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

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

**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 | Ordinal |
| 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 | Interval |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

Ans- Probability=3/8 or 37.5%

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

1. Equal to 1 Ans-Zero
2. Less than or equal to 4 Ans-1/6
3. Sum is divisible by 2 and 3 Ans-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- 7C2  =21

5C2  =10

Probability= 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 | X\*P(x) |
| A | 1 | 0.015 | 0.015 |
| B | 4 | 0.2 | 0.8 |
| C | 3 | 0.65 | 1.95 |
| D | 5 | 0.005 | 0.025 |
| E | 6 | 0.01 | 0.06 |
| F | 2 | 0.12 | 0.24 |
|  |  |  | Total=3.09 |

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**

**Use Q7.csv file**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variance | 0.276948 | 0.927461 | 3.09338 |
| Standard Deviation | 0.52625 | 0.96304 | 1.7588 |
| Range | 2.17 | 3.911 | 8.4 |

This data is not a normally distributed data as mean,median & mode are varying.

Standard deviation & Variance are high for score and weigh.

**Q8) Calculate Expected Value for the problem below**

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

**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**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Car speed | Distance | SP | Weight |
| Skewness | -0.11751 | 0.806895 | 1.61145 | -0.61475333 |
| Kurtosis | -0.50899 | 0.405053 | 2.97733 | 0.950291 |

Skewness and Kurtosis value for car speed are negative so the data has lower tails or less outliers

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



Histogram

* The data is skewed on the right side.
* There are no outliers in the data.
* The mean lies somewhere between 50-100.



Boxplot

* Data is distributed on the right and is positively skewed.
* There are outliers in the data.

**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?

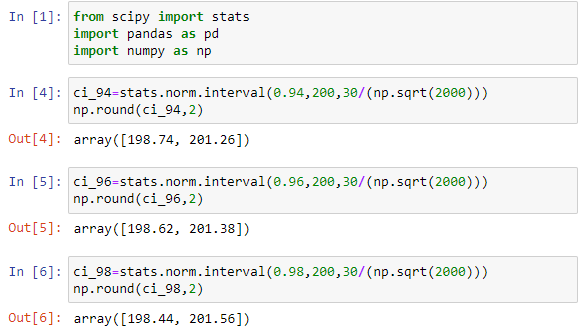
Ans-

PopulationI(N)=3,000,000

Sample(n)=2000

Average weight=200 pounds

Standard deviation=30 pounds



**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

Median=40.5

Variance=23.722

Standard Deviation=4.87

2) Mean & median are not equal so this is not a normally distributed data, the difference is 0.5 though.

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

Ans- Zero Skewness

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

Ans-Positive Skewness

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

Ans-Negative Skewness

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

Ans- Data has a heavier tails than normal distribution.

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

Ans-Data has lighter tails & wide spread than a normal distribution.

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



a) What can we say about the distribution of the data?

Ans-Distribution of the data is not symmetrical.

b) What is nature of skewness of the data?

Ans-Data is negatively skewed.

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

Ans-IQR will be 8 (approximately).  
**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 data are normally distributed.
* Both have no outliers present.
* Variance of boxplot(1) is less than boxplot(2).
* Outliers are absent in both data.

**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- P(MPG>38)= 33/81

P(MPG<40)=67/81

P (20<MPG<50)= 69/81

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

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

* As mean and median are equal,the given data is normally distributed.

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

Dataset: wc-at.csv

* As mean and median are not equal,the data doesn’t follow normal distribution.

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

|  |  |
| --- | --- |
| Confidence Interval | Z scores |
| 60% | 0.84 |
| 90% | 1.645 |
| 94% | 1.88 |

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

|  |  |
| --- | --- |
| Confidence Interval | T scores |
| 95% | 1.71 |
| 96% | 1.82 |
| 99% | 2.49 |

**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- Probability will be 32.16%

