

Глоссарий к Теме 4

Анализ двумерной зависимости. Таблицы сопряженности. Коэффициенты парной связи.

Каузальная зависимость – причинно-следственная связь.

Таблица сопряженности – показывает совместное распределение переменных (номинальной или порядковой шкал) и помогает увидеть наличие/отсутствие взаимосвязи между этими переменными.

Критерий Хи-квадрат Пирсона – это показатель, фиксирующий степень расхождения реальных и ожидаемых частот. Целью применения критерия является установление степени соответствия между наблюдаемыми и ожидаемыми значениями ячеек.

Число степеней свободы (df) – количество значений в итоговом вычислении статистики, способных варьироваться. Рассчитывается по формуле $N=(r-1)*(c-1)$, где r – количество столбцов, c – количество строк.

Асимптотическая значимость (sig) – вероятность случайности связи. Чем меньше эта величина, тем выше статистическая значимость (достоверность) связи.

Точная значимость – p -уровень значимости, вычисляемый точным методом; принимается во внимание, когда условия применения традиционного критерия нарушены (более 25 % ячеек таблицы сопряженности имеют частоту менее 5).

Линейно-линейная связь – статистический критерий, определяющий степень корреляции между переменными.

Номинальная по номинальной – меры связи для двух номинальных переменных.

Коэффициент Сопряженности – мера тесноты связи, основанная на критерии хи-квадрат (меняется между 0 и 1).

Лямбда – мера связи, которая отражает относительное снижение ошибки, когда значения независимой переменной используются для предсказания значений зависимой переменной.

Коэффициент неопределенности – мера связи, указывающая относительное снижение ошибки в случае, когда значения одной переменной используются для предсказания значений другой.

Эта – мера связи между переменными строки и столбца, значения которой изменяются от 0 (отсутствие связи) до 1 (сильная связь).

Гамма – симметричная мера связи между двумя порядковыми переменными (-1 и 1).

D Сомерса – мера связи между двумя порядковыми переменными, (-1 и 1).

Тау-b Кендалла – непараметрическая мера корреляции для порядковых или ранговых переменных, учитывает возможные совпадения значений (связи).

Тау-c Кендалла – непараметрическая мера связи для порядковых переменных, игнорирующая возможные совпадения значений (связи).

Коэффициент Крамера – является мерой связи между двумя переменными, принимает значения от 0 до 1 и более приемлем для таблиц с $df > 2$.

Фи — коэффициент – является мерой связи двух переменных, аналог корреляции Пирсона.