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

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

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

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

**Ans:** 0.375 or (3/8)

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

1. Equal to 1

**Ans:** 0

1. Less than or equal to 4

**Ans:** (5/36) or 0.1389

1. Sum is divisible by 2and 3

**Ans:** (1/6) or 0.167

**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:** (20/42)=(10/21) or 0.476

**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 A – probability of having 1 candy = 0.015. Child B – probability of having 4 candies = 0.20**

**Ans:**

|  |  |  |  |
| --- | --- | --- | --- |
| CHILD | Candies count | Probability | Expected number of candies |
| A | 1 | 0.015 | 7 |
| B | 4 | 0.20 | 20 |
| C | 3 | 0.65 | 5 |
| D | 5 | 0.005 | 1000 |
| E | 6 | 0.01 | 600 |
| F | 2 | 0.120 | 17 |

**Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset**

* **For Points,Score,Weigh>**

**Ans:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.5965 | 3.2172 | 17.8487 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 18.90 |
| Variance | 0.2769 | 0.9274 | 3.0933 |
| Std. Deviation | 0.5262 | 0.9630 | 1.7588 |
| Range | 2.17 | 3.9109 | 8.3999 |

**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:**

there are 9 patients

Probability of selecting each patient = 1/9

Expected Value  =  ∑ ( probability  \* Value )

 ∑ P(x).E(x)

Ex  108, 110, 123, 134, 135, 145, 167, 187, 199

P(x)  1/9  1/9   1/9  1/9   1/9   1/9   1/9   1/9  1/9

Expected Value  =  (1/9)(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) ( 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9)  (  1308)

= 145.33

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

**1] speed and distance**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| Speed | -0.11750 | -0.5089 |
| Distance | 0.8068 | 0.4050 |

Variable Speed has negative skewness as well as negative kurtosis which means the data is not normally distributed.

Variable Distance has positive skewness as well as positive kurtosis which means the data is normally distributed.

**2] SP and Weight(WT)**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| SP | 1.6114 | 2.9773 |
| Weight | -0.6147 | 0.9502 |

Variable SP has positive skewness as well as positive kurtosis which means the data is not normally distributed.

Variable WT has negative skewness as well as positive kurtosis

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

**Ans:**

1]HISTOGRAM:

Following given histogram represents the frequency distribution of the weight in the chickweight dataset.

The given histogram is a left skewed histogram which can be considered as the positively skewed.

According to the histogram the major part of the data lies between the range of 50 to 200.

While the tail of the histogram is to the right side which represents remaining part of the data.



2]BOXPLOT:

Given boxplot shows that the IQR of the data set fall near the lower extreme of the plot.

As the IQR falls closer to lower extreme of the whisker the median of the data set lies in the lower values of data set.

Also the dataset has several outliers which are occurring after upper extreme of the boxplot. Upper whisker is comparatively much longer than lower whisker.



**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:**

1] average weight of an adult at CI 94%: 200+/- 1.26

2] average weight of an adult at CI 98%: 200+/- 1.56

3] average weight of an adult at CI 96%: 200+/- 1.37

**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.**

**Ans:** 1] mean= 41

2] median= 40.5

3] variance= 24.11

4] standard deviation= 4.9103

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

**Ans:** The student has the marks ranging between 40.5 to 41 which has variance of 24.11 marks along with s.d. of 4.103.

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

**Ans:** Then the skewness is called as zero skewness.

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

**Ans:** Then the skewness is called as positively skewed.

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

**Ans:** Then the skewness is called as negatively skewed.

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

**Ans:** Positivevalues of kurtosis indicate that a distribution is peaked and possess thick tails.

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

**Ans:** 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?**

**Ans:** The Inner quartile range (IQR) comes almost in the upper part of the data set. Upper whisker is marginally very less than lower whisker

What is nature of skewness of the data?

**Ans:** The nature of the skewness is positively skewed.

What will be the IQR of the data (approximately)?   
**Ans:** The IQR of the data will be between 10 to 18.

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



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

**Ans:** According to given visualization Boxplot 1 has comparatively much smaller range than the Boxplot 2. However the medians of both the Boxplots are almost identical. Both the Boxplot looks like the data is normally distributed.

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

**Calculate the probability of MPG ofCars for the below cases.**

**Ans: Probability:**

* 1. P(MPG>38)= 33/81= 0.474
  2. P(MPG<40)= 67/81= 0.8271

c. P (20<MPG<50)= 69/81= 0.8518

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

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

**Ans:** With the help of various normality distribution verification technique it is found that “MPG” Data of Cars does not follow normal distribution.

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

**Ans:** 1]With the help of various normality distribution verification technique it is found that “Waist” Data of wc-at follows normal distribution.

2]With the help of various normality distribution verification technique it is found that “AT” Data of wc-at does not follow normal distribution.

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

**Ans:** 1] Z scores of 90% confidence interval= 1.644854

2] Z scores of 94% confidence interval= 1.880794

3] Z scores of 60% confidence interval= 0.8416212

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

**Ans:** 1] t scores of 95% confidence interval= 2.059539

2] t scores of 96% confidence interval= 2.166587

3] t scores of 99% confidence interval= 2.787436

**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**

**Ans:** x = mean of the sample of bulbs =  260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

t - statistics for the data is given as follows:

t= == -0.471

For probability calculations, the number of degrees of freedom is n - 1,

n-1=18-1=17

so here t-distribution is needed with 17 degrees of freedom.

The probability that t < - 0.471 with 17 degrees of freedom assuming the population mean is true, the t-value is less than the t-value obtained With 17 degrees of freedom and a t score of - 0.471, the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300 days.