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

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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

-The probability for three coins are tossed is (HHH,HHT,HTH,THH,HTT,THT,TTH,TTT) =8

-the probability for two heads and one tail is =3 (HTH,THH,HHT)

-the probability is = 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

a)- total number of possibilities for two dice rolled is =36

probability of having sum is equal to 1 is =0

b)- the possibilities for less than or equal to 4 is=

(1,1)(2,1)(1,2)(1,3)(3,1)(2,2)=6

Then the probability is =6/36=1/6

c)- the probability of sum is divisible by 2 and 3 is =24

total probability is =24/36=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?

* Total number of balls drawn is – 2+3+5=7
* The number of drawing 2 balls randomly is

7C2 = 7!/(7-2)!2! =42/2=21

-none of the balls drawn is blue = 5C2

5!/(5-2)!2!=20/2=10

-P()=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

* (1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120)
* 0.015+0.8+1.95+0.025+0.06+.24
* 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.

For Points:

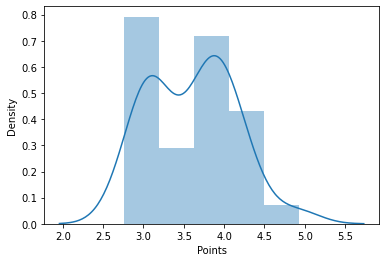
Mean = 3.5965625

Mode= 3.07 and 3.92

Median= 3.69500  
 Variance= 0.2858813

Standard Deviation= 0.534679

Range= 2.17



-It was normal distribution.there is no skeweness here.

It will achieved the histogram at two levels

For Score:

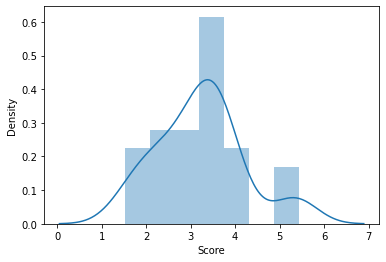
Mean= 3.2172499

Mode=3.44

Median=3.325  
 Variance=0.957378

Standard Deviation=0.97845

Range= 3.91



It was a Normally distributed.

There is no skewness here.

For Weigh:

Mean=17.848

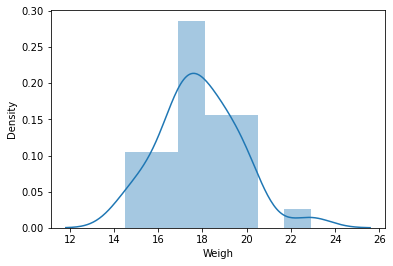
Median=17.710

Mode=17.02 and 18.90

Variance=3.193166

Standard deviation=1.786943

Range=8.4



It was normally distributed and histogram achievs at some 0.21

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?

Expected Value  =   ( probability  \* Value )

  P(x).E(x)

Probability of selecting each patient = 1/9

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

P(x) =1/9 ,1/9  ,1/9 ,1/9 ,1/9  ,1/9  ,1/9 ,1/9 ,1/9

Expected Value  =  (1/9)(108) + (1/9)110  + (1/9)123 + (1/9)134 + (1/9)135 + (1/9) 145 + (1/9(167) + (1/9)187 + (1/9)199

= (1/9) ( 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9)  (  1308)

= 145.33

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

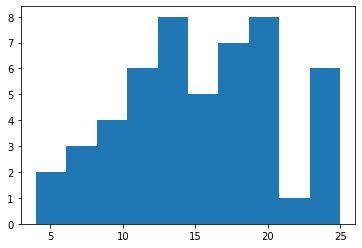
Cars speed and distance

Use Q9\_a.csv

Speed:

Skewness= -0.11750

Kurtosis= -0.57714

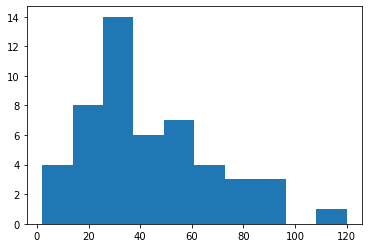
****

There is negative skewness.

Distance:

Skewness= .80689

Kurtosis= .248018



It is positively skewed.

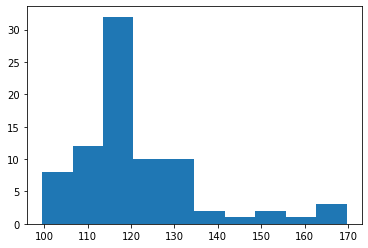
SP and Weight(WT)

Use Q9\_b.csv

For SP

Skewness = 1.61145

Kurtosis = 2.723521

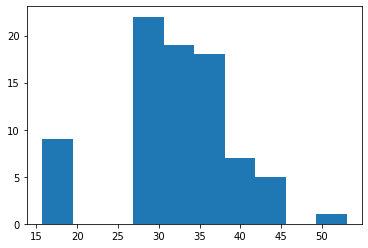


It was positively skeweed.

Weight:

Skewness= -0.6147

Kurtosis= 0.950294



It was normally distributed.It achieves histogram.

There is no skeweness in this one.

Q10) Draw inferences about the following boxplot & histogram



-It was positively skewed.

-it does not follows normal distribution.

The mean is greater than median.

It will be maximum at 200 frequency. And check weight of 100



-It is positively skewed.

It does not follows normal distribution.

It has more than 2 outliers.the mean is greater than median.

**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?

Here 2000 > 30.hence sample is normal and it is sample random sample.

Sample statistic is 200

We have to calculate confidence interval for 94%,98%,96%.

Then alpha =1-(confidence level/100)

For 94% 🡪 alpha=1-(94/100)=0.06

For 98% 🡪 alpha= 1-(98/100)=0.02

For 96% 🡪 alpha= 1-(96/100)=0.04

Critical probability for 94%= P=1-alpha/2=1-0.06/2=0.97

For 98% 🡪P=1-0.02/2=0.99

For 96% 🡪 P=1-0.04/2=0.98

Degree of freedom= n-1

=2000-1

=1999

From t distribution critical value are

For 94% = 1.96

For 98% = 2.326

For 96% = 2.054

Margin of error

For 94% 🡪 critical value(94%)\*standard

=1.96\*0.67=1.3132

For 98% 🡪 2.326\*0.67=1.55842

For 96% 🡪 2.054\*0.67=1.37618

Confidence interval =Sample statistics +/- margin of error

CI of 94% = 200 + 1.3132 200-1.3132

= 201.31 = 198.68

Confidence interval for 94% is between 198.68 to 201.31

CI of 98% = 200 + 1.55842 200-1.55842

= 201.558 = 198.442

Confidence interval for 98% is between 198.442 to 201.558

CI of 96% = 200+ 1.37618 200-1.37618

= 201.37618 = 198.63

Confidence interval for 96% between 198.63 to 201.37

**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.

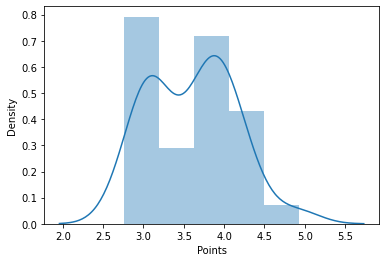
Mean=41

Median=40.5

Variance=25.5294

Standard deviation=5.0526

1. What can we say about the student marks?



By observing histogram the student marks will achieve the histogram at the 39 marks.It was normally distributed,

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

-When the mean,median of data are equal then the skewness will be zero

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

-When the mean>median then the distribution will be positively skewed.

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

-if the median>mean then the distribution will be negatively skewed.

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

A distribution with a positive kurtosis value indicates that the distribution has heavier tails and a sharper peak than the normal distribution.

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

A distribution with a negative kurtosis value indicates that the distribution has lighter tails and a flatter peak than the normal distribution.

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



What can we say about the distribution of the data?

For a given data the distribution is negatively skewed.the box plot will show the median closer to the upper or top quartile. For the given data mean < median.

What is nature of skewness of the data?

* The median is close to 3rd quartile
* The data is negatively skewed

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

IQR = Q3-Q1

= 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.

-Boxplot 1 was positevly skewed.it does not follows Normal distribution.

- Boxplot 2 was follows Normal distribution.

- Both have a median of around 262 wbs.

- both don’t have outliers.

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)

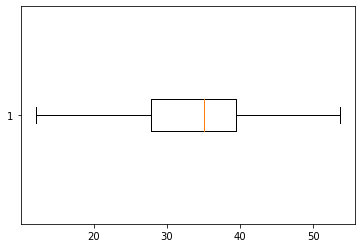
Probability = 0.34667

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

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

MPG of Cars

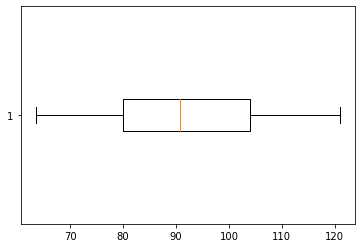


It does not follows Normal distribution.It was negatively skewed.

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

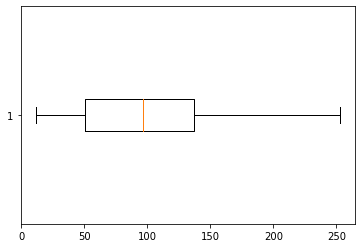
Dataset: wc-at.csv

Waist



It was positively skewed. hence It does not follows Normal distribution.

AT(Adipose tissue)



It was negatively skewed . hence it does not follows Normal distribution.

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

Z scores for

90%=1.65

94%=1.89

60%=0.84

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

T scores for

95%= 2.064

96%=2.064

99%=2.797

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

🡺

t-statistics for the data is given as follows:

t=(x-mean)/(s/sq.rt(n))

x=mean of the sample of bulbs = 260

mean = population mean=270

s=standard deviation of the sample=90

n= number of items in the sample =18

t=(260-270)/(90/sq.rt(18))

t=-10/(90/3sq.rt(2))

t=-10/(30/sq.rt(2))

t=(-1\*sq.rt(2))/3

t=-0.471

degrees of freedom is n-1

t-distribution with 17 degrees of freedom

The probability that t<-0.471 with 17 degrees of freedom assuming the population mean is true. The t-value is less than the t-value obtained with 17degrees of freedom and a t score of -0.471,the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300 days.