|  |  |
| --- | --- |
| 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 | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Nominal |

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 | Interval |
| Type of living accommodation | Ratio |
| Level of Agreement | Ratio |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Interval |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal |
| Years of Education | Interval |

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

Answer: If three coins are tossed then the total probability would be,

(HHH,HTT,HTH,TTT,THT,THH,TTH,HHT)

So the probability of getting two heads and one tail = 3/8.

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

Answer: When Two Dices are rolled, the total probability would be 36 (6x6).

1. Equal to 1 = 0/36
2. Less than or equal to 4 = 6/36
3. Sum divisible by 2 and 3 =

Total probability would be: 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?

Answer: Using combination formula the total outcome would be 7C2 = 21 ways.

The probability of 2 red and 3 green would be 5C2 = 10 ways.

Probability = Number of ways to choose 2 non-blue balls / Total ways to choose 2 balls.

So, the probability is 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: 1x0.015 = 0.015

4x0.20 = 0.8

3x0.65 = 1.95

5x0.005 = 0.025

6x0.01 = 0.06

2x0.120 = 0.4

Total probability = 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.**

Answer: File attached with the mail. (Assignment Q7.csv.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?

Answer: Adding the values and dividing it with total values would give the expected value.

108+110+123+134+135+145+167+187+199 / 9 = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

Answer: File attached with the mail. (Assignment Q9\_a.ipynb)

**SP and Weight(WT)**

**Use Q9\_b.csv**

Answer: File attached with the mail. (Assignment Q9\_b.ipynb)

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



Answer: I'm sorry, but I don't know the answer to that question.

**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: I'm sorry, but I don't know the answer to that question.

**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: Mean: 41.0

Median: 40.5

Variance: 24.1111

Standard deviation: 4.9103

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

Answer: Neutral or symmetrically distributed. There will be no skewness.

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

Answer: If mean > median, the data is positively skewed.

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

Answer: If median > mean, the data is negatively skewed.

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

Answer: More excessive data expected than normal distribution.

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

Answer: Less excessive data expected than normal distribution.

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



What can we say about the distribution of data?

Answer: Data is distributed along the left side of plot.

What is the nature of skewness of the data?

Answer: Left skewed.

What will be the IQR of the data (approximately)?   
Answer: Q3-Q1 =7

Q19) Comment on the below Boxplot visualizations?



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)

Answer: File attached with the mail. (Assignment Q20 and Q21A.ipynb)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer: No,this does not follow Normal Distribution.

Code given in above 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: File attached with the mail. (Assignment Q21B.ipynb)

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

Answer: To find Z score, 1 - ( C % 100)

Z Score at 90% = 0.89

Z Score at 94% = 0.93

Z Score at 60% = 0.59

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

Answer: To find T score, 1 - ( C % 100) and divide by 2.

T Score at 95% = 0.47

T Score at 96% = 0.475

T Score at 99% = 0.49

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

Answer: We use t-score here.

t-score = (260 - 270) / (90 / sqrt(18)) = -0.471

p- value - 0.3218