|  |  |
| --- | --- |
| 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 | ***Ratio*** |
| Hair Color | ***Nominal*** |
| Socioeconomic Status | ***Ordinal*** |
| Fahrenheit Temperature | ***Interval*** |
| Height | ***Ratio*** |
| Type of living accommodation | ***Ordinal*** |
| Level of Agreement | ***Nominal*** |
| IQ(Intelligence Scale) | ***Ratio*** |
| 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 | ***Ratio*** |
| SAT Scores | ***Interval*** |
| Years of Education | ***Ratio*** |

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

**Solution:**

There are 8 outcomes when Three coins are Tossed Which are;

S= [HHH, HHT, HTH, THH, TTH, THT, HTT, TTT]

N(S) = 8

P = P(A) / N(S), Where P(S) is Probability that two head and one tail

P = **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

**Solution:**

One dice has 6 outcomes, when 2 Dice are rolled, So 6x6 = 36 Outcomes

N(S)=36, P = P(A) / N(S),

1. **Equal to 1**: No Outcomes is equal to 1, so P = 0/36 = **0**
2. **Less than or equal to 4:** 6 Outcomes, so P = 6/36 = **1/6**
3. **Sum is divisible by 2 and 3:** 24 Outcomes, so P = 24/36 = **2/3**

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?

**Solution:** Using Probability formula nCr:

Total no of balls (n) = 2 Red + 3 Green + 2 Blue = 7

n(S): Two balls are drawn at random maybe its blue or green or red = nCr 7C2=7\*6/2\*1=21

P(B): find the probability that none of balls drawn is blue so it means it would be 2 red or 2 green or 1 red and 1 green = nCr = 2C2+3C2+2C1\*3C1

= (2\*1/2\*1) + (3\*2/2\*1) + (2\*3/1) =1+3+6=10

P(B)=n(B)/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

**Solution:**

Expected Value E(x) = Σ xi \* P(xi)

Expected number of Candies E(x) = Candies count\* Probability

E(x) = 1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

E(x) = **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**

**Solution:** For Code Refer - Jupyter Notebook PDF Attached

**Mean:**

* Points: 3.596563
* Score: 3.217250
* Weigh: 17.848750

**Median:**

* Points: 3.695000
* Score: 3.325000
* Weigh: 17.710000

**Mode:**

* Points: 3.07, 3.92
* Score: 3.44
* Weigh: 17.02, 18.90

**Variance:**

* Points: 0.2858813
* Score: 0.9573789
* Weigh: 3.193166

**Standard Deviation:**

* Points: 0.534679
* Score: 0.978457
* Weigh: 1.786943

**Range:**

* Points: 2.17
* Score: 3.91105
* Weigh: 8.39

**Inference:** From obtained value we can say that standard deviation is very small so all values are closer to their mean values. There is no outlier detected.

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?

**Solution:**

Expected Value E(x) = Σ xi \* P(xi)

There are 9 patients, Probability of selecting each patient, P = 1/9

E(x) = 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)

E(x) = (1/9) (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

E(x) = (1/9) (1308)

E(x) = 145.33

Expected Value of the Weight of that patient = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Solution:** For Code Refer – Jupyter Notebook PDF Attached

**Skewness:**

* Speed: -0.11
* Dist: 0.80

**Kurtosis:**

* Speed: -0.50
* Dist: 0.40

**SP and Weight (WT)**

**Use Q9\_b.csv**

**Skewness:**

* SP: 1.6
* WT: -0.61

**Kurtosis:**

* SP: 2.97
* WT: 0.95

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

**Solution:**

Inferences on Histogram:

* Histogram is of chick weight Vs frequency
* Histogram is not symmetric
* The Tail is going on the right-hand side, means Data is positively skewed.
* Centre of data for weight group is about 50-150 and then further values becomes rare.
* Centre of data for frequency is about 100(most values between 0 to 100)



Inferences on Boxplot:

* Boxplot showing most values falls in upper quartile and comparatively less values falls in lower quartile
* This Boxplot has 7 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?

**Solution:** For Code Refer – Jupyter Notebook PDF Attached

CI at 94%: [198.7376, 201.2623]

CI at 98%: [198.4381, 201.5618]

CI at 96%: [198.6214, 201.3785]

**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:**

* Mean = 41
* Median = 40.5
* Variance = 25.52
* Standard Deviation = 5.05

There are total 18 students. The Maximum marks obtained is 56 and Minimum marks obtained is 34. 41 marks is obtained by most student i.e.,4 which is also the mode of this dataset.

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

**Solution:** If Mean & Median is equal, then the distribution is symmetric, then the nature of skewness is, **Zero Skewness or No Skewness.**

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

**Solution:** If Mean > Median, then the nature of skewness is, **Positively Skewed**

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

**Solution:** If Median >Mean, then the nature of skewness is, **Negatively Skewed**

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

**Solution:** “Positive Kurtosis” indicates a distribution where more of the values are located in the tails of the distribution rather than around the mean **“Sharp Peak”**.

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

**Solution:** “Negative values of Kurtosis” indicate that a distribution is flat and has thin tails **“Broad Peak”**.

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



What can we say about the distribution of the data?

**Solution:** Most of the data points comes in lower quartile. There are more Data points Between Q1 & Q2 then between Q2 & Q3. So, this is not a **Normal Distribution**.

What is nature of skewness of the data?

**Solution:** As is it can be observed Median > Mean. So, it **Negatively Skewed**.

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

**Solution:** Inter Quartile Range (IQR) = Upper Quartile – Lower Quartile

IOR = 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.

**Solution:**

* Both Boxplots follows Normal Distribution.
* Median of both boxplots are same.
* As they follow normal distribution, skewness is less.
* There are no Outliers present in either boxplot.

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)

**Solution:**

1. P(MPG>38) = 0.3474
2. P(MPG<40) = 0.7294
3. P (20<MPG<50) = 0.8989

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Solution:** For Graph Refer – Jupyter Notebook PDF Attached

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

Dataset: wc-at.csv

**Solution:** For Graph Refer – Jupyter Notebook PDF Attached

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

**Solution:** For Code Refer – Jupyter Notebook PDF Attached

* 90% = 1.64
* 94% = 1.88
* 60% = 0.84

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

**Solution:** For Code Refer – Jupyter Notebook PDF Attached

* 95% = 1.71
* 96% = 2.17
* 99% = 2.79

Q 24**)** A Government 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:** For Code Refer – Jupyter Notebook PDF Attached

Defining Hypothesis: H0 ≤ 260, H1 >260

t\_value is: -0.4714

The P\_value is: 0.3216

Since p\_value > α, Fail to reject H0, bulb have an average life of no more than 260days, No action required.