|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete (Quantitative) |
| Results of rolling a dice | Ordinal (Qualitative) |
| Weight of a person | Continuous (Quantitative) |
| Weight of Gold | Continuous (Quantitative) |
| Distance between two places | Continuous (Quantitative) |
| Length of a leaf | Continuous (Quantitative) |
| Dog's weight | Continuous (Quantitative) |
| Blue Color | Nominal (Qualitative) |
| Number of kids | Discrete (Quantitative) |
| Number of tickets in Indian railways | Discrete (Quantitative) |
| Number of times married | Ordinal (Qualitative) |
| Gender (Male or Female) | Nominal (Qualitative) |

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 | Interval |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ordinal |
| Years of Education | Interval |

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

Ans)

Total possible outcomes: 8

(HHH,TTT,HHT,HTH,THH,TTH,THT,HTT)

Outcomes with 2H and 1T: 3

Therefore, P(2 heads and 1 tail) = 3/8 = 0.375

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

1. Equal to 1 : Not possible
2. Less than or equal to 4 : 0.167
3. Sum is divisible by 2 and 3 : 0.056

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?

Sol)

P= no of ways to draw two non blue balls / no of ways to draw any two balls

= 5C2/7C2 =[ (5x4)/(2x1)]/ [ (7x6)/(2x1)] = 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

Sol)

Expected number of candies = (1 x 0.015) + (4 x 0.20) + (3 x 0.65) + (5 x 0.005) + (6 x 0.01) + (2 x 0.120)

Therefore, the expected number of candies for a randomly selected child is 3.095

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**

**Sol)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **mean** | **Median** | **Mode** | **var** | **std** | **Range** |
| **Points** | **3.596563** | **3.695** | **3.07** | **0.285881** | **0.534679** | **2.76-4.93** |
| **Score** | **3.21725** | **3.325** | **3.44** | **0.957379** | **0.978457** | **1.513-5.424** |
| **Weigh** | **17.84875** | **17.71** | **17.02** | **3.193166** | **1.786943** | **14.5-22.9** |

**Points : Left Skewed : Lot of vehicles have less points.**

**Score : Left Skewed : Mean<median : More vehicles have scored lesser.**

**Weight: Right Skewed : There are few vehicles with very less weight.**

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?

Sol: It would be average of all the patients: 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**

**Sol) Q9\_a.csv :**

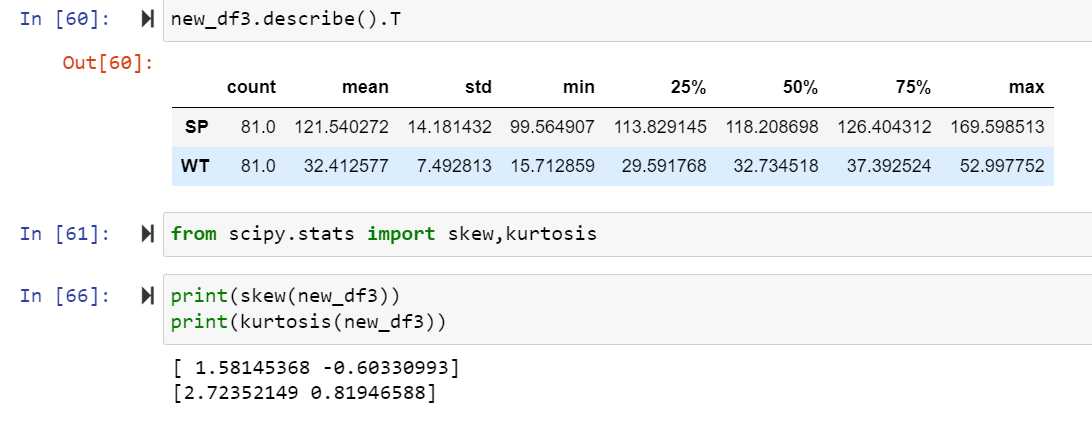
**Text

Description automatically generated**

**There are outliers in distance data as max value is 120 and skewness is positive.**

**Which means some cars are travelling more distance on same speed.**

**Q9\_b.csv**

****

**SP has positive kurtosis and skewness. WT has negative skewness and positive kurtosis.**

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



Sol)

Histogram:

Almost 200 chickweight is between 50 to 100

Almost 100 chickweight is between 0 to 50

And the number 200 gradually decreases when we move towards 400 on interval of 50.

Frequency is very less in range 350- 400 so we can consider it as outliers.

Box Plot:

Data is not unifomly distributed and it is left skewed and there are a lot of outliers.

**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?

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

Sol)

Graphical user interface, text, application, email

Description automatically generated

Standard deviation is less which means that data is clustered around the mean which is 41 marks. No outliers, no one got very less or very high marks.

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

Sol) Skewness will be zero as the data is symmetric.

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

Sol) Skewness is positive if mean is greater than median. The distribution is right skewed, and tail is towards right.

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

Sol) Skewness is negative if median is greater than mean. The distribution is left skewed, and tail is towards left.

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

Sol) Positive kurtosis means data is peaked and distribution has thick tails.

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

Sol) Negative kurtosis means data is not peaked and distribution has thin tails.

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



What can we say about the distribution of the data?

Sol) Data is not distributed normally and there might be outliers towards the left.

What is the nature of skewness of the data?

Sol) Data is left skewed. (negative)

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

Sol) 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.

Sol) Data is less spread and IQR is also less.

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)
  2. P(MPG<40)
  3. P(20<MPG<50)

Sol)

Graphical user interface, text, application, email

Description automatically generated

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Sol) Yes.

Chart, line chart

Description automatically generated

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

Dataset: wc-at.csv

Sol) Yes, both follow normal distribution.

Chart, line chart

Description automatically generated Chart

Description automatically generated

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

Graphical user interface, text, application, email

Description automatically generated

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

Graphical user interface, text, application

Description automatically generated

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 = 18-1= 17

t\_score = (260-270) / (90 / √18) = -1.39

Graphical user interface, text, application, email

Description automatically generated