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

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

**Ans…** 3/8

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

1. Equal to 1

**Ans…** 0

1. Less than or equal to 4

**Ans…** 1/6

1. Sum is divisible by 2 and 3

**Ans…** 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…** 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 |

**Ans….**

Expected no. 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))

**= 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…**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weight** |
| **Mean** | 3.596563 | 3.21725 | 17.84875 |
| **Median** | 3.695 | 3.325 | 17.71 |
| **Mode** | 3.92 | 3.44 | 17.02 |
| **Variance** | 0.285881 | 0.957379 | 3.193166 |
| **Standard Deviation** | 0.534679 | 0.978457 | 1.786943 |
| **Range** | 2.17 | 3.911 | 8.4 |

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...** 145.33 pounds

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans….**

|  |  |  |
| --- | --- | --- |
|  | **Cars speed** | **Distance** |
| **Skewness** | -0.11751 | 0.806895 |
| **Kurtosis** | -0.50899 | 0.405053 |
| **Mean** | 15.4 | 42.98 |
| **Median** | 15 | 36 |
| **Mode** | 20 | 26 |
| **Variance** | 27.95918 | 664.0608 |
| **Standard deviation** | 5.287644 | 25.76938 |

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans….**

|  |  |  |
| --- | --- | --- |
|  | **SP** | **WT** |
| **Skewness** | 1.61145 | -0.61475 |
| **Kurtosis** | 2.977329 | 0.950291 |
| **Mean** | 121.5403 | 32.41258 |
| **Median** | 118.2087 | 32.73452 |
| **Mode** | 118.289 | #N/A |
| **Variance** | 201.113 | 56.14225 |
| **Standard deviation** | 14.18143 | 7.492813 |

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



**Ans….** In above histogram figurer and boxplot figure is positively skewed on right side. That means Mean and median of the data is greater than mode

**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….**

1. The 94% CI is (198.73, 201.27)
2. The 96% CI is (198.61, 201.39)
3. The 98% CI is (198.43, 201.57)

**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 |
| **Median** | 40.5 |
| **Variance** | 25.52941 |
| **Standard deviation** | 5.052664 |

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

**Ans….** The skewness is zero

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….** Distribution is peaked and possesses thick tails

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

**Ans….** Distribution is flat and has thin tails

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

**Ans….**1.A negatively skewed distribution is a type of distribution in which more

values are concentrated on the right side of the distribution graph.

2. The nature of skewness of the data is that Mean of negatively skewed data

will be less than the Median.

3. IOR = (-10-18)

Q19) Comment on the below Boxplot visualizations?



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

**Ans….** 1. Boxplot 1 is positively skewed that’s means Mean and Median is greater than Mode.

2. Boxplot 2 is shows normal distribution that means the skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero.

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)

**Ans….** 1- stats.norm.cdf(38,loc=34.422,scale=9.13) = 0.3475908

* 1. P(MPG<40)

**Ans….** stats.norm.cdf(40,loc=34.422,scale=9.13)= 0.7293527

* 1. P (20<MPG<50)

**Ans….** stats.norm.cdf(50,loc=34.422,scale=9.13) – stats.norm.cdf(20,loc=34.422,scale=9.13144)= 0.01311818

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans….** MPG of Cars is highly negative skewed.

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

**Ans….**

|  |  |
| --- | --- |
| Z score of 60% | 0.8416212 |
| Z score of 90% | 1.644854 |
| Z score of 94% | 1.880794 |

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

**Ans….**

|  |  |
| --- | --- |
| **Confidence Interval** | **t score** |
| 95% | 2.063899 |
| 96% | 2.171545 |
| 99% | 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

**Ans….**

import pandas as pd

import scipy.stats as stats

import numpy as np

a= (260-270)/(90/np.sqrt(18))

a=-0.4714045207910317

p\_value=1-stats.t.cdf(abs(a),df=17)

p\_value=**0.32167253567098353**