|  |  |
| --- | --- |
| 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 | Ratio |
| Weight | Interval |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Ratio |
| Height | Interval |
| Type of living accommodation | Odinal |
| Level of Agreement | Nominal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Nominal |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Interval |

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 2 and 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
* 7C2
* (7\*6) / (2\*1)
* 21
* E = Event of Drawing Balls, None of which is BLUE
* n(E) = number of ways of drawing 2 balls out of (2+3) balls
* 5C2
* (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

* 1 \* 0.015 + 4 \* 0.20 + 3 \* 0.65 + 5 \* 0.005 + 6 \* 0.01 + 2 \* 0.12
* 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,Weigh>

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

**Use Q7.csv file**

**In Jupyter Notebook 🡪 Attached below in the mail.**

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 = (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 = **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**

* **In Jupyter Notebook** 🡪 **Attached below in the mail.**

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



* Histogram : From the figure above it is clear that the distribution is **right-skewed or positively skewed.**

Mean of ChickWeight$weight is about **75**.

Range 50-100 and range 350-400 have the most and the least frequency respectively.



* Boxplot: From the figure above it is clear that the given data contains **outliers**. Distribution of frequency has a little skew.

**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?

* **In Jupyter Notebook** 🡪 **Attached below in the mail.**

**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?

**1) Answer in Jupyter Notebook** 🡪 **Attached below in the mail.**

**2) Average students marks = 41**

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

* When Mean, Median of Data are equal, we can say that it is Perfectly Symmetric (i.e no Skewness – Symmetric)

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

* The nature of skewness is Right/Positively skewed

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

* The nature of skewness is Left/Negatively skewed

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

* Positive Kurtosis value indicates that thinner peak and wider tails than the normal distribution [peakness (sharp peak) and less variation.]

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

* Positive Kurtosis value indicates that lighter peak and wider tails than the normal distribution [less peakness (Broad peak) and more variation.]

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



What can we say about the distribution of the data?

* Data is Skewed distribution. Not normally distributed.

What is nature of skewness of the data?

* Negatively/Left Skewed

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

* IQR = 18 – 10
* 8

Q19) Comment on the below Boxplot visualizations?



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

1) The median of the two boxplots are same approximately 260.

2) The boxplots are not skewed in +ve or –ve direction.

3) Outliers doesn’t exist in both of the boxplots.

4) Both Boxplot follows Normal Distribution.

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)

* **In Jupyter Notebook** 🡪 **Attached below in the mail.**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

* **In Jupyter Notebook** 🡪 **Attached below in the mail.**
* From Histogram plotted with the help of Data, we can say that 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

* **In Jupyter Notebook** 🡪 **Attached below in the mail.**

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

* **In Jupyter Notebook** 🡪 **Attached below in the mail.**

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

* **In Jupyter Notebook** 🡪 **Attached below in the mail.**

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

* **In Jupyter Notebook** 🡪 **Attached below in the mail.**

P = 0.3216 = 32%