|  |  |
| --- | --- |
| 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 | Nominal |
| 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 | Ordinal |
| 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:- No. of possible combinations = 2^n = 2^3 = 8

Possible outcomes with 2 heads and 1 tail are 3

P(2 heads and 1 tail) = 3/8 = 0.375

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) n(s) = 36

There is nothing that we get sum is 1

n(A) = 0

Probability, P(A) = 0/36 = 0

b) n(B) = 3

Probability, P(B) = 3/36 = 0.0833

c) n(C) = 24

Probability, P(C) = 24/36 = 0.666

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:- Sample space = 7c2 = [7!/(7-2)!\*2!] = 21

Probability of none of the balls drawn blue = 5c2 = 10

P(A) = 10/21 = 0.47619

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:- Expected number of candies for a randomly selected child

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

= 0.015+0.8+1.95+0.025+0.06+0.24

= 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**

**Ans:- For Points:- mean= 3.596, median= 3.695, mode= 3.92, variance= 0.2858,**

**std= 0.5346 and range= 2.17**

**From mean and median, we can say that there are no outliers inside a dataset and also there is a less variation.**

**For Score:- mean= 3.2172, median= 3.325, mode= 3.44, variance= 0.9573, std=**

**0.9784 and range= 3.911**

**There is a variation in a dataset.**

**For Weigh:- mean= 17.8487, median= 17.71, mode= 17.02, variance= 3.1931,**

**std= 1.7869 and range= 8.4**

**There is a variation in a dataset.**

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 Value = Sum of(Probability\* value)**

**There are 9 patients**

**Hence probability of one patient randomly = 1/9**

**Expected Value =**

**(1/9)\*108+(1/9)\*110+(1/9)\*123+(1/9)\*134+(1/9)\*135+(1/9)\*145+(1/9)\*167+(1/9**

**)\*187+(1/9)\*199**

**= (1/9) \* 1308**

**= 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:- For Speed:- mean = 15.4, median = 15 and Kurtosis = -0.5089**

**Here, Mean > median So Speed data is right skewed and Kurtosis shows**

**that distribution in not peaked and not too flat.**

**For Distance:- mean = 42.98, median = 36 and Kurtosis = 0.4050**

**Here, Mean > median So Speed data is right skewed and Kurtosis shows**

**that distribution in not peaked and not too flat.**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:- For SP:- skewness = 1.61145 and kurtosis = 2.9773**

**The distribution of SP is right skewed there is peakedness in distribution i.e.**

**positive kurtosis.**

**For WT:- skewness = -0.6147 and kurtosis = 0.9502**

**The distribution of SP is near to normal and kurtosis shows distribution in**

**not too peak and not too flat.**

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

**Ans:- The Histogram shows dataset is positive skewed and hence dataset has**

**some outliers. The majority of our data is lies in between 0 to 200. From Boxplot,**

**we can see that dataset has an outlier.**



**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:- Confidence Interval = x̄ ± z \* ơ / √n

Where, x̄ =mean, Z=value of z-stat at confidence level

Ơ=standard deviation, n=no of sample

a)AT 94%:-

=200±1.96\*30/√2000

=200 ± 1.314

=[198.686, 201.314]

b)AT 98%:-

=200±2.236\*30/√200

=200 ± 1.562 =[198.438,201.562]

c)AT 94%:-

=200±2.04\*30/√2000

=200 ± 1.375

=[198.625, 201.375]

**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. Deviation = 5.052

The average marks obtained by students is 41. The median value is 40.5 that

means center of the marks is 40.5. The variance of students marks shows large

difference between mean and variance. Hence, there is an outlier is present in a

data.

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

Ans:- Zero Skew

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

Ans:- Positive Skew

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

Ans:- Negative Skew

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

Ans:- Positive kurtosis shows peakedness in a data. It means our data is not

spread.

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

Ans:- Negative kurtosis shows that our dataset is spread.

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



What can we say about the distribution of the data?

Ans:- The dataset is not normally distributed. It shows Negative skeweness. The

mean of dataset is between 15 to 16.

What is nature of skewness of the data?

Ans:- Negative Skewed

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

Ans:- IQR = Q3-Q1 = 18.2-10 = 8.2

Q19) Comment on the below Boxplot visualizations?



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

Ans:- For Boxplot 1, data is lies between 240 to 285 and for boxplot 2, data lies

between 180 to 340. The mean of boxplot 1 and boxplot 2 is same.

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)
  3. P (20<MPG<50)

Ans:- a) P(MPG&gt;38):- There are 81 total rows in MPG cars dataset. Amongst them

33 rows contains value MPG&gt;38

P(MPG&gt;38) = 33/81

b) P(MPG&lt;40) = 61/81

c) P(20&lt;MPG&lt;50) = 69/81

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:- mean = 34.24, median = 35.15 and mode = 29.62 Mean, median and

mode are nearly equal. So MPG of cars is normally 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:- For Waist:- mean, median and mode are 90.90, 90.8, 94.5

Hence, Waist is normally distributed.

For AT:- mean, median and mode are 101.89, 96.54, 121

Here Mean > Median. Hence, AT is not normally distributed, it is right

skewed.

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

Ans:- For 90% CI:- AL = (1+ CL)/2 = (1+0.90)/2 = 0.95 , So from Z- table ,

Z-score = 1.645

For 94% CI:- Z-score = 1.89

For 60% CI:- 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

Ans:- degree of freedom= n-1 = 25-1 = 24

For 95% CI, t-score= 2.0639

For 96% CI, t-score=

For 99% CI, t-score= 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:- t = (sample mean-population mean)/((sample std)/sqrt(n))

t = (260-270)/(90/18\*\*0.5)

t = -0.471

For probability calculations, the number of degrees of freedom is n - 1, so

here you need the t-distribution with 17 degrees of freedom.

The probability that t &lt; - 0.471 with 17 degrees of freedom assuming the

population mean is true, the t-value is less than the t-value obtained With

17 degrees of freedom and a t score of - 0.471, the probability of the bulbs

lasting less than 260 days on average of 0.3218 assuming the mean life of

the bulbs is 300 days.