

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

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

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

Ans: Let 'S' be the sample – space. Then $S = \{ HHH, HHT, HTH, THH, HTT, THT, TTH, TTT \}$

→ Let E_1 = Event of getting '2' heads. Then:

$$E_1 = \{ HHT, HTH, THH \}$$

$$|E_1| = 3$$

$$P(E_1) = 3 / 8$$

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: a) favorable outcome (*sum equal to 1*) = 0

Required probability = $0/36 = 0$

b) (1,3) (2,2) (3,1) = 3 outcomes, $3/36$ i.e. $1/12$

c) Hence, number of possible outcomes = $6 \times 6 = 36$ {outcome can be any from (1,2),(1,3)..(1,6),(2,1),(2,2)...(6,6)}

Favorable outcomes are

$\{(1,1),(1,2),(1,3),(1,5),(2,1),(2,2),(2,4),(2,6),(3,1),(3,3),(3,5),(3,6),(4,2),(4,4),(4,5),(4,6),(5,1),(5,3),(5,4),(5,5),(6,2),(6,3),(6,4),(6,6)\}$

(because sum of the numbers should be divisible by 2 or 3)

Number of favorable outcomes = 24

Probability = $24/36 = 2/3$

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?

Ans: total number of balls = $(2+3+2) = 7$

$n(S)$ = no. of ways of drawing 2 balls out of 7

$$= {}^7C_2$$

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

$$= 21$$

(as E = drawing 2 balls, none of which is blue.)

$n(E)$ = no. of ways of drawing 2 balls out of 5 balls

$$= {}^5C_2$$

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

$$= 10$$

so, $P(E) = n(E)/n(S)$

$$= 10/21$$

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

Ans: Expected number of candies for a randomly selected child

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

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

= 3.090

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.

Use Q7.csv file

Ans:

Rows/Columns	Mean	Median	Mode	Variance	S.D	Range
Points	3.596563	3.69	3.92	0.292243	0.540595	2.17
Score	3.21725	3.435	3.44	0.977018	0.988442	3.911
Weigh	17.89355	17.82	17.02	3.233244	1.798122	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?

Ans: 150 is the expected weight of patient.

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

Cars speed and distance

Use Q9_a.csv

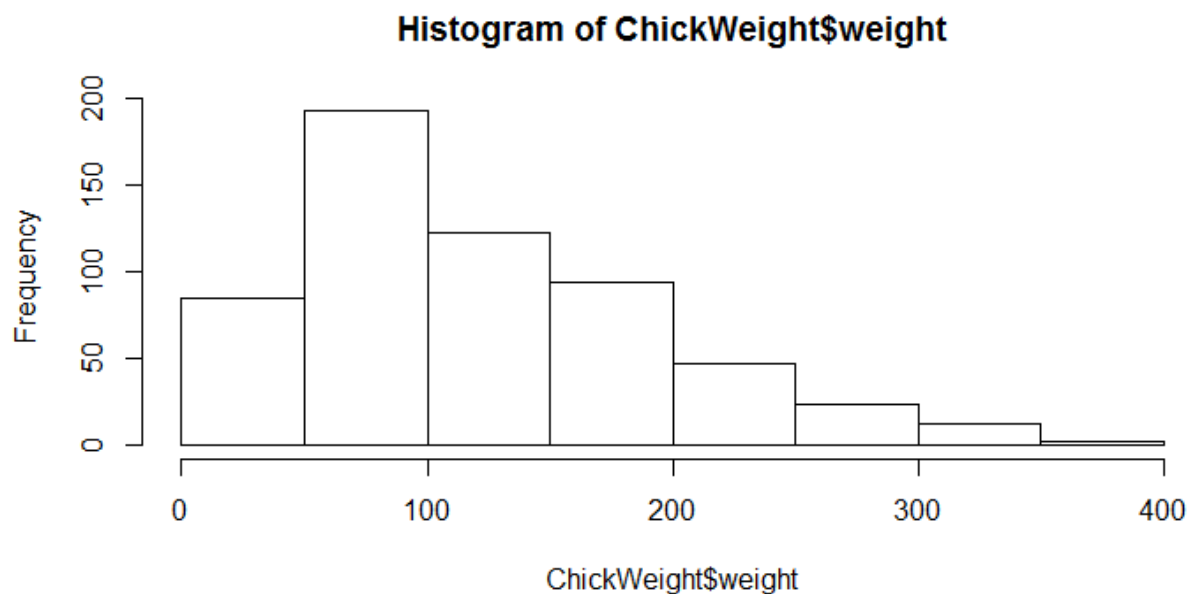
SP and Weight(WT)

Use Q9_b.csv

Ans: Skewness for speed= -0.1139548, skewness value is negative so it is left skewed. Since magnitude is slightly greater than 0 it is slightly left skewed

And for distance= 0.7824835, right skewed (Positive) slight magnitude to right.

Q10) Draw inferences about the following boxplot & histogram

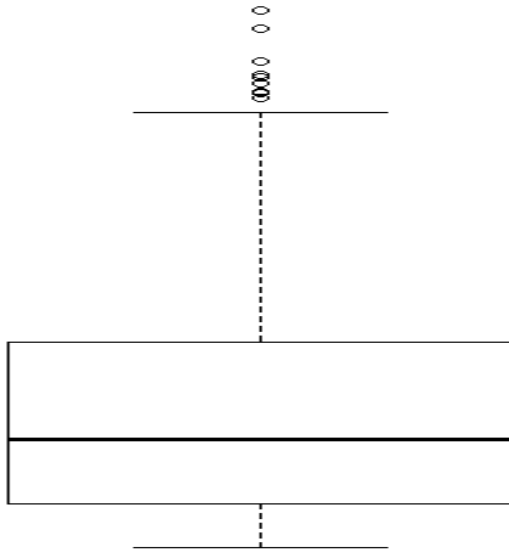


Ans: The most of the data points are concentrated in the range 50-100 with frequency 200.

And least range of weight is 400 somewhere around 0-10.

So the expected value the above distribution is 75.

Skewness- we can notice a long tail towards right so it is heavily right skewed



Ans: Median is less than mean right skewed and we have outlier on the upper side of box plot and there is less data points between Q1 and bottom point.

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: $\bar{X} \pm (Z_{1-\alpha} \cdot \sigma / \sqrt{n})$

Degrees of freedom= 2000-1= 1999

Confidence interval= 94%

$(1 - \sigma/2) = 1 - 0.03 = 0.97$

for confidence interval for 94% is 1.882

Confidence interval for 98%= 2.33

Confidence interval for 96% = 2.05

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.

Ans: Mean= 41, Median= 40, variance= 24.111, Standard deviation= 4.910

2) What can we say about the student marks?

Ans: The range of students marks is between 34 to 60.

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

ANS: Symmetrical

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

ANS: Right Skewed

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

ANS: Left Skewed

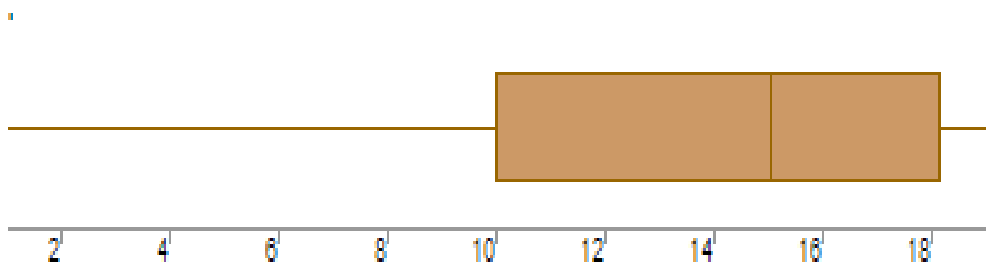
Q16) What does positive kurtosis value indicates for a data ?

ANS: The data is normally distributed and kurtosis value is 0.

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

ANS: The distribution of the data has lighter tails and a flatter peaks than the normal distribution.

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



What can we say about the distribution of the data?

Ans: Let's assume above box plot is about ages of the students in a school. 50% of the people are above 10 yrs. old and remaining are less. And students whose age is above 15 are approx. 40%.

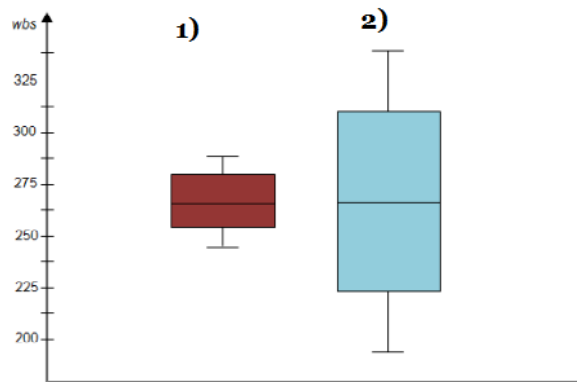
What is nature of skewness of the data?

Ans: Left skewed, median is greater than mean.

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

Ans: Approximately = -8

Q19) Comment on the below Boxplot visualizations?



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

Ans: By observing both the plots whisker's level is high in boxplot 2, mean and median are equal hence distribution is symmetrical.

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: There are 33 observations where the MPG value is greater than 38.

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

Ans: There are 61 observations where the MPG value is less than 40.

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

Ans: There are 73 observations where the MPG is between 20 - 50 .

Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: Yes, the data has been normally distributed for MPG of Cars data set.

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

Dataset: wc-at.csv

Ans: Adipose Tissue (AT) from wc-at dataset follows normal distribution,
Waist from wc-at dataset follows uniform distribution

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

Ans: Z score of 90% confidence interval is 1.65

Z score of 94% confidence interval is 1.55

Z score of 60% confidence interval is 0.85

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

Ans: Z score of 95% confidence interval is 1.96

Z score of 96% confidence interval is 2.5

Z score of 99% confidence interval is 2.47

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: $x = \text{mean of sample of bulbs} = 260$

μ = population mean = 270

s = standard deviation of sample = 90

n = number of items in sample = 18

$$t = \frac{260-270}{\frac{90}{\sqrt{18}}}$$

$$t = \frac{-10}{\frac{90}{3\sqrt{2}}}$$

$$t = \frac{-10}{\frac{30}{\sqrt{2}}}$$

$$t = \frac{-1 \times \sqrt{2}}{3}$$

$$t = -0.471$$

ثThe probability $t < -0.471$ with 17 degrees of freedom assuming population mean is true, the probability of bulbs less than 260 days average of 0.3218 assuming mean life of bulbs is 300 days.