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

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

**ANS: Three Coins are tossed Simultaneously. Then ,**

**All possible outcomes are 2^3=8**

**i.e., (HHH),(THH),(HTH),(HHT),(TTH),(THT),(HTT),(TTT)**

**So, The Probability of Two Heads and One Tail is =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

**ANS: a) The Probability sum is Equal to 1 is “0”.**

**b) Less than or equal to 4:**

**The total possible outcomes are =1/6^2**

**The possibilities are =P(X<=4)=6/36**

**c) Sum is divisible by 2 and 3:**

**Favourable outcomes are (3,3),(2,4),(4,2),(1,5),(5,1),(6,6)**

**Hence,**

**The Probability Sum is divisible by 2 and 3 is 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?

**ANS: The total number of balls are =7**

**The Probability that none of the balls drawn is Blue =5C2/7C2**

**=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

**ANS: Expectation =∑pixi**

**=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+0.006+0.24**

**=3.036**

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

**ANS: POINTS:**

**Mean=3.59563**

**Median=3.695**

**Mode=3.07,3.92**

**Variance=0.2858814**

**Standard Deviation=0.5346787**

**Range=2.170**

**SCORES:**

**Mean=3.21725**

**Median=3.325**

**Mode=3.44**

**Variance=0.957379**

**Standard Deviation=0.9784574**

**Range=3.911**

**WEIGH:**

**Mean=17.84875**

**Median=17.71**

**Mode=(17.02,18.90)**

**Variance=3.193166**

**Standard Deviation=1.786943**

**Range=8.400**

**Inference: Points and Weigh have Bimodal Values.**

* **Refer “ A1-Ans.ipynb “ file.**

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?

**ANS: The probability of choosing one patient = 1/9**

**Expectation = ∑** **pixi**

**=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]**

**=145.333**

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

Cars speed and distance

Use Q9\_a.csv

**ANS: There are 81 observations and 5 variables**

**speed:**

**Skewness= -0.1175**

**Kurtosis= -0.5089**

**distance:**

**Skewness= 0.8086**

**Kurtosis= 0.4050**

**Inferences:**

**The data of speed is Negatively skewed and kurtosis is “<3”**

**The data of distance is Positively skewed and kurtosis is “>3”.**

SP and Weight (WT)

Use Q9\_b.csv

**ANS: SP:**

**Skewness= 1.6114**

**Kurtosis= 2.978**

**WT:**

**Skewness= -0.614**

**Kurtosis= 0.9502**

**Inferences:**

**The data of SP is Positively skewed and kurtosis is “>3”**

**The data of WT is Negatively skewed and kurtosis is “<3”.**

* **Refer “ A1-Ans.ipynb “ file.**

Q10) Draw inferences about the following boxplot & histogram



**ANS: Inferences:**

**Histogram: The Histogram is peak has right skew**

**And it is Positively Skewed (>3)**

**Boxplot: The Middle line indicates that Median or Inter Quartile Range**

**Above the Median is Upper Quartile (Q1)**

**Below the Median is Lower Quartile (Q3)**

**And the Upper whisker > Lower whisker**

**The boxplot has Outliers on the maximum side.**

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?

**ANS**: **Given That,**

**Mean(µ) = 200 & Standard Deviation=30 Then,**

**Confidence interval for 94%=143.576 , 256.423**

**Confidence interval for 98%=130.209 , 269.790**

**Confidence interval for 96%=138.387, 261.612**

* **Refer “ A1-Ans.ipynb “ file.**

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.

**ANS: Mean=41**

**Median=40.5**

**Variance=25.5294**

**Standard Deviation=5.052**

1. What can we say about the student marks?

**ANS: Since, Mean is greater than Median**

**So, the data is Right Skewed.**

* **Refer “ A1-Ans.ipynb “ file.**

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

**ANS: The data will be Perfectly Symmetric**

**i.e., Skewness=0.**

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

**ANS: The data will be Right Skewed**

**i.e., Skewness will be Positively Skewed.**

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

**ANS: The data will be Left Skewed**

**i.e., Skewness will be Negatively Skewed.**

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

**ANS: Positively Skewed (>3)**

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

**ANS: Negatively Skewed (<3)**

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

**ANS: a) The Boxplot does not follows the Normal Distribution. It has no Outliers.**

1. **The Nature of the Skewness of the data is Left Skewed.**
2. **Inter Quartile Range = (Q3-Q1) = 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.

**ANS: For the Boxplot 1 and Boxplot 2, The Inter Quartile Range is Equal.**

**There are no Outliers for the both Boxplots.**

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)

**ANS: a) P(MPG>38)=0.3475**

**b) P(MPG<40)=0.7293**

**c) P(20<MPG<50)=0.8988**

* **Refer “ A1-Ans.ipynb “ file.**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**ANS: Mean=34.4220**

**Median=35.1527**

**Mode=29.6229**

**We can observe that the mean,median,mode are not equal**

**Therefore, the data of MPG doesn’t follow Normal Distribution.**

* **Refer “A1-Ans.ipynb “ file.**

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

Dataset: wc-at.csv

**ANS:** **Adipose Tissue:**

**Mean=101.8940**

**Median=96.54**

**Mode=121,123**

**Waist** **Circumference**:

**Mean=91.9018**

**Median=90.80**

**Mode=94.5,106,108**

**The data of Adipose Tissue and Waist Circumference does not follows the Normal distribution.**

* **Refer “A1-Ans.ipynb” file.**

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

**Ans: Z score of 90% Confidence Interval is 1.645**

**Z score of 94% confidence Interval is 1.880**

**Z score of 60% confidence Interval is 0.841**

* **Refer “A1-Ans.ipynb” file.**

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

**Ans: For 95% Confidence Interval =2.0638**

**For 96% Confidence Interval =2.1715**

**For 99% Confidence Interval =2.7969**

* **Refer “A1-Ans.ipynb” file.**

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

**Ans: T-Score= 260-270/(90/18^0.5)=-0.47**

**P(X<260) =Stats.t.cdf (-0.47,17) =0.32**

* **Refer “A1-Ans.ipynb” file.**

**Submitted By**

**K.KRISHNAVENI**

**(krishnavenikolipaka1000@gmail.com)**