Фамилия и имя: Вареник Наталия

Прикладной статистический анализ данных

Лабораторная работа 1

Задача 2.1 Имеется истинный вектор ответов $\boldsymbol{y}$ и предсказанный некоторой моделью $\hat{\boldsymbol{y}}$. Проверяется гипотеза, что модель ровно в 25% случаев дает заниженный ответ. Для этого рассмотрим случайную величину $\boldsymbol{x} = \hat{\boldsymbol{y}} - \boldsymbol{y}$, тогда проверяемая гипотеза будет состоять в том, что 25% квантиль распределения $F(\boldsymbol{x})$ равна $m_0 = 0$.

Рассмотрим статистику $T(\boldsymbol{x}) = nF_n(\boldsymbol{x}, m_0) = \sum\limits_{i=1}^{n} \mathbb{I}(x_i \leq m_0)$. Где $F_n(\boldsymbol{x}, m_0)$ – эмпирическая функция распределения $\boldsymbol{x}$ в точке m_0, n – размерность $\boldsymbol{x}$. Найдем распределение этой статистики при выполнении нулевой гипотезы. Запишем условие 25-процентной квантили: $P\{x \leq m_0\} = 0.25$. Тогда заметим, что $\sum\limits_{i=1}^{n} \mathbb{I}(x_i \leq m_0) \sim Bi(n, 0.25)$. Критическую область будем искать следующим образом. Если удалось найти такое $K \in \mathbb{N}$, что $2\sum\limits_{k=0}^{K} P\{T(\boldsymbol{x}) = k\} = \alpha$, тогда критическая область будет состоять из хвоста биномиального распределения, соответствующего значению K и ему симметричного относительно среднего значения. При попадении статистики в эти хвости будем отвергать гипотезу. В ином случае, если такого K найти не удалось, то будем искать максимальное $\tilde{K} \in \mathbb{N}$ такое, что $2\sum\limits_{k=0}^{\tilde{K}} P\{T(\boldsymbol{x}) = k\} < \alpha$ и в качестве K будем брать $K_1 = \tilde{K} + 1$. Но тогда вероятность ошибки первого рода будет $> \alpha$. Поэтому введем нормировочный множитель $p_\alpha : \alpha = p_\alpha \cdot \tilde{\alpha}$ и тогда при попадении статистики в хвосты соответствующие значению $K = K_1$ будем отвергать гипотезу с вероятностью p_α.

Распределение статистики в зависимости от истинного процента заниженных ответов p есть $\xi_p = T(\boldsymbol{x}|p) \sim Bi(n, p)$. Тогда мощность данного критерия имеет вид:

$$W = P\{\xi_p \in \Omega_K\} + p_\alpha P\{\xi_p \in \Omega_{K_1}\}.$$