|  |  |
| --- | --- |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical |

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 | Ordinal |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure |  |
| SAT Scores | Ratio |
| Years of Education | Nominal |

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

**Ans:**

The total number of possible outcomes when tossing three coins = 8

Outcomes =3(HHT, HTH, THH)

Probability=3/8 =0.375

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

When two dice are rolled, there are a total of 6 x 6 = 36 possible outcomes.

1. Equal to 1- There is no way to get a sum of 1 with two standard dice since the minimum value on a single die is 1. So, the probability is 0.
2. Less than or equal to 4- Outcomes = 6 ((1, 1), (1, 2), (2, 1), (1, 3), (2, 2), (3,1))

Probability = 6/36 = 0.166

1. Sum is divisible by 2 and 3- Outcomes = 6 ((1,5), (2,4), (3,3), (4,2), (5,1),

(6,6)

Probability = 6/36 = 0.166

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:** Number of ways to choose 2 non-blue balls from the 5 non-blue balls: C (5, 2) = 5! / [(2!) (3!)] = 10 ways. Number of ways to choose 2 balls from all 7 balls: C (7, 2) = 7! / [(2!) (5!)] = 21 ways

Probability (none of the balls drawn is blue) = (Number of ways to choose 2 non-blue balls) / (Number of ways to choose 2 balls from all 7 balls)

Probability (none of the balls drawn is blue) = 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 = (0.015 \* 1) + (0.20 \* 4) + (0.65 \* 3) + (0.005 \* 5) + (0.01 \* 6) + (0.120 \* 2) = 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.

**Ans** Points Score Weigh

Mean 3.596563 3.217250 17.848750

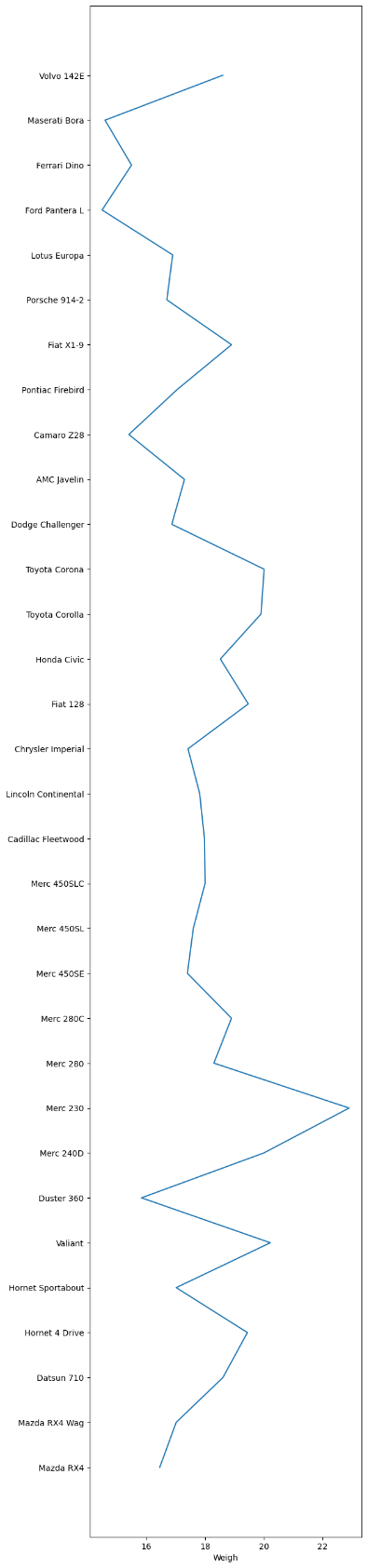
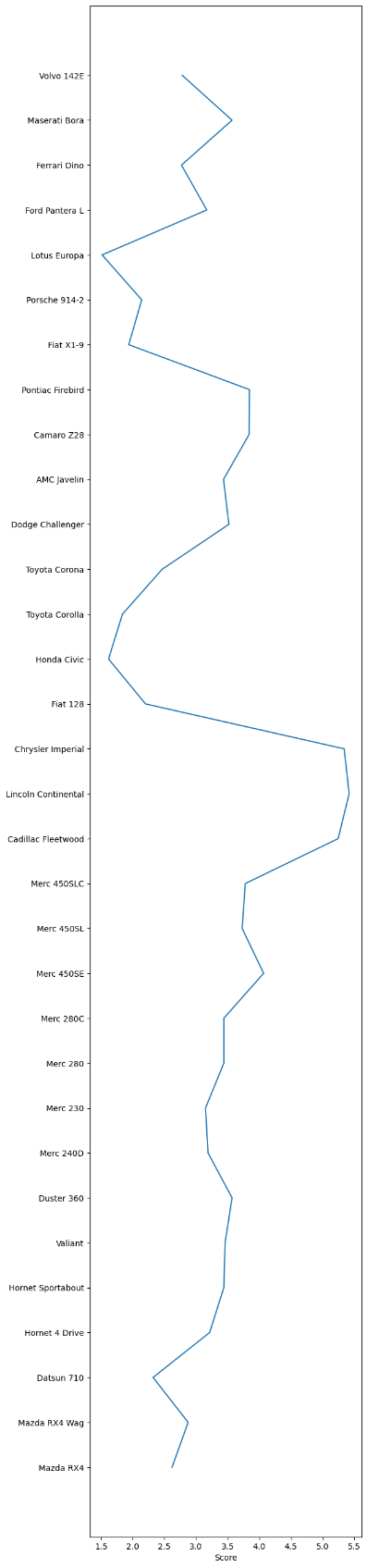
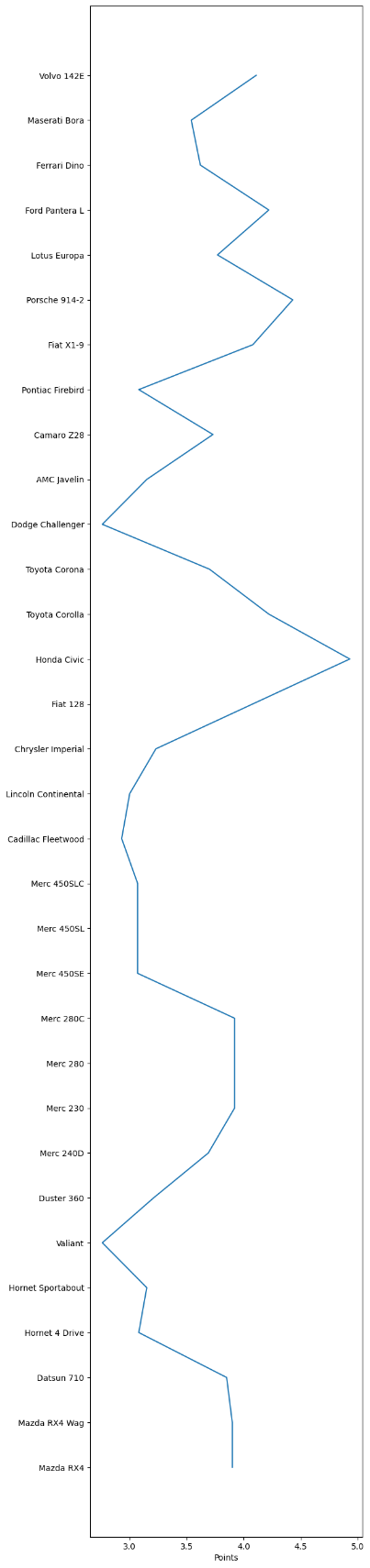
Std 0.534679 0.978457 1.786943

Var 0.28588 0.95738 3.16

Range 2.17 3.91 8.4

Med 3.69500 3.325 17.71

Mode 3.92 3.44 17.02



(a) (b) (c)

Figure 1

In Figure 1 (a) represents, Honda Civic car has the max points and valiant has the minimum points, In Figure 1 (b), Lincoln Continental has the max Score and Lotus Europa has the minimum Score, Meanwhile in Figure 1 (c), Merc 230 has the maximum and Ford Pantera L has the minimum Weight.

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** Possible outcomes = 9

Probability = 1/9

Expected Values= 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)=

145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans**

Skewness of speed -0.12, Skewness of cars dist 0.81

Kurtosis for cars dist: 0.25, Kurtosis for cars speed: -0.57

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans**

Skewness of SP 1.61, Skewness of WT -0.61.

Kurtosis for SP: 2.72, Kurtosis for WT: 0.82

**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?

Ans

For 94% confidence interval Range is [ 198.73 – 201.26]

For 96% confidence interval range is [198.62 – 201.38]

For 98% confidence interval range is [198.44 – 201.56]

**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.

Ans

Mean-41, Standard deviation- 5.05, Variance -25.53, Median-40.5

What can we say about the student marks?

Ans

Students marks ranges between 34-56, where the average is 41 , highest mark obtained is 56 and lowest is 34

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

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

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

Q16) What does positive kurtosis value indicates for a data – The curve is Peaked

Q17) What does negative kurtosis value indicates for a data? The curve is flat and broader.

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?



Ans: No outliers Present. 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

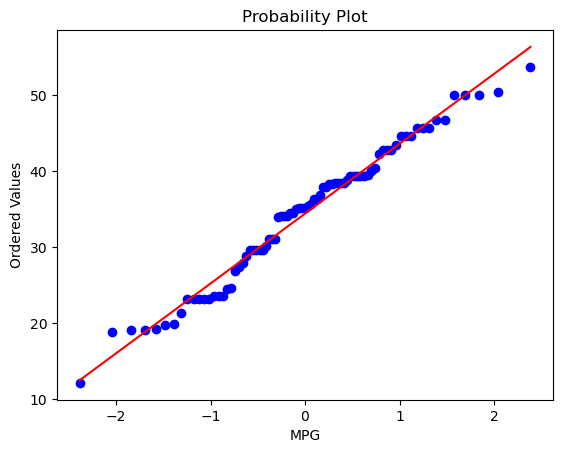
* 1. P(MPG>38) = 0.348
  2. P(MPG<40) =0.729

c. P(20<MPG<50) =0.013

Q 21) Check whether the data follows normal distribution

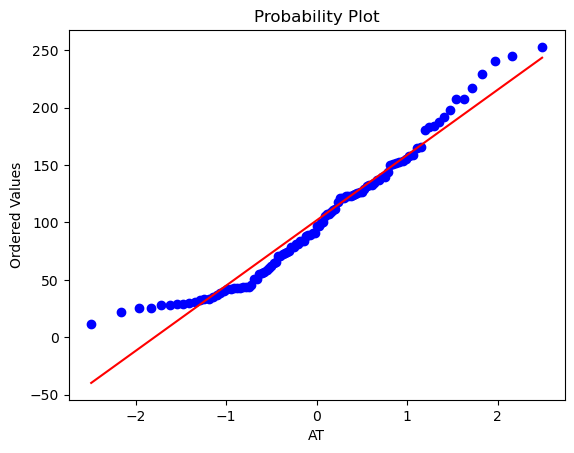
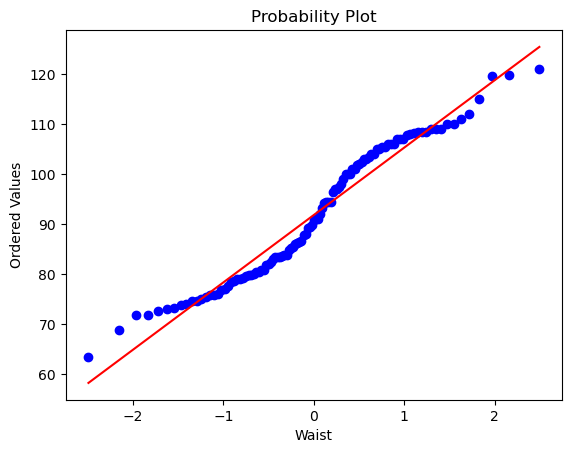
1. Check whether the MPG of Cars follows Normal Distribution

**Ans, 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

**Ans;** Both Doesn’t follow Normal Distribution



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

**Ans;**

Z score for 90% Confidence Interval = -1.6449

Z score for 94% Confidence Interval = -1.8808

Z score for 60% Confidence Interval = -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 96% Confidence Interval = -1.8281

T score for 99% 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;**

tscore = -0.471

stats.t.cdf(tscore, df = 17)

0.32 = 32%