

1. Есть две гипотезы: основная состоит в том, что элементы выборки имеют нормальное распределение, а альтернатива — в том, что элементы выборки имеют распределение Пуассона. Построить критерий, обладающий нулевыми вероятностями ошибок первого и второго рода.
2. Пусть X_1, \dots, X_n — выборка из биномиального распределения с параметрами m и p , где p может принимать лишь значения $1/3$ и $2/3$ с априорными вероятностями $1/5$ и $4/5$ соответственно, а параметр m известен и фиксирован. Построить байесовский критерий.
3. Построить критерий для проверки равенства дисперсий двух независимых нормальных выборок с известными средними, статистика которого имеет при верной основной гипотезе распределение Фишера с n и m степенями свободы.
4. Показать теоретически и проверить экспериментально, что при числе групп $k = 2$ критерий для решения задачи однофакторного дисперсионного анализа совпадает с критерием Стьюдента.
5. Среди учителей младших классов есть мнение, что у маленьких детей способности к математике "прямо пропорциональны" увлеченности настольными играми. Ясно, что в такой формулировке это звучит не очень математично: обе величины сложно "оцифровать". Предложите способ численно оценивать силу такой взаимосвязи, сформируйте соответствующую статистическую гипотезу и предложите метод для ее проверки.

Решение предполагает небольшое рассуждение, "изобретение" или использование какой-нибудь статистики и ее распределение (можно теоретическое, но проще эмпирическое).
6. Недобросовестные исследователи проверяют величины X, Y на независимость с помощью критерия хи-квадрат, разделяя множества значений X и Y на две части каждое, однако подбирая границы для разделения из данных. То есть они разбивают множество значений на 4 части, и при этом максимизируют значение χ^2 .

Оцените с помощью симуляций, для какого процента выборок объема 1000 удастся ложно отвергнуть нулевую гипотезу о независимости с помощью таких манипуляций.
7. Постройте регрессионные модели для данных в файлах

`data_regr_penalty.csv`

- (a) Проверьте, распределены ли данные нормально и осмысленно ли строить для них линейную модель. Если нет, попробуйте модифицировать данные.

- (b) Проверьте, не скоррелированы ли предикторы; если какие-то предикторы не нужны, не включайте их в модель.
- (c) Постройте модель линейной регрессии. В ответе укажите коэффициенты при предикторах и коэффициент детерминации R^2 .
- (d) Все ли коэффициенты значимо отличаются от 0? Как распределены коэффициенты, если зависимости на самом деле нет (укажите название этого распределения). !Если какие-то коэффициенты значимо не отличаются от 0, удалите их из модели.
- (e) Проведите анализ остатков полученной модели и сделайте выводы.