

Q1) Identify the Data type for the Following:

Activity	Data Type
Number of beatings from Wife	Discrete
Results of rolling a dice	Ordinal
Weight of a person	Ratio
Weight of Gold	Ratio
Distance between two places	Ratio
Length of a leaf	Ratio
Dog's weight	Ratio
Blue Color	Nominal
Number of kids	Discrete
Number of tickets in Indian railways	Ordinal
Number of times married	Discrete
Gender (Male or Female)	Nominal

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?

- Let S be the sample space then
 $S = \{H,T\}, \{H,T\}, \{H,T\}$ then the event E is $\{HHT, HTH, THH\}$
Let X be a random variable denoting the two heads and one tail.
 $P(X=2)$ = probability of occurrence of 2 heads and 1 tail.
- $$= P(HHT) + P(HTH) + P(THH)$$
- $$= (1/2 * 1/2 * 1/2) + (1/2 * 1/2 * 1/2) + (1/2 * 1/2 * 1/2) = 3/8$$

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
- two dice are thrown here n
 $n(s) = 36$
a) the sum is equal to 1
is zero because they start with (1,1) likewise other than in the dice we are not having zero.
b) the sum is equal to 4
the possible outcomes are (1,3), (2,2), (3,1) therefore $n(b) = 3/36 = 1/12$
c) Sum is divisible by 2 and 3.
So here the outcome will be on solving we get . $n(c) = 6/36 = 1/6$

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?

- Total ball is 2 yellow + 3 green + 2 blue = 7 ball.
It can be written as 7C_2 ,
So None of the ball is blue that is 7 - 2 (blue ball) is = 5 ball Or 5C_2
So ${}^5C_2 / {}^7C_2 = (5! / 3! * 2!) // (7! / 5! * 2!) = 10/21$

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

➤ Expected number of candies for a randomly selected child

$$\begin{aligned} &= 1 * 0.015 + 4 * 0.20 + 3 * 0.65 + 5 * 0.005 + 6 * 0.01 + 2 * 0.12 \\ &= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24 \\ &= 3.090 \\ &= 3.09 \end{aligned}$$

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

- For Points, Score, Weight
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

Use Q7.csv file

Ans – So here,

We get for

Points:

Mean=3.596563, Median=3.695, Mode=numeric,
Variance=0.2858814, Std deviation=0.5346787

Score:

Mean=3.21725, Median=3.325, Mode=numeric,
Variance=0.957379, Std deviation=0.9784574

Weight:

Mean=17.84875, Median=17.71, Mode=numeric,
Variance=3.193166, Std deviation=1.78694

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 - There are 9 patients

Probability of selecting each patient = $1/9$

Expected Value

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

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

$$= (1/9) \cdot (1308)$$

$$= 145.33$$

Expected Value of the Weight of that patient = 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

ANS –1) By using Q9_a.csv

Skewness-

Car speed= -0.11395480, Distance= 0.7824835

Kurtosis-

Car speed= 2.4228533, Distance= 3.248019

2) By using Q9_b.csv

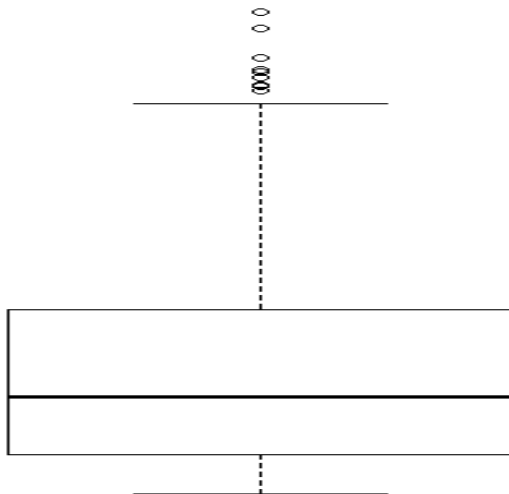
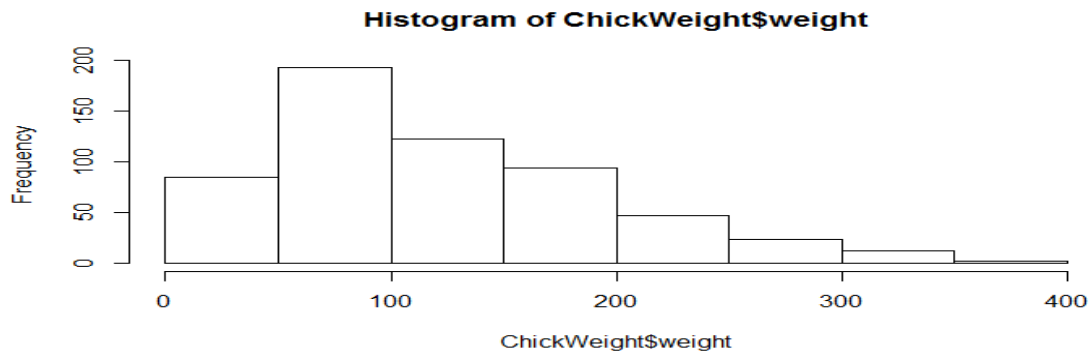
Skewness-

SP =1.581454 **WT**= -0.6033099

Kurtosis-

SP= 5.723521 **WT**= 3.819466

Q10) Draw inferences about the following boxplot & histogram



ANS –

- The most of the data are concentrated in the range 50-100 with the frequency 200. And the least range of weight is 400 somewhere around 0-10. So the expected value of the above distribution is 75 and as it has a long tail at right it is right skewed.
- Median is less and we have outliers on the upper side of box plot and there are 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 – $\bar{X} \pm (Z_{1-\alpha/2} \cdot \sigma / \sqrt{n})$

Degree of freedom = 2000 – 1 = 1999

Confidence interval = 94%

$$(1 - \frac{\alpha}{2}) = 1 - 0.03 = 0.97$$

for 94% is 1.882

for 98% is 2.33

for 96% is 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.

- Mean=41, Median=40, Variance=24.111, Std deviation=4.910

2) What can we say about the student marks

- Mass of students marks between 38-42. Skewness(1.52) is positive because mass of marks in left side of plot.

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

- Symmetrical.

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

- Right skewed.

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

- Skewed.

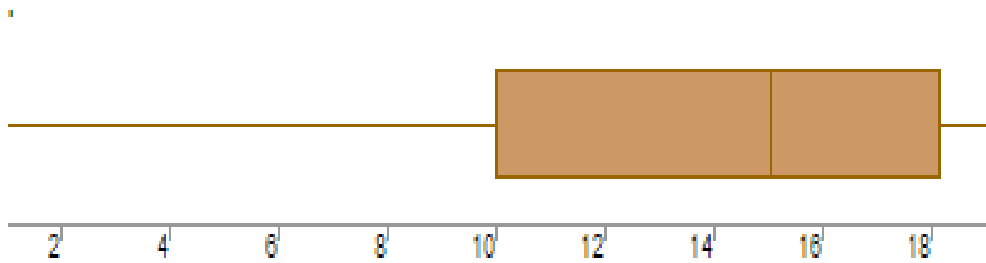
Q16) What does positive kurtosis value indicates for a data ?

- The data is normally distributed and kurtosis value is 0.

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

- The distribution of the data has lighter tails and the 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?

- 50% of the people are above 10 years old and remaining are less and students who's age is above 15 are approx. 40%.

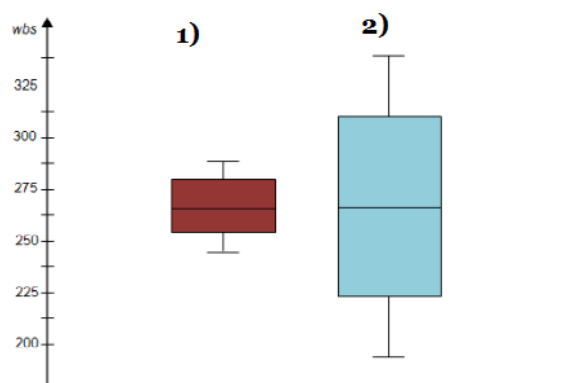
What is nature of skewness of the data?

- Left skewed as median is greater than mean.

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

- Approx = BETWEEN 10-18

Q19) Comment on the below Boxplot visualizations?



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

ANS – As in the boxplot the median is approx. which lies between 275 to 250 and here 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. $P(\text{MPG} > 38)$**
- b. $P(\text{MPG} < 40)$**
- c. $P(20 < \text{MPG} < 50)$**

ANS – A) $P(\text{MPG} > 38) = 1 - \text{pnorm}(38, 34.422, 9.13144) = 0.3475908$

B) $P(\text{MPG} < 40) = \text{pnorm}(40, 34.422, 9.13144) = 0.7293527$

C) $P(20 < \text{MPG} < 50) = \text{pnorm}(50, 34.422, 9.13144) - (1 - \text{pnorm}(20, 34.422, 9.13144)) = 0.01311818$

Q 21) Check whether the data follows normal distribution

ANS - a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

- MPG of cars 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

- Adipose tissue (AT) and Waist circumference (Waist) follows Normal distribution.

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

**ANS – For 90%,
 $\alpha = 5\% = 0.05$
Z at 0.05 is 1.645**

**For 94%,
 $\alpha = 3\% = 0.03$
Z at 0.03 is 1.555**

**For 60%,
 $\alpha = 20\% = 0.2$
Z at 0.2 is 0.253.**

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

ANS – Here the $n=25$

$$(df) = n - 1 = 25 - 1 = 24$$

Then on solving,

$$t \text{ score of } 95\% = 2.0638$$

$$t \text{ score of } 96\% = 2.1715$$

$$t \text{ score of } 99\% = 2.7969$$

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 → **pt(tscore,df)**

df → **degrees of freedom**

ANS – Here,

We have formula,

$$t = \frac{x - \mu}{(s/\sqrt{n})}$$

so on solving,

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

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

$$t = -0.471.$$

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