

ANALIZA TABLICA KONTINGENCIJE

TABLICA KONTINGENCIJE

- tablica koja u retcima i stupcima sadrži frekvencije atributivnih obilježja
- predstavlja empirijsku razdiobu frekvencija obilježja mjerenih nominalnom ili ordinalnom ljestvicom mjerenja

TABLICA S "JEDNIM ULAZOM" (1×k)

- opažanja su klasificirana samo po jednom obilježju

PRIMJER.

	GODINA STUDIJA			UKUPNO
	I	II	III	
BROJ STUDENATA	64	48	32	144

TABLICA S "DVA ULAZA" ($r \times k$)

- opažanja klasificirana po više atributa
- opažanja iz više uzoraka klasificirana po kategorijama jednog atributa

2×2 ...najjednostavnija tablica s "dva ulaza"

obilježje A	obilježje B		UKUPNO
	DA	NE	
DA	n_{11}	n_{12}	n_{1y}
NE	n_{21}	n_{22}	n_{2y}
UKUPNO	n_{x1}	n_{x2}	n_{xy}

- može se promatrati kao:
 - jedan uzorak (sa n_{xy} ispitanika)
 - dva uzorka (sa n_{1y} , n_{2y} ispitanika)

TABLICA S "DVA ULAZA" (r×k)

Stručna sprema	Spol		UKUPNO
	Muški	Ženski	
Nezavršena osnovna škola	4	27	31
Osnovna škola	12	35	47
Srednja škola	46	32	78
Viša škola/bakalaureat	12	25	37
Visoka škola/magisterij	52	18	70
Doktorat	11	4	15
UKUPNO	137	141	278

χ^2 TEST

- ocjena slaganja s poznatom razdiobom
- ocjena razlike razdiobe kategoričkog svojstva u nezavisnim uzorcima
- ocjena razlike dihotomnog svojstva u zavisnim uzorcima

χ^2 TEST ZA OCJENU SLAGANJA S POZNATOM RAZDIOBOM

- uz unaprijed poznatu razdiobu očekivanih frekvencija, test statistika

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

gdje je: O_i opažena frekvencija
 E_i očekivana frekvencija
 k broj kategorija

ima χ^2 razdiobu s

$df = k - 1 - m$ stupnjeva slobode

k ... broj kategorija

m ... broj parametara u modelu koje treba procijeniti

za normalnu razdiobu:

$$m = 2;$$

$$df = k - 1 - 2 = k - 3$$

za binomnu:

$$m = 1$$

$$df = k - 1 - 1 = k - 2$$

ako je zadana razdioba (ništa ne moramo računati iz podataka):

$$m = 0$$


$$df = k - 1$$

● **UZ**

H_0 ... nema razlike u razdiobi O_i i E_i

granični χ^2 za dani α i df

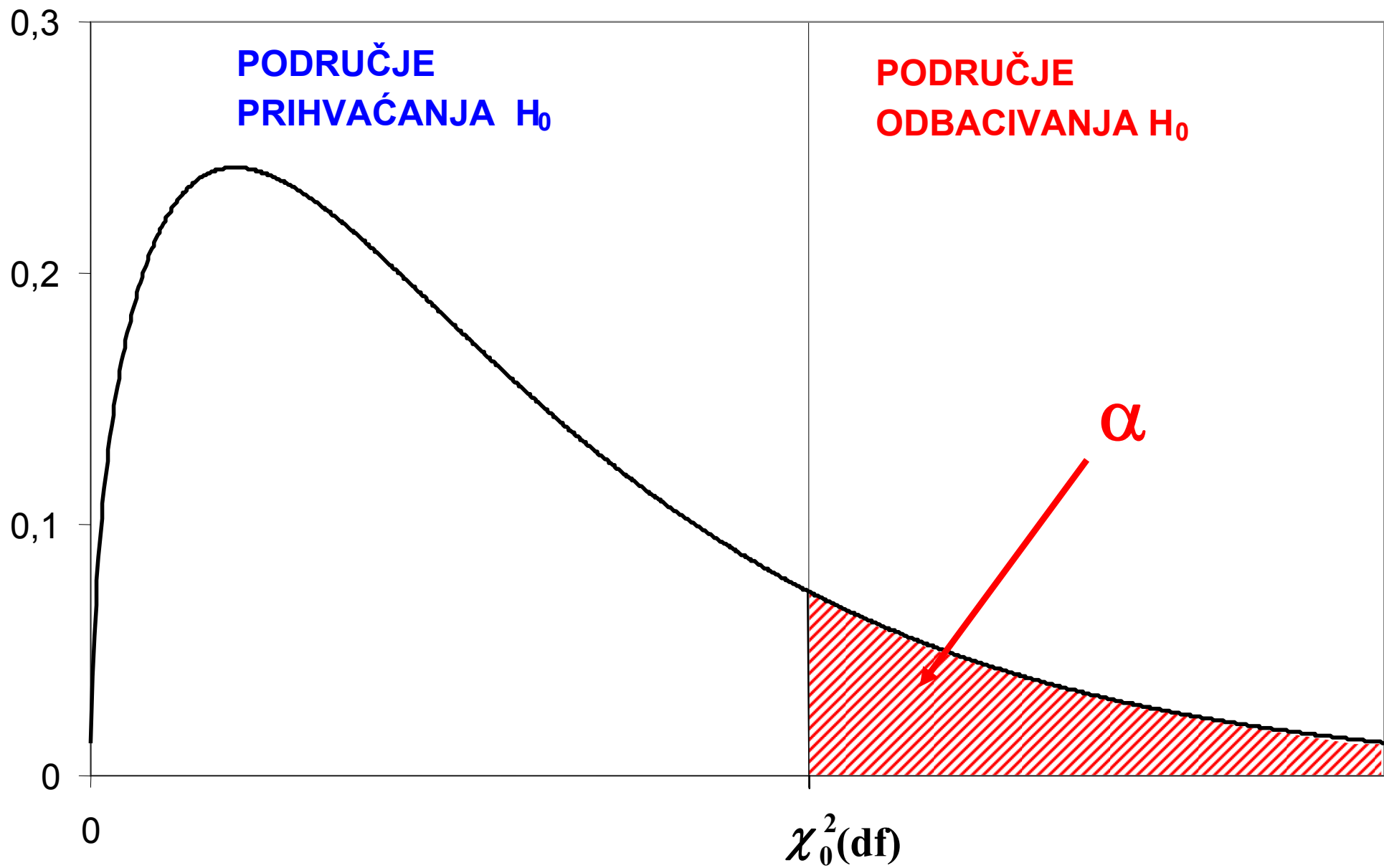
za

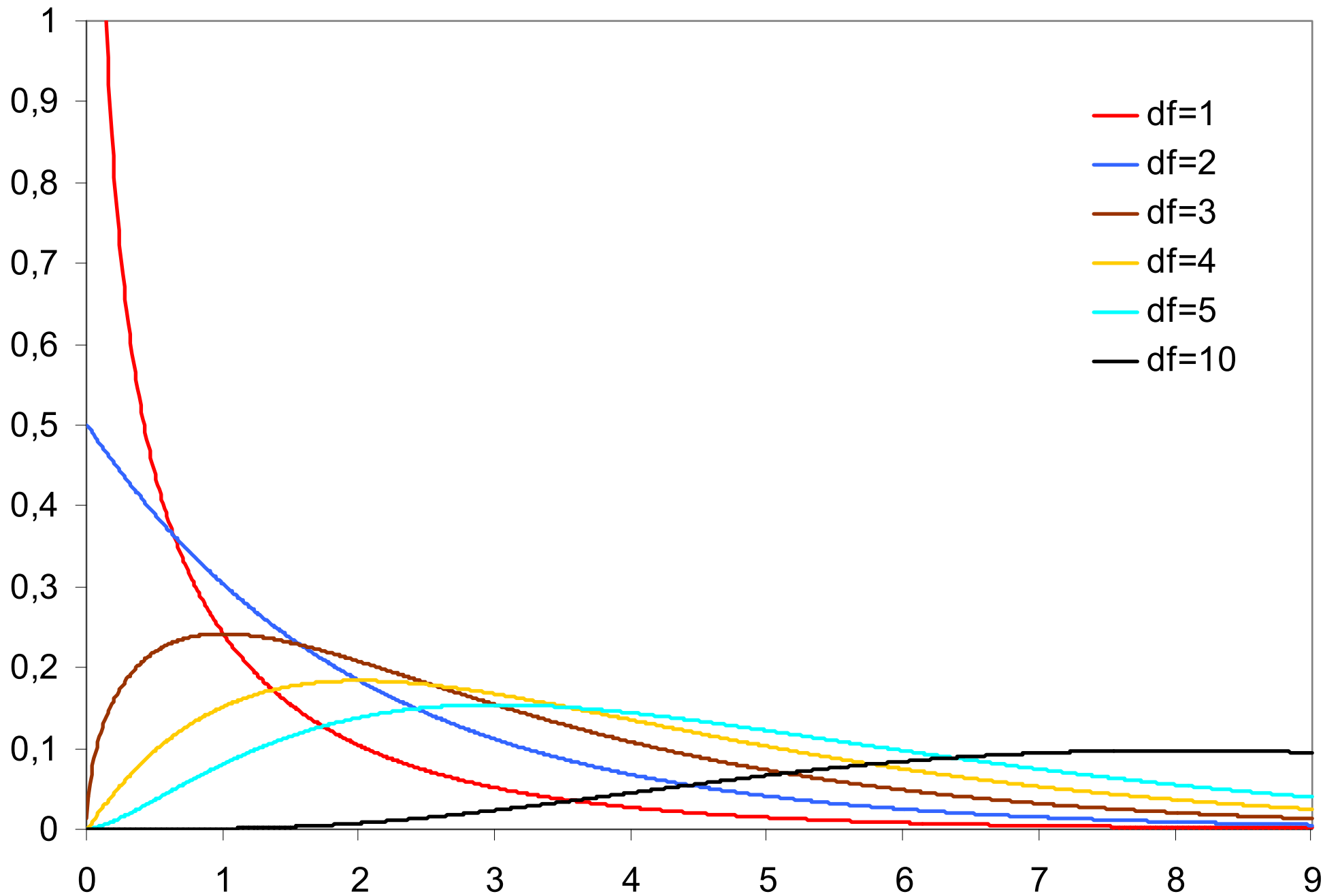

$$\chi^2 > \chi^2_{(1-\alpha)} \Rightarrow P(\chi^2) < P(\chi^2_{(1-\alpha)})$$

ODBACI H_0

$$\chi^2 < \chi^2_{(1-\alpha)} \Rightarrow P(\chi^2) > P(\chi^2_{(1-\alpha)})$$

PRIHVATI H_0





Križanjem dviju vrsta biljki dobivena je u sljedećoj generaciji ova razdioba opaženih genotipova:

genotip	opažene frekvencije
Aa	53
AA	23
aa	24

Odgovara li ova razdioba očekivanoj razdiobi 2:1:1 uz $\alpha=0.01$?

genotip	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
Aa	53	50	3	9	0.18
AA	23	25	-2	4	0.16
aa	24	25	-1	1	0.04
					0.38

$$\chi^2 = 0.38$$

$$df = 3 - 1 = 2$$

χ^2 RAZDIOBA

df \ p	0.99	0.98	0.95	0.9	0.8	0.7	0.5
1	6.635	5.412	3.841	2.706	1.642	1.074	0.455
2	9.210	7.824	5.991	4.605	3.219	2.408	1.386
3	11.345	9.837	7.815	6.251	4.642	3.665	2.366
4	13.277	11.668	9.488	7.779	5.989	4.878	3.357

za $df=2$: $\chi^2_{(1-\alpha)} = \chi^2_{(0.99)} = 9.210$

$$\chi^2 < \chi^2_{(0.99)} \Rightarrow P(\chi^2) > P(\chi^2_{(0.99)})$$

PRIHVATI H_0

χ^2 TEST ZA NEZAVISNE UZORKE

postupak:

- formirati tablicu kontingencije ($r \times k$)
- na osnovu postavljene hipoteze izračunati očekivane frekvencije
- test statistika dana je sa:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

gdje je: r broj redaka
 k broj stupaca

ima χ^2 razdiobu s

$df = (r-1) \cdot (k-1)$ stupnjeva slobode

VAŽNE NAPOMENE

- a) u tablicu smijemo unijeti **SAMO APSOLUTNE FREKVENCije**
- b) uzorci moraju biti **nezavisni**
- c) u 2x2 tablici:
 - ako je $N < 40$:
 - **NITI JEDNA** očekivana frekvencija **ne smije biti < 5**
 - preporuča se uvesti Yatesovu korekciju (umanjiti svaku razliku O-E prije kvadriranja)
- d) ako je u $r \times k$ tablici $E < 5$ u više od 20% polja, **NE MOŽEMO KORISTITI χ^2 TEST**

rješenje:

- spajanje susjednih razreda (frekvencija susjednih polja)
- Fisherov egzaktni test



Pri istraživanju djelovanja nekog cjepiva, opažena je sljedeća učestalost oboljenja kod određene grupe ljudi:

	cijepljeni	necijepljeni	ukupno
oboljeli	3	10	13
nisu oboljeli	144	117	261
ukupno	147	127	274

Postoji li povezanost između učestalosti bolesti i cijepljenja (je li učestalost bolesti jednaka kod cijepljenih i necijepljenih) uz $\alpha = 0.01$?

H_0 ... učestalost je ista kod cijepljenih i necijepljenih
iz $H_0 \Rightarrow$ proporcije oboljelih trebaju biti jednake u obje
skupine

zajednička proporcija oboljelih: $z_{po} = \frac{13}{274} = 0.0474$

zajednička proporcija zdravih: $z_{pz} = \frac{261}{274} = 0.9526$

E oboljelih: u grupi cijepljenih..... $147 * 0.0474 = 6.97$
u grupi necijepljenih.. $127 * 0.0474 = 6.02$

E zdravih: u grupi cijepljenih..... $147 * 0.9526 = 140.03$
u grupi necijepljenih.. $127 * 0.9526 = 120.98$

	cijepljeni	necijepljeni	ukupno
oboljeli	3 (6.97)	10 (6.02)	13
nisu oboljeli	144 (140.03)	117 (120.98)	261
ukupno	147	127	274

*

O_i	E_i	O_i - E_i	(O_i-E_i)²	(O_i-E_i)²/E_i
3	6.97	-3.97	15.7609	2.26
10	6.02	3.98	15.8404	2.63
144	140.03	3.97	15.7609	0.11
117	120.98	-3.98	15.8404	0.13
				$\chi^2 = 5.13$

Yates-ova korekcija:

O_i	E_i	(O_i - E_i)_{corr}	(O_i-E_i)_{corr}²	(O_i-E_i)_{corr}²/E_i
3	6.97	-3.47	12.0409	1.73
10	6.02	3.48	12.1104	2.01
144	140.03	3.47	12.0409	0.09
117	120.98	-3.48	12.1104	0.10
				$\chi^2 = 3.93$

$$\chi^2 = 3.93$$

za $\alpha = 0.01$, $df=1$: $\chi_0^2 = 6.635$

$$\chi^2 < \chi_0^2 \Rightarrow P > 0.01$$

\Rightarrow ne postoji povezanost između učestalosti bolesti i cijepljenja

χ^2 RAZDIOBA

df \ p	0.99	0.98	0.95	0.9	0.8	0.7	0.5
1	6.635	5.412	3.841	2.706	1.642	1.074	0.455
2	9.210	7.824	5.991	4.605	3.219	2.408	1.386
3	11.345	9.837	7.815	6.251	4.642	3.665	2.366
4	13.277	11.668	9.488	7.779	5.989	4.878	3.357

χ^2 TEST ZA ZAVISNE UZORKE (McNemarov test)

- testiranje značajnosti razlike (ili vjerojatnosti povezanosti) između podataka dobivenih na uzorcima parova

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c}$$

b, c ... frekvencije parova koji se ne slažu po prisutnosti obilježja

OBILJEŽJE A		UZORAK I	
		DA	NE
UZORAK II	DA	a	b
	NE	c	d

Skupina od 75 bolesnika praćena je tijekom 20 godina. Ispitanici su s obzirom na težinu svrstani u skupinu normalne i prekomjerne (overweight) težine.

TEŽINA		Nakon 20 godina		Ukupno
		Normalna	Prekomjerna	
Na početku	Normalna	26	14	40
	Prekomjerna	3	32	35
	Ukupno	29	46	75

Je li se težina ispitanika promijenila tijekom promatranog perioda uz $\alpha = 0.05$?

$$\chi^2 = \frac{(|14 - 3| - 1)^2}{14 + 3} = \frac{10^2}{17} = 5.88$$

$$df = 1, \quad \alpha = 0.05$$

$$\chi_0^2 = 3.84$$

$$\chi^2 > \chi_0^2 \Rightarrow p < 0.05$$