

Задание 4.3

Сотников Антон, 674 группа

30 марта 2020 г.

Задача

Правительство города М. испытывает систему обнаружения нежелательных лиц по камерам в метро. В качестве демонстрации работоспособности системы была поставлена цель: найти и задержать опасного рецидивиста по имени Николай Вальный, а также его соучастников. Была собрана выборка из 5000 снимков лица, для которых была проверена гипотеза о несовпадении этого снимка с лицами участников команды Н. Вального. Для 100 фотографий нулевая гипотеза была отвергнута на уровне значимости $\alpha = 0.05$.

Требуется:

1. Определить, в чем недостаток данного подхода и как можно его улучшить.
2. Предложить наилучший, на ваш взгляд, способ для повышения качества данного решения.
3. Какую меру качества контролирует данный способ? Какие гарантии он предоставляет?
4. В чём недостатки данного способа?
5. Как изменилась мощность при использовании предложенного вами способа относительно изначальной процедуры проверки?

6. Известно, что все 5000 фотографий были сделаны для разных людей, и правительство хочет, чтобы система ни в коем случае не упустила преступников. Ответьте на те же вопросы из пунктов 2-4.
7. Как изменилась мощность при использовании предложенного вами способа относительно предыдущего способа?

Решение

1. Недостаток данного подхода заключается в отсутствии поправки на множественную проверку гипотез.
2. На мой взгляд, наилучшим способом является метод Холма.
3. Метод обеспечивает контроль над FWER на уровне $\alpha \forall p_i, T_i$. Также он гарантирует, что FWER не превзойдет α .
4. Главным недостатком является то, что не учитывается характер зависимости статистик.
5. Мощность уменьшилась, т.к. увеличивается число ошибок второго рода.
6. Теперь нужно использовать FDR вместо FWER и соответствующим методом Бенджамини-Хохберга. Метод обеспечивает контроль над FDR на уровне α при условии, что статистики T_i независимы (а они независимы, т.к. люди разные).
7. Число ошибок второго рода упало, мощность возросла.