**Attention!** To complete the test tasks, download and open the data array using the link:

[Копия Данные для тестового задания](https://docs.google.com/spreadsheets/d/1TB8gc40MtI4SK0pwo2XOg0o51OHYTSujBjYK7rKlWm4/edit?gid=1687485228#gid=1687485228)

1. In the "Audience Data" tab, information about users who visited our app in November. What is the MAU of the product?

\*MAU (Monthly Active Users) is a metric used to measure user activity for one month. It shows the number of unique users who have interacted with a product, service, or application at least once in the last month.

7639168141048216529



=COUNTA(UNIQUE(FILTER(B2:B, MONTH(A2:A) = 11, YEAR(A2:A) = 2023)))

**Answer: 7639**

2. Using the "Audience Data" tab, calculate what the DAU will be.

\*DAU (Daily Active Users) is a metric that shows the number of unique users who have interacted with a product, application, or service at least once during the day. DAU helps to understand how many users actively use the product every day.

255490560483



=COUNTA(UNIQUE(FILTER(B2:B16815; A2:A16815 = A2)))

**Answer: 560**

3. Using the "Audience Data" tab, calculate what the first-day retention rate will be for users who joined the product on November 1st.

\*Retention is a metric that shows how many users continue to use a product after a certain period following their initial interaction. Retention can be calculated as the percentage of users who returned to the product after a specific time (e.g., after 1 day, 1 week, 1 month) out of the total number of new users.

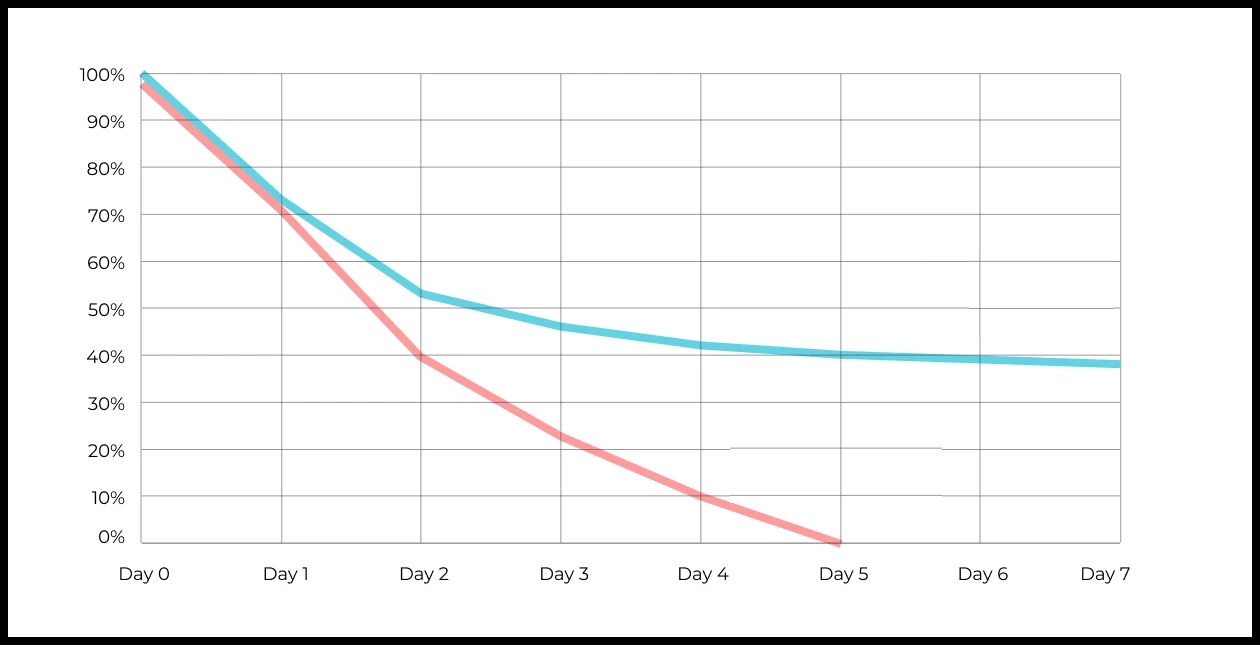
28,3%26,6%38,5%32,7%



=COUNTA(UNIQUE(FILTER(A2:A16815; B2:B16815 = DATE(2023; 11; 1))))) \* 100

**Answer: 26,6%**

4. On the graph, there are retention curves for two products. What conclusions can be drawn by looking at them?



Your answer:

The blue line shows higher user retention: even after a week, a significant portion of users return.

The red line shows a sharp drop in retention, and by the fifth day, almost no one returns.

This means that the first product retains users better for a long time, while the second product loses most of its users in the first few days.

5. In the "Audience Data" tab, there is information about how many ads each user has viewed (view\_adverts). Do you calculate the user conversion per ad view for November? (in users)

\* User conversion is a metric that shows what percentage of users completed a target action relative to the total number of users. In the context of websites, this can be an action such as viewing an ad or clicking on an advertising banner.

41,8%54,7%46,3%39%



= (COUNTA(UNIQUE(FILTER(A2:A16815; (MONTH(B2:B16815) = 11) \* (YEAR(B2:B16815) = 2023) \* (C2:C16815 > 0)))) / COUNTA(UNIQUE(FILTER(A2:A16815; (MONTH(B2:B16815) = 11) \* (YEAR(B2:B16815) = 2023))))) \* 100

**Answer: 46,3%**

6. Using the information from the Audience Data tab, calculate the average number of ads viewed per user in November

4,96,25,32,9



= SUMIFS(C2:C16815; (MONTH(B2:B16815) = 11) \* (YEAR(B2:B16815) = 2023)) / COUNTA(UNIQUE(FILTER(A2:A16815; (MONTH(B2:B16815) = 11) \* (YEAR(B2:B16815) = 2023))))

7. We conducted a survey among 2,000 users. Of them, 500 are "critics," 1,200 are "promoters," and 300 are "neutrals." Calculate the NPS.

\*NPS (Net Promoter Score) — is a metric that measures user loyalty towards a company or product and categorizes them into three groups: Promoters, Passives, and Detractors. NPS is calculated as (% of promoters - % of detractors).

30%43%40%35%  
**Answer: 35%**



8. In the "AB Test Data" tab, there are results of three unrelated AB tests for ARPU (total revenue/total number of users).

1. **Look at the test results and interpret them.**
2. **Write down the p-values you obtained.**
3. **Prepare conclusions and recommendations.**

**Data Columns:**

* **experiment\_num**: Experiment number
* **experiment\_group**: Group the user was assigned to
* **user\_id**: User ID
* **revenue**: Revenue generated by the user through purchasing a paid promotion service.

Your answer:

import pandas as pd

from scipy.stats import ttest\_ind

data = pd.read\_excel("Данные для тествого задания.xlsx", sheet\_name="Данные АБ тестов")

results = []

for experiment\_num in data['experiment\_num'].unique():

exp\_data = data[data['experiment\_num'] == experiment\_num]

test\_group = exp\_data[exp\_data['experiment\_group'] == 'test']['revenue']

control\_group = exp\_data[exp\_data['experiment\_group'] == 'control']['revenue']

t\_stat, p\_value = ttest\_ind(test\_group, control\_group, equal\_var=False)

results.append({

"experiment\_num": experiment\_num,

"test\_mean\_ARPU": test\_group.mean(),

"control\_mean\_ARPU": control\_group.mean(),

"p\_value": p\_value

})

results\_df = pd.DataFrame(results)

print(results\_df)

p-value

0.796006

0.008453

0.001001

9. Calculate the average income per user based on the dataset with the listers

121.2156.470.930.7the average is not applicable here  
**Answer: 30.7%**



10. Based on the dataset with the listers, calculate the median age of the user

27,422827,9327the median is not applicable here



**Answer: 28%**

11. Which chart is best suited to display the spread of prices for goods in different stores?

\*There may be several possible answers.

Line chartPie chartBox with whiskers (box plot)

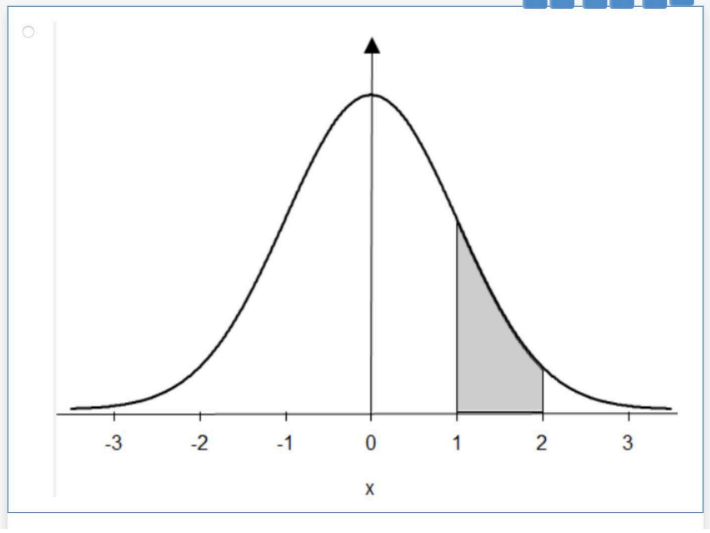


The histogram

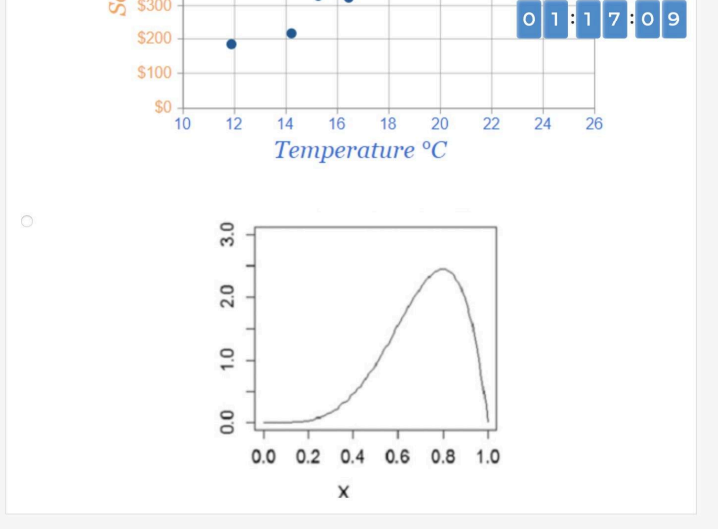


**Answer: Box plot, Historgram**

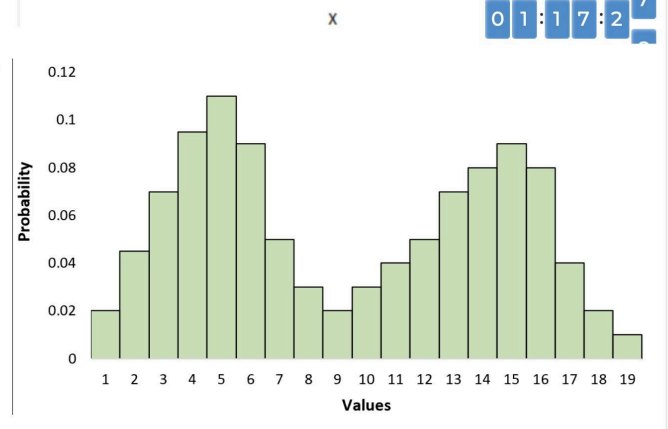
12.What is the bimodal distribution graph?

№1

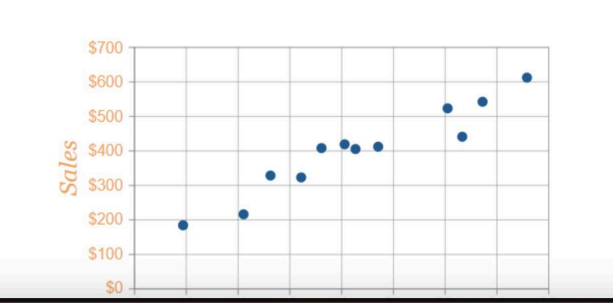


№2



№3

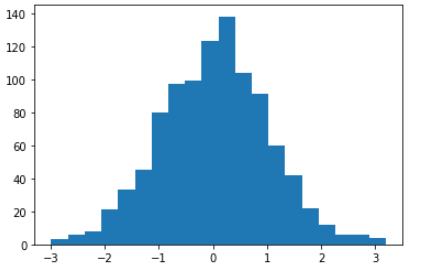


№4

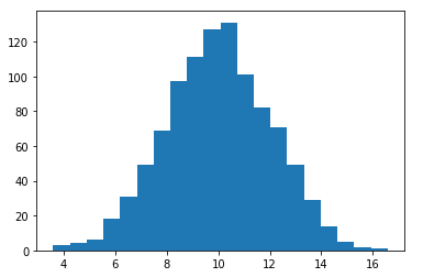


**Answer: 1**

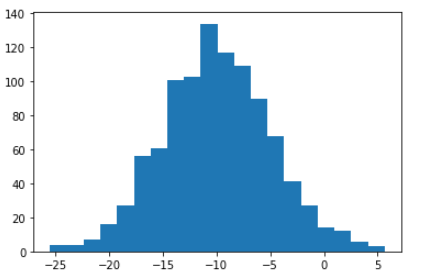
13. Which random variable has the largest data variance according to the following distribution density graph?

№1

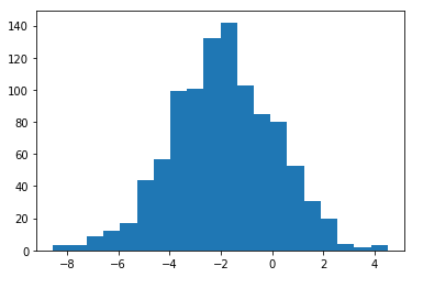


№2



№3

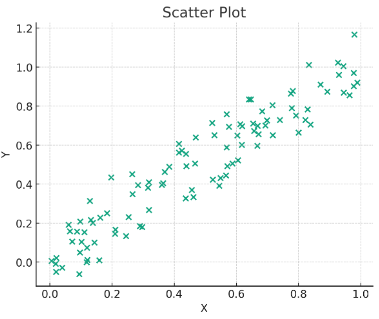


№4

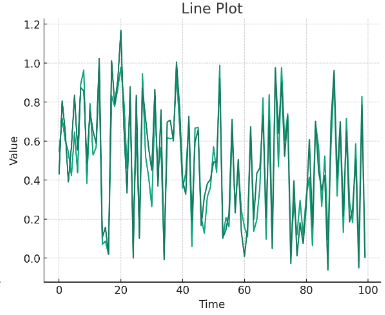


**Answer: 3**

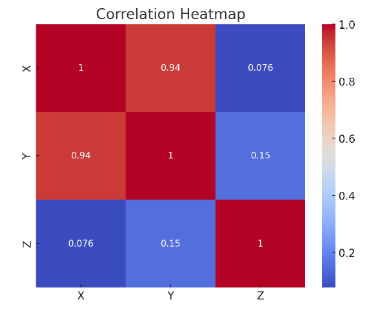
14. On which graph can the correlation be calculated?  
\**There may be several possible answers.*



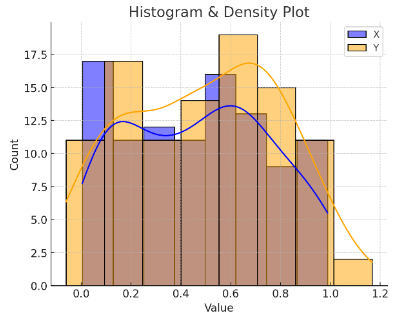














**Answer: 1**

15. What does it mean if, when testing hypotheses, we got p-value = 0.05?

This means that there is no statistically significant difference between the groups



There is a 5% chance of accidentally getting this or an even more extreme result if the null hypothesis is correct



This means that the results of the experiment are 95% accurate



This suggests that the alternative hypothesis is correct with a probability of 95%



**Answer: 2**

16. Which method is most suitable for testing the hypothesis of equality of the average of two samples from a normal distribution?

t-testChi-square testAnalysis of Variance (ANOVA)Pearson Correlation



**Answer: 1**

17. How to interpret quartiles in the distribution of user income?

They show the maximum and minimum income



Divide the data into four equal parts



Indicate the most common income



Graph of the density distribution of matter in the universe



**Answer: 2**

18. The following results were obtained. Colleagues ask you to confirm them and make a final conclusion on the experiment.

* + - * Option A (control group) — 100,047,501 visitors, 1003 payments.
      * Option B (test group) — 100,001,055 visitors, 1099 payments.

What recommendations would you make based on this data?

Your answer:

Without statistical testing, we can tentatively say from this data that Option B has a slightly higher conversion rate.