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

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

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

E= {HTH,HHT,THH} 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

(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)

(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)

(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)

(4,1), (4,2), (4,3), (4,4), (4,5), (4,6)

(5,1), (5,2), (5,3), (5,4), (5,5), (5,6)

(6,1), (6,2), (6,3), (6,4), (6,5),(6,6)

1. 0
2. (1,1), (1,2), (1,3)(2,1), (2,2),(3,1)

C) (1,5) (2,4) (3,3) (4,2) (5,1) )(6,6)

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?

TOTAL BALL IN THE BAG=2RED+3GREEN+2BLUE=7

Two balls are drawn at random,n(Two balls are drawn at random=7c2=21

p(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

Soln:

Expexted no of candies = sum of (X\*p(x))

=sum of candiescount\*probability)

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

=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

**Soln:**

**Mean :**

Points 3.596563

Score 3.217250

Weigh 17.848750

dtype: float64

**Median:**

Points 3.695

Score 3.325

Weigh 17.710

dtype: float64

**Varience :**

Points 0.285881

Score 0.957379

Weigh 3.193166

dtype: float64

**Standard Deviation :**

Points 0.534679

Score 0.978457

Weigh 1.786943

dtype: float64

**Mode :**

Points 3.07 and 3.92

Score 3.44

Weigh 17.02 and 18.90

**Range :**

Points 2.17

Score 3.91

Weigh 8.39 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?

**Soln:**

Expected Value of a Discrete Random Variable is actually the mean of the

statements

= Hence Mean is 1308/9

= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Soln:**

**Skewness :**

**Index 0.000000**

**speed -0.117510**

**dist 0.806895**

**dtype: float64**

**Kurtosis :**

**Index -1.200000**

**speed -0.508994**

**dist 0.405053**

**dtype: float6**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Soln. :**

**Skewness :**

**SP 1.611450**

**WT -0.614753**

**dtype: float64**

**Kurtosis :**

**SP 2.977329**

**WT 0.950291**

**dtype: float64**

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



Soln. :

Here we can see that the major Chick weights fall in the category of 50-

100g(measures in x) as the maximum which is 200.The minimum weights have a

frequency if less than or equal to 5.

The plot is Right skewed which show that there is lesser concentration of

chick weights in the 300-400gram category .

The expected value should be above 46.45

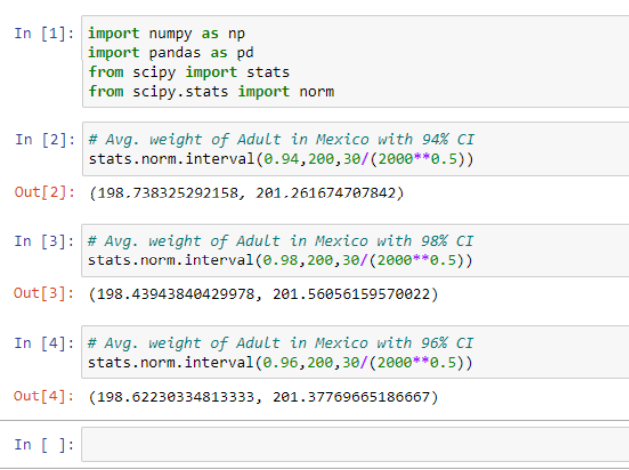
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

Soln:

94% = [ 2004.8, -1995.16]

98% = [2006.39, -1993.6]

96% = [2005.46, -1994.54



**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.

Soln. :

Mean = 41

Median = 40.5

Variance = 25.5294

Standard Deviation = 5.05

1. What can we say about the student marks?

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

Soln:

The skewness will be symmetrical. Hence both the sides of the plot must be

equal in proportion for the data should be normally distributed.

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

Soln:

For data which produces mean > median the skewness will be a +ve

skewness or the data will be right skewed. Most of the data will be lying on the

left side of the plot. Mean always tends to go towards the most skewed part since

skewness influences the mean.

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

soln:

For data which produces median > mean the skewness will be a -ve

skewness or the data will be left skewed. Most of the data will be lying on the

right side of the plot. Mean always tends to go towards the most skewed part

since skewness influences the mean.

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

Soln:

A distribution with a positive kurtosis value indicates that the distribution

has heavier tails than the normal distribution. For example, data that follow a t

distribution have a positive kurtosis

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

Soln:

A distribution with a negative kurtosis value indicates that the distribution

has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

Soln. :

The data is not actually equally distributed across the plane. There might be

outliers influencing the data. Median of the data is 14.7(appx)

25 percent of the data lies between 0-10

50 percent of the data lies between 10-18

25 percent of the data lies after 18-20 approx.

What is nature of skewness of the data?

Soln. :

The data will be left skewed since whisker length on the upper quadrant is

higher than the data on the lower quadrant. Median will be greater than the

mean since data is left skewed.

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

IQR is the inter quartile range.

Here Q1 = 10

Q2 = 14.7

Q3 = 18

IQR = Q3 – Q1 = 8(approx

Q19) Comment on the below Boxplot visualizations?



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

Soln. :

Here when we compare box plot 1 with box plot 2 we can say that the data

in box plot 1 is widely spread. Here the main inference is that since the data range

varies high in box plot 2 it is hard to make a prediction in box plot 2. The median

in the 2 box plots are equal. And the data spread in both of them are symmetrical

Q 20) Calculate probability from the given dataset for the below cases

Soln. :

a. P(MPG>38) = 0.3475939251582705

b. P(MPG<40) = 0.7293498762151616

c. P (20<MPG<50) = 1.2430968797327613e-05

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

Soln. :

Z scores of 90% confidence interval = 1.6448536269514722

Z scores of 94% confidence interval = 1.8807936081512509

Z scores of 60% confidence interval = 0.8416212335729143

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

Soln. :

t-SCORE CALCULATION

T((1,alpha),(n-1))

Here n = 25

n-1 = 24

Hence t score values will be:

95% = qt(0.975,24) = 2.063899

96% = qt(0.98,24) = 2.171545

99% = qt(0.995,24) = 2.79694

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

Soln. :

Sample size = 18 = n

Sample mean = 260 days = x

Sample standard deviation = s = 90days

t-score = -0.4714

P(X>= 260) = 0.32167