

## Задача 4.3

---

Правительство города М. испытывает систему обнаружения нежелаемых 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 состоит в том, что поправка на множественную проверку гипотез отсутствует (нужно ее ввести).
2. Наилучшим способом повышения качества данного решения будет введение метода Холма – один из методов контроля FWER, который является равномерно наиболее мощным.
3. Данный метод контролирует FWER – family-wise error rate (групповая вероятность ошибки первого рода). Процедура обеспечивает  $\text{FWER} \leq \alpha$ .
4. Однако, он не учитывает точный характер зависимости между статистиками  $P_i$ .
5. При использовании данного способа мощность уменьшится относительно изначальной процедуры проверки. Хотя метод Холма обладает достаточно большой мощностью, он может уменьшиться из-за сохранения многих

нулевых гипотез, которые потенциально могут представлять исследовательский интерес и которые, соответственно, следовало бы отклонить.

6. В данном случае необходимо перейти от метода Холма к методу Бенджамини-Хохберга. Это метод контроля FDR - false discovery rate (ожидаемой доли ложных отклонений гипотез). Метод обеспечивает контроль над FDR на уровне  $\alpha$  при условии, что статистики независимы.
7. При использовании метода Бенджамини-Хохберга мощность увеличится.