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

**Name: Jiyashan Imtiyaz Pathan**

**Batch:**  Data Science-Weekday-Pune(Baner)-12th January 2024-02:30 PM to 04:30 PM

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) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Numeric |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Ordinal |

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

Ans:- Sample Space = S

S = {TTT,THT,TTH,HHH,HTH,HHT,THH,HTT}

We have total 8 combinations

From that having two heads and one tail ,we have 3 combinations {HTH,HHT,THH }

Therefore, probability of two heads and one tail = 3/8 = 0.375

Formula: P(event) = number of favorable events **/** total number of possible event

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)

If two dices were rolled, then total possible cases =36

X=Sum

**P(X=1) = 0**

As the minimum sum is 2

b) P(X<=4)=

Numbers of Outcomes = {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}

**P = 6/36 = 1/6 = 0.166**

c)P(Sum is divisible by 2 and 3) =

Numbers of Outcomes = {(1,5),(5,1),(2,4),(3,3),(4,2),(6,6)}

**P = 6/36 = 1/6 = 0.166**

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:- Total number of balls = (2+3+7) = 7

Let S be the Sample Space,

n(S) = Number of Two balls are drawn

=7C2 = (7\*6)/(2\*1) = 42/2 = 21

Let B be the event of drawing no blue balls

N(B) = 2 red +3 green = 5

= 5C2 = (5\*4)/(2\*1) = 20/2 = 10

**P(B) = 10/21 = 0.476**

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:-

Expected Value= ∑Xi\*P(X = Xi)

Expected number of candies for a randomly selected child

=  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

**Expected number of candies for a randomly selected child = 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**

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:- Expected value = Sum(X\*Probability of X)

Probability of 1 patient = 1/9

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**

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



Answer:-

The histogram's peak has a right skew and tail is on right. Mean > Median. We have outliers on the higher side



Answer:-

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?

Answer: -

Given:

Sample mean (x): 200 pounds

Standard deviation (s): 30 pounds

Sample size (n): 2000 men

We'll use the formula for the margin of error (

ME = Z \* s/√n

And then calculate the lower and upper bounds of the confidence intervals:

Lower Bound = x - ME

Upper Bound = X + ME

For a 94% confidence interval:

Z for 94% confidence level is approximately 1.88.

ME = 1.88 \* 30/√2000 = 1.261

Lower Bound = 200 - ME = 200 - 1.261= 198.74

Upper Bound = 200 + ME= 200 + 1.261= 201.26

For a 96% confidence interval:

Z for 96% confidence level is approximately 2.05.

ME = 2.05 \* 30/√2000 = 1.375

Lower Bound = 200 - ME = 200 - 1.375= 198.625

Upper Bound = 200 + ME= 200 + 1.375= 201.375

For a 98% confidence interval:

Z for 98% confidence level is approximately 2.33.

ME = 2.33 \* 30/√2000 = 1.56

Lower Bound = 200 - ME = 200 - 1.56= 198.44

Upper Bound = 200 + ME= 200 + 1.56= 201.56

**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 (Average):

Mean = ∑Scores / Number of Scores

Mean = (34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56) / 18

Mean = 738 / 18 = 41

Median:

Arrange the scores in ascending order: 34, 36, 36, 38, 38, 39, 39, 40, 40, 41, 41, 41, 41, 42, 42, 45, 49, 56.

9th = 40 , 10th= 41

Median = (9th = 10th) / 2 = (40+41) = 40.5

Variance:

Variance = ∑(Score−Mean) ^2 / Number of Scores

Variance = [(34-41)^2+(36-41)^2+(36-41)^2+(38-41)^2+(38-41)^2+(39-41)^2+(39-41)^2+(40- 41)^2+(40-41)^2+(41-41)^2+(41-41)^2+(41-41)^2+(41-41)^2+(42-41)^2+(42-41)^2+(45-41)^2+ (49-41)^2+(56-41)^2] / 18 = 24.111

Standard Deviation:

Standard Deviation = √Variance

Standard Deviation = √24.11 = 4.91

1. What can we say about the student marks?

ANS:-  We don’t have outliers and the data is slightly skewed towards right because mean is greater than median.

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

ANS:-  No skewness is present we have a perfectly symmetrical distribution

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

ANS:- Skewness and tail is towards the Right. It is also called positive skewness

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

ANS:- Skewness and tail is towards the left. It is also called negative skewness

Q16) What does positive kurtosis value indicate for data ?

ANS:- Positive kurtosis means the curve is more peaked and it is Leptokurtic

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

ANS:- Negative Kurtosis means the curve will be flatter and broader

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



What can we say about the distribution of the data?

ANS:- The above Boxplot is not normally distributed the median is towards the higher value

What is nature of skewness of the data?

ANS:- The data is skewed towards the left. The whisker range of minimum value is greater than the maximum

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

ANS:- The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile = 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:-  First there are no outliers. Second, both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

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

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

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