

# Lab 1 task 4.3

---

## 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. Гипотеза проверяется для каждой пары снимок лица в метро - снимок лица команды Н. Вального. Необходимо учитывать поправку на множественную проверку.
2. Нельзя утверждать, что статистика для проверки всех снимков из метро с лицом самого Н. Вального независимы - вдруг у него такое лицо, что очень часто происходит false positive, а у других участников его команды нет. Не учитывая характер зависимости между статистиками, нельзя построить контролирующую FWER процедуру мощнее, чем метод Холма.
3. Контролирует  $\text{FWER} = P(V > 0) \leq \alpha$

4. Для построения мощной процедуры множественной проверки гипотез необходимо учесть структуру зависимости статистик.
5. Вероятность отклонить нулевую гипотезу при поправке уменьшается, даже если она неверна. Поэтому Уменьшается мощность критерия. Но метод Холма наиболее мощный, не учитывающий зависимости между статистиками.
6. Если правительство хочет минимизировать ошибку первого рода, нужно перейти от контроля  $FWER$  к контролю  $FDR$ . Метод можно взять метод Бенджамини-Иекутиели. Если не учитывать зависимости между статистиками, этот метод неулучшаем.
7. Вместо контроля числа ложно отвергнутых гипотез переходим к контролю доли ложно отвергнутых гипотез, значит мощность возрасла.