|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete Data |
| Results of rolling a dice | Discrete Data |
| Weight of a person | Continous Data |
| Weight of Gold | Continous Data |
| Distance between two places | Continous Data |
| Length of a leaf | Continous Data |
| Dog's weight | Continous Data |
| Blue Color | Nominal Data |
| Number of kids | Discrete Data |
| Number of tickets in Indian railways | Discrete Data |
| Number of times married | Discrete Data |
| Gender (Male or Female) | Nominal Data |

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

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

Answer: Total Possible outcomes:8; Total no of fav outcomes=3

P=3/8=0.375

The probability of getting two heads and one tail is 37.5%

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

Answer: a) Equal to 1=0

b)less than or equal to 4=6/36=1/6

c)sum is divisible by 2 and 3=5/36

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: Red ball=2

Green ball=3

Blue ball =2

The probability that none of the balls drawn is blue: 2+3+2=7

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: Expected number of candies=(1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)

(6\*0.01)+(2\*0.120)

=0.015+0.80+1.95+0.025+0.06+0.24

Expected number of candies=3.14

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**

|  |  |  |  |
| --- | --- | --- | --- |
| Column | Points | Score | weight |
| Mean | 3.596562 | 3.21724 | 17.848750 |
| Median | 3.6950 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | **16.46** |
| Variance | 0.2858813 | 0.957378 | 3.193166 |
| STD | 0.534678 | 0.978457 | 1.786943 |
| Min | 2.76 | 1.513 | 14.5 |
| Max | 4.93 | 5.424 | 22.9 |
| Range | **2.17** | **3.911** | **8.4** |

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:145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **Speed** | -0.11 | -0.50 |
| **Distance** | 0.80 | 0.40 |

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **Speed** | 1.61 | 2.97 |
| **Weight** | -0.61 | 0.95 |

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



Answer:

**. Histogram of ChickWeight$weight is Right tail**

**. Histogram of chickweight$weight is positive Skewed**

**.The range of 50-100 are most number of chick weight present**



Answer:

.It has many outliers and positively skewed in given box plot

**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?

|  |  |  |  |
| --- | --- | --- | --- |
|  | 94% | 98% | 96% |
| Upper | 201.04 | 201.38 | 201.17 |
| Lower | 198.96 | 198.62 | 198.83 |

**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.

Mean:41

Median:40.5

Standard Deviation:5.05

Variance:25.53

1. What can we say about the student marks?

.most of the student marks are lies between the range of 35-45.

.most of the student got mark below 41

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

Answer: skewness is zero

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

Answer: skewed positively

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

Answer: skewed Negatively

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

Answer: Distribution of positive kurtosis value indicate for a data is

Peaked in center part.

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

Answer: Distribution of negative kurtosis value indicate for a data is wider or

Flatter In center .

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



What can we say about the distribution of the data?

Answer: the given boxplot visualization of the distribution is left tail

And given data are lies on right side.

What is nature of skewness of the data?

Answer: skewed Negatively.

What will be the IQR of the data (approximately)?   
 Answer: IQR=UQ-LQ

18-10=8(approximately).

Q19) Comment on the below Boxplot visualizations?



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

* Box plot 1 and box plot 2 are equal median 261(app)
* Both are normally distribution
* 1st box plot lower than 2nd box plot

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)

Answer:0.40740

* 1. P(MPG<40)

Answer:0.75308

* 1. P (20<MPG<50)

Answer:0.85185

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Anwer: MPG 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

Answer: bimodal Distribution

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

Answer:

|  |  |  |
| --- | --- | --- |
| Confidence interval | Alpha a | Z score |
| 94% | 0.03 | 1.880 |
| 90% | 0.05 | 1.664 |
| 60% | 0.02 | 0.841 |

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

Answer

|  |  |
| --- | --- |
| 95% | ±2.060 |
| 96% | ±2.167 |
| 99% | ±2.787 |

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

Answer:

t\_score:-0.4714

Degree of freedom(df):17

P(t\_score)=0.3216