# **Python**

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. Note - You have to write at least 2 additional test cases in which your program will run successfully and provide an explanation for the same. Answer. 1 Python -> [Interview\_Assignment\_6\_June\_Deadline/Python\_Ques1.ipynb at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline · GitHub](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Python_Ques1.ipynb)
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 . Answer. 2 Python -> [Interview\_Assignment\_6\_June\_Deadline/Python\_Ques2.ipynb at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline (github.com)](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Python_Ques2.ipynb)
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. Note - Write comments wherever necessary explaining the code written. Link Answer. 3 Python -> [Interview\_Assignment\_6\_June\_Deadline/Python\_Ques3.ipynb at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline (github.com)](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Python_Ques3.ipynb)
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 Answer-4 Python -> [Interview\_Assignment\_6\_June\_Deadline/Python\_Ques\_4.ipynb at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline (github.com)](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Python_Ques_4.ipynb)
5. Write a program to download the data from the given API link and then extract the following data with proper formatting Link – Answer-5 Python -> [Interview\_Assignment\_6\_June\_Deadline/Python\_Ques\_5.ipynb at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline · GitHub](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Python_Ques_5.ipynb)
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 Insights to be drawn  Answer-6 Python -> [Interview\_Assignment\_6\_June\_Deadline/Python\_Ques6.ipynb at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline (github.com)](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Python_Ques6.ipynb)
7. Using the data from Question 4, write code to analyze the data and answer the following questions Note Answer-7 Python -> [Interview\_Assignment\_6\_June\_Deadline/Python\_Ques7.ipynb at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline (github.com)](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Python_Ques7.ipynb)
8. Using the data from Question 5, write code the analyze the data and answer the following questions Note Answer-8 Python-> [Interview\_Assignment\_6\_June\_Deadline/Python\_Ques8.ipynb at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline (github.com)](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Python_Ques8.ipynb)
9. Write a program to read the data from the following link, perform data analysis and answer the following ques Answer-9 Python-> [Interview\_Assignment\_6\_June\_Deadline/Python\_Ques9.ipynb at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline (github.com)](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Python_Ques9.ipynb)

# **statistics**

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? Answer 1. Correlation coefficient of 0.7 indicates a strong positive relation between Students SAT Scores and college GPA. Which means if students SAT scores is high then college GPA is also good and vice-versa
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:
   1. What percentage of individuals in the dataset have heights between 160 cm and 180 cm?

Answer a)

N=1000

mu=170

sd=10

z1=(160-mu)/sd

z2=(180-mu)/sd

area\_z1=0.1587

area\_z2=0.8413

req\_percent=area\_z2-area\_z1

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

* 1. If we randomly select 100 individuals from the dataset, what is the probability that their average height is greater than 175 cm?

Answer b) n=100

prob sample\_mean>175?

as cnetral limit theoram

series consist of mean of various samples from a distribution has mean equal to mean of original distribution and sample sd = pop sd/ sample size\*\*(1/2)

# z score of 175 for this new series

z\_score=(175-mu)/(sd/10)

**prob of mean greater then 175 is almost zero**

* 1. Assuming the dataset follows a normal distribution, what is the z-score corresponding to a height of 185 cm?

z\_score=(185-mu)/(10)

**hence z-score corresponding to a height of 185 cm is 1.5**

* 1. We know that 5% of the dataset has heights below a certain value. What is the approximate height corresponding to this threshold?

**Answer d)**

z score for prob of 0.05 from z score table is "-1.6207"

-1.6207=(x-mu)/sd

-1.6207\*10=x-mu

x=-16.207+170

x=153.793

* 1. Calculate the coefficient of variation (CV) for the dataset.

Answer e)

CV=(sd/mu)\*100

**Coefficient of variation (CV) is 5.88235294117647 %**

* 1. Calculate the skewness of the dataset and interpret the result

**Answer f)**

Skewness = (3 \* (mu - median)) / sd

since data is normally distributed so mean=median=mode

so Skewness=0

hence data is symmetrical with zero skewness

1. Consider the ‘Blood Pressure Before’ and ‘Blood Pressure After’ columns from the data and calculate the following

Answer 3) [Interview\_Assignment\_6\_June\_Deadline/Stats\_Ques3.ipynb at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline (github.com)](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Stats_Ques3.ipynb)

1. 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-4) perfect square between 1 and 20 are:- 1,4,9,16 total numner of outcomes=20 prob=4/20=2/10=0.2
2. 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?

Answer 5) According to Bayes' theorem:

P(A|L) = (P(L|A) \* P(A)) / P(L)

P(L/A)=1-0.95=0.05 P(A)=0.80 #acc to law of total prob

P(L) = P(L|A) \* P(A) + P(L|B) \* P(B) P(L)=0.050.80 + 0.100.20=0.06

P(A|L) = (P(L|A) \* P(A)) / P(L)=(0.05\*0.80) / 0.06= 2/3 P(A|L) =2/3

1. 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.

Answer-6) [Interview\_Assignment\_6\_June\_Deadline/Stats\_Ques6.ipynb at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline (github.com)](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Stats_Ques6.ipynb)

1. The equations of two lines of regression, obtained in a correlation analysis between variables X and Y are as follows: and . 2𝑋 + 3 − 8 = 0 2𝑌 + 𝑋 − 5 = 0 The variance of 𝑋 = 4 Find the 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

## Answer 7)

a) 2X+3Y-8=0---eq1 2Y+X-5=0---eq2 var(X)=4

lets consider eq1 as X on Y and find bxy

2X+3Y-8=0 x=(-3/2)y+4

bxy=-3/2

lets consider eq1 as Y on X and find byx 2Y+X-5=0 Y=(-1/2)x+(5/2) byx=(-1/2)

bxy\*byx=0.75 which is less than 1. Hence our assumption is correct

a) VarY=VarX/bxy= 4/(-3/2)=2.666

b) Coefficient of determination of C and Y

we have to find R square

R square=bxy\*byx=0.75

c) Standard error of estimate of X on Y and of Y on X

Standard Error of Estimate (X on Y) = √(SSR / (n - 2))

For estimating Y based on X (Y on X): Standard Error of Estimate (Y on X) = √(SSR / (n - 2))

1. 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.

Answer. 8 Python - > [Interview\_Assignment\_6\_June\_Deadline/Stats\_Ques\_8.pdf at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline (github.com)](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Stats_Ques_8.pdf)

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

Answer 9) [Interview\_Assignment\_6\_June\_Deadline/Stats\_Qus9.ipynb at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline (github.com)](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Stats_Qus9.ipynb)

1. 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? Answer 10) [Interview\_Assignment\_6\_June\_Deadline/Stats\_Ques10.ipynb at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline (github.com)](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Stats_Ques10.ipynb)
2. Given the data of a feature contributing to different classes <https://drive.google.com/file/d/1mCjtYHiX--mMUjicuaP2gH3k-SnFxt8Y/view?usp> =share\_ a. Check whether the distribution of all the classes are the same or not. b. Check for the equality of variance/ c. Which amount LDA and QDA would perform better on this data for classification and why. d. Check the equality of mean for between all the classes.

Answer 11) [Interview\_Assignment\_6\_June\_Deadline/Stats\_Ques11.ipynb at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline (github.com)](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Stats_Ques11.ipynb)

1. 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.

Answer 12) [Interview\_Assignment\_6\_June\_Deadline/Stats\_Ques12.ipynb at main · ssharma2323/Interview\_Assignment\_6\_June\_Deadline (github.com)](https://github.com/ssharma2323/Interview_Assignment_6_June_Deadline/blob/main/Stats_Ques12.ipynb)