


## Требования к итоговому проекту по статистике

Итоговый проект по статистике направлен на развитие навыков анализа реальных данных, формулирования и проверки гипотез, а также интерпретации результатов. Вам предлагается выполнить одно из двух аналитических заданий (по выбору), каждое из которых требует самостоятельной постановки проблемы, сбора или анализа данных, применения статистических методов и формулировки выводов.

Выберите **один из двух вариантов** (или выполните оба — для максимальной проработки темы):

### Вариант 1. Математико-статистический анализ существующих данных

 **Цель:** Научиться извлекать знания из реальных или открытых данных путём строгого статистического анализа — через построение точечных и интервальных оценок, проверку гипотез и базовое моделирование. Главное – **все вычисления делать вручную на листах бумаги А4**.

#### Что необходимо сделать

##### 1. Выбор темы и набора данных

- Найдите открытый или рабочий датасет (например, Kaggle, Росстат, ВЦИОМ, Госуслуги, opendata.gov.ru и пр.).
- Отберите **небольшую подвыборку размером от 15 до 30 наблюдений** (можно вручную из большого массива), чтобы анализ был компактным и ручным.

##### Примеры тем:

- анализ цен на аренду жилья в одном городе;
- успеваемость студентов по двум предметам;
- рейтинг удовлетворённости клиентов (отзывы);
- параметры автомобилей (мощность, расход, цена).

##### 2. Формулировка исследовательского вопроса и гипотез

Сформулируйте 1–2 **основные гипотезы**, которые можно проверить с помощью статистических критериев.

##### Примеры:

- Средний уровень зарплат в регионе X превышает 60 000 руб.
- Расход топлива у автомобилей с дизельным и бензиновым двигателями различается.
- Удовлетворенность клиентов не зависит от региона обслуживания.

### 3. Статистический анализ (основная часть)

(а) **Вычисление точечных оценок.** Для основных параметров данных рассчитайте:

- среднее арифметическое  $\bar{x}$ ,
- дисперсию  $s^2$  и стандартное отклонение  $s$ ,
- медиану, моду, минимум, максимум,
- доли/проценты (если есть категориальные данные).
- найти оценку по методу ММП
- проверить оценку по методу Рао-Крамера

(b) **Построение доверительных интервалов.** Постройте интервальные оценки для:

- среднего (с известной или неизвестной дисперсией,  $t$  или  $Z$ ),
- дисперсии (используя  $\chi^2$ ),
- доли (для категориальных переменных, при необходимости).

Например:

- 95% доверительный интервал для среднего значения веса/цены/оценки.
- Интервал для разности средних двух групп (если сравниваете группы).

(с) **Проверка статистических гипотез.** Для подтверждения/опровержения гипотез примените:

- **t-критерий** (одновыборочный, двухвыборочный, парный — по задаче),
- **$\chi^2$ -критерий** (для категориальных переменных, например, «удовлетворён/не удовлетворён»),
- **F-критерий Фишера** (если необходимо сравнение дисперсий),
- и/или **проверку гипотезы о доле.**



Обязательно укажите:

- **нулевую и альтернативную гипотезу,**
- **уровень значимости  $\alpha$**  (обычно 0.05),
- **критическую область / p-value,**
- **вывод:** отвергается ли  $H_0$ , и как это интерпретируется.

### 4. Визуализация и интерпретация

Постройте хотя бы 2–3 графика:

- гистограмму или boxplot,
- диаграмму рассеяния,
- график доверительного интервала (если возможно).


Напишите интерпретации:

— Что означают оценки?

— Совпали ли они с ожиданиями?

— Подтвердились ли гипотезы?

## **Вариант 2. Проведение собственного статистического исследования (опрос)**

 **Цель:** самостоятельно провести мини-исследование — от постановки гипотез до анализа собранных данных. Это позволяет освоить полный цикл статистического исследования: формулировку задачи, разработку анкеты, сбор данных, анализ и выводы.

### **Что необходимо сделать**

#### **1. Выбор темы и постановка гипотез**

Выберите актуальную, интересную и измеримую тему из сфер:

- психология и поведение,
- потребительские предпочтения,
- обучение и мотивация,
- социальные установки и т.п.

#### **Примеры тем:**

- Влияние уровня тревожности на выбор жанров сериалов.
- Самооценка и склонность к импульсивным покупкам.
- Зависимость отношения к рекламе от возраста или образования.
- Связь между академической мотивацией и количеством часов сна.
- Кто умнее или шизанутее ШЦТ или ШАД

**Сформулируйте 2–4 гипотезы,** которые можно проверить количественно.  
Примеры:

- $H_0$ : Средний уровень тревожности не зависит от пола.
- $H_0$ : Доля респондентов, регулярно читающих книги, одинакова в группах 18–25 и 26–35 лет.
- $H_0$ : Нет связи между самооценкой и склонностью к эмоциональному перееданию.

## 2. Разработка анкеты

### Требования:

- Минимум 10–15 вопросов.
- Включить:
  - **шкальные** вопросы (например, от 1 до 5 или 1 до 10),
  - **альтернативные** (да/нет),
  - **категориальные** (пол, возраст, статус, доход и т.д.),
  - при желании — **блочные методики** (например, шкала самооценки Дембо–Рубинштейн, шкала тревожности Спилбергера и пр.)

### Инструменты для сбора:

- Google Forms, Yandex Forms, Survio и др.
- Можно опросить студентов, коллег, друзей, родственников (желательно  $\geq 30$  респондентов).

## 3. Анализ собранных данных

### ◆ (а) Описательная статистика

- частоты ответов, доли категорий;
- средние значения по шкальным вопросам;
- стандартные отклонения;
- медианы, мода, разброс, IQR.

### ◆ (b) Проверка гипотез. Выберите подходящие статистические тесты в зависимости от типа переменных:

Тип данных	Пример критерия
Среднее между двумя группами	t-критерий Стьюдента
Связь между двумя шкальными шкалами	корреляция Пирсона/Спирмена
Доля/частота между группами	$\chi^2$ -критерий
Сравнение по 1 шкале с эталоном	одновыборочный t-тест

Обязательно указывайте:

- $H_0$  и  $H_1$ ,
- уровень значимости  $\alpha$ ,
- p-value или критическое значение,
- интерпретацию результата: подтвердилось ли предположение.

#### 4. Визуализация и интерпретация

Постройте хотя бы 2–3 графика:

- круговые диаграммы (категории),
- столбчатые/гистограммы (распределения),
- boxplot или scatterplot (для шкальных данных),
- графики доверительных интервалов — если применимо.

**Напишите интерпретации:**

- Подтверждаются ли гипотезы?
- Какие есть интересные или неожиданные закономерности?
- Какие могли быть искажения или ограничения (размер выборки, субъективность, нерепрезентативность)?

#### 5. Структура финального отчета

##### 1. Введение

- описание темы и обоснование значимости,
- цели исследования,
- гипотезы.

## **2. Методика**

- описание респондентов,
- структура анкеты,
- способ проведения,
- переменные.

## **3. Анализ и результаты**

- описательная статистика,
- графики,
- таблицы,
- результаты проверок гипотез.

## **4. Обсуждение и выводы**

- интерпретация результатов,
- подтверждение/опровержение гипотез,
- практическая ценность и ограничения.

## **5. Приложения**

- текст анкеты,
- графики (если отдельно),
- код (если использовался).

### **■ Примечание:**

**Качество важнее объема.** Главное — логическая связь между гипотезами, методами и результатами. Не нужно избыточных расчетов, если они не отвечают на исследовательский вопрос.

Пример выполненной работы можете посмотреть в корне папки.