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	Nominal
Time on a Clock with Hands	Ratio
Number of Children	Ratio
Religious Preference	Ordinal

Barometer Pressure	Interval
SAT Scores	Ratio
Years of Education	Ratio

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?0.375

TTH HHH
TTT HTH
THH HTT
THT HHT

$P(2H\,1T) = 3/8 = 0.375$

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

a. 0
b. 1/6 = 0.167
c. 2/3 = 0.67

with 2 dices available, there are 36 (6×6) combinations.

Minimum sum (1,1) = 2

a) ∴ P(sum = 1) = 0

b) combinations for sum ≤ 4 are (1,1), (1,2), (1,3), (2,1), (2,2), (3,1)

P(sum ≤ 4) = 6/36 = 1/6

c) Events divisible by 2,
(1,1), (1,3), (1,5), (2,2), (2,4), (2,6),
(3,1), (3,3), (3,5), ...

Total no. = 3 × 6 = 18

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

Total no. = $2 \times 6 = 12$

There are a few repetative events.

∴ Total no. of events divisible by 2 and 3 are = $18 + 6 = 24$

P(sum divisible by 2 and 3) = $24/36 = 2/3$

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? P = 10/21

2R 3G 2B Total 7 balls.

$^7C_2 = \frac{7!}{5!\,2!} = \frac{7 \times 6}{2} = 21$ (Total no. of ways of drawing 2 balls)

Now, since none of the balls can be blue, we have to select 2 balls from red and green.

Total no. of ways of selecting 2 red or green balls = $^5C_2 = \frac{5 \times 4 \times 3!}{3! \times 2} = 10$

∴ P(none of balls is blue) = $10/21$

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

3.09

Expected no. of candies for a randomly selected child

$= 1 \times 0.015 + 4 \times 0.20 + 3 \times 0.65 + 5 \times 0.005 + 6 \times 0.01 + 2 \times 0.12$

$= 3.09$

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

please refer to the jupyter notebook.

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?

Expected weight of a person

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

$= 145.33$

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

Cars speed and distance

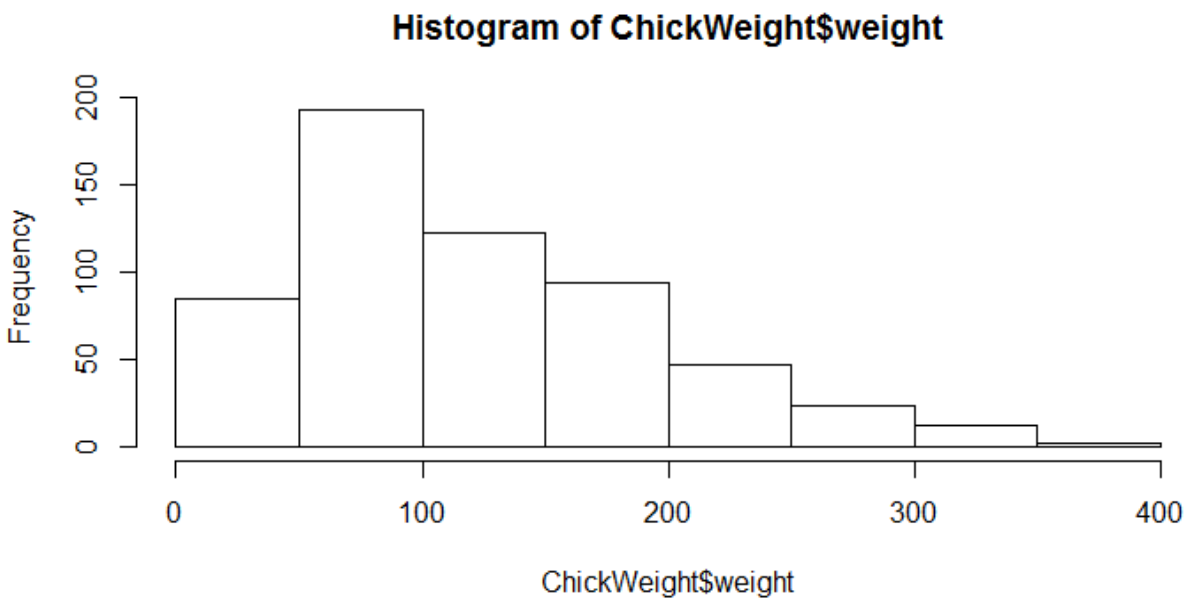
Use Q9_a.csv

SP and Weight(WT)

Use Q9_b.csv

Please refer to the jupyter notebook.

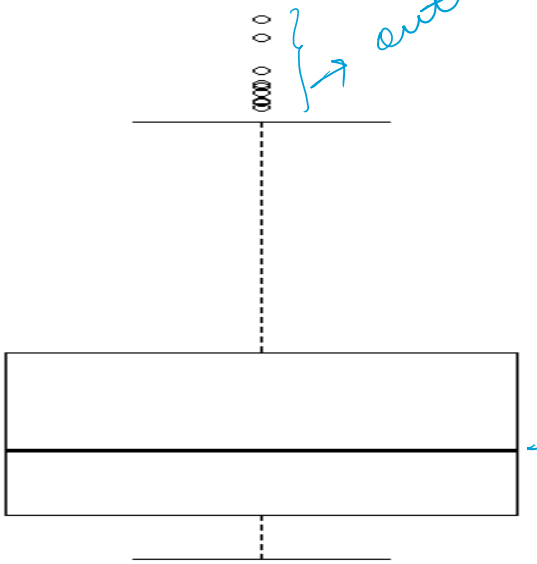
Q10) Draw inferences about the following boxplot & histogram



1. The range of the variable weight is from 0 to 400.

2. The mode of the variable lies between (50-100); since the frequency is max in this range

3. Since the mass of the distribution is concentrated in the right side, it is positively skewed.



1. upper whisker shows that 25% of the data points have very high values.

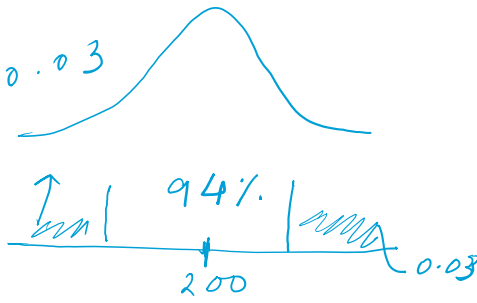
2. But the other 75% of the data points have values on the lower side.

Hence, the median is also on the lower side.

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?

$n = 2000$. $\bar{X} = 200$ $S = 30$

$$Z \text{ score} = \frac{X - \mu}{S/\sqrt{n}}$$



here, we are considering $\bar{X} = \mu = 200$.

1. 94% confidence level.

$$\alpha = 1 - \frac{94}{100} = 6\% = 0.06$$

$$\alpha/2 = 0.03$$

Now, I have to find the values of a and b.

$P(x < a) = 0.03$ Z score $= -1.88$

$P(x < b) = 0.94 + 0.03 = 0.97$ | Z score $= 1.88$

$P(a < x < b) = 0.94$

$$Z_a = \frac{a - 200}{30/\sqrt{2000}} \quad \text{or,} \quad -1.88 \times \frac{30}{44.7} + 200 a$$

$$\text{or, } a = -\frac{56.4}{44.7} + 200 = -1.26 + 200 = 198.74$$

$$Z_b = \frac{b - 200}{30/\sqrt{2000}} \Rightarrow b = 200 + 1.88 \times \frac{30}{44.7}$$

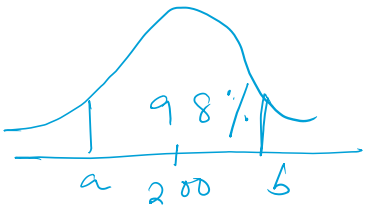
$$= 200 + 1.26 = 201.26$$

∴ The confidence interval is

(198.74, 201.26)

For 98% confidence level,

$\alpha = 2\% = 0.02$



$P(X < a) = 0.01$

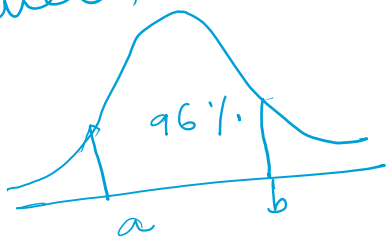
$\therefore P(X < b) = .98 + 0.01 = 0.99$

$CI = 200 \pm Z_{\alpha/2} \frac{30}{\sqrt{2000}} = 200 \pm 2.3 \times \frac{30}{44.7}$

$= 200 \pm 1.54$

$C.I = (198.46, 201.54)$

Now, for 96% confidence level, $\alpha = 4\%$.



$P(X < a) = 0.02 \quad P(X > b) = 0.02$

$C.I = 200 \pm Z_{\alpha/2} \frac{S}{\sqrt{n}} = 200 \pm 2.05 \times \frac{30}{\sqrt{2000}}$

$= 200 \pm 1.37$

$C.I = (198.63, 201.37)$

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.
   Mean= 41
   Median=40.5
   Mode= 41

Refer to the jupyter notebook

Variance= 25.5

Standard Deviation = 5.2

2) What can we say about the student marks?

The range is from 34 to 56. But from the boxplot we can see 49 and 56 are outliers. The average score of the student is 41.

The skewness is 1.5, therefore the distribution is slightly right skewed.

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

Skewness = 0

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

The dist. is positively skewed.

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

The dist is negatively skewed.

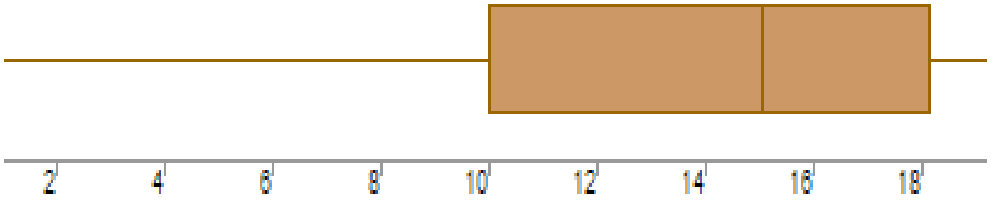
Q16) What does positive kurtosis value indicates for a data ?

It means the dist. is peaked and have long tails. That is, more data points are located in the tails rather than around the mean. It is called Leptokurtic.

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

It means the dist is more flatter with more values around its mean rather than the tails. It is called platykurtic.

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



What can we say about the distribution of the data?

The range of the dist is from 2 to 20, with the median around 15.2 (approx).

IQR → (10, 18)

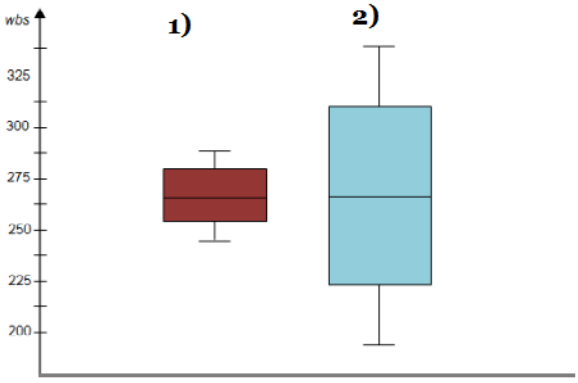
What is nature of skewness of the data?

Here, we can say the mean would be less than the median, so it is negatively skewed.

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

IQR → (10, 18).

Q19) Comment on the below Boxplot visualizations?



Range

1) (212.5, 287.5)

2) (200, 337.5)

Median

262.5 (Same for both).

Even though both the plots have same median, their interquartile range is different.

IQR ① (250, 275) ② (225, 300)

∴ The plot (2) is more widespread.

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

Please refer to the jupyter notebook.

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

Please refer to the jupyter notebook.

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

Check the notebook

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

Check the notebook

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

Check the notebook