|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Discrete |
| Weight of Gold | Discrete |
| Distance between two places | Discrete |
| Length of a leaf | Discrete |
| Dog's weight | Discrete |
| Blue Color | Continuous |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Continuous |

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 | Interval |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Interval |
| Type of living accommodation | Nominal |
| Level of Agreement | Nominal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal |
| Years of Education | Interval |

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

H,H,T ; T,H,H; H,T,H (3/8)

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 :) a – 0, b – 6/36 , c- 2/36

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: 7c2 = 21 , 5c2 = 10 (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

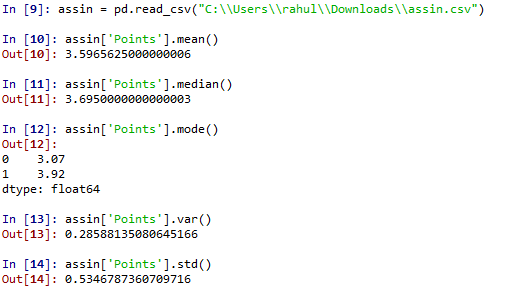
Ans: **3.090** [1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120]

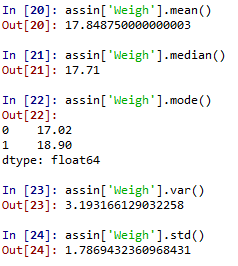
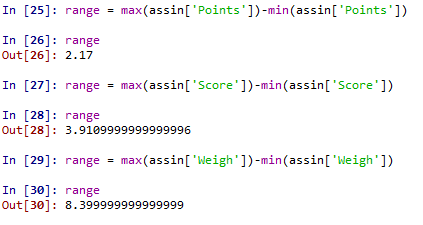
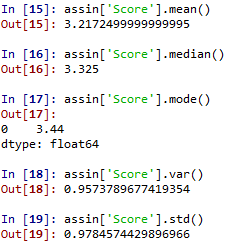
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

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





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

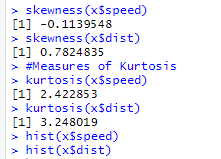
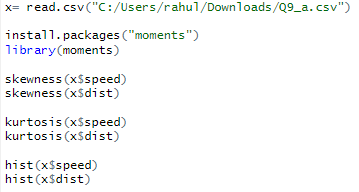
Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

|  |  |  |  |
| --- | --- | --- | --- |
|  | x | f(x) | E(x) |
|  | 108 | 0.083 | 8.964 |
|  | 110 | 0.084 | 9.24 |
|  | 123 | 0.094 | 11.562 |
|  | 134 | 0.102 | 13.668 |
|  | 135 | 0.103 | 13.905 |
|  | 145 | 0.111 | 16.095 |
|  | 167 | 0.128 | 21.376 |
|  | 187 | 0.143 | 26.741 |
|  | 199 | 0.152 | 30.248 |
| Total | 1308 | 1 | E(x) =151.799 |

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

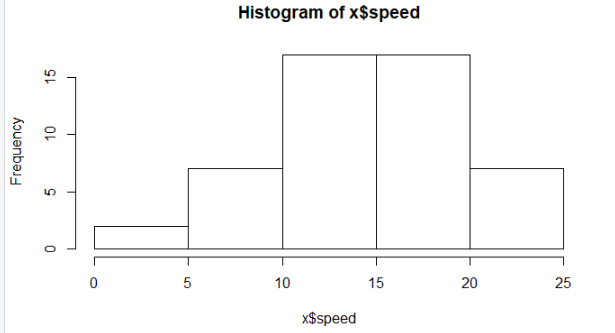
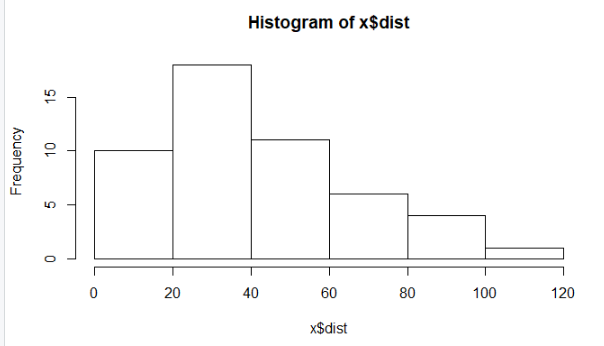
**Cars speed and distance**





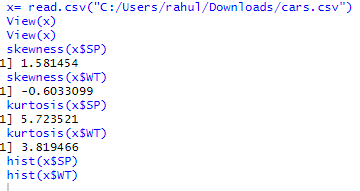
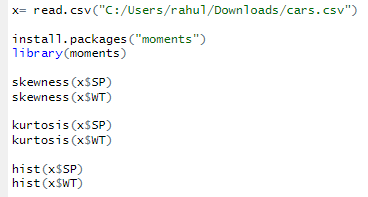
As we can see the speed is negatively skewed means mass of the data in this is lying on the right hand side, as we can see in the histogram given below. For distance it is vice versa.

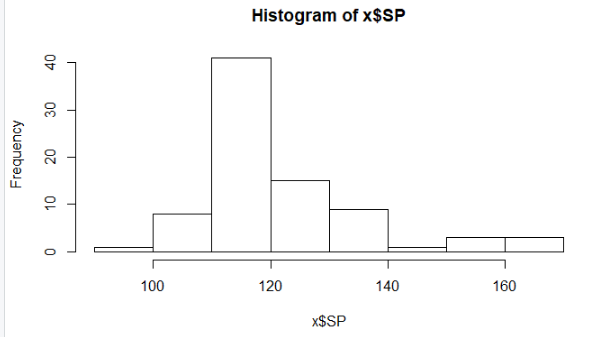
Kurtosis – measures the peakedness of the distribution.



**SP and Weight(WT)**









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

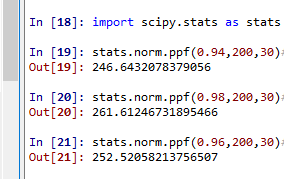


* **From the above histogram we can deduce that the mass of the data is on the left hand side and tail is on the right hence it is positive skewed data.**



* **From the Box Plot we can see that there are many outliers in the data**

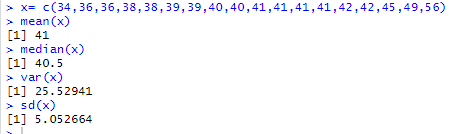
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 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?

1: 

2: The marks are consistently increasing.

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

Ans: Then it’s a perfectly symmetrical data, the mass of the data will be in center.

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

Ans: It is positively skewed.

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

Ans: It is Negatively skewed.

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

Ans: It indicates that the data is more peaked

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

Ans: It indicates that the data is not peaked in nature.

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



1. What can we say about the distribution of the data?
2. What is nature of skewness of the data?
3. What will be the IQR of the data (approximately)?

Ans:

1. The mass of the data is concentrated on the right hand side.
2. It is negatively skewed.
3. Q1 = 10, Q3 = 18, => Q3 – Q1 => 18 – 10 = 8

Q19) Comment on the below Boxplot visualizations?

* Both the plots show similar medians.
* Plot 2 has more variation than plot 1.
* There are no outliers in both plots.



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

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

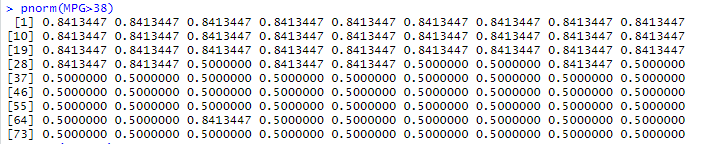
Data \_set: Cars.csv

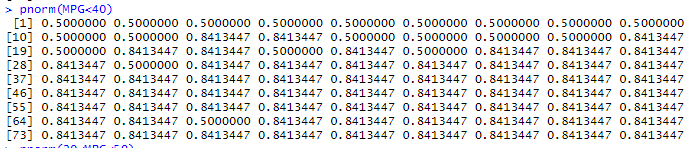
Calculate the probability of MPG of Cars for the below cases.

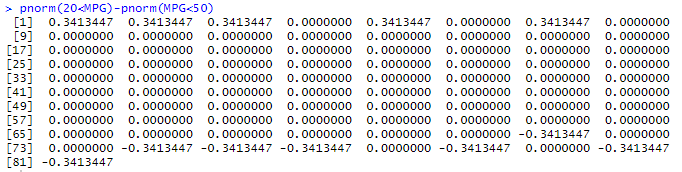
MPG <- Cars$MPG

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

c. P (20<MPG<50)

a. 

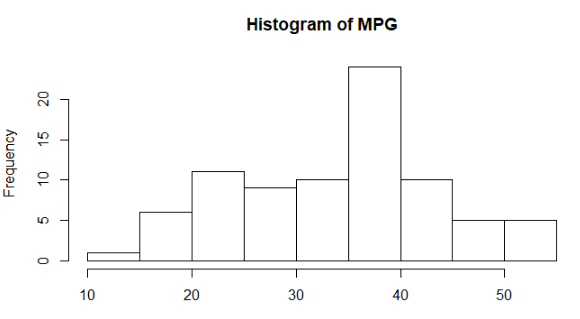
b. 

c. 

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

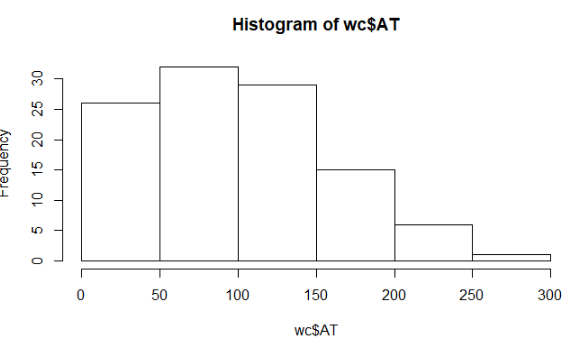
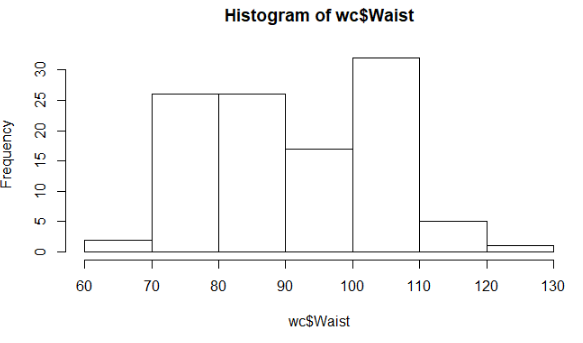
Dataset: Cars.csv



From the above histogram we can deduce that the data is not normally distributed.

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

Dataset: wc-at.csv



From the above histogram we can deduce that the data is not normally distributed.

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

qnorm(0.90) = 1.281552

qnorm(0.94) = 1.554774

qnorm(0.60) = 0.2533471

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

qt(0.95,25) = 1.708141

qt(0.96,25) = 1.824828

qt(0.99,25) = 2.485107

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

Sd = 90

Xavg = 260

df = n-1 = 17

qt(0.950,17)

[1] 1.739607