Бизнис статистика

Непараметарски тестови
(Пирсонов Хи-квадрат тест)

Тестови на согласност

- Во првиот дел од ова предавање ќе разгледаме тестови со кои се проверува дали дадени податоци се во согласност со определена претпоставка за распределбата на обележјето на популацијата.
- Да го разгледаме случајот на тестирање на проста нулта наспроти сложена алтернативна хипотеза или:

H_0: Обележјето X има F_0 распределба

наспроти H_a: Обележјето X нема F_0 распределба

- Го делиме множеството R_X на вредности на обележјето X на r подмножества (класи) $S_1, \ldots, S_r$, кои се по парови дисјунктни подмножества, такви што веројатностите

$$p_k = P_{H_0}\{X \in S_k\}, \qquad k = 1, \ldots, r$$

се сите позитивни и $\sum_{k=1}^{r} p_k = 1$.

- Тестот за совпаѓање на распределби кој ќе го разгледаме, ги споредува очекуваните (теориски) честоти на класите со набљудуваните честоти класите добиени со примерокот. Целта е да се провери дали постои значајна разлика помеѓу очекуваните и набљудуваните честоти.

Тестови на согласност

- Ако, кога е точна нултата хипотеза веројатноста X да прими вредност од класата S_k е

$$p_k = P_{H_0}\{X \in S_k\}, \qquad k = 1, \dots, r$$

тогаш, ако се направени n набљудувања и ако е точна нултата хипотеза очекуваната честота на набљудувањата од примерокот кои се во класата S_k е еднаква на $E_k = np_k$, за $k = 1, \dots, r$.

- Нека $(X_1, \dots, X_n)$ е случаен примерок за X и нека $N_1, \dots, N_r$ се набљудуваните честоти од примерокот за класите $S_1, \dots, S_r$ (фрекфенции на податоците од примерокот во секоја од класите).

Пирсонов χ^2-тест

- Како мерка за разликите помеѓу набљудуваните честоти на примерокот $N_1, \ldots, N_r$ и претпоставените (очекуваните) честоти $np_1, \ldots, np_r$, кога е точна нултата хипотеза, се користи следната статистиката, позната како Пирсонов χ^2,

$$\chi^2 = \sum_{k=1}^{r} \frac{(N_k - np_k)^2}{np_k}$$

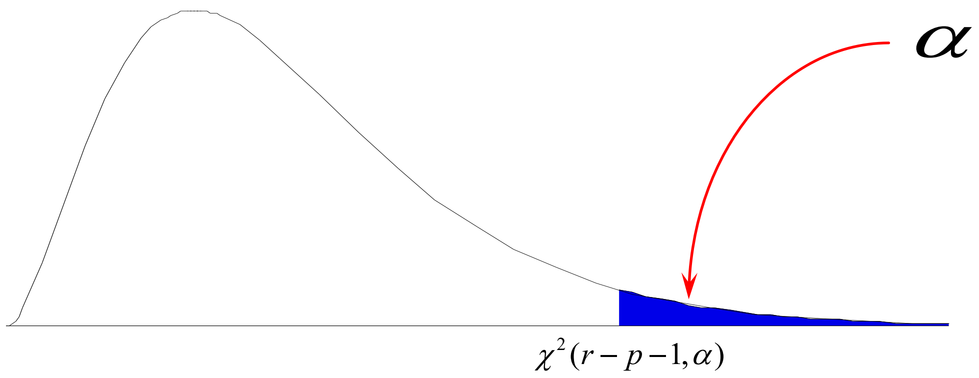
- Се покажува дека, кога е точна нултата хипотеза, оваа статистика има приближно χ^2 распределба со $r - p - 1$ степени на слобода, каде што p е бројот на непознати параметри во претпоставената распределба кои ќе бидат оцените со соодветна статистика од примерокот.
- Притоа, треба $r > p + 1$.
- На пример, ако распределбата во нултата хипотеза е $N(\mu, \sigma^2)$ и притоа и двата параметри се непознати, статистиката χ^2 , ќе има приближно χ^2 распределба со $r - 2 - 1$ степени на слобода.
- Непознатите параметри ќе ги оценуваме со вредностите на соодветните точкасти оценувачи.

Пирсонов χ^2-тест

- При примена на овој тест треба да се внимава на големината на очекуваните фреквенции.
- Ако очекуваните фреквенции np_k, $k = 1, \ldots, r$ се премногу мали, тест статистиката нема да ја одразува разликата на набљудуваните од очекуваните фрекфенции, туку само малата вредност на очекуваните фреквенции.
- Затоа, треба за секоја од класите, очекуваната фреквенција да биде барем 5.
- Бидејќи, ширината на класите не мора да биде еднаква, доколку очекуваната фреквенција за некоја од класите е помала од 5, тогаш таа класа се спојува со соседна класа.
- Притоа, очекуваната (набљудуваната) фрекфенција на новоформираната класа се добива со сумирање на очекуваните (набљудуваните) фреквенции на споените класи.
- Со ова и бројот на класи r ќе се намали за 1.

Пирсонов χ^2-тест

- Ако хипотезата H_0 е точна, очекуваме разликите помеѓу честотите во примерокот и претпоставените честоти да бидат мали. Затоа за големи вредности на статистиката ќе ја отфрлиме нултата хипотеза како неточна.
- Според тоа, критичниот домен на тестот е даден со $C = (\chi^2_{\alpha,r-p-1}, +\infty)$.



Пример 1.

Четири парички се фрлени 160 пати. Притоа, набљудувани се следните честоти на појавувања на грб:

k	0	1	2	3	4	Вкупно
N_k	6	34	67	41	12	160

Да се провери со 5% ниво на значајност дали се работи за фер парички!

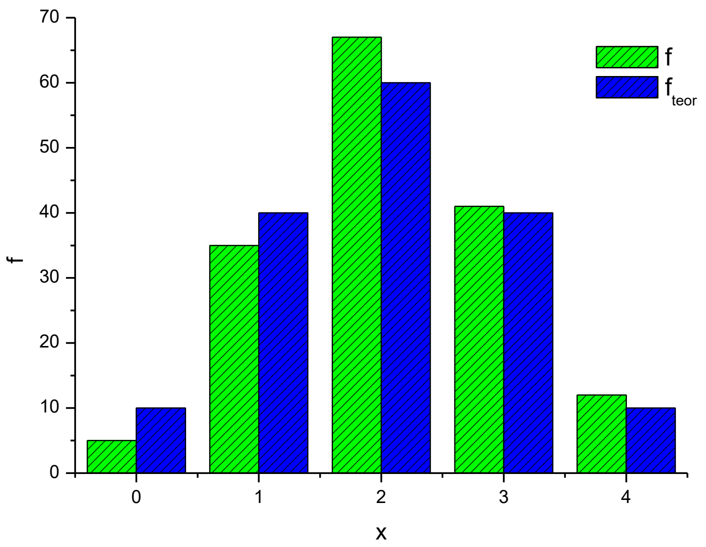
(ако паричките се фер, а X е бројот на појавувања на грб, очекуваме биномна распределба, т.е., $X \sim B(4,1/2)$).

Пример 1 – решение

- H_0: $X \sim B(4,1/2)$ (паричките се “фер”)
- Ако е точна H_0, тогаш

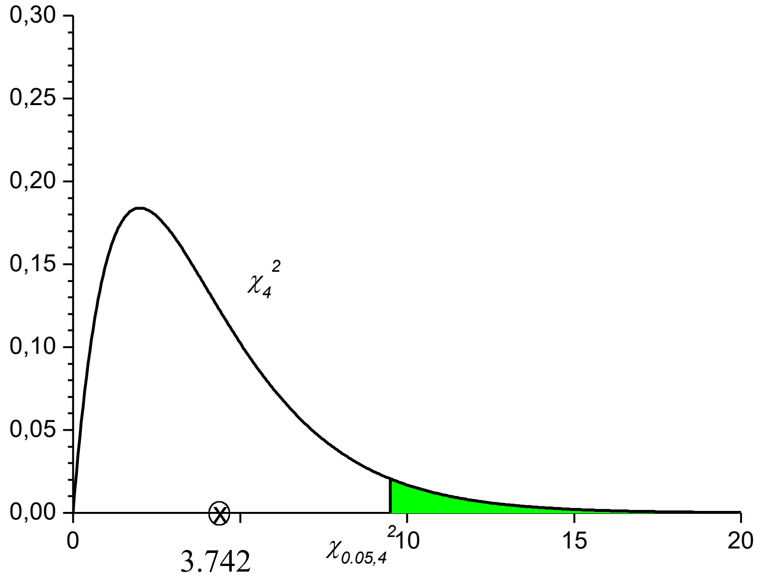
$$p_k = \binom{4}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{4-k} = \binom{4}{k}\left(\frac{1}{2}\right)^4, \quad k = 0,1,2,3,4.$$

S_k	N_k	np_k	$(N_k - np_k)^2/np_k$
0	6	10	1.6
1	34	40	0.9
2	67	60	0.817
3	41	40	0.025
4	12	10	0.4
вкупно	160	160	3.742



Пример 1 – решение

- Бидејќи имаме 5 класи, степените на слобода на тест статистиката ќе бидат 4.
- За 4 степени на слобода и 5% ниво на значајност, критичната вредност (од таблица) е $\chi^2_{0.05,4} = 9.49$, па критичниот домен е $C = (9.49, +\infty)$.
- Пресметаната вредност на тест статистиката 3.742 не припаѓа во критичниот домен. Затоа хипотезата H_0 не ја отфрламе и заклучуваме дека бројот на појавувања на грб, има биномна $B(4,1/2)$ распределба, т.е., паричките се фер.



Пример 2. Биномна распределба, непознато p

Од голем број производи се земаат случајни примероци со 20 парчиња. Експериментот се повторува 100 пати. Сакаме да ја процениме веројатноста за дефект и дали бројот на неисправни производи има биномна распределба. Набљудуваните честоти на неисправни производи во секој од избраните примероци со 20 производи се:

k	0	1	2	3	4	5	6	Вкупно
N_k	14	25	27	23	7	3	1	100

Со ниво на значајност 5% да се провери дали бројот на неисправни производи има биномна распределба.

Пример 2 – решение

Нека X е обележјето - број на неисправни производи во случаен примерок со 20 производи.

Треба да се тестира хипотезата H_0: $X \sim B(20, p)$, каде што параметарот p е непознат и треба да се оцени.

За да најдеме оценка на непознатиот параметарот p, тргнуваме од просекот $\bar{X}$ за кој знаеме дека е добар точкаст оценувач на математичкото очекување EX.

Ако $X \sim B(20, p)$, тогаш $EX = 20 \cdot p$. Оттука, добиваме дека оценувач за непознатиот параметар p е:

$$\hat{P} = \frac{\bar{X}}{20}$$

За примерокот (со обем 100), даден во табелата добиваме:

$$\bar{x} = 1.97, \quad \text{па} \quad \hat{p} = \frac{\bar{x}}{20} = 0.0985.$$

Значи, треба ќе ја тестираме хипотезата H_0: Распределбата на X е $B(20, 0.0985)$

Пример 2 – решение

H_0: $X \sim B(20,\ 0.0985)$

Последната класа треба да биде за $20 \geq k \geq 6$ за да унијата на сите класи биде еднаква на множеството вредности на $B(20,\ 0.0985)$ распределба и сумата на сите веројатности (во третата колона) да биде еднаква на 1.

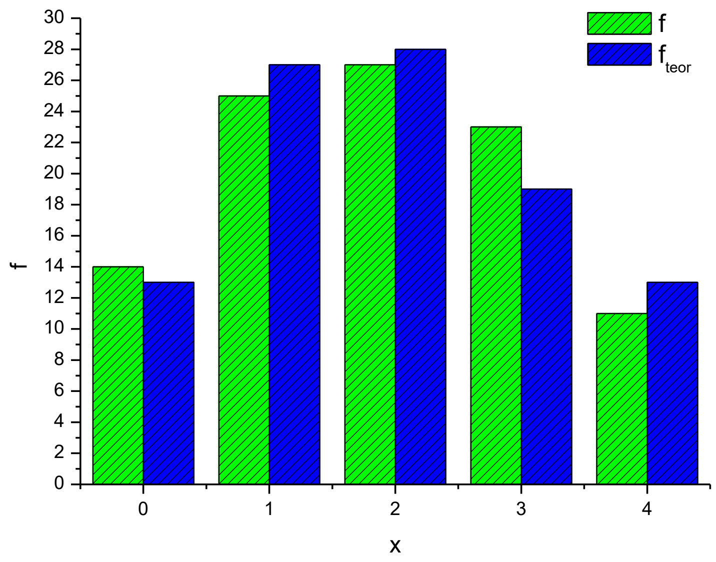
S_k	N_k	p_k	np_k
0	14	0,126	12,6
1	25	0,275	27,5
2	27	0,285	28,5
3	23	0,187	18,7
4	7	0,087	8,7
5	3	0,030	3,0
≥ 6	1	0,010	1,0
вкупно	100	1	100

Очекуваните честоти (np_k) за класите S_5 и S_6 се помали од 5 и затоа ги спојуваме со претходната класа S_4, во класа $S_{k \geq 4}$.

Пример 2 – решение

Значи, имаме пет класи и еден непознат оценет параметар, според тоа степените на слобода се $5 - 1 - 1 = 3$.

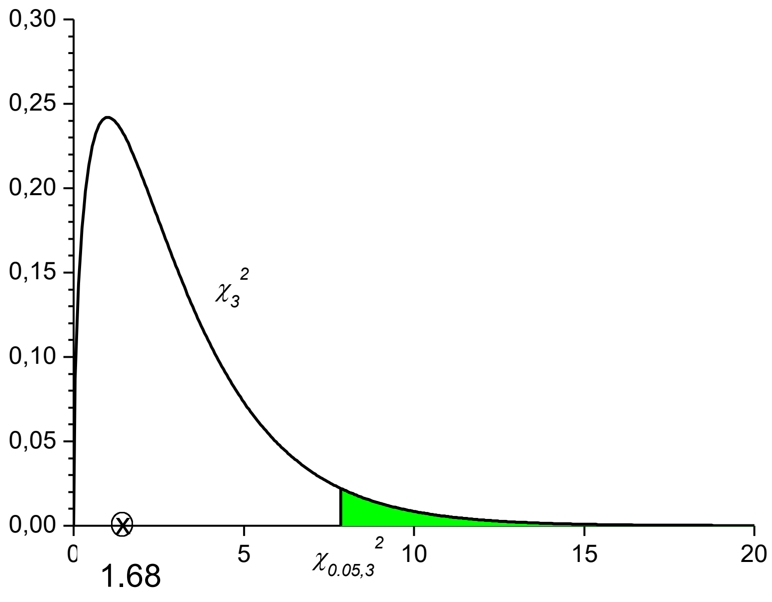
S_k	N_k	np_k	$(N_k - np_k)^2/np_k$
0	14	12.6	0.16
1	25	27.5	0.23
2	27	28.5	0.08
3	23	18.7	0.99
≥ 4	11	12.7	0.23
вкупно	100	100	1.68



Пример 2 – решение

За 3 степени на слобода и 5% ниво на значајност, критичната вредност (од таблица) е $\chi^2_{0.05,3}$= 7.82.

Пресметаната вредност на тест статистиката $\chi^2 = 1.68$ не е во критичниот домен. Затоа хипотезата H_0 не ја отфрламе, т.е. бројот на неисправни производи има биномна $B(20, 0.0985)$ распределба.



Пример 3. Пуасонова распределба

Во следната табела дадени се набљудувани честоти на бројот на расипувања на машини во некој индустриски погон.

Дневен број расипувања k	≤ 5	6	7	8	9	10	11	12	13	14	15	≥ 16
Честота N_k	11	13	20	18	27	34	28	12	13	10	7	7

Треба да се провери хипотезата дека распределбата на бројот на расипувањата во текот на денот е Пуасонова распределба или H_0: $X \sim P(\lambda)$.

Пример 3 – решение

Кога би било познато λ, би ги пресметале претпоставените веројатности на класите

$$p_k = P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$$

и очекуваните честоти np_k. Но, λ е непознато и затоа треба да се оцени од дадениот примерок.

Знаеме дека, ако $X \sim P(\lambda)$ тогаш $EX = \lambda$, а точкаст оценувач за математичко очекување е $\bar{X}$. Според тоа, непознатиот параметар λ ќе го оцениме со вредноста на $\bar{X}$. За дадениот примерок ја добиваме следната оценка:

$$\lambda = \bar{x} = \frac{1}{200}(5 \cdot 11 + 6 \cdot 13 + \ldots + 15 \cdot 7 + 16 \cdot 7) = 9.89 \approx 10.$$

Пример 3 – решение

Следно, ги пресметуваме претпоставените веројатности на класите, т.е. веројатностите X да прими вредност од соодветната класа, ако $X \sim P(10)$.

$$\hat{p}_0 = P\{X \in S_0 \mid \lambda = 10\} = P\{X \leq 5 \mid \lambda = 10\} = \sum_{k=0}^{5} P\{X = k \mid \lambda = 10\} = \sum_{k=0}^{5} \frac{10^k}{k!} \cdot e^{-10} = 0.0671$$

$$\hat{p}_1 = P\{X \in S_1 \mid \lambda = 10\} = P\{X = 6 \mid \lambda = 10\} = \frac{10^6}{6!} \cdot e^{-10} = 0.0631$$

…

$$\hat{p}_{10} = P\{X \in S_{10} \mid \lambda = 10\} = P\{X = 15 \mid \lambda = 10\} = \frac{10^{15}}{15!} \cdot e^{-10} = 0.0347$$

$$\hat{p}_{11} = P\{X \in S_{11} \mid \lambda = 10\} = P\{X \geq 16 \mid \lambda = 10\} = 1 - P\{X < 16 \mid \lambda = 10\} = 1 - \sum_{k=0}^{15} \frac{10^k}{k!} \cdot e^{-10} = 0.0487$$

Пример 3 – решение

k	S_k	N_k	p_k	np_k	$(N_k - np_k)^2 / np_k$
0	≤5	11	0,0671	13,42	0,436
1	6	13	0,0631	12,62	0,011
2	7	20	0,0901	18,02	0,218
3	8	18	0,1126	22,52	0,907
4	9	27	0,1251	25,02	0,157
5	10	34	0,1251	25,02	3,223
6	11	28	0,1137	22,74	1,217
7	12	12	0,0948	18,96	2,555
8	13	13	0,0729	14,58	0,171
9	14	10	0,0521	10,42	0,017
10	15	7	0,0347	6,94	0,001
11	≥16	7	0,0487	9,74	0,771
	Вкупно	200		200	**9,684**

Имаме 12 класи и еден оценет параметар, т.е. 12–1–1 = 10 степени на слобода.

За 10 степени на слобода и 5% ниво на значајност, критичната вредност (од таблица) е $\chi^2_{0.05,10} = 18.307$.

Пресметаната вредност на тест статистиката $\chi^2 = 9.684$ не е во критичниот домен.

Затоа, хипотезата H_0 не ја отфрламе, т.е. распределбата на бројот на расипувањата во текот на денот е Пуасонова.

Пример 4. Експоненцијална распределба

За ненегативната случајна променлива X, реалната оска е поделена на следните класи: $S_1=(-\infty, 1)$, $S_2=[1, 2)$, $S_3=[2, 3)$, $S_4=[3, 4)$, $S_5=[4, 5)$, $S_6=[5, +\infty)$. Врз основа на примерок со обем 1000 определени се фреквенциите на тие класи: 610, 220, 100, 40, 20, 10 соодветно. Со ниво на значајност 5% да се провери хипотезата дека примерокот е земен од експоненцијална $E(1)$ распределба.

Решение:

Ја тестираме хипотезата

H_0: $X \sim E(1)$

H_a: X нема $E(1)$

Ако H_0 е точна, тогаш

$$F_0(x) = \begin{cases} 1-e^{-x}, & x > 0 \\ 0 & \text{инаку} \end{cases}$$

$$p_k = P\{X \in S_k | H_0\}, \quad k = 1, 2, \ldots, 6$$

$$p_1 = P\{X < 1 | H_0\} = F_0(1) = 1 - e^{-1} = 0.632$$

$$p_2 = P\{1 \le X < 2 | H_0\} = F_0(2) - F_0(1) = (1-e^{-2}) - (1-e^{-1}) = 0.233$$

$$p_3 = P\{2 \le X < 3 | H_0\} = F_0(3) - F_0(2) = (1-e^{-3}) - (1-e^{-2}) = 0.086$$

$$p_4 = P\{3 \le X < 4 | H_0\} = F_0(4) - F_0(3) = (1-e^{-4}) - (1-e^{-3}) = 0.031$$

$$p_5 = P\{4 \le X < 5 | H_0\} = F_0(5) - F_0(4) = (1-e^{-5}) - (1-e^{-4}) = 0.011$$

$$p_6 = P\{X \ge 5 | H_0\} = 1 - P\{X < 5 | H_0\} = 1 - F_0(5) = 1 - (1-e^{-5}) = 0.007$$

Пример 4 – решение

S_k	N_k	p_k	np_k	$N_k - np_k$	$(N_k - np_k)^2$	$(N_k - np_k)^2/np_k$
1	610	0.632	632	–22	484	0.77
2	220	0.233	233	–13	169	0.73
3	100	0.086	86	14	196	2.29
4	40	0.031	31	9	81	2.61
5	20	0.011	11	9	81	7.36
6	10	0.007	7	3	9	1.29
$\sum$	1000	1	1000			$\chi^2 = 15.05$

Бројот на степени на слобода е $6 - 1 = 5$ (нема непознат параметар)

$\chi^2_{0.05,\,5} = 11.07$, па критичниот домен е $C = (11.07, +\infty)$.

Вредноста на тест статистиката $\chi^2 = 15.05 \in C$, па заклучуваме дека хипотезата се отфрла т.е. примерокот не е земен од експоненцијална $E(1)$ распределба.

Пример 5

При попис на населението се утврдува кое е највисокото ниво на образование на испитаникот (без постдипломски студии и докторски студии). Категориите се следни: 1 - завршен факултет, 2 - виша школа, 3 - средно училиште, 4 - делумно завршено средно училиште, 5 - осмо одделение, 6 - четврто одделение или помалку. Според некој постар попис од пред 20 години распределбата по категории кај 25 годишното население е следна: 17%, 18%, 32%, 17%, 13%, 3%. Една актуелна студија испитала на случаен начин 200 млади на таа возраст и ги добила следните бројки по категории 40, 35, 83, 26, 16, 0. Со ниво значајност 0.05 да се тестира дали распределбата на населението според образование е иста како пред 20 години.

Пример 5 – решение

Нека X е категоријата за степенот на образование на случајно избрано 25 годишно лице. Треба да ја тестираме хипотезата

H_0: распределбата на X е $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0.17 & 0.18 & 0.32 & 0.17 & 0.13 & 0.03 \end{pmatrix}$

S_k	N_k	p_k	np_k	$(N_k - np_k)^2 / np_k$
1	40	0.17	34	1.059
2	35	0.18	36	0.028
3	83	0.32	64	5.641
4	26	0.17	34	1.882
5	16	0.13	26	3.846
6	0	0.03	6	6
вкупно	200	1	200	18.456

Бројот на степени на слобода е $6 - 1 = 5$ (нема непознат параметар)

$\chi^2_{0.05,\,5} = 11.07$, па критичниот домен е $C = (11.07, +\infty)$.

Вредноста на тест статистиката $\chi^2 = 18.456 \in C$, па заклучуваме дека H_0 се отфрла, т.е., дали распределбата на населението според образованието не е иста како пред 20 години.

Испитување на независност на две категориски обележја

- Пирсоновиот Хи-квадрат тест се користи и при анализа на врски меѓу категориски обележја.
- Честопати, n елементи на примерокот може да се класифицираат според два различни критериуми. Тогаш може да нè интересира дали двата методи на класификација се независни. На пример, може да ја земеме популацијата на дипломирани инженери, и да утврдиме дали почетната плата е независна од академските дисциплини.
- Значи, на исти индивидуи (елементи) се одредуваат две категориски обележја, X и Y. Нека категориите на X се $A_1, A_2, \ldots, A_r$, а категориите на Y се $B_1, B_2, \ldots, B_k$. Сакаме да ја тестираме хипотезата

H_0: Обележјата X и Y се независни

H_a: Обележјата X и Y не се независни

Табели на контингенција

- Ако обемот на примерокот (набљудуваните индивидуи) е n се формира табела на контингенција со r редици кои одговараат на категориите на X и k колони кои одговараат на категориите на Y. Во секое (i, j) поле од табелата се запишува набљудуваната фреквенција n_{ij} на бројот на индивидуите (елементи) кои припаѓаат во категоријата A_i на X и категоријата B_j на Y.

$X \backslash Y$	B_1	B_2	. . .	B_j	. . .	B_k	Вкупно
A_1	n_{11}	n_{12}		n_{1j}		n_{1k}	$n_{1\cdot}$
A_2	n_{21}	n_{22}		n_{2j}		n_{2k}	$n_{2\cdot}$
. .							
A_i	n_{i1}	n_{i2}		n_{ij}		n_{ik}	$n_{i\cdot}$
. .							
A_r	n_{r1}	n_{r2}		n_{rk}		n_{rk}	$n_{r\cdot}$
Вкупно	$n_{\cdot 1}$	$n_{\cdot 2}$		$n_{\cdot j}$		$n_{\cdot k}$	n

$$n_{i\cdot} = \sum_{s=1}^{k} n_{is},$$

$$n_{\cdot j} = \sum_{t=1}^{r} n_{tj},$$

$$n = \sum_{i=1}^{r} \sum_{j=1}^{k} n_{ij}$$

Хи-квадрат тест за независност

- Нека p_{ij} е теориската веројатност случајно избран елемент да биде во категоријата A_i на X и категоријата B_j на Y, т.е., $p_{ij} = P(A_i \cap B_j)$.

- Ако H_0 е точна тогаш $p_{ij} = P(A_i \cap B_j) = p_{i.}\, p_{.j}$, каде што $p_{i.} = P(A_i)$ и $p_{.j} = P(B_j)$.

- Оценувачи за $p_{i.}$ и $p_{.j}$ се

$$\hat{p}_{i.} = \frac{n_{i.}}{n} \quad \text{и} \quad \hat{p}_{.j} = \frac{n_{.j}}{n}$$

- Според тоа, ако H_0 е точна очекуваната честота во полето (i, j) ќе биде:

$$E_{ij} = np_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

Хи-квадрат тест за независност

- За големи вредности на n, кога H_0 е точна статистиката

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{k}\frac{(n_{ij}-E_{ij})^2}{E_{ij}} = \sum_{i=1}^{r}\sum_{j=1}^{k}\frac{\left(n_{ij}-\frac{n_{i}.n._{j}}{n}\right)^2}{\frac{n_{i}.n._{j}}{n}} = n\sum_{i=1}^{r}\sum_{j=1}^{k}\frac{n_{ij}^2}{n_{i}.n._{j}} - n$$

има χ^2 - распределба со $(r-1)(k-1)$ степени на слобода.

- Нултата хипотеза се отфрла доколку пресметаната вредност на тест статистика е поголема од критичната вредност $\chi^2_{\alpha,(r-1)(k-1)}$ која се чита од таблица, т.е. $C = (\chi^2_{\alpha,(r-1)(k-1)}, +\infty)$.

- **Забелешка**: Пирсоновиот тест се користи кога очекуваните честоти во секоја ќелија од табелата на контингенција е барем 5. Ако ова не е исполнето, се препорачуваат други тестови.

Пример 1.

Со ниво на значајност $\alpha = 0.05$, да се тестира хипотезата дека моделот на автомобилот од марката Пежо независи од полот на сопственикот. Резултатите добиени со испитување на 350 случајно избрани сопственици на автомобили се дадени во следната табела на контингенција:

X \ *Y*	*Пежо 106*	*Пежо 206*	*Пежо 306*	*Вкупно*
маж	40	80	30	150
жена	30	120	50	200
Вкупно	70	200	80	350

Пример 1. – решение

Нека обележјето X е полот на сопственикот, а обележјето Y – моделот на Пежо автомобилот. Врз основа на дадениот примерок со обем $n = 350$, ќе ја тестираме хипотезата:

H_0: X и Y се независни обележја.

Притоа, за обележјето X има 2 категории, а обележјето Y 3 категории, што значи дека $r = 2$, а $k = 3$.

X \ Y	*Пежо 106*	*Пежо 206*	*Пежо 306*	*Вкупно*
маж	40	80	30	150
жена	30	120	50	200
Вкупно	70	200	80	350

Пример 1. – решение

Ја пресметуваме вредноста на тест статистиката за податоците од примерокот дадени во табелата на контингенција.

$$\chi^2 = \frac{\left(40 - \frac{70\cdot 150}{350}\right)^2}{\frac{70\cdot 150}{350}} + \frac{\left(80 - \frac{200\cdot 150}{350}\right)^2}{\frac{200\cdot 150}{350}} + \frac{\left(30 - \frac{80\cdot 150}{350}\right)^2}{\frac{80\cdot 150}{350}} +$$

$$+ \frac{\left(30 - \frac{70\cdot 200}{350}\right)^2}{\frac{70\cdot 200}{350}} + \frac{\left(120 - \frac{200\cdot 200}{350}\right)^2}{\frac{200\cdot 200}{350}} + \frac{\left(50 - \frac{80\cdot 200}{350}\right)^2}{\frac{80\cdot 200}{350}}$$

$$= 7.4373$$

X \ *Y*	*Пежо 106*	*Пежо 206*	*Пежо 306*	*Вкупно*
маж	40	80	30	150
жена	30	120	50	200
Вкупно	70	200	80	350

Пример 1. – решение

За определување на критичниот домен за $\alpha = 0.05$, ја читаме од таблица за χ^2 - распределба, вредноста

$$\chi^2_{\alpha,(r-1)(s-1)} = \chi^2_{0.05,(2-1)(3-1)} = \chi^2_{0.05,2} = 5.99$$

Критичниот домен е $C = (5.99, +\infty)$.

Добиената вредноста на тест статистиката е $\chi^2 = 7.4373 \in C$, од каде се заклучува дека нултата хипотеза се отфрла.

Значи, полот на сопственикот зависи од моделот на Пежо автомобилот.

Пример 2.

Врз база на резултатите од анкетата спроведена на 385 вработени, кои се дадени во табелата на контингенција да се провери дали превозното средство кое го користат вработените за да дојдат на работа зависи од нивните месечни примања, со ниво на значајност $\alpha = 0.01$.

Y / *X*	*Велосипед*	*Автобус*	*Автомобил*	*Вкупно*
помалку од 30 000 ден.	85	16	6	107
од 30 000 ден. до 49 999 ден.	102	27	13	142
од 50 000 ден. до 100 000 ден.	36	22	15	73
повеќе од 100 000 ден	15	23	25	63
Вкупно	238	88	59	385

Пример 2. – решение

Врз основа на дадениот примерок со обем $n = 385$, ќе ја тестираме хипотезата:

H_0: X и Y се независни обележја т.е. превозното средство не зависи од примањата

наспроти

H_a: X и Y се зависни обележја т.е. превозното средство зависи од примањата

Притоа, за обележјето X има 4 категории, а за обележјето Y има 3 категории, што значи дека $r = 4$, а $k = 3$.

Во овој пример за пресметување на тест статистиката ќе ја користиме втората формула:

$$\chi^2 = n\sum_{i=1}^{r}\sum_{j=1}^{k}\frac{n_{ij}^2}{n_{i\cdot}\cdot n_{\cdot j}} - n$$

Пример 2. – решение

Y / X	Велосипед	Автобус	Автомобил	Вкупно
помалку од 30 000 ден.	85	16	6	107
од 30 000 ден. до 49 999 ден.	102	27	13	142
од 50 000 ден. до 100 000 ден.	36	22	15	73
повеќе од 100 000 ден	15	23	25	63
Вкупно	238	88	59	385

$$\chi^2 = n\sum_{i=1}^{r}\sum_{j=1}^{k}\frac{n_{ij}^2}{n_{i}.n._{j}} - n = 385\cdot\left(\frac{85^2}{107\cdot 238}+\frac{16^2}{107\cdot 88}+\frac{6^2}{107\cdot 59}+\frac{102^2}{142\cdot 238}+\frac{27^2}{142\cdot 88}+\frac{13^2}{142\cdot 59}+\right.$$

$$\left.+\frac{36^2}{73\cdot 238}+\frac{22^2}{73\cdot 88}+\frac{15^2}{73\cdot 59}+\frac{15^2}{63\cdot 238}+\frac{23^2}{63\cdot 88}+\frac{25^2}{63\cdot 59}\right)-385 = 70.78$$

Пример 2. – решение

За определување на критичниот домен за $\alpha = 0.01$, од таблица за χ^2-распределба ја читаме вредноста

$$\chi^2_{\alpha,(r-1)(s-1)} = \chi^2_{0.01,(4-1)(3-1)} = \chi^2_{0.01,6} = 16.81$$

Критичниот домен е $C = (16.81, +\infty)$.

Добиената вредноста на тест статистиката $\chi^2 = 70.78 \in C$, од каде заклучуваме дека нултата хипотеза се отфрла.

Значи, превозното средство кое го користат вработените за да дојдат на работа зависи од нивните месечни примања.