

**Внимание!** Для выполнения тестовых заданий скачайте и откройте массив данных по ссылке:

<https://docs.google.com/spreadsheets/d/1EOEmGcBpokRfYbiNBDQs5XnWG9QGmOSwYKpKiOkhQR4/edit?usp=sharing>

1. Во вкладке "Данные об аудитории" информация о пользователях, посетивших наше приложение в ноябре. Чему равен MAU продукта?

Ответ: MAU = 7639.

\*MAU (Monthly Active Users) — это метрика, используемая для измерения активности пользователей в течение одного месяца. Она показывает количество уникальных пользователей, которые взаимодействовали с продуктом, сервисом или приложением хотя бы один раз за последний месяц.

2. Используя вкладку "Данные об аудитории", посчитайте, чему будет равен DAU

Ответ: DAU = 560.5.

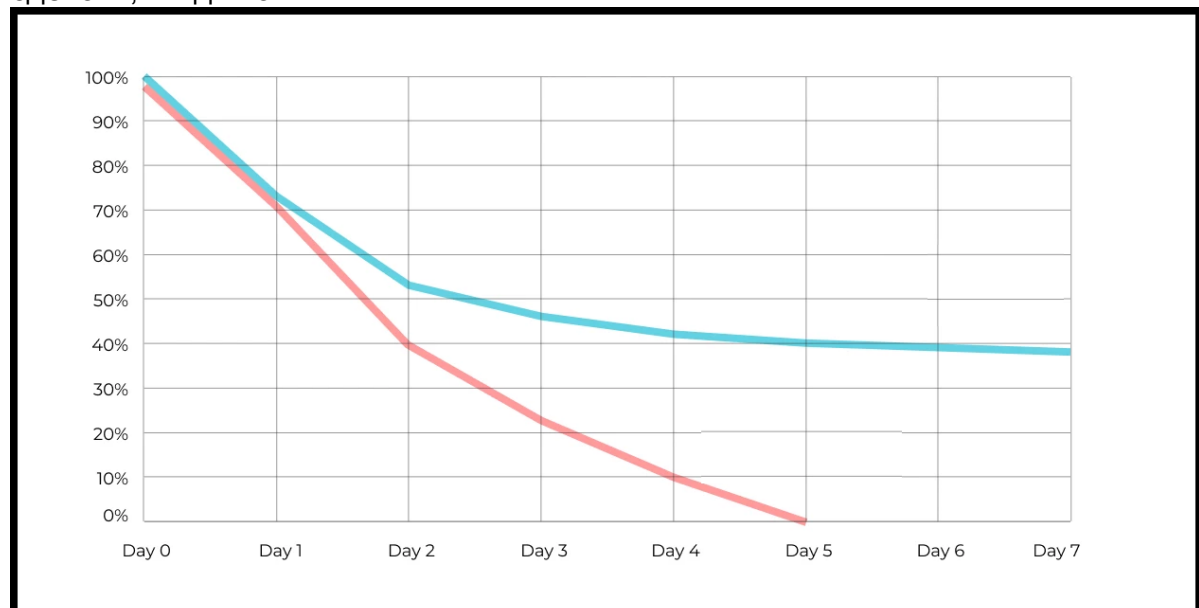
\*DAU (Daily Active Users) — это метрика, которая показывает количество уникальных пользователей, которые взаимодействовали с продуктом, приложением или сервисом хотя бы один раз в течение дня. DAU помогает понять, сколько пользователей активно пользуются продуктом каждый день.

3. Используя вкладку "Данные об аудитории", посчитайте, чему будет равен retention первого дня у пользователей, пришедших в продукт 1 ноября

Ответ: Retention первого дня = 67.6%.

\*Retention (удержание пользователей) — это метрика, которая показывает, сколько пользователей продолжает пользоваться продуктом через определенный промежуток времени после первоначального взаимодействия. Retention можно рассчитать как процент пользователей, вернувшихся в продукт через определенное время (например, через 1 день, 1 неделю, 1 месяц) от количества всех новых пользователей.

4. На графике изображены retention кривые 2 продуктов. Какие выводы можно сделать, глядя на них?



1. Продукт 1 (синяя линия):

- Изначально имеет высокое удержание, которое плавно снижается к 7 дню. Это значит, что удержание этого продукта значительно выше красного и возврат к продукту выше.

2. Продукт 2 (красная линия):

- Удержание значительно быстрее падает в течение первых 3 дней и практически достигает нуля к 7 дню. По сравнению с синим продуктом товар имеет меньшую удерживаемость и соответственно быстрый спад.

Рекомендации:

- Проанализировать причины, по которым пользователи уходят из Продукта 2, и улучшить функционал или пользовательский опыт.
- Рассмотреть успешные аспекты Продукта 1 и применить их к Продукту 2.

5. Во вкладке "Данные об аудитории" есть информация о том, сколько объявлений посмотрел каждый пользователь (view\_adverts). Посчитайте пользовательскую конверсию в просмотр объявления за ноябрь? (в пользователях)

Ответ: Конверсия = 46.3%.

\* Пользовательская конверсия — это метрика, которая показывает, какой процент пользователей выполнил целевое действие по отношению к общему количеству пользователей. В контексте веб-сайтов это может быть действие, такое как просмотр объявления или клик по рекламному баннеру.

6. Используя информацию из вкладки "Данные об аудитории", посчитайте среднее количество просмотренных объявлений на пользователя в ноябре

Ответ: 4.9

7. Мы провели опрос среди 2000 пользователей. Из них 500 «критики», 1200 «сторонники» и 300 «нейтралы». Посчитайте, чему будет равен NPS

Ответ: NPS = 35.0%.

\* NPS (Net Promoter Score) — это метрика, которая измеряет лояльность пользователей к компании или продукту и делит их на три группы: Сторонники (Promoters), Нейтралы (Passives), Критики (Detractors). NPS высчитывается как (% сторонников - % критиков).

8. Во вкладке "Данные АБ-тестов" результаты трех несвязанных АБ тестов для ARPU (общая выручка/общее количество пользователей).

Посмотрите на результаты тестов и интерпретируйте их. Напишите значения p-value, которые вы получили.

Подготовьте выводы и рекомендации.

experiment\_num - номер эксперимента

experiment\_group - группа, в которую попал пользователь

user\_id - id пользователя

revenue - выручка, которую сгенерировал пользователь, купив платную услугу продвижения

## A/B Test Results with ARPU and p-values

#	experiment_num	ARPU_test	ARPU_control	p_value	
1	1	665.7395833333334	722.4602150537635	0.688966190624939	
2	2	332.9291666666667	704.6537634408602	0.0011282266247294027	
3	3	998.66875	663.2064516129033	0.06031548002841902	

Рекомендации:

1. Эксперимент 1:
  - Поскольку разницы в ARPU нет, оставьте контрольный вариант. Тестовый вариант не предоставляет преимуществ.
2. Эксперимент 2:
  - Не внедряйте изменения, тестовый вариант значительно ухудшает ARPU.
  - Проведите дополнительный анализ, чтобы понять, почему тестовый вариант негативно влияет на доход.
3. Эксперимент 3:
  - Проведите тест с увеличенной выборкой, чтобы подтвердить тренд, так как тестовый вариант показывает улучшение ARPU, но без статистической значимости.
  - Возможно, потребуется дополнительный анализ для выявления факторов, влияющих на результат.

Если потребуется добавить подробности или расширить выводы, дайте знать!

9. По датасету с листерами посчитайте средний доход на пользователя

Ответ: Средний доход = 30.70.

10. По датасету с листерами посчитайте медиану возраста пользователя

Ответ: 27 лет

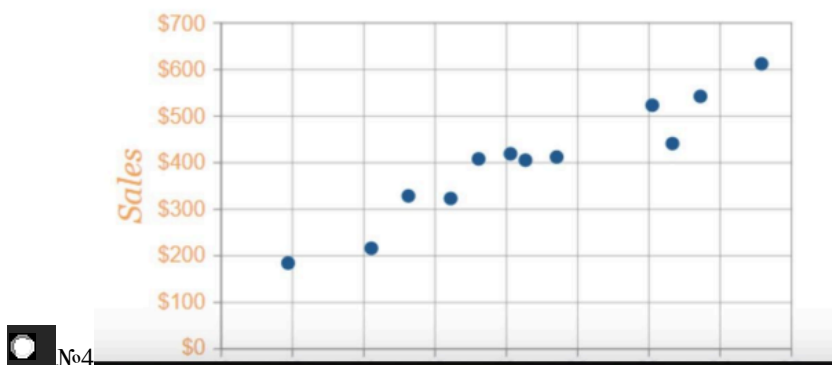
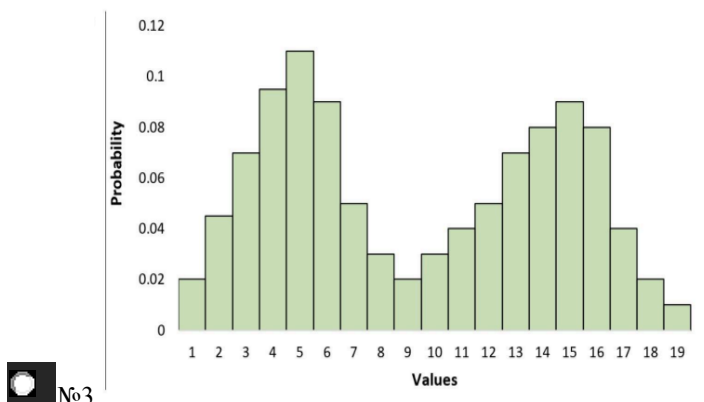
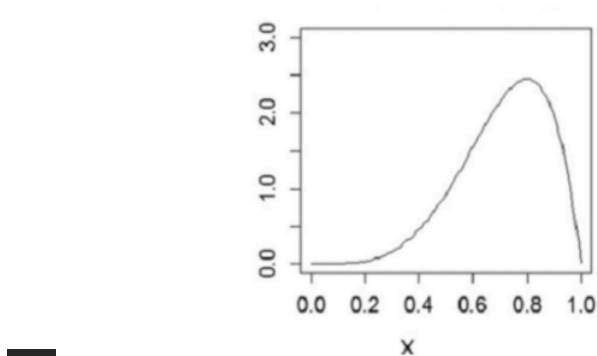
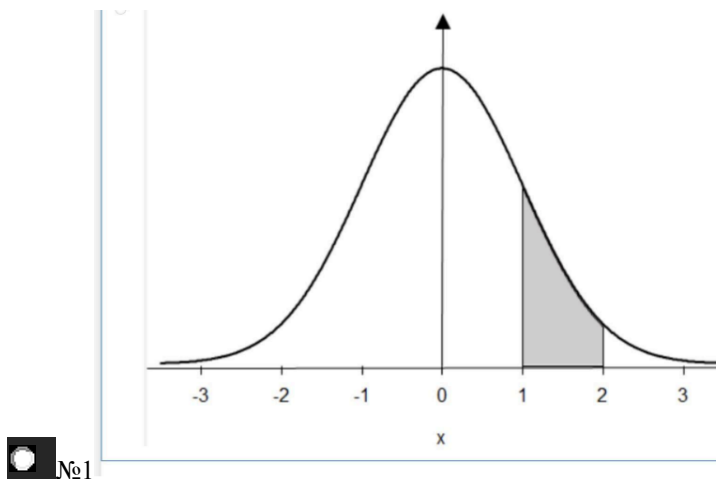
11. Какой график лучше всего подходит для отображения разброса цен на товары в разных магазинах?

\*возможно несколько вариантов ответа

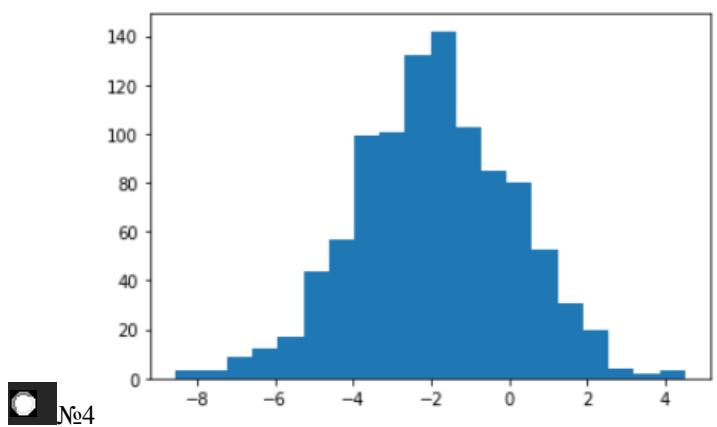
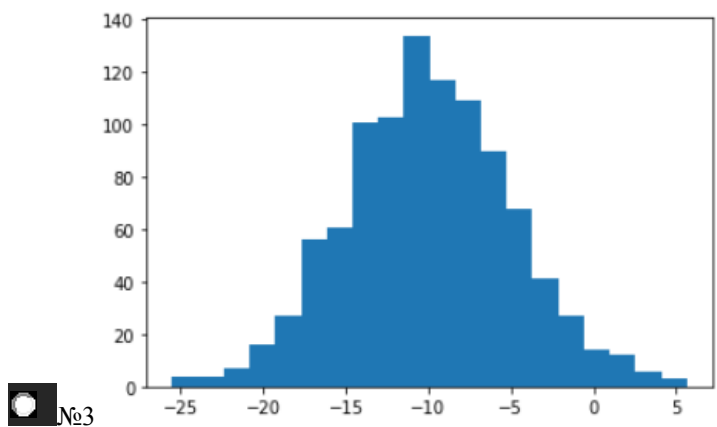
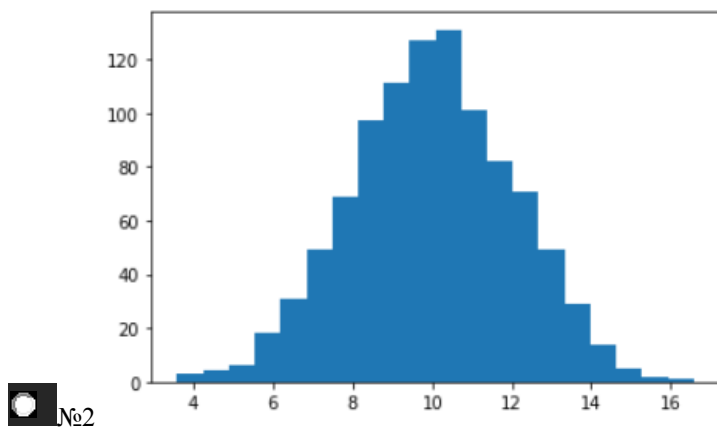
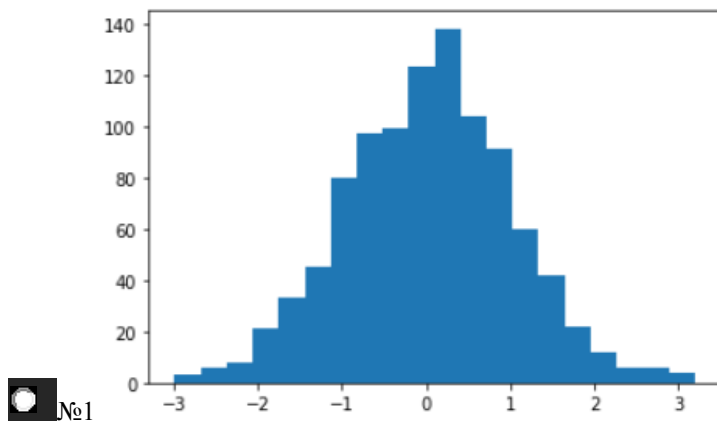
Ответ: гистограмма, Ящик с усами (box plot)

12. На каком графике бимодальное распределение?

Ответ: N3

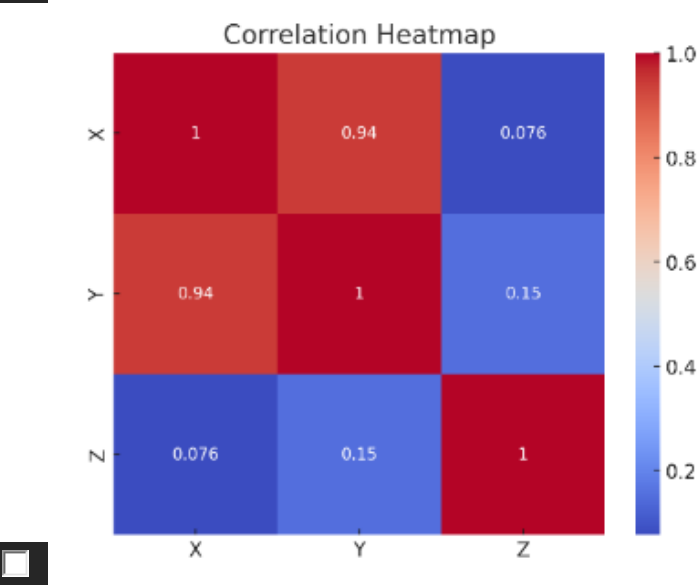
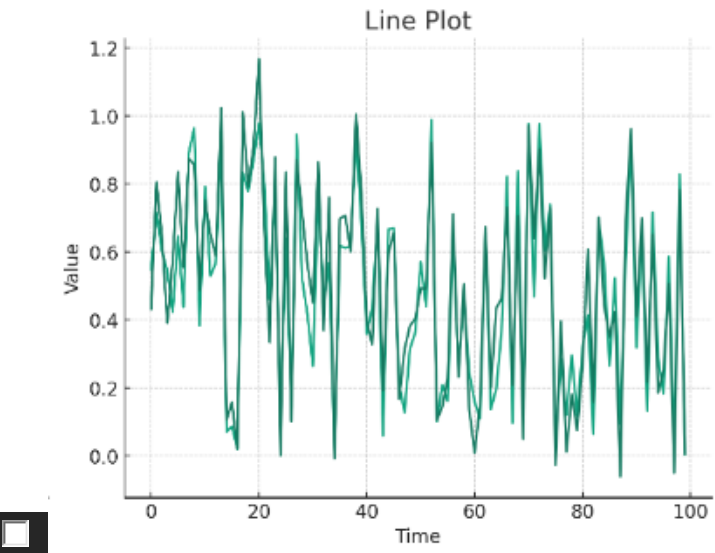
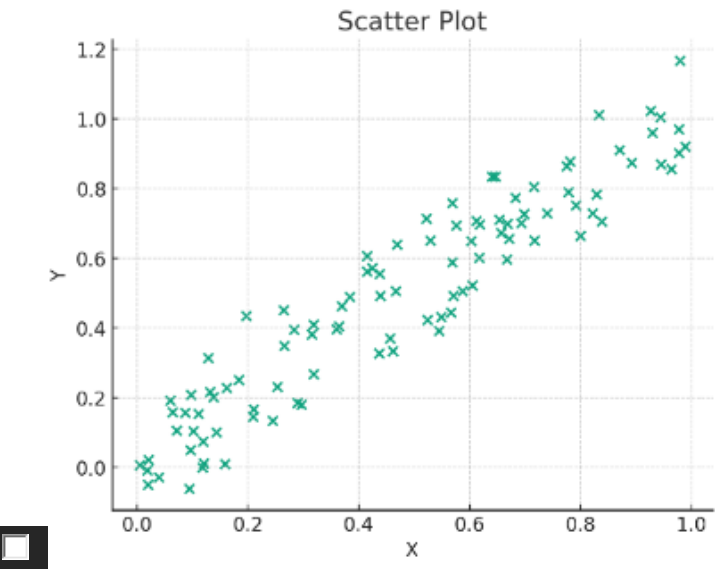


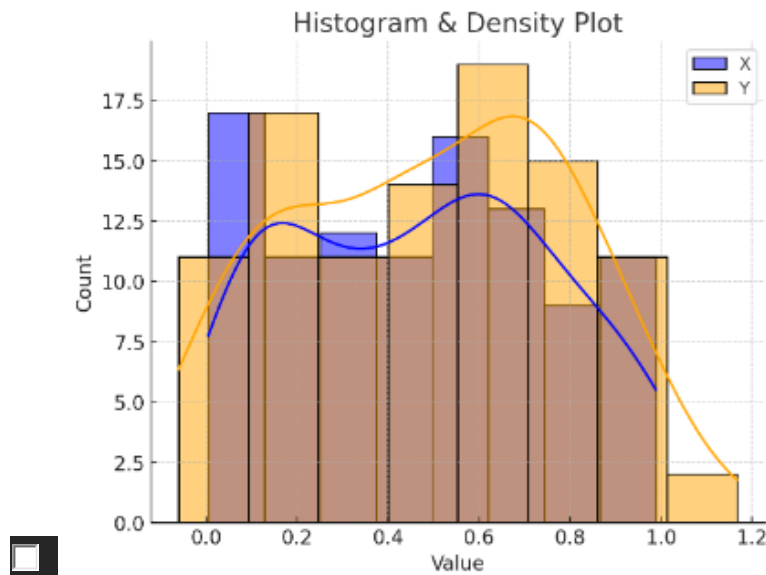
13. Какая случайная величина имеет наибольшую дисперсию данных по следующим графикам плотности распределения?  
 Ответ: N3 потому что разброс и высота графика влияет на наибольшую дисперсию.



14. На каком графике можно посчитать корреляцию?  
 \*возможно несколько вариантов ответа

Ответ: на всех, так как 2 графика - линейные, остальные 2 точечные и уже с
 распределенной корреляцией.





15. Что значит, если при проверке гипотез мы получили  $p\text{-value} = 0.05$ ?

Ответ: Есть 5% вероятность случайно получить такой или еще более экстремальный результат, если нулевая гипотеза верна

$p\text{-value}$  — это вероятность наблюдать текущий результат (или более экстремальный) при условии, что нулевая гипотеза истинна. Значение 0.05 указывает на то, что существует 5% вероятность ошибочно отвергнуть нулевую гипотезу.

16. Какой метод наиболее подходит для проверки гипотезы о равенстве средних двух выборок из нормального распределения?

Ответ: Т-рекс, то есть Т-тест :)

17. Как интерпретировать квантили в распределении доходов пользователей?

Ответ: Делят данные на четыре равные части

Квантили — это точки, которые делят упорядоченные данные на четыре равных части, каждая из которых содержит 25% наблюдений. Например:

25% пользователей имеют доход ниже первого квантили.

50% пользователей находятся между первым и третьим квантилем.

18. Были получены следующие результаты. Коллеги просят вас подтвердить их и сделать окончательный вывод по эксперименту.

- Вариант А (контрольная группа) — 100 047 501 посетитель, 1003 платежа.
- Вариант В (тестовая группа) — 100 001 055 посетителей, 1099 платежей.

Какие рекомендации вы бы дали, основываясь на этих данных?

Данные:

- Вариант А: 100 047 501 посетитель, 1003 платежа.
- Вариант В: 100 001 055 посетителей, 1099 платежей.

Анализ:

1. Конверсия платежей:
  - Вариант А:  $(1003/100047501) \times 100 = 0.0010\%$
  - Вариант В:  $(1099/100001055) \times 100 = 0.0011\%$

2. Выводы:
- Конверсия в варианте В немного выше, чем в варианте А.
  - Однако разница слишком мала, чтобы делать окончательный вывод без проверки статистической значимости (например, с помощью t-теста, ой то-есть Т-теста).

Рекомендация:

Рекомендуется провести статистический тест для проверки значимости разницы. Если разница окажется значимой, можно внедрить Вариант В. Если нет, продолжить использовать Вариант А.