|  |  |
| --- | --- |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| 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 | Interval |
| SAT Scores | Ratio |
| Years of Education | Interval |

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

Total Possible Events: 8

No: of Desired Events: 3

P=3/8 = 0.375

**37.5%** is the chance of getting 2 heads and one tail

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:

1. Equal to 1= **0**
2. Less than or equal to 4 = 6/36 = **0.167**
3. Sum is divisible by 2 and 3 = 6/36 = **0.167**

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?

No: of Red Balls: 2

No: of Green Balls: 3

No: of Blue Balls: 2

Total Number Of Balls = 7

No of ways Getting 2 Balls from 7 = 7C2

No of ways Getting Not 2 Blue balls = 5C2

Probability of getting not 2 Blue balls = 5C2/7C2=20/42

P = **0.476**

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

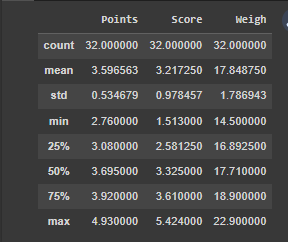
ANS:**3.09**

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**

****

**Variance:**

****

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: **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** | **2.42** |
| Distance | **0.76** | **3.25** |

* “dist” is **positively Skewed** whereas “speed” is **negatively skewed**
* Thus dist has distribution of data concentrated on left whereas speed has distribution on the right as Seen in that graph.
* Both dist and speed has positive Kurtosis

**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**



* Majority of the Chicks weight in range 50-100,followed by 100-150,followed by 150-200
* The data provided is Positively Skewed
* The chicks can be categorized into 3 categories
* (Chicks<50-Under weight)
* (Chicks between 50 and 150-Average weight)
* (Chicks >150-Over weight)



ANS:

* This Data has Outliers
* 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?

|  |  |  |  |
| --- | --- | --- | --- |
|  | **94%** | **96%** | **98%** |
| **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 | 40.50 |
| Variance | 25.53 |
| Std. Deviation | 5.05 |

1. What can we say about the student marks?

* Most of the students got marks between 35-45
* Data has Outliers
* Most of the students got scores below 41

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

Zero Skewness.

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

Positive Skewness

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

Negative Skewness

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

Distribution of the data is peaked in Center part.

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

Distribution of data in wider in Center Part.

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



What can we say about the distribution of the data?

The Data is not symmetric. Data is more concentrated on right side

What is nature of skewness of the data?

Negative Skewness

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

IQR(Inter Quartile Region) = UQ(Upper Quartile)-LQ(Lower Quartile)

IQR = 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.

* Both the plots are **Normally Distributed**
* Both the plots are having same **Median** around **262.5(**approx.)
* Comparing to **first plot,Second plot is having more range**

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)

ANS: 0.40740

* 1. P(MPG<40)

ANS: 0.75308

* 1. P (20<MPG<50)

ANS: 0.85158

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

MPG has 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,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

|  |  |  |
| --- | --- | --- |
|  | Alpha a | Z-Score |
| 60% | 0.20 | 0.253 |
| 90% | 0.05 | 1.555 |
| 94% | 0.03 | 1.645 |

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

|  |  |
| --- | --- |
| 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

t-score:-0.4714

Degree of Freedom(df):17

P(t): 0.3216