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

Q1) Identify the Data type for the Following:

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

Data	Data Type
Gender	Nominal
High School Class Ranking	Interval
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	Ratio
Blood Group	Nominal
Time Of Day	Interval
Time on a Clock with Hands	Interval

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

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

ANS) When three coins are tossed, there are 2^3=8 possible outcomes. They are

HHH,HHT,HTH,THH,HTT,THT,TTH,TTT

Now, we have to find the probability of getting two heads and one tail. There are three outcomes that satisfy this condition: HHT, HTH, and THH.

$$P = \frac{Total\ Number\ of\ Outcomes}{Number\ of\ Favorable\ Outcomes}$$

$$P = \frac{3}{8}$$

Therefore, the probability of getting two heads and one tail when three coins are tossed is $\frac{3}{8}$

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

ANS)

a) When two six-sided dice are rolled, the minimum sum you can get is 2 (rolling a 1 on both dice), and the maximum sum is 12 (rolling a 6 on both dice). There is no possible way to get a sum of 1 with two six-sided dice.

Therefore, the probability of getting a sum equal to 1 when two dice are rolled is 0.

b) The possible combinations for a sum less than or equal to 4 are (1, 1), (1, 2), (2, 1), (1, 3), (3, 1), and (2, 2). There are 6 favorable outcomes, so the probability is $\frac{6}{36} = \frac{1}{6}$

c) The sums divisible by 2 and 3 are 6 and 12. The combinations for each of these sums are (1,5), (2,4), (3,3), (4,2), (5,1), (6,6). Therefore required probability is $\frac{6}{36} = \frac{1}{6}$

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?

ANS)

The 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 = $7C_2 = 21$

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

n(E)= Number of ways of drawing 2 balls out of (2+3) balls = $5C_2 = 10$

Now, to find the probability that none of the balls drawn is blue

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

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*.005+6*.01+2*.0120
= 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)

	Points	Score	Weigh
mean	3.596563	3.217250	17.848750
median	3.695000	3.325000	17.710000
mode	3.92	3.44	17.840
variance	0.285881	0.957379	3.193166

std	0.534679	0.978457	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, 1 23, 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)

$E(X)$ = 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)

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

$$E(X) = \frac{1308}{9}$$

So, the expected value of the weights of patients at the clinic is $\frac{1308}{9}$ pounds, or approximately 145.33 pounds.

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)

a)

	Skewness	Kurtosis
Speed	-0.117510	-0.508994
Distance	0.806895	0.405053

Conclusion:

- Speed is Negatively skewed
- Distance is positively skewed
- Speed is Leptokurtic
- Distance is Leptokurtic

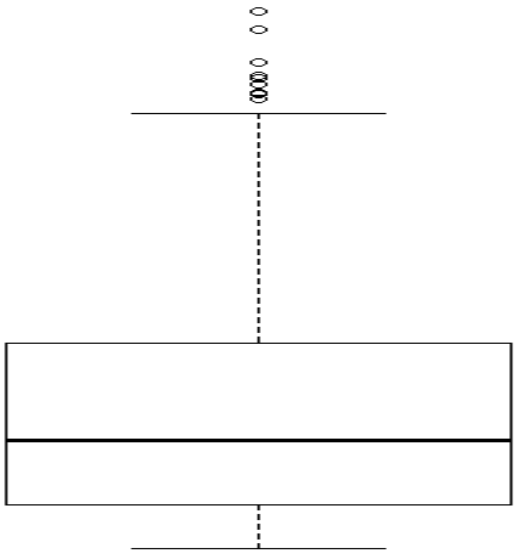
b)

	Skewness	Kurtosis
SP	1.611450	2.977329
WT	-0.614753	0.950291

Conclusion:

- SP is Positively skewed
- WT is Negatively skewed
- SP is Leptokurtic
- WT is Leptokurtic

Q10) Draw inferences about the following boxplot & histogram



ANS)

The histogram shows the greatest frequency in weight occurs between 150-200 and the weight that will least likely occur is between 350-400.

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)

Confidence Interval = Sample Mean ± Margin of Error

1) 94% Confidence Interval:

Critical Value ≈ 1.88(from a standard normal distribution table for a two-tailed test at 94% confidence).

Margin of Error = $1.88\times\frac{30}{\sqrt{2000}}$

upper bound = 201.2596
lower bound = 198.7404

2) 98% Confidence Interval:

Critical Value ≈ 2.33(from a standard normal distribution table for a two-tailed test at 98% confidence).

Margin of Error= $2.33\times\frac{30}{\sqrt{2000}}$

upper bound = 201.5611
lower bound = 198.4389

3) 96% Confidence Interval:

Critical Value≈ 2.576(from a standard normal distribution table for a two-tailed test at 96% confidence).

Margin of Error=$2.576 \times \frac{30}{\sqrt{2000}}$

upper bound = 201.72592

lower bound = 198.27408

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?

ANS)

Mean = $\frac{34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56}{18}$ = 41

Median = $\frac{40+41}{2}$ = 40.5

Variance = 25.52941

Standard deviation = 5.0526

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

ANS) Zero skewness

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

ANS) Positive skewness or Right skewness

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

ANS) Negative skewness or Left skewness

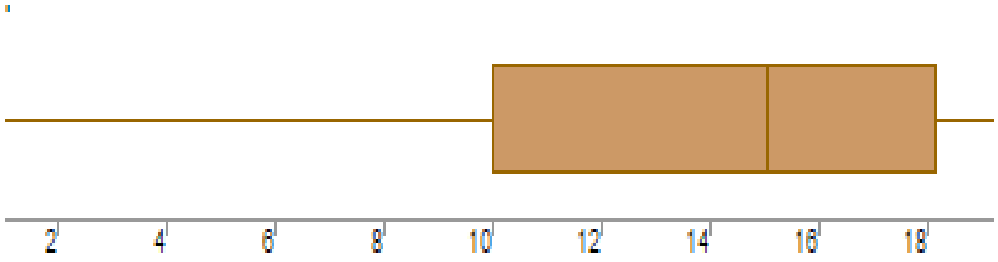
Q16) What does positive kurtosis value indicates for a data ?

ANS) Leptokurtic

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

ANS) Platykurtic

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



What can we say about the distribution of the data?

ANS) The longer part of the boxplot is to the left

What is nature of skewness of the data?

ANS) Left skewness or negative skewness

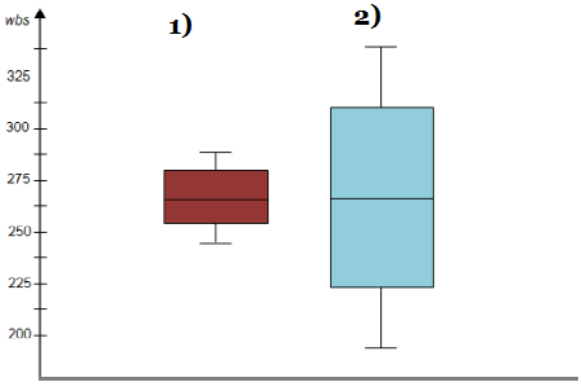
What will be the IQR of the data (approximately)?

ANS) IQR = Upper Quartile - Lower Quartile

= 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.

ANS)

- Boxplot 1 & 2 have the same median.
- Boxplot 2 have greater IQR than Boxplot 1

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)

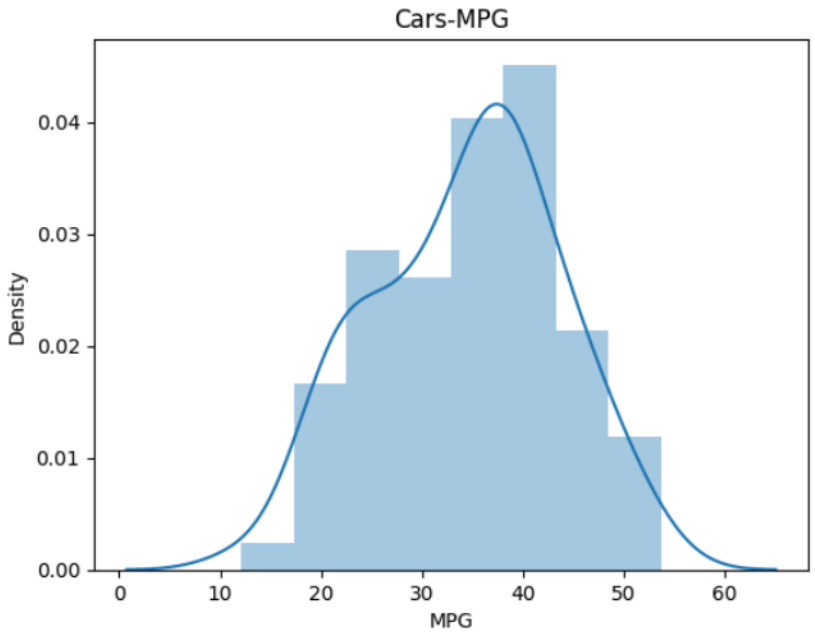
ANS)

a) P(MPG>38) = 34.75%
b) P(MPG<40) = 72.93%
c) P (20<MPG<50) = 89.88%

Q 21) Check whether the data follows normal distribution

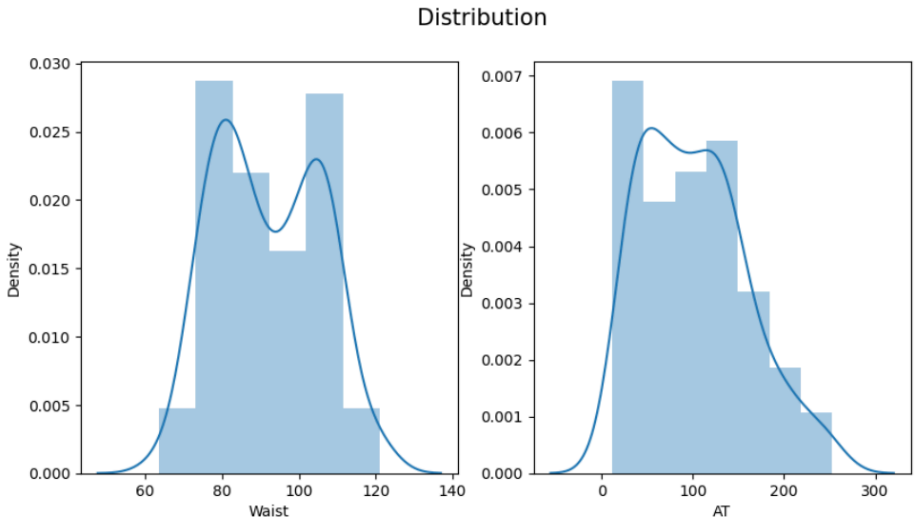
a) Check whether the MPG of Cars follows Normal Distribution
Dataset: Cars.csv

ANS)



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

ANS)



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

ANS)

Confidence interval	Z-score
90%	1.645
94%	1.880
60%	0.841

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

ANS)

Confidence interval	t-score
95%	2.063
96%	2.1715
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)

$\mu = 270$, $\bar{x}$ = 260 , n = 18, s = 90

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

=(260-270)/(90/sqrt(18))

=-10/21.23

=-0.47

Required probability = 0.319