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

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

Ans: Number of Coins = 3

Probability = Number of favorable outcomes / Total number of Outcomes

= 3 / 2\*2\*2

= 3 / 8

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

1. **Equal to 1**

Ans: we have two dice so it cannot be equal to 1, it’s a false statement so the answer is Zero.

1. **Less than or equal to 4**

Ans: Total outcome = 36

Favorable outcomes = (1,1) (1,2) (1,3) (2,1) (2,2) (3,1)

Number of favorable outcomes = 6

Probability = 6 / 36

= 1 / 6

1. **Sum is divisible by 2 and 3**

Ans: **Total number of possible outcome = 36**

Favorable outcomes = sum is divisible by 2 and 3

Sum should be divisible by both 2 and 3

Favorable outcomes = (1 , 5) , (3 , 3) , (4 , 2) , (5 , 1) , (6 , 6)

Number of favorable outcomes = 5

Probability = 5 / 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:

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

Number of ways of drawing 2 balls out of 7 = 7C2 = (7 × 6) / (2 × 1) = 42/2 = 21 Number of balls other than blue = 5

Number of ways of drawing 2 balls out of 5 = 5C2 = (5 × 4) / (2 × 1) = 20/2 = 10

 Required Probability = 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:** Expected number of candies for a randomly selected child

=  1 \* 0.015  + 4\*0.20  + 3 \*0.65  + 5\*0.005  + 6 \*0.01  + 2 \* 0.120

= 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24

= 3.090

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:** By using the Q7.csv file in Python the value of

**Mean** for ‘Points’ = 3.596563

‘Score’ = 3.217250

‘Weigh’ = 17.848750

**Median** for ‘Points’ = 3.695

‘Score’ = 3.325

‘Weigh’ = 17.710

**Mode** for Points = 0 3.07

1 3.92

**Mode** for Score = 0 3.44

**Mode** for Weigh = 0 17.02

1 18.90

**Variance** for Points = 0.285881

Score = 0.957379

Weigh = 3.193166

**Standard deviation** for Points = 0.534679

Score = 0.978457

Weigh = 1.786943

**Points Range =** 2.17

**Score Range =** 3.9110000000000005

**Weigh Range =** 8.399999999999999

**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:** Total number of values = 9

= 1/9[108+110+123+134+135+145+167+187+199]

= 1/9(1308)

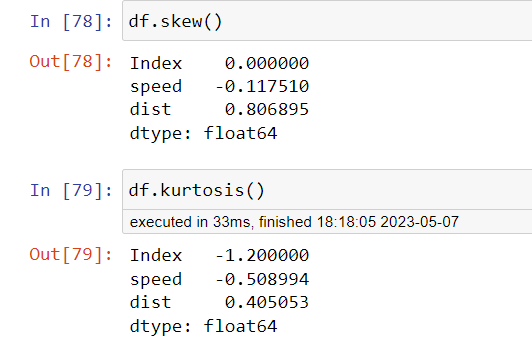
= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

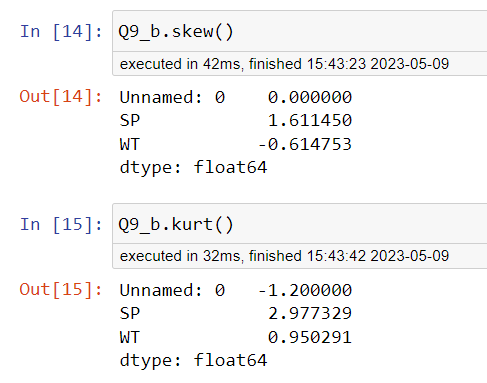
**Ans:**



**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:**

****

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



**Ans:**

**HISTOGRAM**: This graph Histogram showing that the distribution of the data is asymmetric, this is right skewed and we also say that it’s a positively skewed.

**BOXPLOT** : The boxplot showing positive skewness of the distribution.

**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:** Given Sample mean(X) = 200

Standard deviation(s) = 30

Random Sample(n) = 2000

CI at 94% ranges from **198.73 to 201.26**

CI at 98% ranges from **198.43 to 201.56**

CI at 96% ranges from **198.62 to 201.37**

**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: mean = 40

Median = 40.5

Variance = 25.52

Standard deviation = 5.05

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

**Ans:** Zero skewness

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

**Ans:** Positively skewed

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

**Ans:** Negatively skewed

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

**Ans:** Positive values of kurtosis indicate that a distribution is peaked and possess thick tails

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

**Ans:** A distribution with a negative kurtosis value indicates that the distribution has **lighter tails** than the normal distribution

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



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

**Ans**: The distribution of data is at the high end of the scale.

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

**Ans:** The distribution is skewed left.

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

**Ans:** The IQR of the data is 18 to 10

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



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

**Ans:**

Both the boxplots show some symmetry means that they are normally distributed. There is no outlier, median is around260.

**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:**

**Text, application

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**

**Ans:**

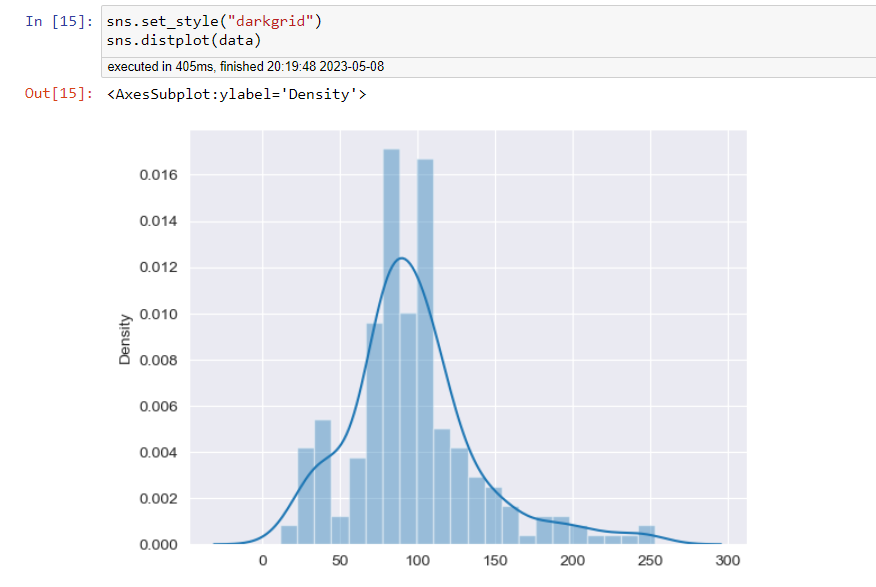
**Chart, histogram

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**

**Ans:**

****

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

**Ans:**

Graphical user interface, text, application

Description automatically generated

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



**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**:

Text

Description automatically generated