

Q1) Identify the Data type for the Following:

Activity	Data Type
Number of beatings from Wife	Discrete
Results of rolling a dice	Discrete
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	Discrete
Number of kids	Discrete
Number of tickets in Indian railways	Discrete
Number of times married	Discrete
Gender (Male or Female)	Discrete

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	Interval
Weight	Ratio
Hair Color	Nominal
Socioeconomic Status	Ordinal
Fahrenheit Temperature	Interval
Height	Ratio
Type of living accommodation	Ordinal
Level of Agreement	Ordinal
IQ(Intelligence Scale)	Ordinal
Sales Figures	Ratio
Blood Group	Nominal
Time Of Day	Interval
Time on a Clock with Hands	Interval

Number of Children	Nominal
Religious Preference	Nominal
Barometer Pressure	Ratio
SAT Scores	Interval
Years of Education	Interval

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

Solution:-

Let S be the sample space then

$$S = \{H,T\}, \{H,T\}, \{H,T\}$$

then the

$$\text{event } E = \{HHT, HTH, THH\}$$

Let X be a random variable denoting the two heads and one tail.

$P(X=2)$ = probability of occurrence of 2 heads and 1 tail.

$$= P(HHT) + P(HTH) + P(THH)$$

$$= \frac{1}{2} * \frac{1}{2} * \frac{1}{2} + \frac{1}{2} * \frac{1}{2} * \frac{1}{2} + \frac{1}{2} * \frac{1}{2} * \frac{1}{2}$$

$$= \frac{1}{8} + \frac{1}{8} + \frac{1}{8}$$

$$= \frac{3}{8}$$

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:- (a) – Sum is equal to 1

If two dices were rolled, then total possible cases = 36

Total Favourable cases (Having sum = 1) = 0

As minimum sum is 2 for outcome (1,1).

Hence, probability is 0

Solution:- (b)- Sum is less than or equal to 4

(total outcomes)- If two dices were rolled, then total possible cases =36

(favourval outcomes)- such that sum is less than or equal to 4 =6

({(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)})

Probability that sum is less than or equal to 4 $\rightarrow \frac{6}{36}$

Solution:- (c)- Sum is divisible by 2 and 3

(total outcomes)- If two dices were rolled, then total possible cases =36

(favourval outcomes)- sum is divisible by 2 and 3 =25

So

Probability that sum is divisible by 2 and 3 is $=\frac{25}{36}$

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 outcomes)-Total number of balls in the bag $\rightarrow (2 + 3 + 2) = 7$

(favourval outcomes)- The number of ways of drawing 2 balls out of 7

$n(s) \rightarrow 7C2$

$$\frac{(7 * 6)}{(2 * 1)} = 21$$

Let E = Event of drawing 2 balls, none of which is blue.

The number of ways of drawing 2 balls out of (2red + 3green) balls

$n(E) \rightarrow 5C2$

$$\frac{(5 * 4)}{(2 * 1)} = 10$$

$$P(E) = \frac{n(E)}{n(S)} = \frac{10}{21}$$

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:- The Expected number of candies for a randomly selected child is **3.09**

Explanation:

$$\begin{aligned} &= 1*0.015 + 4*0.20 + 3*0.65 + 5*0.005 + 6*0.01 + 2*0.12 \\ &= 3.09 \end{aligned}$$

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

Solution:-

Parameters	Points	Score	Weigh
Mean	3.597	3.217	17.85
Median	3.695	3.325	17.71
Mode	3.92	3.44	18.90
Variance	0.276	0.297	3.09
Standard Deviation	0.53	0.97	1.78
Range	[2.76, 4.93]	[1.513, 5.424]	[14.5, 22.9]

Inferences:

- *“Points”&“Score”are negatively skewed and “Weigh”is positively skewed.*
- *“Points” has low Variance, Standard Deviation, and Range.*
- *#for code section refer Assingmentno1.ipynb 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:- Explanation:

No. of data given here is 9, then probability of choosing one patient is $1/9$, so

$$= \frac{(108+110+123+134+135+145+167+187+199)}{9}$$
$$=145.33$$

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

Cars speed and distance

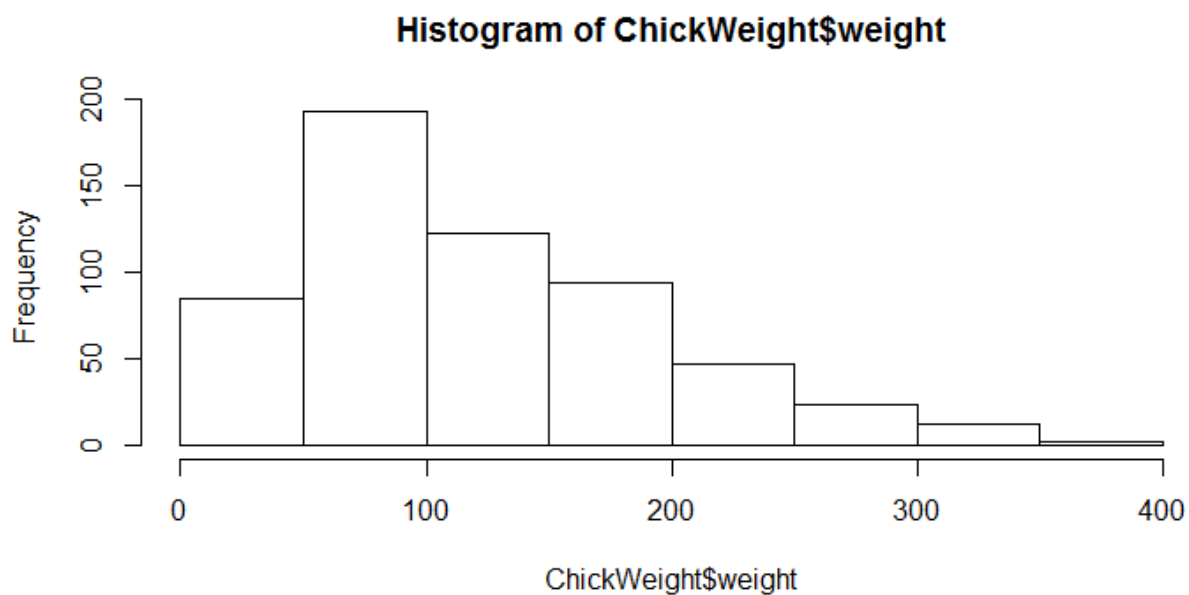
Use Q9_a.csv

SP and Weight(WT)

Use Q9_b.csv

#for code section refer Assingmentno1.ipynb file.

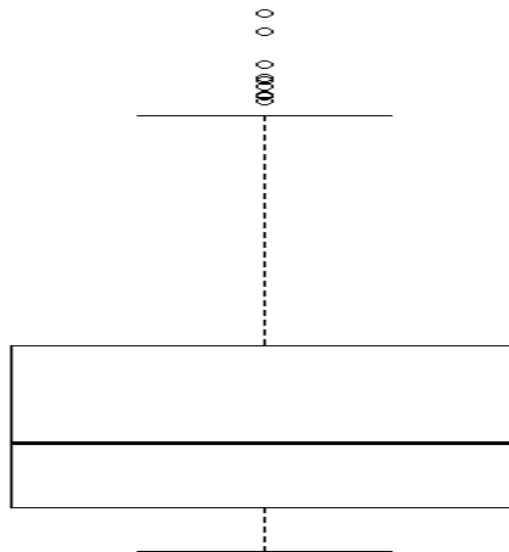
Q10) Draw inferences about the following boxplot & histogram



Solution:- *The inferences about the histogram are :-*

- Data is Positively Skewed as the tail is to the right side.
- The large part of data is concentrated on the left and has a long tail to the right.
- Mode of the data is 100.
- It's unimodal data(has a single mode) and is approximately normal.

Boxplot:



The inferences about the above boxplot are :-

- The boxplot shows that greater density of data present towards left side.
- The data are positively skewed or right skewed.
- The data contains outliers in positive direction.
- The dots represents the presence of outliers in the data.

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:-

t₉₄ is 1.8819

t₉₈ is 2.3282

t₉₆ is 2.0551

94% confidence interval:-----

$$= 200 - 1.882 * (30 / \sqrt{2000}) \text{ to } 200 + 1.882 * (30 / \sqrt{2000})$$

$$= 198.73 \text{ to } 201.26$$

98% confidence interval:----->

$$= 200 - 2.328 * (30 / \sqrt{2000}) \text{ to } 200 + 2.328 * (30 / \sqrt{2000})$$

$$= 198.43 \text{ to } 201.56$$

96% confidence interval:-→

$$= 200 - 2.055 * (30 / \sqrt{2000}) \text{ to } 200 + 2.055 * (30 / \sqrt{2000})$$

$$= 198.62 \text{ to } 201.378$$

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:- 1)-

Mean = 41
Median = 40.5
Variance = 25.52
Std deviation = 5.05

####for code section refer Assingmentno1.ipynb file.

2)-

- The average mark of students in test is 41.
- The maximum and minimum marks are 56 and 34.
- Most of the students mark between 35 to 45.

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

Answer: If the mean, median and mode of data are equal then there is no skewness.

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

Answer: If the mean > median than the distribution is positively skewed.

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

Answer: If the median > mean than the distribution is negatively skewed.

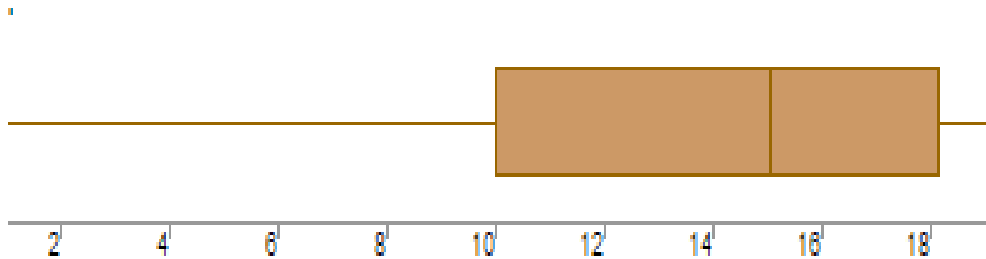
Q16) What does positive kurtosis value indicates for a data ?

Answer: Positive values of kurtosis indicate that a distribution is peaked and contains thick tails.

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

Answer: A negative kurtosis value in a distribution indicates that the distribution has lighter tails than the normal distribution

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



What can we say about the distribution of the data?

Answer:

- The distribution of this data is not Normally distributed
- Most of the values are in left so it is left skewed
- The median value of the distribution is 15
- The Lower and Upper quartile are 10 and 18

What is nature of skewness of the data?

Answer: Most of the values are skewed towards left side so the nature of the skewness is left skewed

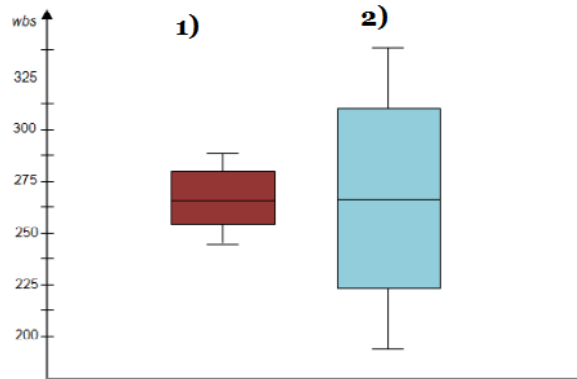
What will be the IQR of the data (approximately)?

Answer: $IQR = QR3 - QR1$

$$= 18 - 10$$

$$IQR = 10$$

Q19) Comment on the below Boxplot visualizations?



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

Answer:

- the median of both the boxplots is the same $\rightarrow (262.5)$
- Both Plots shown above has equal distribution of data above and below the median and apparently it's a normal distribution.
- Kurtosis is negative in 1st plot as it has thinner tails and positive in the 2nd plot.

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. $P(20 < \text{MPG} < 50)$

Answer:

- $P(\text{MPG} > 38) = 0.34$
- $P(\text{MPG} < 40) = 0.72$
- $P(20 < \text{MPG} < 50) = 0.89$

#for code section refer Assingmentno1.ipynb file.

Q 21) Check whether the data follows normal distribution

- a) Check whether the MPG of Cars follows Normal Distribution
Dataset: Cars.csv

Answer: Skewness around -0.5 to 0.5 is acceptable, here skewness of MPG is -0.17 hence it follows normal distribution

#for code section refer Assingmentno1.ipynb file.

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

Answer: Skewness around -0.5 to 0.5 is acceptable, here skewness of Waist is 0.13 and AT is 0.58 hence both follows normal distribution

#for code section refer Assingmentno1.ipynb file.

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

Solution:-

```
Z score of 90% confidence interval is 1.2816  
Z score of 94% confidence interval is 1.5548  
Z score of 60% confidence interval is 0.2533
```

#for code section refer Assingmentno1.ipynb file.

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% confidence interval is 1.7109  
T score for 96% confidence interval is 1.8281  
T score for 99% confidence interval is 2.4922
```

#for code section refer Assingmentno1.ipynb file.

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

[Kuldeep Jeengar Assingment No. 01]

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 \rightarrow pt(tscore,df)

df \rightarrow degrees of freedom

Solution:-

The probabily of gitting an average life of no more than 260 days are 32.181

#for code section refer Assingmentno1.ipynb file.