

Время выполнения задания – 180 минут. Максимальное количество баллов – 100

Задание 1. (14 баллов) Юный аналитик Олег получил заказ от правительства Свердловской области: проанализировать взаимосвязь между погодой и частотой использования общественного транспорта жителями Свердловской области.

1. (4 балла) Сформулируйте потенциальную *гипотезу**, которую Олег мог проверять в ходе своего исследования.
2. (5 баллов) Сформулируйте экономический *механизм**, который оправдывает гипотезу из пункта 1.
3. (5 баллов) Предположим, что в результате анализа данных гипотеза из пункта 1 не подтвердилась. Сформулируйте экономический механизм, который оправдывает гипотезу, противоположную той, что представлена в пункте 1.

*Гипотеза исследования** - ваше предположение о наличии и характере связи между конкретными факторами (переменными), взаимосвязь которых является темой исследования.

*Механизм гипотезы** - подробное обоснование правильности выдвинутой гипотезы.

Задание 2. (15 баллов) Школьница Наташа получила от местного университета массив данных в Excel, который содержит информацию об абитуриентах, поступивших в него в текущем году. Массив содержит следующие переменные (переменные указаны в том порядке, в котором расположены в данных Наташи): порядковый номер абитуриента, имя, фамилия, город, возраст, пол, количество баллов ЕГЭ по математике, количество баллов ЕГЭ по русскому языку, количество баллов ЕГЭ по биологии. Для проверки основной гипотезы исследования Наташа написала следующую команду:

KORREL(G2:G1200, I2:I1200)

1. (3 балла) Какие две переменные стали фокусом исследования, которое проводила Наташа?
2. (3 балла) Какое количество абитуриентов попало в выборку исследования?
3. (3 балла) Предположим, что в результате написания команды, Наташа получила значение -0.02. Какой вывод можно сделать о взаимосвязи переменных из пункта 1 (о направлении этой взаимосвязи и ее силе)?
4. (3 балла) Изобразите, как выглядит диаграмма рассеивания для данной взаимосвязи.
5. (3 балла) Назовите два подходящих типа графика для того, чтобы показать распределение переменной «возраст». Объясните свой выбор (не более двух предложений).

Задание 3. (16 баллов) Аптека «ТюменьЗдрав» испытывает спад выручки за последние полтора месяца. Руководство компании проанализировало данные и выяснило, что группа покупателей, которая стала меньше посещать аптеку - это пенсионеры. Вас наняли для исследования вопроса: «Что влияет на выбор аптеки среди пенсионеров Тюмени?» и попросили провести исследование методом анкетирования, чтобы получить как можно более подробные ответы.

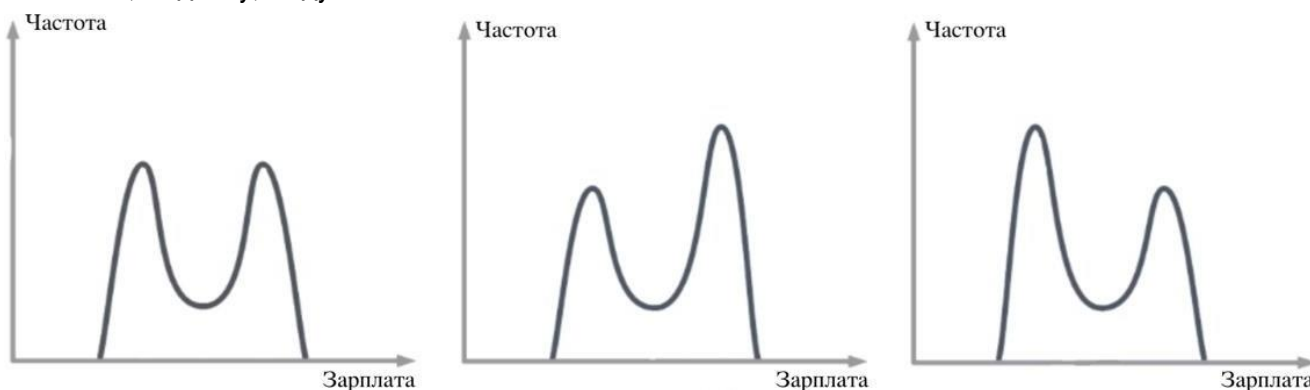
Чтобы не опрашивать всю генеральную совокупность (все пенсионеры Тюмени) и сократить расходы на исследование, руководство аптеки предложило несколько вариантов формирования выборки для анкетирования. Проанализируйте каждый из них и сделайте вывод о том, *репрезентативна** ли данная выборка. Если вы сделали вывод о не репрезентативности, объясните факт наличия смещения (не более одного предложения на пункт).

1. (2 балла) Пенсионеры Тюмени, которые регулярно посещают «ТюменьЗдрав».
2. (2 балла) Пенсионеры Тюмени, которые никогда не посещали «ТюменьЗдрав».
3. (2 балла) 50000 случайно отобранных пенсионеров Тюмени.
4. (2 балла) 50000 случайно отобранных пенсионеров Свердловской области.
5. (2 балла) Пенсионеры Тюмени, проживающие в Ленинском административном районе (где и расположена аптека «ТюменьЗдрав»).
6. (2 балла) Пенсионеры Тюмени, которые имеют проблемы со здоровьем.
7. (2 балла) Врачи поликлиник Тюмени, где обслуживаются пенсионеры.
8. (2 балла) Дети и родственники пенсионеров Тюмени, которые занимаются покупкой лекарств.

*Репрезентативная выборка** -- выборка, свойства которой совпадают со свойствами генеральной совокупности (в ней представлены все подгруппы, которые входят в генеральную совокупность).

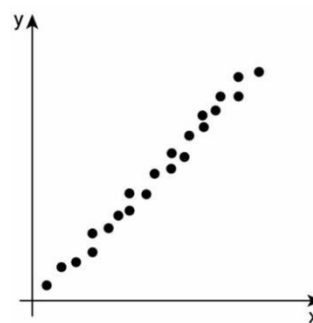
Задание 4. (15 баллов) Долгое время в компании «Альфа» распределение зарплат сотрудников совпадало с нормальным.

1. (3 балла) Изобразите данное распределение на графике схематично, указав на нем: среднее значение, моду, медиану.
2. (3 балла) Аналитики «Альфа» заметили, что в последнем месяце три показателя находятся в следующем соотношении: среднее значение > медиана > мода. Изобразите данное распределение на графике схематично, указав на нем: среднее значение, моду, медиану.
3. (3 балла) Аналитики «Альфа» связались с тремя компаниями-партнерами «Бета», «Гамма», «Омега» и запросили графики распределения зарплат среди их сотрудников. На каждом графике укажите: среднее значение, медиану, моду.



4. (6 баллов) Докажите или опровергните утверждение: «Не бывает такого распределения, для которого медиана была бы строго меньше моды, а мода - строго меньше среднего значения».

Задание 5. (15 баллов) Отличник Паша работал над проектом по анализу данных и исследовал взаимосвязь между доходами человека и количеством совершенных авиаперелетов. Собрав репрезентативные данные по своему городу и построив диаграмму рассеивания, Паша увидел взаимосвязь, представленную на графике. Он вспомнил, что на лекции по анализу данных им рассказывали, что такой график показывает сильную положительную корреляцию, и сделал вывод про свой город: «Увеличение доходов человека положительно влияет на количество авиаперелетов». За такой вывод Паша получил двойку и перестал быть отличником.



1. (5 балла) Определите, в чем состоит ошибка, и как называется феномен, который не смог распознать Паша. Почему он имеет место быть в данном примере?
2. (5 балла) Исправьте ошибку Паши и сформулируйте корректный вывод.
3. (5 баллов) Предложите, как можно было бы изменить исследование, чтобы с выводом Паши нельзя было поспорить.

Задание 6. (25 баллов) Предложите дизайн собственного исследования на интересную для вас тему и раскройте его как можно подробнее, упомянув:

1. (5 баллов) Факторы, которые являются фокусом исследования, гипотезу, механизм. Кому и почему результаты исследования могут быть полезны?
2. (5 баллов) Генеральная совокупность, ее географические и временные рамки с обоснованием их выбора. Способ формирования выборки, обоснование репрезентативности, ее размер.
3. (5 баллов) Источник данных (например, если анализ доступных данных, то какие могут быть потенциальные источники и переменные, если анкетирование – то какие вопросы войдут в анкету). Подумайте, какие дополнительные переменные можно использовать для анализа. Влияет ли на ваш исход что-то еще, помимо основной переменной интереса, и каким образом?
4. (5 баллов) Типы переменных, типы используемых графиков. Какие показатели и модели будут посчитаны/оценены для ответа на главный исследовательский вопрос?
5. (5 баллов) С какими потенциальными трудностями вы можете столкнуться в процессе проведения исследования?