

## Metodologija psiholoških istraživanja

### Obrada frekvencijski 4

19. novembar 2018

#### IV. Obrada podataka

##### A. Frekvencijski nacrti

- Univarijantni frekvencijski nacrti  
**UFN sa više od dve kategorije**
- Bivarijantni frekvencijski nacrti

## 1. Univarijantni frekvencijski nacrti (UFN)

(2) UFN sa više od dve kategorije: 3, 4, itd

- PRIMER: nacrt sa tri kategorije**
  - da li rodovi imenica u srpskom jeziku imaju različitu učestalost?
 

	muški	ženski	srednji	TOTAL
ROD	f1	f2	f3	N

 za prikaz toka statističke analize
- PRIMER: nacrt sa četiri kategorije**
  - da li bebe preferiraju crvenu, zelenu, žutu ili plavu boju?
  - da li pun mesec izaziva nasilno ponašanje?
    - da li su proporcije nasilnih incidenata veće u doba punog meseca nego u doba kada se vidi oko 1/4, 1/2 i 3/4 površine meseca?
- PRIMER: nacrti sa mnogo kategorija**
  - za koju partiju biste glasali kad bi danas bili izbori?

## 1. Univarijantni frekvencijski nacrti (UFN)

tok statističke analize prolazi kroz pomenutih pet faza

- Faza I: Podaci i deskriptivne mere**
  - broj kategorija je  $k=3$ , tj. muški, ženski i srednji rod
  - u uzorak je ušlo  $N=300$  imenica iz nekog teksta
  - opservirane frekvence:
    - muški rod:  $f_1=132$ , ženski rod:  $f_2=126$ , srednji rod:  $f_3=42$

	muški	ženski	srednji	TOTAL
ROD	132	126	42	N
- Faza II: nulta hipoteza, očekivane vrednosti, devijacije**
  - $H_0$ : u jeziku nema razlike između proporcija reči sva tri roda
  - formalno:  $H_0: p_1 = p_2 = p_3 = 1/3$
  - očekivane frekvence su jednake, i iznose po  $N/k = 300/3 = 100$ , dakle:
    - $f_1'=100, f_2'=100, f_3'=100$
  - reziduali (rezidualne frekvence, devijacije):  $d = f - f'$ 
    - $d_1 = 132 - 100 = 32; d_2 = 128 - 100 = 28; d_3 = 42 - 100 = -58$
  - uočiti: reziduali se sabiraju do nule:  $d_1 + d_2 + d_3 = 32 + 26 + (-58) = 0$

## 1. Univarijantni frekvencijski nacrti (UFN)

- Faza III: Test-statistik:  $\chi^2$** 
  - podsetimo se, u slučaju **dve** kategorije:
 
$$\chi^2 = \frac{d_1^2}{f_1'} + \frac{d_2^2}{f_2'}$$
 dva sabiraka
  - u slučaju **tri** kategorije:
 
$$\chi^2 = \frac{d_1^2}{f_1'} + \frac{d_2^2}{f_2'} + \frac{d_3^2}{f_3'}$$
 tri sabiraka
  - u slučaju **k** kategorija:
 
$$\chi^2 = \sum \frac{d_i^2}{f_i'} = \sum \frac{(f_i - f_i')^2}{f_i'}$$
 k sabiraka
  - napomena:** postoje i nešto složenije formule za  $\chi^2$  one sadrže i tzv. 'korekciju za kontinuitet', ali ona uglavnom nije potrebna
  - pregled računa:
 

faze	koraci	muški	ženski	srednji
I	empirija (f)	132	126	42
II	očekivanje (f')	100	100	100
	rezidual (d = f-f')	32	26	-58
	rezidual <sup>2</sup> (d <sup>2</sup> )	1024	676	3364
III	količnik (d <sup>2</sup> /f')	10.24	6.76	33.64
	zbir ( $\chi^2$ )	50.64		
  - uočiti: test ima **dva** stepena slobode: 3 (kat.) - 1 (izgubljen jer se zna zbir)
  - važi: ako u UFN ima **k** kategorija, broj stepeni slobode iznosi  $df = k - 1$ 
    - $\chi^2$ -distribucije su različite za različite  $df$  (za  $df = 1, df = 2, df = 3, \dots$ )

## 1. Univarijantni frekvencijski nacrti (UFN)

- Faza IV: p-vrednost**
  - to je verovatnoća da se nađe dobijeni ili veći  $\chi^2$ , ako je u populaciji  $H_0$  tačna
  - kompjuter izračunava da je u ovom slučaju  $p < 0.001$ , dakle,  $p < 0.05$
- Faza V: Odluka o statističkoj značajnosti**
  - statistički zaključak:** rezultat je statistički značajan
  - u izveštaju treba navesti i **istraživački zaključak** odn. **interpretaciju rezultata**, npr. 'u srpskom jeziku rodovi imenica imaju različite proporcije'

pregled računa:

istraživač samo unosi opservirane frekvence u statistički program (ovde su to: 132, 126, i 42) ceo ostali račun (očekivane vrednosti (f'), reziduali (d), količnici  $d^2/f'$ ,  $\chi^2$ , i p-vrednost) izvodi kompjuter

faze	koraci	muški	ženski	srednji
I	empirija (f)	132	126	42
II	očekivanje (f')	100	100	100
	rezidual (d = f-f')	32	26	-58
	rezidual <sup>2</sup> (d <sup>2</sup> )	1024	676	3364
III	količnik (d <sup>2</sup> /f')	10.24	6.76	33.64
	zbir ( $\chi^2$ )	50.64		
IV	p-vrednost	< 0.05		
V	značajnost	jeste SZ		
	zapis	$\chi^2(2) = 50.64, p < 0.05$		

## 1. Univarijantni frekvencijski nacrti (UFN)

- Vrste testova u UFN sa više od dve kategorije
  - zbog postojanja većeg broja kategorija, u ovakvim nacrtima postoji više vrsta poređenja, a ne samo jedno, kao kod UFN sa dve kategorije
  - razmotrićemo tri vrste poređenja, a za sve se koristi hi-kvadrat:
    - omnibus test:** poređenje svih kategorija zajedno
    - test kontrasta:** poređenje dve kategorije međusobno
    - parcijalni test:** poređenje više od dve ali manje od svih kategorija
- (1) **omnibus test:** poređenje svih kategorija zajedno
  - $H_0$ : proporcije **svih**  $k$  kategorija varijable A u populaciji su jednake
    - upravo taj test smo opisali u predhodnoj analizi
  - formalno:  $H_0: p_1 = p_2 = p_3 = \dots = N/k$
  - PRIMER:** proporcije imenica u srpskom jeziku
    - $H_0$ : ne postoje razlike u proporciji rodova, tj.  $p = 0.333\dots$  za sva tri roda
    - ako je omnibus test SZ, to znači da se gramatički rodovi u srpskom jeziku ne pojavljuju u jednakim proporcijama
    - međutim: još uvek ne znamo koji se rodovi **međusobno** statistički razlikuju
    - da bi to utvrdili, moramo koristiti dodatne, detaljnije testove

## 1. Univarijantni frekvencijski nacrti (UFN) 7

- (2) **test kontrasta**: poređenje dve kategorije međusobno
  - H0: proporcije dve kategorije varijable A u populaciji su jednake
    - kod nacrti sa dve kategorije test kontrasta i omnibus test su jedno isto
  - PRIMERI**: tri testa značajnosti razlika frekvenci imenica: muški i ženski rod (132-126), muški i srednji rod (132-42), ženski i srednji rod (126-42)
  - uočiti: ako je omnibus test značajan, ne znači nužno da je:
    - kategorija sa najvišom frekvencom značajno viša od svih ostalih
    - kategorija sa najnižom frekvencom značajno niža od svih ostalih
  - postojanje ovakvih razlika se mora testirati testovima kontrasta
    - u primeru: važi da f1-f2 nije SZ, dok f1-f3 i f2-f3 jesu SZ
- (3) **parcijalni test**: poređenje više od dve ali manje od svih kategorija
  - retko se koristi, nema standardni naziv
  - H0: proporcija *izvesnog broja* kategorija u populaciji je jednaka
  - PRIMER**: da li postoji razlika u proporcijama nasilnih incidenata u doba kada se vidi 1/4, 1/2 i 3/4 površine meseca? H0: nema razlike
- postoje i drugačiji testovi za UFN, koje ovde ne razmatramo
  - primer: *test normalne raspodele*

## 2. Bivarijantni frekvencijski nacrti (BFN) 8

- BFN: frekvencijski nacrti sa dve kategoričke varijable
  - varijabla A ima *a* kategorija, varij. B ima *b* kategorija, nacrt je tipa *axb*
- PRIMERI**:
  - nacrt 2x2: odnos pola i pušenja; odnos pola i rukosti
  - nacrt 3x3: odnos omiljenosti vrste muzike i uzrasta
- a. Organizacija podataka**

#	POL	PUŠENJE
1.	1	1
2.	2	1
3.	2	2
4.	...	...
...	...	...
- matrica podataka**: tipa objekti x varijable
  - kolone: dve (po jedna za svaku varijablu)
  - redovi: onoliko koliko ima objekata istraživanja
  - čelije: podaci (kategorije objekata)
    - članstvo u kategorijama se često *numerički kodira*
    - PRIMER**: muški: 1, ženski: 2; pušač: 1, nepušač: 2
- b. Deskriptivne statističke mere**
  - frekvence svih *situacija* (kombinacija kategorija)
  - frekvence kategorija za obe varijable, ukupna frekvencija N
  - odgovarajuće proporcije, procenti, šanse, i njihovih odnosi

## 2. Bivarijantni frekvencijski nacrti (BFN) 9

### c. Prikaz rezultata

- rezultati se prikazuju *numerički* i *grafički*
  - razmotrićemo detaljno samo prikaze rezultata nacrti tipa 2x2
- 1. numerički prikaz**
  - matrice frekvence, proporcija, i procenata
    - razmotrićemo prvo matrice (apsolutnih) frekvenci
  - osnovna matrica AB**
    - u BFN zove se takođe: *tabela kontingencije*
    - sadrži frekvence *situacija* (kombinacije kategorija)
    - ove frekvence se dobijaju prebrojavanjem, iz matrice podataka

**PRIMER:** rukost i pol

	desnoruki	levoruki
muškarci	765	80
žene	1168	70

	a <sub>1</sub>	a <sub>2</sub>
b <sub>1</sub>	a	b
b <sub>2</sub>	c	d

a=f<sub>11</sub>  
b=f<sub>12</sub>  
c=f<sub>21</sub>  
d=f<sub>22</sub>

4 frekvence situacija

## 2. Bivarijantni frekvencijski nacrti (BFN) 10

matrica AB upotpunjena matricama A, B, i T

	desnoruki	levoruki	POL
muškarci	765	80	845
žene	1168	70	1238
RUKOST	1933	150	2083

	a <sub>1</sub>	a <sub>2</sub>	B
b <sub>1</sub>	a	b	f <sub>1</sub> =a+b
b <sub>2</sub>	c	d	f <sub>2</sub> =c+d
A	f <sub>a</sub> =a+c	f <sub>a</sub> =b+d	N=a+b+c+d

- marginalne matrice A, B**
  - sadrže *marginalne* frekvence
  - frekvence kategorija
  - dobijaju se sabiranjem odgovarajućih redova i kolona
- matrica T**:
  - sadrži totalnu frekvencu N
  - ukupan broj objekata istraživanja
  - dobija se sabiranjem marginalnih frekvenci redova ili kolona

uočiti: empirijski su ustanovljene samo 4 brojke u matrici AB

ostalih 5 brojki su dobijene obradom empirijskih podataka

## 2. Bivarijantni frekvencijski nacrti (BFN) 11

- matrice *relativnih* frekvenci (proc. i propor.)
- u BFN, *relativne* frekvence se mogu računati ne samo na jedan (kao u UFN) već na **tri** načina:
  - način 1: u odnosu na ukupnu frekvencu N
  - način 2: u odnosu na marginalne frekvence varijable A
  - način 3: u odnosu na marginalne frekvence varijable B
- procenti u odnosu na N**
- procenti u odnosu na var. A**
- procenti u odnosu na var. B**
- uočiti: sve ove brojke su nastale obradom početne 4 brojke (30, 20, 60, 90)
- uvid u relativne frekvence može pružiti dublje razumevanje strukture rezultata
  - npr.: frekvencija žena pušača (60) je dva puta veća nego frekvencija muškaraca pušača (30)
  - ali: to *ne* znači da se može zaključiti da žene dva puta više *puše* od muškaraca!
  - naime: uvid u relativne frekvence pokazuje da je među ženama je *manje* pušača (40%) nego nepušača (60%), dok je među muškarcima je više pušača (60%) nego nepušača (40%)
  - razlog ovih *prividnih* neslaganja: u uzorku ima 3 puta više žena (150) nego muškaraca (50)
  - ali: među pušačima ima ih 2x više (67%:33%), a među nepušačima ima ih 4.5x više (82%:18%)

## 2. Bivarijantni frekvencijski nacrti (BFN) 12

- ovi rezultati mogu se izraziti i u obliku *proporcija* odn. *verovatnoća*:
  - način 1: **zajedničke** *verovatnoće*
  - način 2: **prva vrsta** *uslovnih* *verovatnoća*
  - način 3: **druga vrsta** *uslovnih* *verovatnoća*
- proporcije u odnosu na N**
- zajedničke verovatnoće**
- proporcije u odnosu na var. A**
- uslovne verovatnoće po kolonama**
- proporcije u odnosu na var. B**
- uslovne verovatnoće po redovima**

nač. 1	puš.	nep.	POL
muš.	0.15	0.10	0.25
žene	0.30	0.45	0.75
PUŠ.	0.45	0.55	1

nač. 1	P	N	POL
M	p(M&P)	p(M&N)	p(M)
Z	p(Z&P)	p(Z&N)	p(Z)
PUŠ.	p(P)	p(N)	1

**primer:** koja je verovatn. da je osoba i M i P? p(M&P)=0.15

nač. 2	puš.	nep.	POL
muš.	0.33	0.18	0.25
žene	0.67	0.82	0.75
PUŠ.	1	1	1

nač. 2	P	N	POL
M	p(M P)	p(M N)	p(M)
Z	p(Z P)	p(Z N)	p(Z)
PUŠ.	1	1	1

**primer:** ako je osoba P, koja je verovatn. da je M? p(M|P)=0.33

nač. 3	puš.	nep.	POL
muš.	0.60	0.40	1
žene	0.40	0.60	1
PUŠ.	0.45	0.55	1

nač. 3	P	N	POL
M	p(P M)	p(N M)	1
Z	p(P Z)	p(N Z)	1
PUŠ.	p(P)	p(N)	1

**primer:** ako je osoba M, koja je verovatn. da je P? p(P|M)=0.60

## 2. Bivarijatni frekvencijski nacrti (BFN) 13

### 2. grafički prikaz

- postoje različite 2D i 3D varijante grafičkih prikaza
- kao i kod prikaza UFN, grafikoni BFN sadrže brojne **grafičke i tekstualne elemente**
  - grafički: štapići, osa mere, ose varijabli, mreža
  - tekstualni: oznake varijabli i njenih kategorija, oznake mere iernih jedinica, naslov
- prvo ćemo razmotriti 3D prikaze apsolutnih frekvenci

	puš.	nep.	POL
muš.	30	20	50
žene	60	90	150
PUS.	90	110	200

- opisaćemo prikaz podataka iz 2D glavne matrice AB
  - frekvence a, b, c, d
- podaci iz prostih i marginalnih 1D matrica prikazuje se prema ranije opisanim principima za UFN

## 2. Bivarijatni frekvencijski nacrti (BFN) 14

### 3D (tro-osni) grafikoni

mera

- vertikalna osa: mera prebrojavanja (npr. frekvencija)
- dve horizontalne ose: dve varijable u nacrtu, A i B
  - u ovim grafikonima, uvođenje dubine ima smisla!
- postoje dve varijante, zavisno od toga koja od dve varijable je prikazana na kojoj horizontalnoj osi

Polne razlike u pušenju

Polne razlike u pušenju

- u ovim primerima su različitim bojama označene različite kategorije jedne varijable
  - varijabla Pol: plavo: podaci za muškarce; crveno: podaci za žene
- međutim, to nije nužno kod ovakvih grafikona

## 2. Bivarijatni frekvencijski nacrti (BFN) 15

### 2D (dvo-osni) grafikoni sa svim naprednim stubićima

Polne razlike u pušenju

Polne razlike u pušenju

- jedna osa: osa mere prebrojavanja
  - u primerima: mera je frekvencija, osa mere je vertikalna (ali može biti i horizont.)
- druga osa: osa jedne od dve varijable (bilo A, bilo B)
  - na osi su naznačene kategorije osne varijable
  - levo: osna var. je pol (mušk., žene); desno: osna var. je pušenje (puš., nep.)
- druga varijabla: mora se naznačiti razlikama u boji (ili na neki drugi način)
  - te razlike su objašnjene u legendi grafikona
  - mala tabela koja označava koja boja odgovara kojoj kategoriji druge varijable
  - levo: druga var. je Pušenje (puš., nep.); desno: druga var. Pol (mušk., žene)

## 2. Bivarijatni frekvencijski nacrti (BFN) 16

### 2D (dvo-osni) grafikoni sa naprednim i nastavljenim stubićima

- važe isti principi prikaza (ose, označavanje, legende, itd.)
  - u narednim primerima mera prebrojavanja nije frekvencija nego procent

način br. 2

	puš.	nep.
muš.	33%	18%
žene	67%	82%
PUS.	100%	100%

Procentualne polne razlike u pušenju

način br. 3

	puš.	nep.	POL
muš.	60%	40%	100%
žene	40%	60%	100%

Procentualne polne razlike u pušenju

u ovoj verziji: grafički naglasak je na razlikama između pušača i nepušača  
npr. lako se vidi da među nepušačima ima više žena nego među pušačima

u ovoj verziji: grafički naglasak je na razlikama između muškaraca i žena  
npr. lako se vidi da među muškarcima ima više pušača, a među ženama više nepušača

## 2. Bivarijatni frekvencijski nacrti (BFN) 17

### dalji primeri rezultati američkih predsedničkih izbora 2016. i 2012.

2016

2012

## 2. Bivarijatni frekvencijski nacrti (BFN) 18

### rezultati američkih predsedničkih izbora 2016. i 2012. prikaz: geografski grafikoni

izbori 2016

Klinton (dem.), Trump (rep.)

po saveznm državama

izbori 2012

Obama (dem.), Romni (rep.)

po okruzima

## 2. Bivarijatni frekvencijski nacrti (BFN) 19

rezultati izlaznih anketa američkih predsedničkih izbora  
 mera prebrojavanja: procenti  
 tip grafikona: kombinacija nastavljenih i naporednih štapića  
 varijable: A: kandidat  
 B: pol glasača

izbori 2016 kandidati: Klinton (dem.), Trump (rep.), Ostali

Male	41%	53%
Female	54%	42%

izbori 2012 kandidati: Obama (dem.), Romni (rep.), Ostali

Men: 47%	40%	52%
Women: 53%	55%	44%

## 2. Bivarijatni frekvencijski nacrti (BFN) 20

rezultati izlaznih anketa američkih predsedničkih izbora  
 mera prebrojavanja: procenti  
 tip grafikona: kombinacija nastavljenih i naporednih štapića  
 varijable: A: kandidat  
 B: uzrast glasača

izbori 2016 kandidati: Klinton (dem.), Trump (rep.), Ostali

18-29	50%	37%
30-44	50%	42%
45-64	44%	53%
65 and over	45%	53%

izbori 2012 kandidati: Obama (dem.), Romni (rep.), Ostali

18-29: 19%	60%	37%
30-44: 27%	52%	46%
45-64: 38%	47%	51%
65 and over: 16%	44%	56%

## 2. Bivarijatni frekvencijski nacrti (BFN) 21

rezultati izlaznih anketa američkih predsedničkih izbora 2016  
 kandidati: Klinton (dem.), Trump (rep.), Ostali

mesto stanovanja	City over 50,000	45%	50%	35%
	Suburbs	42%	50%	48%
	Small city or rural	34%	52%	42%
obrazovanje	High school or less	45%	51%	38%
	Some college/associate degree	43%	52%	41%
	College graduate	49%	45%	45%
rasa	White	37%	58%	38%
	Black	65%	29%	8%
	Hispanic/Latino	60%	29%	11%
rasa i obrazovanje	White college graduates	23%	67%	40%
	White without a college degree	71%	23%	6%
	Nonwhite without a college degree	78%	20%	2%
religioznost	Once a week or more	40%	56%	35%
	A few times a month	46%	46%	38%
	A few times a year	48%	47%	39%
	Never	62%	51%	28%

nedostaju marginalne frekvence!

## 2. Bivarijatni frekvencijski nacrti (BFN) 22

primer lošeg prikaza istih rezultata  
 rezultati glasanja na američkim predsedničkim izborima 2012

- prikaz je natpran i veoma nepregledan
- nije jasno razdvojena informacija o polu, uzrastu, rasi, prihodu ...
- prikaz je neprijatan za čitanje
- ometajuća variranja veličine slova u tekstu
- nepotrebno šaren tekst (tri boje)
- koristi se nepotrebna pozadina
- ne sadrži relevantnu informaciju
- otežava čitanje teksta

## 2. Bivarijatni frekvencijski nacrti (BFN) 23

rezultati glasanja za Brexit, 2016

How different age groups voted

18-24	77%	23%
25-34	70%	30%
35-44	48%	52%
45-54	50%	50%
55-64	57%	43%
65+	60%	40%

nedostaju marginalne matrice

## 2. Bivarijatni frekvencijski nacrti (BFN) 24

### d. Struktura rezultata

- znatno složenija (i zanimljivija) nego zbir dva UFN

	primer UFN		
PUS.	90	110	200

(1) Nacrti tipa 2x2

- podsetimo se: u UFN sa dve kategorije možemo uočiti šanse, tj. porediti frekvence u matrici A
  - f1:f2, na pr. pušači:nepušači, 90:110 = 9:11
- ako proširimo UFN u BFN tipa 2x2, dodavanjem var. B, možemo te šanse porediti u tri 1D matrice:
  - u marginalnoj matrici A: fa1:fa2 (npr. šansa puš:nep, i to ukupno, 90:110 = 9:11)
  - u prostoj matrici A/b1: a:b (npr. šansa puš:nep, i to kod muškaraca, 30:20 = 3:2)
  - u prostoj matrici A/b2: c:d (npr. šansa puš:nep, i to kod žena, 60:90 = 2:3)
- analogni odnosi se mogu porediti u kolonama (npr. šansa m:ž kod puš., itd)
  - možete uraditi za domaći zadatak!
- rezultate ćemo analizirati s obzirom na tri pitanja o njihovoj strukturi
  - na svako od tih pitanja moguća su dva odgovora ('da' i 'ne')
  - stoga postoji 2x2x2 = 8 osnovnih tipova struktura rezultata u BFN
  - oni će biti predstavljeni preko osam slučajeva (osam primera rezultata)
  - kasnije ćemo razmotriti statističku značajnost rezultata

	primer BFN		
	puš.	nep.	POL.
muš.	30	20	50
žene	60	90	150
PUS.	90	110	200

## 2. Bivarijantni frekvencijski nacrti (BFN) 25

1. pitanje: da li postoji ili ne postoji razlika frekvenci kategorija varijable A?

- dve mogućnosti: da *ne postoji* i da *postoji*
- (a) razlika *ne postoji*: slučajevi 1, 2, 3, 4: *jednake* frekvence kategorija a1 i a2

puš.	nep.	POL.	šanse	
muš.	50	50	1:1	
žene	50	50	1:1	
PUS.	100	100	200	1:1

puš.	nep.	POL.	šanse	
muš.	60	60	120	1:1
žene	40	40	80	1:1
PUS.	100	100	200	1:1

puš.	nep.	POL.	šanse	
muš.	60	40	100	3:2
žene	40	60	100	2:3
PUS.	100	100	200	1:1

puš.	nep.	POL.	šanse	
muš.	80	40	120	2:1
žene	20	60	80	1:3
PUS.	100	100	200	1:1

- učiti: veoma *različiti* ishodi u osnovnoj matrici AB dovode do *istih* ishoda u marginalnoj matrici A
- dakle: samo na osnovu ishoda u A nemamo punu informaciju o ishodu u AB

(b) razlika *postoji*: slučajevi 5, 6, 7, 8: *različite* frekvence kategorija a1 i a2

puš.	nep.	POL.	šanse	
muš.	60	40	100	3:2
žene	60	40	100	3:2
PUS.	120	80	200	3:2

puš.	nep.	POL.	šanse	
muš.	90	60	150	3:2
žene	30	20	50	3:2
PUS.	120	80	200	3:2

puš.	nep.	POL.	šanse	
muš.	70	30	100	7:3
žene	50	50	100	1:1
PUS.	120	80	200	3:2

puš.	nep.	POL.	šanse	
muš.	100	20	120	5:1
žene	20	60	80	1:3
PUS.	120	80	200	3:2

- učiti: iste napomene kao gore, za slučajeve 1, 2, 3, i 4
- ovakvih numerički različitih matrica ima naravno *bezbroj*
- međutim, s obzirom na analizu struktura rezultata, postoji samo *8 različitih tipova* ovakvih matrica (ovde predstavljenih)

## 2. Bivarijantni frekvencijski nacrti (BFN) 26

2. pitanje: da li postoji ili ne postoji razlika frekvenci kategorija varijable B?

- dve mogućnosti: da *ne postoji* i da *postoji*, prikazane preko *istih* 8 slučajeva
- (a) razlika *ne postoji*: slučajevi 1, 3, 5, 7: *jednake* frekvence kategorija b1 i b2

puš.	nep.	POL.	šanse	
muš.	50	50	100	1:1
žene	50	50	100	1:1
PUS.	100	100	200	1:1

puš.	nep.	POL.	šanse	
muš.	60	40	100	3:2
žene	60	40	100	3:2
PUS.	120	80	200	3:2

puš.	nep.	POL.	šanse	
muš.	70	30	100	7:3
žene	50	50	100	1:1
PUS.	120	80	200	3:2

- učiti: veoma *različiti* ishodi u osnovnoj matrici AB dovode do *istih* ishoda u marginalnoj matrici B
- dakle: samo na osnovu ishoda u B nemamo punu informaciju o ishodu u AB

(b) razlika *postoji*: slučajevi 2, 4, 6, 8: *različite* frekvence kategorija b1 i b2

puš.	nep.	POL.	šanse	
muš.	60	60	120	1:1
žene	40	40	80	1:1
PUS.	100	100	200	1:1

puš.	nep.	POL.	šanse	
muš.	90	60	150	3:2
žene	30	20	50	3:2
PUS.	120	80	200	3:2

puš.	nep.	POL.	šanse	
muš.	80	40	120	2:1
žene	20	60	80	1:3
PUS.	100	100	200	1:1

puš.	nep.	POL.	šanse	
muš.	100	20	120	5:1
žene	20	60	80	1:3
PUS.	120	80	200	3:2

- učiti: iste napomene kao gore, za slučajeve 1, 3, 5, i 7

## 2. Bivarijantni frekvencijski nacrti (BFN) 27

3. pitanje: da li postoji ili ne postoji *korelacija* varijabli A i B u matrici AB?

- (a) korelacija *ne postoji*: slučajevi 1, 2, 5, 6 (ovi slučajevi čine *Grupu I*)  
šanse su *jednake*:  $a:b = c:d = fa1:fa2$

puš.	nep.	POL.	šanse	
muš.	50	50	100	1:1
žene	50	50	100	1:1
PUS.	100	100	200	1:1

puš.	nep.	POL.	šanse	
muš.	60	60	120	1:1
žene	40	40	80	1:1
PUS.	100	100	200	1:1

puš.	nep.	POL.	šanse	
muš.	60	40	100	3:2
žene	60	40	100	3:2
PUS.	120	80	200	3:2

puš.	nep.	POL.	šanse	
muš.	90	60	150	3:2
žene	30	20	50	3:2
PUS.	120	80	200	3:2

- jednake* šanse tj. jednaki odnosi frekvenci u A/b1, A/b2, i A
- muškarci i žene imaju *iste* pušačke karakteristike
- razlike u vrednostima jedne varijable (pol) *nisu* praćene razlikama u vrednosti druge varijable (pušenje)
- dakle: *ne postoji* korelacija A i B

(b) korelacija *postoji*: slučajevi 3, 4, 7, 8 (ovi slučajevi čine *Grupu II*)  
šanse su *različite*:  $a:b \neq c:d \neq fa1:fa2$

puš.	nep.	POL.	šanse	
muš.	60	40	100	3:2
žene	40	60	100	2:3
PUS.	100	100	200	1:1

puš.	nep.	POL.	šanse	
muš.	70	30	100	7:3
žene	50	50	100	1:1
PUS.	120	80	200	3:2

puš.	nep.	POL.	šanse	
muš.	80	40	120	2:1
žene	20	60	80	1:3
PUS.	100	100	200	1:1

puš.	nep.	POL.	šanse	
muš.	100	20	120	5:1
žene	20	60	80	1:3
PUS.	120	80	200	3:2

- različite* šanse tj. različiti odnosi frekvenci u A/b1, A/b2, i A
- muškarci i žene imaju *različite* pušačke karakteristike
- razlike u vrednostima jedne varijable (pol) *jesu* praćene razlikama u vrednosti druge varijable (pušenje)
- dakle: *postoji* korelacija A i B

## 2. Bivarijantni frekvencijski nacrti (BFN) 28

*hijerarhijski prikaz svih 8 slučajeva*

- fa1=fa2 slučajevi 1, 2, 3, 4
  - fb1=fb2 slučajevi 1, 3
    - fa1:fa2 = a:b = c:d slučaj. 1
    - fa1:fa2 ≠ a:b ≠ c:d slučaj. 3
  - fb1≠fb2 slučajevi 2, 4
    - fa1:fa2 = a:b = c:d slučaj. 2
    - fa1:fa2 ≠ a:b ≠ c:d slučaj. 4
- fa1≠fa2 slučajevi 5, 6, 7, 8
  - fb1=fb2 slučajevi 5, 7
    - fa1:fa2 = a:b = c:d slučaj. 5
    - fa1:fa2 ≠ a:b ≠ c:d slučaj. 7
  - fb1≠fb2 slučajevi 6, 8
    - fa1:fa2 = a:b = c:d slučaj. 6
    - fa1:fa2 ≠ a:b ≠ c:d slučaj. 8

učiti: svi slučajevi se međusobno razlikuju na bar jednom od tri kriterijuma, i zajedno iscrpljuju svih 8 mogućnosti

## 2. Bivarijantni frekvencijski nacrti (BFN) 29

numerički pokazatelji strukture rezultata u BFN tipa 2x2

- podsetimo se mera u UFN, u dihotomijama

	mušk.	žene	ZBIR
POL.	25	75	100

totalna frekvencija

	a <sub>1</sub>	a <sub>2</sub>	T
A	f <sub>1</sub>	f <sub>2</sub>	N

totalna frekvencija

frekvencija kategorija

proporcija muškaraca (1/4)

	a <sub>1</sub>	a <sub>2</sub>	T
A	1	3	4

$p1 = f1/N$

proporcija žena (3/4)

	a <sub>1</sub>	a <sub>2</sub>	T
A	1	3	4

$p2 = f2/N$

šansa (1:3 ili 3:1)

	a <sub>1</sub>	a <sub>2</sub>	T
A	1	3	4

$f1:f2$  ili  $f2:f1$  (odn. p1:p2 ili p2:p1, jer se N skrati)

- u BFN tipa 2x2: proporcije i šanse mogu se računati u nekoliko 1D matrica
- to su redovi i kolone osnovne matrice AB, kao i marginalne matrice A i B

## 2. Bivarijantni frekvencijski nacrti (BFN) 30

novе mere za nacrt 2x2: *količnik šansi*, *količnik proporcija*, *fi-koefficijent*

- 1. **količnik šansi**
  - engleski: *odds ratio*, OR
  - $KŠ = (a/b)/(c/d) = ad/bc$
  - učiti: u KŠ se pojavljuju sve četiri situacije u nacrtu 2x2 (a, b, c, d) u specifičnoj algebarskoj kombinaciji
- PRIMER: Slučaj 8
  - šansa da je muškarac pušač: 5:1
  - šansa da je žena pušač: 1:3
  - $KŠ = (100/20)/(20/60) = (5/1)/(1/3) = 15$
- KŠ izražava numerički *odnos* dve šanse: *a/b* i *c/d*
  - šanse su *količni*, a KŠ je *količnik količnika* odn. *količnik drugog reda*
  - KŠ izražava koliko puta je jedna šansa veća ili manja od druge šanse
  - u primeru:
    - šansa da je muškarac pušač je 15 puta veća nego šansa da je žena pušač
    - a šansa da je žena pušač iznosi 1/15 = 0.067 šanse da je muškarac pušač

## 2. Bivarijantni frekvencijski nacrti (BFN) 31

- kako se menja KŠ ako se tabela preuredi?
  - u originalnoj tabeli:  $KŠ = (a/b)/(c/d) = ad/bc$
  - neka se redovi i kolone zamene (tj. redovi postaju kolone, i obrnuto)
    - sada je:  $KŠ = (a/c)/(b/d) = ad/cb$  dakle, dobija se opet ista vrednost!
  - ako dva reda u tabeli zamene mesta, dobija se inverzna vrednost
    - $(c/d)/(a/b) = cb/da = bc/ad = 1/KŠ$
  - ako dve kolone u tabeli zamene mesta, dobija se opet inverzna vrednost
    - $(b/a)/(d/c) = bc/ad = 1/KŠ$

	$a_1$	$a_2$	šanse
$b_1$	a	b	a/b
$b_2$	c	d	c/d

	$b_1$	$b_2$	šanse
$a_1$	a	c	a/c
$a_2$	b	d	b/d

	$a_1$	$a_2$	šanse
$b_2$	c	d	c/d
$b_1$	a	b	a/b

	$a_2$	$a_1$	šanse
$b_1$	b	a	b/a
$b_2$	d	c	d/c

- uočimo: ovakva preuredenja podataka ne menjaju ništa u suštini rezultata!
- stoga: i KŠ i  $1/KŠ$  jednako valjano prikazuju ishod

## 2. Bivarijantni frekvencijski nacrti (BFN) 32

- za KŠ postoje dve osnovne mogućnosti:  $KŠ = 1$  i  $KŠ \neq 1$ 
  - ove dve mogućnosti odražavaju postojanje i nepostojanje korelacije!
- (a)  $KŠ = 1$ , tj.  $(a/b)/(c/d) = 1$ , odakle sledi da je  $a/b = c/d$ 
  - dakle, *ne postoji* korelacija dve varijable
  - PRIMER: Slučaj 6
    - $KŠ = (90/60)/(30/20) = (3/2)/(3/2) = 1$
- (b)  $KŠ \neq 1$ , tj.  $(a/b)/(c/d) \neq 1$ , odakle sledi da je  $a/b \neq c/d$ 
  - dakle, *postoji* korelacija dve varijable
  - PRIMER: Slučaj 8
    - $KŠ = (100/20)/(20/60) = (5/1)/(1/3) = 15$
- zaključak: KŠ je pokazatelj postojanja korelacije dve varijable
- ako je  $KŠ \neq 1$  postoje dve mogućnosti:
  - ako je  $(a/b) > (c/d)$  onda je  $KŠ > 1$
  - ako je  $(a/b) < (c/d)$  onda je  $KŠ < 1$
- međutim, ta razlika nije bitna, jer može nastati preuređivanjem

## 2. Bivarijantni frekvencijski nacrti (BFN) 33

### relativnost KŠ

- setimo se: šanse su *relativne* mere (kao i proporcije i procenti)
  - daju informacije o *odnosu* (relaciji) apsolutnih mera
  - prednost šansi:
    - omogućavaju ili olakšavaju razumevanje nekih aspekata rezultata koje je teže poimati samo uvidom u apsolutne mere
  - mana šansi (slično manama proporcija i procenta):
    - nedostaje informacija o samim *apsolutnim* frekvencama, koja je potrebna za potpuno shvatanje strukture rezultata
- količnik šansi (KŠ) je odnos *odnosa* šansi, tj. *relacija relacija*
  - dakle, KŠ je *dvostruko* relativna mera
  - prednost KŠ:
    - omogućava bolje razumevanje nekih aspekata rezultata koje je teže poimati samo uvidom u apsolutne mere i šanse
  - mana KŠ
    - nedostaje informacija o samim apsolutnim merama i šansama, koja je često nužna za potpuno shvatanje strukture rezultata

## 2. Bivarijantni frekvencijski nacrti (BFN) 34

- u medijima, napisi o nekim istraživanjima se ponekad odnose na rezultate prikazane u obliku KŠ, ali to često nije jasno iskazano
  - takva informacija može da bude jednostrana
- PRIMER: mediji izveštavaju: u gradu se oboljeva 5 puta više od bolesti X nego na selu! Da li se treba odmah iseliti na selo?
  - uočiti: ovakva informacija može da se iskaže u obliku KŠ, tj. kao odnos: *šansa da se oboli u gradu / šansa da se oboli na selu*
  - međutim, da bi se valjano procenilo njeno značenje potrebno je znati i koliko su *same šanse*, a ne samo njihov *odnos*, tj. KŠ
- primer mogućih šansi, ako je bolest veoma retka:
  - šansa oboljenja u gradu: 5:1000000; šansa oboljenja na selu: 1:1000000
  - dakle:  $KŠ = (5/1000000)/(1/1000000) = 5:1$  (uočimo: milioni se skraćuju!)
- vest je dakle tačna, ali je nepotpuna, i stoga je, na izvestan način, *dezinformacija*
  - naime: tačno je, doduše, da je bolest 5 puta češća u gradu nego na selu
  - ali: takođe je tačno da je verovatnoća da se od nje oboli vrlo mala, bilo u gradu bilo na selu, a to nije rečeno u izveštaju u mediju

## 2. Bivarijantni frekvencijski nacrti (BFN) 35

### 2. količnik proporcija (KP)

- srodna mera količniku šansi, ali zasnovana na odnosu *proporcija*
  - takođe je nazvan količnik *verovatnoća*
  - budući da postoje dve vrste proporcija, postoje i dva KP
- KP1: proporcije za kategoriju a1
  - za kat. b1:  $a_1/b_1 = a/(a+b)$
  - za kat. b2:  $c_1/b_2 = c/(c+d)$
  - količnik:  $KP1 = (a/b_1)/(c/b_2)$
- KP2: proporcije za kategoriju a2
  - za kat. b1:  $b_1/b_1 = b/(a+b)$
  - za kat. b2:  $d/b_2 = d/(c+d)$
  - količnik:  $KP2 = (b/b_1)/(d/b_2)$
- KP izražava koliko *puta* je jedna proporcija veća ili manja od druge
  - postoji bliska algebarska veza između KP1, KP2 i KŠ:
    - naime, *količnik* KP1 i KP2 (tj. *količnik trećeg reda*) jednak je KŠ
    - $KP1/KP2 = [(a/b_1)/(c/b_2)] / [(b/b_1)/(d/b_2)] = (a/c)/(b/d) = KŠ$

## 2. Bivarijantni frekvencijski nacrti (BFN) 36

- Slučaj 6:  $KP1 = (90/150)/(30/50) = (3/5)/(3/5) = 1$ 
  - proporcija pušača među muškarcima jednaka je proporciji pušača među ženama
  - $KP2 = (60/150)/(20/50) = (2/5)/(2/5) = 1$
  - proporcija nepušača među muškarcima jednaka je proporciji nepušača među ženama
  - uočiti:  $KP1/KP2 = 1/1 = 1 = KŠ$
- Slučaj 8:  $KP1 = (100/120)/(20/80) = (0.83)/(0.25) = 3.33$ 
  - proporcija pušača među muškarcima je 3.33 puta veća nego među ženama
  - $KP2 = (20/120)/(60/80) = (0.17)/(0.75) = 0.22$
  - proporcija nepušača među muškarcima iznosi oko 1/5 proporcije nepušača među ženama
  - uočiti:  $KP1/KP2 = 3.33/0.22 = 15 = KŠ$
- u medicini, KP se zove *odnos rizika* (engl.: risk ratio) ili *relativni rizik*
  - često se koristi kad se razmatra učestalost neke bolesti, pri čemu se *rizici* odnose na proporcije odn. verovatnoće te bolesti kod dve različite populacije
  - npr.: rizik da se bude pušač kod mušk. (5/6) je 3.33 puta veći nego kod žena (1/4)
  - kao i KŠ, KP daje nepotpunu informaciju, ako se ne navedu i sami rizici

## 2. Bivarijadni frekvencijski nacrti (BFN) 37

- **3.  $\phi$ -koeficijent:** vrsta *koeficijenta korelacije* (KK)
  - KK: matematička veličina koja odražava postojanje korelacije između dve varijable, a koja varira između  $-1$  i  $+1$ 
    - postoji više KK, a  $\phi$  (piše se i  $\varphi$ , čita se 'fi') je jedan od njih
    - $\phi$ -koeficijent je poseban slučaj tzv. *Pirsonovog KK*
      - primenjen je na slučaj kada su obe varijable dihotomije
  - uočiti: KŠ je *pokazatelj* korelacije, ali, pošto ne uzima vrednosti između  $-1$  i  $1$ , nije *koeficijent* korelacije (KK)
    - naime, KŠ varira između  $0$  i  $\infty$  (beskonačno)
  - kao i KŠ, i  $\phi$  uzima u obzir frekvence sve četiri situacije ( $a, b, c, d$ ) ali ih drugačije (i znatno komplikovanije) kombinuje:

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} = \frac{ad - bc}{\sqrt{fa1 * fa2 * fb1 * fb2}}$$

## 2. Bivarijadni frekvencijski nacrti (BFN) 38

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \quad \begin{array}{l} \text{brojilac B} \\ \text{imenilac I} \end{array}$$

- uočiti:  $\phi = B/I$ , pri čemu  $B = ad - bc$  može biti  $>0$ ,  $=0$ , i  $<0$ 
  - za razliku od B, imenilac I može biti samo pozitivan (budući da je koren), i nije od presudne važnosti sa narednu analizu
- podsetimo se tri aspekta korelacije: *postojanje, tip, stepen*
- **1. postojanje korelacije**
  - KK može biti ili  $0$  ( $\phi = 0$ ) ili različit od nule ( $\phi \neq 0$ )
  - te dve mogućnosti odražavaju nepostojanje i postojanje korelacije, naime:
    - ako dve varijable *nisu* korelirane onda je  $\phi = 0$ 
      - naime, kao što znamo, tada je  $a/b = c/d$ , odakle sledi  $ad=bc$ , odakle sledi  $ad-bc = B = 0$ , pa je  $\phi = B/I = 0$
    - takođe: ako dve varijable jesu korelirane onda je  $\phi \neq 0$ 
      - naime, po istoj logici tada je  $B = ad-bc \neq 0$ , pa je i  $\phi \neq 0$

## 2. Bivarijadni frekvencijski nacrti (BFN) 39

- **2. tip odn. smer korelacije:** KK može biti pozitivan ( $\phi > 0$ ) ili negativan ( $\phi < 0$ )
  - biće  $\phi > 0$ , ako je  $B = ad-bc > 0$ 
    - uočiti: tada je  $ad > bc$ , odakle sledi  $a/b > c/d$ , pa je  $KŠ = (a/b) / (c/d) > 1$
  - biće  $\phi < 0$ , ako je  $B = ad-bc < 0$ 
    - uočiti: po istoj logici, tada je  $KŠ < 1$
  - dakle: promenom rasporeda kategorija u tabeli može se promeniti predznak KK
    - stoga *smer* korelacije nije naročito bitan u ovakvim nacrtima
    - važnija je *jačina* korelacije, odn. apsolutna vrednost KK
- **3. stepen odn. jačina korelacije:** zavisi od apsolutne vrednosti KK, tj.  $|\phi|$ 
  - za  $|\phi|$  važi:
    - korelacija je *najjača* ako je  $|\phi| = 1$ , tj. ako je  $\phi = +1$  ili  $\phi = -1$ ,
    - naime, može se matematički dokazati da  $\phi$  ne može biti  $>1$  niti manje  $<-1$
    - $|\phi|$  će biti utoliko veći, ukoliko su *veće* frekvence više koncentrisane na jednoj dijagonali matrice AB, a *manje* frekvence na drugoj dijagonali

• **PRIMERI:**

$\phi = 0$	puš.	nep.	$\phi = 0.2$	puš.	nep.	$\phi = 0.6$	puš.	nep.	$\phi = 1$	puš.	nep.
muš.	50	50	muš.	60	40	muš.	80	20	muš.	100	0
žene	50	50	žene	40	60	žene	20	80	žene	0	100