|  |  |
| --- | --- |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Nominal or categorical |

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 | Ratio |
| Weight | Ordinal |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ratio |
| Height | Ordinal |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Ratio |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| 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 Ans:0
2. Less than or equal to 4 Ans:6/36=1/6
3. Sum is divisible by 2 and 3 Ans:18/36+12/36=30/36=5/6

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: probability that none of the balls drawn is blue is 5C2/7C2=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 a randomly selected child is

Summation(x\*p(x))=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**

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: The expected value of the weight of a patient chosen at random is mean of all the patients i.e

145.34 pounds

**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:

50-100 weight having more frequency 180

350-400 weight having very less frequency 5

Positive skewness

Data is right skewed

Data is not a normal distribution

0-50 weight having 80 frequency

100-150 weight having 120 frequency



Ans:

* 7 Outliers are present in above box plot
* Positive skewness .i.e. data is right skewed
* DATA is not normally distributed
* Q1 is smaller than the Q3

**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: we don’t have the standard deviation for population. So we have to use the T-distribution to determine the CI of the given data

X̄ = 200 pounds,S = 30 pounds, n = 2000

X̄  ± t1-a,n-1 S/√n

Confidence interval for 94%:using R getting the

R code: qt(0.97,1999) = 1.88

Substituting values in the equation

200 ±1.88 30/√2000

Hence the confidence interval for 94% is [198,201]

Confidence interval for 96%:using R getting the

R code: qt(0.98,1999) = 2.05

Substituting values in the equation

200 ±2.05 30/√2000

Hence the confidence interval for 96% is [198.6,201.3]

Confidence interval for 98%:using R getting the

R code: qt(0.98,1999) = 2.328

Substituting values in the equation

200 ±2.328 30/√2000

Hence the confidence interval for 98% is [198.4,201.4]

**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, Variance=25.52941, SD=5.052664

1. What can we say about the student marks?

Ans: Avg of student marks 41

The students marks range from 34 to 56

Mode is 41

Most of students score is between 35 to 42

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

Ans:When the values of mean=median data is perfectly symmentrical

Skewness is zero.

Mean=median data is symmentrically distributed around the center you have 50% observations on left side 50% observations on right side no skewness.

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

Ans:mean>median nature is right skewed.

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

Ans:median<mean nature is left skewed.

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

Ans: A distribution with a positive kurtosis value indicates that the distribution has heavier tails than the normal distribution.

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

Ans: A negative kurtosis means that your distribution is flatter than a normal curve with the same mean and standard deviation

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



What can we say about the distribution of the data?

Ans:

* Skewed
* No outliers
* Q1 greater than Q3
* Median between 15 to 16
* Most of data present in range of 10 to 18
* Not following normal distribution
* Left skewness of data

What is nature of skewness of the data?

Ans:left skewed

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

Ans: IQR = Q3-Q1

=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.

Ans:

* Both the plots infer that their data is normally distributed.
* We can say that box plot 1 is for sample distribution and box plot 2 is for population or a sample with larger size.
* No outliers
* Q1 is 25%,Q3=75%.IQR is 50% for both the box plots. So we can say both the distributions follow normal distribution i.e mean=median=mode.

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

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

Dataset: wc-at.csv

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

Ans:

qnorm (0.95) # Z score for 90% confidence interval is 1.64485

qnorm(0.97) # Z score for 94% confidence interval is 1.8807

qnorm(0.80) # Z score for 60% confidence interval is 1.8416

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

Ans:

qt(0.975,24) #t score of 95% confidence interval is 2.0638

qt(0.98,24) #t score of 96% confidence interval is 2.171

qt(0.995,24) #t score of 99% confidence interval is 2.2.796

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