

Цаплина Н.С.

АЛГОРИТМ ВЫЧИСЛЕНИЯ МОЩНОСТИ КРИТЕРИЯ ХИ-КВАДРАТ

Научный руководитель: канд. техн. наук, доц. Лямец Л.Л.

Кафедра физики, математики и медицинской информатики

Смоленский государственный медицинский университет, г. Смоленск

Актуальность. Критерий хи-квадрат – непараметрический критерий, который выявляет значимые различия между наблюдаемыми и ожидаемыми частотами в статистических распределениях. Обоснование статистических выводов связано с необходимостью вычисления мощности данного критерия. В литературе для медицинских специальностей алгоритм вычисления мощности критерия встречается редко, поэтому его разработка имеет учебно-методическое значение.

Цель: разработка алгоритма для вычисления мощности критерия для таблиц сопряженности и автоматизация соответствующих вычислений при помощи современных информационных технологий.

Материалы и методы. Анализируются два атрибутивных признака А и В. Признак А имеет градаций, признак В – имеет градаций. На первом этапе алгоритма формулируются статистические гипотезы. Нулевая гипотеза – признаки статистически не сопряжены; альтернативная гипотеза – признаки статистически сопряжены. На основании первичных данных строится таблица сопряженности признаков размером . Рассчитываются наблюдаемые и ожидаемые частоты для каждой ячейки таблицы. Статистика критерия рассчитывается по следующей формуле:

В формуле (1) суммирование ведется по ячейкам таблицы, где . Статистика также является параметром нецентрального распределения .

На втором этапе алгоритма для распределения с числом степеней свободы вычисляется квантиль уровня . Если , то квантиль вычисляется из следующего выражения:

На третьем этапе алгоритма вычисляется мощность критерия по формуле:

В выражении (3) через обозначена функция плотности вероятности нецентрального распределения с числом степеней свободы и параметром нецентральности . Через обозначена вероятность ошибки второго рода.

Для автоматизации вычислений по формулам (1), (2) и (3) можно использовать возможности библиотеки *Statsmodels* языка программирования *Python*. Формула (3) может быть реализована импортированием из *statsmodels* объекта *stats.power* методом *solve_power*:

- количество наблюдений,
- уровень значимости,
- это разница между относительными показателями, рассчитывается автоматически.

Результаты и их обсуждение. Разработан и описан алгоритм вычислений мощности критерия для таблицы сопряженности. Для автоматизации вычислений предлагается использовать возможности библиотеки *Statsmodels* языка программирования *Python*

Выводы. Разработанный алгоритм может использоваться в учебном процессе для объяснения сущности вычисления мощности критерия и для анализа статистических данных в научных исследованиях.