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

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

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

Solution: Total Possible outcome 2^3 = 8

Total outcomes HHH, HHT, HTT, THT, HTH, THH, TTT

No. Of Favourable Outcomes = 3

P(2 Head and 1 Tail) = 3/8 = 0.375

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

**Solution:**

1. Total number of outcomes = 6^2 = 36

No. Of outcomes equal to 1 = 0

P(Equal to 1) = 0/36 = 0

b) Less than or equal to 4

Total number of outcomes = 6^2 = 36

No. Of outcomes Less than or equal to 4 = 6

P(Less than or equal to 4) = 6/36 = 0.166

c) Sum is divisible by 2 and 3

Total number of outcomes = 6^2 = 36

No of outcomes Sum is divisible by 2 and 3 = 6

P(Sum is divisible by 2 and 3) = 6/36 = 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?

Solution:

Total no of balls (2 + 3 + 2) = 7

No of ways of drawing 2 balls out of 7 n(S) = (7\*6) / (2\*1) = 21

E = Event of drawing 2 balls, none of which is blue

n(E) = (2 + 3) = 5

= 5C2 = (5\*4) / (2\*1) = 10

P(E) = 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

Solution : 3.09

Q7) 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**

**Solution: [Python Assignment attached.](https://colab.research.google.com/drive/1S7ytT_l6GdXTe4uHmaiXsNJ8Ki4dYb1i?authuser=1)**

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?

**Solution: [Python Assignment attached.](https://colab.research.google.com/drive/1S7ytT_l6GdXTe4uHmaiXsNJ8Ki4dYb1i?authuser=1)**

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

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Solution: [Python Assignment attached.](https://colab.research.google.com/drive/1S7ytT_l6GdXTe4uHmaiXsNJ8Ki4dYb1i?authuser=1)**

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





**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?

**Solution: Python Assignment attached.**

**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?

**Solution: Python Assignment attached.**

1) Find mean, median, variance, standard deviation.

**Solution:**

statistics.mean(marks) 41

statistics.median(marks) 40.5

statistics.mode(marks) 41

statistics.variance(marks) 25.529

statistics.stdev(marks) 5.05266

2) What can we say about the student marks?

**Solution:**

we can say that mean of student marks is 41 which is slightly greater than median. Most of the students got marks in between 41 and 42. Also there are 2 outliers at 49, 56

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

**Solution:** When the mean and median of the data are equal then the nature of skewness has zero skewness.

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

**Solution: T**he nature of skewness when the mean is greater than the median, the distribution is positively skewed.

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

**Solution: T**he nature of skewness when the mean is less than the median, the distribution is negatively skewed.

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

**Solution:** Positive kurtosis value indicates for a data that the data points are more concentrated towards the right-hand side of the distribution.

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

**Solution:** Negative kurtosis value indicates for a data that the data points are more concentrated towards the left-hand side of the distribution.

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



What can we say about the distribution of the data?

**Solution:** The distribution of the data is not a Normal Distribution.

What is nature of skewness of the data?

**Solution: The nature of the Skewness of the data is Left Skewed.**

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

**Solution: Both are normally distributed.**

**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

**Solution:**

* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

* 1. P(MPG>38) = 0.35

statistics.mean(cars.MPG)

34.422

statistics.stdev(cars.MPG)

9.1314

Pnorm(38,34.422, 9.13)

0.652513

1 - 0.652513 = 0.35

**b) P(MPG<40) = 0.73**

Pnorm(40,34.422, 9.13)

0.73

**C) P (20<MPG<50) = 0.898**

Pnorm(50,34.422, 9.13) - Pnorm(20,34.422, 9.13)

0.8989178

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

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Solution: Python Assignment attached.**

From above plot and values we can say that data is fairly symmetrical, and MPG of Cars 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

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

**Solution:**

**90% Confidence**

**qnorm(0.95)**

**1.644854**

**94% Confidence**

**qnorm(0.97)**

**1.880794**

**60% Confidence**

**qnorm(0.80)**

**0.8416212**

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

**Solution: 95% Confidence**

**qt(0.975, df = 24)**

**2.063899**

**96% Confidence**

**qt(0.98, df = 24)**

**2.171545**

**99% Confidence**

**qt(0.995, df = 24)**

**2.79694**

**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

μ = 270, n = 18, x bar = 260, s = 90

T score = (x bar - μ)/(s/sqrt(n))

= (260-270)/(90/sqrt(18))

= -10/21.23

= -0.47

Required probability = 0.32

Pt(-0.47, df=17)

0.3221639