

Politechnika Gdańska



Mariusz Kaszubowski
Katedra Statystyki
Wydział Zarządzania i Ekonomii
Politechnika Gdańska

Korelacja dla zmiennych jakościowych

Ćwiczenia 10

Podstawy statystyki

Podstawowe pojęcia

- statystyka chi-kwadrat
- tablice wielodzielne, kontyngencji, asocjacji
- współczynniki kontyngencji (T Czuprowa, V Cramera, C Pearsona, Yule'a)
- wartość krytyczna rozkładu chi-kwadrat
- stopnie swobody
- poziom ufności, poziom istotności

wartości teoretyczne dla tablic kontyngencji	$\hat{n}_{ij} = \frac{n_i \cdot n_j}{n} \quad , \quad n = \sum_i n_i = \sum_j n_j$
statystyka χ^2	$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}$
współczynnik T Czuprowa	$T = \sqrt{\frac{\chi^2}{n \cdot \sqrt{(k-1) \cdot (l-1)}}}$
współczynnik V Cramera	$V = \sqrt{\frac{\chi^2}{n \cdot \min\{l-1, k-1\}}}$
współczynnik C Pearsona	$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$
współczynnik C Pearsona z poprawką	$C_{popr} = \frac{C}{C_{\max}} \quad , \quad C_{\max} = \frac{1}{2} \cdot \left(\sqrt{\frac{k-1}{k}} + \sqrt{\frac{l-1}{l}} \right)$
współczynnik ϕ Yule'a dla tablic 2x2	$\phi = \sqrt{\frac{\chi^2}{n}}$

Zadanie 1

Poniższa tabela przedstawia zestawienie wyników ankiet przeprowadzonych wśród 400 dorosłych osób na temat regularnych ćwiczeń. Dokonaj analizy korelacji dla poniższej tabeli kontyngencji przyjmując poziom ufności 0,95.

wartości empiryczne			
	ćwiczenia regularne	brak ćwiczeń	suma
mężczyzna	43	173	216
kobieta	112	72	184
suma	155	245	400

Rozwiązanie

wartości teoretyczne

	ćwiczenia regularne	brak ćwiczeń	suma
mężczyzna			
kobieta			
suma			

n_{ij}	\hat{n}_{ij}	$n_{ij} - \hat{n}_{ij}$	$(n_{ij} - \hat{n}_{ij})^2$	$\frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}$
43,00				
173,00				
112,00				
72,00				
suma	X	X	X	

Rozwiązanie c.d.

wartości teoretyczne

	ćwiczenia regularne	brak ćwiczeń	suma
mężczyzna	83,7	132,3	216
kobieta	71,3	112,7	184
suma	155	245	400

n_{ij}	\hat{n}_{ij}	$n_{ij} - \hat{n}_{ij}$	$(n_{ij} - \hat{n}_{ij})^2$	$\frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}$
43,00	83,70	-40,70	1656,49	19,79
173,00	132,30	40,70	1656,49	12,52
112,00	71,30	40,70	1656,49	23,23
72,00	112,70	-40,70	1656,49	14,70
suma	X	X	X	70,24

$$df = 1 \cdot 1 = 1 \quad , \quad 1 - \alpha = 0,95 \Rightarrow \alpha = 0,05$$

$$\chi^2 = 70,24 \quad , \quad \chi_{0,05}^2 = 3,841$$

$$T = V = \varphi = 0,419 \quad , \quad C_{popr} = \frac{C}{C_{max}} = \frac{0,386}{0,707} \approx 0,546$$

Zadanie 2

Poniższa tabela przedstawia zestawienie badań powiązań pomiędzy poziomem inteligencji a paleniem papierosów.

Dokonaj analizy korelacji tych dwóch zmiennych przy poziomie ufności 0,99 oraz 0,90.

wartości empiryczne					
	poziom inteligencji				suma
	niski	średni	wysoki	b. wysoki	
niepalący	283	486	226	38	1033
palący okazyjnie	123	201	58	18	400
palący regularnie	100	147	64	8	319
suma	506	834	348	64	1752

Rozwiązanie

wartości teoretyczne					
	poziom inteligencji				
	niski	średni	wysoki	b. wysoki	suma
niepalący					
palący okazyjnie					
palący regularnie					
suma					

Rozwiązanie c.d.

wartości teoretyczne					
	poziom inteligencji				
	niski	średni	wysoki	b. wysoki	suma
niepalący	298,34	491,74	205,18	37,74	1033,00
palący okazyjnie	115,53	190,41	79,45	14,61	400,00
palący regularnie	92,13	151,85	63,36	11,65	319,00
suma	506,00	834,00	348,00	64,00	1752,00

Rozwiązanie c.d.

$n_{\bar{y}}$	$\hat{n}_{\bar{y}}$	$n_{\bar{y}} - \hat{n}_{\bar{y}}$	$(n_{\bar{y}} - \hat{n}_{\bar{y}})^2$	$\frac{(n_{\bar{y}} - \hat{n}_{\bar{y}})^2}{\hat{n}_{\bar{y}}}$
283,00				
486,00				
226,00				
38,00				
123,00				
201,00				
58,00				
18,00				
100,00				
147,00				
64,00				
8,00				
suma	X	X	X	

Rozwiązanie c.d.

$n_{\bar{y}}$	$\hat{n}_{\bar{y}}$	$n_{\bar{y}} - \hat{n}_{\bar{y}}$	$(n_{\bar{y}} - \hat{n}_{\bar{y}})^2$	$\frac{(n_{\bar{y}} - \hat{n}_{\bar{y}})^2}{\hat{n}_{\bar{y}}}$
283,00	298,34	-15,34	235,43	0,79
486,00	491,74	-5,74	32,91	0,07
226,00	205,18	20,82	433,27	2,11
38,00	37,74	0,26	0,07	0,00
123,00	115,53	7,47	55,87	0,48
201,00	190,41	10,59	112,13	0,59
58,00	79,45	-21,45	460,19	5,79
18,00	14,61	3,39	11,48	0,79
100,00	92,13	7,87	61,92	0,67
147,00	151,85	-4,85	23,55	0,16
64,00	63,36	0,64	0,41	0,01
8,00	11,65	-3,65	13,34	1,15
suma	X	X	X	12,60

Rozwiązanie c.d.

$$df = (3-1) \cdot (4-1) = 6 \quad , \quad 1-\alpha = 0,99 \Rightarrow \alpha = 0,01$$

$$\chi^2 = 12,60 \quad , \quad \chi_{0,05}^2 = 16,812$$

$$T = 0,054 \quad , \quad V = 0,06 \quad , \quad C_{popr} = \frac{C}{C_{\max}} = \frac{0,0845}{0,9856} \approx 0,0857 \quad , \quad \varphi = 0,0848$$

Zadanie 3

Wylosowano 1000 klientów PZU mieszkających w małych i większych miastach. Oceniali oni swoją sytuację materialną. Na podstawie danych z tabeli określ czy istnieje zależność pomiędzy miejscem zamieszkania a sytuacją materialną mieszkańców.

Sytuacja materialna / Miasta	Małe	Duże	Suma
Zła	175	505	
Średnia	35	245	
Dobra	20	20	
Suma			1000

Rozwiązanie

statystyka	Statystyka: sytuacja(3) x miasta(2) (jakosciowa)		
	Chi-kwadr.	df	p
Chi kwadrat Pearso	36,76886	df=2	p=,00000
Chi^2 NW	36,54252	df=2	p=,00000
Fi	,1917521		
Wsp. kontyngencji	,1883211		
V Craméra	,1917521		
tau b i c Kendalla	b=,0715361	c=,0576000	
D(X Y), D(Y X) Sommera	X Y=,08131	Y X=,06293	
Gamma	,1805643		
R rang Spearmana	,0727573	t=2,3046	p=,02139
Wsp. niepewności	X=,0244453	Y=,0338811	X Y=,02840