

Лабораторная работа №1

Акименкова Мария 675

28 марта 2020 г.

Задача 4.3

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?

Вопрос 1.

При таком подходе не используется множественная проверка гипотез. Чтобы улучшить подход, надо ввести поправку, например, Холма или Бонферрони.

Вопрос 2.

Возможно, среди 5000 снимков есть снимки одних и тех же людей, значит статистики зависимы, поэтому лучше всего использовать метод Холма.

Вопрос 3.

Метод обеспечивает контроль над групповой вероятностью ошибки первого рода на уровне α . Гарантия метода в том, что не используя знаний о связи между статистиками, он является наиболее мощным среди таких методов.

Вопрос 4.

Недостаток в том, что этот метод не использует информацию о статистиках.

Вопрос 5.

Мощность равна $1 - \alpha$, следовательно, тк уровни значимости не меньше изначальных, мощность уменьшилась.

Вопрос 6.

Надо уменьшить вероятность ошибки первого рода, поэтому будем использовать FDR (False Discovery Rate) и метод Бенджамини-Хохберга - один из методов контроля ожидаемой доли ложных отклонений гипотез (FDR) который утверждает, что при определенных ограничениях на статистики гипотез T_i для достижения контроля FDR на уровне α достаточно, чтобы отвергались гипотезы H_i , для которых $p_i \leq \frac{i\alpha}{m}$, где m — количество гипотез.

Вопрос 7.

Мощность увеличилась, тк смягчились условия на количество ошибок первого рода.