

# Neparametarski testovi za jedan uzorak

0011 0010 1010 1101 0001 0100 1011

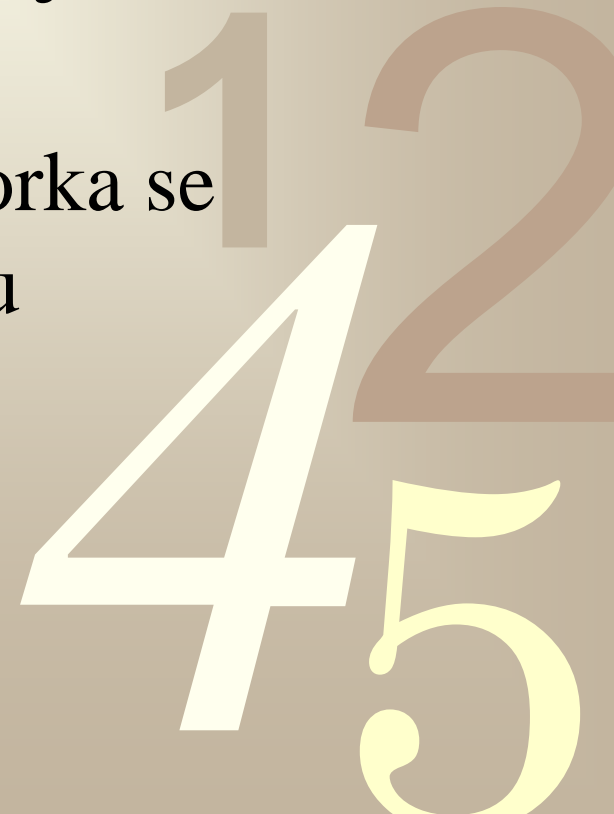
Aleksandra Mandić  
65/2010



# Statistička analiza

0011 0010 1010 1101 0001 0100 1011

- Statističko zaključivanje je proces izvlačenja zaključaka o populaciji na osnovu uzorka
- To jest, rezultati dobijeni iz uzorka se proširuju na populaciju uz meru pouzdanosti



# 2 načina

0011 0010 1010 1101 0001 0100 1011

- Ocenjivanje
- Statistički testovi



# Statistički testovi

0011 0010 1010 1101 0001 0100 1011

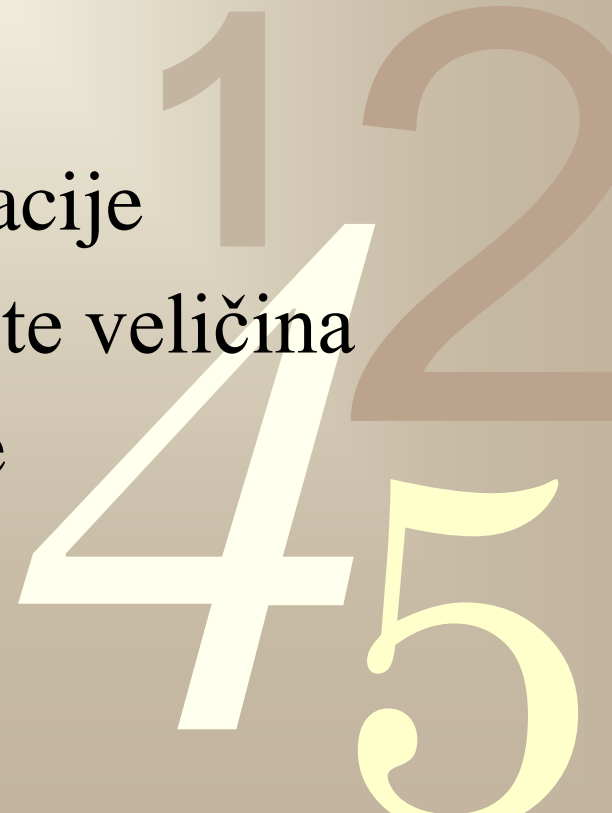
- Parametarski testovi
- Neparametarski testovi



# Neparametraski testovi-prednosti

0011 0010 1010 1101 0001 0100 1011

- Neparametraski testovi nemaju pretpostavku o raspodeli populacije
- Mogu da se koriste za sve 4 vrste veličina
- Kalkulacije su vrlo jednostavne



# Neparametarski testovi-nedostaci

0011 0010 1010 1101 0001 0100 1011

- Zbog slabijih pretpostavki imaju manju moć od parametarskih testova



# Testovi

0011 0010 1010 1101 0001 0100 1011

- Hi-kvadrat test
- Binomni test
- Kolmogorov-Smirnov test
- Test koraka



# Implementacija u SPSS-u

0011 0010 1010 1101 0001 0100 1011

Analyze->Nonparametric Tests->  
Legacy Dialogs







## Hi-kvadrat test

• Naziva se još

The Chi-squared Goodness of Fit Test

• Od interesa je raspodela frekvencija cele populacije

• Posmatra se uzorak izvučen iz neke populacije, pri čemu se pretpostavlja da postoji  $c$  događaja od koji se svaki dešava sa verovatnoćom  $p_j$ .

• Cilj je ispitati tu raspodelu tih frekvencija, to jest da li je  $p_j = p_j^*$  za neke vrednosti

•  $O_j$  - Broj pojavljivanja  $j$ -tog događaja u uzorku

•  $N$  - ukupan broj opservacija

•  $E_j$  - očekivan broj pojavljivanja događaja pri  $H_0$   $E_j = Np_j^*$

• Test statistika  $T$  ima hi-kvadrat raspodelu sa  $c-1$  stepeni slobode

- Pretpostavke  
U pitanju je prost slučajan uzorak.  
(Cochran's Rule)

- Hipoteze

$H_0: p_j = p_j^*$  za  $j=1,2,\dots,c$

$H_1: \text{bar jedna od } p_j \text{ se razlikuje od pretpostavljene vrednosti}$

- Test statistika

$$t = \sum_{j=1}^c \frac{(O_j - E_j)^2}{E_j}$$

- P-vrednost

$p = P(T \geq t)$

- Zaključak



# Primer

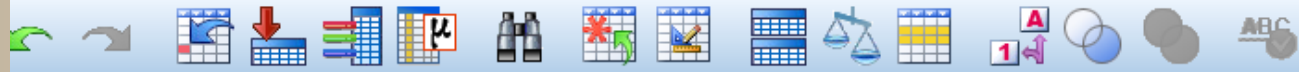
0011 0010 1010 1101 0001 0100 1011

Supervizor rada u jednoj kompaniji sumnja da je verovatnije da će radnici tražiti slobodan dan zbog bolesti određenim danima. Kako bi ovo testirala, sakupila je podatke o brojevima poziva za svaki dan u nedelji. Podaci su dati sledećom tabelom.

dan	Pon-1	Uto-2	Sre-3	Čet-4	Pet-5
Br.poziva	8	9	6	9	8

Da li postoje dokazi da frekvencije poziva nisu ravnomerno raspodeljene po danima, to jest da je supervizor u pravu?

HINT:  $p_j=0.2, j=1,2,\dots,5$



var	var	var	var	var	var	var	var	var	var	var	var	var	var	var
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

### Chi-square Test

Test Variable List: pozivi

Expected Range

Get from data

Use specified range

Lower:

Upper:

Expected Values

All categories equal

Values:

Add

Change

Remove

OK Paste Reset Cancel Help

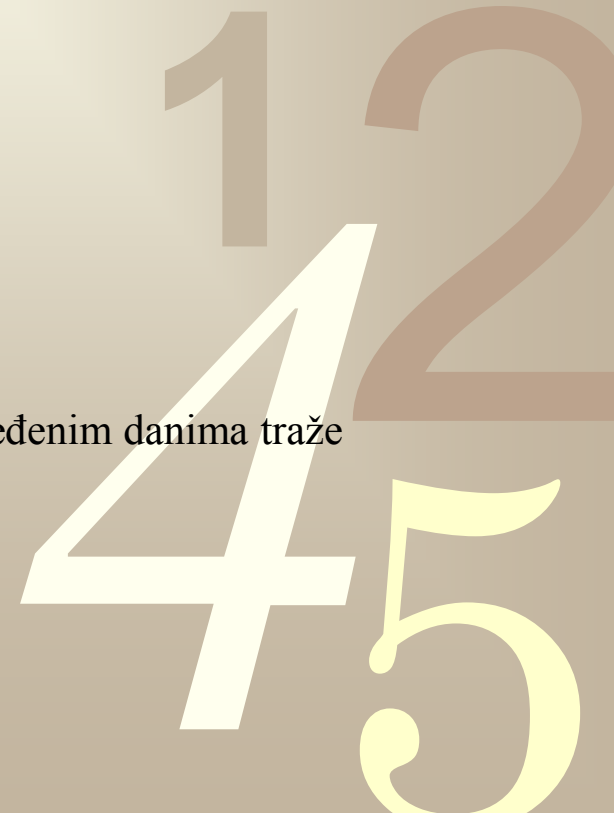
Exact...

Options...

# Test

0011 0010 1010 1101 0001 0100 1011

- Pretpostavke  
U pitanju je prost slučajni uzorak.  
(Cochran's Rule)
- Hipoteze  
 $H_0: p_j = 0.2$  za  $j = 1, 2, \dots, 5$   
 $H_1$ : bar jedna od  $p_j$  se razlikuje od pretpostavljene vrednosti
- Test statistika  
 $t = 0.75$
- P-vrednost  
 $p = 0.945$
- Zaključak  
Ne postoje značajni dokazi da radnici imaju tendenciju da određenim danima traže slobodan dan.



# Izlaz

The screenshot displays the IBM SPSS Statistics Viewer interface. The title bar reads '\*Output1 [Document1] - IBM SPSS Statistics Viewer'. The menu bar includes File, Edit, View, Data, Transform, Insert, Format, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations and analysis. The left pane shows a tree view of the output, with 'Test Statistics' selected. The main pane displays the following results:

[DataSet0]

### Chi-Square Test

#### Frequencies

pozivi

	Observed N	Expected N	Residual
1.00	8	8.0	.0
2.00	9	8.0	1.0
3.00	6	8.0	-2.0
4.00	9	8.0	1.0
5.00	8	8.0	.0
Total	40		

#### Test Statistics

	pozivi
Chi-Square	.750 <sup>a</sup>
df	4
Asymp. Sig.	.945

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 8.0.

## Binomni test

• Ispituje se populacija u kojoj postoje samo dve mogućnosti

• Jedna se može označiti uspehom, druga neuspehom

• Na osnovu uzorka potrebno je ispitati verovatnoću pojavljivanja događaja koji je označen kao uspeh

• Test statistika  $T$  ima binomnu raspodelu sa parametrima  $n$  i  $p^*$ , pri čemu je  $n$  obim uzorka

- Pretpostavke  
U pitanju je prost slučajan uzorak, pri čemu u populaciji postoje samo dve mogućnosti.

- Hipoteze

$$H_0: p=p^*$$

$$H_1: p \neq p^*$$

- Test statistika  
t-broj uspeha u uzorku

- P-vrednost

$$p = P(T \leq t) + P(T \geq n - t)$$

$$p^* = 0.5 \quad p = 2P(T \leq t)$$

- Zaključak



# Primer

0011 0010 1010 1101 0001 0100 1011

- Posmatra se baza 3 sa časova vežbi.
- Specijalno, posmatra se promenljiva pol I cilj je ispitati da li je data prednost nekom polu prilikom ispitivanja.
- Uspeh-ispitanik je osoba muškog pola
- $p^*=0.5$





# Baza

0011

63-baza3.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Broj	Numeric	11	0	Broj ispitanika	None	None	8	Right	Scale	Input
2	pol	Numeric	11	0	pol	{1, muski}...	None	8	Right	Nominal	Input
3	X1	Numeric	11	0	X1-poznavanje Interneta	{1, Veoma slabo}...	9	8	Right	Scale	Input
4	Casovi	Numeric	11	0	X2-koriscenje Interneta_cas/nedelja	None	None	8	Right	Scale	Input
5	X3	Numeric	11	0	X3-Stav prema Internetu	{1, veoma negativan}...	None	8	Right	Scale	Input
6	X4	Numeric	8	0	X4-stav prema zabavi	{1, veoma negativan}...	None	8	Right	Scale	Input
7	X5	Numeric	11	0	X5-stav prema tehnologiji	{1, veoma negativan}...	None	8	Right	Scale	Input
8	Globalizacija	Numeric	8	0	Stav prema globalizaciji	{1, veoma negativan}...	None	10	Right	Scale	Input
9	X7	Numeric	8	0	X7-stav prema elektronskom marketingu	{1, veoma negativan}...	None	8	Right	Scale	Input
10	X8	Numeric	8	0	X8-stav prema elektronskoj trgovini	{1, veoma negativan}...	None	8	Right	Scale	Input
11	X9	Numeric	11	0	X9-stav prema elektronskom bankarstvu	{1, veoma negativan}...	None	8	Right	Scale	Input
12	Tip	Numeric	8	0	Tip_Internet_korisnika	{1, slab korisnik}...	None	8	Right	Ordinal	Input
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											

Data View Variable View

IBM SPSS Statistics Processor is ready



pol	X1	Casovi	X3	X4	X5	Globalizacija	X7	X8	X9	Tip	var
2	5	18	4	5	1	2	5	7	5	3	
2	4	14	5	6	5	4	6	4	6	2	
1	5	17							7	3	
1	6	5							4	1	
1	7	18							2	3	
1	5	13							3	2	
1	4	8							3	2	
2	3	5							4	1	

**Binomial Test**

Test Variable List:

pol [pol]

Exact...  
Options...

Test Proportion: 0.50

Define Dichotomy:

Get from data  
 Cut point

OK Paste Reset Cancel Help

# Test

0011 0010 1010 1101 0001 0100 1011

- **Pretpostavke**  
U pitanju je prost slučajan uzorak, pri čemu u populaciji postoje samo dve mogućnosti.
- **Hipoteze**  
 $H_0: p=0.5$   
 $H_1: p \neq 0.5$
- **Test statistika**  
 $t=50$
- **P-vrednost**  
 $p=1$
- **Zaključak**  
Ne postoje značajni dokazi da je neki pol imao prednost pri ispitivanju.



# Izlaz

The screenshot shows the IBM SPSS Statistics Viewer interface. The title bar reads '\*Output4 [Document4] - IBM SPSS Statistics Viewer'. The menu bar includes File, Edit, View, Data, Transform, Insert, Format, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations and analysis. The left sidebar shows a tree view with 'Output' expanded to 'NPar Tests', which includes 'Binomial Test'. The main window displays the following text:

```
GET  
FILE='C:\Users\Aleksandra\AppData\Local\Temp\63-baza3-1.sav'.  
DATASET NAME DataSet1 WINDOW=FRONT.  
NPAR TESTS  
  /BINOMIAL (0.50)=pol  
  /MISSING ANALYSIS.
```

**NPar Tests**

→ [DataSet1] C:\Users\Aleksandra\AppData\Local\Temp\63-baza3-1.sav

**Binomial Test**

	Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
pol	Group 1 muski	50	.50	.50	1.000
	Group 2 zenski	50	.50		
	Total	100	1.00		

## Kolmogorov-Smirnov Test

- Ovaj test ima istu ulogu kao The Chi-Squared Goodness of Fit Test u smislu da ga interesuje raspodela cele populacije
- Primenjuje se za neprekidne raspodele
- Prednost mu je u tome što ima veću moć i koristi se za ordinalne slučajne veličine
- Takođe, treba ga uzeti u obzir i prilikom malog obima uzorka
- Osnova ovog testa je funkcija raspodele
- Za proveravanje se koristi empirijska funkcija raspodele  $S(x)$
- Centralna teorema matematičke statistike
- Nedostatak je taj što se validno može ispitati samo klasa raspodele, jer procene parametara u softveru nisu najkvalitetnije

- **Pretpostavke**  
U pitanju je prost slučajan uzorak, pri čemu populacija ima funkciju raspodele  $F(x)$
- **Hipoteze**  
 $H_0: F(x) = F^*(x)$   
 $H_1: F(x) \neq F^*(x)$
- **Test statistika**  
 $t = \sup |F^*(x) - S^*(x)|$
- **P-vrednost**  
veoma komplikovan izraz
- **Zaključak**



# Primer

0011 0010 1010 1101 0001 0100 1011

- Posmatra se ista baza 3 kao u prethodnom primeru
- Cilj je testirati raspodelu promenljive Tip, to jest videti da li je ta raspodela normalna



Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	F
eric	11	0	Broj ispitanika	None	None	8	Right	Scale	↳ Inp
eric	11	0	pol	{1, muski}...	None	8	Right	Nominal	↳ Inp
eric	11	0	X1-poznavanje Interneta	{1, Veoma sla...	9	8	Right	Scale	↳ Inp
eric	11	0	X2-koriscenje Interneta... (nedelja	None	None	9	Right	Scale	↳ Inp
eric	11	0	X3-Stav pr				Right	Scale	↳ Inp
eric	8	0	X4-stav pr				Right	Scale	↳ Inp
eric	11	0	X5-stav pr				Right	Scale	↳ Inp
eric	8	0	Stav prem				Right	Scale	↳ Inp
eric	8	0	X7-stav pr				Right	Scale	↳ Inp
eric	8	0	X8-stav pr				Right	Scale	↳ Inp
eric	11	0	X9-stav pr				Right	Scale	↳ Inp
eric	8	0	Tip_Intern				Right	Ordinal	↳ Inp

### One-Sample Kolmogorov-Smirnov Test

Test Variable List:

- Tip\_Internet\_korisni...

Test Distribution:

Normal  Uniform

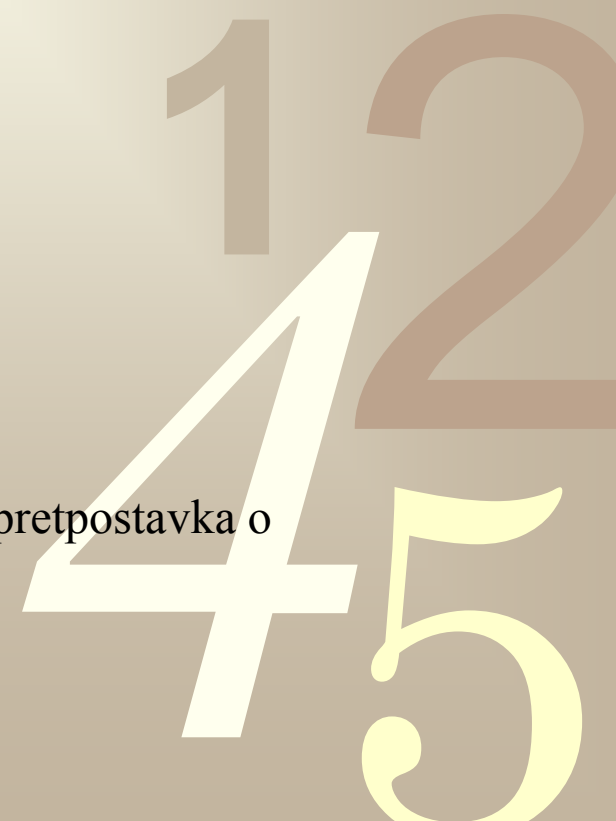
Poisson  Exponential

Buttons: OK, Paste, Reset, Cancel, Help, Exact..., Options...

# Test

0011 0010 1010 1101 0001 0100 1011

- Pretpostavke  
U pitanju je prost slučajni uzorak, pri čemu populacija ima funkciju raspodele  $F(x)$
- Hipoteze  
 $H_0$ : raspodela je normalna  
 $H_1$ : raspodela nije normalna
- Test statistika  
 $t=2.245$
- P-vrednost  
 $p=0$
- Zaključak  
Postoje značajni dokazi za alternativnu hipotezu, pa se pretpostavka o normalnosti odbacuje u korist alternative.





# Izlaz

0011

The screenshot displays the IBM SPSS Statistics Viewer interface. The title bar reads '\*Output8 [Document8] - IBM SPSS Statistics Viewer'. The menu bar includes File, Edit, View, Data, Transform, Insert, Format, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations and analysis. The left-hand 'Output' pane shows a tree structure with 'Output' expanded to 'NPar Tests', which is further expanded to 'Title'. The main window displays the following text:

```
NPAr TESTS
  /K-S (NORMAL) =Tip
  /MISSING ANALYSIS.
```

→ **NPAr Tests**

[DataSet1] C:\Users\Aleksandra\AppData\Local\Temp\63-baza3-1.sav

**One-Sample Kolmogorov-Smirnov Test**

		Tip_Internet_k orisnika
N		100
Normal Parameters <sup>a,b</sup>	Mean	2.03
	Std. Deviation	.810
Most Extreme Differences	Absolute	.225
	Positive	.208
	Negative	-.225
Kolmogorov-Smirnov Z		2.245
Asymp. Sig. (2-tailed)		.000

a. Test distribution is Normal.  
b. Calculated from data.

At the bottom right of the window, the status bar indicates 'IBM SPSS Statistics Processor is ready'.

## Test koraka

•Ovaj test se koristi za ispitivanje slučajnosti podataka, pri čemu je važno da podaci budu podeljeni u dve grupe.

•Pod korakom se podrazumeva podniz elemenata iste kategorije bilo koje dužine

•R-broj koraka

•Test statistika se dobija kao razlika broja koraka i očekivanog broja koraka, podeljena sa standardnom devijacijom broja koraka

•Za velike obime uorka, ova test statistika ima standardnu normalnu raspodelu

•n1-broj elemenata prve grupe

•n2-broj elemenata druge grupe

$$\bar{R} = \frac{2n_1n_2}{n_1+n_2} + 1$$

$$S_R = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}$$

- Hipoteze  
H0: u pitanju je slučajan niz  
H1: niz nije slučajan

- Test statistika

$$t = \frac{(R - \bar{R})}{S_R}$$

- P-vrednost  
 $p=2P(T \geq t)$
- Zaključak



# Primer

0011 0010 1010 1101 0001 0100 1011

- Posmatra se ista baza 3, kao u prethodna dva primera
- Cilj je ispitati da li je niz ispitanika slučajan tako što će se izvršiti podela prema polu ispitanika





Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	R
numeric	11	0	Broj ispitanika	None	None	8	Right	Scale	Input
numeric	11	0	pol	{1, muski}...	None	8	Right	Nominal	Input
numeric	11	0	X1-poznavanje Interneta	{1, Veoma sla...	9	8	Right	Scale	Input
numeric	11	0	X2-koriscen...				Right	Scale	Input
numeric	11	0	X3-Stav pre...				Right	Scale	Input
numeric	8	0	X4-stav pre...				Right	Scale	Input
numeric	11	0	X5-stav pre...				Right	Scale	Input
numeric	8	0	Stav prema...				Right	Scale	Input
numeric	8	0	X7-stav pre...				Right	Scale	Input
numeric	8	0	X8-stav pre...				Right	Scale	Input
numeric	11	0	X9-stav pre...				Right	Scale	Input
numeric	8	0	Tip_Intern...				Right	Ordinal	Input

**Runs Test**

Test Variable List:

pol [pol]

Cut Point

Median  Mode

Mean  Custom: 1.5

OK Paste Reset Cancel Help

# Test

0011 0010 1010 1101 0001 0100 1011

- Hipoteze  
H0: u pitanju je slučajan niz  
H1: niz nije slučajan
- Test statistika  
 $t = -1.005$
- P-vrednost  
 $p = 0.315$
- Zaključak  
Ne postoje dokazi da dati niz podataka nije slučajan.



# Izlaz

The screenshot displays the IBM SPSS Statistics Viewer interface. The main window shows the output of an N-Par Test. The left sidebar contains a tree view with 'Output' expanded to show 'Log', 'NPar Tests', 'Title', 'Notes', 'Active Dataset', 'Runs Test', and 'Runs Test 2'. The main area displays the following text:

```
/RUNS (MEDIAN) =pol  
/RUNS (1.5) =pol  
/MISSING ANALYSIS.
```

→ **NPar Tests**

[DataSet1] C:\Users\Aleksandra\AppData\Local\Temp\63-baza3-1.sav

**Runs Test**

	pol
Test Value <sup>a</sup>	2
Cases < Test Value	50
Cases >= Test Value	50
Total Cases	100
Number of Runs	46
Z	-1.005
Asymp. Sig. (2-tailed)	.315

a. Median

**Runs Test 2**

	pol
Test Value <sup>a</sup>	1.50
Total Cases	100
Number of Runs	46
Z	-1.005
Asymp. Sig. (2-tailed)	.315

a. User-specified.

# Dodatak

0011 0010 1010 1101 0001 0100 1011

## Kartica Exact Test

The screenshot shows the IBM SPSS Statistics Data Editor interface. The main window displays a data grid with 12 rows and 12 columns. The columns are labeled: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role. The rows contain data for variables: Broj, pol, X1, Casovi, X3, X4, X5, Globalizacija, X7, X8, X9, and Tip. A dialog box titled "Binomial Test" is open, showing the "Exact Tests" section. The "Exact" radio button is selected, and the "Number of samples" is set to 10000. The "Time limit per test" is set to 5 minutes. The "Define Dichotomy" section is also visible, with "Get from data" selected.

## Kartica Options

The screenshot shows the IBM SPSS Statistics Data Editor interface, similar to the previous one. The "Binomial Test" dialog box is open, and the "Test Variable List" section is active, showing "Broj ispitivanja" selected. The "Binomial Test: Options" sub-dialog box is also open, showing the "Statistics" section with "Descriptive" and "Quantiles" selected. The "Missing Values" section has "Exclude cases listwise" selected. The "Define Dichotomy" section has "Get from data" selected. The "OK" button is highlighted.