

Q1) Identify the Data type for the Following:

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

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	quantitative
Weight	quantitative
Hair Color	nominal
Socioeconomic Status	ordinal
Fahrenheit Temperature	quantitative
Height	quantitative
Type of living accommodation	nominal
Level of Agreement	ordinal
IQ(Intelligence Scale)	quantitative
Sales Figures	ordinal
Blood Group	nominal
Time Of Day	quantitative
Time on a Clock with Hands	quantitative
Number of Children	discrete
Religious Preference	nominal

Barometer Pressure	quantitative
SAT Scores	quantitative
Years of Education	ordinal

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

SOLUTION – Probability of getting two heads and one tails in toss is $\frac{3}{8}$

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

- a) Equal to 1
- b) Less than or equal to 4
- c) Sum is divisible by 2 and 3

SOLUTION – Total possible outcome = 36

1) sum equal to 1 is 0

2) sum less than or equal to 4 is $\frac{6}{36}$

3) sum is divisible by 2 and 3 is $\frac{6}{36}$

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?

SOLUTION –

Total no. of balls = 7

No. of ways of drawing 2 balls out of 7 = ${}^7C_2 = 21$

No. of ball other than blue = 5

No of ways of drawing 2 balls out of 5 = ${}^5C_2 = 10$

REQUIRED PROBABILITY = $\frac{10}{21}$

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

SOLUTION

Expected no. of candies for randomly selected child =

$$1*0.015 + 4*0.20 + 3*0.65 + 5*0.005 + 6*0.01 + 2*0.12 = 3.09$$

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?

SOLUTION -

Probability of selecting each patient = $\frac{1}{9}$

$$\begin{aligned}\text{Expected value} &= 1/9 (108+110+123+134+135+145+167+187+199) \\ &= 1/9(1308) \\ &= 145.33\end{aligned}$$

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

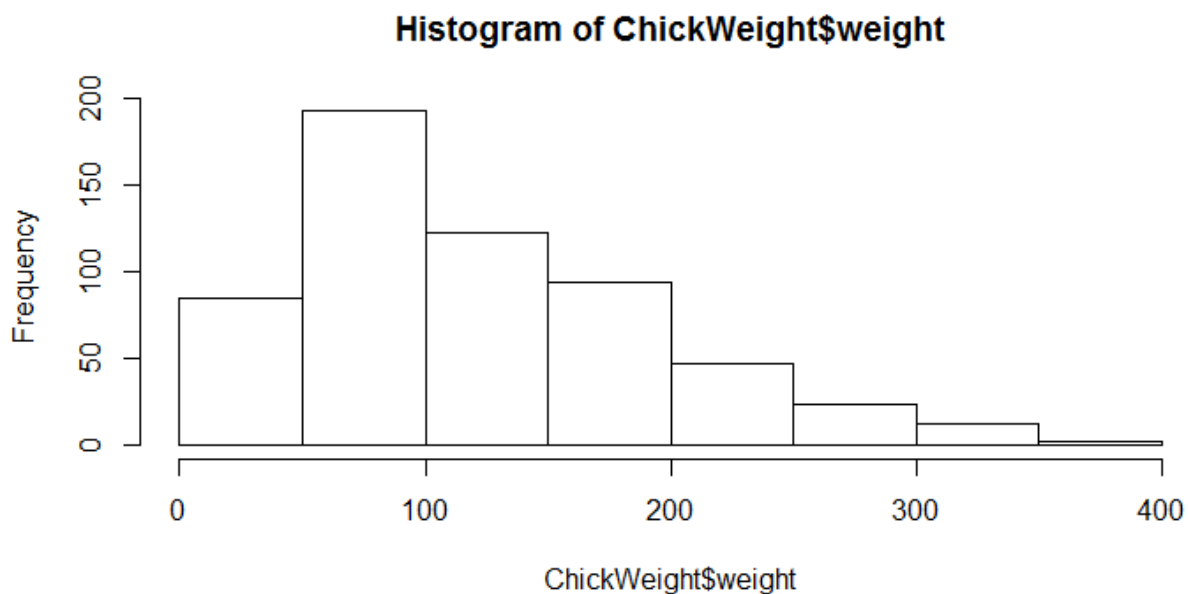
Cars speed and distance

Use Q9_a.csv

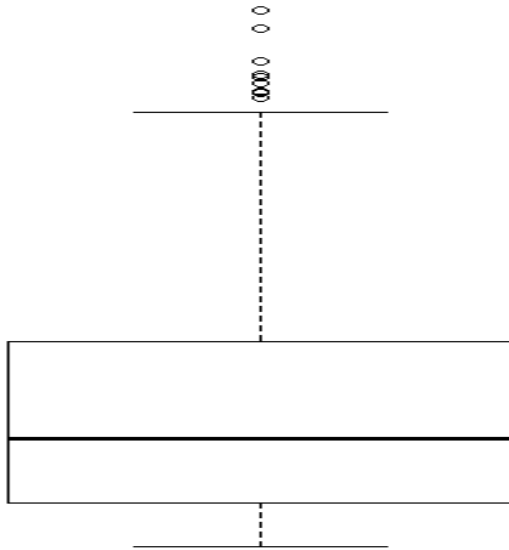
SP and Weight(WT)

Use Q9_b.csv

Q10) Draw inferences about the following boxplot & histogram



Solution – expected value - 46.4



Solution –Median is less than mean right skewed

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?

Solution t- distribution of 94% (198.73,201.27)

96% (198.61,201.39)

98% (198.43,201.57)

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?

SOLUTION

MEAN -41

MEDIAN -40.5

VARIANCE – 25.5

STD – 5.05

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

SOLUTION – ZERO SKEW

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

SOLUTION – POSITIVELY SKEWED

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

SOLUTION – NEGATIVELY SKEWED

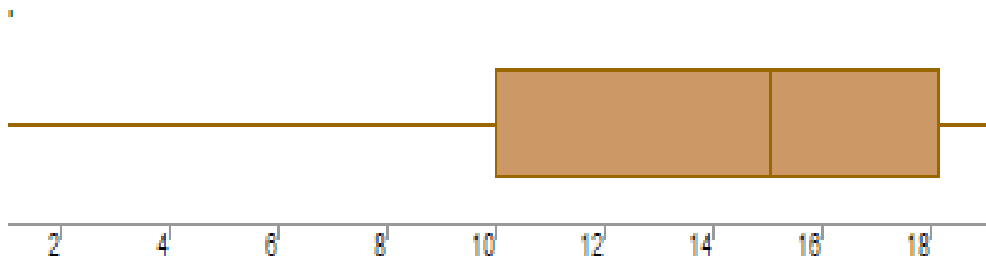
Q16) What does positive kurtosis value indicates for a data ?

SOLUTION – Distribution is peaked

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

SOLUTION – DISTRIBUTION HAS LIGHTER TAILS THAN NORMAL DISTRIBUTION

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

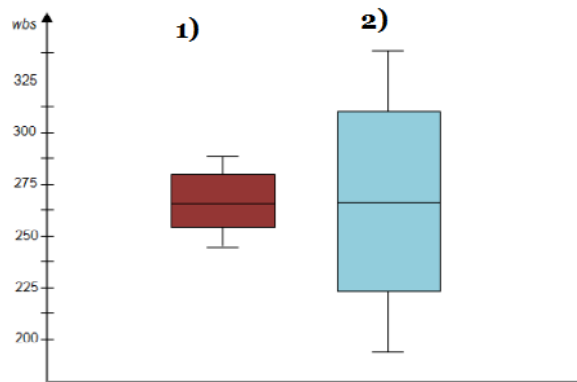


What can we say about the distribution of the data?

What is nature of skewness of the data?

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

Q19) Comment on the below Boxplot visualizations?



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

Solution – z score at 90% - 1.645

Z score at 94% - 1.555

Z score at 60% - .253

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

SOLUTION – t score for 95% - 2.064

T score for 96% - 2.171

T score for 99% - 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