

Лабораторная работа №3.

1. Загрузить данные из файла data.csv. В таблице приведены названия столбцов и соответствующие вопросы:

Название столбца	Вопрос
is_stud	Вы студент?
group	Из какой группы?
sex	Ваш пол
work_status	Вы работаете?
fav_year	Какой год обучения вам понравился больше всего?
visit_freq	Оцените, как часто вы посещаете пары?
usd_pred	Ваш прогноз на курс доллара на 01.01.2025
random_value	Сгенерируйте число от 0 до 1

2. Визуализировать данные таблицы, рассчитать некоторые базовые статистики. Дать описание используемых в п.3-п.4 данных (например, кол-во студентов в группах, кол-во по полу, доля работающих в группах и т.п.)
3. Выдвинуть 2 гипотезы относительно влияния фактора на бинарную метрику. С помощью Биномиального теста (логику теста реализовать самому) проверить гипотезы, дать интерпретацию полученных результатов.

Пример формулировки гипотезы:

Предположим, что каждый студент с некоторой вероятностью p_1 приходит на занятие.

Студентов можно разделить на 2 группы по признаку наличия работы. X_1 – работающие студенты, X_2 – не работающие.

Хотим проверить, одинакова ли частота посещения пар среди групп.

$$H_0 : p_1 = p_2.$$

Гипотеза: **Фактор «студент работает/не работает» не влияет на «посещаемость».**

4. Выдвинуть 2 гипотезы относительно влияния фактора на непрерывную метрику. С помощью t-критерия Стьюдента проверить гипотезы, дать интерпретацию полученных результатов.