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

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

Ans: When 3 coins are tossed the total no. of possible combinations are 2\*\*3 = 8.

These combinations are HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

The no. of combinations which have 2 heads and 1 tail are :

HHT, HTH, THH

Therefore, probability of getting 2 heads and 1 tails in the toss of 3 coins simultaneously is:

P(Two heads and one tail) :

P(HHT) + P(HTH) + P(THH) = 1/8 + 1/8 + 1/8 = 3/8

**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: Two dice are rolled

Sample space = { (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6) }

n(S) = 6 x 6 = 36 (6 rows & 6 columns )

1. Sum is equal to 1 = 0

Probability = 0/ n(s) = 0/36 = 0

1. Sum is less than or equal to 4 =

X = {(1,3), (2,2), (3,1)}

P(X) = n(X)/n(S) = 3/36 = 1/12

1. Sum is divisible by 2 and 3: Form the above sample space,

Y = {(1,5), (2,4), (3,3), (4,2), (5,1), (6,6)}

So, as you can see there are only 6 set which can’t be divisible by 2 and 3: Y = 6

P(Y) = n(Y)/n(S) = 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?**

Ans: Total no. of balls = 2R + 3G + 2B = 7

1. No. of ways of drawing 2 balls out of 7:

7C2 = (7 x 6)/ (2 x 1) = 21

No. of balls other than blue = 5

1. No. of ways of drawing 2 balls out of 5:

5C2 = (5 x 2)/ (2 x 1) = 10

Probability = 5C2 / 7C2 = 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 No. of candies for a randomly selected child =

(1 x 0.015 + 4 x 0.20 + 3 x 0.65 + 5 x 0.005 + 6 x 0.01 + 2 x 0.120) = 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:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Mean** | **Median** | **Mode** | **Variance** | **Standard Deviation** |
|  |  |  |  |  |  |
| **Points:** | 3.596563 | 3.695 | Numeric | 0.285881 | 0.534679 |
|  |  |  |  |  |  |
| **Score:** | 3.217250 | 3.325 | Numeric | 0.957379 | 0.978457 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| **Weight:** | 17.848750 | 17.710 | Numeric | 3.193166 | 1.786943 |

**File in Jupyter Notebook : Question no.7, Assignment 1.ipynb**

**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: Excepted Value = sum( probability \* Value) = sum P(x).E(x)

There are 9 patients: Probability of selecting each patient = 1/9

The weights (X) of patients at a Clinic (in pounds):

E(x) = 108, 110, 123, 134, 135, 145, 167, 187, 199

P(x) = 1/9

Expected Value = (1/9) (108+ 110+ 123+ 134+ 135+ 145+ 167+ 187+ 199)

= (1/9) (1308) = 145.33

Expected Value of the Weight of that patient = 145.33

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

1. **speed and distance {Use Q9\_a.csv}**
2. **SP and Weight(WT) {Use Q9\_b.csv}.**

Ans.

**Skewness**

|  |  |  |
| --- | --- | --- |
|  | **Speed** | **Distance** |
| 1. Q9\_a | -0.11395477 | 0.78248352 |
|  |  |  |
|  | **SP** | **WT** |
| 1. Q9\_b | 1.58145368 | -0.60330993 |
|  |  |  |

Here if the skewness value is negative means left skewed, and if positive then it’s right skewed.

**Kurtosis**

|  |  |  |
| --- | --- | --- |
|  | **Speed** | **Distance** |
| 1. Q9\_a | -0.57714742 | 0.24801866 |
|  |  |  |
|  | **SP** | **WT** |
| 1. Q9\_b | 2.72352149 | 0.819465881 |
|  |  |  |

**File in Jupyter Notebook : Question no.9, Assignment 1.ipynb**

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

****

**Ans:** The most of the data points are in the range of 50-100 with frequency 200 and least range of weight is 400 somewhere around 0-10. So, the expected value of the distribution is 75.

**Skewness:** We can notice a long tail towards right so it is heavily right skewed.

****

**Ans:** Above plot is less than mean right skewed and we have outlier on the upper side of box plot and there is less data points between 1 and bottom point.

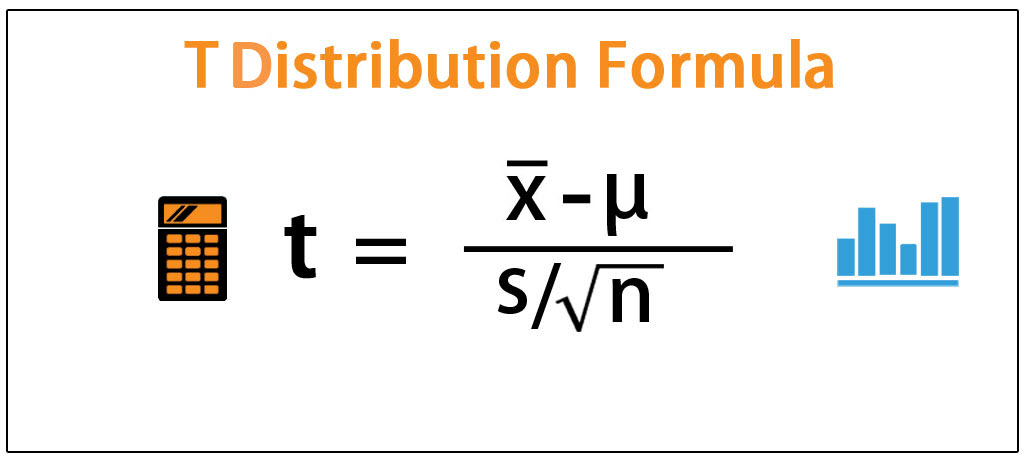
**Q11) Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?**

Ans: The information given is: Since the data is too small so using **t-distribution**

* Sample **mean** of x = 200
* Sample **standard deviation** of s = 30
* Sample **size** of n = 2000

The confidence interval = x(+/-) t\*(s/(n)\*\*1/2)

* **x:** sample mean
* **t:** t-value that corresponds to the confidence level
* **s:** sample standard deviation
* **n:** sample size



Solved in python

|  |  |  |
| --- | --- | --- |
| **C.I.** | **x - t\*(s/(n)\*\*1/2)** | **x + t\*(s/(n)\*\*1/2)** |
| 94% | 198.738325292158 | 201.261674707842 |
|  |  |  |
| 98% | 198.43943840429978 | 201.56056159570022 |
|  |  |  |
| 96% | 198.62230334813333 | 201.37769665186667 |

**File in Jupyter Notebook : Question no. 11, Assignment 1.ipynb**

**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 = 41.0
* Median = 40.5
* Variance = 25.529411764705884
* Standard Deviation = 5.0566382858645

**File in Jupyter Notebook : Question no.12, Assignment 1.ipynb**

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

Ans: Symmetrical

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

Ans: Right skewed

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

Ans: Left skewed

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

Ans: The Data is normally distributed and kurtosis value is 0.

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

Ans: The distribution of the data has lighter tails and a flatter peaks 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: Let’s assume above box plot is about age’s of children in a society. 50% of the children are above 10years old and remaining are less, and children who’s age is above 15 are approx 40%.

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

Ans: Left skewed because ( median > mean )

**What will be the IQR of the data (approximately)?**Ans: Approximately = -8

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

****

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

Ans: By observing both the plots whisker’s level is high in boxplot 2, mean and median are equal hence distribution is symmetrical.

**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)**

**c. P (20<MPG<50)**

Ans:

|  |  |
| --- | --- |
|  | **Probability of MPG of Cars** |
| P(MPG>38) | 0.3475939251582705 |
| P(MPG<40) | 0.72934987621516161 |
| P(20<MPG<50) | 1.2430968797327613e-05 |

**File in Jupyter Notebook : Question no.20, Assignment 1.ipynb**

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

1. **Check whether the MPG of Cars follows Normal Distribution**

**Dataset: Cars.csv**

Ans:

Mean = 34.422075728024666 , Standard Deviation(std) = 9.131444731795982

**File in Jupyter Notebook : Question no.21(a), Assignment 1.ipynb**

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

**Dataset: wc-at.csv**

Ans:

|  |  |  |
| --- | --- | --- |
|  | **Waist** | **AT** |
|  |  |  |
| Mean | 91.901835 | 101.894037 |
| Median | 90.80 | 96.54 |
| Mode | 94.5, 106.0, 108.5 | 121.0, 123.0, NaN |
|  |  |  |

**File in Jupyter Notebook : Question no.21(b), Assignment 1.ipynb**

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

Ans:

* Z-score of 90% CI = 1.644853626914722
* Z-score of 94% CI = 1.8807936081512509
* Z-score of 60% CI = 0.8416212335729143

**File in Jupyter Notebook : Question no.22, Assignment 1.ipynb**

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

Ans:

* t-score of 95%CI = 2.063895616280205
* t-score of 96%CI = 2.1715446760080677
* t-score of 99%CI = 2.796939504772804

**File in Jupyter Notebook : Question no.23, Assignment 1.ipynb**

**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 at x =260 , t = -0.4714045207910317
* p\_value using cdf function = 0.32073040369132255
* p\_value using sf function = 0.32167411684460556

**File in Jupyter Notebook : Question no.24, Assignment 1.ipynb**