|  |  |
| --- | --- |
| 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 | interval |
| 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 | 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?

**probability = [HHH, HHT, HTH]**

**= 3/8**

**= 0.375**

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

1. Equal to 1

**The probability that sum is equal to zero.**

1. Less than or equal to 4

**The probability that sum is less than or equal to 4 are**

**[2+1, 1+1, 1+2, 2+2, 1+3, 3+1] = 6/36**

**=1/6**

**=0.167**

1. Sum is divisible by 2 and 3

**The probability that sum is divisible by 2 and 3 are**

**[1+5, 2+4, 3+3, 5+1, 4+2, 6+6]=6/36**

**=1/6**

**=0.167.**

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?

**probability = 5C2/7C2 = 10/21**

**= 0.476**

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

**The probability of expected number of candies for a randomly selected child are**

**= Xi \* P (Xi) =0.015\*1+0.20\*4+0.65\*3+0.005\*5+0.01\*6+0.120\*2**

**=3.09**

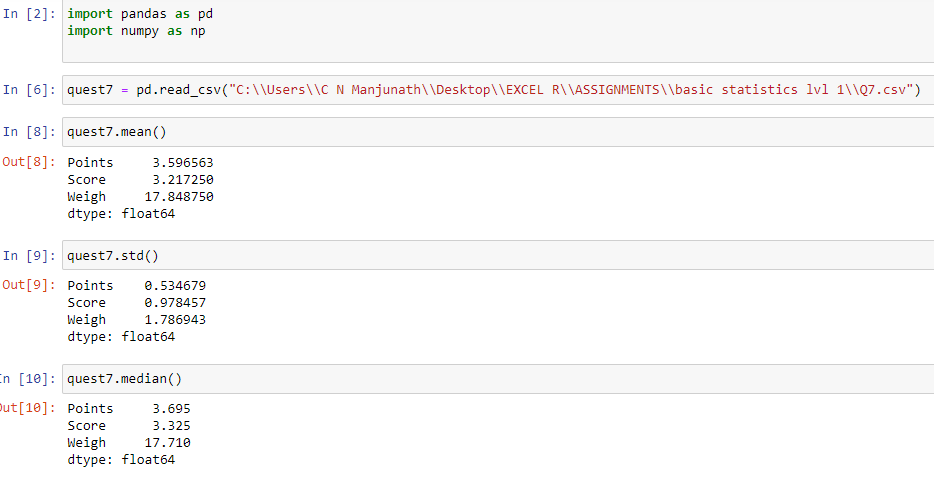
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

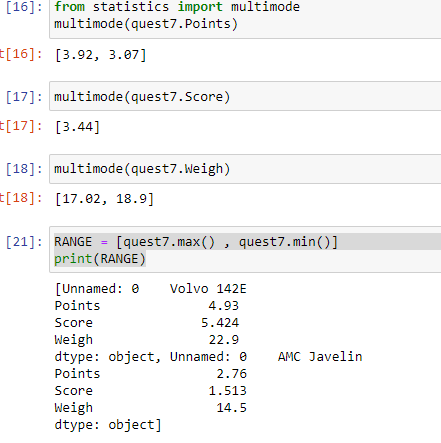
* For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weight** |
| **Mean** | **3.596563** | **3.217250** | **17.848750** |
| **Median** | **3.695** | **3.325** | **17.710** |
| **Mode** | **3.92, 3.07** | **3.44** | **17.02, 18.9** |
| **Standard Deviation** | **0.534679** | **0.978457** | **1.786943** |
| **Range** | **[4.930 , 2.760]** | **[5.424 , 1.513]** | **[22.90 , 14.50]** |

****

****

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 of the weight of that patient are**

**=108+110+123+134+135+145+167+187+199 /9**

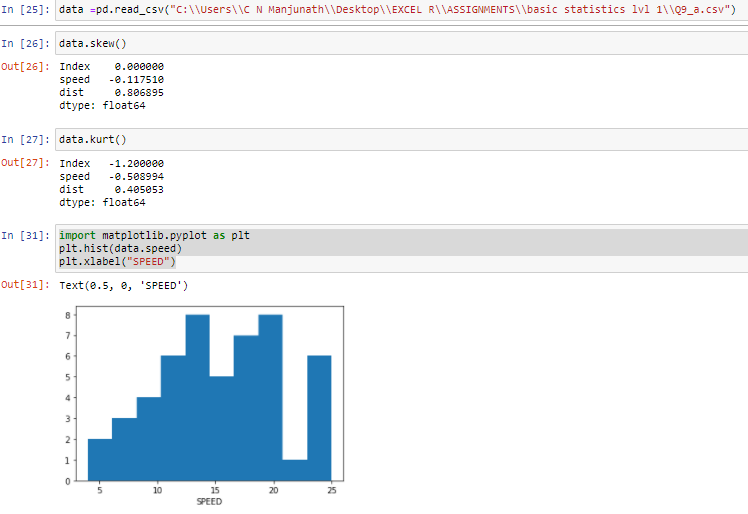
**=145.33**

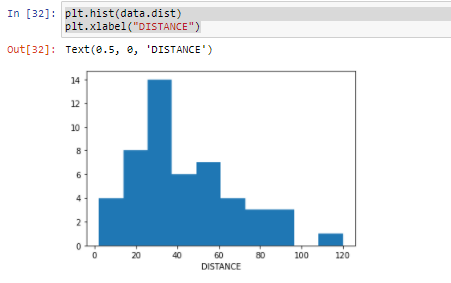
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **kurtosis** |
| **Car speed** | **-0.117510** | **-0.508994** |
| **distance** | **0.806895** | **0.405053** |

****

****

**Inferences drawn:**

* **Speed:**

**Speed has negative skewness.**

**Kurtosis value is less than 1 hence negative kurtosis.**

* **Distance:**

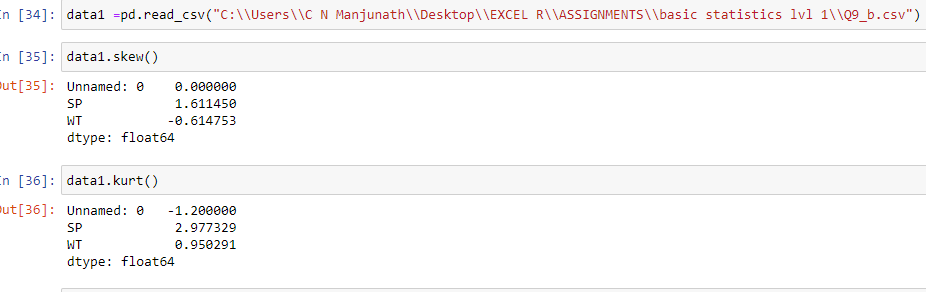
**Distance has positive skewness.**

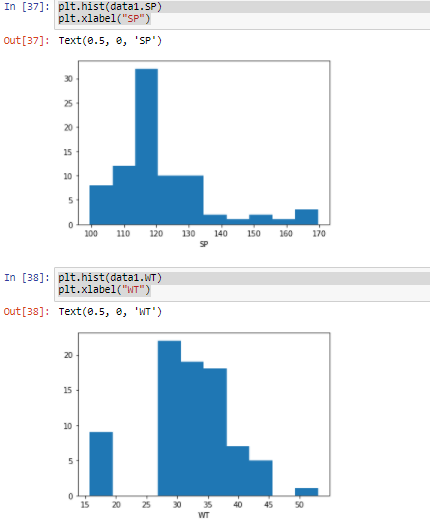
**Kurtosis value is greater than 1 hence data is normal.**

**SP and Weight (WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | **skewness** | **kurtosis** |
| **sp** | **1.611450** | **2.977329** |
| **weight** | **-0.614753** | **0.950291** |

****

****

**INFERENCES ARE DRAWN:**

* **WT:**

**Skewness is negative.**

**Kurtosis is positive.**

* **SP:**

**Skewness is positive.**

**Kurtosis is positive.**

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



* **There are no outliers.**
* **The data is positive skewness.**
* **Mode lies between 50 and 100.**



* **There are more outliers.**
* **Positive skewness.**
* **The distribution of the data is more concentered in upper whisker.**

**11)** 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 ; s =30 , n = 2000 Mu=200**

**s/sqrt(n)= 30/sqrt(2000) = 0.67**



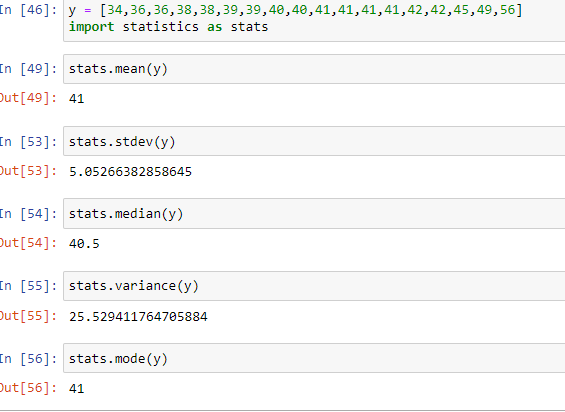
**Q12)** below are the scores obtained by a student in tests

**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.

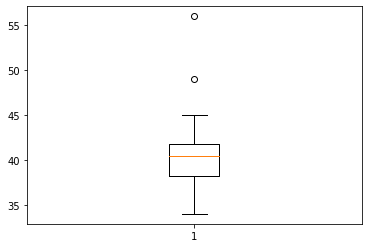
ANS:

|  |  |  |  |
| --- | --- | --- | --- |
| mean | median | variance | Standard deviation |
| 41.000000 | 40.5 | 25.529412 | 5.052664 |



1. What can we say about the student marks?

* **In the data score obtained by the students there is 2 outlier.**
* **The two out layers are 50 and 52.**



* **In the score most of the marks repeated between the 40-45 and the**

**Mode of the data is 41.**

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

**when the mean and median are equal there is an equal distribution the data**

**Will distributed symmetry.**

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

ANS: **when the mean is greater than median means Skewness of the data will be positive.**

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

ANS: **when the median is greater than mean means data will be or Skewness of the data will be negative.**

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

ANS: **The positive value of the kurtosis indicates that a distribution is peaked and possess thick tails.**

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

ANS: **The negative value of the kurtosis indicates that 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?

ANS: **the distribution of the data more concerted near lower whisker.**

What is nature of skewness of the data?

ANS: **The data will be negative skewness, because Q3-Q2<Q2-Q1.**

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

**ANS: The IQR of the data will be 8.**

**Explanation: IQR=Q3-Q1**

**Q3 value of the data will be 18.**

**Q1 value of the data will be 10.**

**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.

**ANS:**

* **Median for both is 262.5**
* **The data is normally distributed for both plates.**
* **2nd boxplot have larger number data than 1st box plot.**

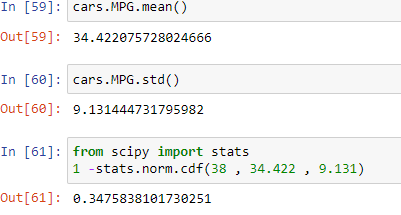
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

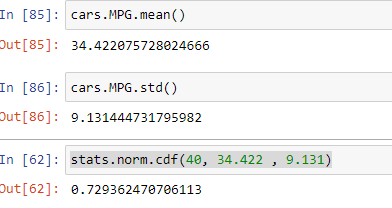
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

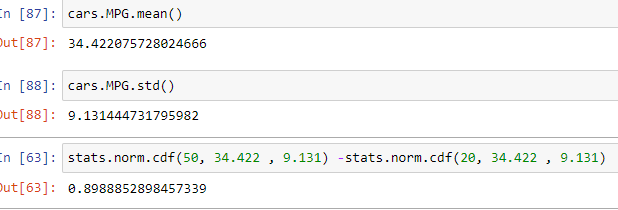
* 1. P(MPG>38)



* 1. P(MPG<40)



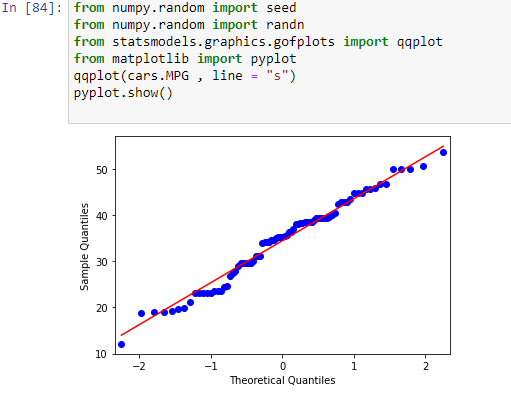
* 1. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

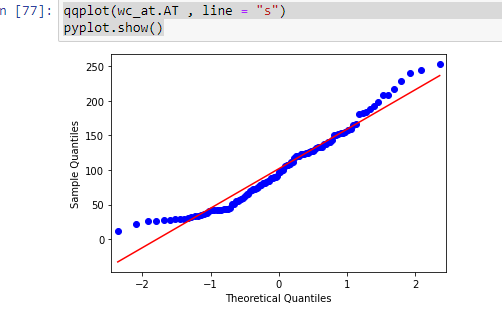
Dataset: Cars.csv



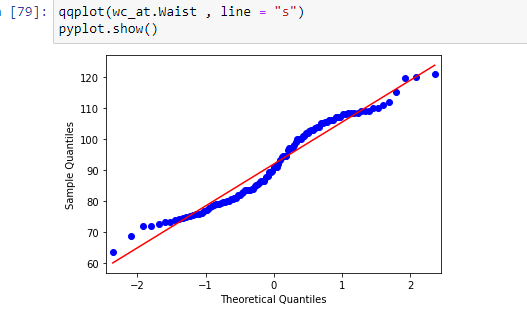
from qqplot data is following normal distribution

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

Dataset: wc-at.csv

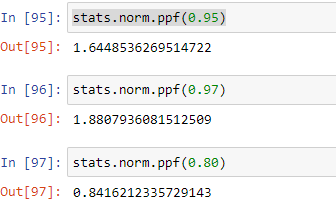


from qqplot data is following normal distribution

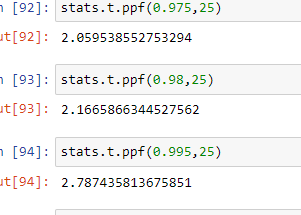


from qqplot data is following normal distribution

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



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



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:

Rode 🡪 pt.(tscore,df)

df 🡪 degrees of freedom

t score= xbar - mu/(S/SQRT(n))

260-270/(90/sqrt(18)) = -0.47

