Инжењерска статистика
Вежба 8

Тест Kolmogorova-интервали

Резултати 1000 мерења одступања од номиналне мере изражени у mm, дати су у табели. Да ли се одступање од номиналне мере понаша по нормалној расподели $\mathcal{N}\left(0,1^2\right)$

Интервал	Фреквенције	Интервал	Фреквенције
$(-\infty,-2.0)$	21	$[0.0,0.5)$	201
$[-2.0,-1.5)$	47	$[0.5,1.0)$	142
$[-1.5,-1.0)$	87	$[1.0,1.5)$	97
$[-1.0,-0.5)$	158	$[1.5,2.0)$	41
$[-0.5,0.0)$	181	$[2.0,+\infty)$	25

6. Прорачун

6.1. Сортирање података по растућем реду је већ урађено

6.2. Одређивање теоријских вредности расподеле на основу таблица

x_i одговара средини интервала

$$\forall x_i \in (-4,+4) \Rightarrow \quad F_{0i} = \Phi(x_i) \sim \mathcal{N}(0,1^2)$$

нпр. $\quad x_2 = -1.75 \Rightarrow \Phi(-1.75) = 0.04006$

6.3. Одређивање емпиријских вредности расподеле на основу формуле

$$F_n(x_k) = \frac{1}{n} \cdot \left(\sum_{i=0}^{k-1} n_i + 0.5 \cdot n_k \right)$$ нпр.

$$F_n(x_4) = \frac{1}{1000}\left(21 + 47 + 87 + \frac{158}{2}\right) = 0.234$$

6.4. Одређивање прорачунске вредности статистике одлучивања

$$D_i = \left|F_n(x_i) - F_0(x_i)\right|$$ нпр. $\quad D_1 = \left|0.0105 - 0.0122\right| = 0.0017$

6.5. Одређивање D_n

$$D_n = \max_x \left| F_n(x) - F_0(x) \right|$$

$$D_n = 0.00885$$

$$\lambda_0 = D_n \sqrt{n} = 0.28$$

$$\lambda_\alpha = \lambda_{0.05} = 1.355$$

7. Одлука

$$\lambda_\alpha > \lambda_0 \Leftrightarrow 1.355 > 0.28$$

M. Нулта хипотеза се не може одбацити.

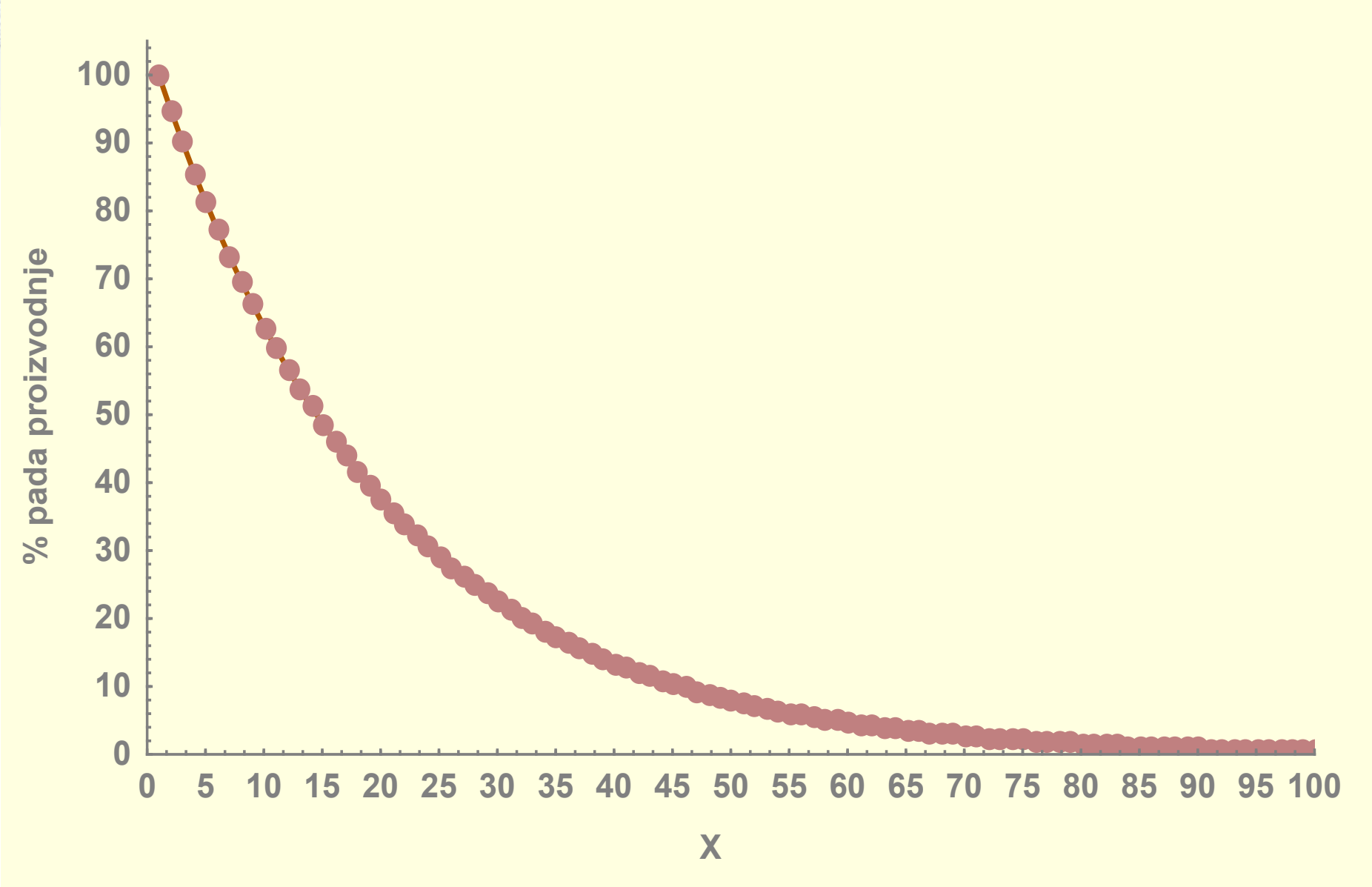
I. Подаци се распоређују по нормалној расподели

	x	$F_0(x)$	$F_n(x)$	$\lvert F_n(x) - F_0(x) \rvert$
1	-2.25	0.01222	0.0105	0.00172
2	-1.75	0.04006	0.0445	0.00444
3	-1.25	0.10565	0.1115	0.00585
4	-0.75	0.22663	0.2340	0.00737
5	-0.25	0.40129	0.4035	0.00221
6	0.25	0.59871	0.5945	0.00421
7	0.75	0.77337	0.7660	0.00737
8	1.25	0.89435	0.8855	$\boxed{0.00885}$
9	1.75	0.95994	0.9545	0.00544
10	2.25	0.98778	0.9875	0.00028

Тест Kolmogorova-интервали

Испитивана је поузданост револвер струга. Мерено је време рада између два отказа (у машинским часовима) и резултати су приказани у табели. Тестом Kolmogorova, користећи 7 интервала, испитати да ли се интензитет отказа понаша по експоненцијалном закону расподеле.

1.945	0.684	0.889	0.252	0.589
0.834	0.222	0.134	0.660	1.496
0.193	0.418	0.368	1.372	0.423
0.426	1.815	1.105	0.719	1.672
1.505	1.261	0.547	1.322	1.153
0.981	1.382	0.080	0.079	0.645
0.480	0.272	0.582	0.986	0.105
0.063	1.811	1.066	1.326	0.564
0.528	0.734	0.476	0.698	2.729
0.975	0.228	0.771	0.577	2.069


% pada proizvodnje
100
90
80
70
60
50
40
30
20
10
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
X

6. Прорачун

6.1. Одређивање граница интервала и формирање фреквенција

6.2. Одређивање теоријских вредности расподеле на основу таблица

x_i одговара средини интервала нпр. $(0.000, 0.375) \Rightarrow x_1 = 0.1875$

$$\bar{x} = \frac{1}{50} \cdot (11 \cdot 0.1875 + \cdots + 1 \cdot 2.4375) = \frac{41.625}{50} = 0.8325 \Rightarrow$$

$$\bar{x} \approx \frac{1}{\lambda}$$

$$\lambda \approx 1.201$$

$$F_0(x) = 1 - e^{-1.201 \cdot x}$$

$$F(0.375) = 1 - e^{-1.201 \cdot 0.375}$$

$$F(0.375) = 0.363$$

	интервал	f_i	x_i	$F_0(x)$	$F_n(x)$	
1	$(0.000, 0.375)$	11	0.1875	0.363	0.22	0.143
2	$[0.375, 0.750)$	17	0.5625	0.594	0.56	0.034
3	$[0.750, 1.125)$	8	0.9375	0.741	0.72	0.021
4	$[1.125, 1.500)$	7	1.3125	0.835	0.86	0.025
5	$[1.500, 1.875)$	4	1.6875	0.895	0.94	0.045
6	$[1.875, 2.250)$	2	2.0625	0.933	0.98	0.047
7	$[2.250, \infty)$	1	2.4375	0.957	1	0.043

6.3. Одређивање емпиријских вредности расподеле на основу формуле

$$F_n(x_k) = \frac{\sum_{i=1}^{k} f_i}{n} \qquad \text{нпр.} \qquad F_n(x_1) = \frac{f_1}{n} = \frac{11}{50} = 0.22$$

6.4. Одређивање прорачунских вредности статистике одлучивања

$$D_i = \left|F_n(x_i) - F_0(x_i)\right| \qquad \text{нпр.} \qquad D_1 = \left|0.363 - 0.22\right| = 0.143$$

6.5. Одређивање D_n

$$D_n = 0.143$$

$$\lambda_0 = D_n\sqrt{n} = 1.011$$

$$\lambda_\alpha = \lambda_{0.05} = 1.355$$

7. Одлука

$$\lambda_\alpha > \lambda_0 \Leftrightarrow 1.355 > 1.011$$

M. Нулта хипотеза се не може одбацити.

I. Подаци се распоређују по експоненцијалној расподели.

	interval	f_i	x_i	$F_0(x)$	$F_n(x)$	$\lvert F_n(x) - F_0(x)\rvert$
1	$(0.000, 0.375)$	11	0.1875	0.363	0.22	$\boxed{0.143}$
2	$[0.375, 0.750)$	17	0.5625	0.594	0.56	0.034
3	$[0.750, 1.125)$	8	0.9375	0.741	0.72	0.021
4	$[1.125, 1.500)$	7	1.3125	0.835	0.86	0.025
5	$[1.500, 1.875)$	4	1.6875	0.895	0.94	0.045
6	$[1.875, 2.250)$	2	2.0625	0.933	0.98	0.047
7	$[2.250, \infty)$	1	2.4375	0.957	1	0.043

Тест Kolmogorov-Smirnov-а

Испитује да ли су две расподеле исте или различите.

Поступак је исти као код теста Kolmogorova, стим што је

Систем хипотеза

$$H_0\left(F_1(x) = F_2(x)\right)$$

$$H_1\left(F_1(x) \neq F_2(x)\right)$$

Статистика одлучивања

$$D_n = \max_x \left|F_1(x) - F_2(x)\right|$$

U* тест Mann-Withney-а

Испитује се да ли су две расподеле исте или различите.

Утицај различите величине узорака је занемарљив.

Пример 4

Дате су две групе података упоредити да ли су им расподеле исте.

x	22	25	25	19	19	21	19	20	24		
y	17	13	19	14	17	15	13	18	14	20	18

Процедура тестирања хипотеза

1. Поређење једнакости две расподеле без обзира на њихов облик

2. Постављање система хипотеза

$$H_0\left(F_1(x)=F_2(y)\right)$$

$$H_1\left(F_1(x)\neq F_2(y)\right)$$

3. Праг значајности: усваја се $\alpha = 0.05$

4. Статистика одлучивања: $Z^* = \dfrac{U^* - \mu_0}{\sigma_0} \sim \mathbb{N}\left(0,1^2\right)$

5. Критична област одбацивања H_0 $|z^*| > z_{\alpha/2}$

H_0	H_1	КО
$H_0 : (F_x = F_y)$	$H_1 : (F_x \neq F_y)$	$\lvert z^*\rvert > z_{\alpha/2}$
	$H_1 : (F_x < F_y)$	$z^* < -z_\alpha$
	$H_1 : (F_x > F_y)$	$z^* > z_\alpha$

6. Прорачун

6.1. Од података се формира један растући низ, без обзира ком узорку припадају

6.2. Подаци се рангирају у табели

13	y	1.5	*1*	19	x	11.5	*11*
13	y	1.5	*2*	19	x	11.5	*12*
14	y	3.5	*3*	19	y	11.5	*13*
14	y	3.5	*4*	20	x	14.5	*14*
15	y	5	*5*	20	y	14.5	*15*
17	y	6.5	*6*	21	x	16	*16*
17	y	6.5	*7*	22	x	17	*17*
18	y	8.5	*8*	24	x	18	*18*
18	y	8.5	*9*	25	x	19.5	*19*
19	x	11.5	*10*	25	x	19.5	*20*

6.3. Рангирани подаци се раздвајају према узорцима (табела) и добијају се суме рангова T_x и T_y

x	11.5	y	1.5
x	11.5	y	1.5
x	11.5	y	3.5
x	14.5	y	3.5
x	16	y	5
x	17	y	6.5
x	18	y	6.5
x	19.5	y	8.5
x	19.5	y	8.5
		y	11.5
		y	14.5
T_x=139		T_y=71	

6.4. Тражи се вредност U^*

$$U^* = T_x - \frac{n_1(n_1+1)}{2} = n_1 n_2 + \frac{n_2(n_2+1)}{2} - T_y$$

$$u^* = 139 - \frac{9 \cdot 10}{2} = 94$$

6.5. Траже се апроксимације средње вредности и варијансе

$$\mu_0 = E(U^*) = \frac{n_1 n_2}{2} = \frac{11 \cdot 9}{2} = 49.5$$

$$\sigma_0^2 = V(U^*) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12} = \frac{9 \cdot 11 \cdot 21}{12} = 173.25$$

6.6. Одређује се вредност статистике одлучивања

$$Z^* = \frac{U^* - \mu_0}{\sigma_0} \sim \mathcal{N}(0, 1^2)$$

$$z^* = \frac{94 - 49.5}{13.162} = 3.378$$

6.7. Одређује се таблична вредност нормалне расподеле

$$z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$$

7. Одлука

$$z_{\alpha/2} > z^* \Rightarrow 3.378 > 1.96$$

М. Нулта хипотеза се не може прихватити.

I. То значи да расподеле нису исте.

Тест знакова за медијану

Користи се за тестирање вредности медијане континуалне расподеле. Медијана је место груписања највећег броја вредности расподеле, односно место које дели податке на два дела. То значи да важи:

$$P\left(x_i \le \tilde{\mu}\right) = P\left(x_i \ge \tilde{\mu}\right) = 0.5$$

Идеја је да се броји број + знакова и да се претпоставља да он одговара биномној расподели

$$B\left(n, p\right), p = 0.5$$

Систем тестирања хипотеза:

H_0	H_1	статистика одлучивања - p-ниво	услов	КО H_0
$\tilde{\mu} = \tilde{\mu}_0$	$\tilde{\mu} \ne \tilde{\mu}_0$	$p_{(\alpha)} = 2P\left(R^+ \le r^+, p = 0.5\right)$	$r^+ < \frac{n}{2}$	$p_{(\alpha)} < \alpha$
		$p_{(\alpha)} = 2P\left(R^+ \ge r^+, p = 0.5\right)$	$r^+ > \frac{n}{2}$	
	$\tilde{\mu} < \tilde{\mu}_0$	$p_{(\alpha)} = P\left(R^+ \le r^+, p = 0.5\right)$		
	$\tilde{\mu} > \tilde{\mu}_0$	$p_{(\alpha)} = P\left(R^+ \ge r^+, p = 0.5\right)$		

$$p_{(\alpha)} = \sum_{i=r^+}^{n} \binom{n}{i} p^i (1-p)^{n-i}, p = 0.5$$

Када је $n > 10 \Rightarrow B(n,p) \sim N$

H_0	H_1	КО	Статистика одлучивања- p - ниво
$\tilde{\mu} = \tilde{\mu}_0$	$\tilde{\mu} \neq \tilde{\mu}_0$	$\lvert z_0 \rvert > z_{\alpha/2}$	$z_0 = \dfrac{R^+ - 0.5n}{0.5\sqrt{n}}$
	$\tilde{\mu} < \tilde{\mu}_0$	$z_0 > z_\alpha$	
	$\tilde{\mu} > \tilde{\mu}_0$	$z_0 < -z_\alpha$	

Пример 5

Мерена је средња потрошња горива код 20 случајно изабраних аутомобила истог модела. Подаци су приказани у табели. Да ли је најчешћа просечна потрошња 10l?

Аут.	1	2	3	4	5	6	7	8	9	10
пот.	10.8	8.4	11.6	10.3	11	8.8	8.9	12.9	11.8	11.3
Аут.	11	12	13	14	15	16	17	18	19	20
пот.	10.8	12	8.9	11.7	8.8	10.3	12.1	11	13.3	8.8

1. Дефинисање теста – вредност медијане

2. Систем хипотеза

$$H_0\left(\tilde{\mu} = 10\right)$$

$$H_1\left(\tilde{\mu} \neq 10\right)$$

3. Ниво значајности $\alpha = 0.05$

4. Статистика одлучивања $r^+ = \sum(+), p_{(\alpha)}$

5. Критеријум одбацивања H_0 $p_{(\alpha)} < \alpha$

6. Прорачун

Аут.	1	2	3	4	5	6	7	8	9	10
пот.	10.8	8.4	11.6	10.3	11	8.8	8.9	12.9	11.8	11.3
r	+	-	+	+	+	-	-	+	+	+
Аут.	11	12	13	14	15	16	17	18	19	20
пот.	10.8	12	8.9	11.7	8.8	10.3	12.1	11	13.3	8.8
r	+	+	-	+	-	+	+	+	+	-

$$r^+ = 14$$

$$p_{(\alpha)} = 2P\left(R^{+} \geq 14, p = 0.5\right) = \sum_{i=14}^{20} \binom{20}{i} (0.5)^{i} (0.5)^{20-i} = 0.1153$$

7. Одлука

$$p_{(\alpha)} = 0.1153 > \alpha = 0.05$$

М. Значи да се нулта хипотеза не може одбацити.
I. Закључује се да је најчешћа потрошња горива 10l

Пример 6

Када се у примеру 1 примени апроксимација биномне расподеле нормалном добија се:

$$z_0 = \frac{r^{+} - pn}{p\sqrt{n}} = \frac{14 - 0.5 \cdot 20}{0.5\sqrt{20}} = 1.789 < 1.96 = z_{\frac{0.05}{2}}$$

па се нулта хипотеза не може одбацити, односно закључује се да је најчешћа потрошња горива 10l

Тест знакова за разлику медијана

Тест знакова за тестирање разлике медијана примењује се у случајевима поређења највећег груписања тачака континуалних асиметричних расподела.

Када су расподеле симетричне, поређење медијана се своди на поређење аритметичких средина

2. Систем тестирања хипотеза је

$$H_0\left(\tilde{\mu}_1 = \tilde{\mu}_2\right)$$

$$H_1\left(\tilde{\mu}_1 \neq \tilde{\mu}_2\right) \Leftrightarrow H_1\left(\tilde{\mu}_d \neq 0\right)$$

3. Ниво значајности може бити 0.1, 0.05 і 0.01 за двострани, односно 0.05, 0.025 і 0.005, за једнострани тест

4. Статистика одлучивања је $R = min\left(R^-, R^+\right)$

5. Критеријум одбацивања нулте хипотезе је $R \leq r_\alpha^*$

Пример 7

Тестира се потрошња горива код два типа аутомобила исте класе. Мерења 12 слулајно изабраних аутомобила сваког типа су приказана у табели. Да ли аутомобили имају исту потрошњу, ако се не зна да ли се потрошња горива понаша по нормалној расподели?

I тип	7.6	9.2	9.5	7.9	5.3	6.9	5.3	9.6	8.4	7.3	10.2	8.2
II тип	6.8	10	8.4	6.6	6.2	6.4	5.9	6.8	8.2	8.9	8.8	7.9

1. Тестира се разлика медијана

2. Систем хипотеза је

$$H_0\left(\tilde{\mu}_1 = \tilde{\mu}_2\right)$$

$$H_1\left(\tilde{\mu}_1 \neq \tilde{\mu}_2\right) \Leftrightarrow H_1\left(\tilde{\mu}_d \neq 0\right) \quad d_i = x_{1i} - x_{2i}$$

3. Ниво значајности – усваја се $\alpha = 0.05$

4. Статистика одлучивања $r = min\left(r^-, r^+\right)$

5. Критеријум одбацивања нулте хипотезе $r \leq r^*_{0.05}$

6. Прорачун

I тип	7.6	9.2	9.5	7.9	5.3	6.9	5.3	9.6	8.4	7.3	10.2	8.2
II тип	6.8	10	8.4	6.6	6.2	6.4	5.9	6.8	8.2	8.9	8.8	7.9
d	0.8	-0.8	0.9	1.3	-0.9	0.5	-0.6	2.8	0.2	-1.6	1.4	0.3
r	+	-	+	+	-	+	-	+	+	-	+	+

$$r^{+} = 8 \qquad r^{-} = 4 \qquad \Rightarrow r = min\left(r^{-}, r^{+}\right) = min\left(4, 8\right) = 4$$

$$r^{*}_{0.05} = 2$$

7. Одлука

$$r = 4 > r^{*}_{0.05} = 2$$

M. Нулта хипотеза се не може одбацити.

I. Закључује се да оба типа возила имају исту потрошњу

Домаћи задатак 6

- Урадити одговарајући задатак из фајла окаченог на сајту предмета

1) Mann-Whitney U* тест
2) Тест знакова за медијану
3) Тест знакова за разлику медијана