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

**Ans : S{HHH, HHT, HTH, THH, TTH, THT, HTT,TTT}**

**P(X) = 3/8 = 0.375**

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

1. Equal to 1 **: P(X1) = 0**
2. Less than or equal to 4 **: P(X2) = 6/36 = 0.166**
3. Sum is divisible by 2 and 3 **: P(X3) = 24/36 = 0.66**

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 : P(X) = 10/21 = 0.47**

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 : E(X) = (1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.001)+(2\*0.120)**

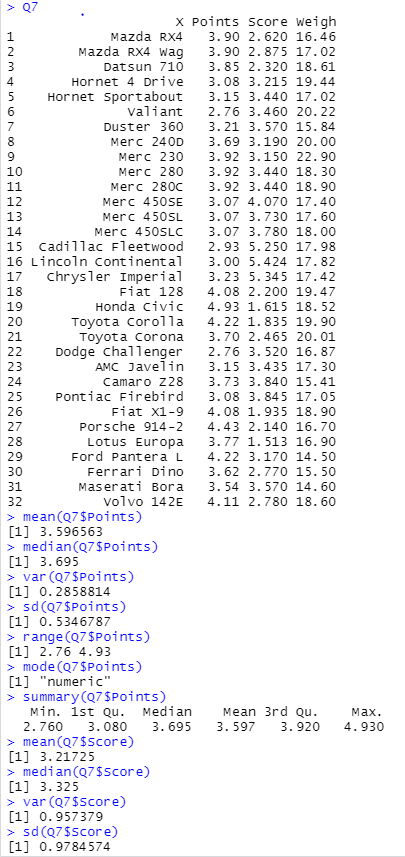
**= 3.09**

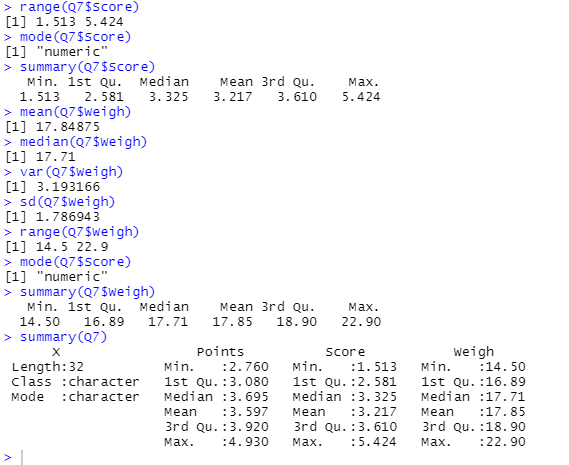
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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**

****

****

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?

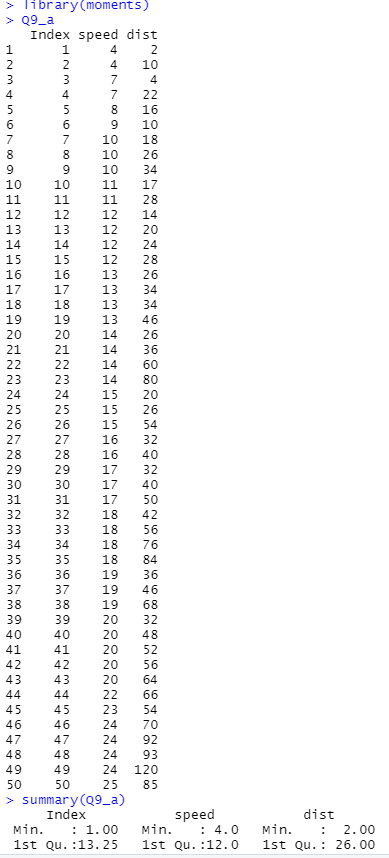
**Ans : E(X) = (108+110+123+134+135+145+167+187+199)**

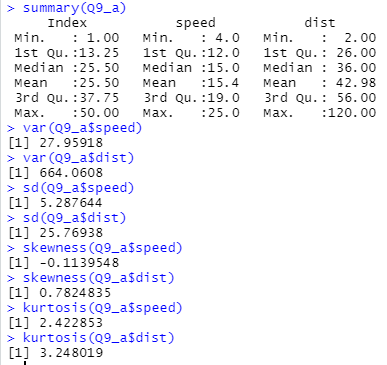
**=145.33**

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

Cars speed and distance

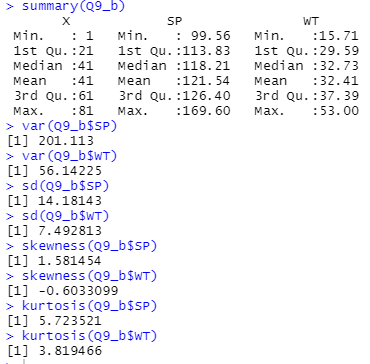
Use Q9\_a.csv





SP and Weight(WT)

Use Q9\_b.csv



Q10) Draw inferences about the following boxplot & histogram



**Ans :- From the above histogram we conclude that the data is right-Skewed.**

**It might have some outliers. Its median lies before its mean value and after its mode value.**



**Ans :- The data is compressed & and has outliers. It is Positively Skewed boxplot.**

**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: n=2000

= 200

s= 30

Confidence Interval Estimate= Z => 200 Z

94% Confidence: qnorm(0.97)

[1] 1.880794=Z

200 1.88\* =**198.74 – 201.26**

98% Confidence: > qnorm(0.99)

[1] 2.326348=Z

200 2.33\* =**198.44-201.56**

96% Confidence: > qnorm(0.98)

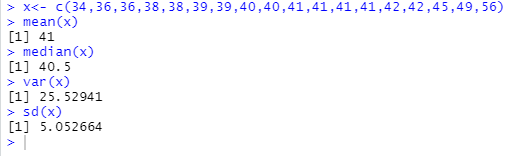
[1] 2.053749

200 2.05\* = **198.62-201.38**

**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?

**Ans : The mean=median only when the distribution is symmetric & distribution has zero skewness. But zero skewness does not imply symmetric distribution necessarily.**

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

**Ans : I the mean is greater than the distribution is positively skewed.**

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

**Ans : If median is greater than mean than the distribution is negatively skewed.**

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

**Ans : Positive valus of kurtosis indicate that a distribution is peaked and possesses thick tails.**

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

**Ans : It indicates that the distribution has lighter tails and i.e. distribution is flatter than the normal curve with same mean and standard deviation.**

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



What can we say about the distribution of the data?

**Ans :- The data is distributed around the median and has no outliers.**

What is nature of skewness of the data?

**Ans :- Left Skewed**

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

**Ans :- 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.

**Ans :- The box plot 1 is more condensed i.e. it varies less then the box plot 2 which has a much larger width. Also the box plot 1 is easier to predict. The medians of both the box plots is nearly equal.**

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

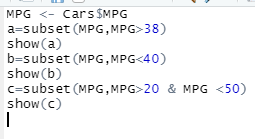
Data \_set: Cars.csv

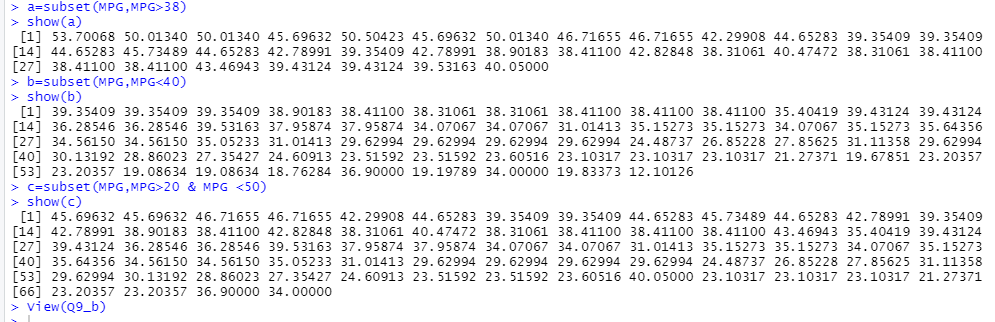
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

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

c. P (20<MPG<50)

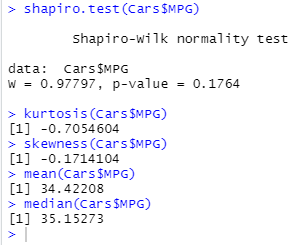




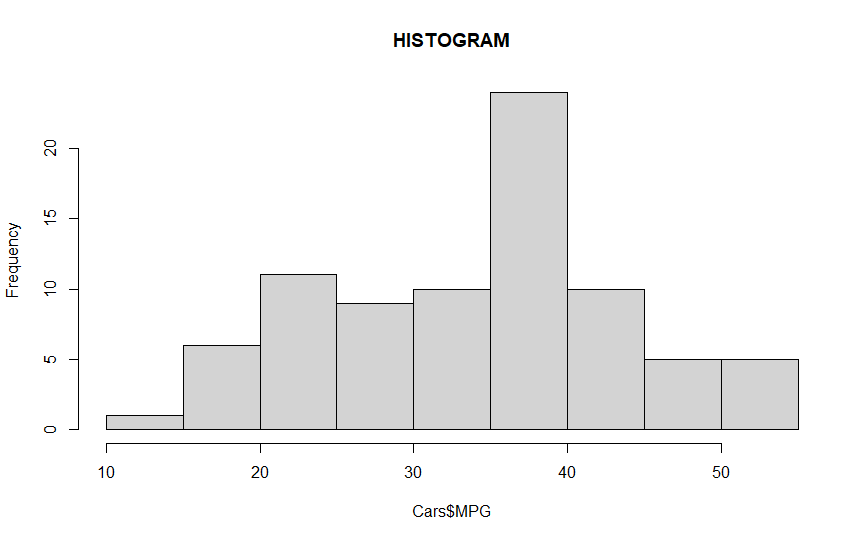
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



**Ans : By the justification from the above calculations we can see that the mean and median are almost similar and that the dataset is 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

Ans) waist follows Normal Distribution from the below QQ-plot

> qqnorm(wc\_at$Waist)

> qqline(wc\_at$Waist)



Adipose Tissue follows normal distribution

qqnorm(wc\_at$AT)

qqline(wc\_at$AT)



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

Ans) 90% > qnorm(0.95)

[1] 1.644854

94% > qnorm(0.97)

[1] 1.880794

60% > qnorm(0.8)

[1] 0.8416212

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

Ans) 95% > qt(0.975,24)

[1] 2.063899

96% > qt(0.98,24)

[1] 2.171545

99% > qt(0.995,24)

[1] 2.79694

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) µ=270, =260, SD=90, n=18, df=n-1=18-1= 17

tscore= = = -10/21.23= -0.47

> pt(-0.47,17)

[1] 0.3221639

Required probability = 0.32=32%