|  |  |
| --- | --- |
| 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 | 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) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| 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?

Ans:-

P (Two heads and one tail) = N (Event (Two heads and one tail)) / N (Event (Three

coins tossed))

= 3/8 = 0.375 = 37.5%

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

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3

Ans:- Number of possible outcomes for the above event is

N (Event (Two dice rolled)) = 6^2 = 36

1. P (sum is Equal to 1) = ‘0’ zero null nada none.
2. P (Sum is less than or equal to 4) = N (Event (Sum is less than or equal to

4)) / N (Event (Two dice rolled))

= 6 / 36 = 1/6 = 0.166 = 16.66%

1. P (Sum is divisible by 2 and 3) = N (Event (Sum is divisible by 2 and 3)) / N

(Event (Two dice rolled))

= 6 / 36 = 1/6 = 0.16 = 16.66%

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 =7 balls

N (Event (2 balls are drawn randomly from bag) = 7! / 2! \* 5!

= (7\*6\*5\*4\*3\*2\*1) /

(2\*1) \* (5\*4\*3\*2\*1)

N (Event (2 balls are drawn randomly from bag) = (7\*6)/ (2\*1) = 21

If none of them drawn 2 balls are blue = 7 – 2 = 5

N (Event (None of the balls drawn is blue) = 5! / 2! \* 3! = (5\*4) / (2\*1)

= 10

P (None of the balls drawn is blue) = N (Event (None of the balls drawn is blue) /

N (Event (2 balls are drawn randomly from

bag)

= 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:- 0.015+0.8+1.95+0.025+0.06+0.24=3.09 so the expected number of candies for a randomly selected child is 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:** Mean:

Points: 3.5965625

Score: 3.2172500000000004

Weigh: 17.848750000000003

Median:

Points: 3.6950000000000003

Score: 3.325

Weigh: 17.71

Mode:

Points: 0 3.07

1 3.92

Name: Points, dtype: float64

Score: 0 3.44

Name: Score, dtype: float64

Weigh: 0 17.02

1 18.90

Name: Weigh, dtype: float64

Variance:

Points: 0.2858813508064516

Score: 0.9573789677419354

Weigh: 3.193166129032258

Standard Deviation:

Points: 0.5346787360709715

Score: 0.9784574429896966

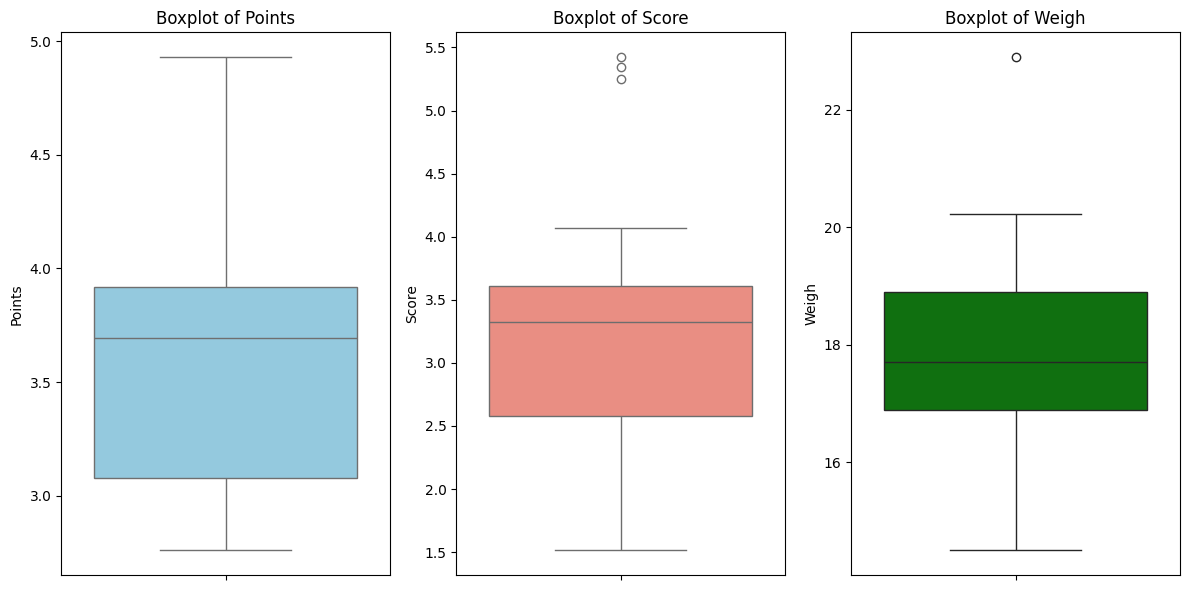
Weigh: 1.7869432360968431

Range:

Points: 2.17

Score: 3.9110000000000005

Weigh: 8.399999999999999



Q8) Calculate Expected Value for the problem below

1. 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:- Expected value = Sum (X \* Probability of X)

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

= 145.33

So the expected weight of a randomly chosen patient at the clinic is approximately 145.33 pounds.

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:** Calculation:

Skewness of Speed: -0.1175

Kurtosis of Speed: -0.5090

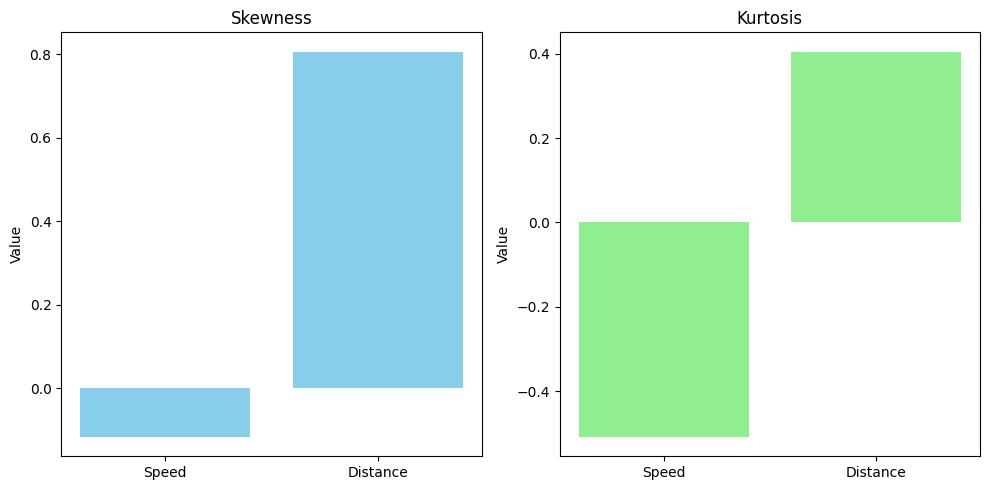
Skewness of Distance: 0.8069

Kurtosis of Distance: 0.4051

**Inferences**:

The speed variable has a slightly negative skewness, indicating a slight left-skewness in its distribution. The kurtosis value suggests that the speed distribution has relatively light tails.

In contrast, the distance variable exhibits positive skewness, indicating a right-skewed distribution. The positive kurtosis value suggests that the distance distribution has heavier tails compared to a normal distribution.



**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:** Calculation:

Skewness of Speed (SP): 1.5815

Kurtosis of Speed (SP): 2.7235

Skewness of Weight (WT): -0.6033

Kurtosis of Weight (WT): 0.8195

Inferences:

The skewness of the speed variable is positive (1.5815), indicating a right-skewed distribution. The kurtosis value (2.7235) suggests that the speed distribution has heavy tails compared to a normal distribution.

Conversely, the skewness of the weight variable is negative (-0.6033), suggesting a left-skewed distribution. The kurtosis value (0.8195) indicates that the weight distribution has lighter tails compared to a normal distribution.

**Q10) Draw inferences about the following boxplot & histogram**



Ans: The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.



Ans: The boxplot has outliers on the maximum 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?

**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:- 1) Mean: The mean of the scores is approximately 41.0

Median: The median of the scores is 40.5.

Variance: The variance of the scores is approximately 24.11

Standard deviation: The standard deviation of the scores is approximately 4.91

1. The student's scores average around 37.67, which means they're doing okay overall. But their scores vary quite a bit from test to test, showing that sometimes they do better or worse than usual.

Ans: we don’t have outliers and the data is slightly skewed towards right because mean is greater than median.

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

Ans:- When the mean and median of a dataset are equal, it indicates a symmetric distribution, and the skewness of the data is minimal or zero.

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

Ans:- When the mean is greater than the median, it indicates a positively skewed distribution.

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

Ans:- When the median is greater than the mean, it suggests a negatively skewed distribution.

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

Ans:- A positive kurtosis value indicates that the data has heavier tails and a sharper peak than a normal distribution

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

Ans:- A negative kurtosis value indicates that the data has lighter tails and a flatter peak compared to a normal distribution.

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



What can we say about the distribution of the data?

Ans: The above Boxplot is not normally distributed the median is towards the higher value

What is nature of skewness of the data?

Ans: The data is a skewed towards left. The whisker range of minimum value is greater than maximum

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

Ans: The Inter Quantile Range = Q3 Upper quartile – Q1 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: First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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

* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

Ans: The probabilities for the given cases are approximately:

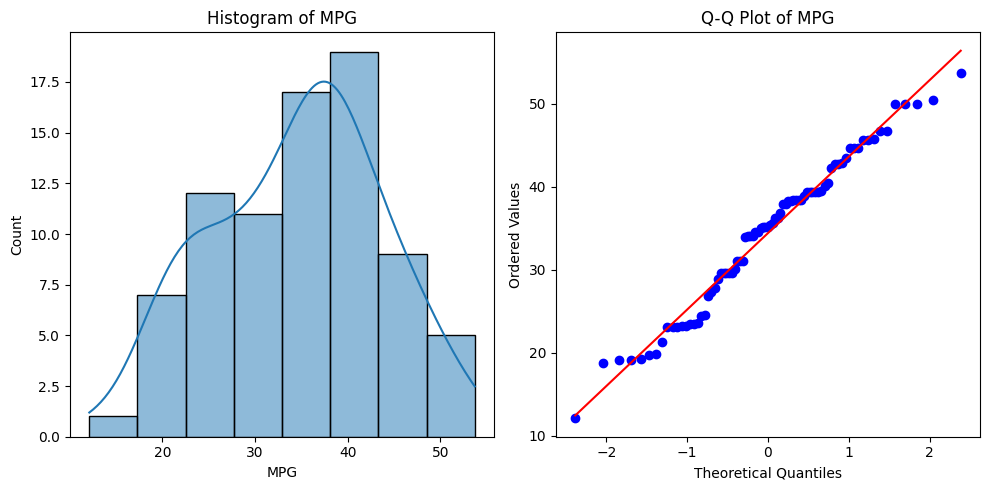
1. P(MPG > 38) ≈ 0.827
2. P(MPG < 40) ≈ 0.938
3. P(20 < MPG < 50) ≈ 0.975

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

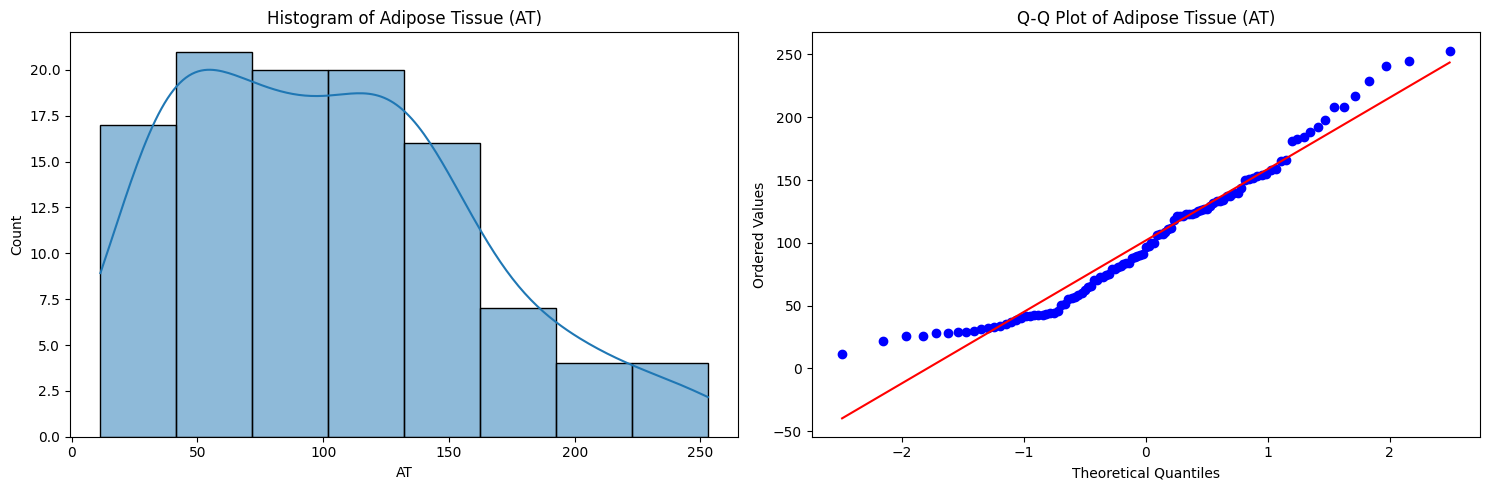
Ans:- a) MPG of cars follows normal distribution

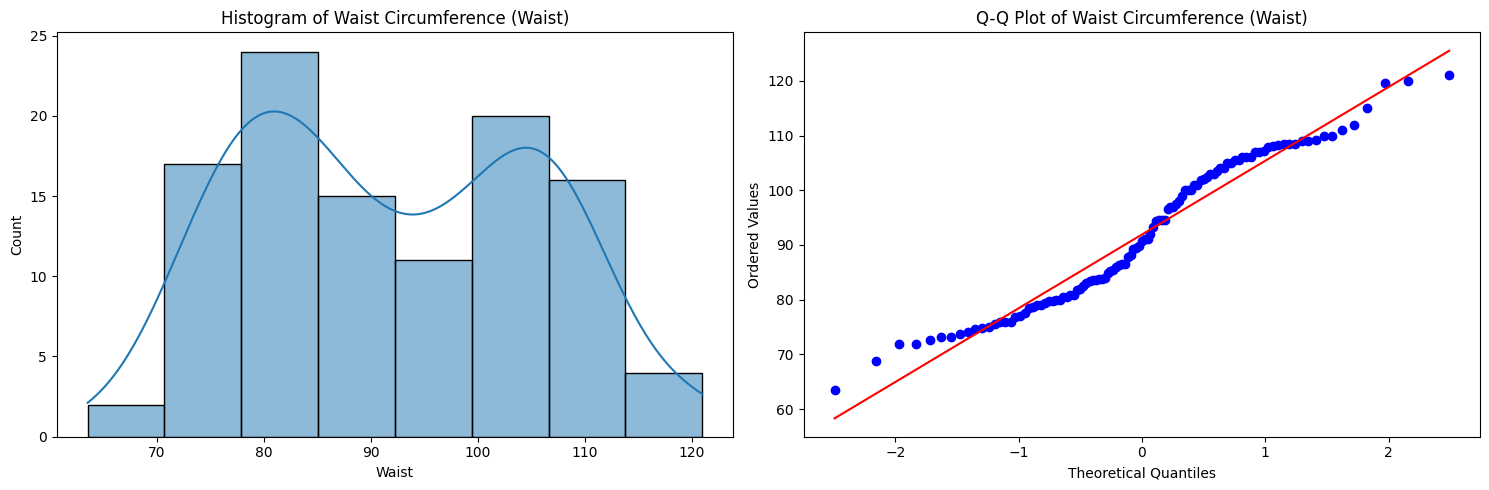


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans:- Adipose Tissue (AT) and Waist does not follow Normal Distribution





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

Ans:- Z score for 60% Conifidence Intervla = -1.6449

Z score for 60% Conifidence Intervla = -1.8808

Z score for 60% Conifidence Intervla = -0.8416

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

Ans:- T score for 95% Confidence Interval = -2.0639

T score for 94% Confidence Inteval = -1.974

T score for 95% Confidence Interval = -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**:-

The resulting probability is approximately 0.3217.

Therefore, there is about a 32.17% chance that 18 randomly selected bulbs would have an average life of no more than 260 days, given the CEO's claim.