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

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

**Ans: P(H)=0.5 and P(T)=0.5**

**Since 3 coins tossed (0.5^3) with probability as 2H and 1T =3**

**Hence 3\*(0.5^3) = 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

**Ans:**

1. **0**
2. **1/6 or 0.17**
3. **5/36 or 0.14**

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: 10 / 21 or 0.48**

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

**Ans: 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**

**Ans:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Mean** | **3.59656** | **3.217** | **17.849** |
| **Mode** | **3.92** | **3.44** | **17.02** |
| **Variance** | **0.28588** | **0.957** | **3.1932** |
| **Std Deviation** | **0.53468** | **0.978** | **1.7869** |
| **Range** | **2.17** | **3.911** | **8.4** |

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?

**Ans: Expected Value = Probability \* Summation (Value) …. ∑P(x)\* E(x)**

**Probability = 1/9**

**E(x)= 108 + 110 + 123+ 134+ 135+ 145+ 167+ 187+199 = 1308**

**Expected value = (1/9) \* (1308) = 145.33**

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

Cars speed and distance

Use Q9\_a.csv

**Ans: Distribution is flatter for speed as the kurtosis is negative and for Distance the distribution is peaked  
 Skewness: Speed’s mean and median are less than its mode and the data is tapered to left side, similarly for distance mean and median is higher than mode and the data is tapered toward right side.**

**Kurtosis - speed -0.508994**

**dist 0.405053**

**Skewness - speed -0.117510**

**dist 0.806895**

SP and Weight(WT)

Use Q9\_b.csv

**Ans: Kurtosis - SP 2.977329**

**WT 0.950291**

**Skewness - SP 1.611450**

**WT -0.614753**

Q10) Draw inferences about the following boxplot & histogram

**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?

Data points (n) = 2000  
sample mean (x) = 200  
Std deviation (sigma) = 30

**94% = (198.738325292158, 201.261674707842)  
98% = (198.43943840429978, 201.56056159570022)  
96% = (198.62230334813333, 201.37769665186667)**

**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:

|  |  |
| --- | --- |
| **Mean** | **41** |
| **Median** | **40.5** |
| **Variation** | **25.5294** |
| **Std Deviation** | **5.05266** |
| **Mode** | **41** |

Q13) What is the nature of skewness when mean, median of data are equal? - 0 skewness

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

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

Q16) What does positive kurtosis value indicates for a data ? – **Distribution is peaked**

Q17) What does negative kurtosis value indicates for a data? – **Distribution is flattered**

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



What can we say about the distribution of the data?

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

What will be the IQR of the data (approximately)? **8**

Q19) Comment on the below Boxplot visualizations?



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

* **The median of both the plots is same**
* **The range of data in plot 2 is widely spread as compared to plot 1**
* **The data in plot 1 is very close to median**
* **The data in plot 1 is slightly left skewed while plot 2 data looks symmetric**

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) = 41%**
  2. **P(MPG<40) = 75%**

**c. P (20<MPG<50) = 85%**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**No, MPG does not follow normal distribution**

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

Dataset: wc-at.csv

**No, AT and WC does not follow normal distribution**

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

90% = 1.645

94% = 1.880

60% = 0.253

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

**95% = 2.064**

**96% = 2.064**

**99% = 2.797**

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**

**t = x-u / (s / N^^2)**

**x = 260**

**u = 270**

**s = 90**

**N = 18**

**t = 260 – 270 / (90 / 18^^2)**

**t = -/+0.471**