|  |  |
| --- | --- |
| 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 | 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 | Nominal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| 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}  
 Occurrence of two heads and one tail = 3

Probability = 3/8 = 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**

**Ans:** Sample space = {(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)}

1. Sum equal to 1 = 0  
   Probability = 0/36 = 0
2. Sum less than or equal to 4 = 6

Probability = 6/36 = 0.16

1. Sum is divisible by 2 and 3 = 24

Probability = 24/36 = 0.66

**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:** Probability of not getting blue ball = 5/7 = 0.7

**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:** Sample size = 6

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

= 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**

**Ans:** Central tendencies for:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | Standard Deviation | Range |
| Points | 3.597 | 3.695 | 3.92 | 0.2858814 | 0.5346787 | 2.17 |
| Score | 3.217 | 3.325 | 3.44 | 0.957379 | 0.9784574 | 3.911 |
| Weigh | 17.85 | 17.71 | 17.02 | 3.193166 | 1.786943 | 8.4 |

* Median for Points is nearly equal to Mean, implying symmetrical distribution. Same goes for Score and Weigh.
* Deviation for the three quantities is low, indicating to the lower spread data. But the deviation of Weigh is comparatively higher implying to the dynamic data.

**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:** Probability of getting one number = 1/9

Expected value =

= (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 \* (1/9))

= 145.33333333333331

**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**

**Ans:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Speed | Distance | SP | WT |
| Skewness | -0.1175 | 0.8068 | 1.6114 | -0.6147 |
| Kurtosis | -0.5089 | 0.4050 | 2.9773 | 0.9502 |

* The car was at average speed most of the time as the mean, mode and median are same since the curve is normal distribution (skewness ~ 0)
* Distance curve is slightly leaned towards right side indicating lesser distance covered at higher values.
* The data count decreases as the SP value increases. (Positive skewed curve)
* Most of the weights are around average value (skewness ~ 0, kurtosis ~ +1)

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



**Ans:** For Histogram:

* Maximum data lies between the range of 0 – 100
* As the weight increases the count of chicks decreases.
* Most of the data is lower than the average chick weight.

For Boxplot:

* Median leaned towards lower quartile implies positive skewed curve
* There are outliers at the higher values

**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:** Data given,

n = 2000

N = 3000000

= 200

s2 = 30

Confidence Interval = Point estimate + Marginal Value

= Z(1-α) \* (σ / )

For 94% = -245.6432 to 246.6432

For 98% = -260.6124 to 261.6124

For 96% = -251.5205 to 252.5205

**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:**

1. mean: 41.0

median: 40.5

variance: 24.11111111111111

standard deviation: 4.910306620885412

1. According to the marks, the student has performed average in most the tests and there’s noticeable deviation in marks achieved.

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

**Ans:** Symmetrical curve, skewness = 0

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

**Ans:** Asymmetrical curve, skewness = positive

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

**Ans:** Asymmetrical curve, skewness = negative

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

**Ans:** Distribution is peaked and possess thick tails

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

**Ans:** Distribution has lighter tails than normal distribution.

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



**What can we say about the distribution of the data?**

**Ans:** Most of the extreme data are in lower range

**What is nature of skewness of the data?**

**Ans:** Negative skewed or Left skewed

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

**Ans:** 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 the boxplots have same median
* The boxplot 1 is slightly right skewed whereas the boxplot 2 has symmetric data distribution.

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

**Ans:** a. P(MPG>38) = 0.4074

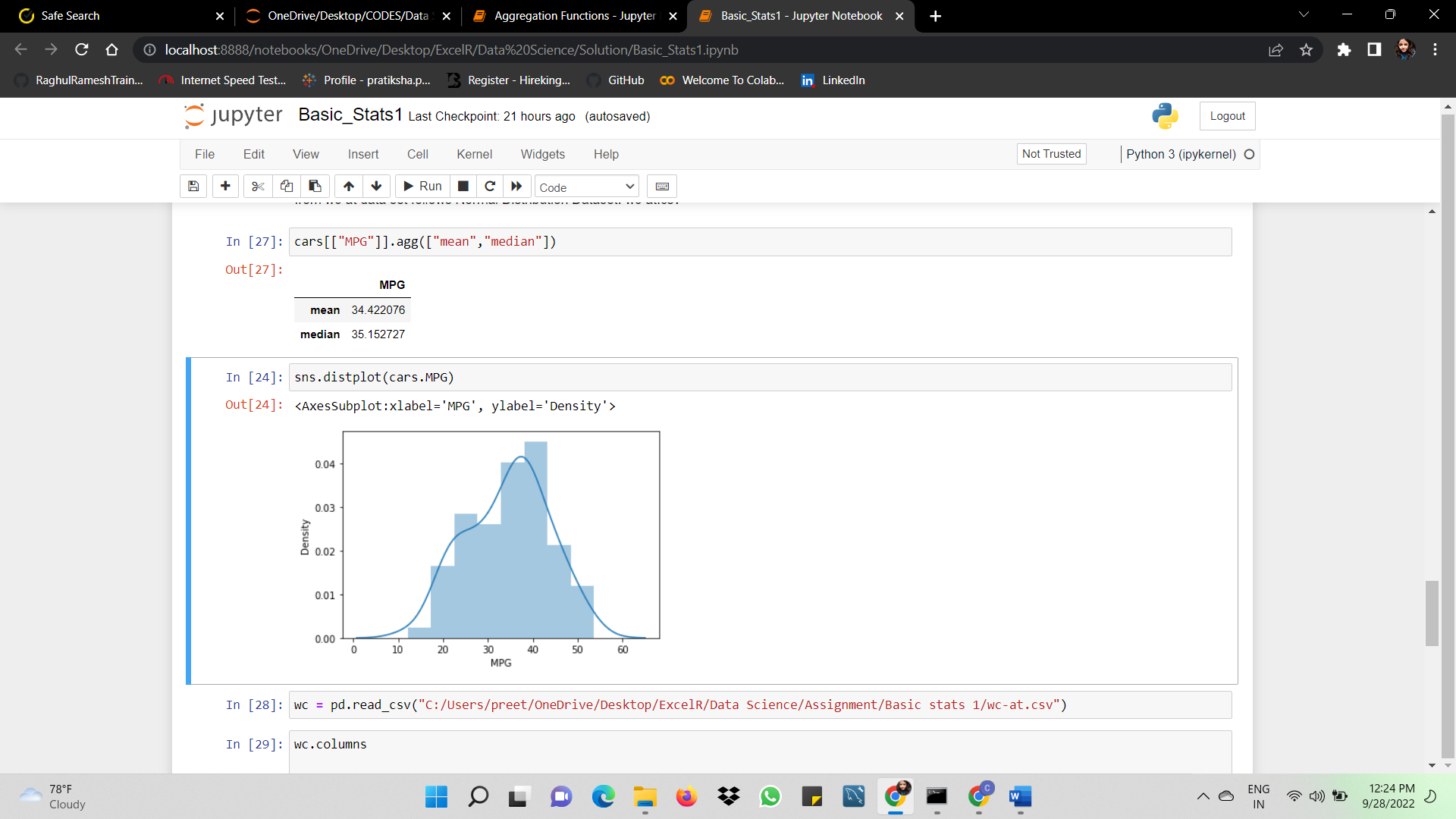
1. P(MPG<40) = 0.7530
2. P (20<MPG<50) = 0.8518

**Q 21) Check whether the data follows normal distribution**

1. **Check whether the MPG of Cars follows Normal Distribution**

**Dataset: Cars.csv**

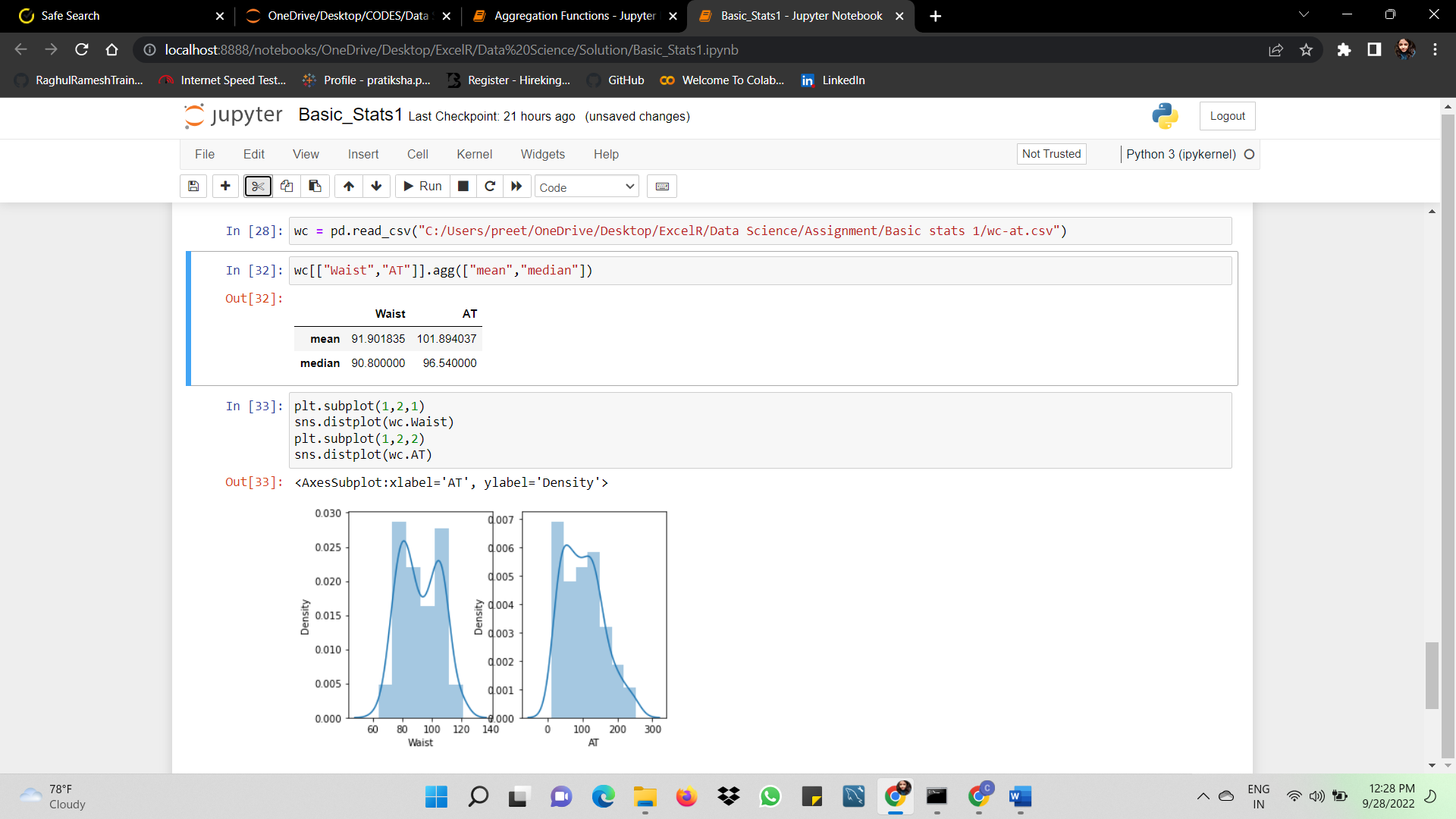
**Ans:** Since mean < median, the MPG of cars is slightly left skewed.



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

**Dataset: wc-at.csv**

**Ans:** In both the data sets, mean > median, hence both are slightly right skewed.

****

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

**Ans:**

|  |  |  |  |
| --- | --- | --- | --- |
| CI | 90% | 94% | 60% |
| Z-score | 1.2815 | 1.5547 | 0.2533 |

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

Ans:

|  |  |  |  |
| --- | --- | --- | --- |
| CI | 95% | 96% | 99% |
| t-score | 1.7081 | 1.8248 | 2.4851 |

**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:** Data given,

µ = 270

n = 18

= 260

s2 = 90

= -0.4714

p-value = 0.3215

