Q1) Calculate Skewness, Kurtosis & draw inferences on the following data

a. Cars speed and distance

****

Ans :

Skewness of Speed : -0.8448909 (moderately skewed)

Skewness of dist : 1.217917 (highly skewed)

Kurtosis of Speed : 2.991396 ( Approximately Mesokurtic)

Kurtosis of dist : 4.816933 (Leptokurtic)

**Understanding :**

**Skewness:**

If skewness is less than −1 or greater than +1, the distribution is **highly skewed**.

If skewness is between −1 and −½ or between +½ and +1, the distribution is **moderately skewed**.

If skewness is between −½ and +½, the distribution is **approximately symmetric**.

**Kurtosis :**

If the distribution is normal then it has kurtosis is equal to 3. This kind of Kurtosis is called as **mesokurtic**.

If the distribution is normal then it has kurtosis is less than 3. This kind of Kurtosis is called as **platykurtic**. Its tails are shorter and thinner, and often its central peak is lower and broader

If the distribution is normal then it has kurtosis is greater than 3. This kind of Kurtosis is called as **leptokurtic**. Its tails are longer and fatter, and often its central peak is higher and sharper.

b. Top Speed (SP) and Weight (WT)

****

Skewness of SP: -0.407692(approximately symmetric)

Skewness of WT : -1.287359 (highly skewed)

Kurtosis of SP : 2.086731 ( Approximately Mesokurtic)

Kurtosis of WT : 3.818812 (Leptokurtic)

Q2) Draw inferences about the following boxplot & histogram



**Ans :** By looking the Histogram, All chicks weight got increases to 100. After when they were waiting for high Chicks weights then number of chicks got gradually decreased.



**Ans :** By looking the Bar Plot, It has the outliers distributed after 4th Quartile (Q4). Here Median is not exactly in the center. Frequency of the Q 1 & Q2 is less and Frequency of Q3 and Q4 were more.

**Q3)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

**Ans:**

N=300000 , n= 2000 , mean = 200 , sd = 30

They have given for standard deviation for sample. So I’m considering it is the T distribution.

Standard error = sd / sqrt(n) = 30/SQRT(2000) = 30/44.72 = 0.67

Mean of sample = 200

Control Interval = [ Xbar + or – T value (Standard error) ]

T value at 98% = 2.328 \* 30 /SQRT(2000) = 1.56

**98 % of Control Interval = [201.56 , 198.43]**

T value at 96% = 2.055 \* 30 /SQRT(2000) = 1.37

**96 % of Control Interval = [201.37 , 198.63]**

T value at 94% = 1.882 \* 30 /SQRT(2000) = 1.26

**94 % of Control Interval = [201.26 , 198.73]**

**Q4)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median, variance, standard deviation.

**Ans :** Mean : 41

Median : 40.5

Variance : 25.529

Standard deviation : 5.052

1. What can we say about the student marks?

**Ans :** The student marks of mean and medians are very closure ( 41 ,40.5). So most of the data dispersed in between the 34 to 42. There were no outliers’ marks in the exam. It seems either subject might be tough or exam might be tough, So that students scored less.

Q5) What is the nature of skewness when mean, median of data are equal?

**Ans :** When mean & median of the data are equal then data distributed in the symmetric and Normal distributed.

Q6) What is the nature of skewness when mean > median?

**Ans :** Nature of skewness is Positive when mean is greater than median.

Q7) What is the nature of skewness when median > mean?

**Ans :** Nature of skewness is Negative when mean is less than median.

Q8) What does positive kurtosis value indicates for a data?

**Ans** : A distribution with positive kurtosis value indicates that the distribution has heavier tail and sharper peak than normal distribution.

Q9) What does negative kurtosis value indicates for a data?

**Ans** : A distribution with negative kurtosis value indicates that the distribution has lighter tail and flatter peak than normal distribution.

Q10) Answer the below questions using the below boxplot visualization.



Q1 = 10 , Q2 = 15 , Q3 = 18

What can we say about the distribution of the data?

**Ans :** Difference is not equal in both sides that means distribution of the data are **not normal distribution.**

What is nature of skewness of the data?

**Ans :** Q2-Q1 = 15 – 10 = 5

Q3 –Q2 = 18 – 15 = 3

Both sides are not equal, Data distributed towards left side negative skewness.

What will be the IQR of the data (approximately)?

**Ans :**  IQR = (Q3 – Q1) = (18 – 10) = 8

Q11) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

Ans : By looking the BoxPlot1 we are considering the Quartile ranges.

Q1 = 250

Q2 = 263

Q3 = 275

Q4 = 285

IQR = Q3-Q1 = 275 – 250 = 25

Median = Q2 = 263

Boxplot2 :

Median = Q2 = 263

Q1 = 225

Q2 = 263

Q3 =300

Q4 = 335

IQR = Q3-Q1 = 300 – 225 = 75

Both boxplots median are same and data distribution is also same. Here boxplot2 is 3 times of Boxplot1.

Q12)



Answer the following three questions based on the boxplot above.

Q1 = 5 , Q2 = 7, Q3 = 11.5

1. What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.

**Ans :** IQR = (Q3 – Q1) = > 11.5 – 5 = 6.5

1. What can we say about the skewness of this dataset?

**Ans :**  Data distributed towards right then it is called as right distribution Or Positive distribution.

1. If it was found that the data point with the value 25 is actually 2.5, how would the new boxplot be affected?

**Ans** : If 25 is there at 2.5 point then data will be distributed in Normal.

Q1 = 2.5 Q2 = 7 and Q3 = 11.5

Q2 - Q1 = 7 – 2.5 = 4.5

Q3 – Q2 = 11.5 – 7 = 4.5

Data are in **Normal Distribution.**

Q13)



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Ans :** Mode of the Y values lies in the 4 to 8 and frequency is 21. The mode is lies on 21.

1. Comment on the skewness of the dataset.

**Ans :** Data distributed in the right side then it is called as Positive Skewness.

1. Suppose that the above histogram and the boxplot in question 2 are plotted for the same dataset. Explain how these graphs complement each other in providing information about any dataset.

**Ans : Histogram :** If we compare both Histogram (Question 2 and Q13 (iii) ). When they were waiting for high Chicks weights then number of chicks got gradually decreased but here numbers of chicks were not gradually decreased but decreased.

**Boxplot :** It has the outliers distributed after 4th Quartile (Q4). Here Median is not exactly in the center. Frequency of the Q 1 & Q2 is less and Frequency of Q3 and Q4 were more. When compare to above dataset it is also looking same.