

Р. Изследва се връзката между пола на потребителя и предпочитан тип кафе. Кафето е два типа: “чисто” и с различни добавки (мляко, сметана и ар.). Направена е случайна извадка от 220 души, които са отговорили на въпроса какъв тип кафе предпочитат. Получените данни са следните:

Таблица 1:

пол	Чисто кафе	Кафе с добавка	сума
мъже	40	70	110
жени	90	20	110
сума	130	90	220

**1. Да се установи има ли връзка между пола и предпочитания тип кафе. Да се оцени силата на тази връзка.**

При това изследване ще приложат хи-квадрат метод. Той показва дали съществува връзка между пола на потребителя и предпочитания тип кафе, но не може да отговори дали тази връзка е причинно следствена или има друг характер.

Етапите през които преминава прилагането на метода са следните:

1/. Дефиниране на хипотези

$H_0$  : Не съществува връзка между пола и предпочитания тип кафе;

$H_1$ : Съществува статистически значима връзка между пола и предпочитания тип кафе.

2/. Определяне равнище на значимост  $\alpha=0,05$

3/.Избор на метод за проверка на хипотезата:

$$\chi^2 = \sum_1^k \left[ \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}} \right]$$

където:

$\hat{f}_{ij}$  са очакваните /теоретичните/ честоти;

$f_{ij}$  са фактическите (емпиричните) честоти.

4/.Определяне на характера на критичната област: Двустранна критична област

5/.Осигуряване на необходимите данни за проверка- когато се прави проверка на хипотези, данните трябва да са от представителни извадки.

6/. Оценка на емпиричната характеристика –  $\chi_{em}$  може да се направи на базата на събраната информация.

7/.Определяне на теоретичната характеристика

Това става с помощта на съответни таблици за стохастичното разпределение на характеристиката при риск за грешка  $\alpha=0,05$ .

8/. Вземане на решение за приемане и отхвърляне на хипотезите.

В таблица 2 са нанесени квадратите, получени от данните от таблица 1.

Таблица 2:

пол	Чисто кафе	Кафе с добавка	сума
мъже	1600	4900	6500
жени	8100	400	8500
сума	9700	5300	15000

Необходимо е да се намери зависимостта между двата признака пол и предпочитан тип кафе. В конкретния случай може да се приложи класическият подход за измерване на корелационни зависимости, но е възможно да се използват и коефициентите на Пирсън и Чупров.

В таблица 3 са нанесени стойностите на  $\phi^2$ .

Таблица 3:

пол	Чисто кафе	Кафе с добавка	сума
мъже	1600/ (130.110)=0,11	4900/(90.110)=0,49	0,6
жени	8100/(130.110)=0,57	400/(90.110)=0,04	0,61
сума	0,68	0,53	1,21

$$\phi^2 = 1,21 - 1 = 0,21$$

$$C = \sqrt{\frac{\phi^2}{1 + \phi^2}} = \sqrt{\frac{0,21}{1 + 0,21}} = \sqrt{\frac{0,21}{1,21}} = \sqrt{0,1736} = 0,42$$

Коефициентът на Чупров:  $K = \frac{\phi^2}{\sqrt{(k_1 - 1)(k_2 - 1)}}$ , където  $k_1$  и  $k_2$  са съответно броя на групите на първия и на втория признак.

$$K^2 = \frac{\phi^2}{\sqrt{(k_1 - 1)(k_2 - 1)}} = \frac{0,21}{\sqrt{(2 - 1)(2 - 1)}} = 0,21$$

$$K = 0,46$$

$$\phi^2 = \frac{\chi^2}{N} \Rightarrow \phi^2 \cdot N = 0,21 \cdot 220 = 46,2$$

$$\chi^2_{\text{емп.}} = 46,2$$

Зависимостта между двата признака е умерена.

Табличната стойност се взема от таблица. Тя е при степени на свобода.

$$\text{За този случай} - f = (2 - 1) \cdot (2 - 1) = 1$$

$$\chi^2_{\text{теор.}} = 3,84$$

Сравняваме  $\chi_e^2 \leq \chi_r^2$ , т.е.  $46,2 > 3,84$ . Приемаме  $H_1$  – алтернативната хипотеза, а отхвърляме нулевата хипотеза, т.е. съществува статистически значима връзка между пола и предпочитания тип кафе.

Често пъти при изучавания се посочват два алтернативни варианта /както е в нашия случай с предпочитания тип кафе /. В тези случаи се