**SOLUTION OF ASSIGNMENT -1**

Ans.Q1) Idnetify the Data type for the Following: anvesh(LMS)- 9390104941

|  |  |
| --- | --- |
| 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 |

Discrete Data: Discrete Data involves whole number and that can’t be divided based on the nature of what they are. For example integers like 1, 10 or 356 like the number of people or children in a class, the number of fingers on our hand. We can’t have 1.9 children in a family.

Continuous Data : Continuous Data involves divided up as much as we want, and measured to many decimal places. For example weight of a car calculated as many decimal places even temperature as 38.965 degrees and speed of airplane is 943.6 km/hr.

Ans.Q2) Identify the Data types, which were among the following:

Nominal, Ordinal, Interval, Ratio.

Nominal Data : Nominal Data are used to label variables without any quantitative value. For examples include male/female, hair color, nationalities, names of people, and so on.

Ordinal Data : Ordinal Data are used to remember that ordinal sounds like order and it’s the order of the variables which matters. Not so much the differences between those values. For example include measure of satisfaction, happiness, sad, emotion and so on.

Interval Data : Interval Data is fun and useful because it’s concerned with both the order and difference between your variables. This allows to measure Standard Deviation and Central Tendency. For example of interval data is temperatures in degrees Celsius. The difference between 10 and 0 is also 10 degrees.

Ratio Data : Ratio Data tells us about the order of variables, the differences between them, and they have that absolute zero. Which allows all sorts of calculations and inferences to be performed and drawn. Ratio Data is very similar to Interval Data, For Ratio Data not possible to have negative values. For instance, height is ratio data. It is not possible to have negative height.

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

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

Answer : Possible outcomes for the three coins tossed : (HHH, HHT, HTH, THH, HTT, THT, TTH, TTT )

Probability that two heads and one tail is : (HHT, HTH, THH)

Hence P = 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

Answer : S = { (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) }

Possible total number of outcomes = 6 x 6 = 36

1. Equal to 1 : When the sum is equal to '1' the Probability will be Zero.

P(E) = n(E) / n(S) = 0 / 36 = 0

(b) Less than or equal to 4

P(AorB)=P(A)+P(B)

Three possible outcomes give a sum less than 4: E = {(1,1),(1,2),(2,1)}, hence. P(A) =3/36

Three possible outcomes give a sum equal to 4: E = {(1,3),(2,2),(3,1)}, hence. P(B)=3/36

P(AorB)=P(A)+P(B) = 3/36+3/36=6/36=1/6

(c) Sum is divisible by 2 and 3

Total number of events =36

Interested number of events = {(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)} =6

P(Sum is divisible by 2 and 3) =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?

Answer : Total number of events {(R,R),(R,G),(R,B),(G,R),(G,G),(G,B),(B,R),(B,G),(B,B)}=9

Interested number of events =4

Probability =4/9

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

Answer : Expected number of candies for a randomly selected child is given by

∑Xp(X)=1\*0.015+4\*0.20+3\*0.65+5\*0.65+5\*0.005+6\*0.01+2\*0120

∑Xp(X)=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**

Answer: Solution in Assignment-1-Q7(Basic Statistics Level-1).ipynb

Q8) Calculate Expected Value for the problem below

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?

Answer: Given,The weights (X) of patients at a clinic (in pounds), are 108, 110, 123, 134, 135, 145, 167, 187, 199

one of the patients is chosen at random.

To Find : Expected Value

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

= (1/9)  x ( 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**

Answer: Solution in Assignment-1-Q9\_a(Basic Statistics Level-1).ipynb

**SP and Weight(WT)**

**Use Q9\_b.csv**

Answer: Solution in Assignment-1-Q9\_b(Basic Statistics Level-1).ipynb

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





Answer : The maximum of the data points is present in the range 50-100 with frequency 200.

minimum range of weight is 400 and its frequency around 0-10.

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

Boxplot : 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?

Answer: Answer: Solution in Assignment-1(Basic Statistics Level-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**

Find mean, median, variance, standard deviation.

1. What can we say about the student marks?
2. What can we say about the student marks?

Answer: Solution in Assignment-1-Q12(Basic Statistics Level-1).ipynb

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

Answer: Symmetrical

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

Answer: Right Skewed

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

Answer: Left Skewed

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

Answer: The data is normally distributed and kurtosis value is 0.

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

Answer: 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.



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

Answer: Let’s assume above box plot is about ages of the students in a school.

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

And students whose age is above 15 are approx. 40%.

(b)What is nature of skewness of the data?

Answer: Left skewed, median is greater than mean.

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

Answer: Approximately= -8  
  
Q19) Comment on the below Boxplot visualizations?



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

Answer: 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)

Answer: Solution in Assignment-1-Q20(Basic Statistics Level-1).ipynb

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer: Solution in Assignment-1-Q21\_a(Basic Statistics Level-1).ipynb

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

Dataset: wc-at.csv

Answer: Solution in Assignment-1-Q21\_b(Basic Statistics Level-1).ipynb

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

Answer: Solution in Assignment-1-Q22(Basic Statistics Level-1).ipynb

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

Answer: Solution in Assignment-1-Q23(Basic Statistics Level-1).ipynb

Q 24**)**A Government companyclaims 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

Answer: Solution in Assignment-1-Q24(Basic Statistics Level-1).ipynb .