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

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained? *3/8*

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

1. Equal to 1 -> *0*
2. Less than or equal to 4 -> *1/6*
3. Sum is divisible by 2 and 3 -> *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? *10/21*

Q6) Calculate the Expected number of candies for a randomly selected child -> *3.09*

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

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.

*Q7.csv file attached*

**Use Q7.csv file**

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? *145.33*

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

**Cars speed and distance**

**Use Q9\_a.csv**

*Refer attached Q9\_a.csv*

**SP and Weight(WT)**

**Use Q9\_b.csv**

*Refer attached Q9\_b.csv*

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



*The weight of chick between 50 to 100 is maximum with frequency 200 followed by the range 100 to 150 subsequently with frequency 125.*



*There are 7 Outlier points which lies outside the box upper quartile value.*

**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?

*Given: Mean (x bar)=200 pounds*

*Population: 3000000 men*

*Sample: 2000 men*

*Std Dev(sigma): 30 pounds*

*To calculate: X bar+- Margin of error*

*Interval estimate formula: x bar+- t\*(sigma/root sample)*

*Calculating the standard error (Sigma x bar)= std dev/sq root sample =30/ sq root 2000 =30/14.14= 0.67*

1. *Calculation value of alpha at 94% C.I. =1-94/100=0.06*

*Cumulative probability=0.06/2=0.03*

*Degree of freedom: n-1= 2000-1=1999*

*Calculate t value at c.p 0.03 and 1999=* **1.881861**

*X bar+- t\*(sigma/root sample) = 200+-1.88\*0.67=200+-1.25*

1. *Calculation value of alpha at 98% C.I. =1-98/100=0.02*

*Cumulative probability=0.02/2=0.01*

*Degree of freedom: n-1= 2000-1=1999*

*Calculate t value at c.p 0.01 and 1999=* **2.328215**

*X bar+- t\*(sigma/root sample) = 200+-2.32\*0.67=200+-1.55*

1. *Calculation value of alpha at 96% C.I. =1-96/100=0.04*

*Cumulative probability=0.04/2=0.02*

*Degree of freedom: n-1= 2000-1=1999*

*Calculate t value at c.p 0.02 and 1999=* **2.05509**

*X bar+- t\*(sigma/root sample) = 200+-2.05\*0.67=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**

*Solved this question using excel sheet, refer attached.*

1. Find mean, median, variance, standard deviation. *41, 40.5, 25.5291, 5.052664*
2. What can we say about the student marks? *Mode for the student marks is 41, the mean and median are almost near, just a difference of 0.5 marks.*

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

Zero skewness

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

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

Q16) What does positive kurtosis value indicates for a data ? *3.95 (Leptokurtic*)

Q17) What does negative kurtosis value indicates for a data? *No Negative Kurtosis*

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



What can we say about the distribution of the data? T*he distribution is more on the either side having higher upper quartile*

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

*In Box plot 1 the value is more concentrated between 255-280 while in Box Plot 2, the value is spread between 225-315*

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.

Refer Cars.csv sheet for calculation

MPG <- Cars$MPG

* 1. P(MPG>38) *0.34827*
  2. P(MPG<40) *0.72907*

c. P (20<MPG<50) 0.89932

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

*Yes, it follows normal distribution but it is negatively skewed.*

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

Dataset: wc-at.csv

*NO*

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

*At 90% interval, Z score 1.65*

*At 94% interval, Z score 1.89*

*At 60% interval, Z score 0.85*

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

*At 95% CI, t score 2.064*

*At 96% CI, t score 2.2*

*At 99% CI, t score 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

*using t score value formula, t value= x-u/sigma(sq rt n)*

*we get t value as -0.47*

*d=n-1=17*

*hence prob=0.322*