

# ЗАДАЧНЫЙ ТУР ЗАКЛЮЧИТЕЛЬНОГО ЭТАПА 2023/2024

## Задача 1 (Странные числа) — 10 %

Конверсия — это доля клиентов, перешедших с одного этапа на другой. Например, на сайт с заявками на кредитные карты зашли 50 человек, после ознакомления с условиями заявку на оформление карты (далее — заявку) оставили только 45 из них. В данном случае конверсия составляет  $90\% = 45/50$ .

Рассмотрим следующую ситуацию. В ноябре сайт посетили 100 мужчин и 100 женщин, при этом из них заявки оставили 10 мужчин и всего 5 женщин.

1. Посчитайте конверсии для мужчин и для женщин из захода на сайт в оформление заявки.
2. Посчитайте общую конверсию для всех пользователей.
3. В декабре была проведена дополнительная рекламная компания, и общее число зашедших на сайт стало больше. При этом конверсия для мужчин стала равна 12%, а для женщин — 7%. Может ли быть такое, что общая конверсия в декабре упала? Если да, то приведите численный пример. Если нет — докажите.
4. При условии увеличения конверсий у мужчин и у женщин до 12% и 7% соответственно в каком интервале будет лежать общая конверсия? Обоснуйте свой ответ.

### Задача 2 (Вкусные булочки) — 20 %

Каждый день Дима звонит в пекарню, чтобы узнать, есть ли сегодня в продаже его любимые булочки с повидлом. За последние 3 дня булочки были в наличии 2 раза, а 1 раз их не было.

Пусть переменная  $X = 0$ , если булочек нет, и  $X = 1$ , если булочки есть. Наличие булочек в конкретный день не зависит от наличия булочек в любой другой день.

1. Сколько наблюдений собрал Дима? Выпишите все значения из его выборки через запятую. Посчитайте для этой выборки среднее значение  $X$ , дисперсию  $X$ . В какой доле случаев булочки были в наличии?
2. Пусть  $p$  — вероятность того, что булочки в наличии ( $p$  не может быть меньше 0 и больше 1. Например,  $p$  может быть равно  $0,2 = 1/5$  — то есть в одном из пяти случаев булочки в наличии). Чему равна вероятность, что сегодня булочки есть, а завтра их не будет? Чему равна вероятность, что за два дня в один день булочки будут, а в другой — не будут? (Напишите два выражения, зависящие от  $p$ .)
3. Чему равна вероятность получения наблюдений как у Димы? (Напишите одно выражение, зависящее от  $p$ .)
4. При каком значении  $p$  вероятность получить выборку как у Димы максимальна? Вычислите его. Как это значение соотносится с наблюдениями Димы?
5. Дима нашел значение  $p$  из предыдущего пункта и сделал вывод, что на следующий день булочки испекут с вероятностью  $p$ . Верный ли вывод сделал Дима? Поясните свой ответ.
6. Рядом с домом Димы открыли новую пекарню, где тоже делают булочки с повидлом. Дима решил сравнить две пекарни. Для этого он собрал выборку за 100 дней: в новой пекарне булочки были в наличии 70 дней, в старой — 60 дней. Какую гипотезу может проверить Дима? Какой механизм, может лежать в основе этой гипотезы? При описании механизма вы можете сами дополнить историю Димы (например, предположить расположение старой пекарни, себестоимость повидла и другое, что могло бы помочь объяснить гипотезу, необязательно рассматривать приведенные примеры). Помогите Диме проверить описанную вами гипотезу. Есть ли разница в производительности между новой пекарней и старой?

### Задача 3 (Веселые горки) — 20 %

Дана функция  $y(x) = 20x^2 - 2000x + 80000$  и значения  $x$  от 0 до 100 с шагом 10: 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100.

1. Найдите значения функции при данных значениях  $x$  (можно использовать калькулятор/Excel/Python). Найдите среднее значение  $x$  и среднее значение  $y$ .
2. Найдите коэффициент корреляции между данными значениями  $x$  и полученными значениями  $y$ , округлите полученное значение до двух знаков после запятой. Интерпретируйте полученное число — что найденное значение позволяет сказать о зависимости между  $x$  и  $y$ ?
3. Найдите коэффициент корреляции между значениями  $x$ , меньшими 50, и соответствующими значениями  $y$ , округлите полученное значение до двух знаков после запятой. Интерпретируйте полученное число. Как оно соотносится с «поведением» функции  $y(x)$  на рассматриваемом промежутке?
4. Приведите пример переменных  $x$  и  $y$ , зависимость между которыми описывается похожим образом:
  - а) с учетом масштаба данных;
  - б) без учета масштаба данных (то есть важна только сама зависимость, а не единицы измерения).Примеры для пунктов а) и б) должны быть различными
5. Предположим, что функция изменилась и теперь выглядит следующим образом:  $y(x) = x^2 - 100x + 4000$ . Не опираясь на вычисления, ответьте на вопрос: коэффициент корреляции теперь станет больше, меньше, чем коэффициент корреляции из пункта 2 или не изменится? Обоснуйте свою гипотезу. Проверьте гипотезу, вычислив новое значение коэффициента корреляции.

#### **Задача 4 (Умный аналитик) — 25 %**

Вы аналитик ведущей игровой компании GameMasters Inc., которая специализируется на разработке мобильных игр. Ваши коллеги разработали обновленный игровой магазин, в котором игроки могут приобретать внутриигровые предметы и суперспособности. Ваша задача — провести сравнение, чтобы определить, как внедрение нового магазина повлияло на поведение пользователей в игре.

Для этого пользователи были разделены на две равные группы случайным образом:

A — пользователи, которым доступен только старый магазин;

B — пользователи, которым доступен только обновленный магазин.

Спустя месяц после запуска по каждому пользователю из каждой группы были посчитаны следующие метрики:

'revenue\_per\_user' — доход, который был получен от пользователя за период;

'orders\_cnt\_per\_user' — количество заказов, которое совершено пользователем за период;

'converted\_from\_main\_screen\_to\_item\_card\_screen' — флаг захода на экран с товарами (0 — если пользователь не заходил на карточку товара, 1 — если заходил).

В таблице приведены значения этих метрик. Также в ней находится столбец 'group', в котором указано, к какой группе (А или В) относится пользователь и столбец 'period' — характеризующий значение метрик до начала теста и во время проведения теста. На карточке товара содержится дополнительная информация, фотография и его характеристики. Однако оплатить товар можно и без захода на карточку товара.

Задача: сравните группы по каждой метрике и сделайте вывод о том, стоит ли продолжить внедрение обновленного магазина или нужно вернуть старый.

#### **Задача 5 (Оптимальный колл-центр) — 25 %**

Каждый день в колл-центр поступает ровно 100 кредитных заявок. Каждая кредитная заявка может быть либо на машину, либо на мотоцикл с равной вероятностью (учитывайте, что распределение заявок между мотоциклами и машинами случайно). Кредитные заявки обрабатывают операторы по одной заявке в день. При этом оператор, обрабатывающий кредитные заявки на машины, НЕ может обрабатывать кредитные заявки на мотоциклы, и наоборот. Каждому оператору нужно платить зарплату, даже если он сидит без дела и не обрабатывает кредитную заявку. Если число кредитных заявок превышает число операторов, то компания теряет заявку из-за длительной обработки, и клиент уходит к конкурентам.

Зарплата одного оператора — 5000 ₺ в день. За одну обработанную кредитную заявку на машину компания зарабатывает 22 000 ₺, на мотоцикл — 11 000 ₺.

1. Какое оптимальное количество операторов нужно выводить на работу каждый день для обработки кредитных заявок по машинам и мотоциклам, чтобы компания заработала как можно больше денег? Сколько всего операторов понадобится?
2. Какую максимальную ожидаемую прибыль в день можно получить?