



Statistična analiza opisnih spremenljivk

doc. dr. Mitja Kos, mag. farm.

Katedra za socialno farmacijo

Univerza v Ljubljani- Fakulteta za farmacijo



Statistični znaki

Proučevane spremenljivke:

- statistični znaki ali statistične spremenljivke

Vrste spremenljivk oz. statističnih znakov:

- **Opisni oz. atributivni:**

Spremenljivke katerih vrednosti opisujemo z besedami

- **Nominalni:** npr. krvna skupina A, B, AB, 0

- **Ordinalni:** npr. stanje pacienta: slabo, srednje dobro, odlično

- **Kvantitativni, številčni oz. numerični:**

Spremenljivke katerih vrednosti izražamo s številkami,

- **Zvezni:** v glavnem merjenje, teoretično katerokoli vrednost znotraj določenega razmika npr. koncentracija glukoze

- **Nezvezni:** vrednosti imajo podane samo s celimi števili npr. število opravljenih izpitov

Urejanje atributivnih spremenljivk



- Kategorije: grupiranje enot v skupine
 - Spremenljivka:
 - Le nekaj vrednosti npr. spol, zakonski stan
 - Mnogo vrednosti + nejasne meje => klasifikacije npr. Mednarodna klasifikacija bolezni (MKB), anatomsko-terpavtska-kemična klasifikacija (ATC)
 - Frekvenca kategorije: koliko enot v kategoriji
-



Vsebina:

- Testiranje neodvisnosti/povezanosti
 - Testiranje normalnosti porazdelitve
 - Testiranje razlik v proporcijah
-



Ivana Kobilca, Kofetarica, 1888.

Testiranje neodvisnosti/povezanosti



- H_0 : med spremenljivkama ni povezave
 - H_1 : med spremenljivkama je povezava

 - Povezanost med budnostjo in pitjem kave?
-



Kontingenčna tabela 2x2

- Ali ohranja kava iz avtomata študente budne?
- Opazovane frekvence:

Placebo

	Kava brez kofeina	Kava s kofeinom	
Budni	8	30	38
Nebudni	65	42	107
	73	72	145



Kontingenčna tabela 2x2

- Pričakovane frekvence, če ni učinka:

	Placebo	Placebo	
	Kava brez kofeina	Kava s kofeinom	
Budni	?	?	38
Nebudni	?	?	107
	73	72	145



Kontingenčna tabela 2x2

- Pričakovane frekvence, če ni učinka:

	Placebo	Placebo	
	Kava brez kofeina	Kava s kofeinom	
Budni	$\frac{38 \times 73}{145} = 19,13$	$\frac{38 \times 72}{145} = 18,86$	38
Nebudni	$\frac{107 \times 73}{145} = 53,87$	$\frac{107 \times 72}{145} = 53,13$	107
	73	72	145

Opazovane / pričakovane frekvence



OPAZOVANE FREKVENCE

	Placebo	
	Kava brez kofeina	Kava s kofeinom
Budni	8	30
Nebudni	65	42

PRIČAKOVANE FREKVENCE

	Placebo	Placebo
	Kava brez kofeina	Kava s kofeinom
Budni	$\frac{38 \times 73}{145} = 19,13$	$\frac{38 \times 72}{145} = 18,86$
Nebudni	$\frac{107 \times 73}{145} = 53,87$	$\frac{107 \times 72}{145} = 53,13$



Hi kvadrat test: χ^2

□ Kontingenčna tabela

□ χ^2_{exp} izračunamo

$$\chi^2_{\text{exp}} = \sum_{i=1}^k \frac{(|f_u - f_p|)^2}{f_p}$$

$$\chi^2_{\text{exp}} = \sum_{i=1}^k \frac{(|f_u - f_p| - 0,5)^2}{f_p}$$

Yatesova korektura (pri 2*2)

□ f_u = ugotovljena frekvenca

□ f_p = pričakovana frekvenca

Testiranje neodvisnosti/povezanosti



- H0: med spremenljivkama ni povezave
- H1: med spremenljivkama je povezava

- Povezanost med budnostjo in pitjem kave?

$$\chi_{\text{exp}}^2 = \sum_{i=1}^k \frac{(|f_u - f_p| - 0,5)^2}{f_p}$$

$$\chi_{\text{exp}}^2 = 16,12$$



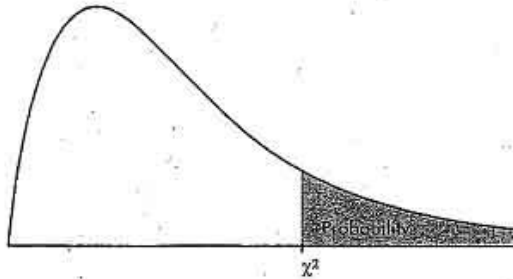
Pogoji hi kvadrat testa

- Vse pričakovane frekvence morajo biti najmanj 1
- 80% celic kontingenčne tabele mora imeti pričakovano frekvenco vsaj 5 ali več



Hi kvadrat tabela

χ^2 CRITICAL VALUES



$$\chi_{tab}^2(df=1; \alpha=0,05) = 3,84$$

Stopinje prostosti:

$$df = (s-1)(v-1)$$

s: število stolpcev
v: število vrstic

TABLE C: χ^2 CRITICAL VALUES

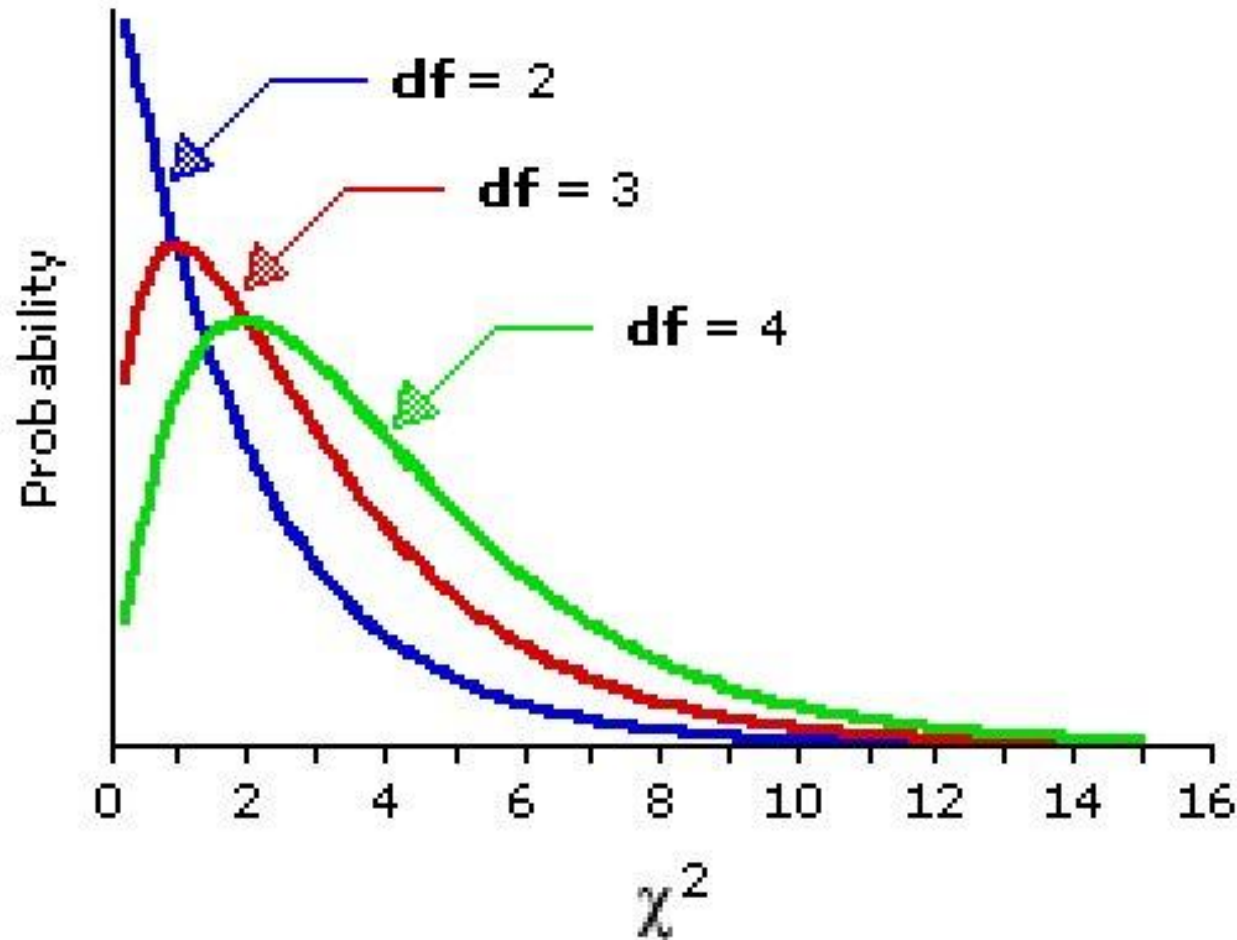
df	Tail probability p										
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.15	29.59

Glej na zgornji meji

df	.99595	.05 ↓	.025	.01
1	0.004	3.841	5.024	6.635
2	0.010	...	0.103	5.991	7.378	9.210



Hi kvadrat porazdelitev



Stopinje prostosti:

$$df = (s-1)(v-1)$$

s: število stolpcev
v: število vrstic



Testiranje hipotez

H_0 : med spremenljivkama ni povezave

H_1 : med spremenljivkama je povezava

$$\chi^2_{\text{exp}} = \sum_{i=1}^k \frac{(|f_u - f_p| - 0,5)^2}{f_p}$$

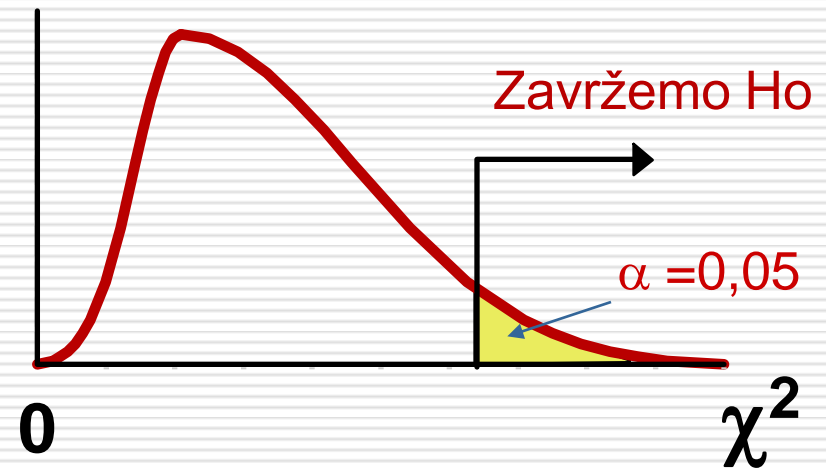
$$\chi^2_{\text{exp}} = 16,12$$

$$\chi^2_{\text{tab}(df=1; \alpha=0,05)} = 3,84$$

$$\chi^2_{\text{exp}} > \chi^2_{\text{tab}}$$

$$p < \alpha; \alpha = 0,05$$

H_0 zavržemo - spremenljivki sta povezani





Testiranje normalnosti

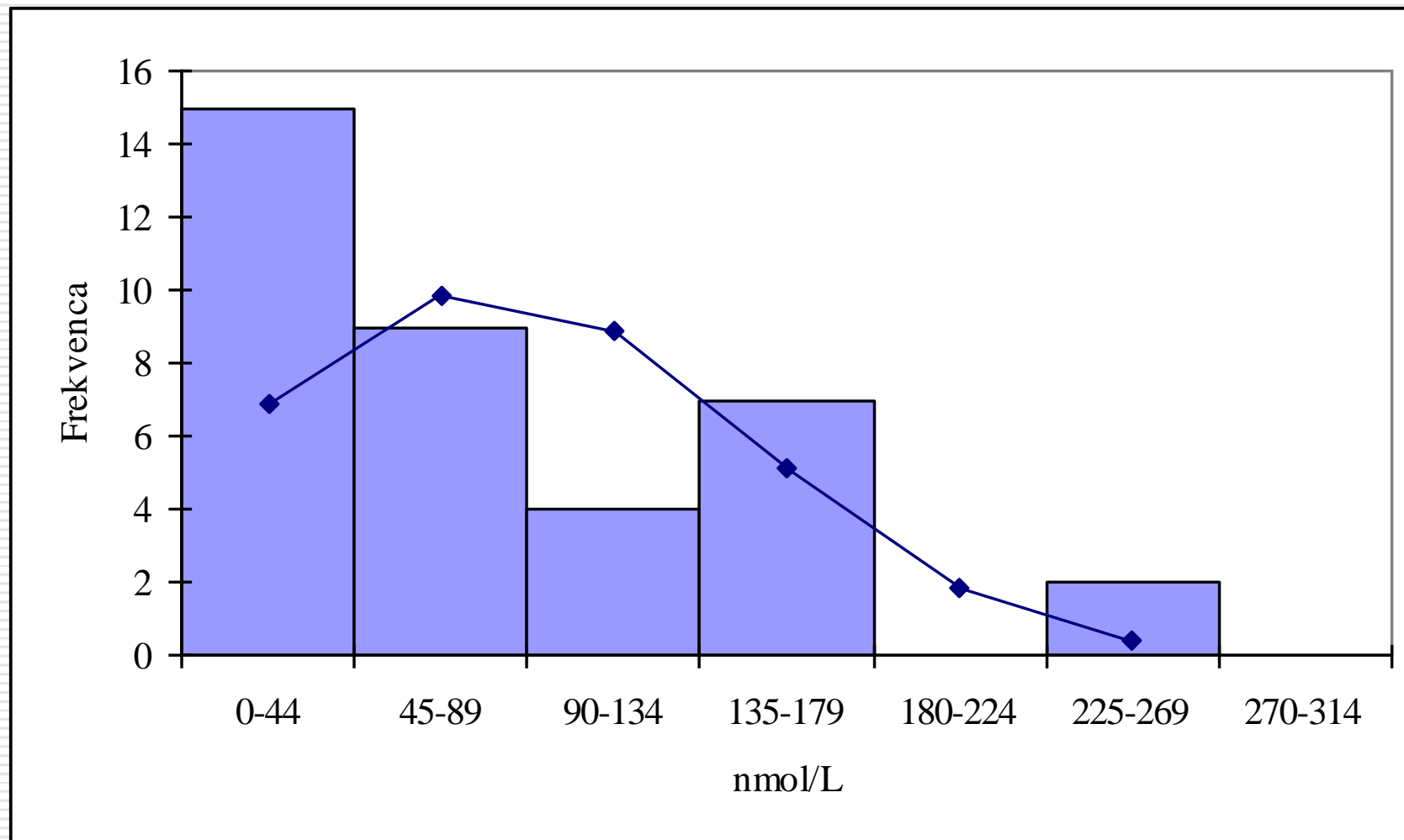
- H_0 : empirična porazdelitev je normalna
- H_1 : empirična porazdelitev ni normalna

Prilagajanje normalne porazdelitve empirični porazdelitvi

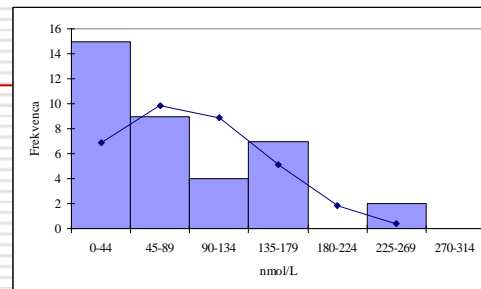


- Normalno porazdelitev (z vsemi karakteristikami) prilagoditi tako, da aritmetična sredina in standardna deviacija enaka tistima, ki ju ima empirična porazdelitev
-

Prilagojena normalna porazdelitev empirični



Prilagajanje normalne porazdelitve empirični porazdelitvi



$$\bar{x} = \frac{\sum_{i=1}^k f_i * \bar{x}_i}{n}$$

$$s = \sqrt{\frac{\sum_{i=1}^k f_i * (\bar{x}_i - \bar{X})^2}{n-1}}$$

$$z = \frac{\bar{x}_i - \bar{x}}{s}$$

$$F^0(x) = 1/2 + F^0(z)$$

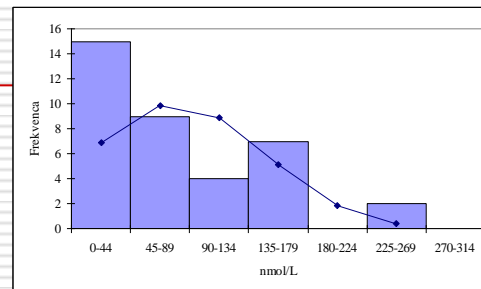
$$f_i' = F_{(x_1 < x < x_2)} = n * F^0_{(x_2 - x_1)}$$

Zap. št.	Razred	Frekvenco f_i	Sredina	Standardizirani odklon z_i	Kumulativna relativna frekvenco $F_0(x)$	Relativna frekvenco $F'(x_i)$	Absolutna frekvenco f_i'
1	0-44	15	22				
2	45-89	9	67				
3	90-134	4	112				
4	135-179	7	157				
5	180-224	0	202				
6	225-269	2	247				
7	270-314	0					

$$F'_{(x_i)} = F^0_{(x_2)} - F^0_{(x_1)} = 1/2 + F^0_{(z_2)} - 1/2 - F^0_{(z_1)}$$

$$F^0_{(x_2)} - F^0_{(x_1)} = F^0_{(z_2)} - F^0_{(z_1)}$$

Prilagajanje normalne porazdelitve empirični porazdelitvi



$$\bar{x} = \frac{\sum_{i=1}^k f_i * \bar{x}_i}{n}$$

$$s = \sqrt{\frac{\sum_{i=1}^k f_i * (\bar{x}_i - \bar{X})^2}{n-1}}$$

$$z = \frac{\bar{x}_i - \bar{x}}{s}$$

$$F^0(x) = 1/2 + F^0(z)$$

$$f_i' = F_{(x1 < x < x2)} = n * F^0_{(x2-x1)}$$

Zap. št.	Razred	Frekvenca fi	Sredina	Standardizirani odklon zi	Kumulativna relativna frekvenca Fo(x)	Relativna frekvenca F'(xi)	Absolutna frekvenca fi'
1	0-44	15	22	-1,23	0,1093	0,1853	6,86
2	45-89	9	67	-0,54	0,2946	0,2650	9,81
3	90-134	4	112	0,15	0,5596	0,2399	8,88
4	135-179	7	157	0,84	0,7995	0,1375	5,09
5	180-224	0	202	1,53	0,9370	0,0494	1,83
6	225-269	2	247	2,21	0,9864	0,0117	0,43
7	270-314	0		2,90	0,9981		

$$F'_{(xi)} = F^0_{(x2)} - F^0_{(x1)} = 1/2 + F^0_{(z2)} - 1/2 - F^0_{(z1)}$$

$$F^0_{(x2)} - F^0_{(x1)} = F^0_{(z2)} - F^0_{(z1)}$$



Testiranje razlik v proporcijah

- $H_0: p_1 = p_2 = p_3$
- $H_1: \text{Vsi } p \text{ niso enaki}$

Testiranje razlik v proporcijah

χ^2 za c proporcev



□ Ali imajo moški, ženske in otroci enako radi jagodni sladoled? Ali je pri stopnji tveganja 0,05 proporc tistih, ki imajo radi sladoled, enak?

Mnenje	Moški	Ženske	Otroci	Skupaj
Radi	12	63	65	140
Nimajo radi	8	17	35	60
Skupaj	20	80	100	200

Testiranje razlik v proporcijah

χ^2 za c proporcev



- $H_0: p_1 = p_2 = p_3$
 - $H_1: \text{Vsi } p \text{ niso enaki}$
 - $\alpha = 0,05$

 - $\chi^2_{\text{exp}}: ?$

 - $\chi^2_{\text{tab}}: ?$
-



Pričakovane frekvence

$$20 \cdot 140 / 200$$

$$80 \cdot 140 / 200$$

$$100 \cdot 140 / 200$$

$f_p \geq 5$ v
vseh
celicah



Mnenje	Moški	Ženske	Otroci	Skupaj
Radi	12/14	63/56	65/70	140
Nimajo radi	8/6	17/24	35/30	60
Skupaj	20	80	100	200

$$20 \cdot 60 / 200$$

$$80 \cdot 60 / 200$$

$$100 \cdot 60 / 200$$

 χ^2

Celica	f_u	f_p	$f_u - f_p$	$(f_u - f_p)^2$	$(f_u - f_p)^2/f_p$
1,1	12	14	-2	4	0,2857
1,2	63	56	+7	49	0,8750
1,3	65	70	-5	25	0,3571
2,1	8	6	+2	4	0,6667
2,2	17	24	-7	49	2,0417
2,3	35	30	+5	25	0,8333
Skupaj	200	200			5,0595



χ^2 za c proporcev

□ $H_0: p_1 = p_2 = p_3$

□ $H_1: \text{Vsi } p \text{ niso enaki}$

□ $\alpha = 0,05$

□ $\chi_{\text{exp}}^2 = \sum_{\text{vse celice}} \frac{(f_o - f_e)^2}{f_e} = 5,0595$

□ $\chi_{\text{tab}}^2: ?$

□ $df = (2 - 1)(3 - 1) = 2$



Hi kvadrat tabela

χ^2 CRITICAL VALUES

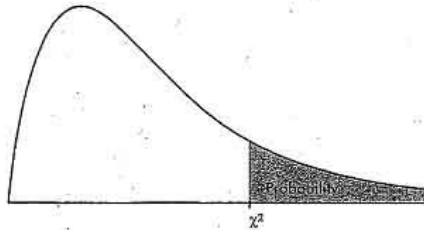


TABLE C: χ^2 CRITICAL VALUES

df	Tail probability <i>p</i>										
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.99	40.79

$$\chi^2_{tab}(df=2; \alpha=0,05) = 5,99$$

Stopinje prostosti:

$$df = (s-1)(v-1)$$

s: število stolpcev
v: število vrstic

Glej na zgornji meji

df	.99595	.05	.025	.01
1	0.004	3.841	5.024	6.635
2	0.010	...	0.103	5.991	7.378	9.210



χ^2 za c proporcev

□ $H_0: p_1 = p_2 = p_3$

□ $H_1: \text{Vsi } p \text{ niso enaki}$

□ $\alpha = 0,05$

□ $\chi_{\text{exp}}^2 = \sum_{\text{vsecelice}} \frac{(f_o - f_e)^2}{f_e} = 5,0595$

□ $\chi_{\text{tab}}^2: 5,99$

□ $df = (2-1)(3-1) = 2$

□ $p > \alpha$

