|  |  |
| --- | --- |
| 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 | Interval |
| 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 | ratio |

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

1. 3/8 =0.375

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

a) zero

b) 1/6 =0.167

c) 1/6 =0.167

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?

1. 10/21 = 0.476

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 |

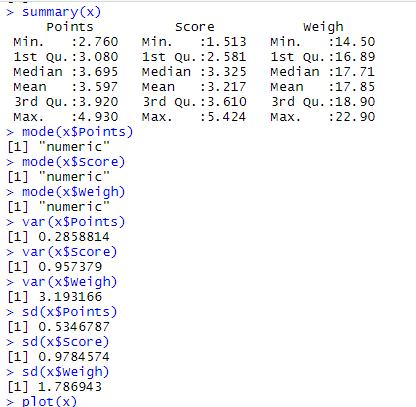
Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

1. 3.09

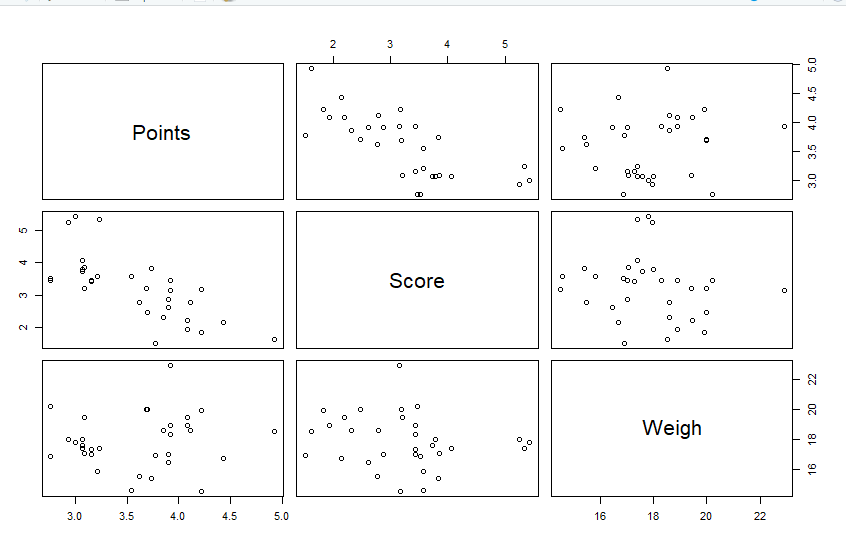
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

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

Inference drawn:

* Mean, Median are nearly equal for all the variables hence given data is normal





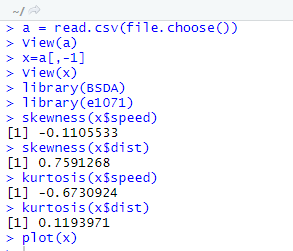
Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

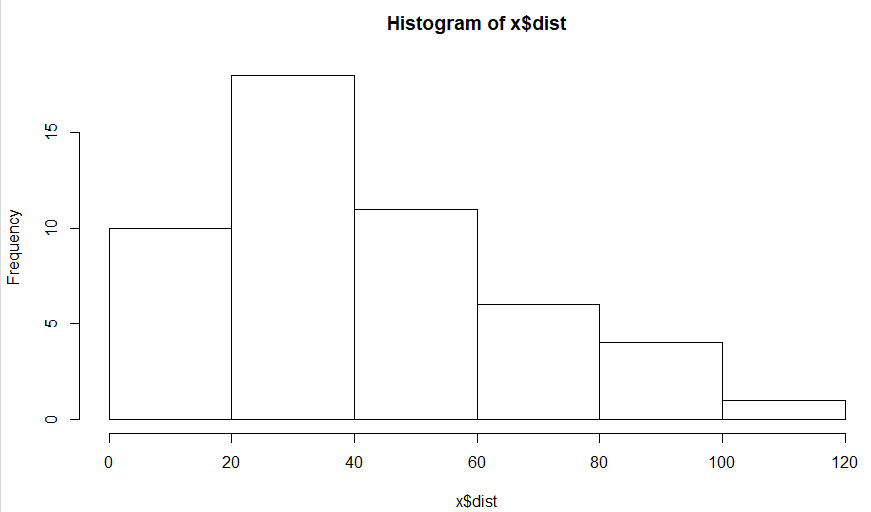
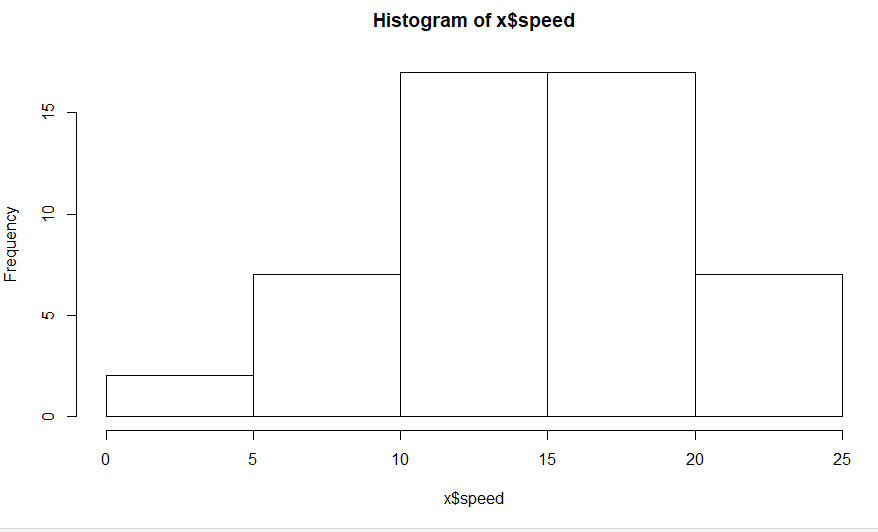
108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

A) 145.33

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

**Cars speed and distance**

****

**Inferences drawn:**

* **Speed:**

**Speed has negative skewness**

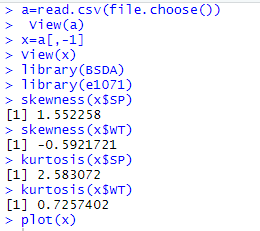
**Kurtosis value is less than 1 hence negative kurtosis**

* **Distance:**

**Distance has positive skewness**

**Kurtosis value is greater than 1 hence data is normal**

**SP and Weight(WT)**

** Inferences drawn:**



* **Wt:**

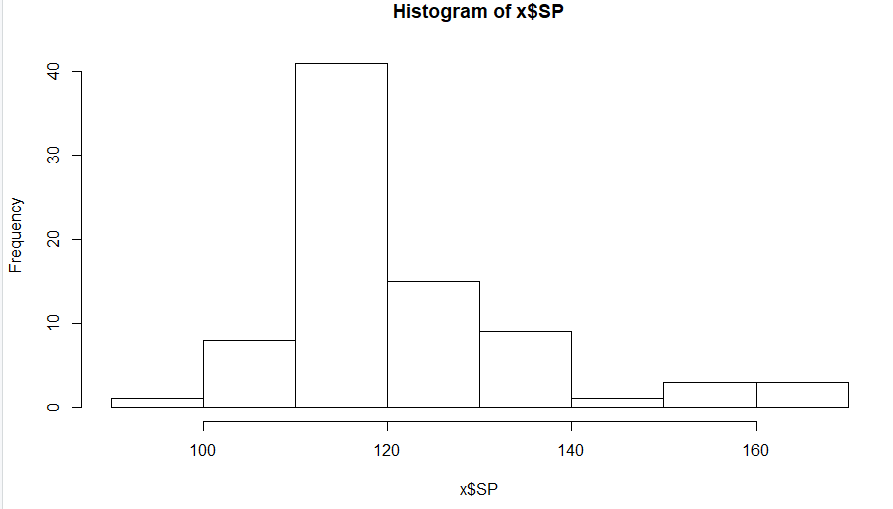
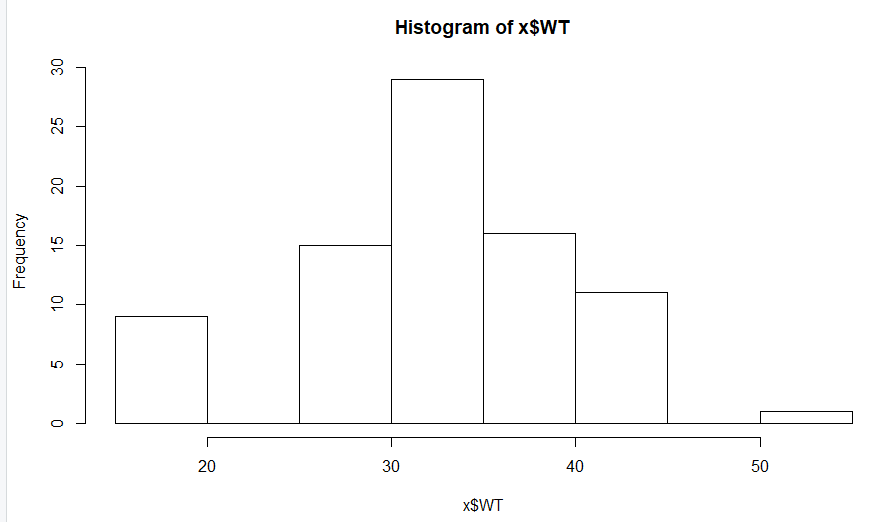
**Skewness is negative**

**Kurtosis is positive**

* **Sp:**

**Skewness is positive**

**Kurtosis is positive**

****

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

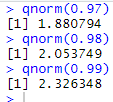


1. Positive skewness/Right skewed
2. Mode lies between 50 and 100
3. There are no outliers



1. Positive skewness
2. There are more number of outliers
3. The distribution of data is more concentrated in upper whisker

**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 ?



**94% Confidence interval**

X-bar = 200,Sd = 30,n = 2000

Interval Estimate = X-bar ± Z\*Sd/sqrt(n)

=200 ± 1.88\*30/sqrt(2000)

=198.74 to 201.26

**96% Confidence interval**

X-bar = 200,Sd = 30,n = 2000

Interval Estimate = X-bar ± Z\*Sd/sqrt(n)

=200 ± 2.05\*30/sqrt(2000)

=198.62 to 201.38

**98% Confidence interval**

X-bar = 200,Sd = 30,n = 2000

Interval Estimate = X-bar ± Z\*Sd/sqrt(n)

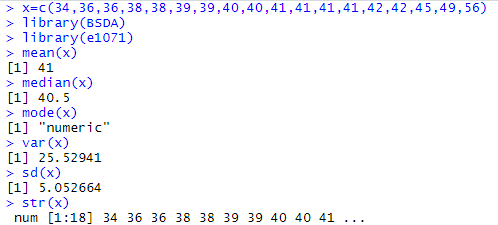
=200 ± 2.33\*30/sqrt(2000)

=198.44 to 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**

1. Find mean, median, variance, standard deviation.
2. What can we say about the student marks?



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

1. No skewness

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

1. Right skewed / positive skewness

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

1. Left skewed / negative skewness

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

1. Heavier tail and sharper peaks

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

1. Lighter tails and flatter peaks

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

What will be the IQR of the data (approximately)?   
  
A) 1.the distribution of data is more concentrated near lower whisker

2. left skewed/ negative skewness

3.IQR = (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.

1. 1. Median for both is 262.5

2. the data is normally distributed for both plots

3. 2nd box plot have large no data than 1st box plot

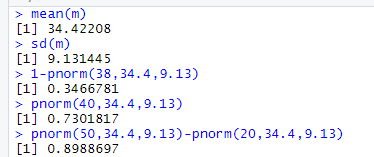
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

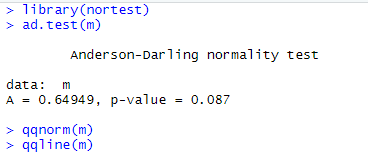
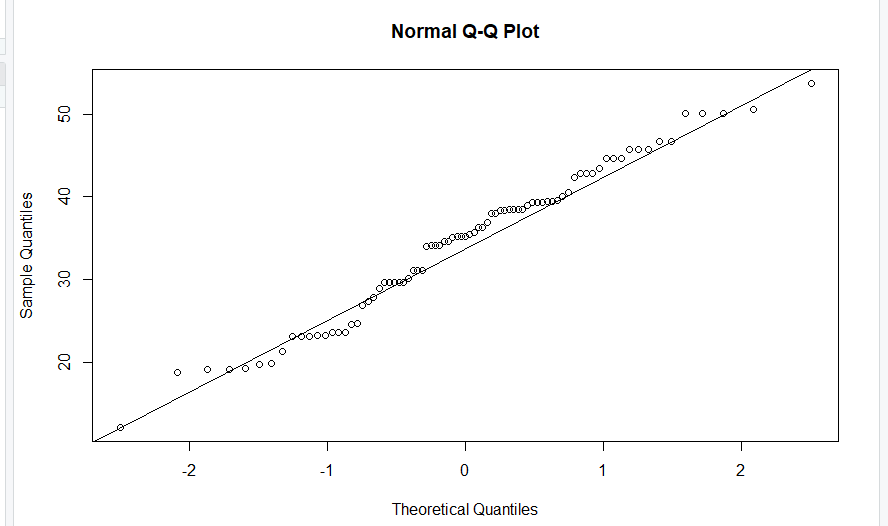
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

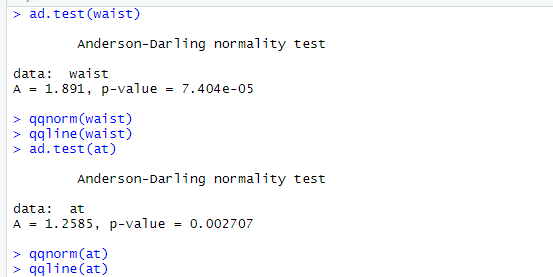
1. Check whether the MPG of Cars follows Normal Distribution

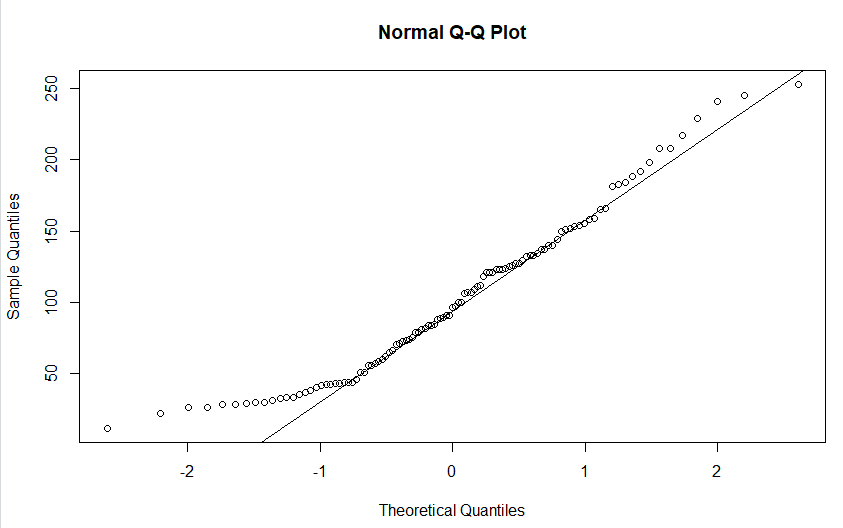
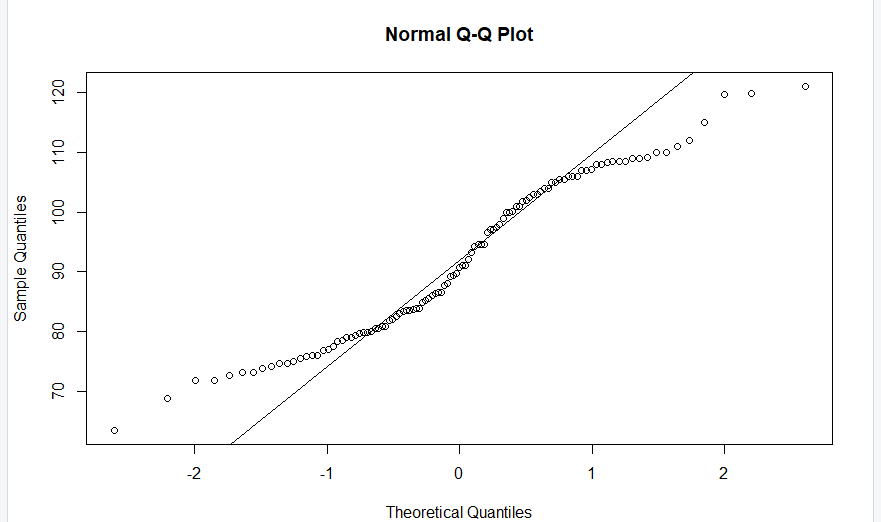
Dataset: Cars.csv



Data is normally distributed

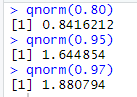
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution



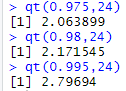


1. Waist data is not normal
2. At data is normal

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



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



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:

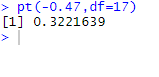
rcode 🡪 pt(tscore,df) df 🡪 degrees of freedom

n=18 , xbar = 260 , s=90 , mu =270

tscore = (xbar – mu )/ sigma

=(xbar -mu)/(s/sqrt(n))

=(260-270)/(90/sqrt(18)) =-0.47



Probability of selected bulbs lifetime not more than 260 days is 32%