6. Имеется таблица сопряженности о зависимости количества товаров в корзине покупателя по данным маркетингового обследования. Требуется проверить наличие зависимости количества товаров от пола покупателя с помощью таблиц сопряженности.

|  |  |  |  |
| --- | --- | --- | --- |
| Пол | Количество товаров в корзине покупателя | | |
| Корзина пуста | В корзине только один товар | В корзине два и более товара |
| Мужской | 88 | 122 | 276 |
| Женский | 164 | 147 | 300 |

Получено выборочное распределение признака «А» в зависимости от B. Результаты приведены в таблице:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A1 | A2 | A3 | ni\* |
| B1 | 88 | 122 | 276 | 486 |
| B2 | 164 | 147 | 300 | 611 |
| n\*j | 252 | 269 | 576 | 1097 |

Для проверки независимости признаков «A» и «B» (т.е. предполагаем, что наполненность корзины не зависит от пола покупателя) проверяем нулевую гипотезу Н0:(pij = pi\*p\*j для всех i, j). Вычислим статистику χ2 набл по формуле:

где nij – наблюдаемые частоты.

Если значение χ2набл попало в критическую область: χ2 > χ2крит(α ; v=2), нулевая гипотеза отвергается с вероятностью ошибки α и признаки считаются зависимыми.

В этом случае имеет смысл измерить полученную связь между X и Y с помощью коэффициентов связи (сопряженности).

Рассчитаем теоретические частоты по формуле:

для всех клеток таблицы

Получим таблицу сопряженности теоретических частот распределения:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A1 | A2 | A3 | ni\* |
| B1 | 111.643 | 119.174 | 255.183 | 486 |
| B2 | 140.357 | 149.826 | 320.817 | 611 |
| n\*j | 252 | 269 | 576 | 1097 |

Вычислим статистику χ2:

= =

По таблице χ2-распределения находим:

χ2крит(0.05;2) = 5.99146

где v = (r-1)(s-1) = (2-1)(3-1) = 2 - число степеней свободы.

Критическая область имеет вид χ2 > χ2крит. Так как вычисленное значение хи-квадрат попадает в критическую область, то гипотеза о независимости отвергается с вероятностью ошибки 0.05.

Т.е. наполненность корзины зависит от пола покупателя.