|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Integer (numerical data) |
| Results of rolling a dice | Integer Numerical data |
| Weight of a person | Continuous data |
| Weight of Gold | Continuous data |
| Distance between two places | Continuous data |
| Length of a leaf | Continuous data |
| Dog's weight | Continuous data |
| Blue Color | Categorical data |
| Number of kids | Discrete data |
| Number of tickets in Indian railways | Discrete data |
| Number of times married | Discrete data |
| Gender (Male or Female) | Categorical 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 |
| Hair Color | Nominal data |
| Socioeconomic Status | Ordinal data |
| Fahrenheit Temperature | Interval data |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal data |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

**Ans: 3/8**

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

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

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?

= 6/7

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

**ANS: 1x0.015+4x0.2+3x0.65+5x0.005+6x0.01+2x0.12=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.

Ans:

|  |  |  |  |
| --- | --- | --- | --- |
|  | point | score | WEIGHT |
| mean | 3.597 | 3.217 | 17.849 |
| median | 3.695 | 3.325 | 17.71 |
| mode | 3.92 | 3.44 | 17.02 |
| varience | 0.276948 | 0.927461 | 3.09338 |
| s.d. | 0.526258 | 0.963048 | 1.758801 |
| range | 2.17 | 3.911 | 8.4 |

**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?

**Probability of selecting on patients from 9 =1/9**

**Expected value=1/9(108+ 110+ 123+ 134+ 135+ 145+ 167+ 187+ 199)**

**=145.3**

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

**Cars speed and distance**

|  |  |  |
| --- | --- | --- |
|  | speed | distance |
| skewness | -0.11395477 | 0.782483517 |
| kurtosis | -0.50899442 | 0.405052582 |

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| skewness | 1.581454 | -0.60331 |
| kurtosis | 2.977329 | 0.950291 |

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



* **Distribution is having positive skewness.**



* **Distribution is positively skewled.**
* **Outlier are present in data**

**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?

**For 94% = 198.95-201.4 pound**

**For 98% =198.62-201.37 pound**

**For 96% = 198.82 -201.174 pound**

**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 |
| mode | 41 |
| variance | 24.11111 |
| s.d. | 4.910307 |

1. What can we say about the student marks?

**Students score is normally distributed.**

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

**=Data is normally distributed.**

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

**=distribution is positively skewed.**

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

**=distribution is negatively skewed**

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

**Tail end of positive kurtosis are flat both side and high frequency having outlier present in data**

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

**Tail end of negative kurtosis are thin both side and low frequency having outlier present in data**

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



What can we say about the distribution of the data?

**=Distribution is negatively skewed.**

What is nature of skewness of the data?

**=negative skewed data**

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.

**IQR of second data is more although both the data look similar and normally distributed. But variance for the first data is small as compared to second data set.**

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) = 0.3475**
  2. **P(MPG<40) = 0.7293**

**c. P(20<MPG<50)= 0.8988**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**MPG 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

**AT column having positive skewness.**

**Waist having jumped up data**

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

**90% confidence interval = 1.64 (TWO TAIL) ±1.28( ONE TAIL)**

**94% confidence interval = 1.88 (TWO TAIL) ±1.55(ONE TAIL)**

**60% confidence interval =0.841 (TWO TAIL) ±0.25(ONE TAIL)**

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

**95% confidence interval = 2.06 (TWO TAIL) ±1.71 ( ONE TAIL)**

**96% confidence interval = 2.06 (TWO TAIL) ±1.82(ONE TAIL)**

**99% confidence interval =2.79 (TWO TAIL) ±2.49(ONE TAIL)**

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

**0.32**