

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	Nominal
Level of Agreement	Ordinal
IQ(Intelligence Scale)	Interval
Sales Figures	Ratio
Blood Group	Nominal
Time Of Day	Ordinal
Time on a Clock with Hands	Interval
Number of Children	Ratio
Religious Preference	Nominal

Barometer Pressure	Interval
SAT Scores	Interval
Years of Education	Ratio

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

Ans:- OUTCOMES: HHH, HHT, HTH, TTT, TTH, THT, HTT, THH.

Probability of getting two heads and one tail :-(HHT,HTH,THH):- $\frac{3}{8}=0.375$

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

Ans) OUTCOMES:(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)

(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)

(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)

(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)

(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)

(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)

- a) 0(ZERO)
- b) (1,1),(1,2),(1,3),(2,1),(2,2),(3,1) :- $\frac{6}{36}=\frac{1}{6}=0.166$
- c) (1,5),(2,4),(3,3),(4,2),(5,1),(6,6) :- $\frac{6}{36}=\frac{1}{6}=0.166$

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) $\frac{{}^5C_2}{{}^7C_2} \rightarrow \frac{(5 \times 4 / 2 \times 1)}{(7 \times 6 / 2 \times 1)} \rightarrow \frac{20}{42} = \frac{10}{21} = 0.476$

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) $\rightarrow 0.015 \times 1 + 0.20 \times 4 + 0.65 \times 3 + 0.005 \times 5 + 0.01 \times 6 + 0.120 \times 2$
 $= 3.09$

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

Ans:- Mean of Points :- 3.596563

Mean of Score :- 3.21725

Mean of Weigh :- 17.84875

Median of Points :- 3.695

Median of Score :- 3.325

Median of Weigh :- 17.71

Mode of Points:- (3.07,3.92)

Mode of Score :- 3.44

Mode of Weigh :- (17.02,18.90)

Variance of Points :- 0.2858814

Variance of Score :- 0.957379

Variance of Weigh :-3.193166

Standard Deviation of Points :- 0.5346787

Standard Deviation of Score :- 0.9784574

Standard Deviation of Weigh :- 1.786943

Range of Points :- 2.76 4.93 (check)

Range of Score :- 1.513 5.424 (check)

Range of Weigh :-14.5 22.9 (check)

Solution: file attached in the name of Assignment 1(Sushanth).ipynb

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) $\rightarrow \frac{1}{9}(108+110+123+134+135+145+167+187+199)$
 $=145.33$

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

Cars speed and distance

Use Q9_a.csv

Ans) Skewness for cars speed :- -0.11750986144663393

Skewness for distance :- 0.8068949601674215

Kurtosis for cars speed :- -0.5089944204057617

Kurtosis for distance :- 0.4050525816795765

SP and Weight(WT)

Use Q9_b.csv

Ans) Skewness for SP :- 1.6114501961773586

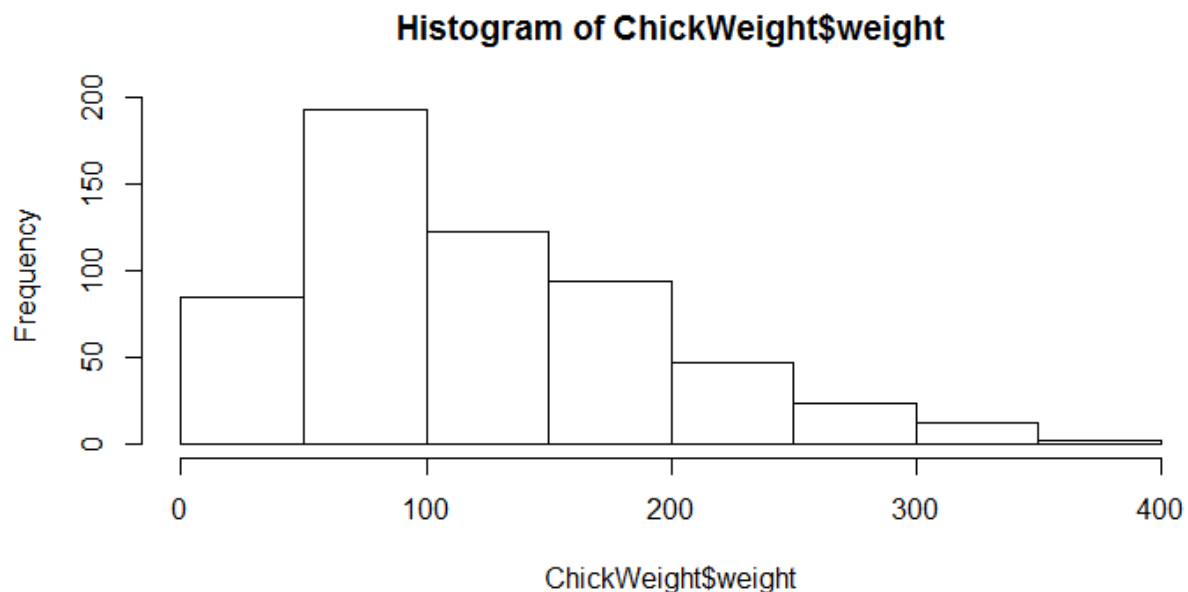
Skewness for Weight(WT) :- -0.6147533255357768

Kurtosis for SP :- 2.9773289437871835

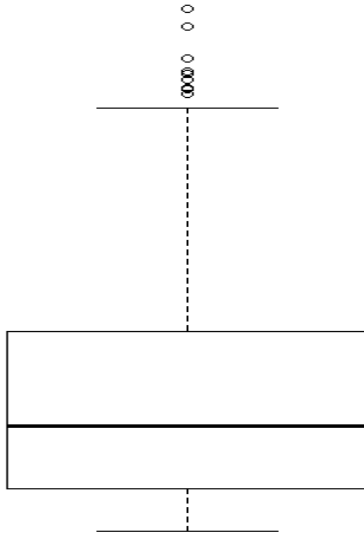
Kurtosis for Weight(WT) :- 0.9502914910300326

Solution: file attached in the name of Assignment 1(Sushanth).ipynb

Q10) Draw inferences about the following boxplot & histogram



Ans) The concentration of the mass is on the left hand side, thus it is right skewed data and I can expect the skewness for my data to be positive.



Ans) We have the Median , Upper Limit Value , Lower Limit value , 1st Quartile Value(Q1) , 3rd Quartile Value(Q3) , Outliers , Upper whisker , Lower whisker .

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?

Ans) mean=200

Sd=30

N=2000

The interval is : $\bar{X} \pm t*s/\sqrt{n}$

Considering a 94% confidence level, using a calculator , with $200-1 = 199$ df, the critical value is $t = 1.8916$

$$\bar{X} - t*s/\sqrt{n} = 200 - 1.8916*30/\sqrt{2000} = 198.73$$

$$\bar{X} + t*s/\sqrt{n} = 200 + 1.8916*30/\sqrt{2000} = 201.27$$

The 94% confidence interval is (198.73,201.27).

Considering a 96% confidence level, using a calculator , with $200-1 = 199$ df, the critical value is $t = 2.0673$

$$\bar{X} - t*s/\sqrt{n} = 200 - 2.0673*30/\sqrt{2000} = 198.61$$

$$\bar{X} + t*s/\sqrt{n} = 200 + 2.0673*30/\sqrt{2000} = 201.39$$

The 94% confidence interval is (198.61,201.39).

Considering a 98% confidence level, using a calculator , with $200-1 = 199$ df, the critical value is $t = 2.3452$

$$\bar{X} - t*s/\sqrt{n} = 200 - 2.3452*30/\sqrt{2000} = 198.43$$

$$\bar{X} + t*s/\sqrt{n} = 200 + 2.3452*30/\sqrt{2000} = 201.57$$

The 94% confidence interval is (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?

Ans :- 1)Mean :- 41

Median :- 40.5

Variance :- 25.52941

Standard Deviation :- 5.052664

2)Using boxplot which is plotted for given data, we can say that most of the students marks lies between 35 to 45.

Solution: file attached in the name of Assignment 1(Sushanth).ipynb

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

Ans:- Symmetric Distribution skewness.

Q14) What is the nature of skewness when $\text{mean} > \text{median}$?

Ans:- Right skewness.

Q15) What is the nature of skewness when $\text{median} > \text{mean}$?

Ans:- Left skewness.

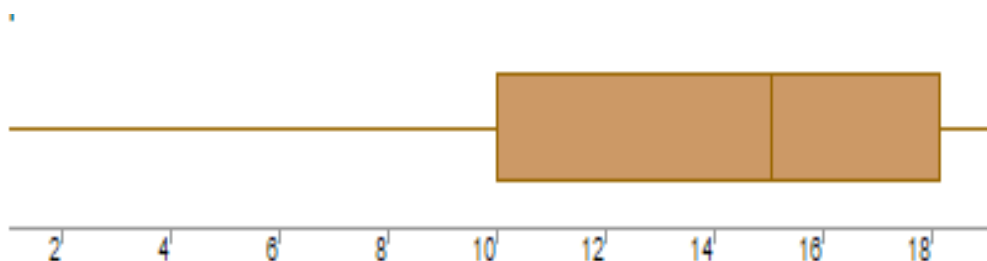
Q16) What does positive kurtosis value indicates for a data ?

Ans:- Sharp peak , Fat tails.

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

Ans:- Blunt peak , Thin tails.

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



What can we say about the distribution of the data?

Ans) First 25% of data value is less than 10 , next 25% of data value lies between 10 to 15.2 , next 25% of data value lies between 15.2 to 18.1 and last 25% of data value is greater than 18.1

What is nature of skewness of the data?

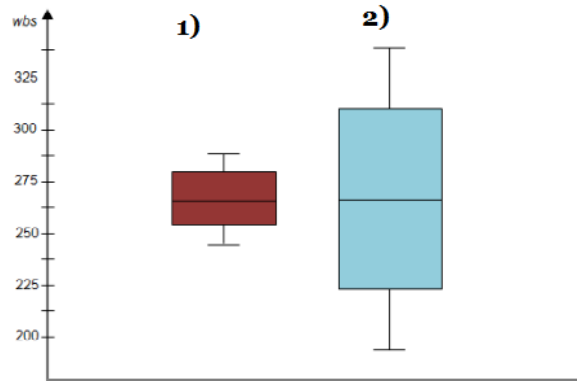
Ans) $\text{median} > \text{mean}$, Skewness is negative.

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

Ans) Here $Q1=10$, $Q2=15.2$, $Q3=18.1$

So, IQR is $18.1-10=8.1$

Q19) Comment on the below Boxplot visualizations?



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

Ans) From 1st boxplot:

The values lies between 245 to 290. The first 25% of values lies between 245 to 255 next 25% of values lies between 255 to 265 and next 25% value is lies between 265 to 280. last 25% values lies between 280 to 290.

From 2nd boxplot:

The values lies between 190 to 340. The first 25% of values lies between 190 to 225 next 25% of values lies between 225 to 255 and next 25% value is lies between 255 to 305. last 25% values lies between 305 to 340.

Therefore, by comparing the both boxplots, we can say that the median value of both distribution is same but in 1st distribution most of values is similar or nearly comparison of 2nd distribution.

The IQR for 1st distribution is $280 - 255 = 25$

The IQR for 2nd distribution is $305 - 225 = 80$

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. $P(\text{MPG} > 38)$

Ans) 0.34759392515827137

b. $P(\text{MPG} < 40)$

Ans) 0.7293498762151609

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

Ans) 0.8988689169682047

Solution: file attached in the name of Assignment 1(Sushanth).ipynb

.Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans) NO, it does not follows normal distribution.

b) Check Whether the Adipose Tissue (AT) and Waist

Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans) Skewness of AT :- 0.584869324127853

Kurtosis of AT :- -0.28557567504584425

Skewness of Waist :- 0.1340560824786468

Kurtosis of Waist :- -1.1026666011768886

NO, it does not follows normal distribution.

Solution: file attached in the name of Assignment 1(Sushanth).ipynb

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

Ans) We get that the z score at 90 % confidence interval is 1.644 , at 94 % confidence interval is 1.880 and at 60 % confidence interval is 0.841

Solution: file attached in the name of Assignment 1(Sushanth).ipynb

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

Ans) T critical value for 95% confidence level with sample size 25 is 2.0638

T critical value for 96% confidence level with sample size 25 is 2.1715

T critical value for 99% confidence level with sample size 25 is 2.7969

Solution: file attached in the name of Assignment 1(Sushanth).ipynb

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 \rightarrow pt(tscore,df)

df \rightarrow degrees of freedom

Ans) Assuming null hypothesis as $h_0 = \text{avg life of bulb} = 260 \text{ days}$

Alternative hypothesis as $h_1 = \text{avg life of bulb} \neq 260 \text{ days}$

The probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 0.32185786141570993

Solution: file attached in the name of Assignment 1(Sushanth).ipynb

