|  |  |
| --- | --- |
| 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 | Nominal |
| High School Class Ranking | Ordinal |
| 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 | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

Solution: When three coins are tossed,

The total number of possible combinations are 2^3 = 8.

These combinations are HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

The number of combinations which have two heads and one tail are: HHT, HTH, TTH =3

Therefore, the probability of getting two heads and one tails is:

* P (Two heads and One Tail) = Number of desired outcomes/ Number of Total Outcomes = 3/8 = 0.375

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

Solution:

1. So, when we roll two dice there are 6 × 6 = 36 outcomes.

There is no any moment when the sum of dice is 1 Therefore, the

* P(sum is equal to 1)= 0/36=0

1. Less than or equal to 4 The set of possible outcomes when we roll a die are {1, 2, 3, 4, 5, 6} So, when we roll two dice there are 6 × 6 = 36 outcomes.

When we roll two dice, the possibility of getting number 4 is (1, 3), (2, 2), and (3, 1).

So, The number of favorable outcomes = 3 Total number of outcomes = 36 Therefore,

* P(sum is Less than or equal to 4)=The number of favorable outcomes / Total number of possibilities = 3 / 36 = 1/12.

1. Sum is divisible by 2 and 3 The set of possible outcomes when we roll a die are {1, 2, 3, 4, 5, 6}

So, when we roll two dice there are 6 × 6 = 36 outcomes.

When we roll two dice, the possibility of getting Sum should be divisible by both 2 and 3 is (1, 5), (2, 4), (3, 3), (4, 2), (5, 1), and (6, 6).

So, The number of favorable outcomes = 6 Total number of outcomes = 36

* Therefore, P(Sum is divisible by 2 and 3)=The number of favorable outcomes / Total number of possibilities = 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?

Solution:

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 =7C2 =(7×6)/(2×1) =21

Let E = Event of drawing 2 balls, none of which is blue.

* So, n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls. =5C2 =(5×4)/ (2×1) =10 Therefore, 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 A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

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

Solution: Expected number of candies for a randomly selected child = 1 \* 0.015 + 4\*0.20 + 3 0.65 + 50.005 + 6 \*0.01 + 2 \* 0.12 = 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24 = 3.090

* So, The expected number of candies for randomly selected child is 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**

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: Expected Value = ∑ (probability \* Value) = ∑ P(x). E(x) There are 9

patients, Probability of selecting each patient = 1/9 Ex: 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 SP and Weight(WT) Use Q9\_b.csv**

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



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

Mean > Median.

* We have outliers on the higher side.



* Answer: 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?

Solution:

The information given is: • Sample mean of . • Sample standard deviation of . • Sample size of . The interval is:

• In which t is the critical value for the two-tailed confidence interval.

Considering a 94% confidence level, using a calculator, with 200 - 1 = 199 df, the critical value is t = 1.8916,

hence:

The 94% confidence interval is (198.73, 201.27). Considering a 96% confidence

level, using a calculator, with 200 - 1 = 199 df, the critical value is t = 2.0673,

hence:

The 96% confidence interval is (198.61, 201.39). Considering a 98% confidence

level, using a calculator, with 200 - 1 = 199 df, the critical value is t = 2.3452,

hence:

* The 98% confidence interval is (198.43, 201.57).

**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: Find mean, median, variance, standard deviation.

1. Mean =41, Median =40.5, Variance =25.52 and Standard Deviation =5.052)
2. What can we say about the student marks?

* => 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?

Answer: No skewness is present we have a perfect symmetrical distribution.

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

Answer: Skewness and tail is towards Right.

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

Answer: Skewness and tail is towards left.

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

Answer: Positive kurtosis means the curve is more peaked and it is Leptokurtic.

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

Answer: Negative Kurtosis means the curve will be flatter and broader.

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



Answer: The above Boxplot is not normally distributed the median is towards the higher value.

What is nature of skewness of the data?

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

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

* 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.



Answer: 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)

c. P (20<MPG<50)

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

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

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

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 :

t\_score = (x - pop mean) / (sample standard daviation / square root of sample size) (260-270) /90/np.sqrt(18))

* t\_score = -0.471

=> stats.t.cdf(t\_score, df = 17)

0.32 = 32%