|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Categorical |
| Results of rolling a dice | Categorical |
| 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 | Categorical |
| Number of tickets in Indian railways | Categorical |
| Number of times married | Categorical |
| Gender (Male or Female) | 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 | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Ordinal |

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

Ans: 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
4. 0
5. 3/36 = 0.083
6. 6/36 = 0.17

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 = 0.476

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.

**Use Q7.csv file**

**ANS:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | MEAN | MEDIAN | MODE | Variance | Std. Dev | Range |
| Points | 2.08 | 3.69 | 3.92 | 0.29 | 0.53 | 2.17 |
| Score | 3.22 | 3.32 | 3.44 | 0.96 | 0.98 | 3.91 |
| Weigh | 17.85 | 17.71 | 17.02 | 3.19 | 1.79 | 8.40 |

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: Total – 9 Patients

Probability of choosing 1 patient – 1/9

Expected value = 1/9 (108 + 110+ 123 + 134 + 135 + 145 + 167 + 187 + 199)

= 1/9 x 1308 = 145

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

**Cars speed and distance**

**Use Q9\_a.csv**

**ANS:**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| speed | -0.11 | -0.50 |
| dist | 0.78 | 0.40 |

Both Skewness & Kurtosis is negative for Speed

Both Skewness & Kurtosis positive for Distance

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| SP | 1.58 | 2.98 |
| WT | -0.60 | 0.95 |

Both Skewness & Kurtosis positive for SP

Skewness is negative & Kurtosis is positive for WT

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



Ans: The histogram is right skewed, 200 chicks is having weight from 50-100



Ans: Outlier is present and box plot is right skewed.

**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: Sample size = 2000

Mean = 200

St devi = 30

|  |  |  |  |
| --- | --- | --- | --- |
|  | 94% | 98% | 96% |
| Upper | 201.04 | 201.38 | 201.17 |
| Lower | 198.96 | 198.62 | 198.83 |

**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 | 40.50 |
| Variance | 25.53 |
| Std Deviation | 5.05 |

1. What can we say about the student marks?

Ans: Here we can say 56 is an outlier.

Many students have got 41 marks.

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

Ans: Normal Distribution, Bell Curve we will get.

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

Ans: Positively skewed.

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

Ans: Negatively skewed.

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

Ans: It means the distribution is thinner, Center of the distribution would be thinner and tail would be vice-versa.

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

Ans: It means the distribution is thicker, Center of the distribution would be thicker and tail would be vice-versa.

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



What can we say about the distribution of the data?

Ans: Not normally distributed.

What is nature of skewness of the data?

Ans: Negatively skewed.

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

Ans: IQR = Q3 – Q1 = 18 – 10 = 8

Q19) Comment on the below Boxplot visualizations?



Ans: 1) The Max is around – 287.5

The Min is around – 237.5

The mean is – 262.5

IQR = Q3 – Q1 = 275 – 250 = 25

No outlier is present.

1. The Max is around – 350

The Min is around – 187.5

The mean is – 262.5

IQR = Q3 – Q1 = 312.5 – 212.5 = 100

No outlier is present.

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

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. 34.8%

b. 72. 89%

c. 89.9%

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: Mean – 34.42

Median – 35.12

Mode - 29.62

As Mean, median & mode are not equal so MPG of cars is followed the Normal Distributions.

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

Dataset: wc-at.csv

Ans: Same here also the mean, median & mode are not equal so both of them also not follow the normal distribution.

|  |  |  |
| --- | --- | --- |
|  | Waist | AT |
| Mean | 91.90183486 | 101.8940367 |
| Median | 90.8 | |  | | --- | | 96.54 | |
| Mode | 94.5 | 121 |

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

Ans: A = (1 + CI)/2 where A is area under the normal distribution curve and CL represents the confidence level

A = (1 + 0.9)/2 = 1.9/2 = 0.95

And by Z table it is 1.645

So the z score for 90% CI would be 1.645

Same for 94%

A = (1+ 0.94)/2 = 1.94/2 = 0.97 and by z table it is close to 1.88 & 1.89

So its average is (1.88 + 1.89)/2 = 1.885

So the z score for 94% CI would be 1.645

Same for 60%

A = (1 + 0.6)/2 = 1.6/2 = 0.8 and by z table it is close to 0.84 & 0.85

So its average is (0.84 + 0.85)/2 = 0.845

So the z score for 60% CI would be 0.845

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

Ans: If the sample size is 25 then degree of freedom = (n – 1) = (25 – 1) = 24

So for 95% CL, the significance value would be 1 - 0.95 = 0.05 the t score would be 1.711.

For 96% CL, the significance value would be 1 - 0.96 = 0.04 the t score would be 1.828.

For 99% CL, the significance value would be 1 - 0.99 = 0.01 the t score would be 2.492.

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

Sol: Population mean = 270

Sample size = 18 so df = 18-1 = 17

Sample mean = 260

St. dev = 90

Formula to calculate t score is

[](https://www.statisticshowto.com/wp-content/uploads/2013/11/t-score.jpg)

= 260 – 270

90

(18)1/2

= -10

90/4.2426

= -10/21.2134

= -0.4714