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

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

**Ans:** Total number of outcomes = 2^3=8

Total combinations are: {HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}

Having two heads and one tail are = {HHT, HTH, TTH} = 3 numbers

Probability of two heads and one tail = 3/8 = 0.375 = 37.5%

**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:** Total number of outcomes are = 6^2=36

The possible sums of two dice rolls are (2,3,4,5,6,7,8,9,10,11,12)

1. 0% Probability of that sum is equal to 1.
2. There are 3 possible number of combinations possible when two dies rolled the sum is less than or equal to 4. Probability=3/36=0.08=8.33%
3. The only two sum’s (6 and 12) are divisible by 2 and 3. The 6 sum having 5 possible ways and 12 sum having 1 possible way to get, then total possible ways to get the sum which is divisible by 2and 3. Probability=6/36=0.1666=16.6%

**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?**

**Ans:** Two balls are drawn every time randomly

Total number of balls are 2+3+2 = 7

Number of balls without blue colour = 2+3 =5

Number of ways of drawing 2 balls out of total balls= 7C2 = 21

Number of ways of drawing 2 balls out of total balls without blue balls = 5C2 = 10

Probability that none of the balls drawn is blue= 10/21=47.6%

**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:**

|  |  |  |  |
| --- | --- | --- | --- |
| CHILD | Candies count | Probability | Total Number of Candies |
| A | 1 | 0.015 | 0.015 |
| B | 4 | 0.2 | 0.8 |
| C | 3 | 0.65 | 1.95 |
| D | 5 | 0.005 | 0.025 |
| E | 6 | 0.01 | 0.06 |
| F | 2 | 0.12 | 0.24 |
| **Total** |  |  | **3.09** |

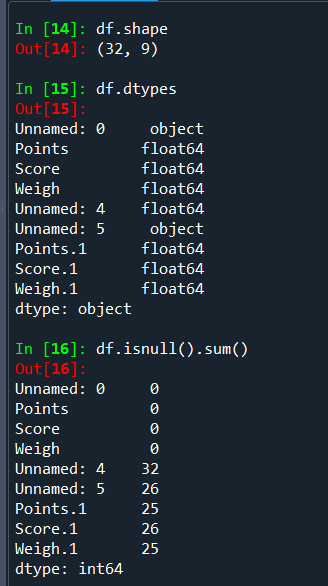
**Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset**

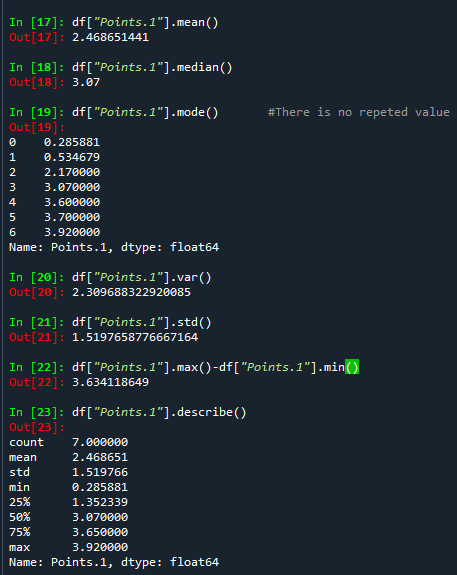
* For Points, Score, Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**Ans:**

****

****

**A screen shot of a graph

Description automatically generated with medium confidence**

**A picture containing text, screenshot

Description automatically generated**

**A screen shot of a graph

Description automatically generated with medium confidence**

**A picture containing text, screenshot

Description automatically generated**

**A screen shot of a graph

Description automatically generated with medium confidence**

**A screenshot of a computer screen

Description automatically generated with low confidence**

**A screen shot of a graph

Description automatically generated with medium confidence**

**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:** in python Spyder

A screenshot of a computer

Description automatically generated with medium confidence

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:**

**A screen shot of a graph

Description automatically generated with medium confidence**

**A picture containing text, screenshot, font

Description automatically generated**

**SP and Weight(WT)**

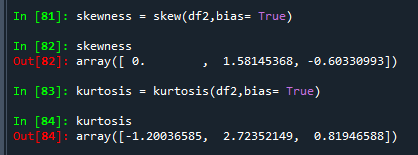
**Use Q9\_b.csv**

**A screen shot of a graph

Description automatically generated with medium confidence**

**A screen shot of a graph

Description automatically generated with medium confidence**

****

SP and Weight(WT)

Use Q9\_b.csv

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



**Ans:**

The histogram shows that the data is Positive skewed (Mode<median<mean)



A: From the above box plot we conclude that the total outliers are on top side

**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?**

**Ans:** Answer in python Spyder

A screen shot of a computer program

Description automatically generated with low confidence

**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?

**Ans:** Answer in python Spyder

A screen shot of a computer

Description automatically generated with medium confidence

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

**Ans**: Data is in symmetric shape and skewness is Zero

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

Ans: Data is not in symmetric shape and skewness is on right side so it is called as negative skewness

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

Ans: Data is not in symmetric shape and skewness is on left side so it is called as positive skewness

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

Ans: It indicates thinner peak and wider tails.

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

**Ans:**  It indicates thinner peak and wider tails.

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



What can we say about the distribution of the data?

Ans: Data is not in symmetric and negative skewness

What is nature of skewness of the data?

Ans: Negative skewness

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

Ans: 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.

Ans:

* Box plot-2 having high spread in data and whisker lengths compared with Box Plot-1.
* Both the boxes having same median

**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)

**Ans:** in python Spyder

A screenshot of a computer program

Description automatically generated with medium confidence

**Q 21) Check whether the data follows normal distribution**

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

A screenshot of a computer program

Description automatically generated with medium confidence

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

Dataset: wc-at.csv

A screen shot of a computer program

Description automatically generated with low confidence

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

**Ans:** Answer in python Spyder

A screen shot of a computer program

Description automatically generated with low confidence

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

**Ans:**

A screenshot of a computer program

Description automatically generated with medium confidence

**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:** Answer in python Spyder

A screenshot of a computer program

Description automatically generated with medium confidence