|  |  |
| --- | --- |
| 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) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ratio |
| 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?

**Solution:**

Total number of ways=2\*2\*2=8

Total number of favorable outcomes=3

Probability=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

**Solution:**

Total number of ways=6\*6=36

1. Favorable outcomes=0

Probability=0/36

1. Favorable outcomes=6

Probability=6/36=1/6

1. Favorable outcome=4

Probability=4/36=1/9

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:**

Probability when one ball is drawn=5/7

Probability when second ball is drawn=4/6

Overall Probability=5/7\*4/6=20/42=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:**

Expected number of candies= (1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*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**

**Solution:**

**Points:**

Mean= 3.5965

Median= 3.6950

Mode= 3.07, 3.92

Variance= 0.2858

Standard Deviation= 0.5346

Range=2.76-4.93

We can see that Mean<Median<Mode then it shows that it has negatively skewness.

**Score:**

Mean= 3.2172

Median= 3.325

Mode= 3.44

Variance= 0.9573

Standard Deviation= 0.9784

Range= 1.513-5.424

We can see that Mean<Median<Mode then it shows that it has negatively skewness.

**Weigh:**

Mean= 17.8487

Median= 17.71

Mode= 17.02, 18.90

Variance= 3.1931

Standard Deviation= 1.7869

Range= 14.5-22.9

We can see that Mode<Median<Mean then it shows that it has positively skewness.

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:**

Probability of each patient if chosen at random=1/9=0.11

Expected value= (108\*0.11)+(110\*0.11)+(123\*0.11)+(134\*0.11)+(135\*0.11)+(145\*0.11)+(167\*0.11)+(187\*0.11)+(199\*0.11)= 143.88

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Solution:**

Skewness of speed=-0.1175

Kurtosis of speed= -0.5089

Skewness of distance=0.8068

Kurtosis of distance= 0.4050

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Solution:**

Skewness of SP= 1.6114

Kurtosis of SP= 2.9773

Skewness of WT= -0.6147

Kurtosis of WT= 0.9502

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



**Solution:**

This histogram is right skewed. Its skewness is positive.



This is a box plot and it contains the information about 25th percentile(Q1), 50th percentile(Q2), 75th percentile(Q3), Lower whisker length, Upper whisker length and outliers.

**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:**

Confidence interval for 94%= (199.36,200.63)

Confidence interval for 98%= (199.34,200.65)

Confidence interval for 96%= (199.35,200.64)

**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.

**Solution:**

Mean=41.0

Median=40.5

Variance=25.529

Standard Deviation=5.052

1. What can we say about the student marks?

**Solution:**

Here the mean is greater than median, and the standard deviation is 5.052.

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

**Solution:** The nature of skew is symmetrical.

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

**Solution:** The nature of skew is positive.

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

**Solution:** The nature of skew is negative.

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

**Solution:** It indicates that the distribution is peaked and possesses thick tails.

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

**Solution:** It 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?

**Solution:** 50th percentile lies between 14 and 16. There are no outliers. There is also a bigger difference between the Upper whisker length and the Lower whisker length.

What is the nature of skewness of the data?

**Solution:** Left skewed is the nature of skewness of the data.

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

Q1=10

Q3=18

IQR= Q3-Q1

=18-10  
 =8

IQR of the data is 8.

Q19) Comment on the below Boxplot visualizations?



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

**Solution:**

The median is the same for both the boxplots. There is a difference between the Interquartile range(IQR) for both the plots and there are no outliers.

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)

**Solution:**

1-nd.cdf(38)= 0.328

* 1. P(MPG<40)

**Solution:**

nd.cdf(39.9)= 0.743

* 1. P (20<MPG<50)

**Solution:**

nd.cdf(50)-nd.cdf(20)= 0.902

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Solution:**

Yes, it follows 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

**Solution:**

Yes, it follows normal distribution.

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

**Solutions:**

from scipy.stats import norm

z\_90=norm.ppf(0.95)= 1.6448

z\_94=norm.ppf(0.97)= 1.8807

z\_60=norm.ppf(0.8)= 0.8416

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

**Solution:**

from scipy.stats import t

t\_95=t.ppf(0.975,24)= 2.0638

t\_96=t.ppf(0.98,24)= 2.1715

t\_99=t.ppf(0.995,24)= 2.7969

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

**Solution:**

(Sample mean – population mean)/ (standard deviations/root(number of sample))

Therefore, answer=-0.4