|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Descrete |
| Results of rolling a dice | Descrete |
| 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 | Descrete |
| Number of kids | Descrete |
| Number of tickets in Indian railways | Descrete |
| Number of times married | Descrete |
| Gender (Male or Female) | Descrete |

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 | Ordinal |
| 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 | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Nominal |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Answer :{HHH,THH,TTT,HHT,HTT,HTH,THT}

Total possible events:8

No of desired events:3

Probability:3/8=0.375

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 2and 3

Answer-(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)

a)0/36=0

b)3/36=0.167

c)5/36=0.139

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:Total number of balls=7

n(S)= Probability of drawing 2 balls out of 7

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

N(E)=ways of drawing 2 balls other than blue

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

Then

P(E)=n(E)/n(S)

=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:-the expected number of candies for 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**

Answer:-

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | MEAN | MEDIAN | MODE | VARIANCE | Std. Dev | RNGE |
| Points | 3.59 | 3.70 | 3.92 | 0.29 | 0.53 | 2.17 |
| Score | 3.22 | 3.33 | 3.44 | 0.96 | 0.98 | 3.91 |
| Weigh | 17.85 | 17.71 | 17.02 | 3.19 | 1.79 | 8.40 |

* No case has the variable Mean = Median = Mode
* Thus as seen in the graph dataset “score” and “Weigh” has outliers

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:-

Probabilty of selecting each patient=1/9

The expected weight of patient

=(1/9)\*(108+110+123+134+135+145+167+187+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**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| SP | 1.55 | 5.72 |
| WT | -0.59 | 3.82 |

* “SP” is positively skewed where as “WT” is negatively skewed

Thus SP has distribution of data concentrated on the left whereas WT has distribution on the right. As seen in the graph

* Both WT and SP has positive Kurtosis

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



Answer:- **Histogram**:

a)Chick weight data is positively skewed.

b)More than 50% Chick Weight is between 50 to 150.

c)Most of the chick weight is between 50 to 100.



a)Data has outliers

c)Data has outliers on top

b)Data is positively skewed

**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:-

|  |  |  |  |
| --- | --- | --- | --- |
|  | 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.

Answer:-

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.50 |
| Variance | 25.53 |
| Std Deviation | 5.05 |

2)What can we say about the student marks?

Answer:- a) Data has outlier

b)Majority of the students scored between 35 – 45 Marks

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:-Positively skewed

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

Answer:-Negativley skewed

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

Answer:-High and narrow peak on central part of the data

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

Answer:- Wider peak on central part of the data

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



What can we say about the distribution of the data?

Answer:- = Data is not symmetric

What is nature of skewness of the data?

Answer:-The data is negatively skewed

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

Answer:-IQR of the Data is 8

Q19) Comment on the below Boxplot visualizations?



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

Answer:-

1. Data is Normally Distributed.
2. No Outliers.
3. Center around 262.5.
4. Comparatively, first graph has less range and second graph has high range

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

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

Answer:- = 0.8518519

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer:-MPG of cars is normally distributed

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

Dataset: wc-at.csv

Answer:- = Both AT and Waist doesn’t follow Normal Distribution

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

Answer:-

|  |  |
| --- | --- |
| 90% | ±1.711 |
| 94% | ±1.828 |
| 60% | ±2.492 |

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 companyclaims 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 = -0.4717

Degree of freedom = 17

P(t) = 0.3216725