

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	IV	V	VI	
BROJ STUDENATA	64	48	32	28	18	15	205

TABLICA S "DVA ULAZA" (r×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 x 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):

$m = 0$
 $df = k - 1 - 0 = k - 1$

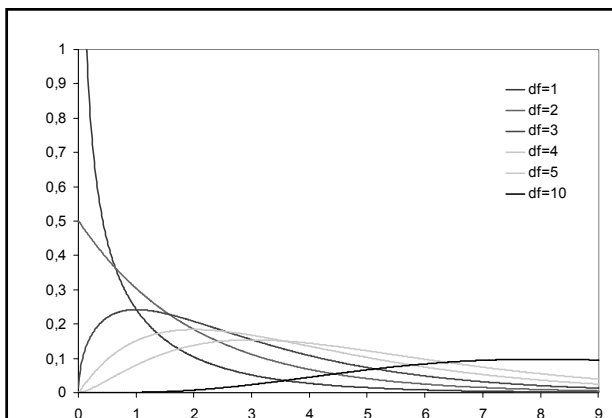
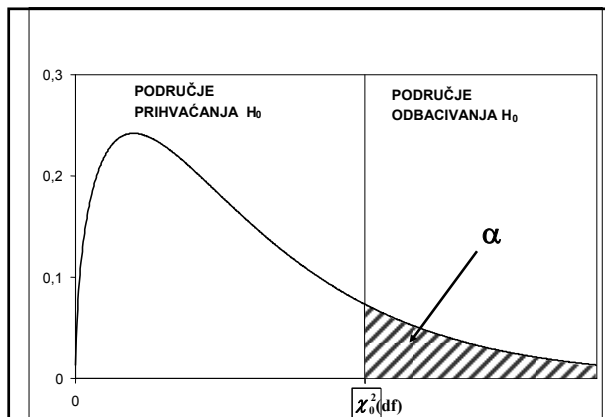
- uz

H_0 ... nema razlike u razdiobi O_i i E_i

granični χ^2 za dani α i df

za $\chi^2 > \chi^2_\alpha \Rightarrow P(\chi^2) < P(\chi^2_\alpha)$ ODBACI H_0

$\chi^2 < \chi^2_\alpha \Rightarrow P(\chi^2) > P(\chi^2_\alpha)$ 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

$\chi^2 = 0,38$ $df = 3 - 1 = 2$

α	0,01	0,02	0,05
1	6,635	5,412	3,841
2	9,210	7,824	5,991
3	11,345	9,837	7,815

za $df = 2$:
 $\chi^2_{(\alpha)} = \chi^2_{(0,01)} = 9,210$

$\chi^2 < \chi^2_{(0,01)} \Rightarrow P(\chi^2) > P(\chi^2_{(0,01)})$ **PRIHVATI H₀**

Kako dobiti točnu P vrijednost?

- kalkulator vjerojatnosti
 - Excel
 - WEB
 - specijalizirani programi za analizu podataka

MF Medicinski fakultet Osijek
 Katedra za medicinsku statistiku i medicinsku informatiku 14

Excel

MF Medicinski fakultet Osijek
 Katedra za medicinsku statistiku i medicinsku informatiku 15

http://vassarstats.net/

Calculators for Statistical Table Entries

z to P chi-square to P t to P r to P F to P

Fisher r-to-z transformation critical values of Q odds & log odds

For values of degrees of freedom (df) between 1 and 20, inclusive, this unit will calculate the proportion of the relevant sampling distribution that falls to the right of a particular value of chi-square. To proceed, enter the values of chi-square and df in the designated cells and click the «Calculate» button.

Chi-Square	df	P
0.38	2	0.826959

Calculate Reset

Entered values of df must be between 1 and 20, inclusive.

MF Medicinski fakultet Osijek
 Katedra za medicinsku statistiku i medicinsku informatiku 16

Statistica

MF 17

MedCalc – radni list

MF Medicinski fakultet Osijek
 Katedra za medicinsku statistiku i medicinsku informatiku 18

Izješavanje rezultata

"Nije uočena razlika između razdiobe opaženih genotipova i očekivane razdiobe 2:1:1 (χ^2 test, $P = 0,827$)."

ili

"Nije uočena razlika između razdiobe opaženih genotipova i očekivane razdiobe 2:1:1 ($\chi^2 = 0,380$, 2 stupnja slobode, $P = 0,827$)."

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 **2 x 2** tablici:

- **NITI JEDNA** očekivana frekvencija **ne smije biti < 5**
- Yatesova korekcija (umanjiti svaku razliku $O - E$ prije kvadriranja za 0,5) ili "N-1" Hi-kvadrat test (modificirana kratka formula)

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	5	10	15
Nisu oboljeli	159	117	276
Ukupno	164	127	291

Postoji li povezanost između učestalosti bolesti i cijepljenja (je li učestalost bolesti različita 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: $zpo = \frac{15}{291} = 0,051546$

ukupno oboljeli

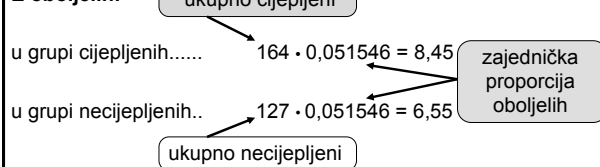
sveukupno ispitanika

zajednička proporcija zdravih: $zpz = \frac{276}{291} = 0,948454$

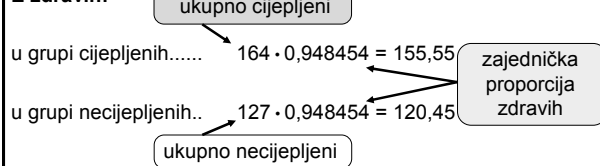
ukupno zdravi

sveukupno ispitanika

E oboljelih:



E zdravih:



	Cijepljeni		Necijepljeni		Ukupno
Oboljeli	5 (8,45)	10 (6,55)			15
Nisu oboljeli	159 (155,55)	117 (120,45)			276
Ukupno	164	127			291

*

O _i	E _i	O _i - E _i	(O _i -E _i) ²	(O _i -E _i) ² /E _i
5	8,45	-3,45	11,9025	1,41
10	6,55	3,45	11,9025	1,82
159	155,55	3,45	11,9025	0,08
117	120,45	-3,45	11,9025	0,10
				χ² = 3,41

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
5	8,45	-2,95	8,7025	1,03
10	6,55	2,95	8,7025	1,33
159	155,55	2,95	8,7025	0,06
117	120,45	-2,95	8,7025	0,07
				χ² = 2,49

χ² = 2,49
 za α = 0,01 i df = 1: **χ₀² = 6,635**
χ² < χ₀² ⇒ P > 0,01
⇒ ne postoji povezanost između učestalosti bolesti i cijepljenja

α \ k	0,01	0,02	0,05	χ ² RAZDIOBA
1	6,635	5,412	3,841	
2	9,210	7,824	5,991	
3	11,345	9,837	7,815	

točna P vrijednost.....

B4		fx = CHIDIST(B2;B3)	
	A	B	
1			
2	Vrijednost χ ²	3.41	
3	Broj stupnjeva slobode	1	
4	P vrijednost	0.064802	bez korekcije

B4		fx = CHIDIST(B2;B3)	
	A	B	
1			
2	Vrijednost χ ²	2.49	
3	Broj stupnjeva slobode	1	
4	P vrijednost	0.114572	s korekcijom

2 x 2 tablice – kratka formula za χ² test
 - za tablicu 2 x 2 sa sljedećim oznakama:

Obilježje A	Obilježje B		UKUPNO
	0	1	
0	a	b	m
1	c	d	n
UKUPNO	r	s	N

$$\chi^2 = \frac{N(ad - bc)^2}{mnr s}$$

"N - 1" χ² test

Obilježje A	Obilježje B		UKUPNO
	0	1	
0	a	b	m
1	c	d	n
UKUPNO	r	s	N

$$"N - 1" \chi^2 = \frac{(N - 1)(ad - bc)^2}{mnr s}$$

χ^2 test (kratka formula) i "N - 1" χ^2 test za podatke iz primjera

	Cijepljeni	Necijepljeni	Ukupno
Oboljeli	5 (a)	10 (b)	15 (m)
Nisu oboljeli	159 (c)	117 (d)	276 (n)
Ukupno	164 (r)	127 (s)	291 (N)

$$\chi^2 = \frac{N(ad - bc)^2}{mnr s} = \frac{291 \cdot (5 \cdot 117 - 10 \cdot 159)^2}{15 \cdot 276 \cdot 164 \cdot 127} = \frac{293917275}{86227920} = 3,409$$

$$\text{"N - 1" } \chi^2 = \frac{(N - 1)(ad - bc)^2}{mnr s} = \frac{290 \cdot (5 \cdot 117 - 10 \cdot 159)^2}{15 \cdot 276 \cdot 164 \cdot 127} = \frac{292907250}{86227920} = 3,397$$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 31

točna P vrijednost.....

B4		f _x =CHIDIST(B2;B3)
	A	B
1		
2	Vrijednost χ^2	3.409
3	Broj stupnjeva slobode	1
4	P vrijednost	0.064842

kratka formula

B4		f _x =CHIDIST(B2;B3)
	A	B
1		
2	Vrijednost χ^2	3.397
3	Broj stupnjeva slobode	1
4	P vrijednost	0.065315

"N - 1" χ^2

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 32

MedCalc – za tablice 2 x 2 koristi "N - 1" χ^2 test

Results

Chi-squared	3.397
DF	1
Significance level	P = 0.0653
Contingency coefficient	0.107

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 33

Izveštavanje rezultata

"Nije uočena povezanost učestalosti bolesti i cijepljenja (χ^2 test, P = 0,065)."

ili

"Nije uočena razlika razdiobe obolijevanja u skupinama cijepljenih i necijepljenih ispitanika (χ^2 test, P = 0,065)."

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 34

Izveštavanje rezultata

"Nije uočena povezanost učestalosti bolesti i cijepljenja ($\chi^2 = 3,397$, 1 stupanj slobode, P = 0,065)."

ili

"Nije uočena razlika razdiobe obolijevanja u skupinama cijepljenih i necijepljenih ispitanika ($\chi^2 = 3,397$, 1 stupanj slobode, P = 0,065)."

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 35

χ^2 TEST ZA ZAVISNE UZORKE (McNemarov test)

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

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

Yatesova korekcija

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

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 36

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 $\alpha = 0,05$?

McNemarov test – pomoću tablica

$$\chi^2 = \frac{((14-3|-1))^2}{14+3} = \frac{10^2}{17} = 5,882$$

df = 1, $\alpha = 0,05$

$$\chi_0^2 = 3,841$$

$$\chi^2 > \chi_0^2 \Rightarrow P < 0,05$$

α	0,01	0,02	0,05
k			
1	6,635	5,412	3,841
2	9,210	7,824	5,991
3	11,345	9,837	7,815

točna P vrijednost....

• za $\chi^2 = 5,882$ i 1 stupanj slobode:

	A	B
1		
2	Vrijednost χ^2	5.822
3	Broj stupnjeva slobode	1
4	P vrijednost	0.015827

MedCalc – McNemarov test (binomni egzaktni test)

McNemar test (paired proportions)

2x2 table

	Pos.	Neg.	
Pos.	26	14	53.3%
Neg.	3	32	46.7%
	38.7%	61.3%	

Results

Difference: 14.67%

95% CI: 2.96% to 20.94%

Exact probability: P = 0.0127

Comment:

Buttons: Clear, Test, Exit

Izveštavanje rezultata

"Težina ispitanika se promijenila tijekom promatranog perioda (McNemarov test, P = 0,013)."

ili

"Težina ispitanika se promijenila tijekom promatranog perioda (razlika proporcija 14,7 %, 95 % raspon pouzdanosti od 3,0 % do 20,9 %, McNemarov test, P = 0,013)."

Ocjena Procjene I Analiza Tablica Kontingencije

STANDARDNA POGREŠKA ARITMETIČKE SREDINE (SEM-standard error of the mean)

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

STANDARDNA POGREŠKA PROPORCIJE

$$s_p = \sqrt{\frac{p \cdot q}{n}}$$

RASPON POUZDANOSTI ARITMETIČKE SREDINE

$$\bar{x} - z \cdot s_{\bar{x}} \leq \mu \leq \bar{x} + z \cdot s_{\bar{x}}$$

POGREŠKA PROCJENE

RASPON POUZDANOSTI PROPORCIJE

$$p - z \cdot s_p \leq \Pi \leq p + z \cdot s_p$$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 43

ZADATAK 1:

U studiji objavljenoj 2004. godine izmjerena je težina 98 prijevremeno rođene djece i dobivena aritmetička sredina 1.31kg sa standardnom devijacijom od 0.42kg. Izračunajte pogrešku procjene aritmetičke sredine i raspon pouzdanosti uz pouzdanost od:

- 90%
- 95%
- Kolika je pogreška procjene za standardnu devijaciju od 0.20 kg i 0.60kg, uz pouzdanost 95%?
- Kolika je pogreška procjene uz pouzdanost od 95% i standardnu devijaciju od 0.6 za uzorke veličine 200 i 300 prijevremeno rođene djece?

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 44

ZADATAK 1:

n=98 $\bar{x} = 1.31$ s=0.42

- 90%
- 95%

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 45

ZADATAK 1:

n=98

- Kolika je pogreška procjene za standardnu devijaciju od 0.20 kg i 0.60kg?
 $s_1 = 0.20$
 $s_2 = 0.60$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 46

ZADATAK 1:

- Kolika je pogreška procjene uz pouzdanost od 95% i standardnu devijaciju od 0.6 za uzorke veličine 200 i 300 prijevremeno rođene djece?
 $n_1 = 200$
 $n_2 = 300$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 47

ZADATAK 2:

Farmaceutska tvrtka predlaže novi lijek za ublažavanje simptoma PMS-a. U prvim kliničkim istraživanjima lijek se pokazao učinkovit kod 7 od 10 žena.

- izračunajte pogrešku procjene proporcije populacije uz pouzdanost od 95%
- konstruirajte 95% raspon pouzdanosti za proporciju populacije
- izračunajte pogrešku procjene i konstruirajte 95% raspon pouzdanosti proporcije populacije za istu proporciju dobivenu iz uzorka od 100 ispitanica.

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 48

ZADATAK 2:

$n=10$ $p=7/10=0.7$ $q=1-0.7=0.3$

a) izračunajte pogrešku procjene proporcije populacije uz pouzdanost od 95%

b) konstruirajte 95% raspon pouzdanosti za proporciju populacije

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 49

ZADATAK 2:

c) izračunajte pogrešku procjene i konstruirajte 95% raspon pouzdanosti proporcije populacije za istu proporciju dobivenu iz uzorka od 100 ispitanica.

$n=100$ $p=0.7$ $q=1-0.7=0.3$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 50

ZADATAK 3:

Ispitivana je ćud (benignost/malignost) tumora mozga prema lokalizaciji. Od 100 bolesnika s benignim tumorom, tumor je bio kod 21 lociran na frontalnom, kod 28 na temporalnom, a kod ostalih na drugim režnjevima mozga. Od 50 bolesnika s malignim tumorom kod 19 se radilo o tumoru frontalnog, kod 2 temporalnog a kod 29 o tumoru ostalih režnjeva mozga. Ocijenite postoji li povezanost malignosti s lokalizacijom tumora na mozgu na razini značajnosti od 0.05.

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 51

OPAŽENE FREKVENCije	Frontalni	Temporalni	Ostali	Ukupno
Benigni				
Maligni				
Ukupno				

OČEKIVANE FREKVENCije	Frontalni	Temporalni	Ostali	Ukupno
Benigni				
Maligni				
Ukupno				

Formulas shown in the spreadsheet:

- $=A2-B2$
- $=C2^2$ ili $=POWER(C2;2)$
- $=D2/B2$
- $=CHIDIST(E8;2)$

The spreadsheet has columns A-E and rows 1-8. Row 1 labels are O_i , E_i , $O_i - E_i$, $(O_i - E_i)^2$, and $(O_i - E_i)^2 / E_i$. Row 8 contains the totals 150, 150, and the calculated χ^2 and P values.

MedCalc – χ^2 test za nezavisne uzorke

Tests-> Chi-squared test...

Chi-squared test

Test data	M1	M2	M3	M4	M5	M6
N1	21	28	51			
N2	19	2	29			
N3						
N4						
N5						

Results

Chi-squared	13.519
DF	2
Significance level	P = 0.0012
Contingency coefficient	0.288

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 54

Izveštavanje rezultata

"Ćud tumora povezana je s lokalizacijom (χ^2 test, $P=0,001$)."

ili

"Razdioba lokalizacija tumora u skupinama ispitanika s malignim i benignim tumorom se razlikuje (χ^2 test, $P=0,001$)."

ZADATAK 4:

Pri križanju dviju jedinki tipa Aa i tipa Bb teorijske vjerojatnosti pojavljivanja kombinacija jesu:

$$p(AB)=9/16$$

$$p(Ab)=3/16$$

$$p(aB)=3/16$$

$$p(ab)=1/16.$$

Ako su u 160 nezavisnih promatranja dobivene frekvencije 86, 35, 26, 13 testirati hipotezu da su podaci suglasni sa teorijskom raspodjelom uz nivo značajnosti 0.01.

tip	O _i	p _i	E _i =p _i *n	O _i -E _i	(O _i -E _i) ²	(O _i -E _i) ² /E _i
AB	86	9/16				
Ab	35	3/16				
aB	26	3/16				
ab	13	1/16				
n=	160	1	160			$\chi^2=$

H₀ nema razlike

k =

m =

df =

$\alpha = 0.01$

P=