

ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ г. СЕМЕЙ
Кафедра: Общественное здравоохранение и информатики

СРС

На тему : Анализ таблиц сопряженности с помощью хи квадрат

Выполнила: Жумаканова Т.М
Специальность: Стоматология
Курс: 2 (207)
Проверила; Атабаева А.К

Семей 2017

Главная часть

Описание таблиц сопряженности

Таблица сопряженности - средство представления совместного распределения двух переменных, предназначенное для исследования связи между ними. Таблица сопряженности является наиболее универсальным средством изучения статистических связей, так как в ней могут быть представлены переменные с любым уровнем измерения.

Строки таблицы сопряженности соответствуют значениям одной переменной, столбцы - значениям другой переменной (количественные шкалы предварительно должны быть сгруппированы в интервалы). На пересечении строки и столбца указывается частота совместного появления f_{ij} соответствующих значений двух признаков x_i и y_j . Сумма частот по строке $f_{i\cdot}$ называется маргинальной частотой строки; сумма частот по столбцу $f_{\cdot j}$ - маргинальной частотой столбца. Сумма маргинальных частот равна объему выборки n ; их распределение представляет собой одномерное распределение переменной, образующей строки или столбцы таблицы.

В таблицах сопряженности могут быть представлены как абсолютные, так и относительные частоты (в долях или процентах). Относительные частоты могут рассчитываться по отношению:

к маргинальной частоте по строке

к маргинальной частоте по столбцу

к объему выборки

Таблицы сопряженности используются для проверки гипотезы о наличии связи между двумя признаками (Статистическая связь, Критерий "хи-квадрат"), а также для измерения тесноты связи (Коэффициент фи, Коэффициент контингенции, Коэффициент Крамера)

Введем следующие обозначения:

- наблюдаемая частота (i,j)

- ожидаемая частота при H_0

- наблюдаемая частота (i,j)

- ожидаемая частота при H_0

$$n_{ij} = \sum_{(x,y)} [x=i][y=j]$$

$$n_{.j} = \sum_i n_{ij}$$

$$n_{i.} = \sum_j n_{ij}$$

$$n = \sum_i \sum_j n_{ij}$$

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

Статистика

$$\chi^2 = \sum_{(i,j)} \frac{\left(n_{ij} - \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}\right)^2}{\frac{n_{i\cdot} \cdot n_{\cdot j}}{n}} = n \left(\sum_{i=1}^K \sum_{j=1}^L \frac{n_{ij}^2}{n_{i\cdot} \cdot n_{\cdot j}} - 1 \right) \sim \chi^2_{KL - (K-1) - (L-1) - 1} = \chi^2_{(K-1)(L-1)}$$

Хи-квадрат (с поправкой Йейтса)

Метод аналогичен предыдущему, но вводится поправка Йейтса, которая делает общую оценку более умеренной. Метод можно применять, если процент доли меньше 10%, но более 5%.

$$Z = \frac{|\hat{p}_1 - \hat{p}_2| - \frac{1}{2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Список литературы

- 1Лапач С. Н. , Чубенко А. В., Бабич П. Н. Статистика в науке и бизнесе. — Киев: Морион, 2002.
- 2Г. Аптон. Анализ таблиц сопря