|  |  |
| --- | --- |
| 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 | Discrete |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical |

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 | Ordinal |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Nominal |
| SAT Scores | Ordinal |
| Years of Education | Ratio |

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

Ans: **Probability of event to happen P(E) = Number of favorable outcomes/Total Number of outcomes**

**P(E) = 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

Ans: a) probability is 0 (because when dice are rolled the minimum sum would be 2 not 0)

b) P(Less than or equal to 4) = 6/36 = 1/6 =0.167

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

Ans: Total Outcome = 7C2 = 21

Total number of favourable event = 5C2 = 10

Hence, P = 10/21 =0.4761

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: By Multiplying probabilities of all candies = 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**

Ans:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | Standard Deviation | Range |
| Points | 3.59 | 3.70 | 3.92 | 0.29 | 0.53 | 2.17 |
| Score | 3.21 | 3.33 | 3.44 | 0.95 | 0.98 | 3.91 |
| Weigh | 17.85 | 17.71 | 17.02 | 3.19 | 1.79 | 8.40 |

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 = probability\*value)=1308/9=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 | -0.57 |
| **distance** | 0.78 | 0.24 |

**As I can see by using distplot, the skewness of distance is positively skewed moreover I can see that the kurtosis of distance is positive which means the curve is peaked(leptokurtic), coming to the speed the skewness is negatively skewed and the kurtosis is negative indicates platykurtic**

**SP and Weight(WT)**

|  |  |  |
| --- | --- | --- |
|  | **skewness** | **kurtosis** |
| **SP** | **1.58** | **2.72** |
| **WT** | **-0.60** | **0.81** |

**Using distplot(seaborn library), SP is positively skewed whereas WT is negatively skewed, the kurtosis for both SP and WT is positive**

**Use Q9\_b.csv**

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



`**1)** From above histogram I observed that majority of data concentrated at the left, the Chickweight mostly lied in between 50-100, from 100 the frequencies gradually decreasing with the increased chickweight

**2)** By the histogram I observed that the graph is positively skewed



1. Boxplot has outliers
2. It is positively skewed because the median is at the bottom of data where median is less than mean

**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?

n = 2000

mean of sample = 200

standard deviation = 30

|  |  |  |  |
| --- | --- | --- | --- |
|  | 94% | 98% | 96% |
| **Upperlimit** | 201.3 | 201.56 | 201.38 |
| **Lowerlimit** | 198.69 | 198.43 | 198.62 |

**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 |
| variance | 25.5 |
| Standard deviation | 5.05 |

1. What can we say about the student marks?

* Majority have scored between 35-45
* It has outliers and it is not normally distributed

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

When mean,median are equal then the skewness = 0

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

If mean is greater than median then the curve is positively skewed

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

If median is greater than mean then the curve is negatively skewed

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

Positive kurtosis value indicates the high and narrow peak of data

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

Negative kurtosis indicates wider peak of the data

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



What can we say about the distribution of the data?

We can say that the distribution of data is not symmetry and it is concentrated on right side

What is nature of skewness of the data?

Negatively skewed

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

IQR = Q3-Q1= 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.

* Boxplot 1 and Boxplot 2 are symmetrically distributed and it has no outliers
* Both boxplots have similar centre 262.5

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)

33/81=0.4074

* 1. P(MPG<40)

0.7530864

* 1. P (20<MPG<50)

0.8518519

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Follows 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

Both AT and Waist doesn’t follow Normal Distribution

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

|  |  |
| --- | --- |
| **90%** | ±1.645 |
| **94%** | ±1.89 |
| **60%** | ±0**.**75 |

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

V = n-1 = 25-1=24

Alpha/2 = (1-cl)/2

|  |  |
| --- | --- |
| **95%** | ±2.064 |
| **96%** | ±2.493 |
| **99%** | ±2.797 |

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

t-score = -0.4714,  
Degree of freedom = 17   
P(t) = 0.3216725