**Placement Assignment**

**Python:**

**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?**

**Answer :** The correlation coefficient range from -1 to 1 , a value close to 1 means that they correlation between them is positive which in turn mean that as one goes up the other too goes up. So SAT score and CGPA are positively correlated.

**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**

**Answer:**

Mean Hight = 170cm

Standard Deviation(STD) = 10cm

**a** : As this is a normally distributed and we know the 68% of data lies between one STD,so 68% of individuals in the dataset have heights between 160cm to 180cm.

**b :**  the probability of getting a average hight of 175cm is very close to 0.

**c :** The Z-score fro 185cm is 1.5.

**d :** The approx height corresponding to this threshold is 153.5cm

**e :** Coefficient of variation is 100

**f** : skewness = (3\*(mean-median))/std which results to 0. This mean the data is symmetric and follows a normal distribution. And it is not skewed to the left or the right implies the spread is even.

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

**data and calculate the following**

**https://drive.google.com/file/d/1mCjtYHiX--mMUjicuaP2gH3k-SnFxt8Y/view?usp=share\_**

**a. Measure the dispersion in both and interpret the results.**

**b. Calculate mean and 5% confidence interval and plot it in a graph**

**c. Calculate the Mean absolute deviation and Standard deviation and interpret**

**the results.**

**d. Calculate the correlation coefficient and check the significance of it at 1% level**

**of significance.**

**Answer: Github link**

**Q-4. 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)?**

**Answer:** The probability is 4/20 = 0.2

**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?**

**Answer:** Let the total taxi be 100

Total Company A taxi = 80

Total Company B taxi = 20

|  | Company A | Company B | Total |
| --- | --- | --- | --- |
| Succes | 76 | 18 | 94 |
| Fail | 4 | 2 | 6 |
|  | 80 | 20 |  |

So the probability of that taxi being from Company A is 4/6 = 0.66 , that is 66%

**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.**

[**https://drive.google.com/file/d/1mCjtYHiX--mMUjicuaP2gH3k-SnFxt8Y/view?usp=share\_**](https://drive.google.com/file/d/1mCjtYHiX--mMUjicuaP2gH3k-SnFxt8Y/view?usp=share_)

**Answer:** Its not a perfect normal distribution.

{githublink}

**Q-7. The equations of two lines of regression, obtained in a correlation analysis**

**between variables X and Y are as follows:**

**2𝑋 + 3Y − 8 = 0 and 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:** Don't know

**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**

**Answer:** to be done

**Q-9. Given the score of students in multiple exams,est the hypothesis that the mean scores of all the students are the same. If not, name the student with the highest score.**

**Answer:** The mean score is not same, highest scores is Karan

**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?**

**Answer: git hub**

**Q-11. Given the data of a feature contributing to different classes**

**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 :** Wrong data provided

**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**

**Answer:**

H0: There is no significant difference in the mean improvement scores between Group A and Group B.

H1: There is a significant difference in the mean improvement scores between Group A and Group B.

T-test = (mean1-mean2)/squr( std1^2/n + std2^2/n))

Mean1 = 2.5

Std1 = 0.8

Mean2 = 2.2

Std2 = 0.6

Putting in the formula we get 1.643

significance level = 0.05

Degree of freedom = 58

T-value = -+ 2.001

As the 1.643 < 2.001 we fail to reject the H0 so no significant difference .