

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	Discrete Data- Nominal
High School Class Ranking	Discrete Data-Nominal
Celsius Temperature	Continuous-Interval
Weight	Continuous-Ratio
Hair Color	Discrete Data-Ratio
Socioeconomic Status	Continuous-Interval
Fahrenheit Temperature	Continuous-Ratio
Height	Continuous-Ratio
Type of living accommodation	Discrete-Ordinal
Level of Agreement	Discrete-Interval
IQ(Intelligence Scale)	Discrete-Interval
Sales Figures	Discrete-Interval
Blood Group	Discrete-Ratio
Time Of Day	Continuous-Interval
Time on a Clock with Hands	Continuous-Interval
Number of Children	Discrete-Interval
Religious Preference	Discrete-Ratio
Barometer Pressure	Discrete-Interval
SAT Scores	Discrete-Ratio
Years of Education	Discrete-Nominal

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

Total number of possible combinations = $2 \times 2 \times 2 = 8$

Total Combinations = { HHH, HHT, HTH, THH, TTH, THT, HTT, TTT }

2-H & 1-T Combinations = { HHT, HTH, TTH }

Probability = 2-H & 1-T combinations / No of possible combinations

$$= 3/8 = 0.375$$

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

Equal to 1

Less than or equal to 4

Sum is divisible by 2 and 3

Ans Total number of combination = $6 \times 6 = 36$

a) Equal to 1 = 0 Combinations

b) Less than or Equal to 4

The possible combinations are

die 1	die 2	sum
-------	-------	-----

1	1	2
---	---	---

1	2	3
---	---	---

1	3	4
---	---	---

2	1	3
---	---	---

2	2	4
---	---	---

3	1	4
---	---	---

$$= 6/36 = 1/6$$

c) Sum is divisible by 2 and 3

Sum should be divisible by both 2 and 3

Favorable outcomes = (1, 5), (3, 3), (4, 2), (5, 1), (6, 6)

Therefore, Number of favorable outcomes = 5

Probability = $\frac{5}{36}$

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?

Total number of balls = $(2 + 3 + 2) = 7$

Let S be the sample space.

Then, $n(S)$ = Number of ways of drawing 2 balls out of 7

$= {}^7C_2$

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

$= 21$

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

$\therefore n(E)$ = Number of ways of drawing 2 balls out of (2 + 3) balls.

$= {}^5C_2$

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

$= 10$

$\therefore P(E) = \frac{n(E)}{n(S)} = \frac{10}{21} = 0.47$

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) Expected number of candies for a randomly selected child

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

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

$$= 3.090$$

$$= 3.09$$

Expected number of candies for a randomly selected child = 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.

Pls check the ipynb included in mail

Use Q7.csv file

Q8) Calculate Expected Value for the problem below

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?

Number of patients = 9

Probability of selecting each patient = $1/9$

108, 110, 123, 134, 135, 145, 167, 187, 199

$1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9$

Expected

value= $108*(1/9)+110*(1/9)+123*(1/9)+134*(1/9)+135*(1/9)+145*(1/9)+167*(1/9)+187*(1/9)+199*(1/9)$

=145.33

Expected value of weight of that person=145.33

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

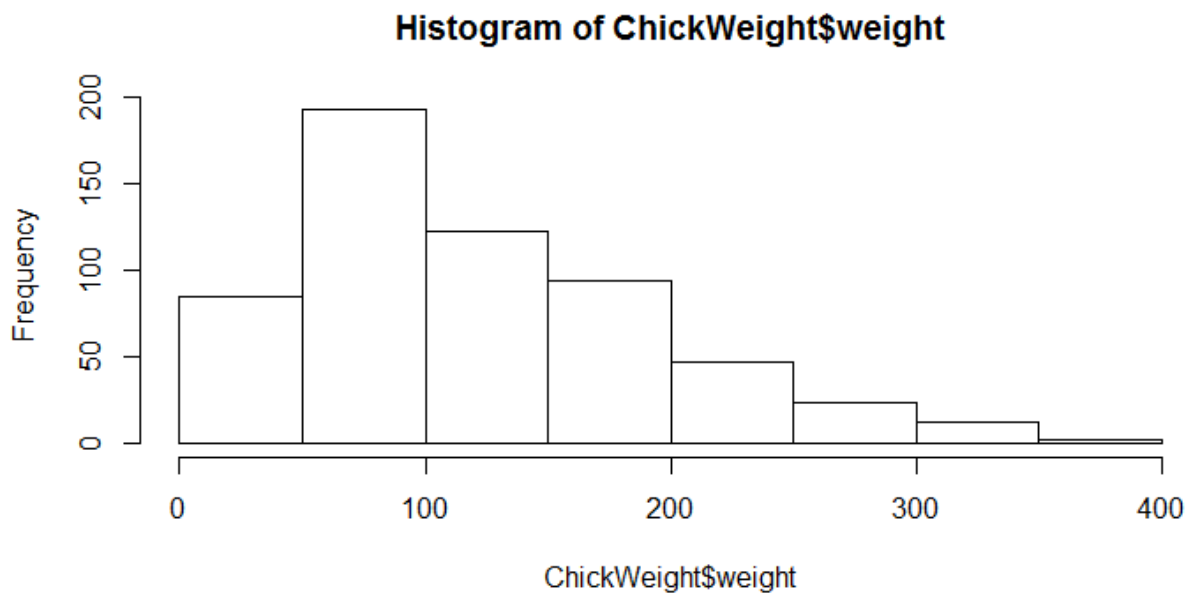
Cars speed and distance

Use Q9_a.csv

SP and Weight(WT)

Use Q9_b.csv

Q10) Draw inferences about the following boxplot & histogram



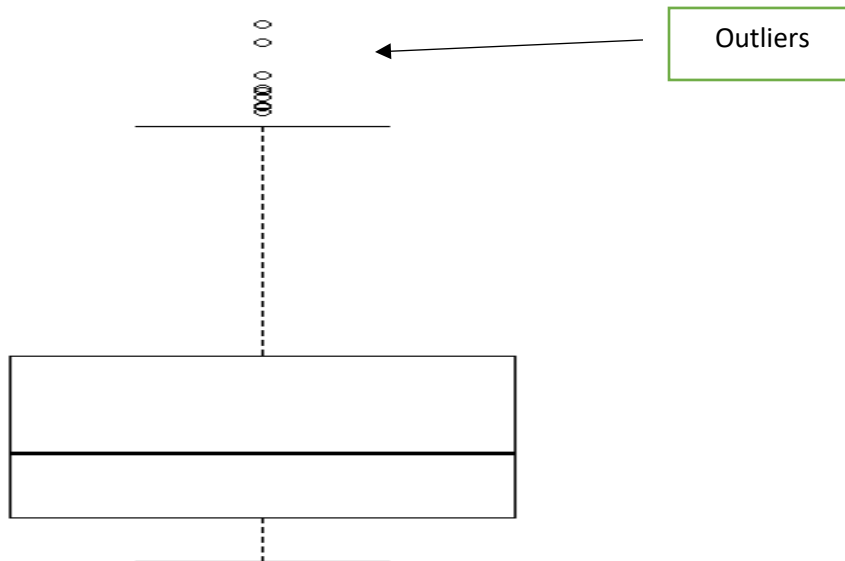
Inferences:

The Histogram is used to plot continuous data

From range 0 to 100 the frequency is increasing and later decreasing

Used for continuous data that is chickweight

Has positive skewness



Inferences:

Boxplot is mainly used to detect the outliers

The points that cross the max line are called outliers

Data is distributed towards minimum range

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?

Kindly check the ipynb included

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

Find mean, median, variance, standard deviation.

What can we say about the student marks?

Kindly check the ipynb file included

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

Skewness=0 , if mean = median and distribution is symmetric

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

Skewness=positively skewed if mean>median

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

Skewness=negatively skewed if median>mean

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

Kurtosis>3

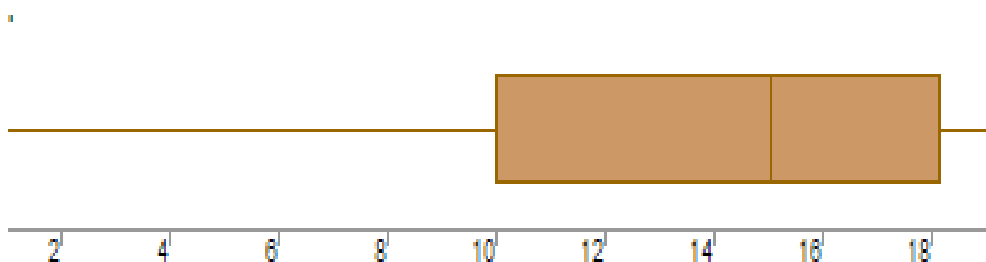
Indicates the distribution has heavier tails and sharper peak than normal distribution

More numbers are located in the tails rather than near mean

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

Has thinner tails and wider peak than normal distribution

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



What can we say about the distribution of the data?

Data is valued from 1 to 19 without any outliers present

What is nature of skewness of the data?

The data is negatively or left skewed

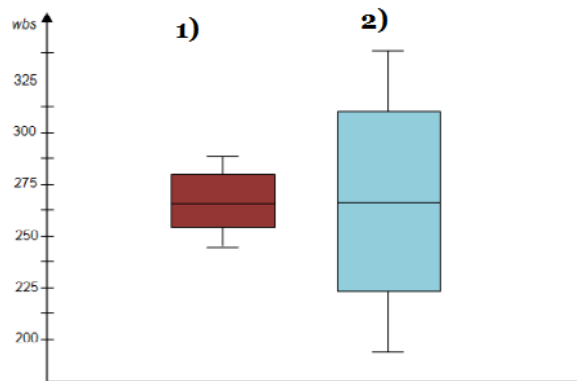
What will be the IQR of the data (approximately)?

$q1=10$

$Q2=18$

$Iqr=q2-q1=8(\text{approx.})$ = Inner quartile range (Where the range is upper quartile- inner quartile)

Q19) Comment on the below Boxplot visualizations?



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

Both have same mean i.e 265

Both are symmetrically distributed

Range of dataset1 is smaller than range of dataset2

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`

$P(MPG > 38)$

$P(MPG < 40)$

c. $P(20 < MPG < 50)$

Q 21) Check whether the data follows normal distribution

Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

From histogram, the graphs do not contain a bell shaped curve hence it will not follow normal distribution

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

Dataset: wc-at.csv

From histogram, the graphs do not follow normal distribution

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

Check the ipynb file included

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

Check the ipynb file included

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

Check the ipynb file included