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

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

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

Ans> probability that two heads and on tail :- 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?

Ans> 10/21

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.

Ans:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | point | score | weight | |
| mean | | **3.596563** | **3.217250** | **17.848750** | |
| median | | **3.695** | **3.325** | **17.710** | |
| mode | | 3.891 | 3.54 | 17.43 | |
| Variance | | **0.285881** | **0.957379** | **3.193166** | |
| Standard devation | | 0.534679 | 0.978457 | 1.786943 | |
| range | 2.76,4.93 | | 1.513,5.424 | | 14.5,22.9 |

**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?

Ans= 145.33 pounds

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

Cars speed and distance

Use Q9\_a.csv

Ans=

|  |  |  |
| --- | --- | --- |
|  | speed | distance |
| skewness | -0.117510 | 0.806895 |
| kurtosis | -0.508994 | 0.405053 |

Use Q9\_b.csv

|  |  |  |
| --- | --- | --- |
|  | sp | wight |
| skewness | 1.611450 | -0.614753 |
| kurtosis | 2.977329 | 0.950291 |

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



* The histogram is positively skewed
* The box plot is positive skewed
* The outlier is at the upper side



**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= **for 94% the confidence interval is :(201.2611 to 198.73)**

**For 98% the confidence interval is :(201.558 to 198.4415)**

**For 96% the confidence interval is :(201.375 to 198.6124)**

**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,variance=25.52 ,std=5.05**

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

**Ans= distribution is symmetric AND the skewness is eqal to zero**

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

Ans= **the distribution is positively skewed**

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

Ans= **the distribution is negatively skewed**

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

**Ans= peaked and possess thick tails**

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

**Ans= distribution is flat and has thin tails**

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



What can we say about the distribution of the data?

Ans= **not normal distribution**

What is nature of skewness of the data?

Ans= left side skewed

What will be the IQR of the data (approximately)?   
ans= 18-10=8 is IQR

Q19) Comment on the below Boxplot visualizations?



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

Ans= There is no outliers.

Average is same for both plots

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)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans= Normal distributed

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

Dataset: wc-at.csv

Ans= Adipose tissue (AT) ans Waist circumference (Waist) they are normal distributed

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

|  |  |  |  |
| --- | --- | --- | --- |
|  | 90% | 94% | 60% |
| Z scores | **1.64485362** | **1.8807** | **0.84162** |

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

|  |  |  |  |
| --- | --- | --- | --- |
|  | 95% | 96% | 99% |
| T scoer | 2.0638 | 2.1715 | 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

ans= tscore=-0.471

0.321