|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continous) |
| Weight of Gold | Continous |
| Distance between two places | Continous |
| Length of a leaf | Continous |
| Dog's weight | Continous |
| 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 | Nominal |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Nominal |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Ordinal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Ratio |
| Number of Children | Nominal |
| Religious Preference | Ordinal |
| Barometer Pressure | Intervak |
| SAT Scores | Ordinal |
| Years of Education | Ordinal |

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

ans: There are 2 power 3 possiblities and only three of them 2 heads and 1 tail

answer will be 3/8

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

* Equal to 1
* Less than or equal to 4
* Sum is divisible by 2 and 3

answer : probablity will be 0/36

2 .three possible outcome are there give a sum equal to 4

{(1,3),(2,2),(3,1))} so ans will be 3/36 or 1/12

3. no of favourable outcomes 24

probablity = 24/36 or we say 2/3

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 A bag contains no of bags which total ball is : 7

5/7 and 2/3 (as random) =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: using expected value formula

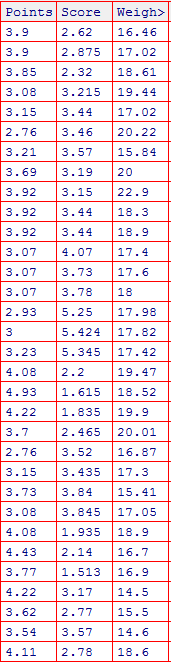
1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

=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.



#average values

> mean(rd$Points)

[1] 3.596563

> mean(rd$Score)

[1] 3.21725

> mean(rd$Weigh)

[1] 17.84875

#all mid values

> median(rd$Points)

[1] 3.695

> median(rd$Score)

[1] 3.325

> median(rd$Weigh)

[1] 17.71

#mode shows data types because mode cant find on discrete data

> mode(rd$Points)

[1] "numeric"

> mode(rd$Score)

[1] "numeric"

> mode(rd$Weigh)

[1] "numeric"

#Units get squared

> var(rd$Points)

[1] 0.2858814

> var(rd$Score)

[1] 0.957379

> var(rd$Weigh)

[1] 3.193166

# units Back to Normal

> sd(rd$Points)

[1] 0.5346787

> sd(rd$Score)

[1] 0.9784574

> sd(rd$Weigh)

[1] 1.786943

#There are range Points score and weight

> range(rd$Points)

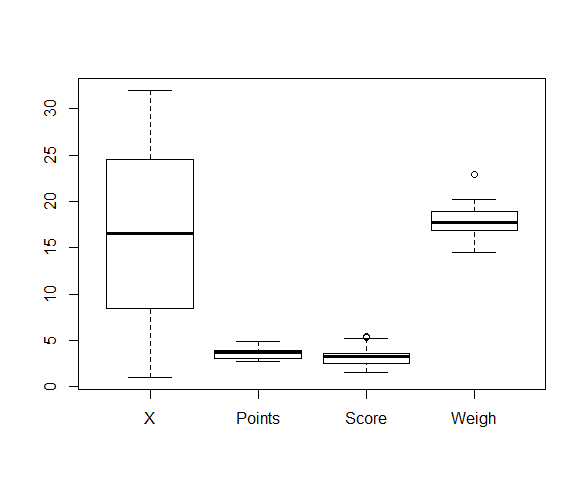
[1] 2.76 4.93

> range(rd$Score)

[1] 1.513 5.424

> range(rd$Weigh)

[1] 14.5 22.9



Here is box plot: that shows where data point Tends.

Q8) Calculate Expected Value for the problem below

* 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 from above it could be any :

Mean calculate : (108+110+123+134+135+145+167+187+199)

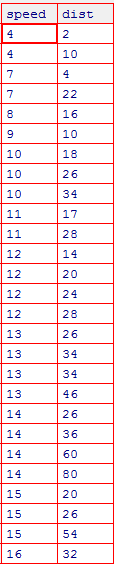
=145.33

mean(rd$V1)

[1] 145.3333

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

**Cars speed and distance**



**Ans:**

> skewness(rd$speed)

[1] -0.1139548 # negative Skewness

> skewness(rd$dist)

[1] 0.7824835 #Positive Skewness

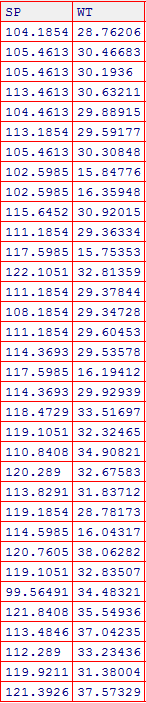
> kurtosis(rd$speed)

[1] 2.422853 #Positive Kurtosis

> kurtosis(rd$dist)

[1] 3.248019 #Positive Kurtosis

**SP and Weight(WT)**



Ans : Left Skewed for SP and Negative kurtosis

Left Skewed for WT and Positive kurtosis

Skewness:

Calculation of speed and wt (for sp )

-0.42675

Calculation of speed and wt (for WT)

-1.34755

Kurtosis:

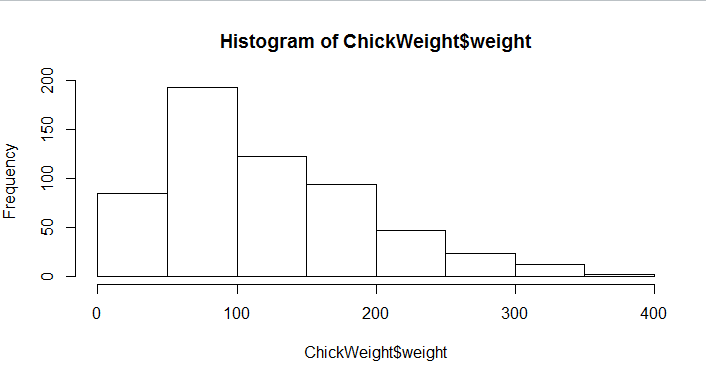
Calculation of speed and wt(for Sp)

-0.86374

Calculation of speed and wt(for Wt)

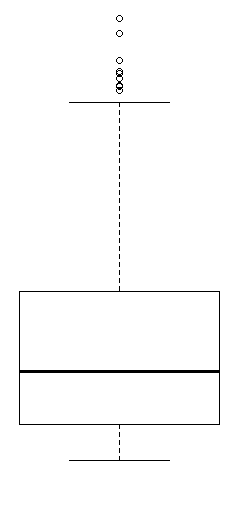
1.152952

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

**Histogram**

Ans : It show a right skewed and it also tell about mean average between 100-150

There are no ouliers in the data



**Boxplot:**

Ans : And this diagram show about the outliers in a upper extreme but the fact is it didn’t affect median part.

Data is Distributed on the right and Positively Skewed

**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; sample std = 30; x = 200

94% is (-198,201)

98% is (198,201)

96% is (198,201)

**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**

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

Ans:-

A:- Mean value is 41.12

Standard deviation is 5.18

Variance is 26.86

Standard deviation is 5.18

B. It shows left skewed & data towards right And Most of Student were passed

Some of are in average .

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

Ans: Zero Skewness

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

Ans: Positive Skewness

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

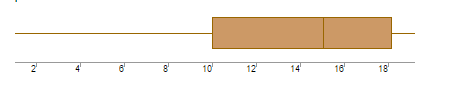
Q16) What does positive kurtosis value indicates for a data ?

Ans: Thin Peak

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

Ans: Wider Peak

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



What can we say about the distribution of the data?

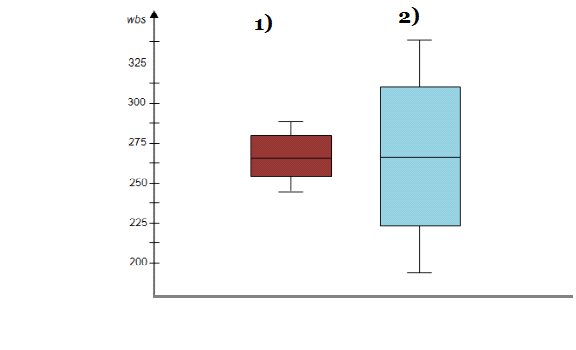
Ans: The data is asymmetrical Distributed

What is nature of skewness of the data?

Ans: Negatively skewed Distribution

What will be the IQR of the data (approximately)?   
Ans: Inter Quartile Range is 8

Q19) Comment on the below Boxplot visualizations?



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

* First box plot are between 280 to 250

And also there is no outliers there

Last it has Normally Distributed

->Second box plot are between 220 to 320

And it has right Skewed

Last there is No ouliers there

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

* P(MPG>38)
* P(MPG<40)

c. P (20<MPG<50)

ans: Probablity of P(MPG>38):

pnorm(38,34.4,9.13) = 0.653321

Probablity of P(MPG <40):

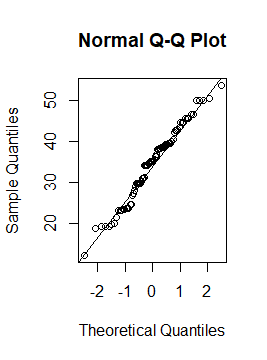
1 - pnorm(40,34.4,9.13) = 0.2698183

Probablity of P(20<MPG<50) :

> pnorm(20,34.4,9.13) - pnorm(50,34.4,9.13)

[1] -0.8988697

Q 21) Check whether the data follows normal distribution

* Check whether the MPG of Cars follows Normal Distribution 

Dataset: Cars.csv

Ans: Yes it Is Normally Distributed

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

Dataset: wc-at.csv

Ans: Both are Normally Distribution

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

Ans: -> 90 % confidence Interval is 1.65

94 % confidence Interval is 1.88

60 % confidenc Interval is 0.84

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

Ans: 95 % confidence Interval is 2.063

96 % confidence Interval is 2.171

99 % confidence Interval is 2.796

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: The first thing we need to do is compute the t statistic

using this formula

t = [ x - μ ] / [ s / sqrt( n ) ]

t = (260-270)[90/sqrt(18)]

t = -10/21.22641 = -0.4711112

and the degree of freedom are equal to 18-1 =17

The t statistic is equal to 0.4711112

Using r code I find : pt(0.47,17) is 0.677

-----------------------------------------------------\*-------------