|  |  |
| --- | --- |
| 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** |

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) | **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 | **Ordinal** |
| SAT Scores | **Interval** |
| Years of Education | **Ratio** |

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

**Soln= (H,T,H)+(T,H,H)+(H,H,T)**

**1/8 + 1/8 + 1/8 = 3/8**

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

**Soln= a) There is no outcomes which corresponds sum is equal to one. That is 0/36. Probability is 0.  
b) (1,3)(2,2)(3,1) = 3 outcomes, 3/36 =1/12**

**c) 6/36= 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?

**Soln=** **10/21 = 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 | 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

**Soln = 1\*0.015 + 4\*0.20 + 3\*0.65 +5\*0.005 + 6\*0.01 + 2\*0.120**

**=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**

**Soln = 1) Points**

**Mean = 3.5965**

**Median = 3.695**

**Mode = 3.92**

**Variance = 0.2858**

**Standard Deviation = 0.5346**

**Range = 2.17**

**2) Score**

**Mean = 3.21**

**Median = 3.325**

**Mode = 3.44**

**Variance = 0.9573**

**Standard Deviation = 0.9784**

**Range = 3.9110**

**3) Weigh**

**Mean = 17.8487**

**Median = 17.710**

**Mode = 17.02**

**Variance = 3.1931**

**Standard Deviation = 1.7869**

**Range = 8.39**

* **No case has the variable Mean = Median = Mode**
* **Thus as seen in the graph dataset “score” and “Weigh” has outliers**

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?

**Soln = 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Soln = 1) Skewness**

**Speed = -0.1139**

**Distance = 0.7824**

**2) Kurtosis**

**Speed = 2.4228**

**Distance = 3.2480**

* **Distance is positively skewed where as Speed is negatively skewed**
* **Thus distance has distribution of data concentrated on the left whereas speed has distribution on the right. As seen in the graph**
* **Both distance and speed has positive Kurtosis**

**SP and Weight (WT)**

**Use Q9\_b.csv**

**Soln = 1) Skewness**

**SP = 1.5814**

**WT = -0.6033**

**2) Kurtosis**

**SP = 5.7235**

**WT = 3.8194**

* **SP is positively skewed where as WT is negatively skewed**
* **Thus SP has distribution of data concentrated on the left whereas WT has distribution on the right. As seen in the graph**
* **Both WT and SP has positive Kurtosis**

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



**Soln = The most of the data are concerted in the range of 50-100 with frequency 200, And the least range of weight is 400 around the 0 to 10. So the expected value the above distribution is 75.**

**Skewness = we can notice that a long tail towards right so it is heavily right skewed.**



**Soln = Median is less than mean right skewed and we have outlier on the upper side of the boxplot.**

**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?

**Soln = X+-(Z1-α. σ/sqrt(n)**

**Degrees of freedom =(n-1)= 2000-1= 1999**

**Confidence interval = 94%**

**(1- σ/2) = 1-0.03) = 0.97**

**For the confidence interval for 94% is 1.882**

**For the confidence interval for 98% is 2.33**

**For the confidence interval for 96% is 2.05**

**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?

**Soln = Mean = ΣX/N = 738/18 = 41**

**Median = (n+1)th/2 = 18+1/2 = 19/2 = 9.5 value = 40**

**Variance = (∑ (x − x̅)^2/ n – 1) = 434/17 = 25.52**

**Standard deviation = sqrt(variance) = sqrt(25.52) = 5.0517**

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

**Soln** = **Perfectly symmetrical and skewness is zero.**

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

**Soln = Right skewed distribution.**

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

**Soln = Left skewed distribution.**

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

**Soln = Sharper peak greater than 3.**

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

**Soln = Flatter peak.**

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



What can we say about the distribution of the data?

**Soln = The Data is not symmetric. Data is more concentrated towards right side.**

What is nature of skewness of the data?

**Soln = Skewness = Negative**

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

**Soln = IQR data is 8(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.

**Soln = a) Data is Normally Distributed. No outliers. Center around 262.5, comparatively, first graph has less range**

**b) Data is Normally Distributed. No outliers. Center around 262.5 comparatively, second graph has more range.**

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)

c. P (20<MPG<50)

**Soln = a. P(MPG>38) = 0.3475**

**b. P(MPG<40) = 0.7293**

**c. P(20<MPG<50) = 1.2430**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Soln = MPG is 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

**Soln = Both AT and Waist doesn’t follow Normal Distribution.**

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

**Soln = 90% CI = +/-1.645**

**= 94% CI = +/-1.881**

**= 60% CI =+/- 0.8416**

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

**Soln = 95% = +/-2.0638**

**96% =+/- 2.1715**

**99% =+/- 2.7969**

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

**Soln = t-score = -0.4714,  
 Degree of freedom = 17   
 P(t) = 0.3216725**