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

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 data- Nominal |
| High School Class Ranking | Discrete data- Nominal |
| Celsius Temperature | Continuous- Interval |
| Weight | Continuous- Ratio |
| Hair Color | Discrete data- 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 | Discrete- Ratio |
| Barometer Pressure | Continuous- Ratio |
| SAT Scores | Continuous- Ratio |
| Years of Education | Discrete- Nominal |

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

Sol: P(H H T) + P(H T H) + P(T H H)

=1/8 + 1/8 +1/8

=3/8

Binomial distribution

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

Sol: a)There is no outcomes which corresponds sum is equal to one i.e. 0/36. Probability is 0

b) (1,3) (2,2) (3,1) = 3 outcomes, 3/36 i.e. 1/12

c) 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?

Sol: 10/21

P(2R,3G,2B)

P(5/7,4/6)= 20/42 i.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

Sol: So expected number of candies for randomly selected child = 1\*0.015+4\*0.120+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 inference

Sol: Solution is in Q7\_soln

Points: Mean =3.5965, Median=3.695, Mode=”numeric”, Variance=0.2858814, Standard deviation: 0.5346787

Score: Mean =3.271725, Median=3.325, Mode=”numeric”, Variance=0.957379, Standard deviation: 0.9784574

Weigh: Mean =17.84875, Median=17.71, Mode=”numeric”,

Variance= 3.193166 , Standard deviation: 1.786943

**Range: 2.17**

**Note: The value of mean of both the points and score are almost**

similar

**Use Q7.csv file**

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?

Sol: Expected Value  =  ∑ ( probability  \* Value )

 ∑ P(x).E(x)

there are 9 patients

Probability of selecting each patient = 1/9

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

P(x) 1/9  1/9   1/9  1/9   1/9   1/9   1/9   1/9  1/9

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

= (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 speed and distance**

**Use Q9\_a.csv**

**Sol : Solution is in Q9\_a\_sol**

Skewness :- speed= -0.117510, distance= 0.806895

Kurtosis :- speed= -0.508994, distance= 0.405053

**Solution is in Q9\_b\_sol**

**SP and Weight(WT)**

Skewness :- SP= 1.611450, Weight(WT)= -0.614753

Kurtosis:- SP= 2.977329, Weight(WT)= 0.950291

**Use Q9\_b.csv**

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



**Sol :** The most of the data point are concenrated in the range 50-100 with frequency 200.

And least range of weight is 400 somewere around 0-10

So the expected value the above distribution is 75.

Skewness – we can notice a long tail towards right so it is heavily right skewed.



**Sol : Median is less than mean right skewed and we have outlier on the upper side of box plot and there is less data points between Q1 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?

**Sol : 94% Confidence Interval**

T = x+t σ/ **√** n

= 200+1.88\*30/**√2000**

= 200+1.2611

t = 201.2611

200-1.2611

= 198.73

t-score = (201.2611,198.73)

**98% Confidence Interval**

t = 200+2.326\*30/**√2000**

= 200+1.56032

t = 201.56032

200-1.56032

= 198.43968

t-score =(201.56032,198.43968)

96% Confidence Interval

t = 200+2.05\*30/**√2000**

= 200+1.375

t = 201.375

200-1.375

= 198.625

t-score =(201.375,198.625)

**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?

**Sol : 1)**  Mean = 41, Median = 40, variance = 24.11, Standard deviation = 4.910

2)Marks of the student is average 41

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

**Sol :** Symetrical

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

**Sol :** Skewed

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

**Sol :** Left skewed

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

**Sol :** The data is normally distributed and kurtosis value is 0

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

**Sol :** The distribution of the data has lighter tails and flatter a peak than the normal distribution

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



What can we say about the distribution of the data?

**Sol :** Let’s assume above the box plot is about age’s of the student in a school.

50% of the people are above 10yrs old and remaining are less.

And student who’s age is above 15 and approx. 40%.

What is nature of skewness of the data?

**Sol :** Left skewed, median is greater than mean.

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

Q19) Comment on the below Boxplot visualizations?



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

Sol : 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)

**Sol : a) P(MPG>38)**

**0.347593925**

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

**0.7293498762**

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

**1.243095879e-05**

**Solution is in Q20\_sol**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Sol: The Histogram looks roughly symmetric and unimodal. This is consistent with data that is approximately Normally Distributed

Graph is in python Q21\_a\_sol

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

Dataset: wc-at.csv

Sol : Solution is in Python Q21\_b\_sol

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

Sol : 90% Confidene interval

Area = 1+C.L/2 = 1+0.90/2= 1.90/2= 0.95

= 0.95 = 1.6+0.04 = 1.64

Z-score of 90% Confidence Interval is 1.64

94% Confidence Interval

Area = 1+C.L./2 = 1+0.94/2 = 1.94/2=0.97

= 0.97 = 1.8+0.08=1.880

Z-Score of 94% Confidence Interval is 1.880

60% Confidence Interval

Area = 1+C.L./2 = 1+0.60/2 =1.6/2 =0.8

= 0.8 = 0.8+0.84 = 0.884

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

Sol : 95% Confidence Interval

= 2.06390

96% Confidence Interval

= 2.171544

99% Confidence Interval

= 2.79693

Solution in python Q23\_sol

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

**Sol :** 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