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)	**Ratio**
Sales Figures	**Interval**
Blood Group	**Nominal**
Time Of Day	**Ratio**
Time on a Clock with Hands	**Ratio**
Number of Children	**Ordinal**
Religious Preference	**Nominal**

Barometer Pressure	**Ratio**
SAT Scores	**Ratio**
Years of Education	**Interval**

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

Answer: (HHH, HHT, HTH, THH, TTH, THT, HTT, TTT)

Probability of two head and one tail = 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

Answer:

(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) sum of the two dice is not equal to 1 so, answer is 0

b) sum is less than and equal to 4 is = 6/36 = 1/6 = 0.16

c) There is one more answer of (c) , outcomes are (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?

Answer:

Total balls are 7

Two balls are drawn randomly = ${}^{N}C_{r} = {}^{7}C_{2}$ = 7*6*5*4*3*2*1/2*1(5*4*3*2*1) =21

Here consider only red and green balls = ${}^{N}C_{r} = {}^{5}C_{2}$ = 5*4*3*2*1/2*1(3*2*1) = 10

Probability = 10/21 = 0.4761

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

Answer:

The Expected number of candies for a randomly selected child

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

= 0.015+0.8+1.95+0.025+0.06+0.24

= **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.

Answer: -

	Points	**Score**	**Weight**
Mean	3.596563	3.21725	17.84875
Median	3.695	3.325	17.71
Mode	3.92	3.44	17.02
Variance	0.28588	0.95738	3.19317
Standard Deviation	0.53468	0.97846	1.786943
Range	2.17	3.911	8.4

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?

Answer: -

E(x)	108	110	123	134	135	145	167	187	199
P(x)	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9

Probability = (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)

= 145.33 pounds

Q9) Calculate Skewness, Kurtosis & draw inferences on the following data Car's speed and distance. (Q9_a.csv).

Answer: -

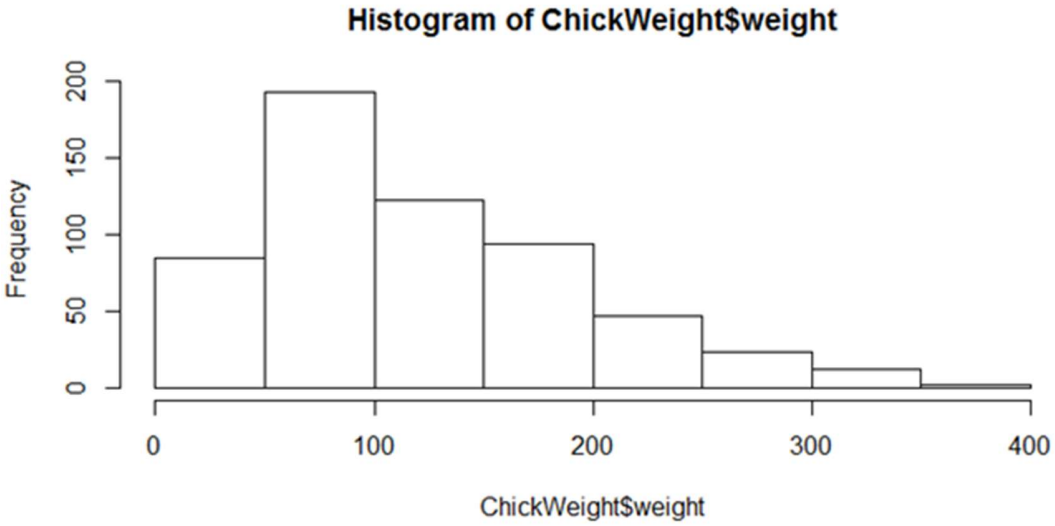
	SPEED	DIST
Skewness	-0.117509861	0.80689496
Kurtosis	-0.50899442	0.405052582
Mean	15.4	42.98
Median	15	42.98
Mode	20	26

SP and Weight (WT) (Use Q9_b.csv)

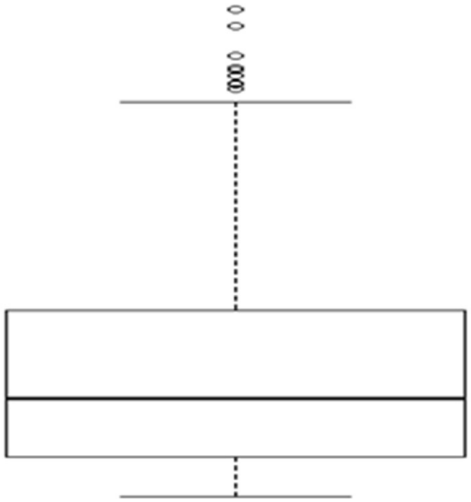
Answer: -

	SP	WT
Skewness	1.611450196	-0.614753326
Kurtosis	2.977328944	0.950291491
Mean	121.5402722	32.41257691
Median	118.2086984	32.73451818
Mode	118.2889958	

Q10) Draw inferences about the following boxplot & histogram



Answer: This Histogram shows Positive skewed distribution, because Mean>Median>mode.



Answer: In this boxplot many(higher) outliers available of dataset.

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?

Answer:

$$CI = \bar{x} \pm z\frac{s}{\sqrt{n}}$$

Where,

CI = confidence interval, $\bar{x}$ = sample mean = 200, z = confidence level value

s = sample standard deviation = 30, n = sample size = 2000

	94%	96%	98%
CI	134.89, 265.102	130.209, 269.790	122.725, 277.274

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?

Answer:

1.

	Mean	**Median**	**Variance**	**Standard Deviation**
34,36,36,38,38, 39,39,40,40,41, 41,41,41,42,42, 45,49,56	41	40.5	24.11	4.91

2. Multi modal

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

Answer: Symmetric Distribution

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

Answer: positive Skew Distribution

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

Answer: Negative Skew Distribution

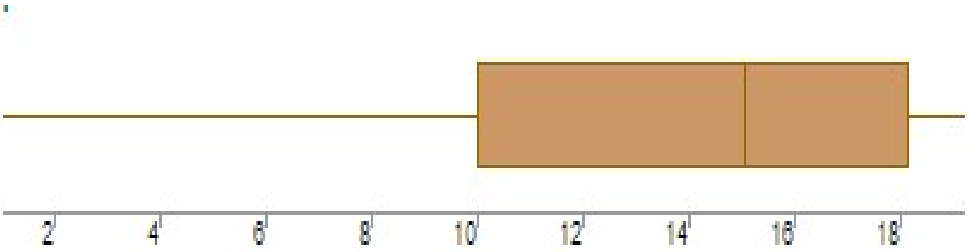
Q16) What does positive kurtosis value indicates for a data?

Answer: distribution is peaked and possesses thick tails.

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

Answer: distribution is flat and has thin tails

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



- What can we say about the distribution of the data?

  Answer: Non-Symmetric distribution (median > mean)

- What is nature of skewness of the data?

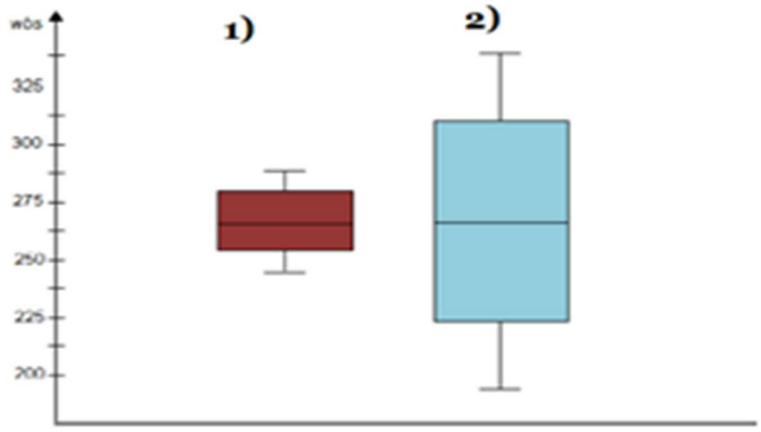
  Answer: Negative Skewed

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

  Answer:
  $Q_1 = 10$, $Q_3 = 18$

  $Q_3 - Q_1 = 18 - 10 = 8$

Q19) Comment on the below Boxplot visualizations?



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

Answer:

- In first boxplot the median is closer to Q_1 so that this positively skewed i.e., Mean & median is greater than mode.
- In second boxplot median is in the middle where mean = median so there is symmetric distribution.

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(MPG>38)
b. P(MPG<40)
c. P (20<MPG<50)

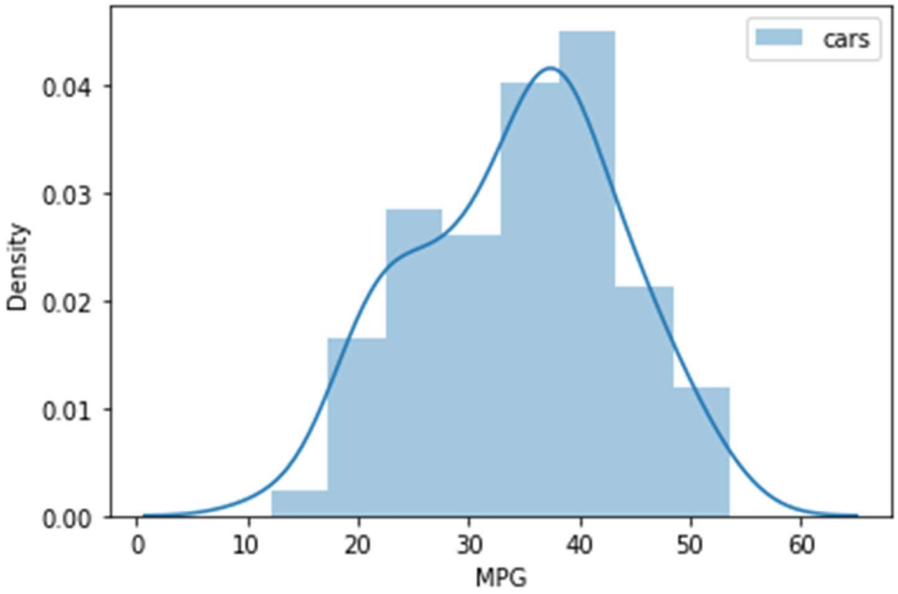
Answer:

a. P(MPG>38) = 0.34750288929863415
b. P(MPG<40) = 0.7294349739243934
c. P(20<MPG<50) = 0.8988805681995043

Q 21) Check whether the data follows normal distribution

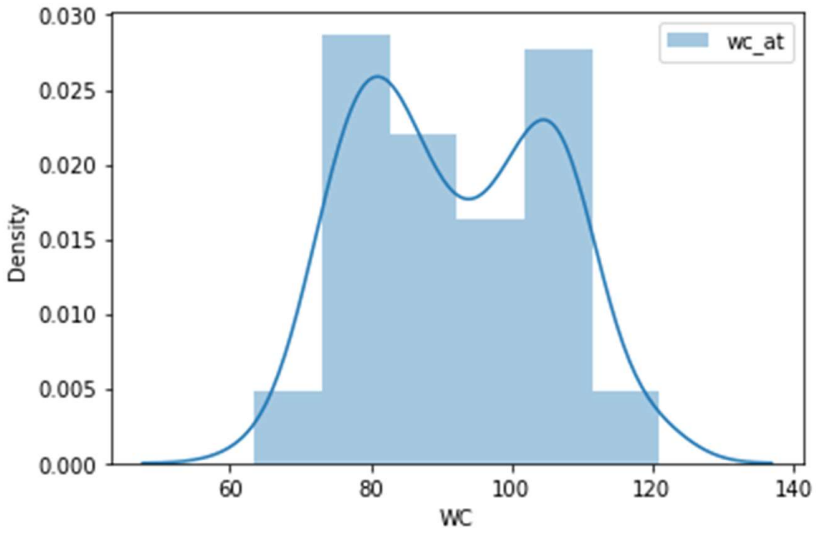
a) Check whether the MPG of Cars follows Normal Distribution
Dataset: Cars.csv

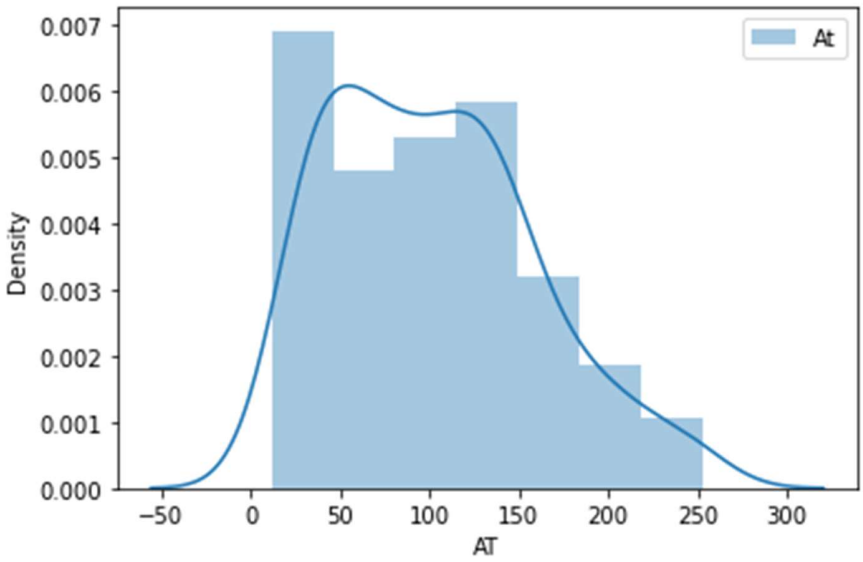
Answer: as per the calculation data has not Normal Distribution.



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

Answer: As per the calculation data has not Normal Distribution.





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

Answer:

percentage	Confidence Interval
90	1.959963984540054
94	1.8807936081512509
60	0.8416212335729143

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

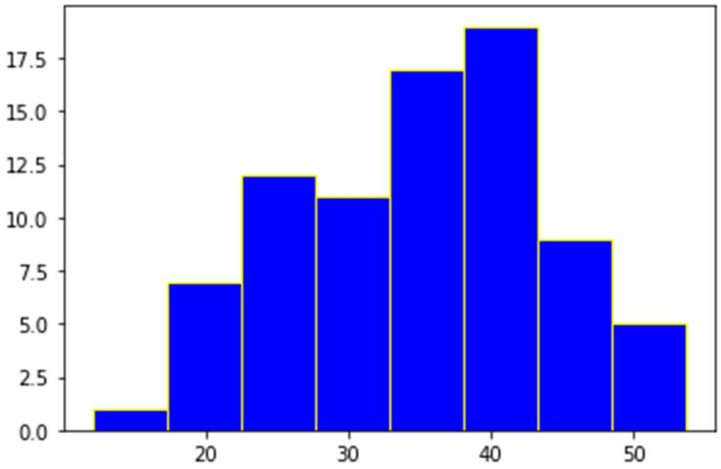
Answer:

percentage	Confidence Interval
95	2.0638985616280205
96	1.8807936081512509
99	0.8416212335729143

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.

Answer:

Probability = 0.6783274643290165



This graph shows that data is non-symmetric and negative skewed.