Бизнис статистика

Тестови за совпаѓање на распределби и независност на обележја

Задача 1

- Една анкета на потрошувачи од цела држава, спроведена во некое списание го поставила прашањето: „Како би го оцениле нивото на квалитетот на услуги што ги нудат фирмите во ИТ секторот?“ Распределбата на одговорите на ова прашање била следна: 1 - одлична 8%, 2 - прилично добра 47%, 3 - добра 34%, 4 - неквалитетна 11%. Менаџерката на една ИТ продавница сака да открие дали резултатите од оваа потрошувачка анкета соодветствуваат и на клиентите на ИТ продавниците во нејзиниот град. За таа цел, таа интервјуира 207 случајно избрани потрошувачи додека ги напуштаат ИТ продавниците во разни делови на градот.

Задача 1 - продолжение

- Добиените одговори од оваа студија се дадени во следната табела.

Одговор	Честота N_k
1 - одлична	21
2 - прилично добра	109
3 - добра	62
4 - неквалитетна	15

- Дали со ниво на значајност $\alpha = 0.05$, менаџерката може да заклучи дека распределбата на одговорите во нејзиниот град се совпаѓа со распределбата на одговорите добиени со националното истражување.

Задача 1: Решение

- Нека X е категорија за оценката на нивото на квалитетот на услуги што ги нудат фирмите во ИТ секторот во градот на менаџерката
- Ги поставуваме следните хипотези

H_0: распределбата на X е $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0.08 & 0.47 & 0.34 & 0.11 \end{pmatrix}$

H_a: распределбата на X не е како претходно дадената.

S_k	N_k	p_k	np_k	$(N_k - np_k)^2 / np_k$
1	21	0.08	16,56	1,19
2	109	0.47	97,29	1,41
3	62	0.34	70,38	1,00
4	15	0.11	22,77	2,65
вкупно	207	1	207	6.25

- Значи, вредноста на тест статистиката е $\chi^2 = 6.25$.

Задача 1: Решение

- Бројот на степени на слобода е 4 – 1 = 3 (нема непознат параметар). Оттука, $\chi^2_{0.05,\,3} = 7.81$, па критичниот домен е $C = (7.81, +\infty)$.
- Вредноста на тест статистиката $\chi^2 = 6.25 \notin C$, па заклучуваме дека H_0 не се отфрла.
- Податоците собрани во примерокот од 207 купувачи во ИТ супермаркети укажуваат на тоа дека распределбата на одговори на купувачите на супермаркети во градот на менаџерот не е значително различна од распределбата на одговорите на националното истражување.
- Менаџерката на продавницата може да заклучи дека ставовите на нејзините клиенти не се разликуваат битно од ставовите кои се изнесени во истражувањето.

Задача 2.

За да се провери еден генератор на случајни броеви од множеството {1, 2, …, 12}генерирани се 18447 броеви. Фреквенциите на добиените броеви се дадени во табелата. Со ниво на значајност $\alpha = 0.01$ да се тестира хипотезата дека генераторот е случаен т.е. дека распределбата на генерираните броеви е рамномерна.

број	Фрек.
1	1610
2	1585
3	1649
4	1590
5	1540
6	1397
7	1410
8	1350
9	1495
10	1564
11	1602
12	1655

Задача 2: Решение

Нека X е обележјето – генерираниот број
H_0: X има рамномерна распредба на множеството {1, 2, …, 12}
H_a: X нема рамномерна распредба на множеството {1, 2, …, 12}

- За да имаме рамномерна распределба на множеството {1, 2, …, 12}, потребно е веројатноста за генерирање на секој од броевите во ова множество да е иста, т.е. $p_k = 1/12$, $k = 1,2,\ldots,12$.

број	N_k	p_k	np_k	$(N_k - np_k)^2 / np_k$
1	1610	1/12	1537.25	3.443
2	1585	1/12	1537.25	1.483
3	1649	1/12	1537.25	8.124
4	1590	1/12	1537.25	1.810
5	1540	1/12	1537.25	0.005
6	1397	1/12	1537.25	12.796
7	1410	1/12	1537.25	10.533
8	1350	1/12	1537.25	22.809
9	1495	1/12	1537.25	1.161
10	1564	1/12	1537.25	0.465
11	1602	1/12	1537.25	2.727
12	1655	1/12	1537.25	9.019
вкуп.	18447	1	18447	74.376

- Вредноста на тест статистиката е $\chi^2 = 74.376$.

Задача 2: Решение

- Бројот на степени на слобода е 12 – 1 = 11 (нема непознат параметар). Оттука, $\chi^2_{0.01,\,11} = 24.72$, па критичниот домен е $C = (24.72, +\infty)$.
- Вредноста на тест статистиката $\chi^2 = 74.376 \in C$, па заклучуваме дека H_0 се отфрла.
- Може да се заклучи дека генерираните броеви немаат рамномерна распределба, т.е., генераторот не е случаен.

Задача 3

Кога зборувавме за Пуасонова распределба рековме дека со неа може да се моделира број на појавувања на некој настан во фиксни временски интервали. За да се провери дали е ова точно, случајно се избрани 84 пет минутни интервали во текот на работните денови од една недела и следен е бројот на пристигнати клиенти во тие интервали во една локална банка. Врз база на добиените податоци, со ниво на значајност $\alpha = 0.05$ да се тестира хипотезата дека бројот на пристигнати клиенти има Пуасонова распределба.

Број на пристигнувања	N_k
0	7
1	18
2	25
3	17
4	12
5	5

Задача 3: Решение

- Бидејќи параметарот на Пуасоновата распределба не е познат, истиот треба да се оцени.
- 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 $\bar{X}$. Пресметката на $\bar{X}$ е дадена во следната табела.

x_k	N_k	$x_k N_k$
0	7	0
1	18	18
2	25	50
3	17	51
4	12	48
5	5	25
Вкупно	84	192

$$\lambda = \bar{x} = \frac{1}{n}\sum_{k=1}^{r} x_k N_k$$

$$= \frac{192}{84} = 2.3$$

Задача 3: Решение

Сега, ги поставуваме следните хипотези:

H_0: X има $P(2.3)$ распределба

H_a: X нема $P(2.3)$ распределба

- Бидејќи множеството вредности $R_X = \{0,1,2,\ldots\}$, класите S_k ќе ги избереме на следниот начин:

$$S_0 = \{0\}, S_1 = \{1\}, S_2 = \{2\}, S_3 = \{3\}, S_4 = \{4\}, S_5 = \{5,6,\ldots\}.$$

- За веројатностите на секоја од класите, се добива:

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

за $k = 1,2,3,4$, а веројатноста на последната класа е

$$1-(p_1 + p_2 + p_3 + p_4)$$

Задача 3: Решение

S_k	N_k	p_k	np_k	$(N_k - np_k)^2 / np_k$
0	7	0.1003	8.42	0.2411
1	18	0.2306	19.37	0.0970
2	25	0.2652	22.28	0.3329
3	17	0.2033	17.08	0.0003
4	12	0.1169	9.82	0.4841
≥5	5	0.0838	7.04	0.5907
	84		84	1.7462

Значи, вредноста на тест статистиката е $\chi^2 = 1.7462$

Задача 3: Решение

- Бројот на степени на слобода е 6 – 1 – 1 = 4 (има еден непознат параметар). Оттука, $\chi^2_{0.05,\,4} = 9.49$, па критичниот домен е $C = (9.49, +\infty)$.
- Вредноста на тест статистиката $\chi^2 = 1.7462 \notin C$, па заклучуваме дека H_0 не се отфрла.
- Може да се заклучи дека бројот на клиенти кои влегуваат во банката нема Пуасонова распределба.

Задача 4

- Со ниво на значајност $\alpha = 0.01$, да се провери дали типот на пијалак, кој е нарачан со оброк во ресторан во Охрид, не зависи од возраста на потрошувачот. Да се користат податоците од резултатите од анкетата спроведена на 309 случајно избрани гости во Охрид.

X \ Y	Кафе	Сок	Друго	Вкупно
21-34	26	95	18	139
35-55	41	40	20	101
>55	24	13	32	69
Вкупно	91	148	70	309

Задача 4: Решение

- Нека обележјето X е возраста, а обележјето Y – видот на пијалокот. Врз основа на дадениот примерок со обем n = 309, ќе ја тестираме хипотезата:

$$H_0: X \text{ и } Y \text{ се независни обележја.}$$

наспроти $$H_a: X \text{ и } Y \text{ не се независни статистички обележја}$$

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

Задача 4: Решение

- За вредноста на тест статистиката, се добива:

$$\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{k}\frac{(n_{ij}-\frac{n_i.n._j}{n})^2}{\frac{n_i.n._j}{n}}=\frac{(26-\frac{139\cdot 91}{309})^2}{\frac{139\cdot 91}{309}}+\frac{(95-\frac{139\cdot 148}{309})^2}{\frac{139\cdot 148}{309}}+\frac{(18-\frac{139\cdot 70}{309})^2}{\frac{139\cdot 70}{309}}$$

$$+\frac{(41-\frac{101\cdot 91}{309})^2}{\frac{101\cdot 91}{309}}+\frac{(40-\frac{101\cdot 148}{309})^2}{\frac{101\cdot 148}{309}}+\frac{(20-\frac{101\cdot 70}{309})^2}{\frac{101\cdot 70}{309}}+\frac{(24-\frac{69\cdot 91}{309})^2}{\frac{69\cdot 91}{309}}+\frac{(13-\frac{69\cdot 148}{309})^2}{\frac{69\cdot 148}{309}}$$

$$+\frac{(32-\frac{69\cdot 70}{309})^2}{\frac{69\cdot 70}{309}}=59.405$$

Задача 4: Решение

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

$$\chi^2_{\alpha,(r-1)(s-1)} = \chi^2_{0.01,(3-1)(3-1)} = \chi^2_{0.01,4} = 13.28.$$

- Оттука, критичниот домен е $C = (13.28, +\infty)$.
- Добиената вредноста на тест статистиката е $\chi^2 = 59.405 \in C$, од каде се следува дека нултата хипотеза се отфрла.
- Може да се заклучи дека двете категориски обележја: пијалоците и возраста - не се независни. Типот на пијалоци што клиентот го нарачува со ручек е поврзан, т.е. зависен од неговата возраст.

Задача 5

Кандидатите за претседател на една држава се една жена и еден маж. Во случаен примерок од 1000 гласачи се добиени резултатите сместени во следната табела на контингенција. Со ниво на значајност 0.01 да се провери дали определбата на гласачите зависи од полот на гласачот.

Кандидат / Гласачи	Жена	Маж	вкупно
жени	220	270	**490**
мажи	260	250	**510**
вкупно	**480**	**520**	**1000**

Задача 5: решение

Нека X е обележјето пол на гласачот, а Y е определбата на гласачите (полот на кандидатот за претседател за кој ќе гласа).
Ги тестираме хипотезите

H_0: X и Y се независни статистички обележја
H_a: X и Y не се независни статистички обележја

Се користи тест статистиката

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

Задача 5: решение

- Со користење на алтернативната формула за χ^2, се добива:

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

$$= 1000\left(\frac{220^2}{490\cdot 480}+\frac{270^2}{490\cdot 520}+\frac{260^2}{480\cdot 510}+\frac{250^2}{520\cdot 510}\right) - 1000$$
$$= 3.704$$

- $\chi^2_{0.01,(2-1)(2-1)} = \chi^2_{0.01,\,1} = 6.63$, па критичниот домен е

$$C = (6.63, +\infty).$$

- $\chi^2 = 3.704 \notin C$, па заклучуваме дека нулата хипотезата не се отфрла т.е. определбата на гласачите не зависи од полот на гласачот.