|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Continuous |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous |
| Weight of Gold | Discrete |
| Distance between two places | Discrete |
| Length of a leaf | Discrete |
| Dog's weight | Discrete |
| Blue Color | Continuous |
| Number of kids | Continuous |
| Number of tickets in Indian railways | Continuous |
| Number of times married | Continuous |
| Gender (Male or Female) | Continuous |

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) | 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?

Answer : 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 2 and 3 =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?

Answer 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

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.

Mean: Mean is the Average Value of Data Set

Median: Median is the Middle Value of the Data set

Mode: Most Occurring Value in Data set is Called Mode

Standard Deviation: Standard deviation is the mostly used measure of variation, which describes how spread out the data is.

Range: The range is the difference between the smallest and the largest value of the data

Mean Value is - mean(Q7$Points)

[1] 3.596563

Median Value is - > median(Q7$Points)

[1] 3.695

Mode Value is - mode<-(Q7$Points)

> mode

[1] 3.90 3.90 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 3.92 3.07 3.07 3.07 2.93 3.00 3.23 4.08 4.93 4.22 3.70 2.76 3.15

[24] 3.73 3.08 4.08 4.43 3.77 4.22 3.62 3.54 4.11

Mean Value for Score: > mean(Q7$Score)

[1] 3.21725

Median Value for Score: > median(Q7$Score)

[1] 3.325

Mode Value for Score: mode<-(Q7$Score)

> mode

[1] 2.620 2.875 2.320 3.215 3.440 3.460 3.570 3.190 3.150 3.440 3.440 4.070 3.730 3.780 5.250 5.424 5.345 2.200 1.615

[20] 1.835 2.465 3.520 3.435 3.840 3.845 1.935 2.140 1.513 3.170 2.770 3.570 2.780

Mean for Weigh- mean(Q7$Weigh)

[1] 17.84875

Median for weigh- median(Q7$Weigh)

[1] 17.71

Mode for Weigh- mode<-(Q7$Weigh)

> mode

[1] 16.46 17.02 18.61 19.44 17.02 20.22 15.84 20.00 22.90 18.30 18.90 17.40 17.60 18.00 17.98 17.82 17.42 19.47 18.52

[20] 19.90 20.01 16.87 17.30 15.41 17.05 18.90 16.70 16.90 14.50 15.50 14.60 18.60

>

Standard Deviation for Points: sd(Q7$Points)

[1] 0.5346787

Standard Deviation for Score: > sd(Q7$Score)

[1] 0.9784574

Standard Deviation for Weigh> sd(Q7$Weigh)

[1] 1.786943

Range for Points:> range(Q7$Points)

1] 2.76 4.93

Range for Score:[> range(Q7$Score)

[1] 1.513 5.424

Range for Weigh:[> range(Q7$Weigh)

[1] 14.5 22.9

**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?

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Skewness: Its Measure the Symmetry of the distribution**

**Kurtosis: Its Measure the Preakness of the Distribution**

**> skewness(Q9$speed)**

**[1] -0.1139548**

**> skewness(Q9$dist)**

**[1] 0.7824835**

**> kurtosis(Q9$speed)**

**[1] 2.422853**

**> kurtosis(Q9$dist)**

**[1] 3.248019**

**SP and Weight(WT)**

**> skewness(Q9\_B$SP)**

**[1] 1.581454**

**> skewness(Q9\_B$WT)**

**[1] -0.6033099**

**> kurtosis(Q9\_B$SP)**

**[1] 5.723521**

**> kurtosis(Q9\_B$WT)**

**[1] 3.819466**

**Use Q9\_b.csv**

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

Chart, histogram

Description automatically generated

Answer:This Histogram Shows the Frequency Distribution. How Many observations take the value within the interval

**Boxplot:**

Shape, rectangle

Description automatically generated

Answer :This Box plot shows the distance between Largest and smallest Value And How Outliers are there

**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?

Mean =200

SD = 30

**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?

**Answer 2 :49 and 56 are the outliers**

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

Q14) What is the nature of skewness when mean > median ?Skewness is Positive(Right Skewed Distribution)

Q15) What is the nature of skewness when median > mean? Skewness is Negative(Left skewed Distribution)

Q16) What does positive kurtosis value indicates for a data ? Distribution is peaked and it having thick tails

Q17) What does negative kurtosis value indicates for a data? Distribution is Flat and it having thin tails

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

Chart, box and whisker chart

Description automatically generated

What can we say about the distribution of the data? distribution is negatively skewed

What is nature of skewness of the data? Negative Skewed

What will be the IQR of the data (approximately)? IQR will be 10 to 18

Q19) Comment on the below Boxplot visualizations?

Chart, box and whisker chart

Description automatically generated

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

Boxplot 2.

Answer :1. Boxplot 2 is Having Maximum Value Above 325 And Minimum Value 200

2.BoxPlot 1 is Having Maximum Value above 275 And Minimum Value is 200

3. Median Value is Same for Boxplot 1 And 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)- 0.6524161898269749
  2. P(MPG<40)- 0.729362470706113

c. P (20<MPG<50)- 0.8988852898457339

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer: Yes its Following the Normal Distribution

Graph is in Python file

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

Dataset: wc-at.csv  
Answer : Yes Both AT And Waist is following the Normal Distribution

Graph is in Python File

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

Answer are in Python File

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

Answer are in Python File

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