|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | NUMERICAL(DISCRETE) |
| Results of rolling a dice | NUMERICAL (DISCRETE) |
| Weight of a person | NUMERICAL (CONTINUOUS) |
| Weight of Gold | NUMERICAL (CONTINUOUS) |
| Distance between two places | NUMERICAL (CONTINUOUS) |
| Length of a leaf | NUMERICAL (CONTINUOUS) |
| Dog's weight | NUMERICAL (CONTINUOUS) |
| Blue Color | CATEGORICAL(NOMINAL) |
| Number of kids | NUMERICAL (DISCRETE) |
| Number of tickets in Indian railways | NUMERICAL (DISCRETE) |
| Number of times married | NUMERICAL (DISCRETE) |
| Gender (Male or Female) | CATEGORICAL(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 | NOMINAL |
| Level of Agreement | ORDINAL |
| IQ(Intelligence Scale) | INTERVAL |
| 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 | RATIO |

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

ANS: Sample space ={HHH,HHT,HTH,THH,HTT,THT,TTH,TTT}=8

Number of times two heads and one tail are obtained = 3

The probability that two heads and one tail are obtained= **3/8**

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

Sample space= 6\*6=36

1. Equal to 1

ANS:**0/36= 0**

1. Less than or equal to 4

ANS:**1/6**

**Sum of numbers < or equal to 4={(1,1), (1,2), (2,1), (1,3), (3,1), (2,2)}= 6**

**Therefore, probability= 6/36= 1/6**

1. Sum is divisible by 2 and 3

ANS: **1/6**

Sum divisible by 2 and 3= {**(1,5), (5,1), (2,4), (4,2), (3,3), (6,6)}**

**Therefore probability = 6/36 = 1/6**

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?

ANS: **10/21**

**Sample space = 7 balls**

**Probability first drawn ball is not blue = 5/7**

**Probability second drawn ball is not blue = 4/6**

**Therefore,** the probability that none of the balls drawn is blue = 5/7\*(4/6)

= 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

ANS:**3.09**

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\*0.120) = 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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Mean** | **Median** | **Variance** | **STD** | **Range** | **Mode** |
| **Points** | **3.596563** | **3.695000** | **0.285881** | **0.534679** | **(2.760, 4.930)** | **3.92** |
| **Score** | **3.217250** | **3.325000** | **0.957379** | **0.978457** | **(1.5130, 5.4240)** | **3.44** |
| **Weigh** | **17.848750** | **17.710000** | **3.193166** | **1.786943** | **(14.50, 22.90)** | **17.02** |

* **In all 3 Points, weigh and Score mean, median, mode are almost same.**
* **The data seems to be normal distribution.**
* **However, variance is great for Weigh compared to scores and points.**

**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**

The Expected Value of the Weight of that patient =

(108+ 110+123+ 134+ 135+ 145+ 167+ 187+ 199)/9= 145.33

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

Cars speed and distance

Use Q9\_a.csv

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **Speed** | **-0.114** | **2.423** |
| **distance** | **0.7825** | **3.248** |

* **Speed is negatively skewed whereas distance is positively skewed.**
* **The maximum data of** speed **is distributed on the right side of mean and is distributed on left side of the mean for** distance**.**
* **Both speed and distance have positive kurtosis implies ,they have thin tails.**

SP and Weight (WT) Use Q9\_b.csv

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **SP** | **1.5815** | **5.724** |
| **WT** | **-0.6033** | **3.82** |

* **WT is negatively skewed whereas SP is positively skewed.**
* **The maximum data of** WT **is distributed on the right side of mean and is distributed on left side of the mean for** SP**.**
* **Both SP and WT have positive kurtosis implies ,they have thin tails.**

Q10) Draw inferences about the following boxplot & histogram



**Histogram:**

* **It is asymmetric distribution.**
* **It is right tailed, that is positively skewed.**
* **Outliers are present in the right end.**



**Boxplot:**

* **It indicates the presence of outlier on the right extreme 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?

**ANS:**

**94% CI= (143.576, 256.424)**

**98% CI= (138.388, 261.612)**

**96% CI= (130.21, 269.79)**

**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:25.52 STD:5.05**

1. What can we say about the student marks?

* **The average marks of students are 41.**
* **Min and max marks are 34 and 56 respectively.**
* **Maximum students have scored marks between 38 to 42.**

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

* **It’ll be normal distribution with no skewness.**

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

* **It’ll be negatively skewed.**

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

* **It’ll be positively skewed.**

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

* **Plot of data will have sharp thin peak and data will be denser in short range.**

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

* **Plot of data will have broader peak and data will be spread over wide range.**

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



What can we say about the distribution of the data?

**ANS:**

* **The data is negatively skewed with outliers present on the left side of median.**
* **The median is present around 15.**

What is nature of skewness of the data?

**ANS**: **NEGETIVELY SKEWED**

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

**ANS**: 8

**IQR = (75% value -25% value)= (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:**

* **Both 1 and 2 plot are normal distributed with their median between 250 to 275.**
* **Both of the plots have no outliers having no skewness.**
* **The spread range of data is large for 2 compared 1 plot.**
* **The range is greater for 2 plot than 1 plot.**

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) ANS:**0.34759394041453007**

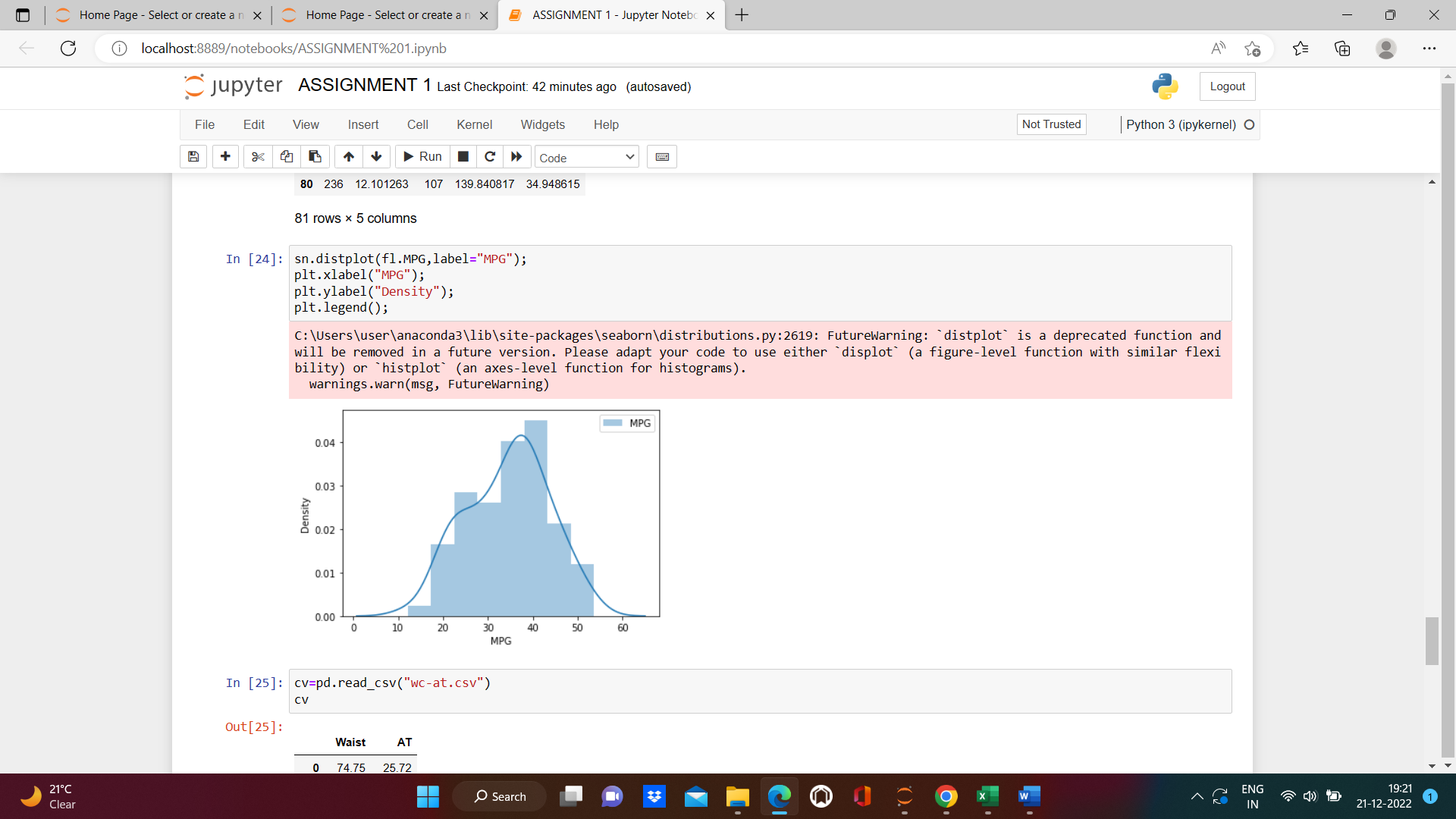
* 1. P(MPG<40) **ANS:** **0.7293498604157946**
  2. P (20<MPG<50) **ANS: 0.8988689076273199**

Q 21) Check whether the data follows normal distribution

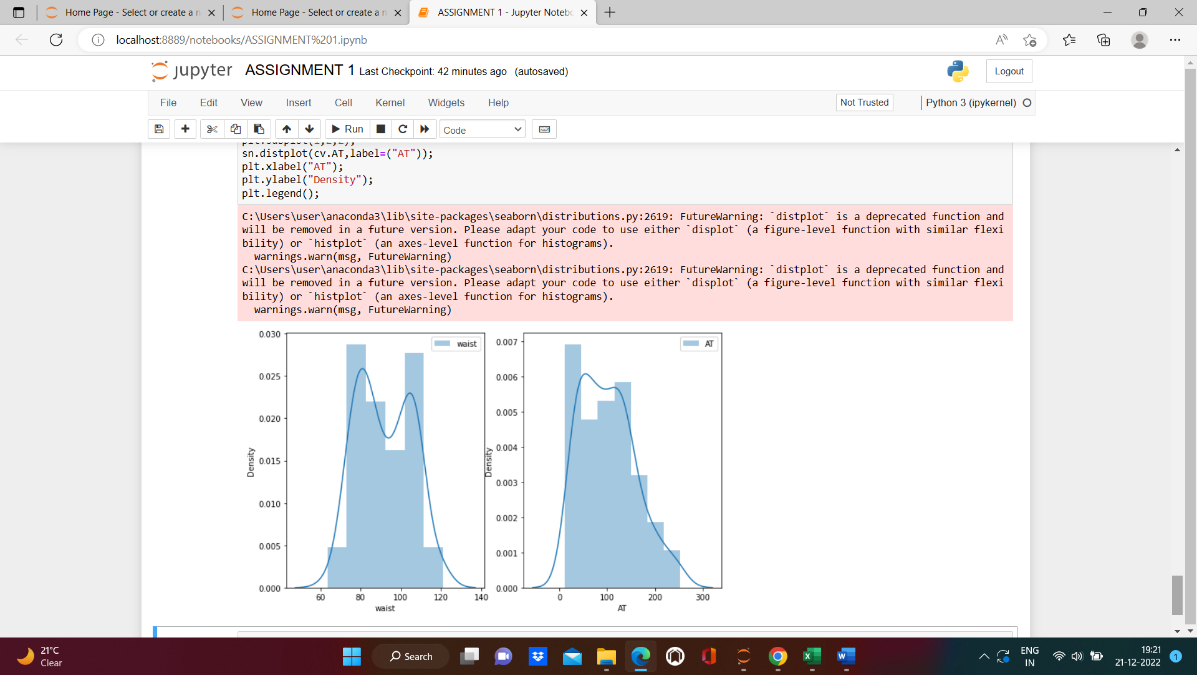
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**It is not 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

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

**ANS:**

**For 90% confidence Interval 1.282**

**For 94% confidence Interval 1.555**

**For 60% confidence Interval 0.253**

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

**ANS:**

**For 95% confidence Interval 1.711**

**For 96% confidence Interval 1.828**

**For 99% confidence Interval 2.429**

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

**ANS:**

**Given**  =270, n=18, sample mean=260 , s=90

T score= ((260-270)/(90/sqrt(18)))= -0.471

**T score=-0.471**

**p-value = 32.2% (0.322)** probability is there that randomly selected bulbs would have an average life of no more than 260 days.

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom