|  |  |
| --- | --- |
| 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) | ratio |
| Sales Figures | ratio |
| Blood Group | nominal |
| Time Of Day | nominal |
| Time on a Clock with Hands | interval |
| Number of Children | nominal |
| Religious Preference | nominal |
| Barometer Pressure | interval |
| SAT Scores | interval |
| Years of Education | ratio |

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

ANS:P(E)=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

ANS: a) Total outcomes = 0

Sum of two dice can never be 1. So probability = 0

b) Possible outcomes = {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}

Required probability =6/36=0.167

c) Sum is dividable by 2 and 3

E = {(5,1),(4,2),(3,3),(2,4),(1,5),(6,6)}

P(E) = n(E) /n(s) = 5/36

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: P(E)= 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:The expected number of candies for randomly selected child is approximately 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**

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 weight= sum of (Weights x Probabilities) =143.88

Expected weight of a randomly chosen patient =144pounds

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

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



ANS: About the histogram, the number of chickWeight around 50-100 is more . whereas the the number of chickWeight from 350-400 is very less, we can say it as a outlier. Our most data is concentrated in 0-250.when we draw a normally distributed graph, tracing the tips of histogram then .it is right tailed skewed distribution .hence it’s a positively skewed.



ANS:In this boxplot, the box represents the maximum data . and the dark line is the mean of our data. There are less data outside the box which is represented by dotted lines .and there are 7 outliers which are represented as circles .

**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%, 200 pounds 1.64 = 198.36 to 201.64

For 98%, 200 pounds 2.03 = 197.97 to 202.03

For 96%, 200 pounds 1.79 = 198.21 to 201.79

**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: Mean:41

Median:40.5

Mode:41

Standard deviation:5.0526

Variance:25.529

1. What can we say about the student marks?

ANS: Maximum mark is 56. whereas the minimum marks is 34.

Most common mark scored by students is 41.

There are more number of students who scored less than averagemark . The student’s marks are skewed to the right, which means that there are more scores at the higher end of the distribution

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

ANS: It is normal distribution.It is said to be symmetrical. Hence skewness is zero

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

ANS: Right tailed skewed distribution , positively skewed

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

ANS: Left tailed skewed distribution ,negatively skewed

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

ANS: Heavier tails and a more peaked distribution .

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

ANS: The distribution has lighter tail than normal distribution data.

Hence it does not have a sharp peak , it has a flat like surface in

Peak.lighter tails than normal distribution.

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



What can we say about the distribution of the data?

ANS: this data has left tail. The mean of this data is 15. the maximum data is covered around 10-18.

What is nature of skewness of the data?

ANS:It is negatively skewed .

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

ANS: 8

Q19) Comment on the below Boxplot visualizations?



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

ANS: Both box plots looks like normally distributed data. Boxplot (1) is smaller compared to boxplot(2). as the size of boxplot(1)is less compared to boxplot(2). the mean of both boxplots are 262.5.the length of data in boxplot(1) is 232-282.the length of the data in boxplot(2) is 200-350 . the maximum data covered in boxplot(1) is 250-275 , whereas in boxplot(2) around 225 - 300.

**Boxplot 1** has a higher median and IQR than **Boxplot 2**, indicating a distribution of higher values and greater variability. And there is no outliers in both the boxplot.

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:The MPG data is normally distributed. The distribution of MPG values is bell-shaped and symmetrical.

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

Dataset: wc-at.csv

ANS:The Adipose Tissue (AT) and Waist Circumference (WT) does not follows the Normal distribution.

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

ANS: z-score of 90% confidence interval =1.645

z-score of 94% confidence interval = 1.881

z-score of 60% confidence interval = 0.842

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

ANS:t-score of 95% confidence interval = 2.064

t-score of 96% confidence interval = 2.171

t-score of 99% confidence interval = 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

ANS: t = -0.4714045207910317

Probability: 0.32167411684460556