|  |  |
| --- | --- |
| 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 | Ratio |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Nominal |
| SAT Scores | Ordinal |
| Years of Education | Ratio |

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

ANS : Total possible events = 8

Number of desired events = 3

Probability = 3/8

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

1. Equal to 1 ANS : 0
2. Less than or equal to 4 ANS : 1/6 =0.167
3. Sum is divisible by 2 and 3 Ans : 5/6 =0.139

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 : 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

Ans : Expected number = E(x)

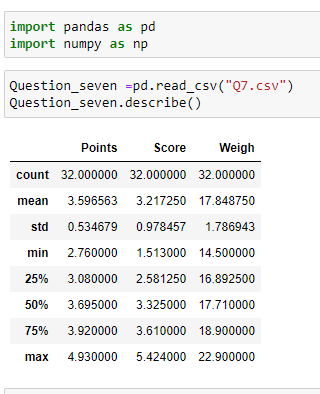
E(x)= 1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120 = 3.09

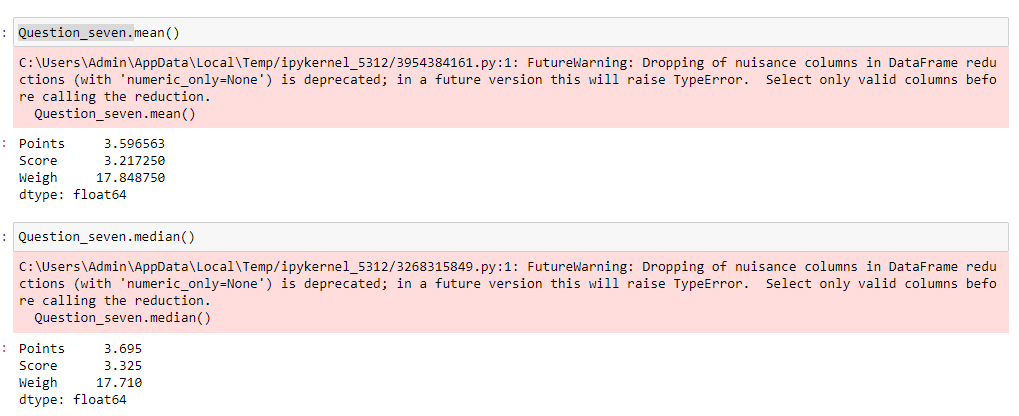
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

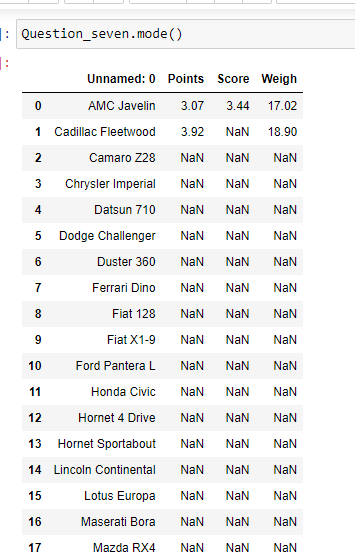
* For Points,Score,Weigh>

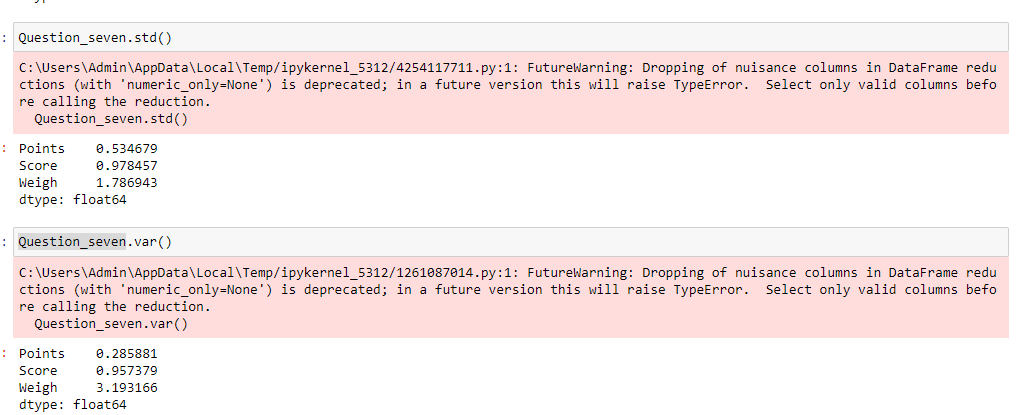
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

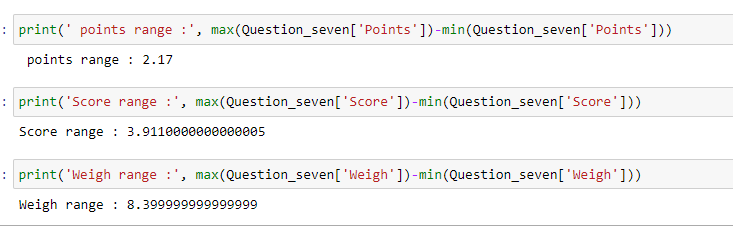
**Use Q7.csv file**











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

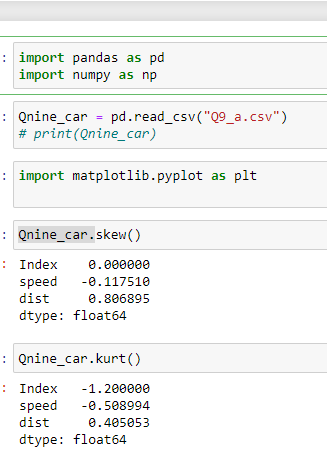
Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

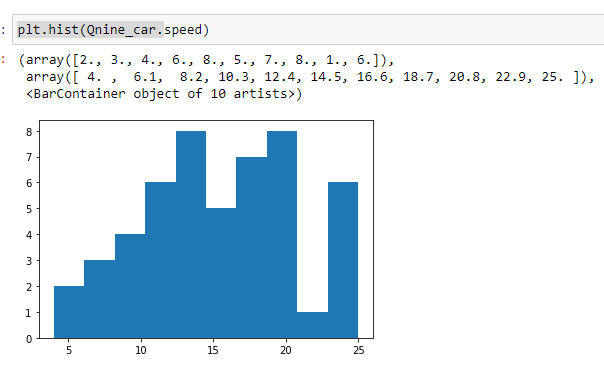
Ans : 145.33

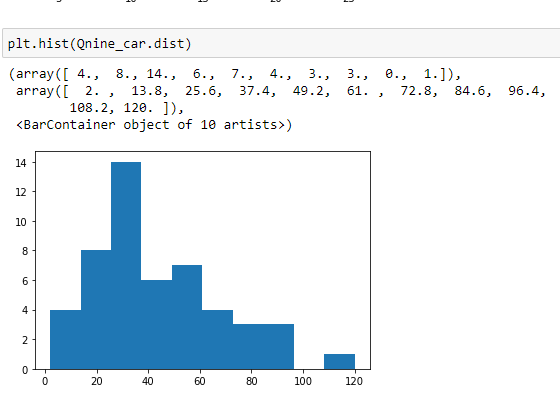
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

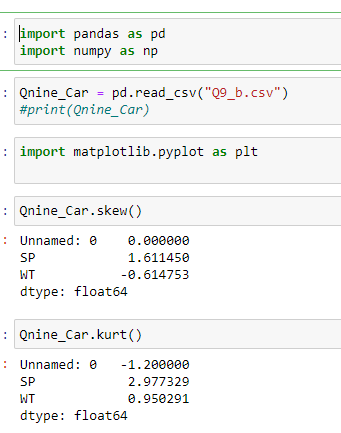


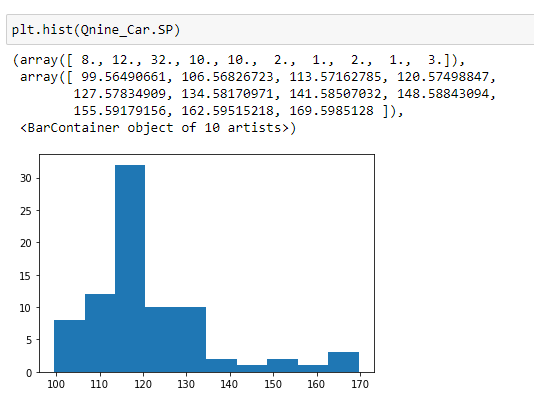


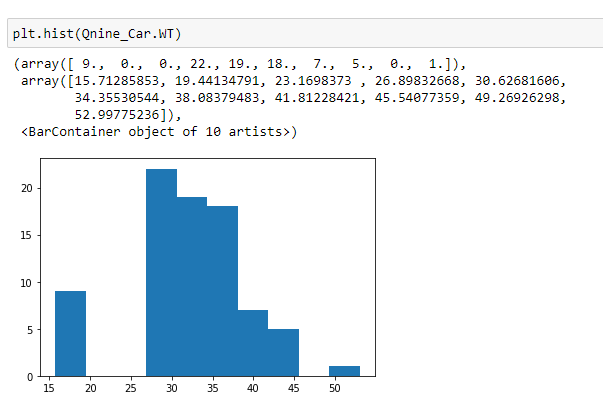


**SP and Weight(WT)**

**Use Q9\_b.csv**







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



Ans: Inference : The distribution is right skew(positive), Mean > Median



Ans: The distribution has lot of outliers towards upper extreme.

**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 ; sd=30 ; mean = 200

Confidence interval:𝒎𝒆𝒂𝒏± Z𝒂/𝟐 𝒔𝒅/√𝒏

**For 94%:**

=200±Za/2\*30/sqrt of 2000

[Za/2= 0.94/2(check z table)=1.88]

=200±1.88\*30/sqrt of 2000

=198.74 – 201.26

= -2.52

**For 98%:**

=200±2.33\*30/sqrt (2000)

=198.44-201.56

= -2.82

**For 96%:**

=200 ± 2.05\*30/sqrt (2000)

=198.62-201.38

= -2.64

**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?

Ans : Mean = 41 ; Median = 40.5

Variance = 25.52941

Standard Deviation = 5.0526

1. Mean > Median.

The distribution is slightly skewed towards right.

No outliers are present.

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

Ans : Skewness = 0

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

Ans : Skewness = positive. (Data is distributed more on left side)

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

Ans : Skewness = negative. (Data is distributed more on right side)

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

Ans : High and Narrow peak on central part of data.

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

Ans : Wider peak on central part of data.

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



What can we say about the distribution of the data?

Ans : The data is not symmetric .Data is distributed more towards right side.

What is nature of skewness of the data?

Ans : Skewness = Negative

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

Ans : IQR data is 8 . (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 :

🡪 Data is normally distributed , no outliers , center around 262.5

(comparatively first graph has less range.)

🡪 Data is normally distributed , no outliers , center around 262.5

(Comparatively second graph has more range.)

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

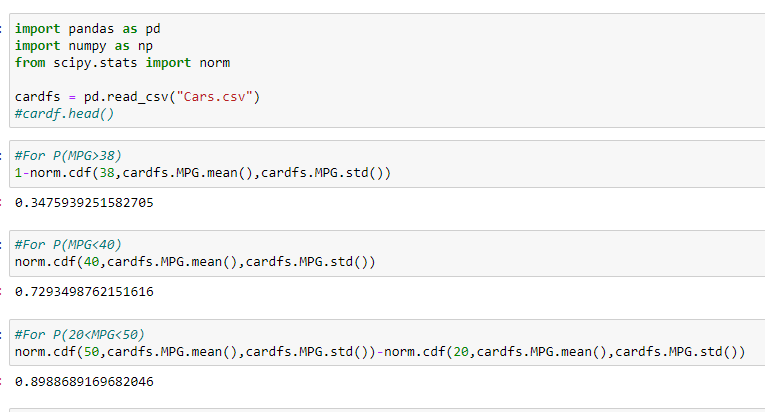
Data \_set: Cars.csv

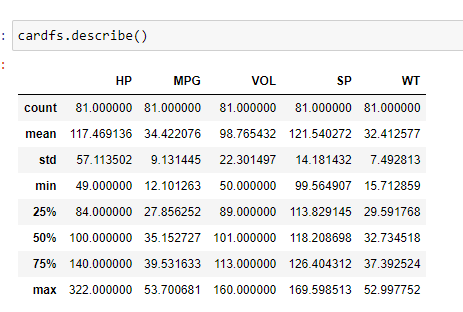
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

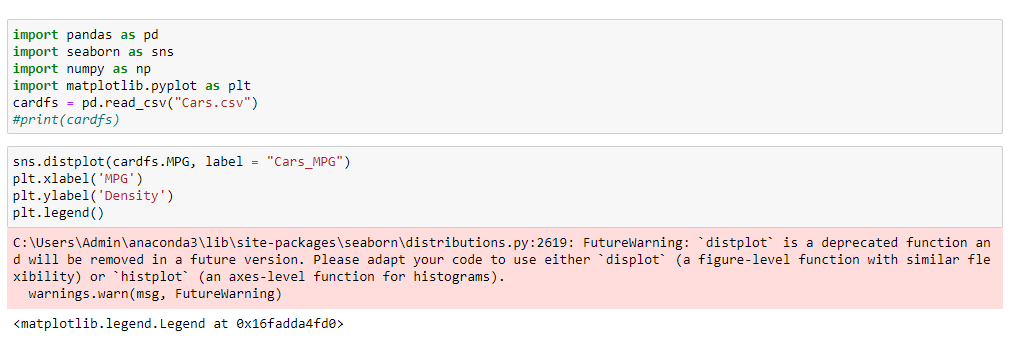


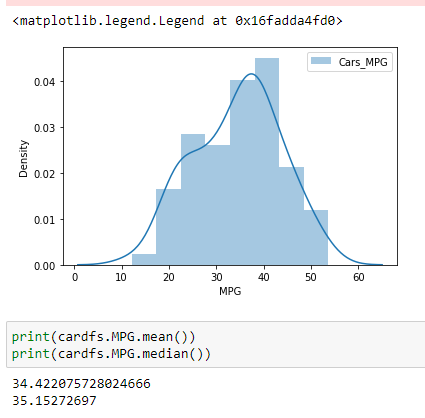


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

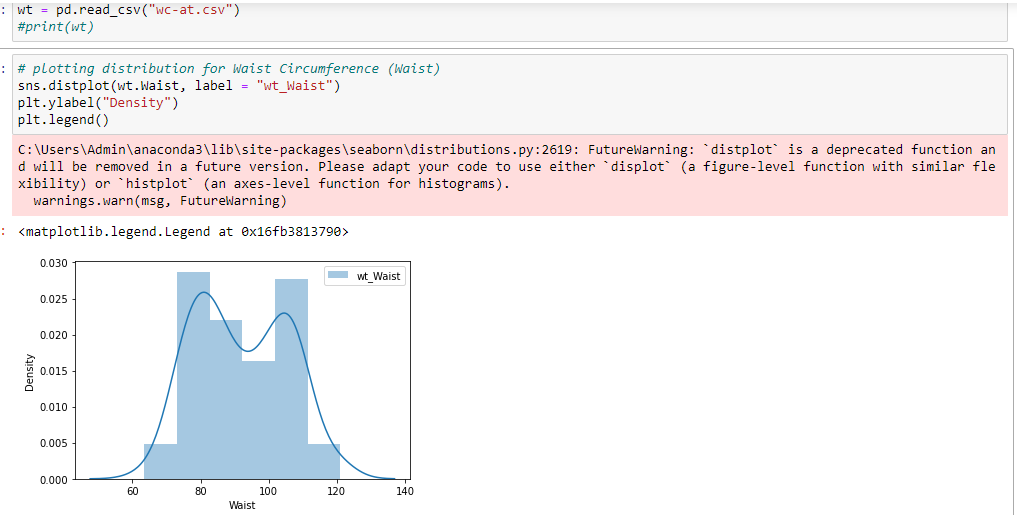
Dataset: Cars.csv

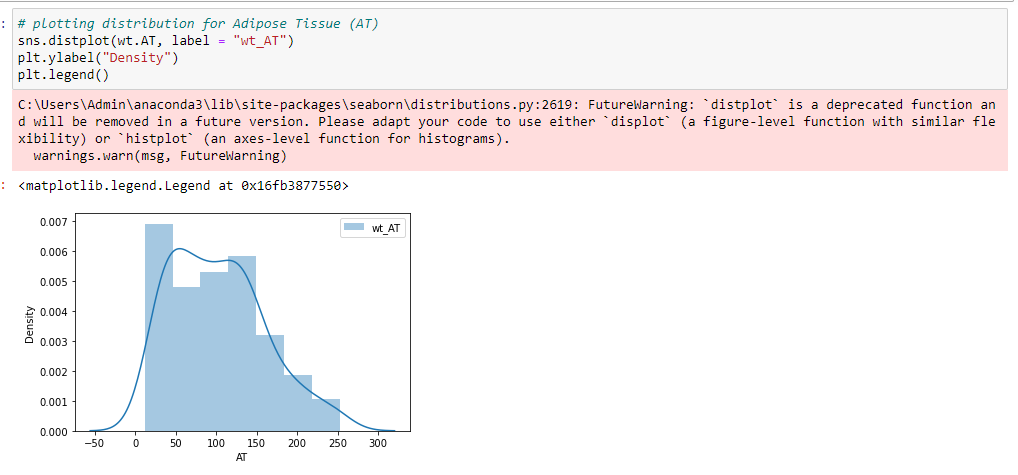


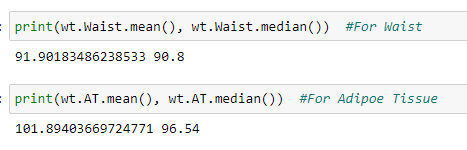


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

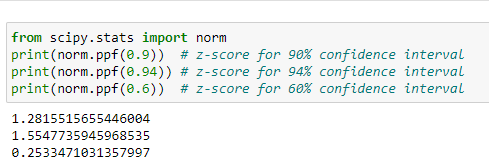
Dataset: wc-at.csv



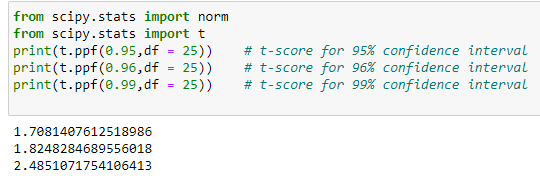




Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval



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:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

Ans:

t-score= -0.4714

Degree of freedom= 17

P(t) = 0.3216725

