

**Metodologija psiholoških istraživanja**

**Obrada frekvencijski 4**

**IV. Obrada podataka**

A. Frekvencijski nacrti

- 1. Univarijatni frekvencijski nacrti  
UFN sa više od dve kategorije
- 2. Bivarijatni frekvencijski nacrti

19. novembar 2018

**1. Univarijatni 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?
  - koristićemo ovaj primer za prikaz toka statističke analize
- **PRIMERI: 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 incidenta veće u doba punog meseca nego u doba kada se vidi oko 1/4, 1/2 i 3/4 površine meseca?
- **PRIMER: nacrt sa mnogo kategorija**
  - za koju partiju biste glasali kad bi danas bili izbori?

**1. Univarijatni 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:  $f_3=42$

ROD	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):  $d = f - f'$ 
    - $d_1 = 132-100 = 32$ ;  $d_2 = 126-100 = 26$ ;  $d_3 = 42-100 = -58$
  - uočiti: reziduali se sabiraju do nule:  $d_1 + d_2 + d_3 = 32 + 26 + (-58) = 0$

**1. Univarijatni frekvencijski nacrti (UFN)**

- **Faza III: Test-statistik:  $\chi^2$** 
  - podsetimo se, u slučaju dve kategorije:
  - u slučaju tri kategorije:
  - u slučaju k kategorija:
  - napomena: postoje i nešto složenije formule za  $\chi^2$  one sadrže i tzv. 'korekciju za kontinuitet', ali ona uglavnom nije potrebna
  - pregled računa:
- 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  $df = 1$ ,  $df = 2$ ,  $df = 3$ , ...)

**1. Univarijatni 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 ( $d$ ), kolikičnici  $d^2/f'$ ,  $\chi^2$ , i p-vrednost) izvodi kompjuter

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

**1. Univarijatni 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 jednakе
  - upravo taj test smo opisali u prethodnoj 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 proporcijama 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. Univarijatni frekvencijski nacrti (UFN)

- (2) **test kontrasta:** poređenje dve kategorije međusobno
  - H0: proporcije dve kategorija varijable A u populaciji su jednake
    - kod nacrt-a sive dve kategorije test kontrasta i omnibus test su jedno isto
  - PRIMERI:** tri testa značajnosti razlika frekvenc 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 incidenta 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*

7

## 2. Bivarijatni frekvencijski nacrti (BFN)

- 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**
  - matrica podataka: tipa objekti x varijable
    - kolone: dve (po jedna za svaku varijablu)
    - redovi: onoliko koliko ima objekata istraživanja
    - ćelije: 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 frekvencia N
  - odgovarajuće proporcije, procenti, šanse, i njihovi odnosi

8

## 2. Bivarijatni frekvencijski nacrti (BFN)

### c. Prikaz rezultata

- rezultati se prikazuju *numerički* i *grafički*
  - razmotrićemo detaljno samo prikaze rezultata nacrt-a 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 dobiju prebrojavanjem, iz matrice podataka

PRIMER: rukost i pol	konkretna matrica		apstraktna matrica		$a=f_{11}$ $b=f_{12}$ $c=f_{21}$ $d=f_{22}$
	desnoruki	levoruki	$a_1$	$a_2$	
muškarci	765	80	845		
žene	1168	70	1238		
PUS.	1933	150	2083		

4 frekvence  
situacija

9

## 2. Bivarijatni frekvencijski nacrti (BFN)

matrica AB upotpunjena matricama A, B, i T

konkretna matrica			apstraktna matrica		
desnoruki	levoruki	POL	$a_1$	$a_2$	B
muškarci	765	80	845		$b_1 = a+b$
žene	1168	70	1238		$b_2 = c+d$
RUKOST	1933	150	2083		$N = a+b+c+d$

2 frekvence kategorija varijable A      1 totalna frekvencia  
2 frekvence kategorija varijable B      4 frekvence situacija

- marginalne matrice A, B**
  - sadrži marginalne frekvence
    - frekvence kategorija
    - dobijaju se sabiranjem odgovarajućih redova ili 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  
ostali 5 brojki su dobijene obradom empirijskih podataka

10

## 2. Bivarijatni frekvencijski nacrti (BFN)

- 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

način 1	puš.	nep.	POL	način 2	puš.	nep.	POL	način 3	puš.	nep.	POL
muš.	15%	10%	25%	muš.	33%	18%	25%	muš.	60%	40%	100%
žene	30%	45%	75%	žene	67%	82%	75%	žene	40%	60%	100%
PUS.	45%	55%	100%	PUS.	100%	100%	100%	PUS.	45%	55%	100%
- 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 muškaricama 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%)

11

## 2. Bivarijatni frekvencijski nacrti (BFN)

- ovi rezultati mogu se izraziti i u obliku proporcija odn. verovatnoće:
    - način 1: zajedničke verovatnoće
  - proporcije u odnosu na N**

nač. 1	puš.	nep.	POL	nač. 1	P	N	POL
muš.	0.15	0.10	0.25	M	p(M&P)	p(N&P)	p(M)
žene	0.30	0.45	0.75	Ž	p(Ž&P)	p(Ž&N)	p(Ž)
PUS.	0.45	0.55	1	PUS.	p(P)	p(N)	1
  - način 2: prva vrsta uslovnih verovatnoća
    - uslovne verovatnoće po kolonama

nač. 2	puš.	nep.	POL	nač. 2	P	N	POL
muš.	0.33	0.18	0.25	M	p(M P)	p(M N)	p(M)
žene	0.67	0.82	0.75	Ž	p(Ž P)	p(Ž N)	p(Ž)
PUS.	1	1	1	PUS.	1	1	1
  - način 3: druga vrsta uslovnih verovatnoća
    - uslovne verovatnoće po redovima

nač. 3	puš.	nep.	POL	nač. 2	P	N	POL
muš.	15%	10%	25%	M	p(P M)	p(N M)	1
žene	30%	45%	75%	Ž	p(P Ž)	p(N Ž)	1
PUS.	45%	55%	100%	PUS.	p(P)	p(N)	1
- primer:** koja je verovatn. da je osoba P, koja je verovatn. da je M i P?  
p(M&P)=0.15
- primer:** ako je osoba P, koja je verovatn. da je M?  
p(M|P)=0.33
- primer:** ako je osoba M, koja je verovatn. da je P?  
p(P|M)=0.60

12

## 2. Bivarijatni frekvencijski nacrti (BFN)

### 2. grafički prikaz

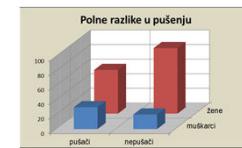
- postoje različite 2D i 3D varijante grafičkih prikaza
  - kao i kod prikaza UFN, grafikoni BFN sadrže brojne **grafische** i **tekstualne elemente**
    - grafiki:** štapići, osa mere, ose varijabli, mreža
    - tekstualni:** oznake varijabli i njenih kategorija, oznake mere i mernih jedinica, naslov
  - prvo ćemo razmotriti 3D prikaze apsolutnih frekvenci
- |      | puš. | nep. | POL |
|------|------|------|-----|
| muš. | 30   | 20   | 50  |
| žene | 60   | 90   | 150 |
| PUŠ. | 90   | 110  | 200 |
- opisamo prikaz podataka iz 2D glavne matrice AB
    - frekvence a, b, c, d
  - podaci iz prostih i marginalnih 1D matrica prikazuju se prema ranije opisanim principima za UFN

13

## 2. Bivarijatni frekvencijski nacrti (BFN)

### 3D (tro-osni) grafikoni

- 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

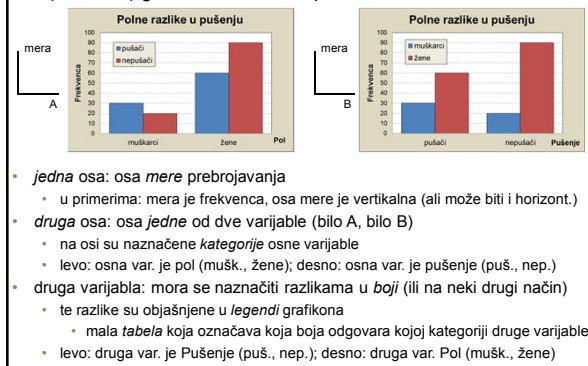


- u ovim primerima su različitim bojama označene različite kategorije jedne varijable
  - varijabla Pol: plavo: podaci za muškarce; crveno: podaci za žene
- međutim, to nije nužno kod ovakvih grafikona

14

## 2. Bivarijatni frekvencijski nacrti (BFN)

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



15

## 2. Bivarijatni frekvencijski nacrti (BFN)

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

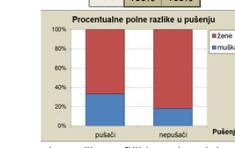
- važe isti principi prikaza (ose, označavanje, legende, itd.)
- u narednim primerima mera prebrojavanja nije frekvencija nego procent

način br. 2

procenti u odn. na var. A	puš.	nep.
muš.	33%	18%
žene	67%	82%
PUŠ.	100%	100%

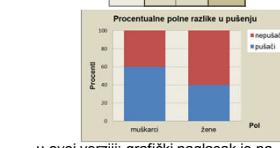
način br. 3

procenti u odn. na var. B	puš.	nep.	POL
muš.	60%	40%	100%
žene	40%	60%	100%



u ovoj verziji: grafički naglasak je na razlikama između pušača i nepušaca

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 nepušaca

16

## 2. Bivarijatni frekvencijski nacrti (BFN)

dalji primjeri  
grafikona

rezultati američkih predsedničkih  
izbora 2016. i 2012.



17

## 2. Bivarijatni frekvencijski nacrti (BFN)

rezultati američkih predsedničkih izbora 2016. i 2012.  
pričak: geografski grafikoni

izbori 2016

Clinton (dem.), Trump (rep.)

po  
saveznim  
državama

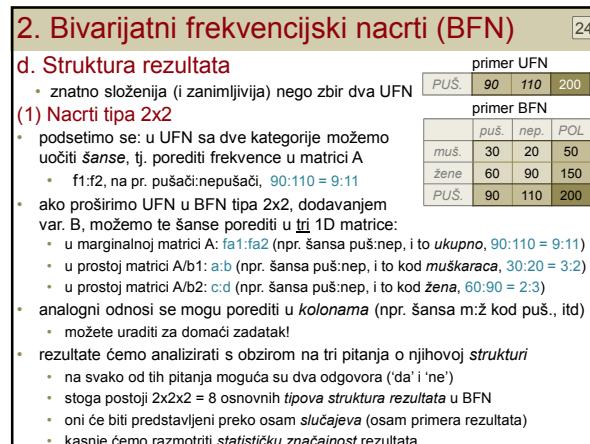
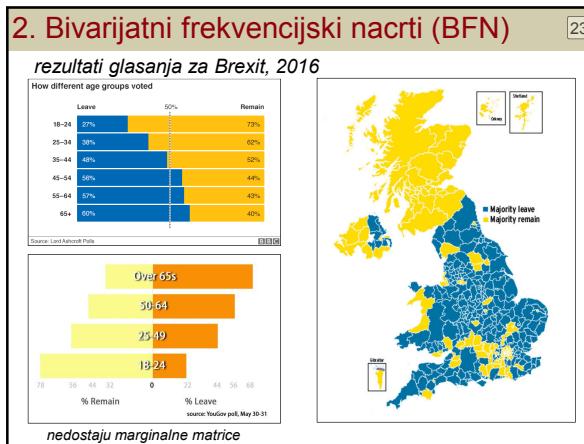
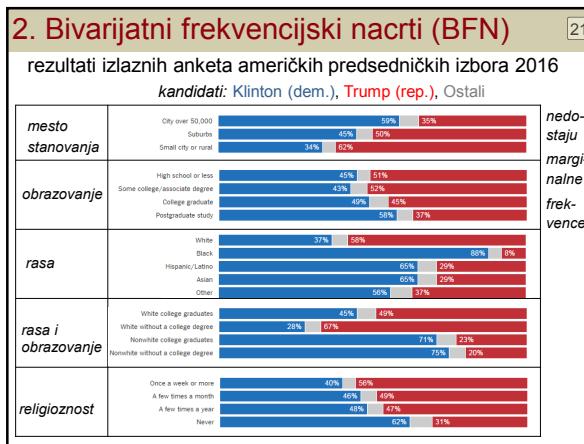
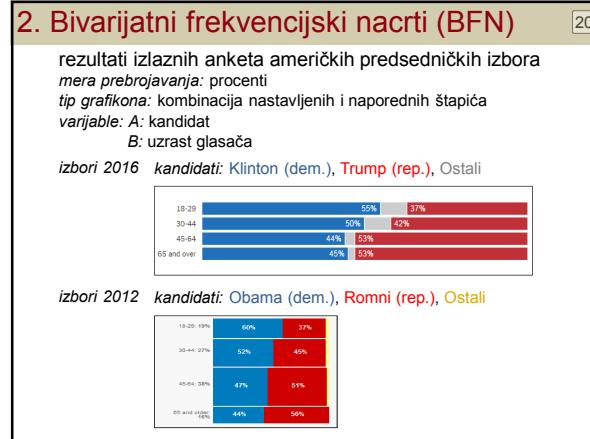
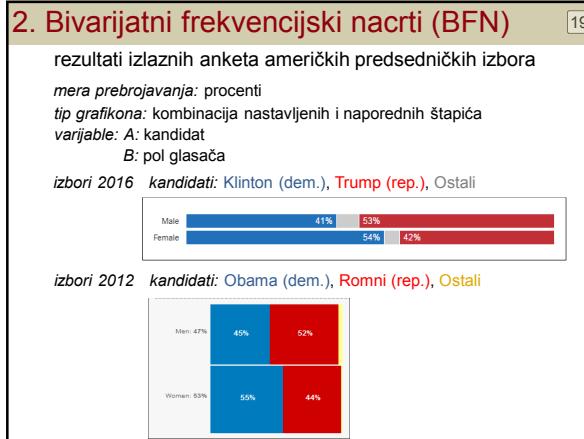


izbori 2012

Obama (dem.), Romney (rep.)

po  
okruzima









## 2. Bivarijatni frekvencijski nacrti (BFN)

37

- 3.  $\phi$ -koeficijent:** vrsta *koeficijenta korelacije* (KK)
  - KK: matematička veličina koja odražava postojanje korelacijske 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*
      - primjenjen je na slučaj kada su obe varijable dihotomije
  - uočiti: KŠ je *pokazatelj* korelacijske, ali, pošto ne uzima vrednosti između **-1** i **1**, nije *koeficijent* korelacijske (KK)
    - naime, KŠ varira između **0** i **∞** (beskonačno)
  - kao i KŠ, i  $\phi$  uzima u obzir frekvence sve četiri situacije (**a,b,c,d**) ali ih drugačije (i znatno komplikovanije) kombinuje:

$$\varphi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} = \frac{ad - bc}{\sqrt{fa_1 * fa_2 * fb_1 * fb_2}}$$

## 2. Bivarijatni frekvencijski nacrti (BFN)

38

- 2. tip odn. smer korelacije:** KK može biti pozitivan ( $\phi > 0$ ) ili negativan ( $\phi < 0$ )
  - blće  $\phi > 0$ , ako je  $B = ad-bc > 0$ 
    - uočiti: tada je  $ad > bc$ , odakle sledi  $a/b > c/d$ , pa je  $K\check{S} = (a/b) / (c/d) > 1$
  - blće  $\phi < 0$ , ako je  $B = ad-bc < 0$ 
    - uočiti: po istoj logici, tada je  $K\check{S} < 1$
- dakle: promenom rasporeda kategorija u tabeli može se promeniti predznak KK
  - stoga *smer* korelacijske nije naročito bitan u ovakvim nacrtima
  - važnija je *jačina* korelacijske, 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* frekvencije više koncentrisane na jednoj dijagonali matrice AB, a *manje* frekvencije na drugoj dijagonali
- PRIMERI:**

$\phi = 0$	puš.	nep.
muš.	50	50
žene	50	50

$\phi = 0.2$	puš.	nep.
muš.	60	40
žene	40	60

$\phi = 0.6$	puš.	nep.
muš.	80	20
žene	20	80

$\phi = 1$	puš.	nep.
muš.	100	0
žene	0	100

## 2. Bivarijatni frekvencijski nacrti (BFN)

39

- 2. tip odn. smer korelacije:** KK može biti pozitivan ( $\phi > 0$ ) ili negativan ( $\phi < 0$ )
  - blće  $\phi > 0$ , ako je  $B = ad-bc > 0$ 
    - uočiti: tada je  $ad > bc$ , odakle sledi  $a/b > c/d$ , pa je  $K\check{S} = (a/b) / (c/d) > 1$
  - blće  $\phi < 0$ , ako je  $B = ad-bc < 0$ 
    - uočiti: po istoj logici, tada je  $K\check{S} < 1$
- dakle: promenom rasporeda kategorija u tabeli može se promeniti predznak KK
  - stoga *smer* korelacijske nije naročito bitan u ovakvim nacrtima
  - važnija je *jačina* korelacijske, 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* frekvencije više koncentrisane na jednoj dijagonali matrice AB, a *manje* frekvencije na drugoj dijagonali
- PRIMERI:**

$\phi = 0$	puš.	nep.
muš.	50	50
žene	50	50

$\phi = 0.2$	puš.	nep.
muš.	60	40
žene	40	60

$\phi = 0.6$	puš.	nep.
muš.	80	20
žene	20	80

$\phi = 1$	puš.	nep.
muš.	100	0
žene	0	100