Python

**Question 1:** -

Write a program that takes a string as input, and counts the frequency of each word in the string, there might be repeated characters in the string. Your task is to find the highest frequency and returns the length of the highest-frequency word.

Ans : <https://github.com/jagannath20-dev/ineuron-Data-Science-Interview-Assessment/blob/main/Frequency_of_a_String.py>

https://github.com/jagannath20-dev/ineuron-Data-Science-Interview-Assessment/blob/main/python\_Question\_1.ipynb

**Question 2: -**

Consider a string to be *valid* if all characters of the string appear the same number of times. It is also *valid* if he can remove just one character at the index in the string, and the remaining characters will occur the same number of times. Given a string, determine if it is *valid*. If so, return YES , otherwise return NO .

Ans :

<https://github.com/jagannath20-dev/ineuron-Data-Science-Interview-Assessment/blob/main/validation_String.py>

https://github.com/jagannath20-dev/ineuron-Data-Science-Interview-Assessment/blob/main/python\_Question\_2.ipynb

**Question 3:**

Write a program, which would download the data from the provided link, and then read the data and convert that into properly structured data and return it in Excel format.

Ans. <https://github.com/jagannath20-dev/ineuron-Data-Science-Interview-Assessment/blob/main/python_Question_3.ipynb>

**Question 4:**

Write a program to download the data from the link given below and then read the data and convert the into the proper structure and return it as a CSV file.

Ans.

https://github.com/jagannath20-dev/ineuron-Data-Science-Interview-Assessment/blob/main/python\_Question\_4.ipynb

**Question 5:**

Write a program to download the data from the given API link and then extract the following data with proper formatting

Ans. <https://github.com/jagannath20-dev/ineuron-Data-Science-Interview-Assessment/blob/main/python_Question_5.ipynb>

**Question 6:**

Using the data from Question 3, write code to analyze the data and answer the following questions Note 1. Draw plots to demonstrate the analysis for the following questions for better visualizations. 2. Write code comments wherever required for code understanding

Ans. <https://github.com/jagannath20-dev/ineuron-Data-Science-Interview-Assessment/blob/main/python_Question_6.ipynb>

**Question 7:**

Using the data from Question 4, write code to analyze the data and answer the following questions Note -

1. Draw plots to demonstrate the analysis for the following questions for better visualizations 2. Write code comments wherever required for code understanding

Insights to be drawn - ● Get all the Earth meteorites that fell before the year 2000 ● Get all the earth meteorites co-ordinates who fell before the year 1970 ● Assuming that the mass of the earth meteorites was in kg, get all those whose mass was more than 10000kg

● Get all the Earth meteorites that fell before the year 2000

● Get all the earth meteorites co-ordinates who fell before the year 1970   
● Assuming that the mass of the earth meteorites was in kg, get all those whose mass was more  
 than 10000kg

Ans. <https://github.com/jagannath20-dev/ineuron-Data-Science-Interview-Assessment/blob/main/Python_Question_7.ipynb>

**Question 8 :**

Using the data from **Question 5,** write code the analyze the data and answer the following questions **Note -**

1. Draw plots to demonstrate the analysis for the following questions and better visualizations

**Insights to be drawn -**

● Get all the overall ratings for each season and using plots compare the ratings for all the seasons, like season 1 ratings, season 2, and so on.  
● Get all the episode names, whose average rating is more than 8 for every season  
● Get all the episode names that aired before May 2019  
● Get the episode name from each season with the highest and lowest rating  
● Get the summary for the most popular ( ratings ) episode in every season

Ans. <https://github.com/jagannath20-dev/ineuron-Data-Science-Interview-Assessment/blob/main/python_Question_8.ipynb>

**Question 9 -**    
Write a program to read the data from the following link, perform data analysis and answer the following questions   
**Insights to be drawn -**● Get all the cars and their types that do not qualify for clean alternative fuel vehicle  
● Get all TESLA cars with the model year, and model type made in Bothell City.  
● Get all the cars that have an electric range of more than 100, and were made after  
 2015  
● Draw plots to show the distribution between city and electric vehicle type

Ans. <https://github.com/jagannath20-dev/ineuron-Data-Science-Interview-Assessment/blob/main/python_Question_9.ipynb>

* **Question 10 -**    
  Write a program to count the number of verbs, nouns, pronouns, and adjectives in a given particular phrase or paragraph, and return their respective count as a dictionary.   
  Ans. https://github.com/jagannath20-dev/ineuron-Data-Science-Interview-Assessment/blob/main/Python\_Question\_10.ipynb

Statistics

### **Q-1.** A university wants to understand the relationship between the SAT scores of its applicants and their college GPA. They collect data on 500 students, including their SAT scores (out of 1600) and their college GPA (on a 4.0 scale). They find that the correlation coefficient between SAT scores and college GPA is 0.7. What does this correlation coefficient indicate about the relationship between SAT scores and college GPA?

Ans . In Reality, if the SAT score is high and their College GPA is high then the chances of admission into colleges is also high as well as if the SAT score is low and their College GPA is low then the chances of admission into colleges is also Low i.e if one variable is increases with respect to this another variable also increases their exists a Positive Correlation.

Similarly, If one variable decreases with respect to this variable another variable decreases their also exists a positive Correlation.

Given that

500 students among 1600 students, including their SAT scores (out of 1600) and their college GPA (on a 4.0 scale).

correlation coefficient between SAT scores and their GPA is 0.7

i.e their exists a strong positive correlation between SAT scores and their College GPA

.**Q-2.** Consider a dataset containing the heights (in centimeters) of 1000 individuals. The mean height is 170 cm with a standard deviation of 10 cm. The dataset is approximately normally distributed, and its skewness is approximately zero. Based on this information, answer the following questions:

a. What percentage of individuals in the dataset have heights between 160 cm and 180 cm?  
 b. If we randomly select 100 individuals from the dataset, what is the probability that their average height is greater than 175 cm?

c. Assuming the dataset follows a normal distribution, what is the z-score corresponding to a height of 185 cm?  
 d. We know that 5% of the dataset has heights below a certain value. What is the approximate height corresponding to this threshold?  
 e. Calculate the coefficient of variation (CV) for the dataset.  
 f. Calculate the skewness of the dataset and interpret the result.

Ans.

Given mean (μ) = 170 cms

Standard deviation (σ ) = 10 cms

Skewness = 0

1. Z1 - Score = (x − μ )/σ

Z1 - score = (160 - 170)/ 10

Z 1 -score = -1

Z 2 - score = (x − μ )/σ

= (180-170)/10

= + 1

From Z - Tables the value for -1 is 15.866, and the value for +1 is 84.13

Z Score = 84.13 - 15.87

= 68.26 %

Therefore the percentage of individuals in the dataset have heights between 160 cm and 180 cm is 68.26 %

1. Sample size (n) = 100

Z - Score = (x − μ )/(σ /√n)

= 175 - 170/(10/10)

Z - Score = 5

Since the Z - Score is quite large, the probability associated with it will be very close to 1.

Therefore the probability that the average height of a randomly selected sample of 100 individuals is greater than 175 cms is approximately 1.

1. Z - score = 185-170/10

= 15/10 = 1.5

Z - Score = 1.5

Therefore the z- score corresponding o a height of 185 is 1.5 cm

1. From the cumulative probability of a z- score calculator is - 1.645

(x − μ )/σ

X = μ + (z \* σ)

X = 170 + (-1.645 \* 10)

X = 170 - 16.45

X = 153.55

Therefore the approximate height corresponding to the threshold below which 5 % of the dataset falls is around 153.55 cms

1. Coefficient of variation = (standard deviation/mean)\*100 %

Given average height = 170

Standard deviation = 10

CV = (10/170)\* 100%

CV = 5. 88 %

Therefore the Coefficient of Variation is 5.88%

f .

Since the dataset is normally distributed and symmetrical, the mean and median are equal, and the skewness will be close to zero. In this case, the skewness is:

skewness = ( 3\* (mean - median))/standard deviation

Skewness = (3\*(170-170))/10

Skewness = 0

**Q-3.** Consider the ‘Blood Pressure Before’ and ‘Blood Pressure After’ columns from the data and calculate the following

Ans.

Blood Pressure Before Blood Pressure After

Mean 133.9100 128.3600

Standard Deviation 6.598278 6.888022

A group of 20 friends decide to play a game in which they each write a number between 1 and 20 on a slip of paper and put it into a hat. They then draw one slip of paper at random. What is the probability that the number on the slip of paper is a perfect square (i.e., 1, 4, 9, or 16)?

Ans. Number of favorable outcomes (perfect squares): 4 (1, 4, 9, 16)

Total number of possible outcomes = 20

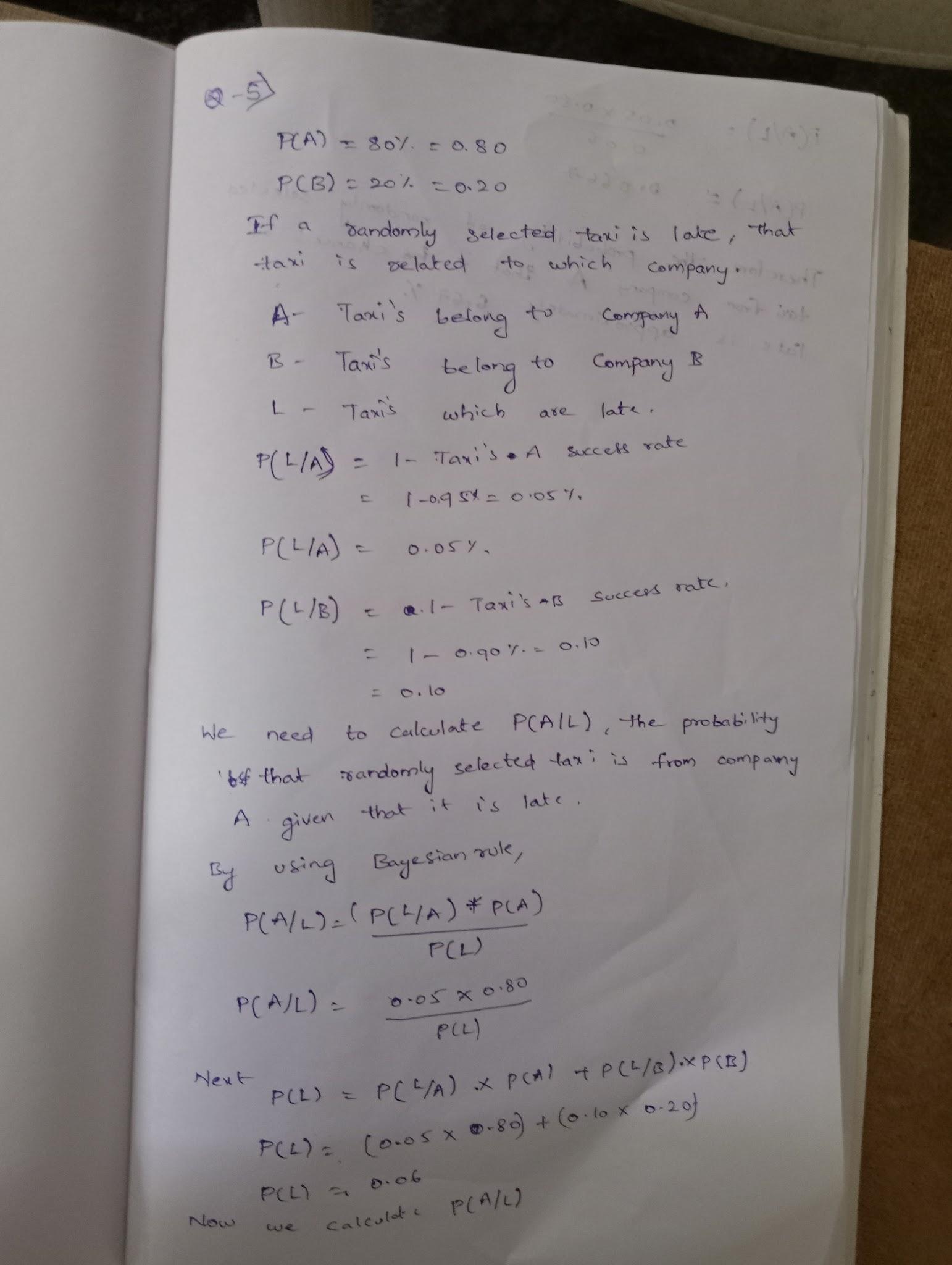
Therefore, the probability of drawing a perfect square number is :

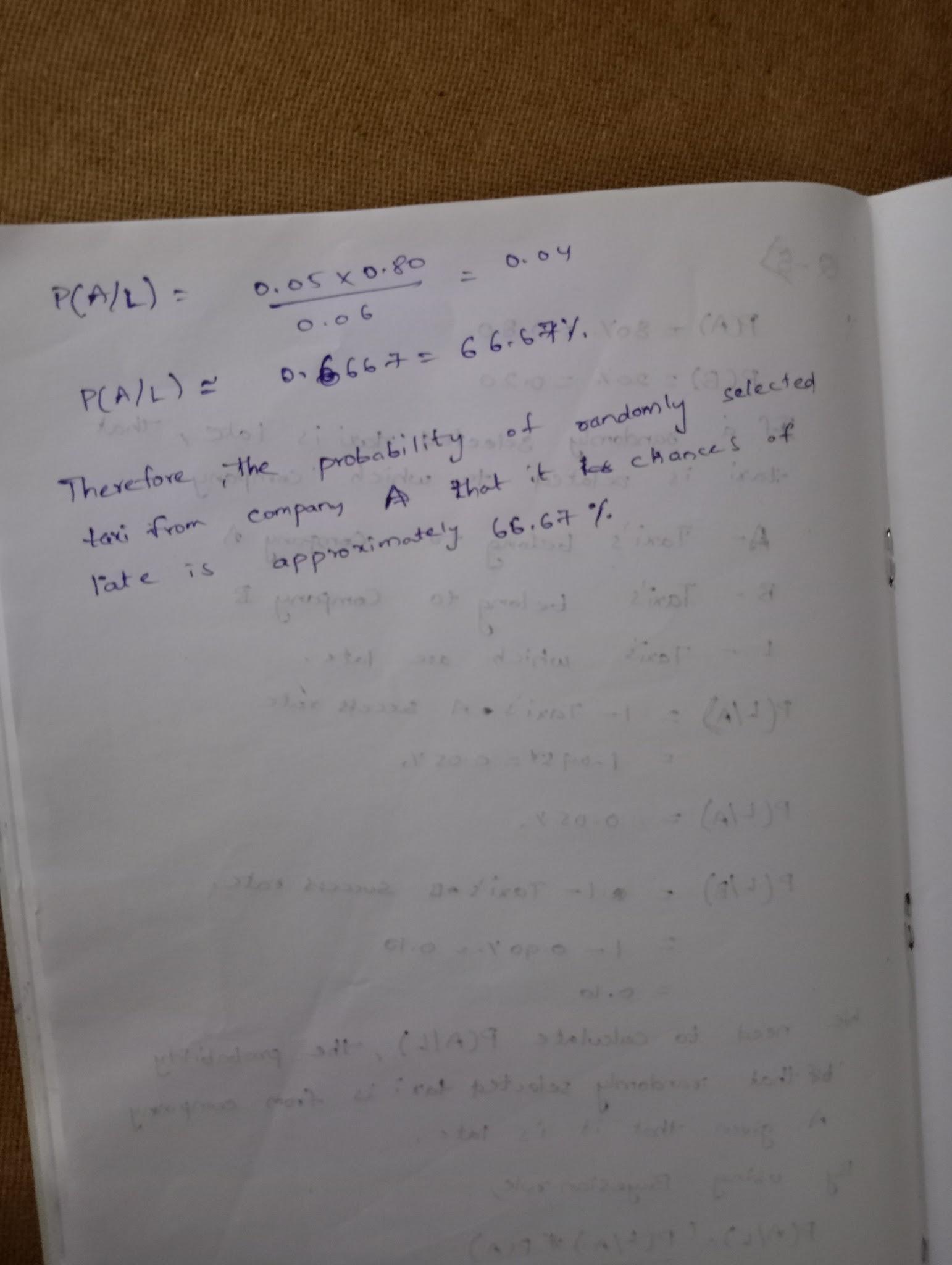
Probability = No of Favorable outcomes = 4

Probability = 4 / 20 = ⅕ = 0.2 = 20%

So, Probability of drawing a perfect square number is 20%.

**Q-5**. A certain city has two taxi companies: Company A has 80% of the taxis and Company B has 20% of the taxis. Company A's taxis have a 95% success rate for picking up passengers on time, while Company B's taxis have a 90% success rate. If a randomly selected taxi is late, what is the probability that it belongs to Company A?

Ans. 

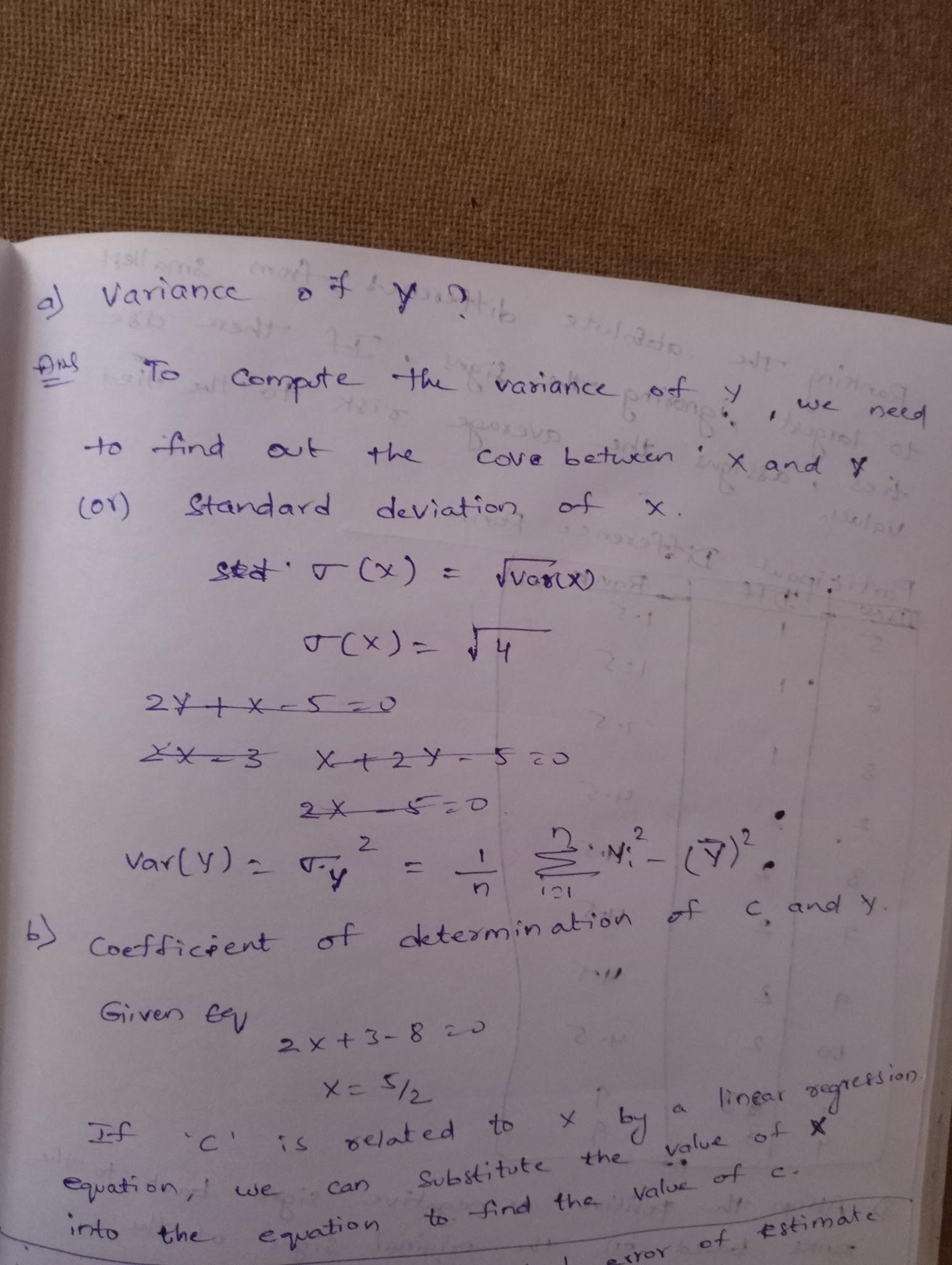
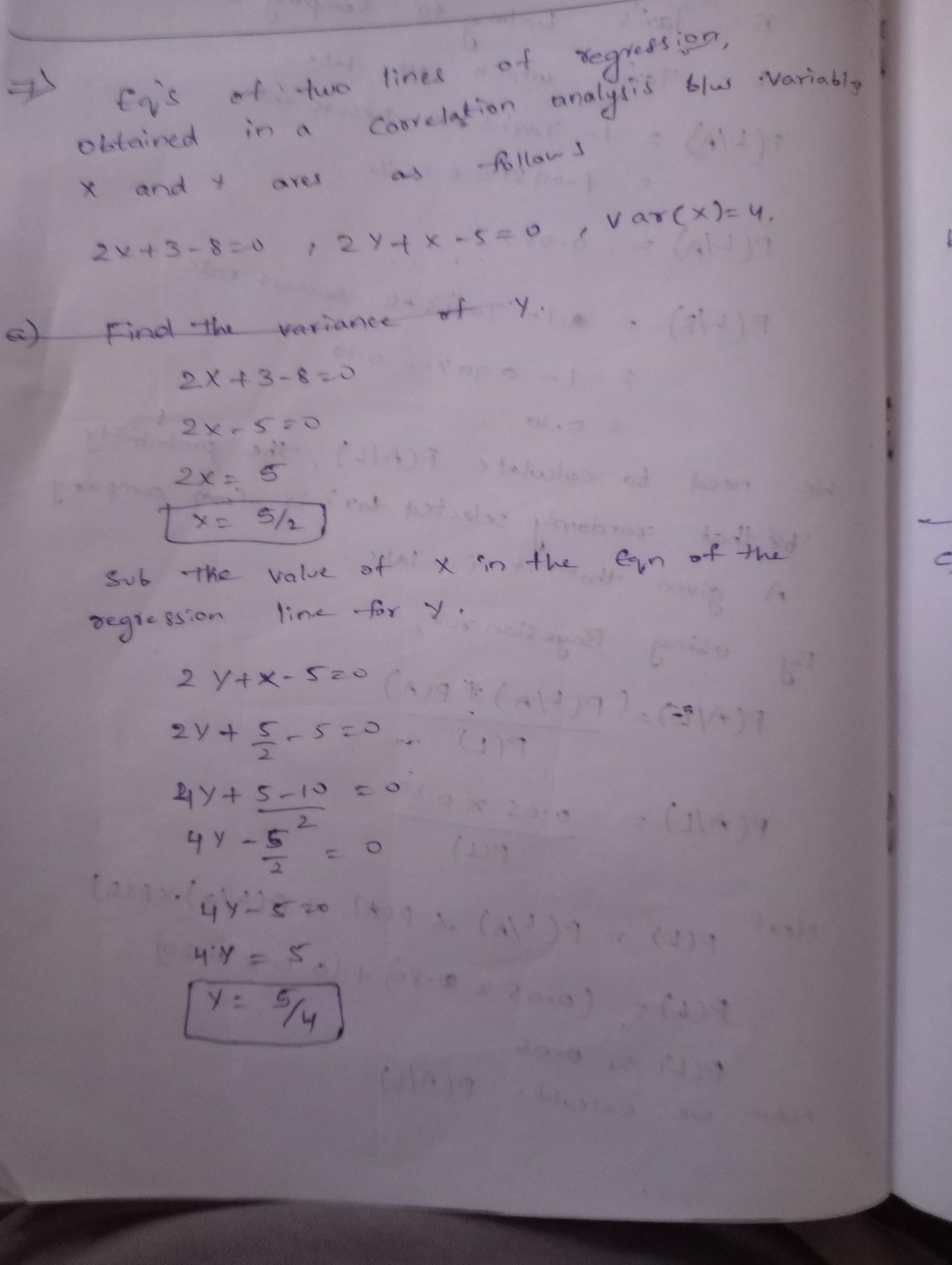
**Q-6.** A pharmaceutical company is developing a drug that is supposed to reduce blood pressure. They conduct a clinical trial with 100 patients and record their blood pressure before and after taking the drug. The company wants to know if the change in blood pressure follows a normal distribution.

Ans.

**Q-7.** The equations of two lines of regression, obtained in a correlation analysis between variables X and Y are as follows:

and. 2𝑋+3−8=02𝑌+𝑋−5=0Thevarianceof𝑋=4Findthe a. Variance of Y

b. Coefficient of determination of C and Y  
 c. Standard error of estimate of X on Y and of Y on X.



**Q-8.** The anxiety levels of 10 participants were measured before and after a new therapy. The scores are not normally distributed. Use the Wilcoxon signed-rank test to test whether the therapy had a significant effect on anxiety levels. The data is given below: Participant Before therapy After therapy Difference

Ans : w > 19

We reject the null hypothesis and conclude that the therapy had a significant effect on anxiety levels.

**Q-9**. Given the score of students in multiple exams

Test the hypothesis that the mean scores of all the students are the same. If not, name the student with the highest score.

Ans. v1 /v2 = 7.26

Here we reject the null Hypothesis

**Q-10.** A factory produces light bulbs, and the probability of a bulb being defective is 0.05. The factory produces a large batch of 500 light bulbs.

a. What is the probability that exactly 20 bulbs are defective?  
 b. What is the probability that at least 10 bulbs are defective?  
 c. What is the probability that at max 15 bulbs are defective?  
 d. On average, how many defective bulbs would you expect in a batch of 500?

a. Probability that exactly 20 bulbs are defective: 0.051616192536641056

b. Probability that at least 10 bulbs are defective: 0.9998316463654902

c. Probability that at most 15 bulbs are defective: 0.01985837716300623

d. Average number of defective bulbs in a batch of 500: 25.0

**Q-12.** A pharmaceutical company develops a new drug and wants to compare its effectiveness against a standard drug for treating a particular condition. They conduct a study with two groups: Group A receives the new drug, and Group B receives the standard drug. The company measures the improvement in a specific symptom for both groups after a 4-week treatment period.

a. The company collects data from 30 patients in each group and calculates the mean improvement score and the standard deviation of improvement for each group. The mean improvement score for Group A is 2.5 with a standard deviation of 0.8, while the mean improvement score for Group B is 2.2 with a standard deviation of 0.6. Conduct a t-test to determine if there is a significant difference in the mean improvement scores between the two groups. Use a significance level of 0.05.

b. Based on the t-test results, state whether the null hypothesis should be rejected or not. Provide a conclusion in the context of the study.

Ans.

Fail to reject the null hypothesis. There is no significant difference in the mean improvement scores between Group A and Group B.

### 

### 

### 

### 

### 