5.3 Тестирање на непараметарски хипотези

Задача 5.20. Со Пирсоновиот χ^2 критериум, со ниво на значајност $\alpha = 0,05$, провери ја хипотезата дека податоците

x_k	(1; 1,2)	(1,2; 1,4)	(1,4; 1,6)	(1,6; 1,8)	(1,8; 2,0)
m_k	33	23	20	15	9

одговараат на обележје X со функција на распределба

$$F(x) = \begin{cases} 0 & , x < 1 \\ \sqrt{x-1} & , 1 \leq x < 2 \\ 1 & , x \geq 2 \end{cases} .$$

Решение. Ја тестираме хипотезата $H_0 : X$ има функција на распределба $F(x)$, против $H_1 : X$ нема функција на распределба $F(x)$. Врз основа на дадените податоци, ја пресметуваме вредноста на Пирсоновата χ^2 тест статистика со пополнување на следната табела:

x_k	(1; 1,2)	(1,2; 1,4)	(1,4; 1,6)	(1,6; 1,8)	(1,8; 2,0)
m_k	33	23	20	15	9
p_k	0,447213	0,185242	0,142141	0,119831	0,105573
np_k	44,7213	18,5242	14,2141	11,9831	10,5573
$(m_k - np_k)^2$	137,389	20,0328	33,4766	9,10169	2,42518
$\frac{(m_k-np_k)^2}{np_k}$	3,07212	1,08144	2,35517	0,759544	0,229716

каде $n = 100$, а веројатностите p_k, $k = 1, ..., 5$ се пресметуваат откако $\mathbb{R}$ ќе се разбие на подинтервали соодветни на дадените податоци т.е. $(-\infty; 1,2]$, $(1,2; 1,4]$, $(1,4; 1,6]$, $(1,6; 1,8]$, $(1,8; +\infty)$, па веројатноста p_k е веројатност да X прима вредности од k-тиот подинтервал, под претпоставка дека H_0 е точна, односно:

$$p_1 = P\{X \in (-\infty; 1,2] | H_0\} = F(1,2) - F(-\infty) = \sqrt{1,2-1} - 0 = 0,447213,$$
$$p_2 = P\{X \in (1,2; 1,4] | H_0\} = F(1,4) - F(1,2) = \sqrt{1,4-1} - \sqrt{1,2-1} = 0,185242,$$

и.т.н. Тогаш, Пирсоновата χ^2 статистика (збирот на броевите од последната редица од табелата) е

$$\overline{\chi}^2 = \sum_{k=1}^{5} \frac{(m_k - np_k)^2}{np_k} = 7,49799.$$

Критичната област го има обликот

$$C = \{x \mid \overline{\chi}^2 = \sum_{i=1}^{r} \frac{(m_i - np_i)^2}{np_i} > \chi^2_{r-1,\alpha}\},$$

каде $r = 5$ е бројот на подинтервали, $\alpha = 0,05$, па од таблица наоѓаме $\chi^2_{r-1,\alpha} = \chi^2_{5-1;0,05} = \chi^2_{4;0,05} = 9,488$. Бидејќи, $\overline{\chi}^2 = 7,49799 < 9,488 = \chi^2_{r-1,\alpha}$ следи дека $x \notin C$, па хипотезата H_0 се прифаќа.

Задача 5.21. Во текот на една година забележан е бројот на итни повици неделно за медицинска интервенција и добиени се следните резултати:

број на итни повици	0	1	2	3	4	5
број на недели	6	10	20	10	6	0

Со Пирсоновиот χ^2 критериум, со ниво на значајност $\alpha = 5\%$, провери ја хипотезата дека бројот на итни повици неделно има Поасонова распределба.

Решение. Нека X е број на итни повици неделно за медицинска интервенција. Треба да се тестира хипотезата $H_0 : X$ има $\mathcal{P}(\lambda)$ распределба, против $H_1 : X$ нема $\mathcal{P}(\lambda)$ распределба, каде $\lambda > 0$ е непознат параметар кој треба прво да се оцени. Максимално подобен оценувач за λ е средината на примерокот, па за оценувач за λ ја земаме аритметичката средина на податоците т.е. $\lambda = \overline{x} = 2$. Тогаш, H_0 преминува во $H_0 : X \sim \mathcal{P}(2)$. Врз основа на дадените податоци, ја пресметуваме Пирсоновата χ^2 тест статистика со пополнување на следната табела:

x_k	0	1	2	3	4	5
m_k	6	10	20	10	6	0
p_k	0,135335	0,270671	0,270671	0,180447	0,090223	0,052653
np_k	7,03742	14,0749	14,0749	9,38324	4,6916	2,73796
$(m_k - np_k)^2$	1,07624	16,6047	35,1069	0,380388	1,71192	7,4964
$\frac{(m_k-np_k)^2}{np_k}$	0,152931	1,17974	2,49429	0,0405391	0,364891	2,73796

пришто $n = 52$, а за да ги пресметаме веројатностите p_k, $k = 1, ..., 6$, просторот $\mathbb{R}$ го разбиваме на подинтервали соодветни на дадените податоци т.е. $(\infty, 0]$, $(0, 1]$, $(1, 2]$, $(2, 3]$, $(3, 4]$, $(4, +\infty)$, па затоа:

$$p_1 = P\{X \in (-\infty; 0]|H_0\} = P\{X = 0|H_0\} = \frac{2^0}{0!}\, e^{-2} = 0,135335,$$

$$p_2 = P\{X \in (0, 1]|H_0\} = P\{X = 1|H_0\} = \frac{2^1}{1!}\, e^{-2} = 0,270671,$$

и.т.н. последната веројатност е

$$p_6 = P\{X \in (4, +\infty)|H_0\} = 1 - P\{X \in \{0,1,2,3,4\}|H_0\} = 1 - 0,947347 = 0,052653.$$

Тогаш, Пирсоновата χ^2 статистика е

$$\overline{\chi}^2 = \sum_{k=1}^{6} \frac{(m_k - np_k)^2}{np_k} = 6,97035.$$

Критичната област го има обликот

$$C = \{x \mid \overline{\chi}^2 = \sum_{i=1}^{r} \frac{(m_i - np_i)^2}{np_i} > \chi^2_{r-j-1,\alpha}\},$$

каде $r = 6$ е бројот на подинтервали, $j = 1$ е бројот на параметри кои требаше да се оценат, $\alpha = 5\% = 0,05$, па од таблица наоѓаме $\chi^2_{r-j-1,\alpha} = \chi^2_{6-1-1;0,05} = \chi^2_{4;0,05} = 9,488$. Бидејќи, $\overline{\chi}^2 = 6,97035 < 9,488 = \chi^2_{r-j-1,\alpha}$ следи дека $x \notin C$, па хипотезата H_0 се прифаќа.

Задача 5.22. Врз основа на податоците од следната табела:

x_k	(0, 5)	(5, 10)	(10, 15)	(15, 20)	(20, 25)
m_k	15	75	100	50	20

со Пирсоновиот χ^2 критериум, со ниво на значајност $\alpha = 0,05$, провери ја хипотезата дека податоците одговараат на обележје со нормална распределба.

Решение. Врз основа на дадените податоци за обележјето X, треба да се тестира хипотезата дека $H_0 : X$ има $\mathcal{N}(a, \sigma^2)$ распределба, против $H_1 : X$ нема $\mathcal{N}(a, \sigma^2)$ распределба, каде $a \in \mathbb{R}$ и $\sigma^2 > 0$ се непознати параметри кои треба прво да се оценат. Максимално подобни оценки за a и σ^2 се аритметичката средина и дисперзијата на податоцоте т.е. $a = \overline{x} = 12,21$ и $\sigma^2 = \overline{s}^2 = 25,4351$, односно $\sigma = 5,04$. Тогаш, H_0 преминува во $H_0 : \mathcal{N}(12,21; 5,04^2)$. Врз основа на дадените податоци, ја пресметуваме вредноста на Пирсоновата χ^2 тест статистика со пополнување на следната табела:

x_k	(0, 5)	(5, 10)	(10, 15)	(15, 20)	(10, 25)
m_k	15	75	100	50	20
p_k	0,07636	0,25361	0,37887	0,23059	0,06057
np_k	19,8536	65,9386	98,5062	59,9534	15,7482
$(m_k - np_k)^2$	23,5574	82,109	2,23144	99,0702	18,0778
$\frac{(m_k-np_k)^2}{np_k}$	1,18656	1,24523	0,0226528	1,65245	1,14793

каде $n = 260$, а веројатностите p_k, $k = 1, ..., 5$ се пресметуваат откако $\mathbb{R}$ ќе се разбие на подинтервали соодветни на дадените податоци т.е. $(-\infty, 5]$, $(5, 10]$, $(10, 15]$, $(15, 20]$, $(20, +\infty)$, па веројатноста p_k е веројатност да X прима вредности од k-тиот подинтервал, под претпоставка дека H_0 е точна, односно:

$$\begin{aligned}
p_1 &= P\{X \in (-\infty; 5] | H_0\} = P\{X \leq 5 | H_0\} = P\{\frac{X - 12,21}{5,04} \leq \frac{5 - 12,21}{5,04} | H_0\}, \\
&= 0,5 + \Phi_0(\frac{5 - 12,21}{5,04}) = 0,5 + \Phi_0(-1,43) = 0,5 - \Phi_0(1,43) = \\
&= 0,5 - 0,42364 = 0,07636, \\
p_2 &= P\{X \in (5; 10] | H_0\} = P\{5 < X \leq 10 | H_0\} = \\
&= P\{\frac{5 - 12,21}{5,04} < \frac{X - 12,21}{5,04} \leq \frac{10 - 12,21}{5,04} | H_0\} =, \\
&= \Phi_0(\frac{10 - 12,21}{5,04}) - \Phi_0(\frac{5 - 12,21}{5,04}) = \Phi_0(-0,44) - \Phi_0(-1,43) = \\
&= -\Phi_0(0,44) + \Phi_0(1,43) = -0,17003 + 0,42364 = 0,25361,
\end{aligned}$$

и.т.н. затоа што, кога H_0 е точна имаме дека $\frac{X-12,21}{5,04} \sim \mathcal{N}(0,1)$, па вредностите за функцијата $\Phi_0(x)$ ги читаме од таблицата за стандардна нормалана распределба. Тогаш, Пирсоновата χ^2 статистика е

$$\overline{\chi}^2 = \sum_{k=1}^{5} \frac{(m_k - np_k)^2}{np_k} = 5,25482.$$

Критичната област го има обликот

$$C = \{x \mid \overline{\chi}^2 = \sum_{i=1}^{r} \frac{(m_i - np_i)^2}{np_i} > \chi^2_{r-j-1,\alpha}\},$$

каде $r = 5$ е бројот на подинтервали, $j = 2$ е бројот на параметри кои требаше да се оценат, $\alpha = 0,05$, па од таблица наоѓаме $\chi^2_{r-j-1,\alpha} = \chi^2_{5-2-1;0,05} = \chi^2_{2;0,05} = 5,991$. Бидејќи, $\overline{\chi}^2 = 5,25482 < 5,991 = \chi^2_{r-j-1,\alpha}$ следи дека $x \notin C$, па хипотезата H_0 се прифаќа.

Задача 5.23. Во една таблица на 200 случајни едноцифрени броеви, цифрите 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 се појавуваат 18, 23, 22, 17, 20, 25, 14, 19, 22, 20 пати соодветно. Со ниво на значајност $\alpha = 0,05$, провери ја хипотезата дека во таблицата на случајни едноцифрени броеви, цифрите 0-9 се рамномерно распределени.

Решение. Нека X е вредност на случајно избрана цифра од 0 до 9. Врз основа на дадените податоци, треба да ја тестираме хипотезата $H_0 : X$ има

дискретна рамномерна распределба на множеството $\{0, 1, ..., 9\}$, против H_1 : X нема дискретна рамномерна распределба на множеството $\{0, 1, ..., 9\}$.
Ја пресметуваме Пирсоновата χ^2 тест статистика со пополнување на следната табела:

x_k	0	1	2	3	4	5	6	7	8	9
m_k	18	23	22	17	20	25	14	19	22	20
p_k	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1
np_k	20	20	20	20	20	20	20	20	20	20
$(m_k - np_k)^2$	4	9	4	9	0	25	36	1	4	0
$\frac{(m_k-np_k)^2}{np_k}$	0,2	0,45	0,2	0,45	0	1,25	1,8	0,05	0,2	0

пришто $n = 200$, а за да ги пресметаме веројатностите p_k, $k = 1, ..., 10$, просторот $\mathbb{R}$ го разбиваме на подинтервали соодветни на дадените податоци т.е. $(\infty, 0]$, $(0, 1]$, $(1, 2]$, ..., $(7, 8]$, $(8, +\infty)$, па затоа

$$p_1 = P\{X \in (-\infty; 0]|H_0\} = P\{X = 0|H_0\} = 0,1,$$
$$p_2 = P\{X \in (0, 1]|H_0\} = P\{X = 1|H_0\} = 0,1,$$

и.т.н. односно $p_k = 0,1$, $k = 1, ..., 10$. Тогаш, Пирсоновата χ^2 статистика е

$$\overline{\chi}^2 = \sum_{k=1}^{10} \frac{(m_k - np_k)^2}{np_k} = 4,6.$$

Критичната област го има обликот

$$C = \{x \mid \overline{\chi}^2 = \sum_{i=1}^{r} \frac{(m_i - np_i)^2}{np_i} > \chi^2_{r-1,\alpha}\},$$

каде $r = 10$ е бројот на подинтервали, $\alpha = 0,05$, па од таблица наоѓаме $\chi^2_{r-1,\alpha} = \chi^2_{10-1;0,05} = \chi^2_{9;0,05} = 16,919$. Бидејќи, $\overline{\chi}^2 = 4,6 < 16,919 = \chi^2_{r-1,\alpha}$ следи дека $x \notin C$, па хипотезата H_0 се прифаќа.

Задача 5.24. Со критериумот на Колмогоров и ниво на значајност $\alpha = 0,01$, врз основа на податоците од табелата:

x_k	(0, 20)	(20, 40)	(40, 60)	(60, 80)	(80, 100)
m_k	2	8	13	17	10

тестирај ја хипотезата дека податоците одговараат на обележје со $\mathcal{U}(0, 100)$ распределба.

Решение. Нека X е обележјето, треба врз основа на дадените податоци да се тестира хипотезата $H_0 : X \sim \mathcal{U}(0, 100)$, против $H_1 : X$ нема $\mathcal{U}(0, 100)$ распределба. Ако H_0 е точна, густината на распределба X е

$$p(x) = \frac{1}{100},\ 0 < x < 100,$$

а функцијата на распределба на X е

$$F(x) = \begin{cases} 0 & , x \leq 0 \\ \frac{x}{100} & , 0 < x < 100 \\ 1 & , x \geq 100 \end{cases}.$$

За дадените податоци имаме $n = 50$, па емпириската функција на распределба $F_n(x)$ е

$$F_n(x) = \begin{cases} 0 & , x < 20 \\ 0,04 & , 20 \leq x < 40 \\ 0,2 & , 40 \leq x < 60 \\ 0,46 & , 60 \leq x < 80 \\ 0,8 & , 80 \leq x < 100 \\ 1 & , x \geq 100 \end{cases}.$$

Статистиката на Комогоров ја пресметуваме со помош на следната табела:

x_k	(0, 20)	(20, 40)	(40, 60)	(60, 80)	(80, 100)
m_k	2	8	13	17	10
a_k	20	40	60	80	100
$F_n(a_k)$	0,04	0,2	0,46	0,8	1
$F(a_k)$	0,2	0,4	0,6	0,8	1
$\lvert F_n(a_k) - F(a_k)\rvert$	0,16	0,2	0,14	0	0

каде a_k е десните граници на интервалите кај интервално зададените податоци. Тогаш, вредноста на статистиката на Колмогоров е (најголемата вредност во последниот ред од табелата)

$$d_n = \max_{1 \leq k \leq r} |F_n(a_k) - F(a_k)| = 0,2.$$

Критичната област го има обликот

$$C = \{x \mid d_n > d_{n,\alpha}\},$$

каде вредноста $d_{n,\alpha} = d_{50;0,01} = 0,226$ се чита од таблица. За дадените податоци имаме дека $d_n = 0,2 < 0,226 = d_{n,\alpha}$, од каде следи дека $x \notin C$, па хипотезата H_0 се прифаќа.

Ирена Стојковска

Задача 5.25. Извршени се мерења на висините на учениците во две различни училишта А и Б и при тоа се добиени следните резултати:

висина (во cm)	(160, 170)	(170, 180)	(180, 190)	(190, 200)
училиште А	25	65	8	2
училиште Б	16	34	6	4

Со ниво на значајност $\alpha = 0,01$, тестирај ја хипотезата дека распределбите на висините на учениците во училиштата А и Б се еднакви.

Решение. Нека X е висината на учениците во училиштето А, Y е висината на учениците во училиштето Б. Ја тестираме хипотезата $H_0 : X$ и Y имаат иста распределба, против $H_0 : X$ и Y немаат иста распределба. За дадените податоци имаме $n_1 = 100$ и $n_2 = 60$, па нивните емпириски функции на распределба се:

$$F_{n_1}(x) = \begin{cases} 0 & , x < 170 \\ 0,25 & , 170 \leq x < 180 \\ 0,90 & , 180 \leq x < 190 \\ 0,98 & , 190 \leq x < 200 \\ 1 & , x \geq 200 \end{cases} , \quad F_{n_2}(x) = \begin{cases} 0 & , x < 170 \\ 0,27 & , 170 \leq x < 180 \\ 0,83 & , 180 \leq x < 190 \\ 0,93 & , 190 \leq x < 200 \\ 1 & , x \geq 200 \end{cases}$$

соодветно. Статистиката на Комогоров-Смирнов ја пресметуваме со помош на следната табела:

x_k	(160, 170)	(170, 180)	(180, 190)	(190, 200)
m_k^1	25	65	8	2
m_k^2	16	34	6	4
a_k	170	180	190	200
$F_{n_1}(a_k)$	0,25	0,90	0,98	1
$F_{n_2}(a_k)$	0,27	0,83	0,93	1
$\lvert F_{n_1}(a_k) - F_{n_2}(a_k)\rvert$	0,02	0,07	0,05	0

каде a_k е десните граници на интервалите кај интервално зададените податоци. Тогаш, вредноста на статистиката на Колмогоров-Смирнов е (најголемата вредност во последниот ред од табелата)

$$d_{n_1,n_2} = \max_{1 \leq k \leq r} |F_{n_1}(a_k) - F_{n_2}(a_k)| = 0,07.$$

Критичната област го има обликот

$$C = \{x \mid d_{n_1,n_2} > d_{\frac{n_1 n_2}{n_1+n_2},\alpha}\},$$

каде вредноста $d_{\frac{n_1 n_2}{n_1+n_2},\alpha} = d_{37,5;0,01} = 0,2605$ се чита од таблица. За дадените податоци имаме дека $d_{n_1,n_2} = 0,07 < 0,2605 = d_{\frac{n_1 n_2}{n_1+n_2},\alpha}$, од каде следи дека $x \notin C$, па хипотезата H_0 се прифаќа.

Ирена Стојковска

Задача 5.26. Врз основа на податоците дадени во табелата, каде X е оценката по математика во последната година од средното образование и Y е оценката на испитот по математика на факултет, добиени од анкета спроведена на 150 студенти:

$X \backslash Y$	5	6	7	8	9	10
2	2	1	0	1	0	0
3	5	35	2	5	0	1
4	3	1	10	15	6	6
5	1	0	0	6	18	32

со ниво на значајност $\alpha = 0,05$, тестирај ја хипотезата дека обележјата X и Y се независни.

Решение. Треба да ја тестираме хипотезата $H_0 : X$ и Y се независни, против $H_1 : X$ и Y не се независни. Критичната област има облик

$$C = \{(x,y) \mid d = nf^2 > \chi^2_{(r-1)(s-1),\alpha}\},$$

каде $n = 150$ е вкупниот број на студенти, f^2 е отстапувањето од статистичката независност кое се пресметува според $f^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{f_{ij}^2}{g_i h_j} - 1 = 1,077947$, каде r е бројот на различни вредности кои ги прима X, тоа се $a_1, ..., a_r$, потоа s е бројот на различни вредности кои ги прима Y, тоа се $b_1, ..., b_s$, додека f_{ij}, g_i, h_j се честотите на (a_i, b_j), a_i, b_j соодветно. Вредноста $\chi^2_{(r-1)(s-1),\alpha} = \chi^2_{(4-1)(6-1);0,05} = \chi^2_{15;0,05} = 24,996$ ја читаме од таблица за χ^2 распределба. Бидејќи

$$d = nf^2 = 150 \cdot 1,077947 = 161,69205 > 24,996 = \chi^2_{(r-1)(s-1),\alpha},$$

следи дека $(x,y) \in C$, па хипотезата H_0 за независност на X и Y ја отфрламе.

Задачи за самостојна работа

Задача 5.27. Седум монети се фрлаат истовремено 1526 пати при што секој пат се разгледува бројот на паднати "грбови" и при тоа добиени се следните резултати:

број на паднати "грбови"	0	1	2	3	4	5	6	7
број на фрлања	12	78	270	456	386	252	69	3

Со Пирсоновиот χ^2 критериум, со ниво на значајност $\alpha = 5\%$, провери ја хипотезата дека бројот на паднати "грбови" при едно фрлање на 7 монети има биномна $\mathcal{B}(7, \frac{1}{2})$ распределба.

Ирена Стојковска

Задача 5.28. При заокружување на 45 броеви, направени се следните апсолутни грешки:

$$0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.07, 0.08, 0.09, 0.09, 0.10, 0.11, 0.13, 0.14, 0.14,$$
$$0.15, 0.16, 0.17, 0.17, 0.18, 0.21, 0.22, 0.22, 0.26, 0.27, 0.31, 0.32, 0.32, 0.33, 0.33,$$
$$0.34, 0.34, 0.35, 0.35, 0.36, 0.37, 0.37, 0.39, 0.41, 0.41, 0.42, 0.43, 0.44, 0.45, 0.48.$$

Со Пирсоновиот χ^2 критериум и ниво на значајност $\alpha = 0,05$, тестирај ја хипотезата дека апсолутната грешка има непрекината рамномерна распределба.

Задача 5.29. На случајно избрани 50 студенти измерена им е висината во cm и добиени се следните резултати:

$$146, 151, 151, 155, 156, 157, 158, 158, 158, 159, 160, 160, 161, 162, 163, 163, 164,$$
$$164, 165, 165, 167, 169, 169, 170, 171, 172, 172, 174, 174, 174, 175, 175, 175, 175,$$
$$177, 177, 178, 179, 179, 179, 179, 180, 180, 182, 185, 186, 187, 187, 189, 202.$$

Спореди ги, со ниво на значајност $\alpha = 0,05$, Пирсоновиот и критериумот на Колмогоров за тестирање на хипотезата дека висината на еден студент има нормална $\mathcal{N}(170, 10^2)$ распределба.

Задача 5.30. Случајно се избрани испитаници од две различни држави А и Б, кои се изјасниле за тоа колку часови дневно гледаат телевизија и добиени се следните резултати:

време на гледање TV (во h)	0	1	2	3	4	5	6	7	8
држава А	3	10	25	40	15	5	0	1	1
држава Б	1	8	30	44	10	6	1	0	0

Со ниво на значајност $\alpha = 0,05$, тестирај ја хипотезата дека распределбите на бројот на часови дневно поминати пред телевизорот за жителите од државите А и Б се еднакви.

Задача 5.31. При едно испитување направено на пазарот за посетеноста на супермаркетите, во еден супермаркет, во 40 случајно избрани денови, измерени се просечниот процент на намалување на цените и бројот на посетители во минута, при што се добиени следните резултати:

$$(0,12), (0,13), (0,13), (1,14), (1,14), (2,14), (2,14), (2,16), (3,17), (3,17),$$
$$(4,18), (4,18), (4,18), (4,18), (5,19), (5,19), (6,19), (6,19), (6,19), (7,19),$$
$$(7,20), (7,20), (7,20), (7,20), (7,20), (7,20), (8,21), (9,21), (10,21), (11,21),$$
$$(11,22), (13,22), (13,22), (14,23), (15,25), (16,26), (17,26), (18,27), (22,29), (26,30).$$

Со 5% ниво на значајност тестирај ја хипотезата за независност на просечниот процент на намалување на цената и бројот на посетители во минута.

Ирена Стојковска