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

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

Ans =

Total possible outcome = [HHH, HHT, HTH, THH, TTH, THT, HTT, TTT]

 two heads and one tail = 3

i.e., HHT, HTH, TTH

= 3/8 = 0.375

= 37.5%

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

Total possible outcome = 6 \* 6 = 36

1. Equal to 1

Ans = P(sum is equal to 1) = 0

1. Less than or equal to 4

Ans = 6/26 = 1/6 = 0.16

= (16.66%)

1. Sum is divisible by 2 and 3

Ans = 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 balls = 2 + 3 + 2 = 7

Let S be the sample space, n(S) = number of ways drawing 2 balls out of 7

= 7 C 2

= (7 \* 6) / (2 \* 1)

= 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.

= 5C2

= (5 \* 4) / (2 \* 1)

= 10

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

1. Mean

Points mean : 3.5965625

Score mean : 3.2172500000000004

Weigh mean : 17.848750000000003

1. Median

Points median : 3.6950000000000003

Score median : 3.325

Weigh median : 17.71

1. Mode

Points mode : 3.92

Score mode : 3.44

Weigh mode : 17.02

1. Variance

Points variance : 0.2858813508064516

Score variance : 0.9573789677419355

Weigh variance : 3.1931661290322575

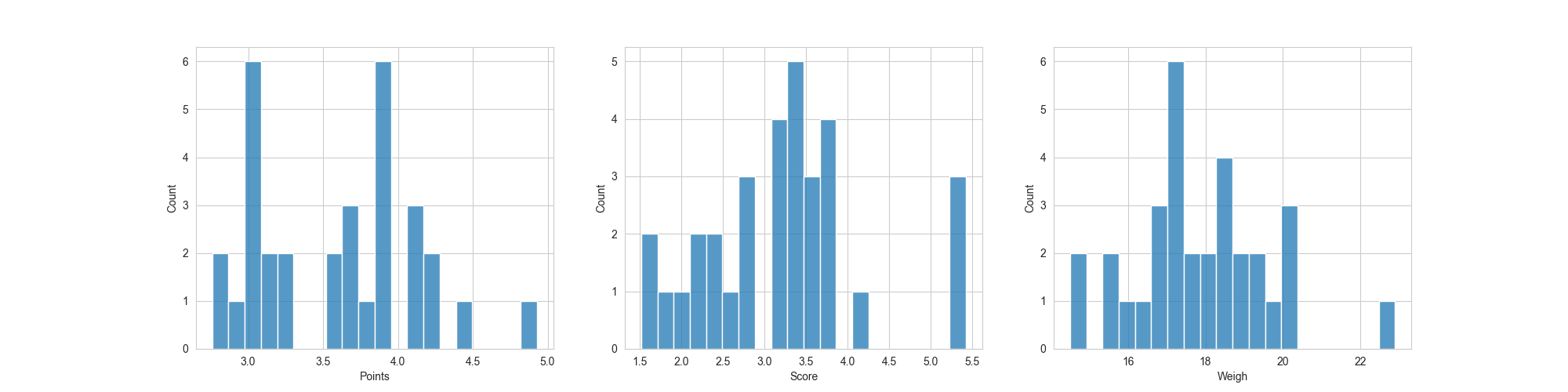
1. Standard deviation

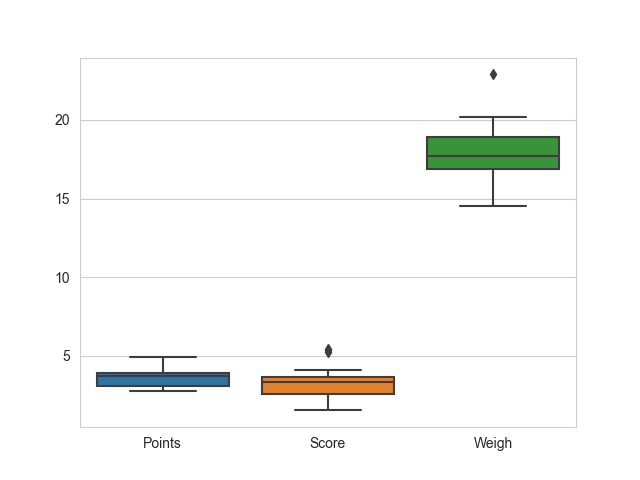
Points standard deviation : 0.5346787360709715

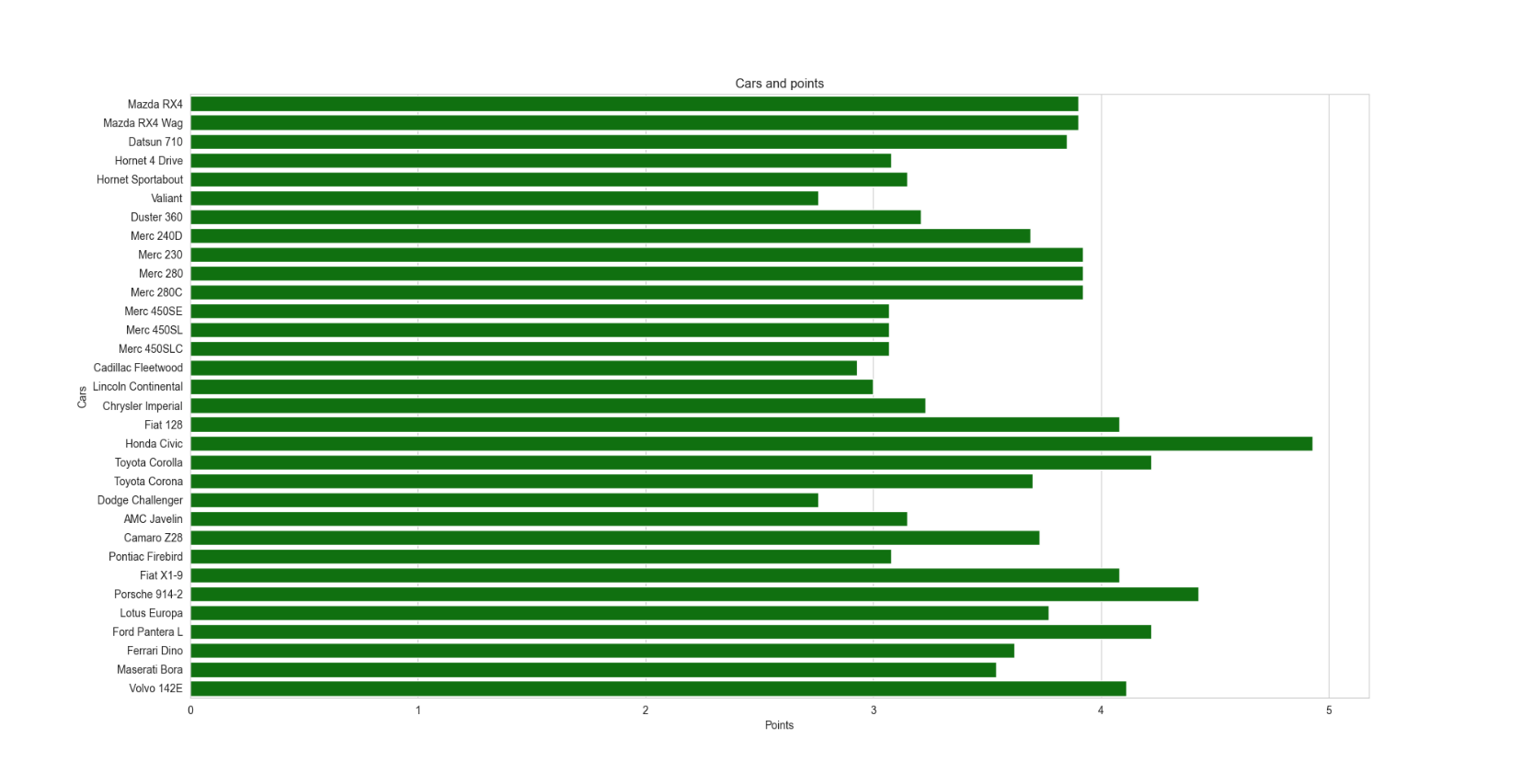
Score standard deviation : 0.9784574429896967

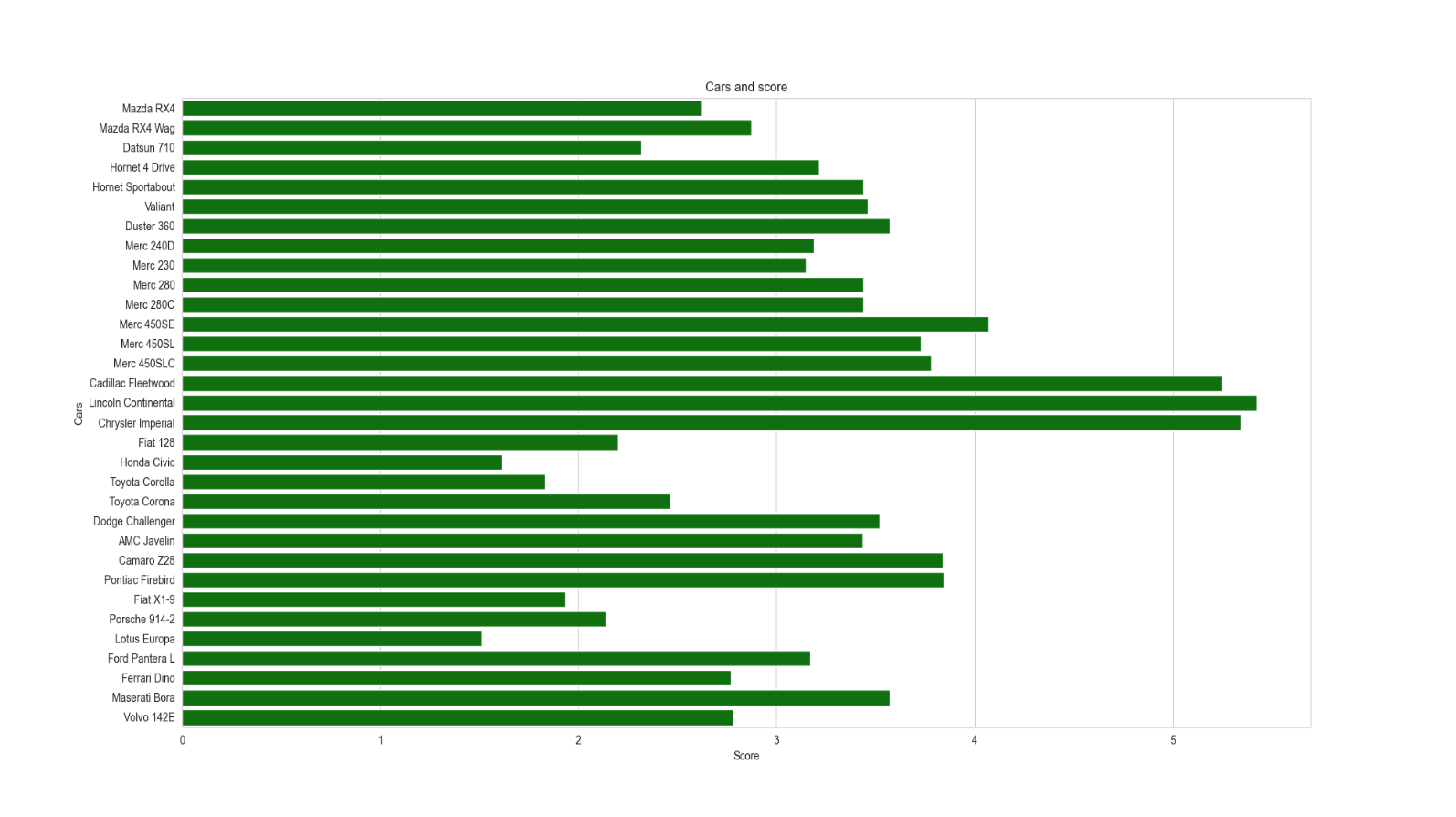
Weigh standard deviation : 1.786943236096843

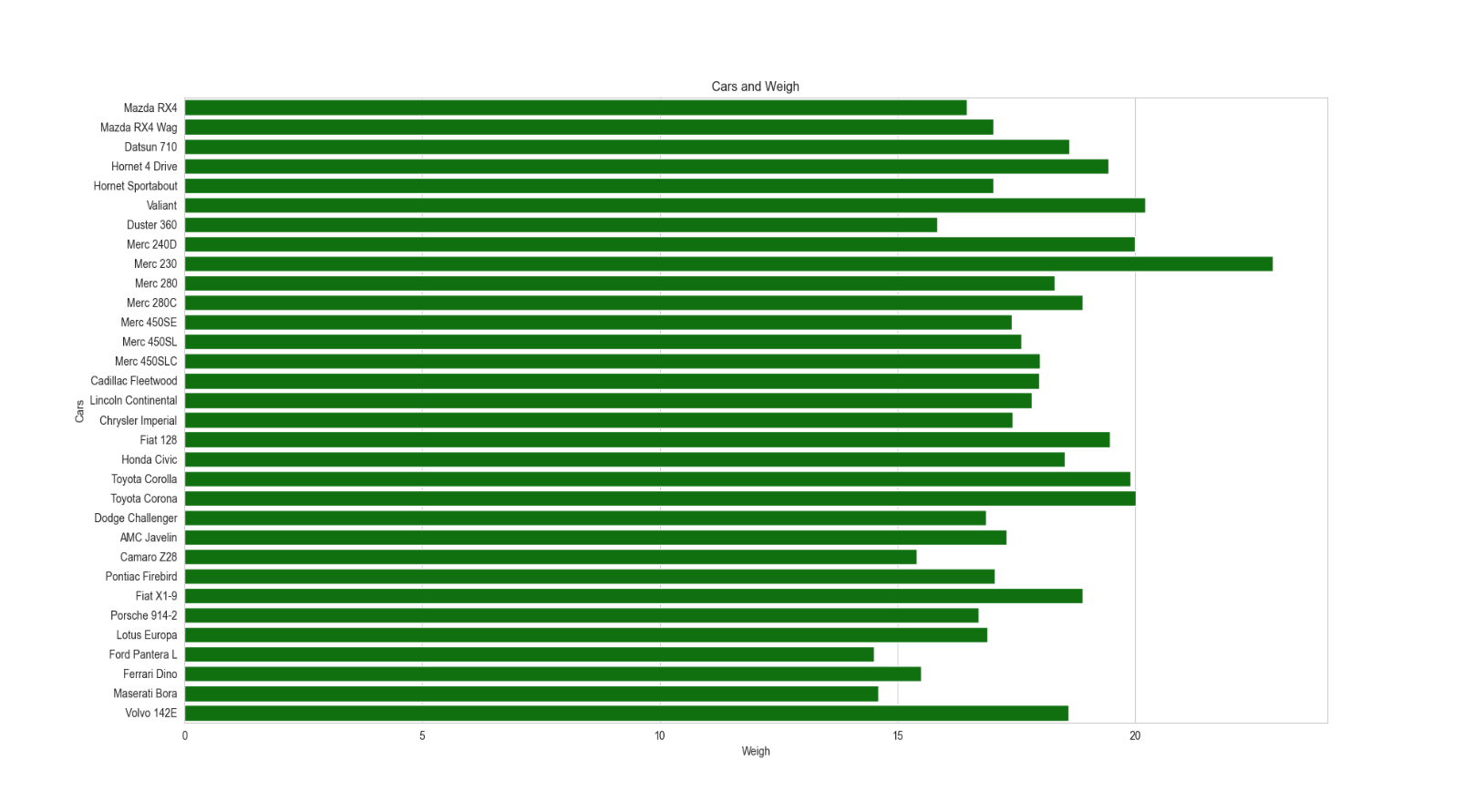
1. Statistical Inference











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

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

**Car’s speed and distance**

**Use Q9\_a.csv**

Ans =

Speed Skewness : -0.12

Speed Kurtosis : -0.51

dist Skewness : 0.81

dist Kurtosis : 0.41

**SP and Weight (WT)**

**Use Q9\_b.csv**

Ans =

SP Skewness : 1.61

SP Kurtosis : 2.98

WT Skewness : -0.61

WT Kurtosis : 0.95

**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)** 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 =

1. 94% CI = (198.7376089443071, 201.2623910556929)
2. 98% CI = (198.4381860483216, 201.5618139516784)
3. 96% CI = (198.6214037429732, 201.3785962570268)

**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 : 40.5

variance : 25.529411764705884

standard deviation: 5.05266382858645

1. What can we say about the student marks?

Ans =

1) 41% marks have fallen to more students

2) Between 35% and 45% the maximum number of marks has been scored

by the student

1. A single student has scored 56%

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

Ans = No skewness is present we have a perfect symmetrical distribution.

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

Ans = Skewness and tail is towards right.

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

Ans = Skewness and tail is towards left

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

Ans = Positive kurtosis means the curve is more peaked and it is leptokurtic

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

Ans = Negative kurtosis means the curve will be flatter 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?



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) = 0.348
  2. P(MPG<40) = 0.729

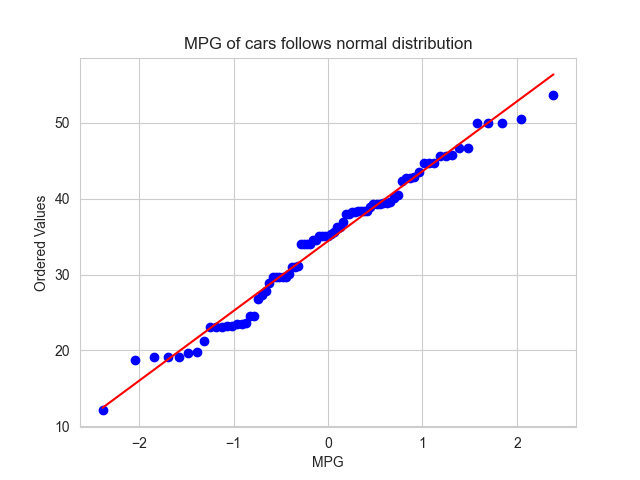
c) P(20<MPG<50) = 0.013000000000000012

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

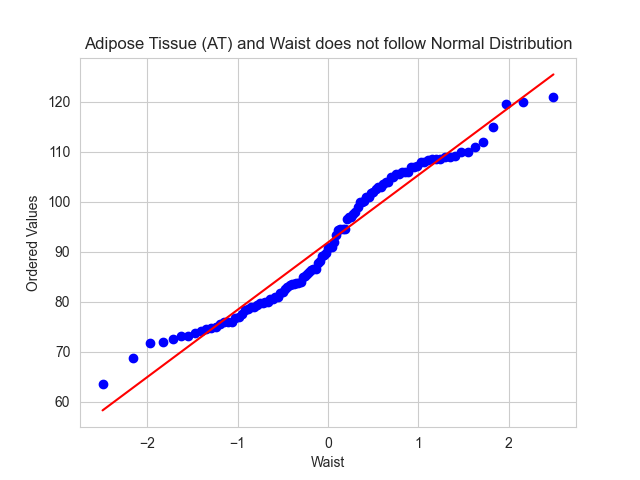
Ans = MPG of cars follows normal distribution



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

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 =

1) Z\_Score 90% CI = -1.6449

2) Z\_score 94% CI = -1.8808

3) Z\_score 60% CI = -0.8416

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

Ans =

1) T\_score 95% CI = -2.0639

2) T\_score 96% CI = -1.974

3) T\_score 99% CI = -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

Ans =

sample\_mu = 260

pop\_mu = 270

sample\_sigma = 90

n = 18 # no of sample

t\_score = (sample\_mu - pop\_mu)/(sample\_sigma/n\*\*0.5)

print('T\_score :',t\_score)

print ()

# Find P (18 >= 260) for null hypothesis

P\_18\_is\_gretter\_260 = 1 - stats.t.cdf (abs (-0.4714), df=17)

Print ('P (18 >= 260) :',P\_18\_is\_gretter\_260)

T\_score : -0.4714045207910317

P (18 >= 260): 0.32167411684460556