

# NEPARAMETRIJSKE METODE III.

## *Analiza relacija između nominalnih varijabli*

*Cramerovim  $\phi$  koeficijentom* izražava se povezanost između dviju nominalnih varijabli. Koeficijent se kreće u intervalu od 0 do 1 pri čemu 0 označava nultu, a 1 potpunu povezanost analiziranih varijabli.

Prije izračunavanja Cramerovog  $\phi$  koeficijenta potrebno je provesti hi-kvadrat test za dva ili više nezavisnih uzoraka. Ako se hi-kvadrat testom utvrdi da je razlika između uzoraka statistički značajna onda je i Cramerov  $\phi$  koeficijent statistički značajan.

# NEPARAMETRIJSKE METODE III.

## *Analiza relacija između nominalnih varijabli*

Cramerov  $\phi$  koeficijent izračunava se formulom:

$$\phi = \sqrt{\frac{\chi^2}{N(s-1)}}$$

gdje je

- ➔  $\chi^2$  - hi-kvadrat vrijednost
- ➔  $s$  - broj kategorija varijable s manjim brojem kategorija
- ➔  $N$  - suma opaženih frekvencija

# NEPARAMETRIJSKE METODE III.

## *Analiza relacija između nominalnih varijabli*

*Primjer:* Od 26 studenata koji su pristupili praktičnom dijelu ispita iz KM-a položilo ih je 10, a od 14 studentica položilo ih je 5.

<b>SPOL</b>	<b>NISU POLOŽILI</b>	<b>POLOŽILI</b>	<b>UKUPNO</b>
MUŠKARCI	16 (62%)	10 (38%)	26
ŽENE	9 (64%)	5 (36%)	14
UKUPNO	25	15	40

Kolika je povezanost spola i uspješnosti na praktičnom dijelu ispita iz KM-a i da li je statistički značajna na razini pogreške 5% ?

# NEPARAMETRIJSKE METODE III.

## *Analiza relacija između nominalnih varijabli*

Hi-kvadrat testom izračunati su sljedeći parametri

$$\chi^2=0,03 \quad df=1 \quad p=0,864$$

te su uvršteni u formulu za Cramerov  $\phi$  koeficijent

$$\phi = \sqrt{\frac{\chi^2}{N(s-1)}} = \sqrt{\frac{0,03^2}{40(2-1)}} = \sqrt{\frac{0,009}{40}} = 0,005$$

Cramerov  $\phi$  koeficijent iznosi 0,005. Proporcija pogreške ( $p=0,864$ ) upućuje na zaključak da povezanost između varijabli nije statistički značajna.

# STATISTICA 7

## *Višedimenzionalno grupiranje kvalitativnih podataka*

Višedimenzionalno grupiranje kvalitativnih podataka izvodi se slijedom koraka: padajući izbornik *Statistics* → *Basic Statistics/Tables* → *Tables and banners*. Odabirom opcije *Specify Tables (select variables)* potrebno je odabrati varijable prema kojima će se izvršiti grupiranje.

**Zadatak** - Pokrenite datoteku *TA.sta*! Utvrdite frekvencije tjelesno aktivnih i tjelesno neaktivnih muškaraca te tjelesno aktivnih i tjelesno neaktivnih žena u inicijalnom testiranju (*TA-I*)!

# STATISTICA 7

## *Izračunavanje Cramerovog $\phi$ koeficijenta*

Hi-kvadrat test za  $2 \times 2$  tablice izvodi se slijedom koraka: padajući izbornik *Statistics*  $\rightarrow$  *Nonparametrics*  $\rightarrow$  *2  $\times$  2 Tables* ( *$\chi^2/N^2/Phi^2$ , McNemar, Fisher exact*). U predloženu tablicu potrebno je unijeti opažene frekvencije te odabrati opciju *Summary*. Iz tablice se može očitati  $\chi^2$  vrijednost i odgovarajuća pogreška statističkog zaključka (*Chi-square (df=1)*) ili Yatesova korigirana vrijednost (*Yates corrected Chi-square*), te kvadrirana vrijednost Cramerovog  $\phi$  koeficijenta (*Phi-square*).

**Zadatak** - Na temelju prethodno provedenog grupiranja podataka hi-kvadrat testom utvrdite kolika je povezanost između spola i tjelesne aktivnosti u inicijalnom testiranju (*TA-I*) i testirajte statističku značajnost povezanosti!

## NEPARAMETRIJSKE METODE III.

*Analiza relacija između ordinalnih i kvantitativnih varijabli*  
*Spearmanov koeficijent rang korelacije* ( $\rho$ ) je neparametrijski koeficijent koji odražava veličinu povezanosti između dviju ordinalnih varijabli. Ako se Spearmanovim koeficijentom korelacije želi utvrditi povezanost između dviju kvantitativnih varijabli prethodno ih je potrebno svesti na ordinalnu skalu, odnosno pripadajuće rezultate transformirati u rangove.

Spearmanova rang korelacija izračunava se na temelju rangova, a kreće se između  $-1$  i  $1$  pri čemu  $-1$  označava potpunu negativnu,  $0$  nultu, a  $1$  potpunu pozitivnu povezanost.

# NEPARAMETRIJSKE METODE III.

*Analiza relacija između ordinalnih i kvantitativnih varijabli*

Spearmanov koeficijent rang korelacije ( $\rho$ ) izračunava se formulom:

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^n (x_{1i} - x_{2i})^2}{n(n^2 - 1)}$$

gdje je

- ➔  $x_{1i}$  - rang entiteta  $i$  u varijabli  $x_1$ , a  $i=1, \dots, n$
- ➔  $x_{2i}$  - rang entiteta  $i$  u varijabli  $x_2$ , a  $i=1, \dots, n$
- ➔  $n$  - broj entiteta



# NEPARAMETRIJSKE METODE III.

## *Analiza relacija između ordinalnih i kvantitativnih varijabli*

**Primjer:** Deset učenika je testirano nekim testom motorike na početku ( $x_1$ ) i na kraju ( $x_2$ ) školske godine. Spearmanovim koeficijentom rang korelacije je potrebno utvrditi povezanost između rezultata učenika u inicijalnom i u finalnom stanju.

ISP.	$x_1$	$x_2$
1	115	120
2	145	195
3	120	110
4	92	122
5	82	99
6	116	178
7	100	90
8	70	105
9	104	188
10	110	145

# NEPARAMETRIJSKE METODE III.

## *Analiza relacija između ordinalnih i kvantitativnih varijabli*

Originalni rezultati ispitanika transformiraju se u rangove. Izračuna se razlika između ranga u prvoj i ranga u drugoj varijabli za svakog ispitanika. Izračunate razlike se kvadriraju i zbroje. Suma kvadrata razlika u rangovima u ovom primjeru iznosi 68.

ISP.	$x_1$	$x_2$	$x_1 - x_2$	$(x_1 - x_2)^2$
1	7	5	2	4
2	10	10	0	0
3	9	4	5	25
4	3	6	-3	9
5	2	2	0	0
6	8	8	0	0
7	4	1	3	9
8	1	3	-2	4
9	5	9	-4	16
10	6	7	-1	1

# NEPARAMETRIJSKE METODE III.

## *Analiza relacija između ordinalnih i kvantitativnih varijabli*

U ovom se primjeru Spearmanov koeficijent rang korelacije ( $\rho$ ) izračunava na sljedeći način:

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^n (x_{1i} - x_{2i})^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 68}{10(10^2 - 1)} = 0,59$$

Spearmanov koeficijent rang korelacije iznosi 0,59.

Često je potrebno utvrditi i statističku značajnost izračunatog koeficijenta rang korelacije.

# NEPARAMETRIJSKE METODE III.

## *Analiza relacija između ordinalnih i kvantitativnih varijabli*

Statistička značajnost Spearmanovog koeficijenta rang korelacije testira se pomoću t-distribucije s  $df=n-2$  stupnjeva slobode. Vrijednost koja se uspoređuje s kritičnom t-vrijednosti izračunava se formulom

$$t = \rho \sqrt{\frac{n-2}{1-\rho^2}}$$

gdje je

- ➔  $t$  - vrijednost koja se distribuira prema t-distribuciji s  $df=n-2$
- ➔  $\rho$  - Spearmanov koeficijent rang korelacije
- ➔  $n$  - broj entiteta

## NEPARAMETRIJSKE METODE III.

*Analiza relacija između ordinalnih i kvantitativnih varijabli*

**Primjer:** Spearmanov koeficijent rang korelacije izračunat na uzorku od 10 entiteta iznosi 0,59. Potrebno je utvrditi statističku značajnost izračunatog koeficijenta.

$$t = \rho \sqrt{\frac{n-2}{1-\rho^2}} = 0,59 \sqrt{\frac{10-2}{1-0,59^2}} = 2,07$$

$$df = n - 2 = 10 - 2 = 8$$

Minimalna pogreška s kojom je moguće zaključiti da je izračunati koeficijent statistički značajan iznosi 0,072 odnosno 7,2%.

# STATISTICA 7

## *Izračunavanje Spearmanovog koeficijenta rang korelacije*

Izračunavanje Spearmanovog koeficijenta rang korelacije izvodi se slijedom koraka: padajući izbornik *Statistics* → *Nonparametrics* → *Correlations (Spearman, Kendall tau, gamma)* → *Spearman rank R*. U dijaloškom okviru koji se pokreće odabirom opcije *One variable list* potrebno je označiti dvije ili više varijabli na temelju kojih se želi izračunati korelacijska matrica.

**Zadatak** - U datoteci *Judo.sta* izračunajte korelacijsku matricu na temelju svih varijabli.

# STATISTICA 7

## *Testiranje statističke značajnosti Spearmanove rang korelacije*

Testiranje statističke značajnosti Spearmanovog koeficijenta rang korelacije izvodi se slijedom koraka: padajući izbornik *Statistics* → *Nonparametrics* → *Correlations (Spearman, Kendall tau, gamma)* → *Spearman rank R*. U traku *p-level for highlighting* potrebno je upisati željenu proporciju pogreške statističkog zaključka. Sve korelacije koje su statistički značajne na razini upisane pogreške će u korelacijskoj matrici biti označene crvenom bojom.

**Zadatak** - U datoteci *Judo.sta* testirajte statističku značajnost korelacije između varijabli *SDM* i *BML* uz pogrešku  $p=1\%$  .

## NEPARAMETRIJSKE METODE III.

*Analiza relacija između dihotomne i kvantitativne varijable*

*Point-biserialni koeficijent korelacije* je mjera povezanosti između *dihotomne varijable* i kvantitativne varijable.

*Dihotomna varijabla* je naziv za nominalnu varijablu s dva moguća pojava oblika obilježja (npr. *spol*).

Point-biserialni koeficijent korelacije izračunava se jednako kao i Pearsonov koeficijent korelacije, uz uvjet da se oznake kategorija dihotomne varijable prethodno zamijene proizvoljno odabranim numeričkim kodovima (npr. za varijablu *spol*: *muškarac*=1 , *žena*=2).



# NEPARAMETRIJSKE METODE III.

## *Analiza relacija između dihotomne i kvantitativne varijable*

Ako je zadovoljen prethodno navedeni uvjet, Point-biserialni koeficijent korelacije ( $r_{pb}$ ) moguće je izračunati formulom

$$r_{pb} = \frac{\sum_{i=1}^n (z_{x_i} z_{y_i})}{n}$$

gdje je

- ➔  $z_{x_i}$  - standardizirani kod entiteta  $i$  u dihotomnoj varijabli  $x$
- ➔  $z_{y_i}$  - standardizirani rezultat entiteta  $i$  u kvantitativnoj varijabli  $y$
- ➔  $n$  - broj entiteta

# NEPARAMETRIJSKE METODE III.

## *Analiza relacija između dihotomne i kvantitativne varijable*

Point-biserialni koeficijent korelacije kreće se u intervalu između  $-1$  i  $1$ . Pri utvrđivanju veličine povezanosti analiziranih varijabli dovoljno je interpretirati apsolutnu vrijednost izračunatog koeficijenta. Apsolutna vrijednost jednaka  $1$  označava potpunu, a  $0$  nikakvu povezanost između varijabli.

Pošto je jedna od varijabli nominalna (ne postoje razlike u vrijednosti između kategorija) negativna vrijednost Point-biserialnog koeficijenta korelacije ne upućuje na negativnu povezanost varijabli.

## NEPARAMETRIJSKE METODE III.

### *Analiza relacija između dihotomne i kvantitativne varijable*

Negativna vrijednost Point-biserialnog koeficijenta korelacije upućuje na zaključak da entiteti iz kategorije čija je oznaka zamijenjena većom numeričkom vrijednošću postižu niže, a entiteti iz kategorije čija je oznaka zamijenjena manjom numeričkom vrijednošću više rezultate u kvantitativnoj varijabli.

Nasuprot tome, pozitivna vrijednost koeficijenta upućuje na zaključak da entiteti iz kategorije čija je oznaka zamijenjena većom numeričkom vrijednošću postižu više, a entiteti iz kategorije čija je oznaka zamijenjena manjom numeričkom vrijednošću niže rezultate u kvantitativnoj varijabli.

# NEPARAMETRIJSKE METODE III.

## *Analiza relacija između dihotomne i kvantitativne varijable*

**Primjer:** Šest učenika (*m*) i pet učenica (*z*) je testirano nekim testom motorike *x*. Potrebno je utvrditi povezanost spola i rezultata na provedenom testu motorike Point-biserialnim koeficijentom korelacije.

ISP.	<i>spol</i>	<i>x</i>
1	m	165
2	m	195
3	z	110
4	m	138
5	z	99
6	z	143
7	z	120
8	m	118
9	m	198
10	m	145

# NEPARAMETRIJSKE METODE III.

## *Analiza relacija između dihotomne i kvantitativne varijable*

Originalne oznake ispitanika u dihotomnoj varijabli se zamijene numeričkim kodovima ( $k_{spol}$ ) te se izračunaju standardizirane vrijednosti u dihotomnoj ( $z_{spol}$ ) i u kvantitativnoj ( $z_x$ ) varijabli. Potom se izračunaju umnošci z-vrijednosti za svakog ispitanika. Zbroj umnožaka z-vrijednosti u ovom primjeru iznosi -5,71.

ISP.	$k_{spol}$	$z_{spol}$	$z_x$	$z_{spol} \cdot z_x$
1	1	-0,77	0,64	-0,50
2	1	-0,77	1,53	-1,18
3	2	1,16	-0,97	-1,13
4	1	-0,77	-0,15	0,12
5	2	1,16	-1,30	-1,51
6	2	1,16	0,00	0,00
7	2	1,16	-0,68	-0,79
8	1	-0,77	-0,74	0,57
9	1	-0,77	1,61	-1,25
10	1	-0,77	0,06	-0,04

## NEPARAMETRIJSKE METODE III.

### *Analiza relacija između dihotomne i kvantitativne varijable*

Prethodno izračunat zbroj umnožaka standardiziranih vrijednosti entiteta u dihotomnoj i kvantitativnoj varijabli se uvrsti u formulu

$$r_{pb} = \frac{\sum_{i=1}^n (z_{x_i} z_{y_i})}{n} = \frac{-5,71}{10} = -0,57$$

Point-biserialni koeficijent korelacije iznosi  $-0,571$  što ukazuje na veliku povezanost spola i rezultata na provedenom motoričkom testu. Negativan predznak koeficijenta upućuje na zaključak da pripadnici muškog spola postižu više rezultate u ovom motoričkom testu nego žene.

# NEPARAMETRIJSKE METODE III.

## *Analiza relacija između dihotomne i kvantitativne varijable*

Statistička značajnost Point-biserialnog koeficijenta korelacije testira se pomoću t-distribucije s  $df=n-2$  stupnjeva slobode. Vrijednost koja se uspoređuje s kritičnom t-vrijednosti izračunava se formulom

$$t = r_{pb} \sqrt{\frac{n-2}{1-r_{pb}^2}}$$

gdje je

- ➔  $t$  - vrijednost koja se distribuira prema t-distribuciji s  $df=n-2$
- ➔  $r_{pb}$  - Point-biserialni koeficijent korelacije
- ➔  $n$  - broj entiteta

# NEPARAMETRIJSKE METODE III.

*Analiza relacija između dihotomne i kvantitativne varijable*

*Primjer:* Point-biserialni koeficijent korelacije izračunat na uzorku od 10 entiteta iznosi -0,57. Potrebno je utvrditi statističku značajnost izračunatog koeficijenta.

$$t = r_{pb} \sqrt{\frac{n-2}{1-r_{pb}^2}} = -0,57 \sqrt{\frac{10-2}{1+0,57^2}} = -2,90$$

$$df = n - 2 = 10 - 2 = 8$$

Minimalna pogreška s kojom je moguće zaključiti da je izračunati koeficijent statistički značajan iznosi 0,02, odnosno 2%.



# STATISTICA 7

## *Izračunavanje Point-biserialnog koeficijenta korelacije*

Izračunavanje Point-biserialnog koeficijenta korelacije izvodi se slijedom koraka: padajući izbornik *Statistics* → *Basic Statistics/Tables* → *Correlation matrices*. U dijaloškom okviru koji se pokreće odabirom opcije *One variable list* potrebno je označiti dvije ili više varijabli na temelju kojih se želi izračunati korelacijska matrica. Prije opisanog postupka potrebno je zamijeniti tekstualne oznake dihotomne varijable numeričkim kodovima pomoću dijaloškog okvira *Replace* koji se pokreće kombinacijom tipki *Ctrl+H*.

**Zadatak** - U datoteci *Pejcic-318.sta* izračunajte korelaciju između spola (*SPOL*) i kožnog nabora nadlaktice (*ANN*)!

# STATISTICA 7

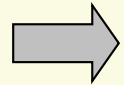
## *Testiranje statističke značajnosti koeficijenta $r_{pb}$*

Testiranje statističke značajnosti Point-biserialnog koeficijenta korelacije izvodi se slijedom koraka: padajući izbornik *Statistics* → *Basic Statistics/Tables* → *Correlation matrices* → *Options*. U traku *p-level for highlighting* potrebno je upisati željenu proporciju pogreške statističkog zaključka. Sve korelacije koje su statistički značajne na razini upisane pogreške će u korelacijskoj matrici biti označene crvenom bojom.

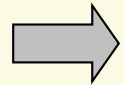
**Zadatak** - U datoteci *Pejcic-318.sta* testirajte statističku značajnost korelacije između varijabli *SPOL* i *ANN* uz pogrešku  $p=5\%$ .

# NEPARAMETRIJSKE METODE III.

## *Literatura za pripremanje kolokvija*



Petz, B. (2002). *Osnovne statističke metode za nematematičare*. Jastrebarsko: Naklada Slap, str. 199-206, 223-225, 227-229.



Šošić, I. (2004). *Primijenjena statistika*. Zagreb: Školska knjiga, str. 424-427.