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

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?

Total Possible Events: 8

(HHH,HTH,HTT,TTT,THT,THH,HHT,TTH)

No of desired events: 3

P = 3/8

P= 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

When two dice are rolled the total number of possible outcomes are

N= {(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),}

N = Total = 36

1. Equal to 1

Possible Outcomes = 0/36 = 0

1. Less than or equal to 4

Possible Outcomes = (1,1), (1,2), (1,3), (2,1), (2,2), (3,1)

P = 6/36 = 0.166

3. Sum is divided by 2 & 3

Possible Outcomes = (1,5), (2,4), (3,3), (4,2), (5,1), (6,6)

P = 6/36 = 0.166

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?

Total number of outcomes to draw 2 balls at random from 7 balls,

n(P) = 7C2 = = = 21

Let E be an event

Number of chances to draw two other than blue ball is,

n(E) = 2C2 + 3C2 + 2C1\*3C1

= 1+3+2\*3 = 10

Probability of none of the balls drawn is blue is,

P = =

P = 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

Let Candies count be X

Probability of candies =P(X)

Expected number = Summation of [X\*P(X)]

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

= 3.09

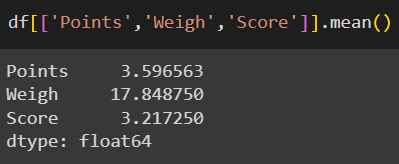
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

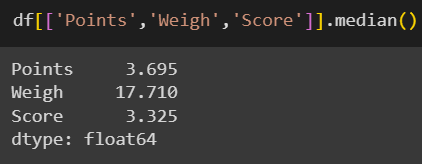
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

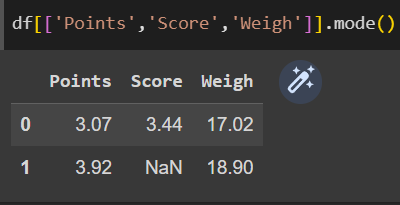
1. **Mean**

****

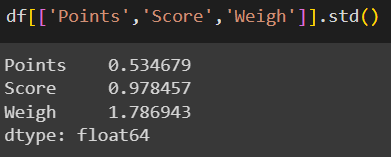
1. **Median**



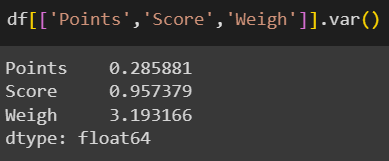
1. **Mode**



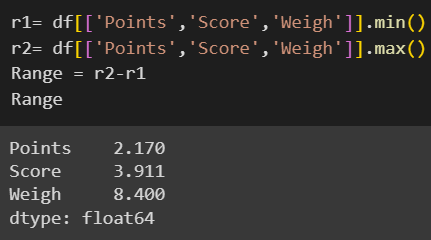
1. **Standard Deviation**

****

1. **Variance**

****

1. **Range**

****

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 = Sum of (probability \* value)

=∑ P(x). E(x)

= there are 9 patients

Probability of selecting each patient =

E(x) = {108, 110, 123, 134, 135, 145, 167, 187, 199}

P(x) = {1/9,1/9,1/9,1/9,1/9,1/9,1/9,1/9,1/9}

Expected value = {(1/9)(108)+(1/9)(110)+(1/9)(123)+(1/9)(134)+(1/9)(135)+

(1/9)(145)+ (1/9)(167)+(1/9)(187)+(1/9)(199)}

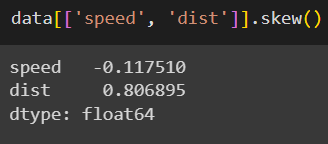
Expected Value = **145.33**

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

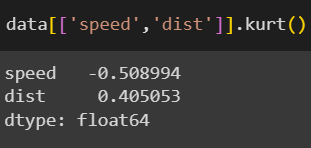
**Cars speed and distance**

**Use Q9\_a.csv**

**Skewness:** It measures the asymmetry of the distribution.

****

**Kurtosis:** It measures the tailedness of distribution

****

**Inferences:**

Distance is positively skewed.

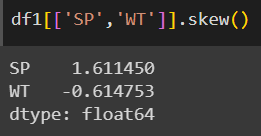
Car Speed is negatively skewed.

Both Distance and Car Speed has positive kurtosis.

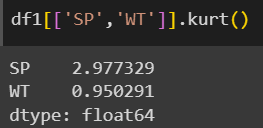
**SP and Weight(WT)**

**Use Q9\_b.csv**

**Skewness:**

****

**Kurtosis:**

****

**Inferences:**

SP is positively skewed.

WT is negatively skewed.

Both WT and SP has positive kurtosis.

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



The above data is positively skewed

Majority of the Chick Weight is between 50 to 150

We can conclude that there are 3 frequency i.e. >50 ,51-150, 150-200.

Mean>Median



The above data has Outliers.

It is positively skewed.

Outlier is on the right side.

**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?

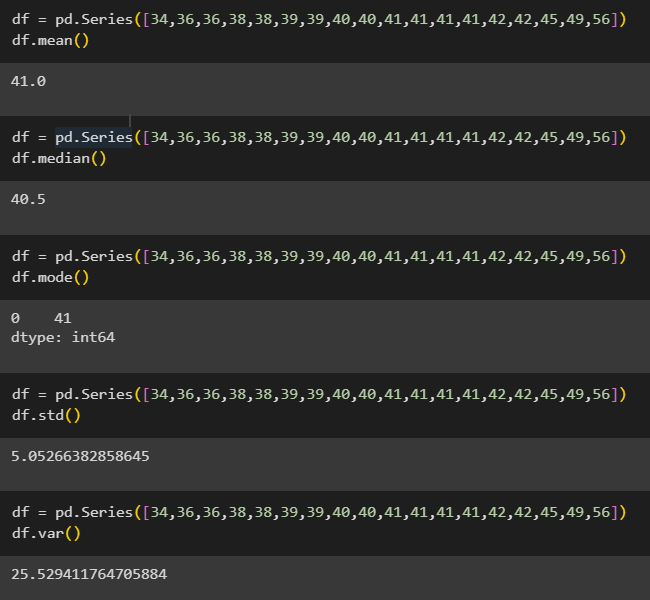
|  |  |  |  |
| --- | --- | --- | --- |
|  | 94% | 98% | 96% |
| Upper | 201.04 | 201.38 | 201.17 |
| Lower | 198.96 | 198.62 | 198.83 |

**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:**



1. What can we say about the student marks?

**Ans:**

Mean > Mean.

The distribution is positively skewed.

The data consist of outliers

Majority student scored between 35 to 45 marks

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

No skewness is present, it is always zero and perfectly symmetric bell-shaped curve.

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

Skewness is positive and data will be more on right.

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

Skewness is negative and data will be more on left.

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

Positive kurtosis means it is high and narrow peak on central part and leptokurtic.

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

Negative kurtosis means it is broader peak on central part.

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



1. What can we say about the distribution of the data?

The above boxplot is not normally distributed and more concentrated on right side.

1. What is nature of skewness of the data?

Skewness is negative. The whisker range of minimum value is greater than the maximum value.

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

IQR = Upper Quartile – Lower Quartile

IQR = 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 the Boxplot is normally distributed in both the graph.
* There are no outliers in both the graph.
* Range is less in first graph as compared to second graph.
* No skewness is present at the max and min whisker range.
* Both the plot shares the same median.
* Median ranges from 250 to 275.

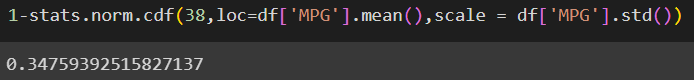
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

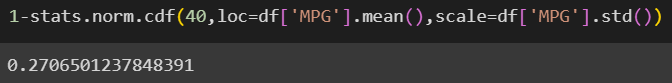
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

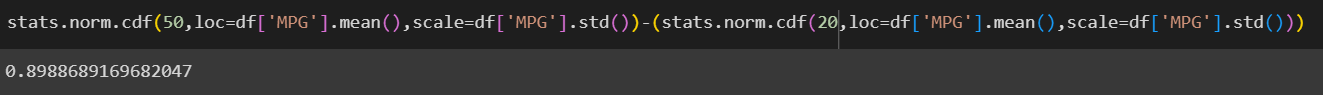
* 1. **P(MPG>38)**



* 1. **P(MPG<40)**



**c. P(20<MPG<50)**

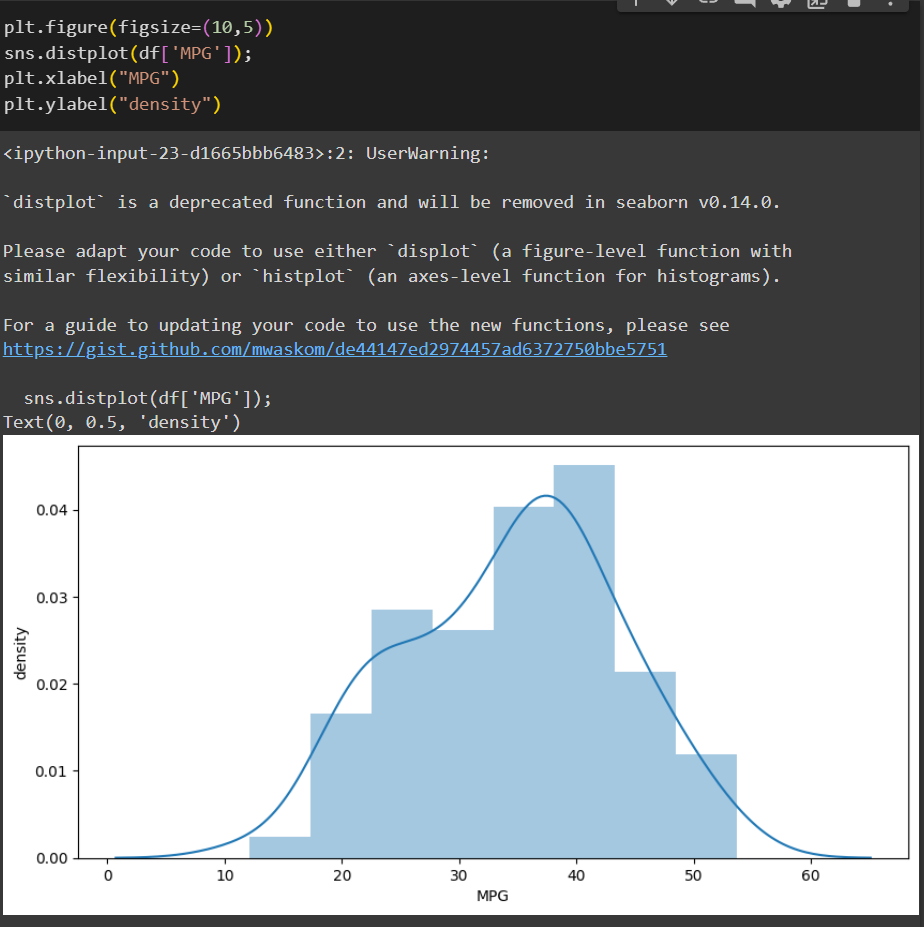


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

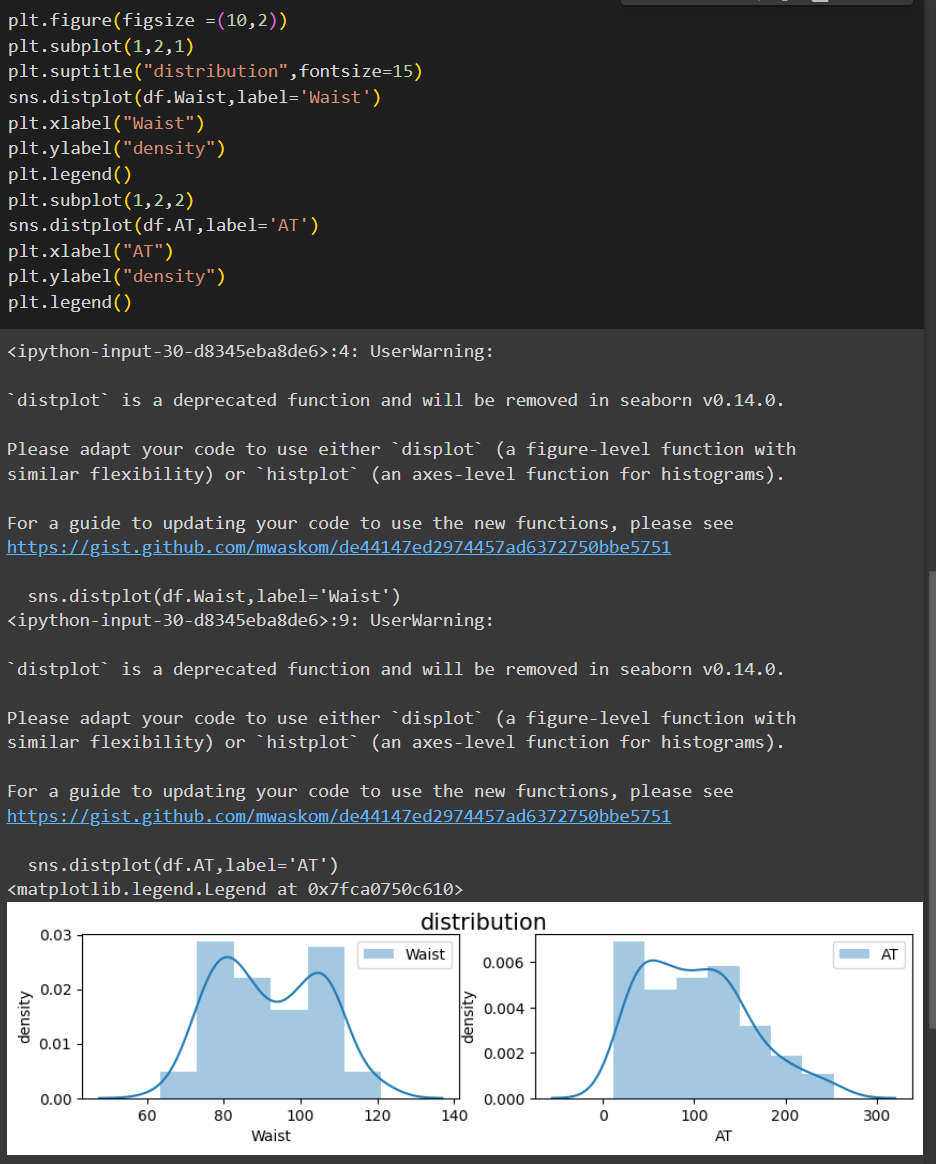
Dataset: Cars.csv

MPG of cars has normal distribution.



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

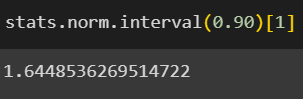
Dataset: wc-at.csv

**Ans:** Both Adipose Tissue(AT) and Waist Circumference(Waist) doesn’t follows normal distribution.

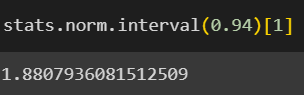
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**Ans:**

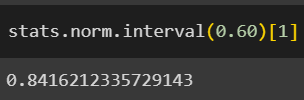
**Z score of 90% Confidence Interval:**

****

**Z score of 94% Confidence Interval:**

****

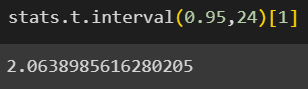
**Z score of 60% Confidence Interval:**

****

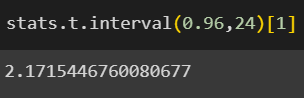
Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25.

**Ans:**

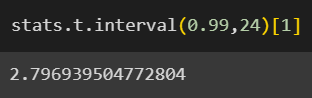
**95% Confidence Level:**



**96% Confidence Level:**

****

**99% Confidence Level:**

****

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:**

df = n-1 = 18-1 = 1

