

Статистическая грамотность и методы анализа данных

Задания для самоконтроля

Задание 1: Измерительные шкалы и типы данных

- **Вопрос:** Определите тип измерительной шкалы и тип данных для следующих примеров:
 - 1.Оценки студентов на экзамене (A, B, C, D, F).
 - 2.Возраст участников опроса в годах.
 - 3.Тип автомобиля (седан, хэтчбек, внедорожник).
 - 4.Температура в градусах Цельсия.

Задание 2: Обобщающие статистические показатели

- **Вопрос:** Найдите среднее значение, медиану и моду для следующих данных: 4, 7, 2, 9, 3, 7, 5.

Задание 3: Абсолютные и относительные величины

- **Вопрос:** В городе проживает 500 000 человек. За последний год численность населения увеличилась на 5%. Найдите абсолютное и относительное изменение численности населения.

Задание 4: Переменные

- **Вопрос:** Классифицируйте следующие переменные как зависимые или независимые:
 1. Время обучения и результат экзамена.
 2. Вес и рост человека.
 3. Количество часов сна и уровень концентрации.

Задание 5: Процентные изменения

- **Вопрос:** В компании было 200 сотрудников. Через год число сотрудников увеличилось до 240. Рассчитайте процентное изменение численности сотрудников.

Задание 6: Анализ рядов распределения

- **Вопрос:** Найдите диапазон, дисперсию и стандартное отклонение для следующих данных: 6, 8, 10, 12, 14.

Задание 7: Показатели центра

- **Вопрос:** Найдите среднее арифметическое, медиану и моду для следующих данных: 1, 2, 2, 3, 4, 4, 4, 5.

Задание 8: Показатели вариации

- **Вопрос:** Найдите межквартильный размах (IQR) для следующих данных: 3, 7, 8, 5, 12, 14, 21, 13, 18, 9.

Задание 9: Показатели формы распределения

- **Вопрос:** Определите коэффициент асимметрии для следующих данных: 10, 20, 30, 40, 50, 60, 70.

Задание 10: Анализ процентных изменений

- **Вопрос:** Доход компании в 2022 году составил 1 000 000 рублей, а в 2023 году увеличился до 1 200 000 рублей. Рассчитайте абсолютное и относительное изменение дохода.

Задание 11: Постановка гипотезы и источники данных

- **Задача:** Вам необходимо исследовать влияние уровня образования на доходы населения в вашем регионе.
- 1. Постановка гипотезы:** Сформулируйте гипотезу, которую вы будете проверять.
 - 2. Источники данных:** Определите, какие источники данных вы будете использовать для сбора информации (например, официальные статистические данные, опросы, базы данных компаний и т.д.).
 - 3. Сбор данных:** Опишите, какие данные будут собраны и как они будут использованы для проверки гипотезы.

Задание 12: Особенности сбора первичных и вторичных данных

- **Задача:** Вы собираете данные для исследования потребительских предпочтений на рынке смартфонов.

- 1.Первичные данные:** Опишите, какие методы вы будете использовать для сбора первичных данных (например, опросы, интервью, наблюдения) и почему.
- 2.Вторичные данные:** Определите, какие вторичные данные могут быть полезны для вашего исследования (например, отчеты о продажах, статистические данные) и где их можно найти.
- 3.Анализ данных:** Опишите, как вы будете обрабатывать и анализировать собранные данные.

Ответы по заданиям

Задание 1: Измерительные шкалы и типы данных

• **Вопрос:** Определите тип измерительной шкалы и тип данных для следующих примеров:

1. Оценки студентов на экзамене (A, B, C, D, F).
2. Возраст участников опроса в годах.
3. Тип автомобиля (седан, хэтчбек, внедорожник).
4. Температура в градусах Цельсия.

• **Ответ:**

1. Порядковая шкала, качественные данные.
2. Интервальная шкала, количественные данные.
3. Номинальная шкала, качественные данные.
4. Интервальная шкала, количественные данные.

Задание 2: Обобщающие статистические показатели

- **Вопрос:** Найдите среднее значение, медиану и моду для следующих данных: 4, 7, 2, 9, 3, 7, 5.
- **Ответ:**
- Среднее значение: $(4 + 7 + 2 + 9 + 3 + 7 + 5) / 7 = 5.29$
- Медиана: 5 (после упорядочивания: 2, 3, 4, 5, 7, 7, 9)
- Мода: 7

Задание 3: Абсолютные и относительные величины

- **Вопрос:** В городе проживает 500 000 человек. За последний год численность населения увеличилась на 5%. Найдите абсолютное и относительное изменение численности населения.
- **Ответ:**
- Абсолютное изменение: $500\,000 * 0.05 = 25\,000$ человек
- Относительное изменение: 5%

Задание 4: Переменные

- **Вопрос:** Классифицируйте следующие переменные как зависимые или независимые:

1. Время обучения и результат экзамена.
2. Вес и рост человека.
3. Количество часов сна и уровень концентрации.

- **Ответ:**

1. Независимая: время обучения, зависимая: результат экзамена.
2. Независимая: вес, зависимая: рост (или наоборот, в зависимости от контекста).
3. Независимая: количество часов сна, зависимая: уровень концентрации.

Задание 5: Процентные изменения

- **Вопрос:** В компании было 200 сотрудников. Через год число сотрудников увеличилось до 240. Рассчитайте процентное изменение численности сотрудников.
- **Ответ:** Процентное изменение = $((240 - 200) / 200) * 100\% = 20\%$

Задание 6: Анализ рядов распределения

- **Вопрос:** Найдите диапазон, дисперсию и стандартное отклонение для следующих данных: 6, 8, 10, 12, 14.
- **Ответ:**
- Диапазон: $14 - 6 = 8$
- Дисперсия: $((6-10)^2 + (8-10)^2 + (10-10)^2 + (12-10)^2 + (14-10)^2) / 5 = 8$
- Стандартное отклонение: $\sqrt{8} \approx 2.83$

Задание 7: Показатели центра

- **Вопрос:** Найдите среднее арифметическое, медиану и моду для следующих данных: 1, 2, 2, 3, 4, 4, 4, 5.
- **Ответ:**
- Среднее арифметическое: $(1 + 2 + 2 + 3 + 4 + 4 + 4 + 5) / 8 = 3.125$
- Медиана: $(3 + 4) / 2 = 3.5$
- Мода: 4

Задание 8: Показатели вариации

- **Вопрос:** Найдите межквартильный размах (IQR) для следующих данных: 3, 7, 8, 5, 12, 14, 21, 13, 18, 9.
- **Ответ:**
- Упорядоченные данные: 3, 5, 7, 8, 9, 12, 13, 14, 18, 21
- Q1 (медиана первой половины): 7.5
- Q3 (медиана второй половины): 14
- $IQR = Q3 - Q1 = 14 - 7.5 = 6.5$

Задание 9: Показатели формы распределения

- **Вопрос:** Определите коэффициент асимметрии для следующих данных: 10, 20, 30, 40, 50, 60, 70.
- **Ответ:** Коэффициент асимметрии = $(\sum(x - \bar{x})^3 / n) / (\sigma^3)$
- $\bar{x} = (10 + 20 + 30 + 40 + 50 + 60 + 70) / 7 = 40$
- $\sigma = \sqrt{(\sum(x - \bar{x})^2) / n} = \sqrt{(900 + 400 + 100 + 0 + 100 + 400 + 900) / 7} \approx 21.60$
- Коэффициент асимметрии ≈ 0 (для симметричного распределения).

Задание 10: Анализ процентных изменений

- **Вопрос:** Доход компании в 2022 году составил 1 000 000 рублей, а в 2023 году увеличился до 1 200 000 рублей. Рассчитайте абсолютное и относительное изменение дохода.
- **Ответ:**
- Абсолютное изменение: $1\,200\,000 - 1\,000\,000 = 200\,000$ рублей
- Относительное изменение: $((1\,200\,000 - 1\,000\,000) / 1\,000\,000) * 100\% = 20\%$
- Эти задания помогут слушателям курса закрепить основные понятия статистических концепций и показателей.

Задание 11: Постановка гипотезы и источники данных

- **Задача:** Вам необходимо исследовать влияние уровня образования на доходы населения в вашем регионе.
- 1. Постановка гипотезы:** Сформулируйте гипотезу, которую вы будете проверять.
 - 2. Источники данных:** Определите, какие источники данных вы будете использовать для сбора информации (например, официальные статистические данные, опросы, базы данных компаний и т.д.).
 - 3. Сбор данных:** Опишите, какие данные будут собраны и как они будут использованы для проверки гипотезы.
- **Пример ответа:**
 - 1. Гипотеза:** Люди с высшим образованием имеют более высокий доход по сравнению с людьми без высшего образования.
 - 2. Источники данных:** Официальная статистика, опросы населения, базы данных компаний.
 - 3. Сбор данных:** Собрать данные о доходах и уровне образования из официальных статистических источников и провести опрос среди населения для получения дополнительной информации.

Задание 12: Особенности сбора первичных и вторичных данных

- **Задача:** Вы собираете данные для исследования потребительских предпочтений на рынке смартфонов.
 1. **Первичные данные:** Опишите, какие методы вы будете использовать для сбора первичных данных (например, опросы, интервью, наблюдения) и почему.
 2. **Вторичные данные:** Определите, какие вторичные данные могут быть полезны для вашего исследования (например, отчеты о продажах, статистические данные) и где их можно найти.
 3. **Анализ данных:** Опишите, как вы будете обрабатывать и анализировать собранные данные.
- **Пример ответа:**
 1. Первичные данные: Проведение опросов и интервью с потребителями для получения информации о предпочтениях и ожиданиях от смартфонов.
 2. Вторичные данные: Использование отчетов о продажах, статистических данных о рынке смартфонов, обзоров и рейтингов в интернете. Эти данные можно найти в отчетах аналитических компаний и на специализированных веб-сайтах.
 3. Анализ данных: Объединение и очистка данных в Excel с использованием Power Query, анализ с помощью сводных таблиц и визуализация результатов.

Задание 13: Методы создания базы данных и чистка данных

- **Задача:** Вы создали базу данных в Excel для анализа продаж вашего интернет-магазина за последний год.
- 1.Создание базы данных:** Опишите, как вы будете структурировать данные в Excel (например, какие столбцы и строки будут использоваться, какие данные будут включены).
 - 2.Чистка данных:** Приведите пример, какие ошибки могут встретиться в данных (например, дублирующиеся записи, пропущенные значения) и как вы будете их исправлять с помощью Power Query или SPSS.
 - 3.Обработка данных:** Опишите, какие шаги вы предпримете для подготовки данных к анализу (например, нормализация данных, преобразование форматов, создание новых вычисляемых полей).

Задание 13: Методы создания базы данных и чистка данных

- **Задача:** Вы создали базу данных в Excel для анализа продаж вашего интернет-магазина за последний год.
- 1. **Создание базы данных:** Опишите, как вы будете структурировать данные в Excel (например, какие столбцы и строки будут использоваться, какие данные будут включены).
- 2. **Чистка данных:** Приведите пример, какие ошибки могут встретиться в данных (например, дублирующиеся записи, пропущенные значения) и как вы будете их исправлять с помощью Power Query или SPSS.
- 3. **Обработка данных:** Опишите, какие шаги вы предпримете для подготовки данных к анализу (например, нормализация данных, преобразование форматов, создание новых вычисляемых полей).
- **Пример ответа:**
 1. Создание базы данных: Столбцы будут включать информацию о дате покупки, названии товара, категории товара, количестве проданных единиц, цене за единицу, общей сумме продажи и идентификаторе покупателя.
 2. Чистка данных: Проверка на наличие дублирующихся записей, удаление дубликатов. Заполнение пропущенных значений с использованием средних значений или медиан, использование Power Query для автоматической очистки данных.
 3. Обработка данных: Преобразование форматов дат в единый формат, создание новых полей для анализа (например, доход на покупателя, количество покупок в месяц), нормализация данных для устранения выбросов и аномалий.