|  |  |
| --- | --- |
| 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 | discrete |
| Number of kids | discrete |
| Number of tickets in Indian railways | discrete |
| Number of times married | discrete |
| Gender (Male or Female) | discrete |

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 | Ratio |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| 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 : Total Possible Events : 8 , No of desired events: 3

P = 3/8

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

1. Equal to 1- Ans : 0
2. Less than or equal to 4- Ans : 1/6
3. Sum is divisible by 2 and 3- Ans:6/36

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?

Ans : 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

Ans : 3.09

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.07 , 3.92 | 3.44 | 17.02 , 18.90 |
| Variance | 0.2858814 | 0.957379 | 3.193166 |
| Standard Deviation | 0.5346787 | 0.9784574 | 1.786943 |
| Range | 2.76 4.93 | 1.513 5.424 | 14.5 22.9 |

**Mean**

Average value of the column

Sum of the values/number of values

**Median**

First order the numbers from smallest to largest. Then median is the middle number.

**Mode**

The mode is the most **common number** in a set

**Variance**

The **variance**is a measure of [variability](https://www.scribbr.com/statistics/variability/). Variance tells you the degree of spread in your data set.

**Standard Deviation**

The [standard deviation](https://www.scribbr.com/statistics/standard-deviation/) is derived from variance and tells you, on average, how far each value lies from the mean. It’s the square root of variance.

**Range**

The **range** is the spread of your data from the lowest to the highest value in the distribution. A large range means high variability, a small range means low variability in a distribution.

**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: 145.333**

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

|  |  |  |
| --- | --- | --- |
|  | **Cars speed** | **Distance** |
| **Skewness** | -0.1139548 | 0.7824835 |
| **Kurtosis** | 2.422853 | 3.248019 |

**SP &WT**

|  |  |  |
| --- | --- | --- |
|  | **SP** | **WT** |
| **Skewness** | 1.581454 | -0.6033099 |
| **Kurtosis** | 5.723521 | 3.819466 |

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



* It is a right skewed histogram- Most of the data are accumulated on the left side.
* Skewness will be a negative value.
* For this plot mean > median



* The given boxplot is a right skewed plot.
* There are 7 outlier values.
* Lower whisker range is lesser than upper whisker range.

**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?

confidence interval 94% (143.5761, 256.4238)

96% (138.3875, 261.6124)

98% (130.2095, 269.7904)

**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 | 25.52941 |
| Standard deviation | 5.052664 |

1. **What can we say about the student marks?**

Ans: The given set of data has a high range of variability.

Most of the students got mark in the range of 38 to 42. The student who got 34 marks and 56 marks are the lower and upper extremes.

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

Skewness=0, plot will be perfect symmetrical

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

Right skewed, skewness will be positive

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

Left skewed, skewness will be negative

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

The plot will have a sharp peak.

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

The plot will have a flat peak.

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



**What can we say about the distribution of the data?**

The range of left whisker is bigger than the range of right whisker value.

**What is nature of skewness of the data?**

Right skewed data.

**What will be the IQR of the data (approximately)?**IQR = Q1-Q3

IQR = 8 (approx.)

**Q19) Comment on the below Boxplot visualizations?**



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

* Lower extreme of 1st boxplot is 237.5 which is greater than the Lower extreme of 2nd boxplot 200.
* Upper extreme of 1st boxplot is 287.5 which is lesser than the Upper extreme of 2nd boxplot,350.
* IQR value of 1st boxplot is lower than that of 2nd boxplot.
* Median value of both boxplots are equal , which is 225.5

**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

P(MPG>38) = 0.6524

P(MPG<40) = 0.7293

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

We can say the given data follows Normal distribution only if it obeys the following criteria

* Mean = 0
* Kurtosis = 3
* Skewness = 0
* Standard deviation = 1

For MPG of Cars

* Mean = 34.4220
* Kurtosis = -0.6116
* Skewness = -0.1779
* Standard deviation = 9.1314

**Hence we can say that the MPG of Cars does not follow 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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Mean | Skewness | Kurtosis | Standard Deviation |
| AT | 101.8940 | 0.5848 | -0.2855 | 57.2947 |
| Waist | 91.9018 | 0.1340 | -1.1026 | 3.5591 |

We can say the given data follows Normal distribution only if it obeys the following criteria

* Mean = 0
* Kurtosis = 3
* Skewness = 0
* Standard deviation = 1

As data in both columns AT and Waist does not follow the above criteria, we can say those **do not follow Normal Distribution**

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

Z Score for 90% confidence interval =1.6448

94% confidence interval =1.8807

60% confidence interval = 0.8416

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

T Score for 95% confidence interval = 2.063

96% confidence interval = 2.171

99% confidence interval = 2.796

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

Ans : t-score = -0.4714, and Degree of freedom = 17

P(t) = 0.3216725