|  |  |
| --- | --- |
| 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 | **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 | **Interval** |
| Time on a Clock with Hands | **Interval** |
| Number of Children | **Nominal** |
| Religious Preference | **Nominal** |
| Barometer Pressure | **Interval** |
| SAT Scores | **Interval** |
| Years of Education | **Interval** |

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

Ans> The Sample Space for above Problem is

S = (HHH,HHT,HTH,THH,HTT,THT,TTH,HHH)

Interested event = (HHT,HTH,THH)

Therefore, Reqd. Probability = **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 2and 3

Ans> Sample Space =

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 11 | 12 | 13 | 14 | 15 | 16 |
| 21 | 22 | 23 | 24 | 25 | 26 |
| 31 | 32 | 33 | 34 | 35 | 36 |
| 41 | 42 | 43 | 44 | 45 | 46 |
| 51 | 52 | 53 | 54 | 55 | 56 |
| 61 | 62 | 63 | 64 | 65 | 66 |

1. **Zero**
2. **Interested Event=(11,12,13,21,22,31)**

**Therefore Reqd. Probability = 6/36**

1. **Interested Event = (15,24,33,42,51,66)**

**Therefore Reqd. Probability = 6/36**

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> Total no of events: ncr7c2=21

None of the balls is blue=5c2=10

Solution : Probability of none of the balls drawn is blue=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

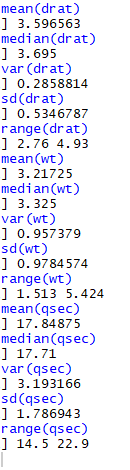
**Expected Value = ∑xP(x) = 1\*0.015+ 4\*0.20 +3\*0.65 +5\*0.005 + 6\*0.01 + 2\*0.120**

**= 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.



|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.59 | 3.22 | 17.85 |
| Median | 3.69 | 3.33 | 17.71 |
| Variance | 0.29 | 0.96 | 3.19 |
| Standard Deviation | 0.53 | 0.98 | 1.79 |
| Range | 2.76 - 4.93 | 1.513 - 5.424 | 14.5 - 22.9 |

Inference Drawn:

* The mean is useful for spotting trends in the data because we can compare means over a time period to spot trends. The mean is the most common measure of central tendency.
* The **median** divides a sample of data in half; it is the middle score. The median is a useful statistic if we think our data have some extreme cases. The median is not impacted by extreme cases, but the mean is.

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>**EV = ∑X/n = 1308/9 = 145.33**

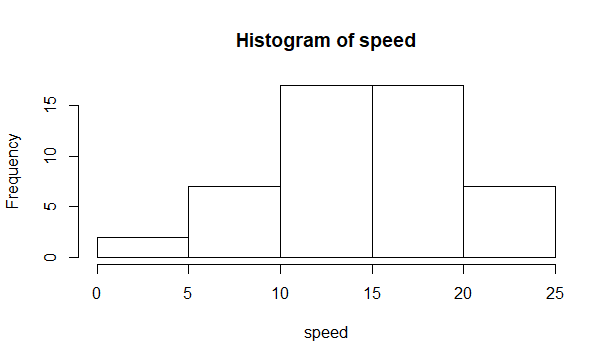
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

****

For Speed Column

|  |
| --- |
| > attach(a)  > skewness(speed)  [1] -0.1139548 ## left skewness  > kurtosis(speed)  [1] 2.422853 ## negative Kurtosis |
|  |
| |  | | --- | |  | |



Skewness is negative, that tells us that the distribution is skewed towards left. Mean of distribution is less than the Median. Kurtosis Value is less than 3, that tells us that the distribution has broad peak and thin tails as evident from the histogram.

For dist column

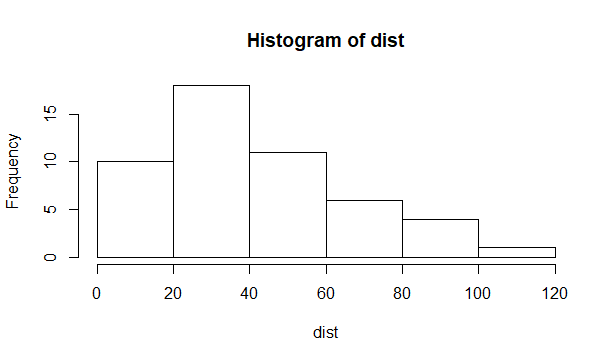
> attach(a)

> skewness(dist)

[1] 0.7824835 ## positive/Right skewness

> kurtosis(dist)

[1] 3.248019 ##positive Kurtosis



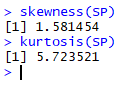
Skewness is positive, that tells us that the distribution is skewed towards right. Mean of distribution is more than the Median. Kurtosis Value is more than 3, that tells us that the distribution has sharp peak and wide tails as evident from the histogram.

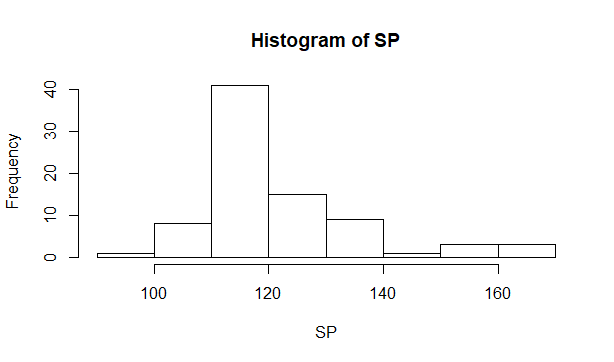
**For**

**SP and Weight(WT**

****

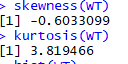
For SP:

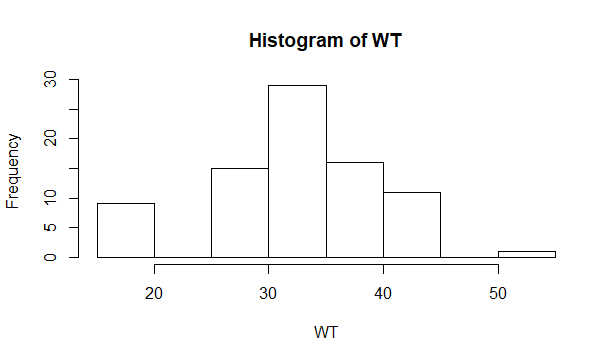
****



Skewness is positive, that tells us that the distribution is skewed towards right. Mean of distribution is more than the Median. Kurtosis Value is more than 3, that tells us that the distribution has sharp peak and wide tails as evident from the histogram.

**For WT:**





Skewness is negative, that tells us that the distribution is skewed towards left. Mean of distribution is less than the Median.Kurtosis Value is more than 3, that tells us that the distribution has sharp peak and wide tails as evident from the histogram.

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



**The Distribution is Right Skewed. Mean > Median.**



**The above boxplot suggests that the distribution has lots of outliers towards upper extreme.**

**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>**94% Confidence:**

X-bar = 200

Sd = 30

n = 2000

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

=200 ± 1.88\*30/sqrt(2000)

=**198.74 – 201.26**

>**98% Confidence:**

X-bar = 200

Sd = 30

n = 2000

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

=200 ± 2.33\*30/sqrt (2000)

=**198.44-201.56**

>**96% Confidence:**

X-bar = 200

Sd = 30

n = 2000

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

=200 ± 2.05\*30/sqrt (2000)

=**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?

M**ean = 41**

**Median = 40.5**

**Variance = 25.53**

**Standard Deviation = 5.05**

**2. Mean > Median, This implies that the distribution is slightly skewed towards right. No outliers are present.**

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

**Skewness = 0, Symmetric**

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 adata ?

**Sharp Peak, Thick Tails**

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

**Broad Peak, Thin Tails**

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



What can we say about the distribution of the data?

**Not a Normal Distribution**

What is nature of skewness of the data?

**Left Skewed**

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

**Both are Normally Distributed**

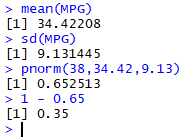
**Q 20)** Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

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

MPG<- Cars $MPG

* 1. (MPG>38)



P(MPG>38) = **0.35**

* 1. (MPG<40)



P(MPG<40) = **0.73**

* 1. (20<MPG<50)

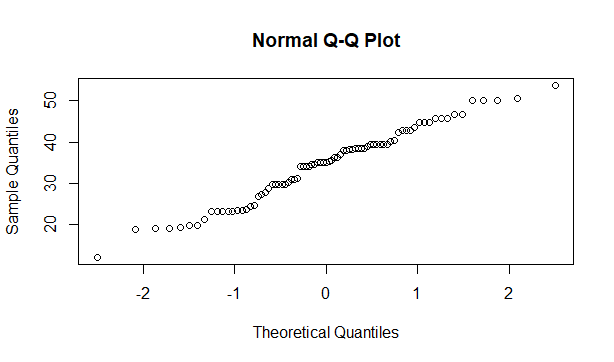


P(20<MPG<50) = 0.898

**Q 21)** Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

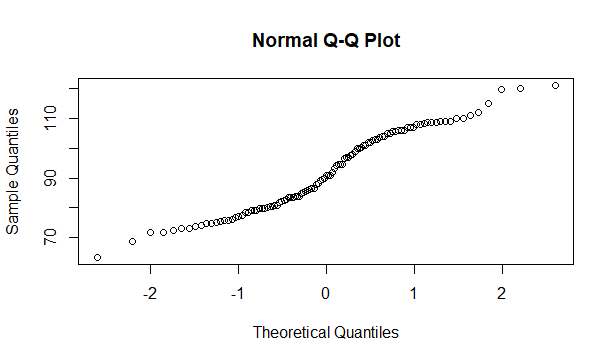
Dataset: Cars.csv

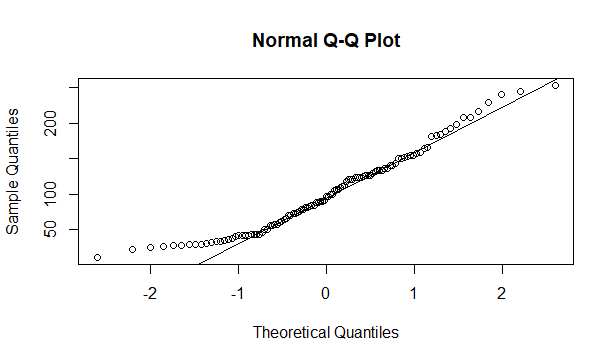


**Follows Normal Distribution as indicated by the qqplot.**

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

Dataset: wc-at.csv

**Wc-at $waist follows Normal Distribution**

**Wc-at$AT follows Normal Distribution**

**Q 22)** Calculate the Z scoresof 90% confidence interval,94% confidence interval, 60% confidence interval

**90% Confidence:**

****

**94% Confidence:**



**60% Confidence:**



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

**95% Confidence:**



**96% Confidence:**

****

**99% Confidence:**

****

**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**

**µ = 270, n= 18, x bar = 260, s = 90**

**t score = (x bar - µ)/(s/sqrt(n))**

**=(260 – 270)/(90/sqrt(18))**

**=-10/21.23**

**=-0.47**

**Required Probability = 0.32**

