|  |  |
| --- | --- |
| 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 | Nominal |
| 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 | Ratio |
| Number of Children | Nominal |
| 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?

Ans: 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: a) 0, it is not possible to two dice are rolled and the sum of outcomes is equal to 1.

b) probability = 6/36 =1/6

c) 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  
Let S be the sample space  
Then, n(S) = Number of ways of drawing 2 balls out of 7

n(S)=7C2

n(S)=(7×6)/(2×1)

n(S)=21

Let E = Event of 2 balls, none of which is blue  
∴ n(E) = Number of ways of drawing 2 balls out of (2 + 3) balls

n(E)=5C2

n(E)=(5×4)/(2×1)

n(E)=10

∴ probability that none of the balls drawn is blue P(E)=n(E)/n(S)

=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

=3.09

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:-**

Point = 3.5965

Score = 3.2172

Weigh = 17.8487

**Median:-**

Points = 3.6950

Score = 3.325

Weigh = 17.71

Mode:-

Points = 3.07, 3.92

Score = 3.44

Weigh = 17.02, 18.90

Variance:-

Points = 0.2858

Score = 0.9573

Weigh = 3.1931

Standard Deviation:-

Points = 0.5346

Score = 0.9784

Weigh = 1.7869

Range:-

Points= (2.76 , 4.93) = 2.17

Score= (1.51 , 5.42) = 3.91

Weigh= (14.5 , 22.9) = 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: Expected value = (total weight of the patient/total no. of patient)

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

= 1308/9

= 145.33

Expected Value of weight of that patient = 145.33 pound

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans: Skewness:-**

Speed = -0.1175

Distance = 0.8068

**Kurtosis:-**

Speed = -0.5889

Distance = 0.4050

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans: Skewness:-**

Speed = 1.6114

Weight = -0.6147

Kurtosis:-

Speed =2.9773

Weight =0.9502

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



Ans:- The histograms peak has right skew and tail is on right.

Mean > Median, We have outliers on the higher side.



Ans: The boxplot has outliers on the maximum side.

**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: Give that :

sample size = 2000

Population size = 3000000

Mean of sample (x̅) = 200

Standard deviation of sample (౮) =30

Standard Error = (x̅ /√౮)

For 94% confidence interval range is (198.738,201.261)

For 98% confidence interval range is (198.439,201.560)

For 96% confidence interval range is (198.622,201.377).

**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: 1) Mean = 41

Median = 40.5

Variance = 25.529

Standard deviation = 5.052

2) We don’t have outliers and the data is slightly skewed towards right because mean is greater than median.

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

Ans: The distribution is perfectly symmetrical.

:The curve is following normal distribution.

: The skewness of the distribution is zero.

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

Ans: The distribution is positively skewed (right skewed).

: The tail of the curve is on right side.

: The skewness is +ve.

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

Ans: The distribution is negatively skewed (left skewed).

: The tail of the curve is on left side.

: The skewness is -ve.

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

Ans: It indicates a sharper peak of the curve.

: not following a normal distribution.

: most observations comes on the tails.

: flatter tails.

: kurtosis value > 3.

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

Ans: It indicates a flatter peak of the curve.

: not following a normal distribution.

: less observations comes on the tails.

: thinner tails.

: kurtosis value < 3.

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



a)What can we say about the distribution of the data?

b)What is nature of skewness of the data?

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

Ans: a)  The above Boxplot is not normally distributed the median is towards the higher value.

b)The data is a skewed towards left. The whisker range of minimum value is greater than maximum.

c)The Inter Quantile Range (IQR) = Q3 Upper quartile – Q1 Lower Quartile

= 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.

Ans: 1)There are no outliers.

2)Both the box plot shares the same median that is approximately in a range between 275 to 250.

1. They are normally distributed with zero .
2. No skewness neither at the minimum or maximum whisker range.

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: a) P(MPG>38) = 0.6524060748417295

1. P(MPG<40) = 0.7293498762151616
2. P(20<MPG<50) = 0.8988689169682046

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: Yes, MPG of Cars 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

Ans: Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set doesn’t follow Normal Distribution.

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

Ans: Z score of 90% confidence interval = 1.64

Z score of 94% confidence interval = 1.88

Z score of 60% confidence interval = 0.84

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

Ans: sample size (n) = 25

Degree of freedom = n-1

= 25-1

= 24

t score of 95% confidence interval = 2.06

t score of 96% confidence interval = 2.17

t score of 99% confidence interval = 2.79

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 - statistics for the data is given as follows:



x = mean of the sample of bulbs =  260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18









t = - 0.471

degree of freedom (df) = n-1

= 18-1

So df = 17

Code for python

P\_value = 1-stats.t.cdf(abs(t-score),df=n-1)

P\_value = 1-stats.t.cdf(abs(-0.471),df=17)

Then, P\_value = 0.32167411684460556

So the probability of 18 randomly selected bulbs would have an average life of no more than 260 days is 0.321.