

Задача 2.1

Кубраков Никита

March 2020

1 Задача 2.1

Пусть задано две вектора ответов: y — истинный вектор ответов для некоторой выборки, а также есть вектор ответов \hat{y} - некоторой предсказательной модели. Наблюдатель хочет проверить гипотезу о том, что ровно в 25 процентах случаев модель дает заниженные оценки. Предложите метод проверки данной гипотезы: запишите задачу формально, предложите статистику для решения данной задачи на уровне значимости $\alpha = 0.05$. Также найдите зависимость мощности данного критерия в зависимости от истинного процента заниженных ответов.

2 Решение

1. Перейдем от векторов ответов, к вектору, в котором на i -ом месте стоит 1, если модель показала результат меньше истинного значения и 0 иначе.
2. Теперь для этого вектора будем проверять гипотезу о том, что его элементы приходят из распределения Бернули с параметром $p = 0.25$. Для этого можно воспользоваться биномиальным тестом, все что нам нужно - найти для биномиального распределения с параметрами n (длина вектора) и $p = 0.25$ α -квантиль. Если он окажется больше количества единиц в векторе (это и есть наша статистика), то данные не противоречат гипотезе, иначе мы отвергаем ее.
3. Построю графики мощности при 3 различных длинах векторов. Для этого найду для каждого n те количества единиц в векторе, когда отклоняется гипотеза о том, что $p = 0.25$ то есть те, где p-value меньше 0.05. Например, для $n = 16$ это 0 и 8 - 16. А теперь при различных p просуммирую вероятности для $\text{binom}(n, p)$ принимать эти значения. Это и будут графики мощности, они приведены ниже.

