BADANIE SIŁY ZALEŻNOŚCI

I. Yula's coefficient  $\phi = \sqrt{\frac{\chi^2}{N}}$

I. V-Cramer's coefficient  $V = \sqrt{\frac{\chi^2}{n(\min(k; p) - 1)}}$

II. Pearson's contingency coefficient  $C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$

III. Czuprow's coefficient  $t = \sqrt{\frac{\chi^2}{n\sqrt{(k-1)(p-1)}}$

Siła <0,1)  
Brak  
kierunku!

# ZADANIE 1

Dane dostępne są w pliku „Characteristics.sta”. Przeanalizuj zależności pomiędzy kolorem oczu a kolorem włosów.

1. Utwórz tabelę kontyngencji
2. Test Chi kwadrat. Poziom istotności 0.01.
3. Jeżeli zależność istnieje, określ jej siłę

# ZADANIE 1

Summary Frequency Table (Characteristics)

Marked cells have counts > 5

(Marginal summaries are not marked)

Eye Color	Hair Color brown	Hair Color red	Hair Color black	Hair Color blonde	Row Totals
blue	22	11	8	0	41
green	9	7	6	0	22
brown	15	4	13	5	37
All Grps	46	22	27	5	100

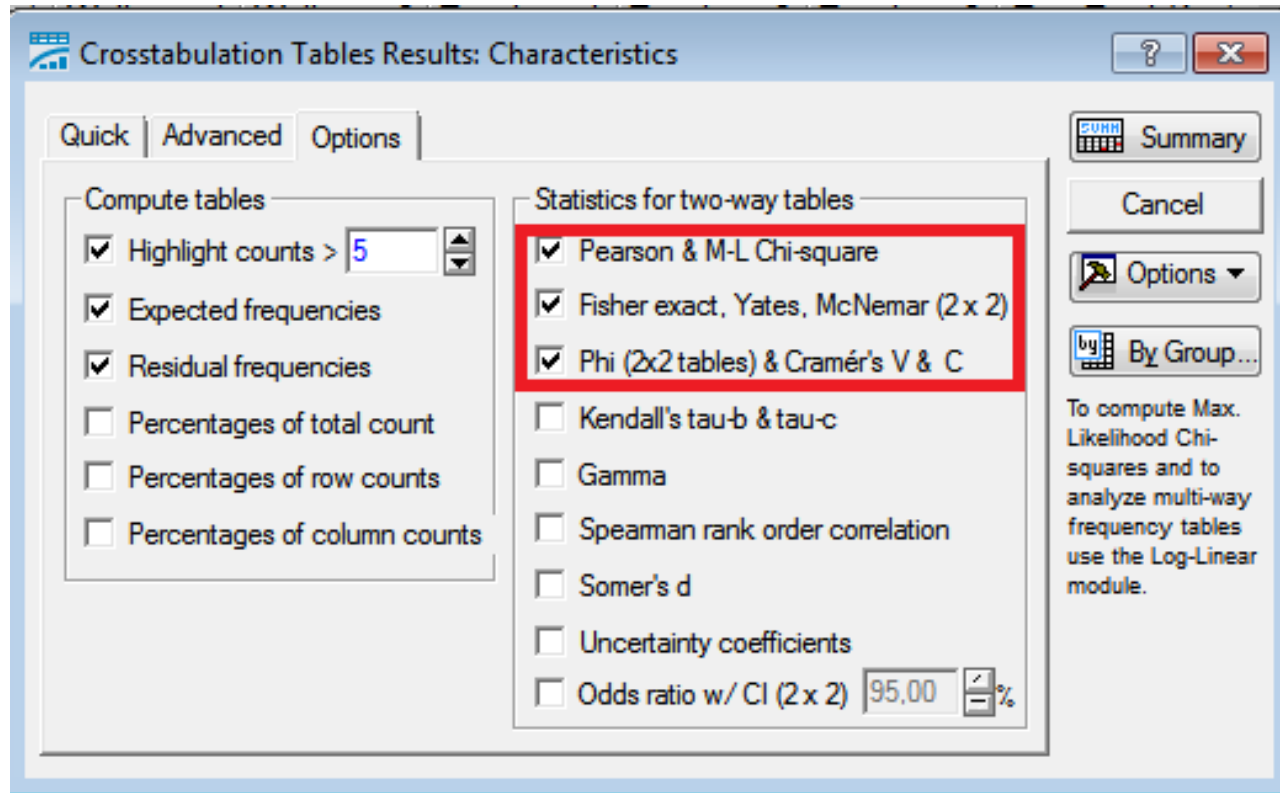
Summary Table: Expected Frequencies (Characteristics)

Marked cells have counts > 5

Pearson Chi-square: 14,6631, df=6, p=,023045

Eye Color	Hair Color brown	Hair Color red	Hair Color black	Hair Color blonde	Row Totals
blue	18,86000	9,02000	11,07000	2,050000	41,0000
green	10,12000	4,84000	5,94000	1,100000	22,0000
brown	17,02000	8,14000	9,99000	1,850000	37,0000
All Grps	46,00000	22,00000	27,00000	5,000000	100,0000

# ZADANIE 1



# ZADANIE 1

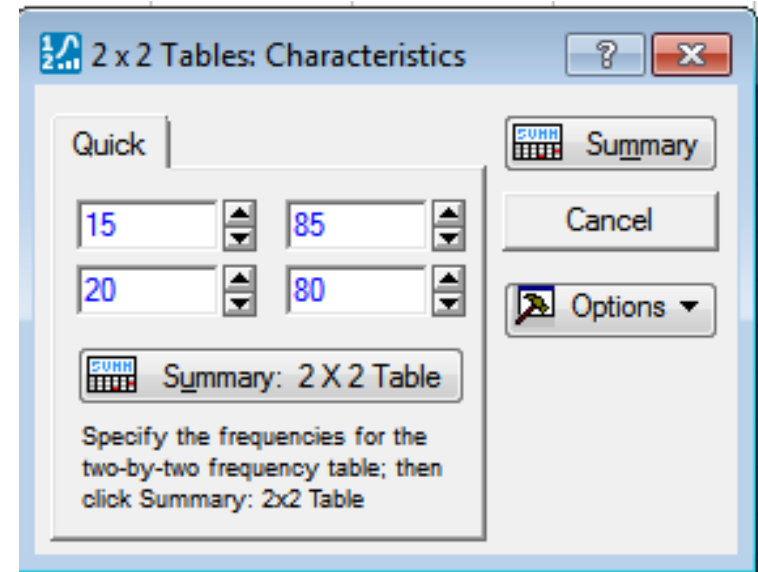
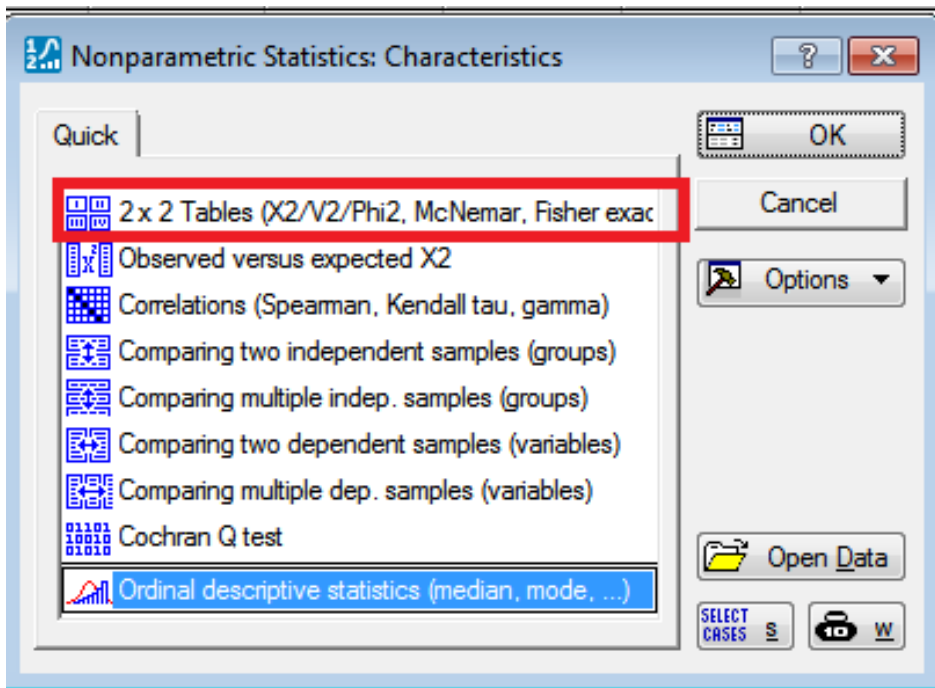
Statistic	Statistics: Eye Color(3) x Hair Color(4) (Characteristics)				
	Chi-square	df	p		
<b>Pearson Chi-square</b>	<b>14,66310</b>	df=6	p=,02305		
M-L Chi-square	16,43634	df=6	p=,01159		
Phi	,3829243				
Contingency coefficient	,3576030				
Cramér's V	,2707684				

## ZADANIE 2

Naukowcy badali zależność pomiędzy kolorem oczu o lewo- i prawo-ręcznością. Czy występuje dowód, jakoby gen związany z kolorem oczu był połączony z genem odpowiedzialnym za prawo- lub lewo-ręczność? Przyjmij poziom istotności 0,05.

		Lewo-, prawo- ręczność	
		Lewo-	Prawo-
Kolor oczu	Niebieskie	15	85
	Brązowe	20	80

# STATISTICA- TASK 3. (2)



# STATISTICA- TASK 3. (3)

	2 x 2 Table		
	Column 1	Column 2	Row Totals
Frequencies, row 1	15	85	100
Percent of total	7,500%	42,500%	50,000%
Frequencies, row 2	20	80	100
Percent of total	10,000%	40,000%	50,000%
Column totals	35	165	200
Percent of total	17,500%	82,500%	
Chi-square (df=1)	,87	p= ,3521	
V-square (df=1)	,86	p= ,3533	
<b>Yates corrected Chi-square</b>	<b>,55</b>	<b>p= ,4566</b>	
Phi-square	,00433		
Fisher exact p, one-tailed		p= ,2285	
two-tailed		p= ,4570	
McNemar Chi-square (A/D)	43,12	p= ,0000	
Chi-square (B/C)	39,01	p= ,0000	



# LITERATURE

McClave, J., Benson, G. & Sincich, T. (2008). Statistics for Business & Economics. Pearson International Edition, p. 553-594



DZIĘKUJĘ ZA UWAGĘ