|  |  |
| --- | --- |
| 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 | Interval |
| 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 | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Interval |

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

HHH

HHT

HTH

HTT

THH

THT

TTH

TTT

Possible outcomes are HHT, HTH, THH

P(Two heads and one tail) = 3/8

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

1. Equal to 1

Ans: Dice starts from (1, 1)

Probability is 0.

1. Less than or equal to 4

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1,1 | 2,1 | 3,1 | 4,1 | 5,1 | 6,1 |
| 1,2 | 2,2 | 3,2 | 4,2 | 5,2 | 6,2 |
| 1,3 | 2,3 | 3,3 | 4,3 | 5,3 | 6,3 |
| 1,4 | 2,4 | 3,4 | 4,4 | 5,4 | 6,4 |
| 1,5 | 2,5 | 3,5 | 4,5 | 5,5 | 6,5 |
| 1,6 | 2,6 | 3,6 | 4,6 | 5,6 | 6,6 |

Ans: Number of possible outcomes are (1,1), (1,2),(1,3), (2,1), (2,2), (3,1)

=6/36 = 1/6

1. Ans: Sum is divisible by 2 and 3 are (1, 5),(2, 4), (3, 3), (4,2), (5,1), (6,6)

=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?

2R, 3G, 2B

P(None of the balls drawn is blue)

= 5C2 / 7C2

= 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:

|  |  |  |  |
| --- | --- | --- | --- |
| CHILD | Candies count (x) | Probability P(x) |  |
| A | 1 | 0.015 | 0.015 |
| B | 4 | 0.20 | 0.8 |
| C | 3 | 0.65 | 1.95 |
| D | 5 | 0.005 | 0.025 |
| E | 6 | 0.01 | 0.06 |
| F | 2 | 0.120 | 0.24 |
| Total |  |  | 3.09 |

= 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.



Solution:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh> |
| Max | 4.93 | 5.424 | 22.9 |
| Min | 2.76 | 1.513 | 14.5 |
| Mean | 3.613548 | 3.210065 | 17.95355 |
| Median | 3.7 | 3.215 | 17.82 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variance | 0.28587 | 0.987584 | 2.936444 |
| Standard deviation | 0.534668 | 0.993773 | 1.713605 |
| Range | 2.17 | 3.911 | 8.4 |

All the three datasets - Points, Scores and weighs are left skewed, since mean<median<mode.

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:

|  |  |  |
| --- | --- | --- |
| X | P(X) |  |
| 108 | 1/9 = 0.1111 | 12 |
| 110 | 1/9 | 12.22222 |
| 123 | 1/9 | 13.66667 |
| 134 | 1/9 | 14.88889 |
| 135 | 1/9 | 15 |
| 145 | 1/9 | 16.11111 |
| 167 | 1/9 | 18.55556 |
| 187 | 1/9 | 20.77778 |
| 199 | 1/9 | 22.11111 |
| Total |  | 145.3333 |

Expected Value = = 145.33

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

**Cars speed and distance**



|  |  |  |
| --- | --- | --- |
|  | Speed | Dist. |
| Skew | -0.89542 | 1.290763 |
| Kurt | 0.249561 | 2.464546 |

Skewness for speed is -0.89542negatively skewed

Skewness for dist. is 1.2907 positively skewed

Kurtosis for speed is 0.2495 close zero we can say mesokurtic.

Kurtosis for dist is 2.4645 leptokurtic (>0)

**SP and Weight(WT)**



|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| Skew | -0.42676 | -1.3436 |
| Kurt | -0.86373 | 1.159038 |

Skewness for SP is -0.4267 negatively skewed

Skewness for WT is -1.3436 negatively skewed

Kurtosis for SP is -0.8637 Platykurtic (<0)

Kurtosis for WT is 1.159 Leptokurtic (>0)

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



Most of the data paints falls in the range 50 to 100 with frequency 200 and it gradually decreases towards right with the least frequency 0 in the range 350 to 400.

Here the tail on the right side of the distribution is longer. Hence, we say that the data is positively skewed.



Most of the observations are on the low end of the scale.Here, the mean is greater than median the distribution is skewed right.

**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:

To estimate the mean weight in the population, we have the mean weight 200 as the sample statistic.

1. In this case, the confidence level is defined as 94%

Standard Error =

Degrees of freedom = n-1 = 1999

Critical Value

= = 0.06

Critical Probability,

The critical value is the Z score having 1999 degrees of freedom and a probability equal to 0.97. From the Standardized Normal distribution table we find that the critical value is 1.88.

Margin of Error: Critical value \* Standard Error =1.88\*0.6708 = 1.26

Margin of Error = 1.26.

Now specifying the Confidence Interval

Sample Statistic ± Margin of Error

= Sample Statistic ± Margin of Error

= 200 ± 1.26

= (198.74, 201.26)

ii) In this case, the confidence level is defined as 98%

Standard Error =

Degrees of freedom = n-1 = 1999

Critical Value

= = 0.02

Critical Probability,

The critical value is the Z score having 1999 degrees of freedom and a probability equal to 0.99. From the Standardized Normal distribution table we find that the critical value is 2.33.

Margin of Error: Critical value \* Standard Error =2.33\*0.6708 = 1.56

Margin of Error = 1.56.

Now specifying the Confidence Interval

Sample Statistic ± Margin of Error

= Sample Statistic ± Margin of Error

= 200 ± 1.56

= (198.44, 201.56)

1. In this case, the confidence level is defined as 96%

Standard Error =

Degrees of freedom = n-1 = 1999

Critical Value

= = 0.04

Critical Probability,

The critical value is the Z score having 1999 degrees of freedom and a probability equal to 0.98. From the Standardized Normal distribution table we find that the critical value is 2.06.

Margin of Error: Critical value \* Standard Error =2.06\*0.6708 = 1.38

Margin of Error = 1.38.

Now specifying the Confidence Interval

Sample Statistic ± Margin of Error

= Sample Statistic ± Margin of Error

= 200 ± 1.38

= (198.62, 201.38)

**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?

Solution:

1. mean, median, variance, standard deviation

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.5 |
| Mode | 41 |
| Variance | 24.11111 |
| Standard deviation | 4.910307 |

1. Average score obtained by the students in tests 41.

Here, the mean, median and mode are almost equal.

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

Solution: Symmetric.

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

Solution: Right skewed.

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

Solution: Left skewed.

Q16) What does positive kurtosis value indicates for adata ?

Solution: The distribution has heavier tails and a sharper peak than the normal distribution.

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

Solution: The distribution has lighter tails and a flatter 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?

Here we see that the 50% of the data are from 10 to 18. Median is approximately 15. The distribution is not normal.

What is nature of skewness of the data?

Negatively skewed

What will be the IQR of the data (approximately)?   
  
Approximately IQR = 8

Q19) Comment on the below Boxplot visualizations?



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

The distributions of both the plots are symmetrical.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

Solution:

1. P(MPG>38) = 33/81 = 0.407
2. P(MPG<40) = 61/81 = 0.753

c. P (20<MPG<50) = 69/81 = 0.851

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

|  |  |
| --- | --- |
| Min | 12.10126 |
| Q1 | 27.85625 |
| Median | 35.15273 |
| Q3 | 39.53163 |
| Max | 53.70068 |

|  |  |
| --- | --- |
| Mean | 34.42208 |
| Median | 35.15273 |
| Mode | 29.62994 |

From the above box plot, we see that the data is negatively skewed.

It is negatively skewed since mean is less than median.

Hence the data does not follow 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

1. Waist

|  |  |
| --- | --- |
| Min | 63.5 |
| Q1 | 80 |
| Median | 90.8 |
| Q3 | 104 |
| Max | 121 |

|  |  |
| --- | --- |
| Mean | 91.90183 |
| Median | 90.8 |
| Mode | 94.5 |

From the above box plot, we see that the data is negatively skewed.

It is negatively skewed since mean is less than mode.

Hence the data does not follow normal distribution.

1. AT

|  |  |
| --- | --- |
| Min | 11.44 |
| Q1 | 50.88 |
| Median | 96.54 |
| Q3 | 137 |
| Max | 253 |

|  |  |
| --- | --- |
| Mean | 101.894 |
| Median | 96.54 |
| Mode | 121 |

From the above box plot, we see that the data is positively skewed.

Hence the data does not follow normal distribution.

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

Solution:

**For 90% confidence interval**

Critical Value

= = 0.1

Critical Probability,

The critical value is the Z score and a probability equal to 0.95. From the Standardized Normal distribution table we find that the critical value is 1.64.

**For 94% confidence interval**

Critical Value

= = 0.06

Critical Probability,

The critical value is the Z score and a probability equal to 0.97. From the Standardized Normal distribution table we find that the critical value is 1.88.

**For 60% confidence interval**

Critical Value

= = 0.4

Critical Probability,

The critical value is the Z score and a probability equal to 0.8. From the Standardized Normal distribution table we find that the critical value is 0.84.

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

Solution: We need to use t distribution table,

For 95% confidence with df = n-1 = 24

Here, t = 2.064

For 96% confidence with df = n-1 = 24

Here, t = 2.172

For 99% confidence with df = n-1 = 24

Here, t = 2.797

Q 24**)**AGovernment 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 260days

Hint:

rcode🡪pt(tscore,df)

df🡪 degrees of freedom

Solution:

We have, = 260, = 270, n = 18, s = 90

Computing the test statistic,

Now, the test statistic t = -0.4714, degrees of freedom = n – 1 = 17

Under area under normal curve table displays the cumulative probability for -0.471 is 0.319. Hence, if the true bulb life were 270 days, there is a 31.9% chance that the average bulb life for 18 randomly selected bulbs would be less than or equal to 260 days.