|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continous |
| Weight of Gold | Continous |
| Distance between two places | Continous |
| Length of a leaf | Continous |
| Dog's weight | Continous |
| 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:

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 | 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?

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 2and 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?

(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

1. Expected number of candies for a randomly selected child is 1(0.015)+4(0.20)+3(0.65)+5(0.005)+6(0.01)+2(0.120)=3.090

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**

**Points:**

Mean of Points=3.5965625

Median of points=3.695

Mode of points=3.92

Variance of points=0.28588135080645

Standard Deviation of points=0.53467873607097

Range of points=2.17

**Score:**

Mean of score=3.21725

Median of score=3.325

Mode of score=3.44

Variance of score=0.95737896774194

Standard Deviation of score=0.9784574429897

Range of score=3.911

**Weigh:**

Mean of weigh=17.84875

Median of weigh=17.71

Mode of weigh=17.02

Variance of weigh=3.193166

Standard Deviation of weigh=1.786943

Range of weigh=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?

Expected value of the weight of the patient is 1/9(108+110+123+134+135+145+167+187+199) =145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**

Skewness of speed: -0.11751

Kurtosis of speed: -0.50899

Skewness of distance: 0.806895

Kurtosis of distance: 0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

Skewness of weight: -0.61475

Kurtosis of weight: 0.950291

Skewness of SP: 1.61145

Kurtosis of SP: 2.977329

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



Most of the data points are concentrated on 50-100 and the outlier lies on 350-400 .



There are so many outliers exists

**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?

Confidence interval for 94% is 200±1.26248398

Confidence interval for 98% is 200±1.56166987

Confidence interval for 96% is 200±1.37853591

**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.

Mean: 41

Median: 40.5

Variance: 25.52941

Standard Deviation: 5.052664

1. What can we say about the student marks?

Mean, Median and Mode are equal so distribution can be assumed to be approximately symmetrical.

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

No skewness

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

Positively skewed (right)

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

Negatively skewed (Left)

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

Distribution is peaked and possess thick tails.

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

Lighter tails and flatter peeks than normal distribution

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



What can we say about the distribution of the data?

Asymmetric Distribution

What is nature of skewness of the data?

Left skewed

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

8

Q19) Comment on the below Boxplot visualizations?



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

Mean and Median are same so boxlpot1 and boxplot 2 are symmetrical and wide spread in boxplot2

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 $ 1MPG

* 1. P(MPG>38) -> 33/81
  2. P(MPG<40)->61/81

c. P (20<MPG<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

Yes, MPG of cars follows Normal Distribution because p value is 0.1764 and w is 0.97797

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

Dataset: wc-at.csv

Waist is normally distributed(mean and median are appr. Similar)

AT is not normally distributed

(P is less than 0.005)

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

Z score of 90% CI is 1.645

Z score of 94% CI is 1.89

Z score of 60% CI is 0.85

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

t scores of 95% confidence interval­­ ->2.064

t scores of 96% confidence interval ->2.172

t scores 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

df(-0.471,17)=0.3218