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

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

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

1. Ans: HHH
2. HHT
3. HTH
4. THH
5. HTT
6. THT
7. TTH
8. TTT

Out of these outcomes, there are three outcomes that have two heads and one tail: HHT, HTH, and THH.

Therefore, the probability of getting two heads and one tail is 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 2and 3

Ans: ) When two dice are rolled, the minimum possible sum is 2 (when both dice show a 1), and the maximum possible sum is 12 (when both dice show a 6).

It is not possible to get a sum of 1 with two six-sided dice because the minimum value on each die is 1. The lowest possible sum is 2.

So, the probability of getting a sum of 1 with two dice is zero.

1. Less than or equal to 4

Ans) Total possible outcomes when rolling two dice = 6 \* 6 = 36

These combinations are (1,1), (1,2),(1,3), (2,1),(2,2) and (3,1)

So, Probability = 6/ 36 =>1 / 6

1. Sum is divisible by 2and 3

Ans) (2, 4), (4, 2), (3, 3), (6, 6), and (1, 5), (5, 1).

Probability = 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?

Ans: Probability of drawing a non-blue ball on the first draw = (Number of non-blue balls) / (Total number of balls) = 5/7

Probability of drawing a non-blue ball on the second draw, given that the first ball was non-blue = (Number of remaining non-blue balls) / (Total number of remaining balls) = 4/6 = 2/3

Overall Probability = (Probability of Step 1) \* (Probability of Step 2) = (5/7) \* (2/3) = 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 = 0.2

Ans: ) Expected Number of Candies =

(Probability of Child A \* Candies Count for Child A) +

(Probability of Child B \* Candies Count for Child B) +

(Probability of Child C \* Candies Count for Child C) +

(Probability of Child D \* Candies Count for Child D) +

(Probability of Child E \* Candies Count for Child E) +

(Probability of Child F \* Candies Count for Child F)

Expected Number of Candies = (0.015 \* 1) + (0.20 \* 4) + (0.65 \* 3) + (0.005 \* 5) + (0.01 \* 6) + (0.120 \* 2)

Now we should calculate each part of the equation:

Expected Number of Candies = 0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.24

Expected Number of Candies = 4.095

So, the expected number of candies selected by child is 4.095..

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: used Q7.csv file**

Pointsmean**=**3**.**5965625

Points median =3.695

Points var=0.28588

Points std=0.53467

Points max=4.93

Points min=2.76

scoremean**=**3**.**2172500000

score median =3.325

score var = 0.957378

score std = 0.978457

score max = 5.424

score min = 1.513

weighmean**=** 17.8487500

weigh median = 317.71

weigh var = 3.193166

weigh std = 1.786943

weigh max = 22.9

weigh min = 14.5

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: List the weights of the patients: X = [108, 110, 123, 134, 135, 145, 167, 187, 199]

P= X/ n is the formula

1193 / 9

P = 132.5

So, the expected value of the weight of a randomly chosen patient is approximately 132.5 pounds.

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:** ) For Cars Speed Skewness value= -0.12 and Kurtosis value= 0.81

For Cars Distance Skewness value = 0.81 and Kurtosis value = 0.41

ForSPSkewness **=** 1**.**61kurtosis **=** 0**.**95

For WT Skewness = 1.61 Kurtosis = 0.95

**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.

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: ) Interval = Mean ± (Z \* (Standard Deviation / √Sample Size))

For a 94% confidence interval:

Confidence Interval = 200 ± (1.88 \* (30 / √2,000))=199.29

For a 98% confidence interval:

Confidence Interval = 200 ± (2.33 \* (30 / √2,000))=198.79

For a 96% confidence interval:

Confidence Interval = 200 ± (1.75 \* (30 / √2,000))=199.00

**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?

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

Ans: In summary, when the mean and median are equal, the data follows a symmetric distribution, and there is no skewness in the data.

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

Ans: Skewness and tail is towards Right

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

Ans: ) Skewness and tail is towards left

Q16) What does positive kurtosis value indicates for adata ?

Ans: **Leptokurtic:** Distributions with positive kurtosis are referred to as "leptokurtic" distributions. They have a sharper and more pronounced peak around the mean.

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

Ans: ) **Flattened Peak:** Negative kurtosis is often associated with a distribution that is less peaked (has a flatter central peak) than a normal distribution. This indicates that the data values are more spread out and less concentrated around the mean.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

Ans: \*The above Boxplot is not normally distributed the median is towards the higher value

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

\*The Inter Quantile Range = 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: \*First there are no outliers.

\*Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to 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 ofCars for the below cases.

MPG<- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

ans: Ans)

P(MPG<40)= 0.729 P(MPG>38)= 0.34

P(20<MPG<50)=1.2430968797327491e-05

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

Ans: MGP 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 does not follows normal distribution

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

Ans: Lower Limit = Sample Mean - (Z-Score \* (Standard Deviation / √Sample Size))

Upper Limit = Sample Mean + (Z-Score \* (Standard Deviation / √Sample Size))

Z score for 60% Confidence Interval = -1.6449

Z score for 60% Confidence Interval=-1.8808

Z score for 60% Confidence Interval=-0.8416

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

Ans: Lower Limit = Sample Mean - (t-Score \* (Standard Deviation / √Sample Size))

Upper Limit = Sample Mean + (t-Score \* (Standard Deviation / √Sample Size))

T score for 95% Confidence Interval = -2.0639

T score for 96% Confidence Interval =2.16658

T score for 99% Confidence Interval = -2.7969

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

ans: sample\_mean - 260 # Sample mean

population\_mean - 270 # Claimed population mean

sample\_size - 18 # Sample size

standard\_deviation - 90 # Standard deviation of the sample

Calculate the t-score

standard\_error - standard\_deviation / sqrt(sample\_size)

t\_score - (sample\_mean - population\_mean) / standard\_error

* t\_score = -0.471 pt(t\_score, df = 17) 0.32 = 32%......