|  |  |
| --- | --- |
| 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 | Interval |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| 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 | Ordianal |
| Years of Education | Nominal |

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

If Three Coin are tossed={HHH,HHT,THH,THT,HTH,TTH,HHT,TTT}=8

Then Probablity For Two Heads And One Tail={HHT,THH,HTH,THH,HTH} =3

Probablity=3/8=0.375

Probablity=37.5%

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:-

If Two Dice Are rolled=6\*6=36

Then Probablity For:-

a)Equal To 1=There is no term Equal To 1.

b)Less than or equal to 4=(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)=No. Of Event/Total No. of Event=6/36=0.16=16%

c)Sum is divisible by 2 and 3={(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)}=No. of Event/Total No. Of Event=6/36=0.16=16%

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 Balls=2+3+2=7

Let s be the sample event=

n(S)=Probablity If Two Balls Are Drawn at random=7C2=7\*6/2\*1=42/2=21

n(E)=Probablity If none of the balls drawn is Blue=5C2=5\*4/2\*1=20/2=10

Probablity For Event=P(S)=n(E)/n(S)=10/21=0.47=47%

\*Conclusion:-So Probablity That None of The Balls Drawn is Blue=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:-

Expected number of candies for a randomly selected child:-

1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120=3.09

\*Conclusion:-Expected number of candies for a randomly selected child:-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.

ANS:-

For The Given Dataset:-

1)Mean for:-

i)Points=3.59

ii)Score=3.21

iii)Weight=17.84

2)Median for:-

i)Points=3.69

ii)Score=3.32

iii)Weight=17.71

3)Mode for:-

i)Points=3.92,3.07,3.9,4.22

ii)Score=3.44,3.57

iii)Weight=17.02

4)Varience for:-

i)Points=0.28

ii)Score=0.95

iii)Weight=3.19

5)Standard Deviation:-

i)Points=0.53

ii)Score=0.97

iii)Weight=1.78

6)Range:- Max.-Min.=

i)Points=4.93-2.76=2.17

ii)Score=5.42-1.51=3.91

iii)Weight=22.90-14.50=8.4

1. For Points dataset:-

a) The data is concentrated around Median

b) There are no Outliers In Point dataset

c) The distribution is Left skewed Data

1. For Score dataset:

a) The data is concentrated around Median

b) There are 3 Outliers: 5.250, 5.424, and 5.345 in Score Dataset

c) The distribution is Right skewed Data

1. For Weigh dataset:

a) The data is concentrated around Median

b) There is 1 Outlier: 22.90 in Weigh Dataset

c) The distribution is Left skewed Data

**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:-

Expected Value = ∑ (probability \* Value)

 ∑ P(x).E(x)

Here There is 9 patients:-

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

Probablity For Each 9 Patients=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 + 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**

**>Skewness for cars speed=**-0.11395477012828319

From this we can conclude that skewness for cars speed is left skewed

**>Skewness for cars distance=**0.7824835173114966

**From this we can conclude that skewness for cars distance is rightly skewed**

**>Kurtosis for cars speed=**-0.5771474239437371

**From this we can conclude that kurtosis for cars speed is -VE which means Data is Almost Equal**

**>Kurtosis for cars distance=**0.24801865717051808

**From this we can conclude that kurtosis for cars distance is +VE which means Data is distributed around High pick**

**-SP and Weight(WT)**

**Use Q9\_b.csv**

**>>Skewness for cars SP=**1.5814536794423764

From this we can conclude that skewness for cars SP is Right skewed

**>>Skewness for cars WT=**-0.6033099322115126

**From this we can conclude that skewness for cars WT is Left skewed**

>**>Kurtosis for cars SP=**2.7235214865269244

**From this we can conclude that kurtosis for cars SP is +VE which means Data is distributed around High Pick**

**>>Kurtosis for cars WT=**0.8194658792266849

**From this we can conclude that kurtosis for cars WT is +VE which means Data is Distrbuted around High Pick**

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



-The Most Of The Data Point Is Within The Range Around 50-100

-The High Frequency In The Histogram Is Around 200

-In This Data Tail is Going To Right side So we can Say That Data IS Right-Skewed Data



-In This Boxplot Median Is Less Than Mean value

-There are 8 Outliers in The Boxplot

-There is Less Data Points near Lower Quartile

**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?

-Given=S=2000

Mean=200

Standard Deviation=30

As We Take Samples From Population So we Go For t-Distribution

-For 94% confidence interval=

weight=stats.t.interval(0.97,df=1999,loc=200,scale=30)

weight=(134.85077199775026, 265.14922800224974)

-For 98% confidence interval=

Weight=stats.t.interval(0.99,df=1999,loc=200,scale=30)

Weight=(122.6512694143423, 277.3487305856577)

-for 96% confidence interval=

Weight=stats.t.interval(0.98,df=1999,loc=200,scale=30)

Weight=(130.15355671679083, 269.84644328320917)

**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?

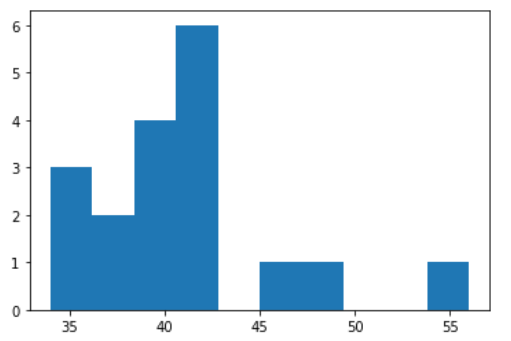
-Mean=41.0

-Median=40.5

-Variance=25.52

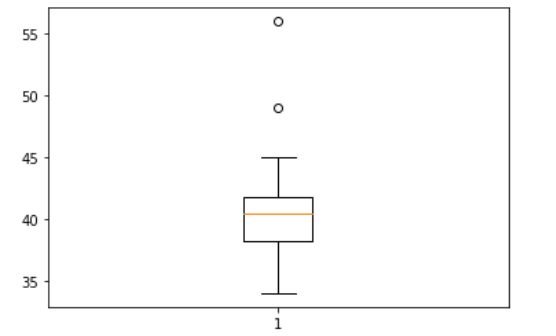
-Standard Deviation=5.05

Histogram :-



Most Students Mark’s are In Between 36 To 42

Boxplot:-



There are Two Outliers:-49 and 56

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

ANS:- It Is Symmetrical

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

ANS:- If mean>median=Right Skewed

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

ANS:- If median>mean=Left skewed

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

ANS:- The Data Have High Pick At The Centre

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

ANS:-The data is Distributed Equally

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



What can we say about the distribution of the data?

ANS:- If The Above Data Is About Student’s Marks Then

50% Students Scores Marks In-between 10-15

And The Remaining Student’s Have Score In-between 16-18

What is nature of skewness of the data?

ANS:-Here Mean<Median Then The Data Is Left Skewed Data

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

IQR of the data=Upper Quartile-Lower Quartile

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:-

In Fig.1 and Fig.2 Have No Outliers And Data Is Normally Distributed.Since Mean And Median Are Equal

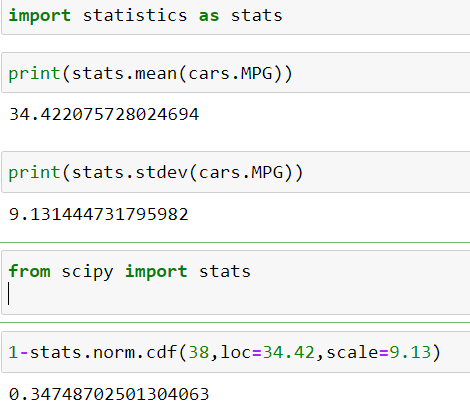
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

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

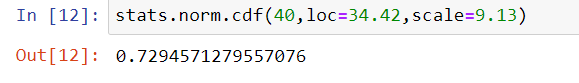
MPG <- Cars$MPG

* 1. (MPG>38)



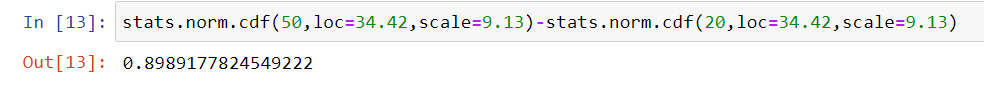
Probability For (MPG>38) is 0.34=34%

* 1. P(MPG<40)



Probability For P(MPG<40) is 0.72=72%

* 1. P(20<MPG<50)



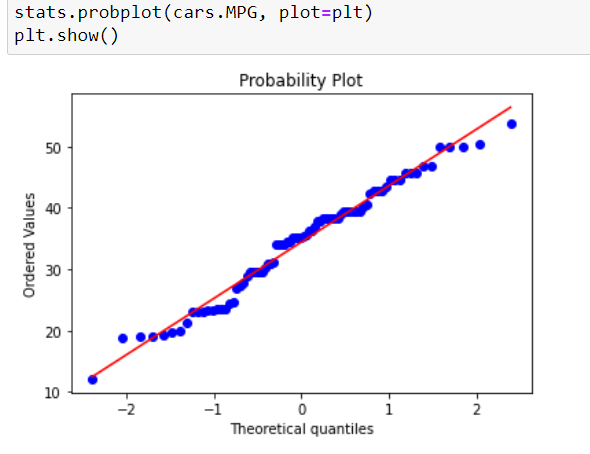
Probability For P(20<MPG<50) is 0.89=89%

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

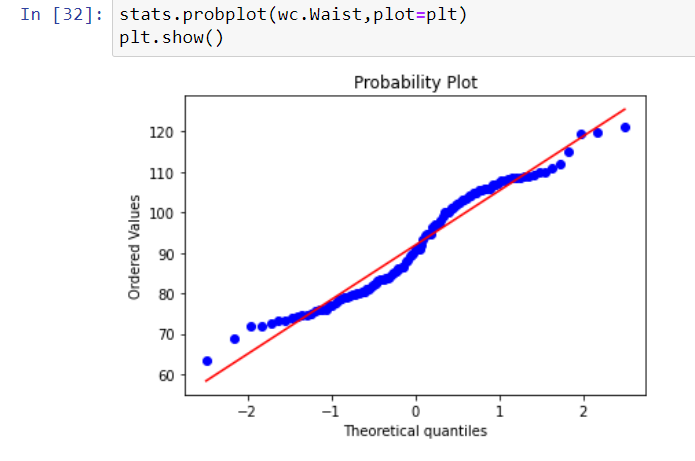
Interference:-MPG of Cars Follows Normal Distribution



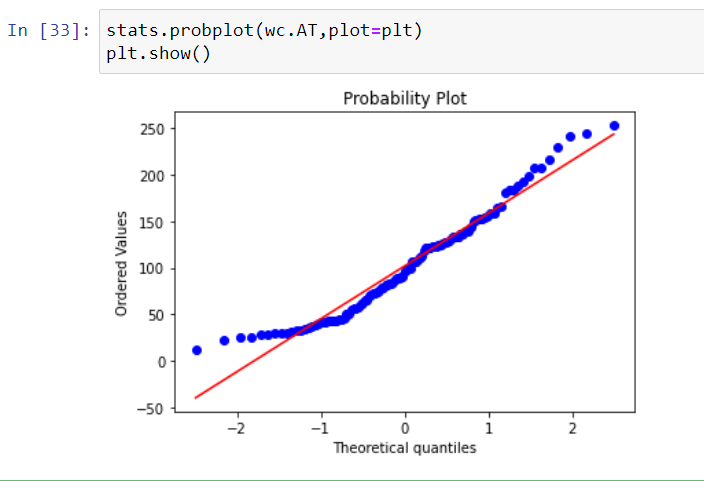
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Interference:-from wc-at data Waist Circumference(Waist) Follows Normal Distribution



Interference:-from wc-at data Adipose Tissue(AT) Follows Normal Distribution



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

Z-Scores Of 90% Confidence Interval=1.64

Z-Scores Of 94% Confidence Interval=1.88

Z-scores of 60% Confidence Interval=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% confidence interval=2.06

96% confidence interval=2.17

99% confidence interval=2.79

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

Sample Size=S=18

Mean=260

Standard Deviation=90

As We Have Samples We Go For Samples Then We Will Apply t-statistics

Don’t Know How To Solve This Question.