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

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

(**HT)(HT)(HT)**

**{(HHH)(HTH)(HHT)(HTT)(THH)(TTH)(THT)(TTT)**

**Probability of two heads and one tail 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

**(1,2,3,4,5,6)(1,2,3,4,5,6))**

**The sum is equal to 1P(E)=0**

**(1,1)(1,2)(1,3)(2,1)(2,2)(3,1)**

**The sum is less than or equal to 4 P(E)=6/36=1/6**

**(1,1)(1,2)(1,3)(1,4)(1,5)(1,6)**

**(2,1)(2,2)(2,3)(2,4)(2,5)(2,6)**

**(3,1)(3,2)(3,3)(3,4)(3,5)(3,6)**

**(4,1)(4,2)(4,3)(4,4)(4,5)(4,6)**

**(5,1)(5,2)(5,3)(5,4)(5,5)(5,6)**

**(6,1)(6,2)(6,3)(6,4)(6,5)(6,6)**

**The probability of Sum is divisible by 2 and 3 is P(E)=6/36**

**(1,5)(2,4)(3,3)(4,2)((5,1)(6,6)**

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?

**No.of balls in a bag=7**

**No.of ways of balls are drawn out of 7=7C2=7!/(7-2)!\*2!=7\*6/2=21**

**No.of ways of balls are drawn none of blue=5C2=10**

**probability that none of the balls drawn is blue P(E)=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

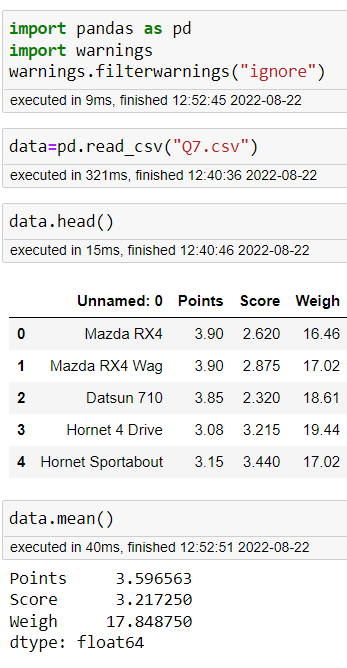
**Expected number of candies for a randomly selected child=1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120=3.09**

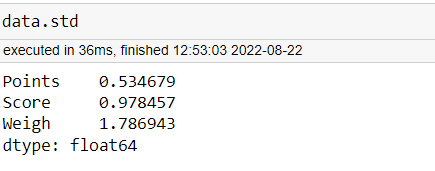
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

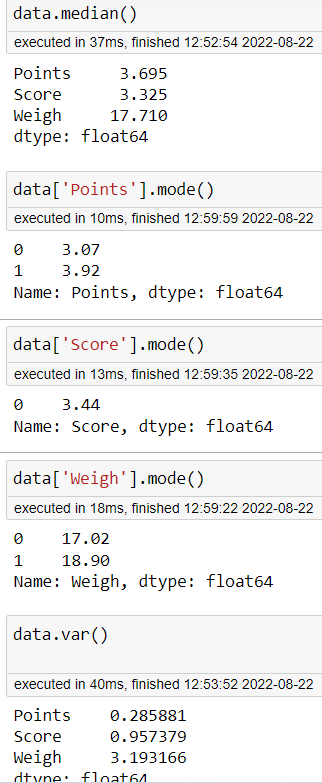
* For Points,Score,Weigh>

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?

**Probability of selecting one patient=1/9=0.111**

**Expected Value of the Weight of that patient=0.111\*108+0.111\*110+0.111\*123+0.111\*134+0.111\*135+0.111\*145+0.111\*167+0.111\*187+0.111\*189=144.08**

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

Cars speed and distance

Use Q9\_a.csv

**SP and Weight(WT)**

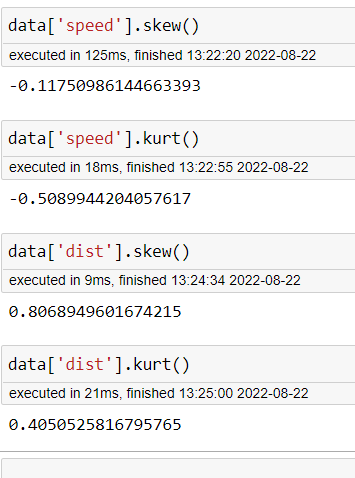
**Cars data Speed is negatively skewed.**

**Speed Skewness= -0.1139548**

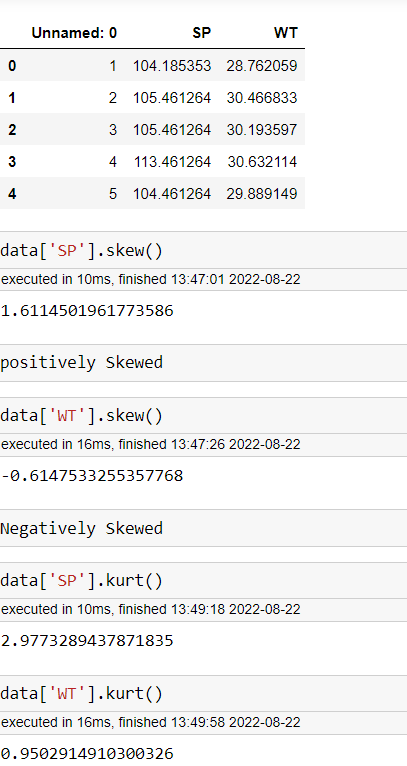
**Cars data Weight is positively skewed**

**Weight skewness=0.7824835**

**kurtosis(Cars$SP)=5.723521**



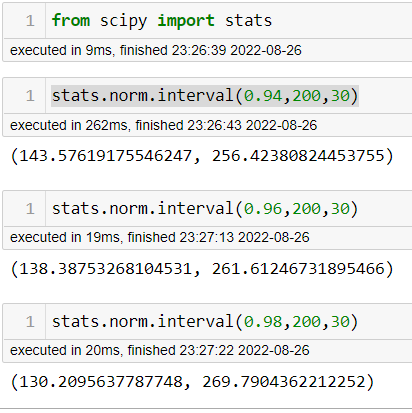
**Use Q9\_b.csv**

****

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



**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?



**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(Score)

[1] 41

> median(Score)

[1] 40.5

> mfv(Score)

[1] 41

> var(Score)

[1] 25.52941

> sd(Score)

[1] 5.052664

> range(Score)

[1] 34 56

1. What can we say about the student marks?

Students Scores are positively Skewed.

Score skewness=1.542885

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

**Perfectly symmetric and skewness for data equals to zero.**

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

**The data is right skewed and skewness is positive.**

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

**The data is left skewed and skewness is negative.**

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

**Positive values of kurtosis indicates that distribution is peaked and possesses thick tails. An extreme positive kurtosis(>3) indicates a distribution where more of the numbers are located in the tails of the distribution instead of around the mean.**

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

**Negative excess values of kurtosis(<3) indicate that distribution is flat and has thin tails**.

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



What can we say about the distribution of the data?

Q1=10,Q2=15,Q3=18

Range=19-1=18

What is nature of skewness of the data?

The data is Left skewed or 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.

Both box plots medians are same.

Boxplot1 and 2 both are symmetric.

Boxplot1 Boxplot2

Min Min

Q1 -250 Q1-225

Q2 -265 Q2-265

Q3 -275 Q3-320

IQR=10 IQR=55

Range=237-287=50 Range=335-200=150

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)

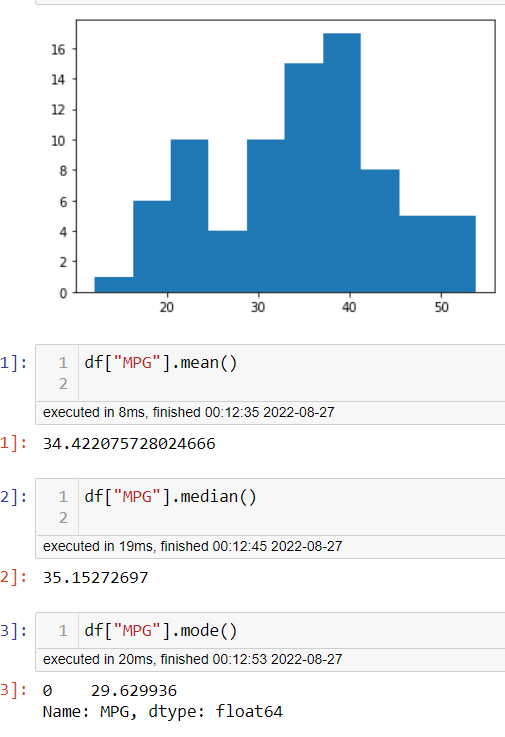
c. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

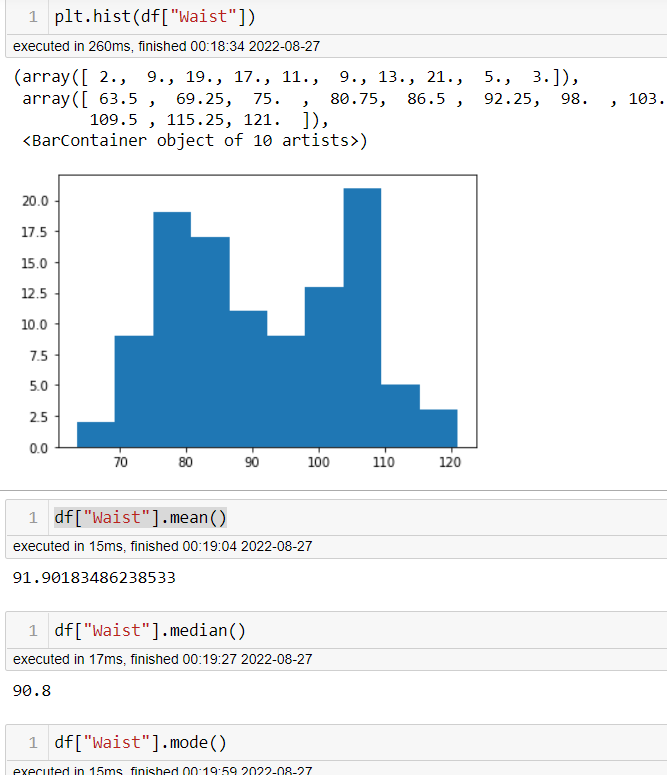
Dataset: Cars.csv



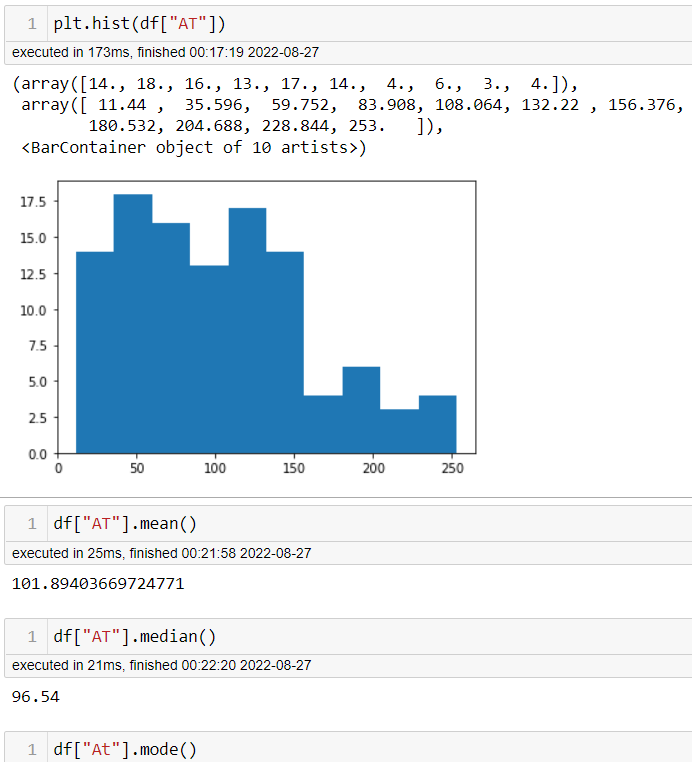
It’s not a normal distribution.

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

Dataset: wc-at.csv

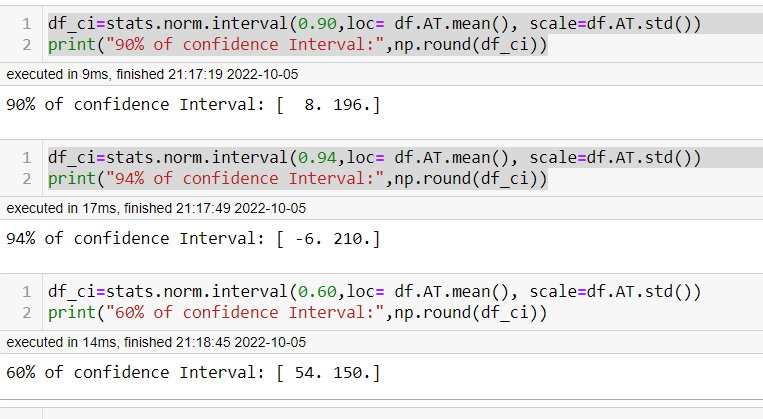


Waist is not follows the Normal distribution.



AT also not follows the 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:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom