

Q1) Identify the Data type for the Following:

Activity	Data Type
Number of beatings from Wife	interval
Results of rolling a dice	ratio
Weight of a person	ratio
Weight of Gold	interval
Distance between two places	interval
Length of a leaf	interval
Dog's weight	ratio
Blue Color	nominal
Number of kids	interval
Number of tickets in Indian railways	ratio
Number of times married	interval
Gender (Male or Female)	nominal

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

Data	Data Type
Gender	nominal
High School Class Ranking	ordinal
Celsius Temperature	ratio
Weight	Ratio/ordinal
Hair Color	nominal
Socioeconomic Status	ordinal
Fahrenheit Temperature	ratio
Height	ordinal
Type of living accommodation	ordinal
Level of Agreement	ordinal
IQ(Intelligence Scale)	ratio
Sales Figures	ratio
Blood Group	nominal
Time Of Day	interval
Time on a Clock with Hands	ratio
Number of Children	nominal
Religious Preference	ordinal

Barometer Pressure	ratio
SAT Scores	ratio
Years of Education	ordinal

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

Ans: **Hhh,hht,hth,thh,tth,tht,htt,ttt : 3/8**

Q4) Two Dice are rolled, find the probability that sum is

- a) Equal to 1: **0/36**
- b) Less than or equal to 4: **(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)=6/36=1/6**
- c) Sum is divisible by 2 and 3: **(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)=6/36=1/6**

Q5) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

Ans: **${}^5C_2/{}^7C_2=10/21$**

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

CHILD	Candies count	Probability
A	1	0.015
B	4	0.20
C	3	0.65
D	5	0.005
E	6	0.01
F	2	0.120

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

Ans: Expected value= summation(probability*value)

$$=(0.015*1)+(0.20*4)+(0.65*3)+(0.005*5)+(0.01*6)+(0.120*2)=3.09$$

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

- For Points,Score,Weigh>
Find Mean, Median, Mode, Variance, Standard Deviation, and Range
and also Comment about the values/ Draw some inferences.

Use Q7.csv file

Q8) Calculate Expected Value for the problem below

a) The weights (X) of patients at a clinic (in pounds), are
108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Ans:

$$=(1/9*108)+(1/9*110)+(1/9*123)+(1/9*134)+(1/9*135)+(1/9*145)+(1/9*167)+(1/9*187)+(1/9*199)$$

$$=(1/9)(1308)=145.33$$

Q9) Calculate Skewness, Kurtosis & draw inferences on the following data

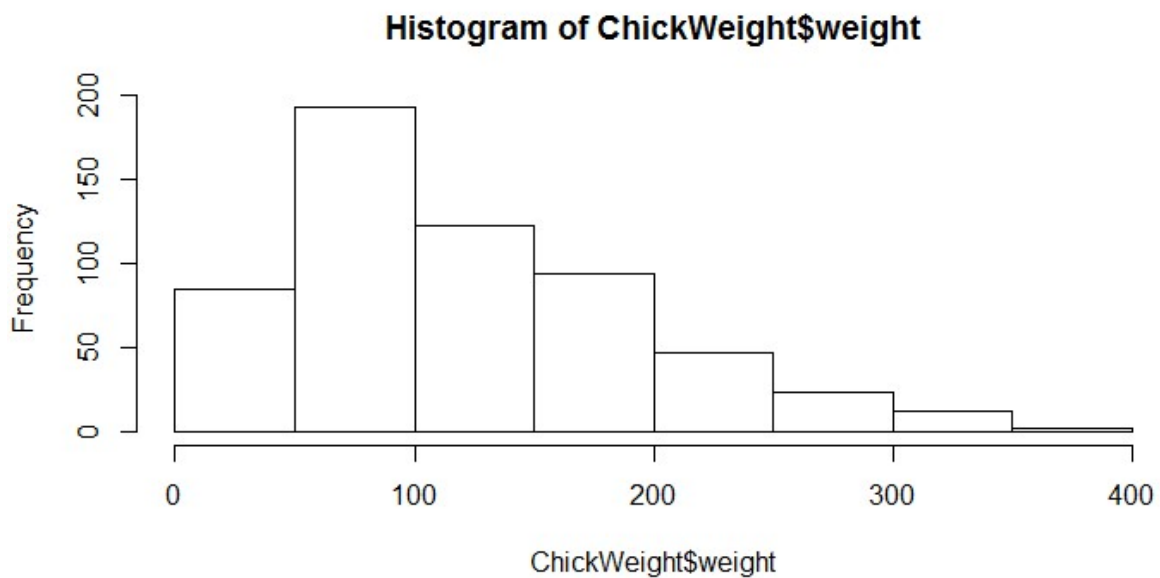
Cars speed and distance

Use Q9_a.csv

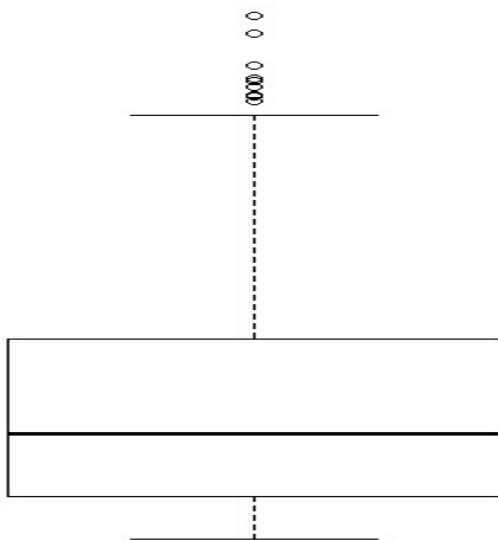
SP and Weight(WT)

Use Q9_b.csv

Q10) Draw inferences about the following boxplot & histogram



Ans: It is +ve skewness and right skewness distribution and 50-100 having more frequency. Since it is RSD the mean is greater than median



Ans: It has 7 outliers as the Q1 is less than Q2 then it is positively skewness and it is RSD

Q11) 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?

Q12) 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.
- 2) What can we say about the student marks?

Ans: 1.Mean= 41; median= 40.5 ; variance = 25.529; stand. Devation= 5.05

2. marks of the students range from 35.95-46.05

Q13) What is the nature of skewness when mean, median of data are equal?

Ans: Normal distribution

Q14) What is the nature of skewness when mean > median ?

Ans: right skewed distribution

Q15) What is the nature of skewness when median > mean?

Ans: Left skewed distribution

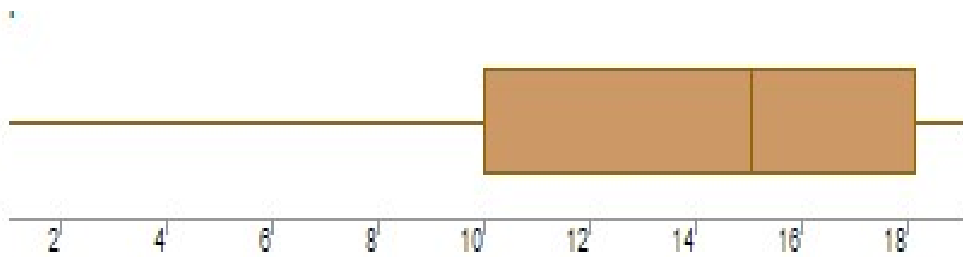
Q16) What does positive kurtosis value indicates for a data ?

Ans: It has high peaks right skewed distribution

Q17) What does negative kurtosis value indicates for a data?

Ans: It has data distribution as wide not peak Left skewed distribution

Q18) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

Ans: The median lies closer to the third quartile and the whisker at the upper end is shorter than it is negative skew(left skew)

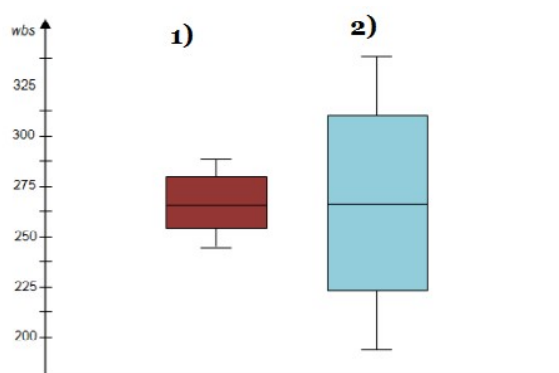
What is nature of skewness of the data?

Ans: Left skew

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

Ans: $18-10=8$

Q19) Comment on the below Boxplot visualizations?



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

Ans: The dispersion of the data is less in plot 1 than in plot 2. The length of whiskers gives an idea of the overall spread of the data. The extreme values give the range of distribution. Hence plot 1 has a longer range than plot 2.

Q 20) Calculate probability from the given dataset for the below cases

Data _set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars\$MPG

- a. $P(\text{MPG} > 38)$
- b. $P(\text{MPG} < 40)$
- c. $P(20 < \text{MPG} < 50)$

Q 21) Check whether the data follows normal distribution

- a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

- b) Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Q 22) Calculate the Z scores of 90% confidence interval, 94% confidence interval, 60% confidence interval

For 90%

$$A(L) = 1 + \text{confidence interval} / 2$$

$$= 1 + 0.90 / 2 = 0.950$$

$$\text{From t table } 0.05 + 1.6 = 1.65$$

For 94%

$$A(L) = 1 + \text{confidence interval} / 2$$

$$= 1 + 0.94 / 2 = 0.970$$

$$\text{From t table } 0.09 + 1.8 = 1.89$$

For 60%

$$A(L)=1+\text{confidence interval}/2$$

$$=1+0.60/2=0.80$$

$$\text{From t table } 0.8+0.05 = 0.85$$

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

95%:

$$\text{Alpha}=0.05, n-1=25-1=24=\text{df}$$

$$T(\text{alpha}/2)=0.025$$

From t table df=24, alpha=0.025 is 2.064

96%

$$\text{Alpha}=0.04, n-1=25-1=24=\text{df}$$

$$T(\text{alpha}/2)=0.02$$

From t table df=24, alpha=0.02 is 2.164

99%

$$\text{Alpha}=0.001, n-1=25-1=24=\text{df}$$

$$T(\text{alpha}/2)=0.005$$

From t table df=24, alpha=0.005 is 2.797

Q 24) A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode → pt(tscore,df)

df → degrees of freedom

Ans:

mean:270

random sample:18

sample mean:260

standard deviation:90

$t = ((260 - 270) / (90 / \sqrt{18}))$

t= 0.4714