**Q1) Identify the Data type for the Following:**

|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete Data |
| Results of rolling a dice | Discrete Data |
| Weight of a person | Continuous Data |
| Weight of Gold | Continuous Data |
| Distance between two places | Continuous Data |
| Length of a leaf | Continuous Data |
| Dog's weight | Continuous Data |
| Blue Color | Discrete Data |
| Number of kids | Discrete Data |
| Number of tickets in Indian railways | Discrete Data |
| Number of times married | Discrete Data |
| Gender (Male or Female) | Discrete Data |

**Q2) Identify the Data types, which were among the following**

Nominal, Ordinal, Interval, Ratio.

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

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

**Soln. :**

S : Three Coins are Tossed.

S = {HHH, THH, HTH, HHT, HTT, THT, TTH, TTT}

n(S) = 8

A : Getting Two Heads and One Tail

A = {THH, HTH, HHT}

n(A) = 3

P(A) =

**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**

**Soln. :**

S : Two dice are rolled.

S = { (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(S) = 36

1. Let,

A : There is no outcomes which corresponds sum equal to 1 = { }

n(A) = 0

P(A) = 0

1. Let,

B : Sum is less than or equal to 4

B = { (1,1), (1,2), (1,3), (2,1), (2,2), (3,1) }

n(B) = 6

P(B) = 6/36 = 1/6

1. Let,

C : Sum is divisible by 2 or 3

C = { (1,1), (1,2), (1,3), (1,5), (2,1), (2,2), (2,4), (2,6), (3,1), (3,3), (3,5), (3,6), (4,2),

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

n(C) = 24

P(C) = 24/36 = 2/3

**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?**

**Soln. :**

Total Balls in Bag = 2Red + 3Green + 2Blue = 7

Two balls are drawn at random, n(Two balls drawn at random = 7C2 = 21

P(None of balls drawn is blue) = 5C2/7C2 = 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

**Soln.:**

Expected number of candies = Sum\_of ( X\* p(x) )

= Sum of (Candies count \* Probability )

= 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**

**Soln. :**

**Mean :**

Points 3.596563

Score 3.217250

Weigh 17.848750

dtype: float64

**Median:**

Points 3.695

Score 3.325

Weigh 17.710

dtype: float64

**Varience :**

Points 0.285881

Score 0.957379

Weigh 3.193166

dtype: float64

**Standard Deviation :**

Points 0.534679

Score 0.978457

Weigh 1.786943

dtype: float64

**Mode :**

Points 3.07 and 3.92

Score 3.44

Weigh 17.02 and 18.90

**Range :**

Points 2.17

Score 3.91

Weigh 8.39

**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?**

**Soln. :**

## Expected Value of a Discrete Random Variable is actually the mean of the statements

= Hence Mean is 1308/9

= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Soln. :**

**Skewness :**

Index 0.000000

speed -0.117510

dist 0.806895

dtype: float64

**Kurtosis :**

Index -1.200000

speed -0.508994

dist 0.405053

dtype: float64

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Soln. :**

**Skewness :**

SP 1.611450

WT -0.614753

dtype: float64

**Kurtosis :**

SP 2.977329

WT 0.950291

dtype: float64

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



**Soln. :**

Here we can see that the major Chick weights fall in the catogory of 50-100g(measures in x) as the maximum which is 200.The minimum weights have a frequency if less than or equal to 5.

The plot is Right sqewed which show that there is lesser concentration of chick weights in the 300-400gram category .

The expected value should be above 46.45



**Soln. :**

Median is less than mean right skewed and we have outlier on the upperside of box plot and there is less data points between Q1 and bottom point.

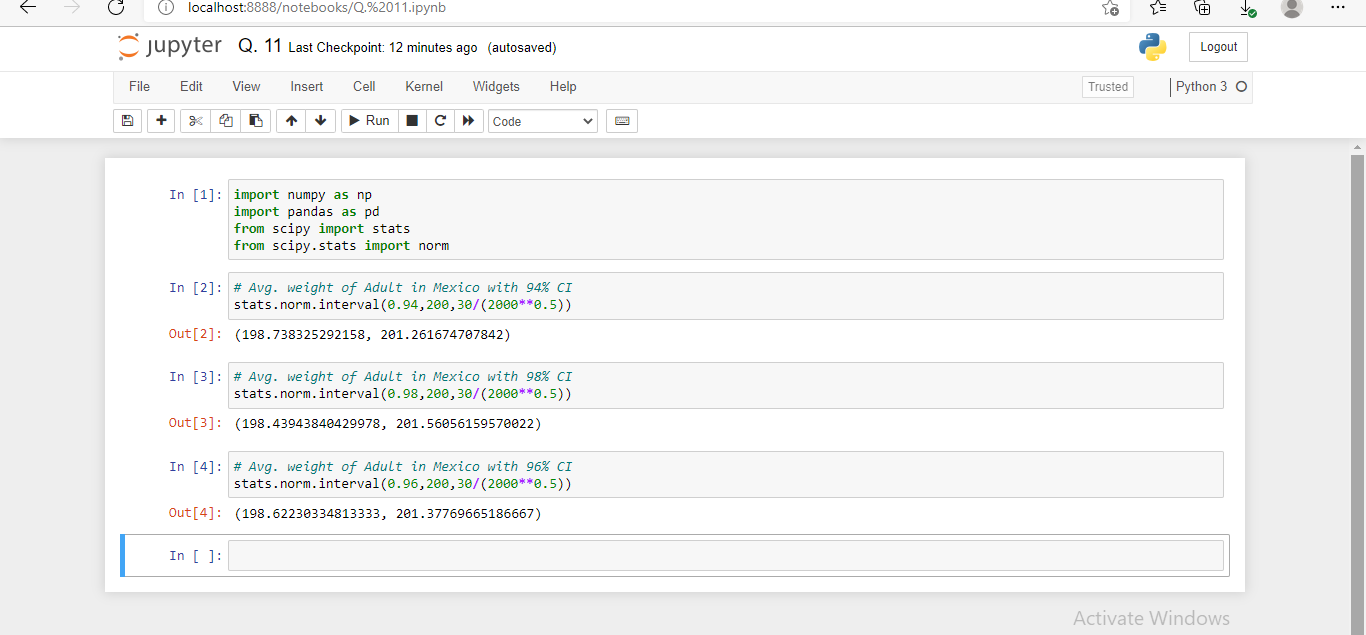
**Q11) Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?**

**Soln. :**

94% = [ 2004.8, -1995.16]

98% = [2006.39, -1993.6]

96% = [2005.46, -1994.54]



**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.**

**Soln. :**

Mean = 41

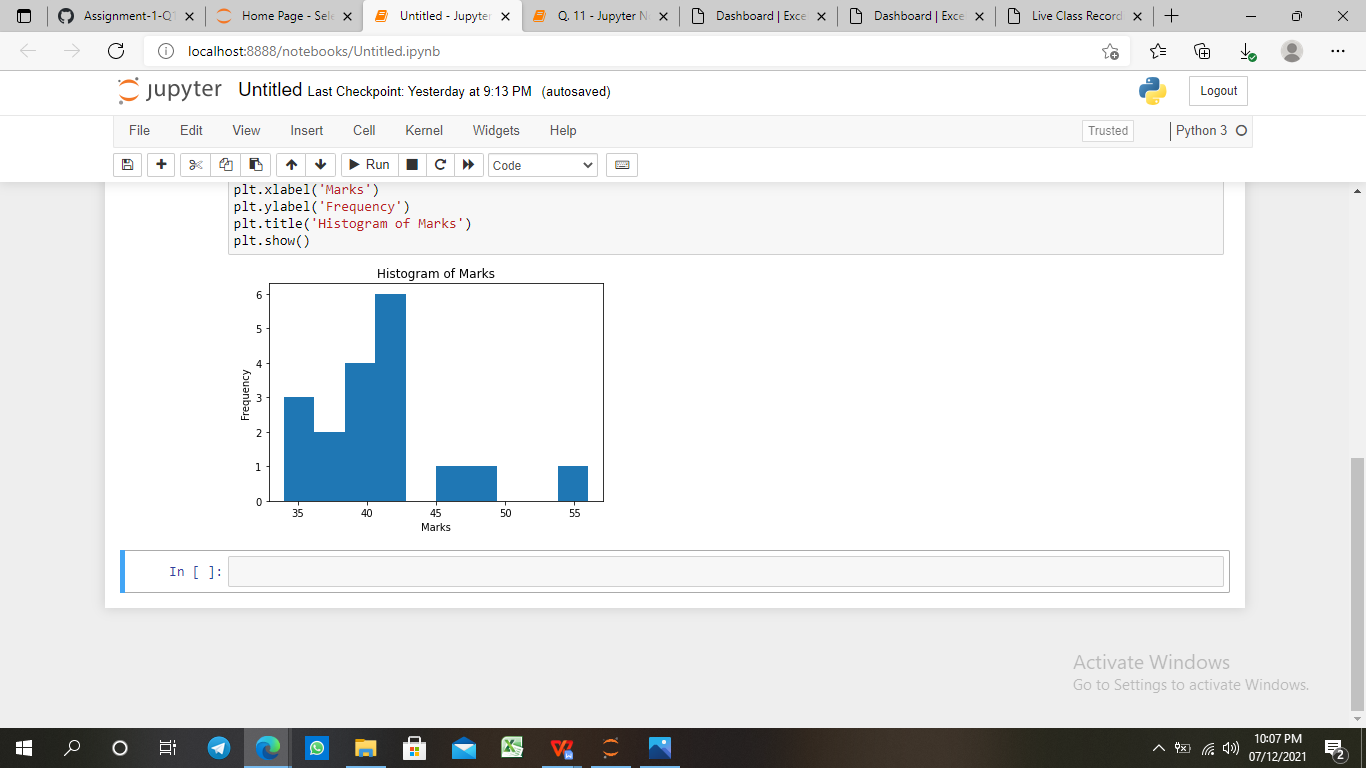
Median = 40.5

Variance = 25.5294

Standard Deviation = 5.05

1. **What can we say about the student marks?**

**Soln. :**



This class contains students that are actually mediocre. Most of the students in the class are having an average percentage of 65 and there are only a few students securing value above 90%.

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

**Soln. :**

The skewness will be symmetrical. Hence both the sides of the plot must be equal in proportion for the data should be normally distributed.

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

**Soln. :**

For data which produces mean > median the skewness will be a +ve skewness or the data will be right skewed. Most of the data will be lying on the left side of the plot. Mean always tends to go towards the most skewed part since skewness influences the mean.

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

**Soln. :**

For data which produces median > mean the skewness will be a -ve skewness or the data will be left skewed. Most of the data will be lying on the right side of the plot. Mean always tends to go towards the most skewed part since skewness influences the mean.

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

**Soln. :**

A distribution with a positive kurtosis value indicates that the distribution has heavier tails than the normal distribution. For example, data that follow a t distribution have a positive kurtosis value.

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

**Soln. :**

A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

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



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

**Soln. :**

The data is not actually equally distributed across the plane. There might be outliers influencing the data . Median of the data is 14.7(app x)

25 percent of the data lies between 0-10

50 percent of the data lies between 10-18

25 percent of the data lies after 18-20 approx.

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

**Soln. :**

The data will be left skewed since whisker length on the upper quadrant is higher than the data on the lower quadrant. Median will be greater than the mean since data is left skewed.

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

IQR is the inter quartile range.

Here Q1 = 10

Q2 = 14.7

Q3 = 18

IQR = Q3 – Q1 = 8(approx)

**Q19) Comment on the below Boxplot visualizations?**



**Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.**

**Soln. :**

Here when we compare box plot 1 with box plot 2 we can say that the data in box plot 1 is widely spread. Here the main inference is that since the data range varies high in box plot 2 it is hard to make a prediction in box plot 2. The median in the 2 box plots are equal. And the data spread in both of them are symmetrical

**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**

**Soln. :**

a. P(MPG>38) = 0.3475939251582705

b. P(MPG<40) = 0.7293498762151616

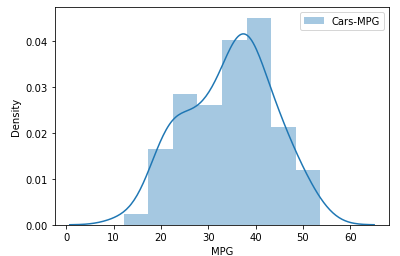
c. P (20<MPG<50) = 1.2430968797327613e-05

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

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

**Dataset: Cars.csv**

**Soln. :**

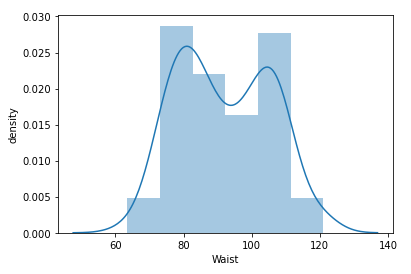


It’s not normal distribution, slightly left skewed

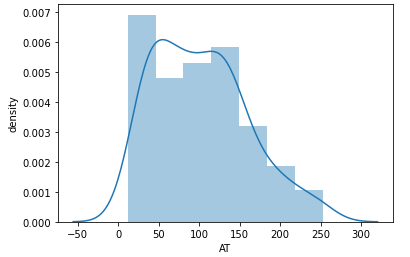
1. **Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution**

**Dataset: wc-at.csv**

**Soln. :**



It’s normal distribution



It’s not normal distribution, slightly right skewed

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

**Soln. :**

Z scores of 90% confidence interval = 1.6448536269514722

Z scores of 94% confidence interval = 1.8807936081512509

Z scores of 60% confidence interval = 0.8416212335729143

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

**Soln. :**

**t-SCORE CALCULATION**

T((1,alpha),(n-1))

Here n = 25

n-1 = 24

Hence t score values will be:

95% = qt(0.975,24) = 2.063899

96% = qt(0.98,24) = 2.171545

99% =  qt(0.995,24) = 2.79694

**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**

**Soln. :**

Sample size = 18 = n

Sample mean = 260 days = x

Sample standard deviation = s = 90days

t-score = -0.4714

P(X>= 260) = 0.32167