|  |  |
| --- | --- |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| 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 | Nominal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Nominal |
| IQ(Intelligence Scale) | Nominal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Interval |

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

Possible outcomes

H H H

T H H

T T H

T T T

H T H

H H T

H T T

T H T

P (2H, 1T) =3/8

**Q4)** Two Dice are rolled, find the probability that sum is

1. Equal to 1: - P(sum=1)=0
2. Less than or equal to 4: -- P(sum<=4)=3/36

1,3

3,1

2,2

1. Sum is divisible by 2 and 3: --P(sum/2=0 and sum/3=0)=6/36

**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?

R, G

R, B

G, B

R, R

G, G,

B, B

P (Not Blue) =3/6

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

Expected number of candies for a randomly selected child =∑x. P(x)=1\*0.015+2\*0.120+3\*0.65+4\*0.20+5\*0.005+6\*0.01=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 Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**(Solution in Assignment1-Final.ipynb)**

**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 random patient E(x)=108+110+123+134+135+145+167+187+199/9=145.4

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

1. Car’s speed and distance

**Use Q9\_a.csv**

**(Solution in Assignment1-Final.ipynb)**

b. SP and Weight (WT)

**Use Q9\_b.csv**

**(Solution in Assignment1-Final.ipynb)**

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



**Inferences:**

The histogram for chick weight is right skewed distribution and is asymmetrical and looks like platykurtic.

The box plot shows outliers above whiskers and 25 percentiles,50 percentiles(median) and 75 percentile in the Inter Quartile range.

The distribution is assymetrical and right skewed based on the median showing on box plot.

**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?

Sample:

n=2000 men

=200 pounds

=30 pounds

Population:

N=3000000

Confidence Interval=Point estimate± Margin of Error

**(Solution in Assignment1-Final.ipynb)**

**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?

**(Solution in Assignment1-Final.ipynb)**

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

When mean, median of data is equal then it is symmetrical distribution like Normal distribution.

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

When mean>median then it is as asymmetrical called right skewed distribution.

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

When mean>median then it is as asymmetrical called left skewed distribution.

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

Positive values of kurtosis indicate that a distribution is peaked and possess thick tails. Leptokurtic distributions have positive kurtosis values.

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

Negative values of kurtosis indicate that a distribution is flat and has thin tails. Platykurtic distributions have negative kurtosis values.

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



1. What can we say about the distribution of the data?

The distribution is not normal distribution and is asymmetrical. the median will be >mean. Many values are not around the mean.

1. What is nature of skewness of the data?

The distribution is left skewed.

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

IQR ~8

**Q 19)** Comment on the below Boxplot visualizations?



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

**Inferences:**

The box plot 1 and boxplot 2 has same median value.

The box plot 2 has greater IQR than boxplot1.

The data of boxplot 2 and boxplot 1 looks symmetrical with no outliers.

The data of boxplot 2 is more spread than boxplot1 data.

**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)

c. P (20<MPG<50)

**(Solution in Assignment1-Final.ipynb)**

**Q 21)** Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**(Solution in Assignment1-Final.ipynb)**

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

Dataset: wc-at.csv

**(Solution in Assignment1-Final.ipynb)**

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

**(Solution in Assignment1-Final.ipynb)**

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

**(Solution in Assignment1-Final.ipynb)**

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

**solution:**

n=18

sample mean=260days

sample std=90days

population mean=270days

find t-score

degrees of freedom n-1=17

t-score==-0.417

P(x<=260) =0.3409

Stats.t.cdf(-0.417,17)=0.3409