**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 Color | Discrete |
| 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 | 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 | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

When 3 coins are tossed, sample space = [(HHH), (THH), (HTH), (HHT), (HTT), (THT), (TTH), (TTT)]

Total no. of outcomes = 8

No. of times two heads & a tail is obtained = 3

Hence, probability = Favorable outcomes/Total outcomes

= 3/8 = 37.5%

**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**

No. of outcomes when 2 dice are rolled = 62 = 36

a) Out of total outcomes sum is equal to 1 = 0 times

Hence, probability = Favorable outcomes/Total outcomes

= 0/36 =0

b) Sum 4 = 6 times [(1,1),(1,2),(2,1),(2,2),(1,3),(3,1)]

Hence, Probability = Favorable outcomes/Total outcomes

= 6/36 = 1/6 = 16.67%

c) For sum to be divisible by 6, sum = multiples of 6 i.e. 6, 12

No. of Favorable outcomes = 6[(2,4),(4,2),(3,3),(1,5),(5,1),(6,6)]

Hence, Probability = Favorable outcomes/Total outcomes

= 6/36 = 1/6 = 16.67%

**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?**

Balls in the bag = [R, R, G, G, G, B, B]

Total outcomes = 7C2 = 7! /2!\*5! = 21

[(R,R),(R,G),(R,G),(R,G),(R,B),(R,B),(R,G),(R,G),(R,G),(R,B),(R,B), (G,G),(G,G),(G,B),(G,B),(G,G),(G,B),(G,B),(G,B),(G,B),(B,B)]

Favorable outcomes = 5C2 = 10

[(R,R),(R,G),(R,G),(R,G), (R,G),(R,G),(R,G), (G,G),(G,G), (G,G)]

Probability = Favorable outcomes / Total Outcomes = 10/21

= 0.4762 = 47.62%

**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**

Expected Value =

Expected no. of candies for randomly selected child

= 1\*0.015 + 4\*0.2 + 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)**

Solved in Python

**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?**

Expected Value =

Since, all the values of weight of patient are different; value of P(x) is same for everyone i.e. 1/9. Therefore,

Expected Value = mean =

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

= 1308/9 = 145.33 pounds.

**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**

Solved in python

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



Inferences on the Histogram: -

1. Mode of the data is between 50 & 100 as frequency of the bar is highest(~200)
2. The data is positive or right skewed.
3. The data has outliers on the upper extreme side.

Inferences on the boxplot:-

1) In the given boxplot, we can see a number of dots above upper extreme which are outliers.

2) The length of upper whisker is longer than lower whisker which means data is right skewed.



**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?**

Solved in Python

**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?**

Solved in Python

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

When mean and median is equal, skewness is equal to zero i.e. data is not skewed or nearly normally distributed.

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

When mean>median, the data is right or positively skewed.

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

When median>mean, the data is left or negatively skewed.

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

Positive kurtosis indicates that most of the data-points are near the mean value and away from the extremes.

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

Negative kurtosis indicates that most of the data points are away from mean and close to either extreme.

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



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

Value of Q1 is around 10 and that Q3 is around 18. This means Interquartile range is 18 - 10 = 8. Median of the data is 15 which is on the right side of graph. Left whisker of boxplot is longer than right whisker which denotes that the given data is left or negatively skewed.

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

Left whisker of boxplot is longer than right whisker which denotes that the given data is left or negatively skewed.

**What will be the IQR of the data (approximately)?**IQR **=** Q3 **-** Q1 = 18 – 10 = 8

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



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

Answer: - Inferences from above distribution -

* Both the box plots are normally distributed on different range.
* The range of the first graph is lesser than that of second boxplot.
* Mean of both the boxplots is about 260-265.
* We can also infer that standard deviation of 1st data would be lesser than the 2nd.

**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)**

Solved in Python

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

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

**Dataset: Cars.csv**

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

**Dataset: wc-at.csv**

Solved in Python

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

Solved in Python

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

Solved in Python

**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**

Solved in Python