Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| 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 |

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

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

Answer: -

Three coins are tossed the total number of possible combinations are 23 = 8

Following is the combination: [HHH, HHT, HTH, THH, TTH, THT, HTT, TTT]

Probability: - **Two heads and one tail[**HTH,HHT,THH**]**

When three coins are tossed at once, the probability of getting two heads and

one tail is 3/8, or 0.375

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

Answer: - Total number of possible outcomes=62 = 36

1. Equal to 1: - = 0 (i.e., it is not possible for the sum to always exceed 1)

Required probability

1. Less than or equal to 4: - 6 (i.e. (1,1) (1,2) (1,3) (2,1) (2,2) (3,1))

Required probability

1. Sum is divisible by 2 and 3: - 6(i.e. (1,5) (5,1) (3,3) (2,4) (4,2) (6,6))

Required probability

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?

Answer: - Total ball: 2 yellow, 3 green, and 2 blue = 7 balls

It can be expressed as 7C2, which means that none of the balls that are 7- 2 (the blue ball) are equal to 5 balls. 5C2

So 5C2/7C2 = (5! /3! \*2!) / (7! /5! \*2!) = 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: - The 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 \* 0.12

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

= 3.090

3.09 is the expected number of candies for a randomly selected child.

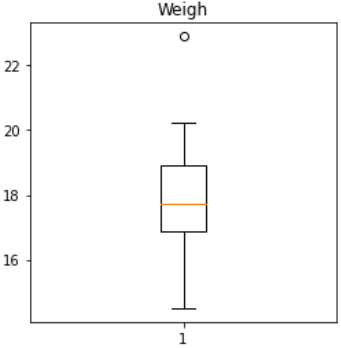
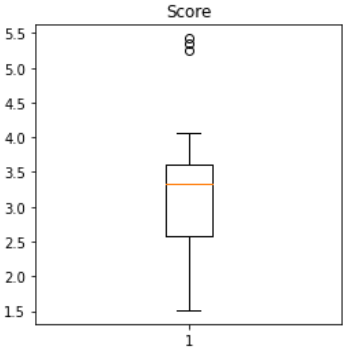
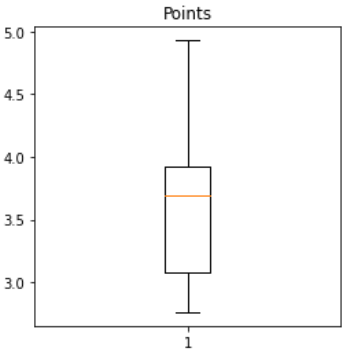
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points, Score, Weigh>

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

Answer: -

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | **3.596563** | **3.217250** | **17.848750** |
| Median | **3.695** | **3.325** | **17.710** |
| Mode | **3.07** | **3.44** | **17.02** |
| Variance | **0.285881** | **0.957379** | **3.193166** |
| Standard  Deviation | **0.534679** | **0.978457** | **1.786943** |
| Range | **2.76,4.93** | **1.513,5.424** | **14.5,22.9** |

****

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: - Sum (X \* Probability of X) = Expected value

= (1/9) (108) + (1/9)(110)+ (1/9)(123)+ (1/9)(134)+ (1/9)(145)+ (1/9)(167)+

(1/9) (187) + (1/9)(199)

= 145.33

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

**Car’s speed and distance**

**Use Q9\_a.csv**

**Answer: -**

**Cars Speed Skewness value= -0.12 and Distance skewness value= 0.81**

**Cars Speed Kurtosis value= -0.51 and Distance kurtosis value= 0.41**

**SP and Weight (WT)**

**Use Q9\_b.csv**

**Answer: -**

**Cars Speed Skewness value= 1.61 and Distance skewness value= -0.61**

**Cars Speed Kurtosis value= 2.98 and Distance kurtosis value= 0.95**

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



Answer :- The peak of the histogram has a right skew, and the tail is on the right.

Mean > Median

The histogram has a Outliers in the end



**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?

Answer: Using the t-distribution,

The 94% confidence interval is (198.73, 201.27)

The 96% confidence interval is (198.61, 201.39)

The 98% confidence interval is (198.43, 201.57)

**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.

Answer:-

Mean :- **41.0**

Median :- **40.5**

Variance :- **25.529412**

standard deviation :- **5.052664**

1. What can we say about the student marks?

Answer:- There are no outliers, and the data is slightly skewed to the right because the mean is greater than the median.

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

Answer:- We have a perfectly symmetrical distribution because there is no skewness

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

Answer:- The mean of positively skewed data will be greater than the median

the distribution is skewed to the right.

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

Answer:- The data distribution is skewed to the left, when the mean is frequently less than the median.

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

Answer:- Positive kurtosis values indicate that the distribution is peaked and has thick tails.

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

Answer:- A negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

Q18) Answer the below questions using the below box plot visualization.



What can we say about the distribution of the data?

* The median in the above Box plot is not normally distributed; it is higher.

What is the nature of skewness of the data?

* The data has a leftward skew. The minimum value whisker range is larger than the greatest value

What will be the IQR of the data (approximately)?   
 - The interquartile range is calculated as Q3 Upper Quartile - Q1 Lower Quartile = 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.

Answer:- For comparative purposes, there are no outliers. Second, both the box plot and the whisker plot have the same median, which is roughly between 275 and 250, and they are normally distributed with zero to no skewness at either the minimum or maximum whisker range.

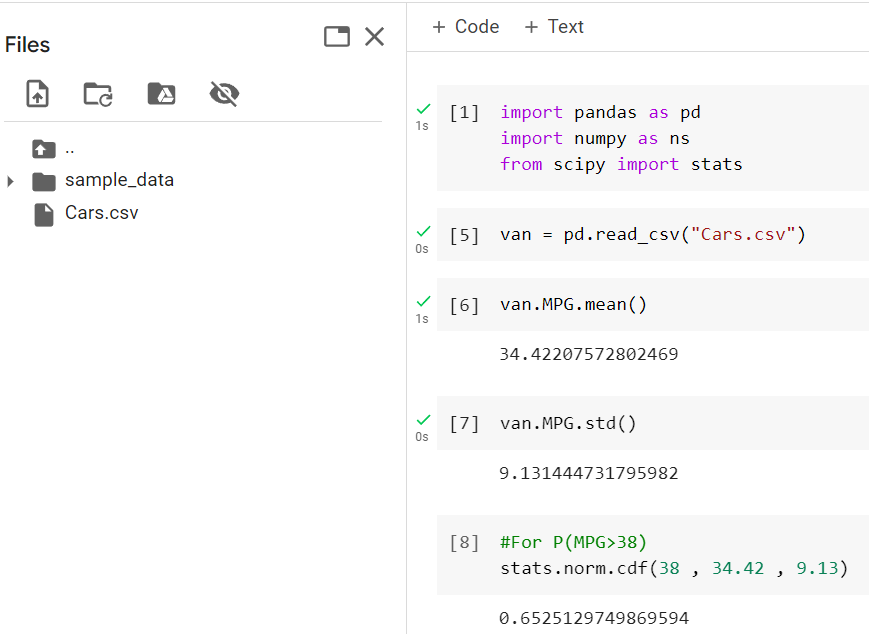
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

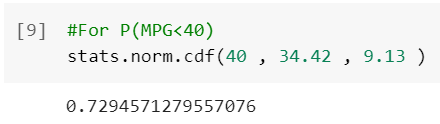
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

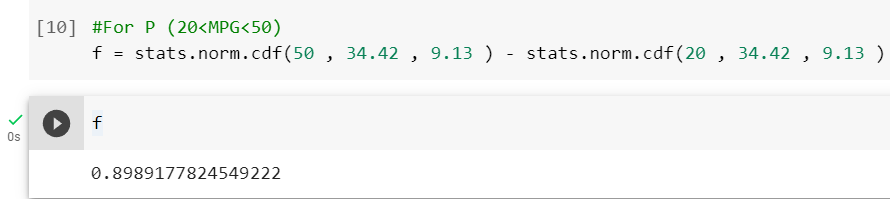
* 1. P(MPG>38)



* 1. P(MPG<40)



c. P (20<MPG<50)

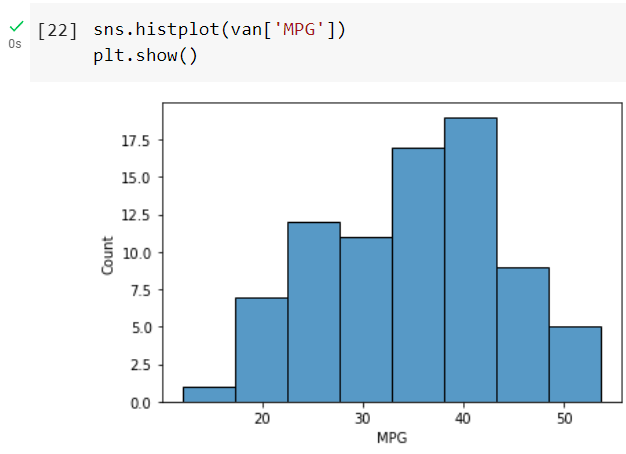


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

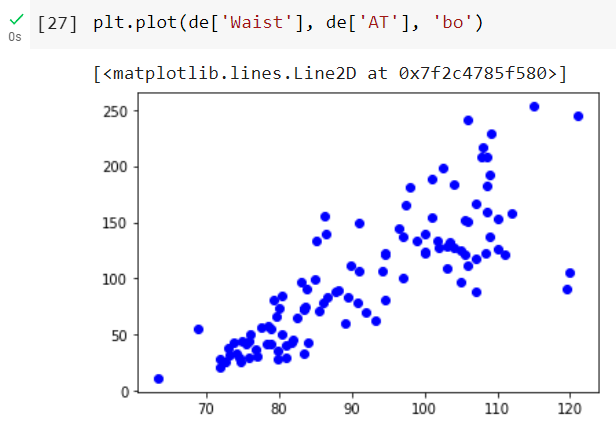
Answer:- MPG of cars not follows normal distribution



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

Dataset: wc-at.csv

Answer:- Adipose Tissue (AT) and Waist does not follow Normal Distribution



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

Answer:- z = 1.555

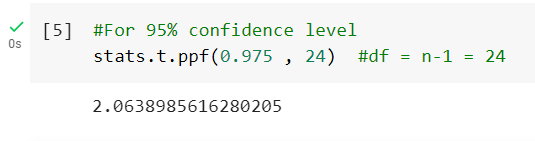
Therefore, we get that the z score at 90 % confidence interval is 1.645

For z score at 94 % confidence interval is 1.555

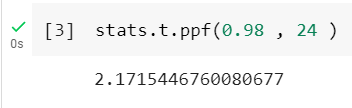
For z score 60 % confidence interval is 0.253

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

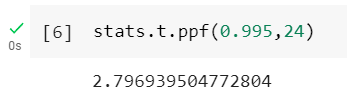
For 95% confidence interval



For 96% confidence interval



For 99% confidence interval



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

Answer: - Sample Mean = 260

Mean Population = 270

Sample Size = 18

Std. deviation Sample = 90

T\_Score = (Sample mean – Population mean)/(sample\_std / (sample size\*\*0.5))

(Degree of Freedom) Df = sample size – 1

Probability = stats.t.cdf(T\_Score,df)

Print(“Probability:”, probability)

**Probability: 0.32167253567098364**