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

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 | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinaldl |
| 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?

|  |
| --- |
| HHH, HTH, HHT, HTT, THT, THH, TTH, TTT |
| P= EXPECTED OUTCOME / TOTAL OUTCOMES (3/8) |
|  |
| Answer : 3/8 |



Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1 is Zero(0) - because lowest possible sum is 2
2. Less than or equal to 4 is 1/6
3. Sum is divisible by 2 and 3 is 1/36

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?

Answer: 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

Answer : 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?

Answer : 145.4

**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**



**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?

Answers : 94% = [22, 38]

98% = [20, 40]

96% = [21, 39]

**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.

df1.marks.mean()

Out[68]:

41.0

In [70]:

df1.marks.median()

Out[70]:

40.5

In [71]:

df1.marks.var()

Out[71]:

25.529411764705884

In [72]:

df1.marks.std()

Out[72]:

5.05266382858645

1. What can we say about the student marks?

Average marks of students is 40.5

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

Normal distribution or a symmetrical distribution or zero Skew

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

Positive Skewed or Right Skewed

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

Negative Skewed or left Skewed

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

 More peaked than normal Distribution.

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

Wider peaked and thinner tails

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



What can we say about the distribution of the data?

Most of the data points lies between 10 to 18.2

Mean of this data set is 15.2 (approximately)

What is nature of skewness of the data?

Negative Skewed or left Skewed.

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

IQR = 18.2 -10 = 8.2

Q19) Comment on the below Boxplot visualizations?



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

Boxplot 1 most of the data points lies between Q1 254 to Q3 279

Boxplot 2 most of the data points lies between Q1 225 to Q3 309

Median of Boxplot 1 and Boxplot 2 is same 261

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

Yes, it is 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

No, it is not Normal Distributed

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

81.59% confidence interval

82.64% confidence interval

72.57% confidence interval

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

82.42% t scores confidence interval

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

Ho: x>=260 fail to reject null

Ha: x<260

inter **=** (260**-**270)**/**(90**/**np.sqrt(18))

In [77]:

inter

Out[77]:

-0.4714045207910317

In [75]:

stats.t.cdf(inter, df**=**17)

Out[75]:

0.32167253567098364

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom