|  |  |
| --- | --- |
| 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 | Ordinal |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ordinal |
| Type of living accommodation | Nominal |
| 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 | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Continuous |
| SAT Scores | Ordinal |
| Years of Education | Ordinal |

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

Answer:

{HHH, THH, TTH, TTT, HTT, HHT, THT,HTH}

Out of these outcomes we have the probability of getting Two heads and

One tail is 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

Total possible outcomes = 36

1. Equal to 1 = 0 (sum always exceed to 1)
2. Less than or equal to 4 = 1/12

The possibility of getting number 4 is (1,3), (2,2), (3,1)

Total number of possibilities = 3/36 = 1/12

1. Sum is divisible by 2 and 3 = 5/36

Outcomes = (1,5), (3,3), (4,2), (5,1), (6,6)

Total number of possibilities = 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:

Total number numbers of balls = (2+3+2) = 7

n(S) = Number of ways of drawing 2 balls out of 7

= (7\*6) / (2\*1)

= 21

Let E = Event of drawing 2 balls, none of which is blue.

n(E) = Number of ways of drawing 2 balls out of (2+3) balls

= (5\*4) / (2\*1)

= 10

P(E) = n(E) / n(s) = 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

Answer:

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, Weight

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

Answer : In jupyter notebook

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?

Answer:

(probability \* value)

There are 9 patients,

Probability of selecting each patients = 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) (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9) (1308)

= 145.33

**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**

Answer : In jupyter notebook

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



Inference:

* In this histogram, the data is right skewed that is positively right skewed.
* More than 50% chick weight is between 50 to 150.



Inference:

* At upper extreme there are some outliers

**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?

Answer : In jupyter notebook

**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 : In jupyter notebook

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

Answer:

When mean, median of data are equal then there is no skewness that is zero

Skewness.

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

Answer:

When mean is greater than median then the distribution is positively

Skewed.

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

Answer:

When median is greater than mean then the distribution is negatively

Skewed.

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

Answer:

Positive kurtosis value indicates the distribution is peaked.

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

Answer:

Negative kurtosis value indicates the distribution is wider peak and

Thinner tails.

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



What can we say about the distribution of the data?

Answer:

The above boxplot is not normally distributed the median is towards the

Higher value.

What is nature of skewness of the data?

Answer:

The data is negatively left skewed, the whisker range of minimum value is

Greater than maximum.

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

Answer:

The inter quartile range = Upper quartile – lower quartile

= 18 – 10

= 8

Inter quartile range = 8

Q19) Comment on the below Boxplot visualizations?



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

Answer:

- There are no outliers in both the boxplots.

- Both boxplot have same median that ranges between 275 to 250.

- Both are normally distributed with zero skewness neither at minimum nor

At maximum whisker range.

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)

Answer : In jupyter notebook

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer:

The MPG of cars follows normal distribution.

1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist)

From wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Answer:

Adipose Tissue (AT) and Waist circumference (waist) does not follow

Normal distribution.

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

Answer:

For 90% confidence interval,

We have significance level as 5% as it is a two tailed test,

Alpha = 5% = 0.05

From Z table,

Z = 1.645

For 94% confidence interval,

We have significance level at 3% as it is a two tailed test,

Alpha = 3% = 0.03

From Z table,

Z = 1.555

For 60% confidence interval,

We have significance level at 20% as it is a two tailed test,

Alpha = 3% = 0.2

From Z table,

Z = 0.253

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

Answer : In jupyter notebook

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) S

df 🡪 degrees of freedom

Answer : In jupyter notebook