|  |  |
| --- | --- |
| 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 | nominal |
| Number of kids | discrete |
| Number of tickets in Indian railways | discrete |
| Number of times married | discrete |
| Gender (Male or Female) | nominal |

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

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

3/8 or 0.375

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

1. Equal to 1- 0/36=0
2. Less than or equal to 4 - 6/36
3. Sum is divisible by 2and 3 – 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?

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.

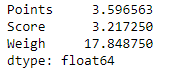
Child B – probability of having 4 candies = 0.20

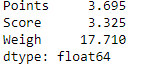
ANSWER: 3.09

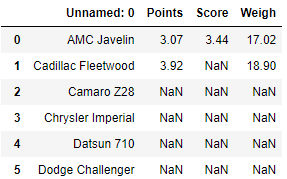
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

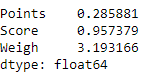
* For Points,Score,Weigh>

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

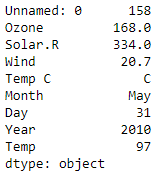
MEAN:

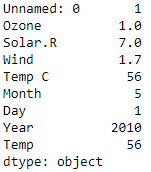
MEDIAN: 

MODE: 

VARIANCE: 

STANDARD DEVIATION: std.PNG

MAX: 

MIN:

**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?

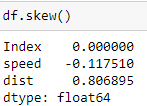
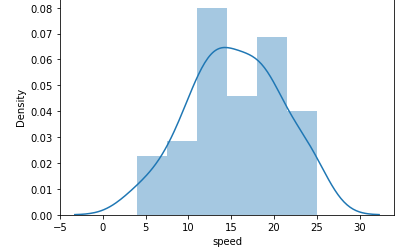
ANSWER:145.33

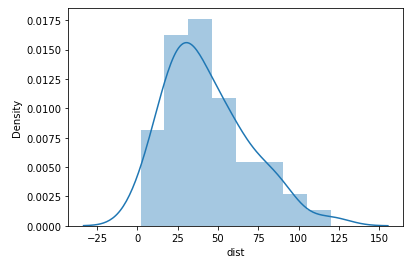
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**SKEWNESS:**

** **

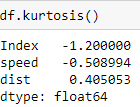
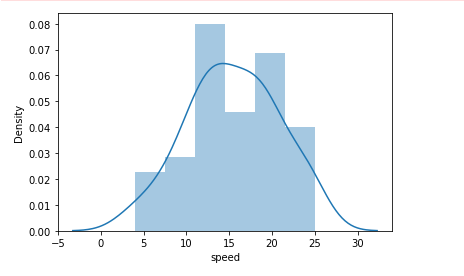
****

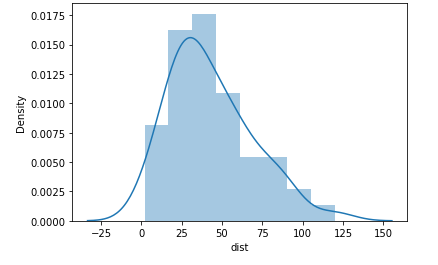
**INFERENCES:**

***SPEED***: The skewness value is negative, then it is stated as negative skewed. It has lower number of data points having low values

***DISTANCE:*** The skewness value is Positive, then it is stated as positive skewed. It has higher number of data points having low values

**KURTOSIS**

** **

****

**INFERENCES:**

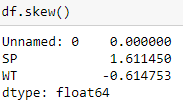
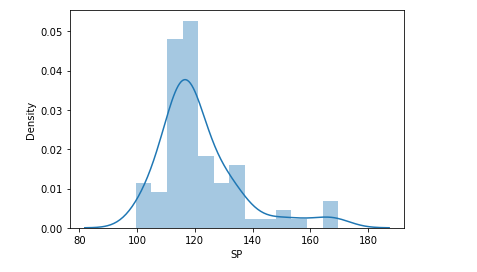
***SPEED*:** It has negative values of kurtosis indicate that a distribution is flat and has thin tails.

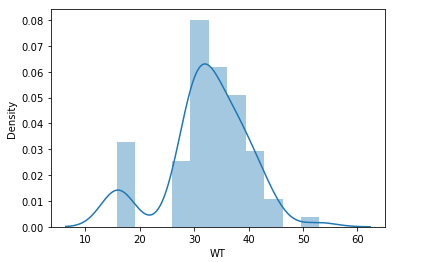
***DISTANCE:*** It has positive values of kurtossis indicate that a distribution is peaked and has thick tails.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**SKEWNESS:**

** **

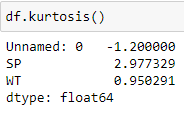
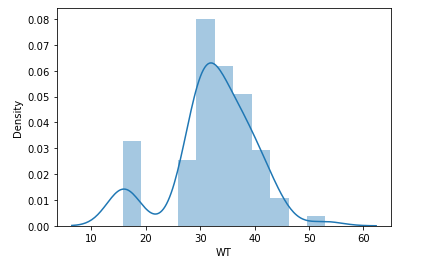
****

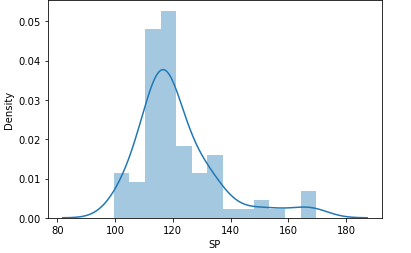
**INFERENCES:**

***SP:***  The value of the skewness is positive then it is stated as positive skewed or right skewed.

***WT:*** The value of the skewness is negative then it is stated as negative skewed or left side distribution.

**KURTOSIS:**

** **

****

**INFERENCES:**

***SP:*** The distribution is peaked and posses thick tails. In the dat obtained kurtosis is possitive then middle values in that dataframe are so high comparedto the left and ride side values.

***WT:*** The kurtosis indicate that a distribution is peakedand posses thick tailsthe ditribution is less than 3 which means to produce less outliers.

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



1. There could be outliers in the right hand side. There ia an extreme frequency on the right hand side.



1. The above boxplot, the distribution has lots of outliers in extreme frequency

**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?

|  |  |  |
| --- | --- | --- |
| Confidence interval | Z-Scores | Range |
| 94% | 1.8808 | 198.74-201.26 |
| 96% | 2.0537 | 198.62-201.37 |
| 98% | 2.3263 | 198.43-201.56 |

**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: 25.29411

STANDARD DEVIATION: 5.052663

1. What can we say about the student marks?
2. Mean is greater than Median, so the distribution is skewed to right hand side

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

ANS: Zero

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

ANS: In a distribution which is skewed to the left, the value of the mean is less than the median. Again the skewness is in the direction of the long tail, which is from the right side thus skewed to right. The large values tend to pull the mean to the right so its little larger than the median

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

ANS: In a distribution which is skewed to the left, the value of the mean is less than the median, the skewness will be in the direction of long tail, which is from the left side thus skewed to left. The small values tend to pull the mean to the left so its little lower than the median

Q16) What does positive kurtosis value indicates for adata ?

ANS: Positive excess values of kurtosis(>3) indicate that a distribution is peaked and possess thick tails. Leptokurtic distributions have positive kurtosis values. A lepokurtic distribution has a higher peak and taller tails than normal distribution

Q17) What does negative kurtosis value indicates for a data? ANS: Negative excess values of kurtosis(<3) indicate that a distribution is flat and possess thin tails. Leptokurtic distributions have negative kurtosis values. A platykurtic distribution is flatter when compared with the normal distribution, with fewer values in its shorter tails.

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



What can we say about the distribution of the data?

1. Not normaly distributed

What is nature of skewness of the data?

1. The nature of the skewness is negative

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

A)10- 18

Q19) Comment on the below Boxplot visualizations?



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

1. Both box plots mentioned above are normally distributed

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

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG< Cars$MPG

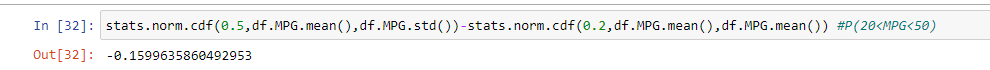
* 1. P(MPG>38)

mpg38.PNG

* 1. P(MPG<40)

mpg40.PNG

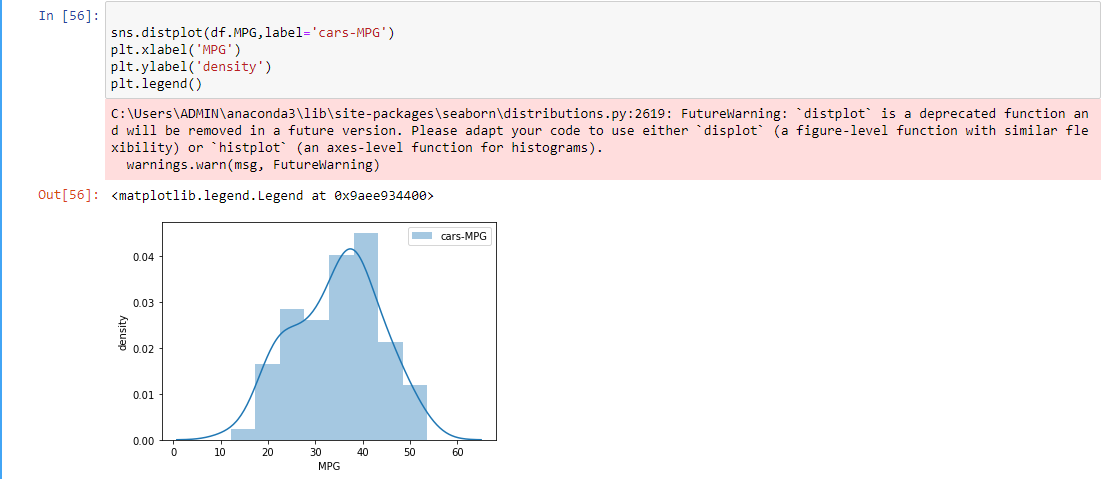
c. P (20<MPG<50)

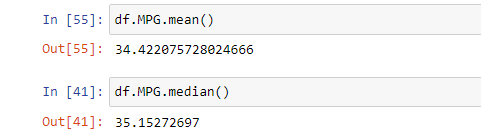


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

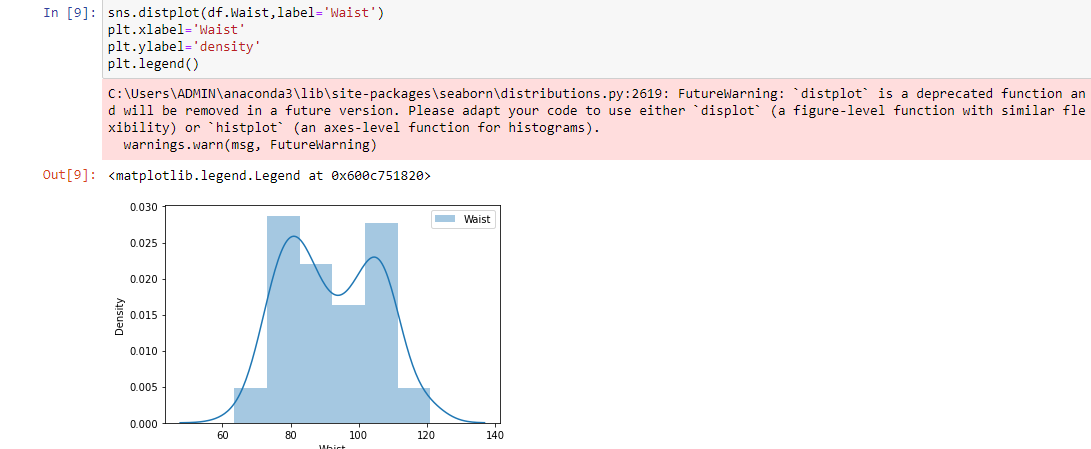




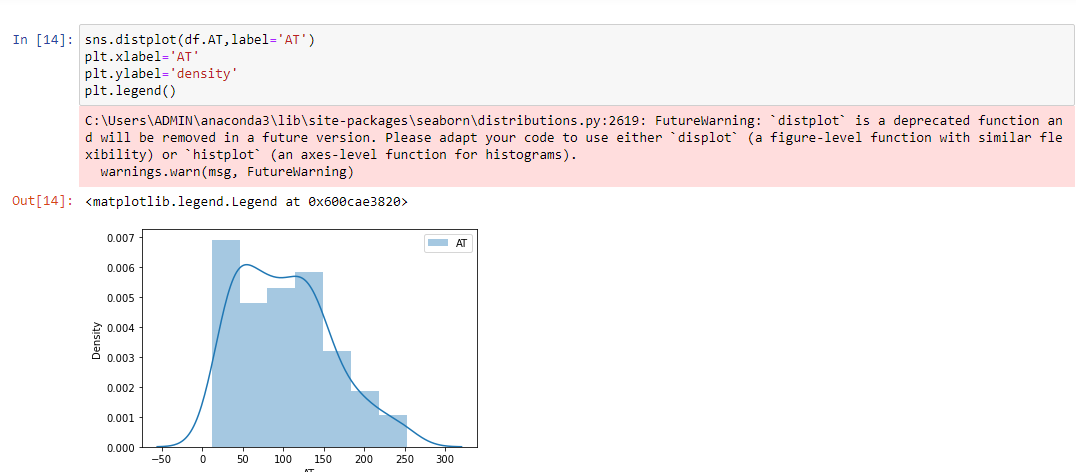
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

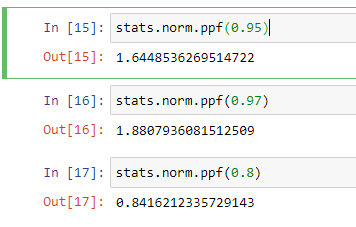
WAIST:



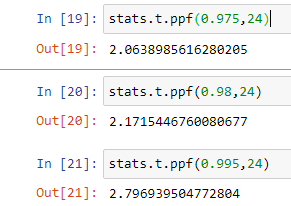
AT:



Q 22) Calculate the Z scoresof 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 companyclaims 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

A)n=18, s=90, xbar=260, mean=270

Tscore=(xbar-mean)/(s/sqrt(n))=270-260/90/sqrt(18))

=-10/21.23

=0.4714045

