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

Q1) Identify the Data type for the Following:

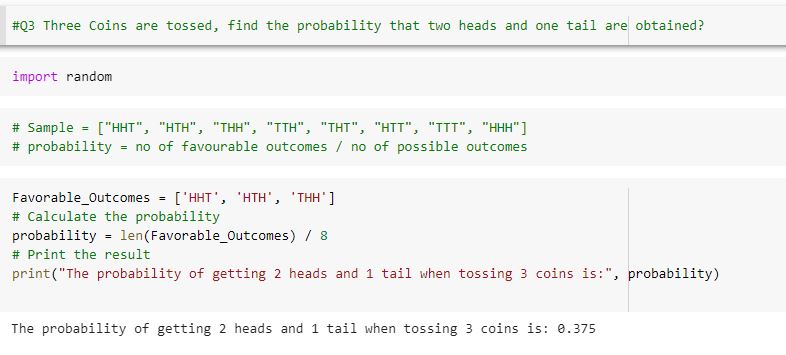
Q2) Identify the Data types, which were among the following

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

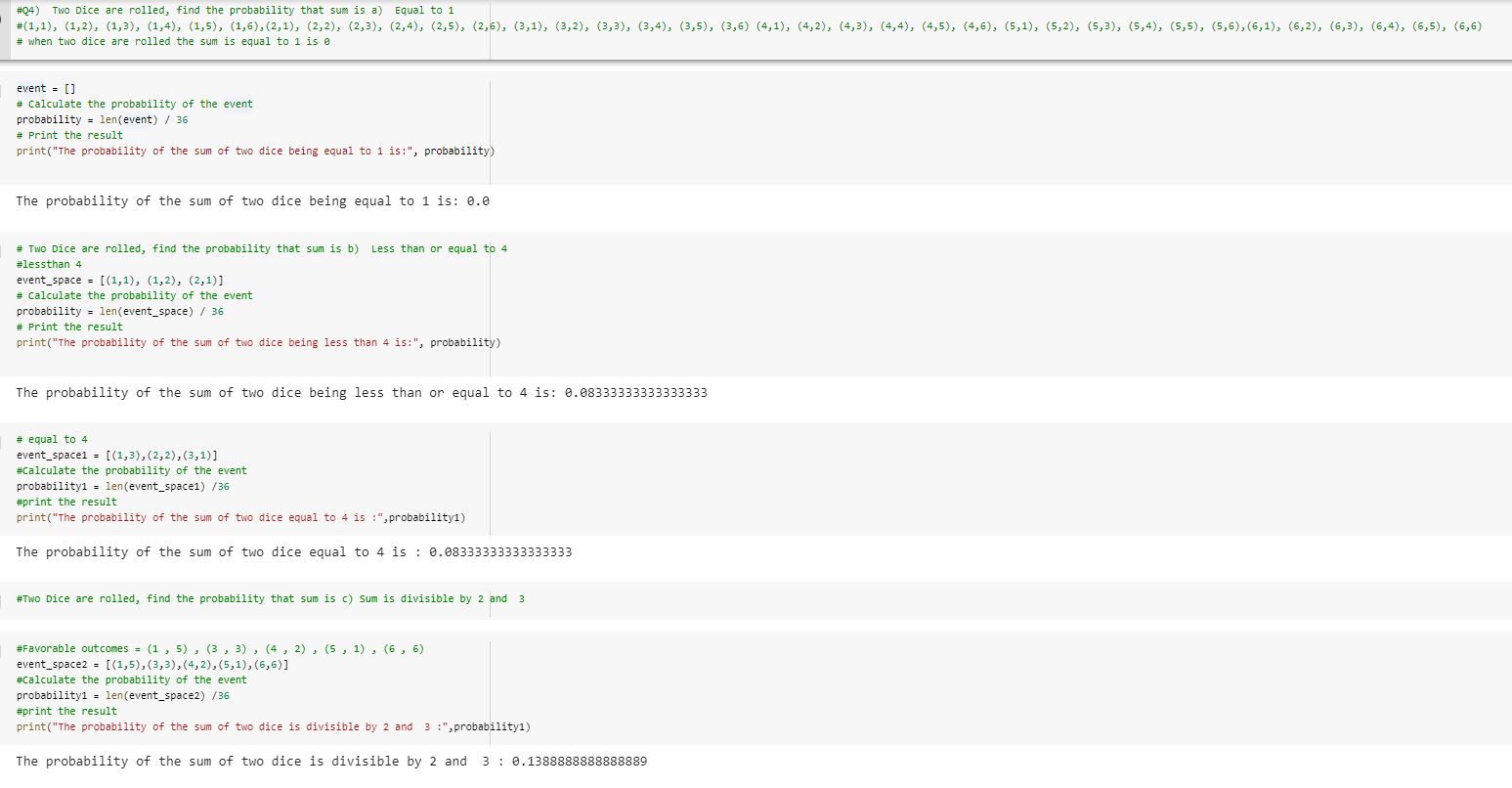
Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

Solution:

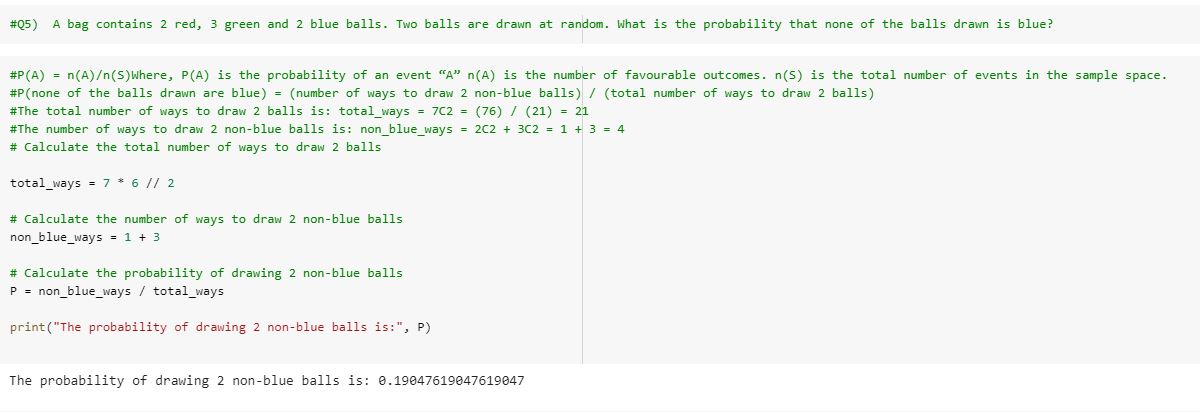


Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3



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?

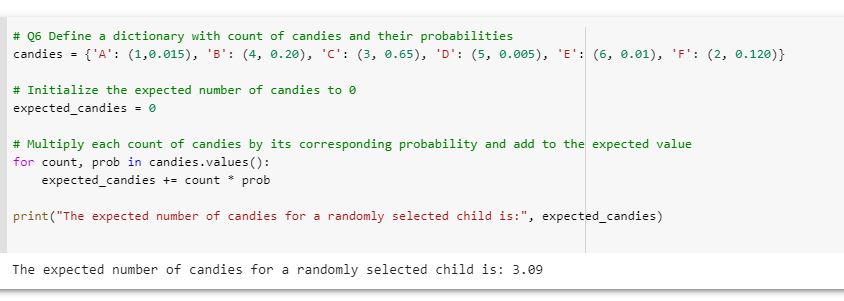


Q6) Calculate the Expected number of candies for a randomly selected child

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

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**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.596 | 3.217 | 17.848 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.07 | 3.0 | 17.02 |
| Variance | 0.276 | 0.927 | 3.093 |
| SD | 0.534 | 0.978 | 1.786 |
| Range | 2.17 | 3.911 | 8.39 |

**Comments:**

1) Differences between measures: There are differences between the mean, median, and mode, indicating that the distribution of the data may not be symmetrical. In particular, the mode is lower than both the mean and median,

suggesting that the data may be negatively skewed.

2) Variation in scores: There is a relatively large range in scores,

with a difference of 3.911 between the highest and lowest scores.

This indicates that there is a significant amount of variation in the scores.

3) Outliers: The range in scores is relatively large, which could suggest the presence of outliers in the data that are affecting the mean and/or median.

This could be investigated further using a box plot or other methods.

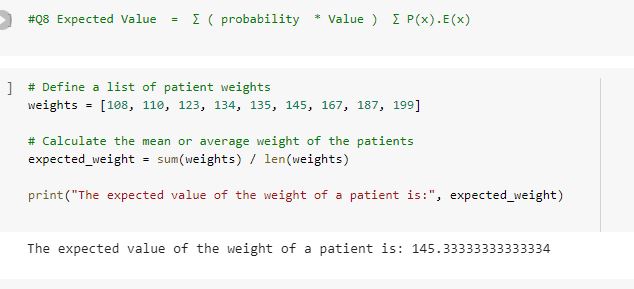
Attached the Assignment\_1\_7Q.ipynb

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?



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

Cars speed and distance

Use Q9\_a.csv

speed distance

Kurtosis - 0.577 0.243

Skewness -0.11 0.78

**Comment:**

**Speed:** A kurtosis of -0.577 indicates a relatively flat distribution with fewer extreme values than a normal distribution. A skewness of -0.11 indicates a slightly negative (left) skew towards lower values, but this skewness value is relatively close to 0, indicating a somewhat symmetrical distribution.

**Distance:** A skewness of 0.78 indicates a positive (right) skew towards higher values, meaning that the distribution is more spread out towards higher values than lower values. A kurtosis of 0.243 indicates a distribution that is slightly more peaked than a normal distribution, with more values in the tails than in the center.

SP and Weight(WT)

Use Q9\_b.csv

speed Weight

Kurtosis 2.73 0.819

Skewness 1.581 -0.603

**Comment:**

**Speed:** A kurtosis of 2.73 indicates a distribution that is much more peaked than a normal distribution, with a high concentration of values in the center and heavy tails. This suggests that there are more extreme values in the tails of the distribution than in a normal distribution.

A skewness of 1.581 indicates a strong positive (right) skew towards higher values, meaning that the distribution is more spread out towards higher values than lower values.

**Weight:** A kurtosis of 0.819 indicates a distribution that is slightly more peaked than a normal distribution, with more values in the tails than in the center. This suggests that the data has more extreme values than a normal distribution.

A skewness of -0.603 indicates a negative (left) skew towards lower values, meaning that the distribution is more spread out towards lower values than higher values.

Attached the Assignment9Q(1).ipynb

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



**Comment:** The Histogram represents the data skewed to the right side , positively skewed have a longer right tail. Here Mean is greater than median



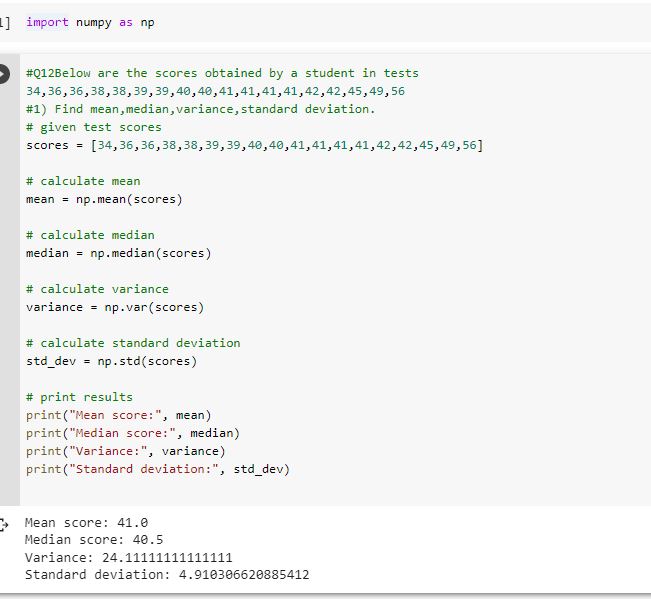
**Comment:** Boxplot explains the data is skewed right side.

**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?

****

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?

**Comment:**

Based on the results, we can see that the student scored an average of 41 marks, with a minimum score of 34 and a maximum score of 56. The distribution of scores is somewhat spread out, as indicated by the relatively large standard deviation of 6.4. The variance of 40.94 indicates that there is quite a bit of variability in the scores, with some scores being much higher or lower than the mean. However, since the mean and median are equal, this suggests that the distribution of scores is roughly symmetric, with no significant outliers. Overall, the student's marks are fairly good, but there is room for improvement.

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

Solution: If the distribution is symmetric, then the mean is equal to the median, and the distribution has zero skewness.

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

Solution: The mean of positively skewed data will be greater than the median. Right skewed.

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

Solution: In a negatively skewed distribution, the exact opposite is the case: the mean of negatively skewed data will be less than the median. left skewed.

Q16) What does positive kurtosis value indicates for adata ?

Solution: A distribution with a positive kurtosis value indicates that the distribution has heavier tails than the normal distribution.

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

Solution: 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?

Solution: Not normally distributon

What is nature of skewness of the data?

Solution: Negative skewness

What will be the IQR of the data (approximately)?   
Solution: IQR = Q3-Q1

approximately 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.

Comment: Both distributions are approximately symmetric, because both half-boxes are almost the same length on the left side and on the right side.

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

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

Solution : Attached Q20.ipynb file

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

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Comment :**

1) About 68% of the data falls within one standard deviation of the mean (i.e., between mean - std and mean + std).

2) About 95% of the data falls within two standard deviations of the mean (i.e., between mean - 2std and mean + 2std).

3) About 99.7% of the data falls within three standard deviations of the mean (i.e., between mean - 3std and mean + 3std).

If the dataset follows this rule reasonably well, it is likely to be normal or approximately normal.

Attached Assignment21Q.ipynd file

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

Dataset: wc-at.csv

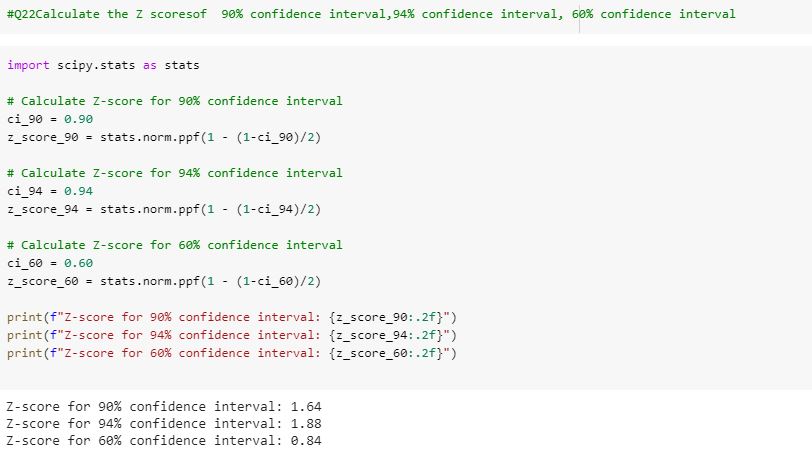
**Comment:**

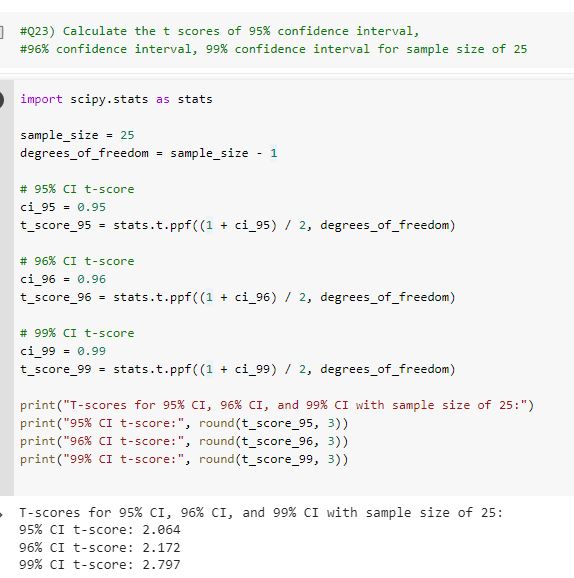
**Waist:** It is not possible to determine if a distribution is normal solely based on the mean and standard deviation. However, if the data is approximately symmetric and unimodal, then it may suggest that the data follows a normal distribution.

**Adipose Tissue(AT):** Based on these guidelines, a standard deviation of 57.29 is quite large compared to the mean of 101.894037, which suggests that the data is more spread out than it would be in a typical normal distribution.

Attached the Assignment21Q.pynb file

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



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

Q 24**)**A Government companyclaims 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

