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

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

When three coins are tossed the total number of possible combinations are

2x2x2 = 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 which makes them 3 in number.

P (two heads and one tail) = 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

Ans:

a) The minimum sum can be obtained when rolling a two dice is 2

the probability of obtaining a sum of 1 is 0

OUTPUT =0

b) (1,1),(2,1),(1,2),(2,2),(3,1),(1,3)

=6/36 =1/6=0.16666

OUTPUT=1/6=0.16666

(c) (1,5),(2,4),(4,2),(5,1),(6,6),(3,3) =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?

Total balls = 7

No of ways of drawings 2balls out of 7 = 7x6/2 = 21

No of ways of drawings 2balls out of 5 = 5x4/2 = 10

P (none of the balls drawn is blue) = 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

Ans; To calculate the expected no of candies for a randomly selected child,we need to multiply each candy count by it’s corresponding probability

The expected no of candies(E)can be calculated as

E=(1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120)

OUTPUT=3.09

The expected number of candies for a randomly selected child is approximately 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.

**Points Score Weigh**

|  |  |  |  |
| --- | --- | --- | --- |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |

**Points Score Weigh P S W**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sum | 377.6872 |  | 342.2023 |  | 9391.468927 | Range | 2.17 | -1.845 | -1.7 |
| Variance | 11.80272 |  | 10.69382 |  | 293.483404 |  |  |  |  |
| SD | 3.435509 |  | 3.270141 |  | 17.13135733 |  |  |  |  |

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

Expected value = Sum of weight /no of patients

=108+ 110+ 123+ 134+ 135+ 145+ 167+ 187+ 199/9

=145.33 pounds

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

**Cars speed and distance**

**Use Q9\_a.csv**

**(IN EXCEL Q9 a,)**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**(IN EXCEL Q9 ANS)**

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



A) It is the graph of the histogram between chickweight$weight & frequency

In the graph we can see the graph is at high peak towards “Right side” so we can get the positive value

When the mean is greater than the median, the distribution is positively skewed (or) Right-skewed ,the tail is in the right side of the histogram



INFERENCE:-The box plot shows that data is “Negative skewed data”. Since the median of the data is near to the minimum value. And for these data, there are outliers also.

**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?

(IN EXCEL Q11 ANS)

**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?

Ans;

1)

* Mean = (34 + 36 + 36 + 38 + 38 + 39 + 39 + 40 + 40 + 41 + 41 + 41 + 41 + 42 + 42 + 45 + 49 + 56) / 18 Mean = 729 / 18 Mean = 40.5
* 34, 36, 36, 38, 38, 39, 39, 40, 40, 41, 41, 41, 41, 42, 42, 45, 49, 56

The median = (40 + 41) / 2

Median = 81 / 2

Median = 40.5

* Variance = [(34 - 40.5)² + (36 - 40.5)² + ... + (56 - 40.5)²] / 18

Variance ≈ 80.25

* Standard Deviation = √Variance

Standard Deviation ≈ √80.25

Standard Deviation ≈ 8.96

2)Now, regarding what we can say about the student's marks:

The mean and median are both 40.5, which indicates that the scores are roughly centered on this value.

The standard deviation of approximately 8.96 suggests that there is some variation in the scores, but it's not extremely high.

The range of scores is from 34 to 56, with most scores clustered between 36 and 49.

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

when the mean and median are equal in a dataset, it indicates that the data is symmetric and has no skewness.

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

when the mean is greater than the median, it suggests a positively skewed distribution with a tail on the right side.

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

when the median is greater than the mean, it suggests a negatively skewed distribution with a tail on the left side.

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

A positive kurtosis value indicates that the data has a distribution with heavier tails and a more peaked central peak, implying a higher likelihood of extreme values or outliers.

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

A negative kurtosis value indicates that the probability distribution of a dataset has lighter tails and is less peaked in the center compared to a normal distribution, potentially suggesting a lower probability of extreme values or outliers.

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



What can we say about the distribution of the data?

Median is closer to upper quartile

Mean <Median

What is nature of skewness of the data?

Negatively skewed

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

8

Q19) Comment on the below Boxplot visualizations?



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

A)

1)

Inference:

In the boxplot (1) the data move towards the “Left skewed data”

The variance is also “Less” ,compared to the boxplot(2)

2)

In the boxplot(2) the data move towards the “Normal data”

The data is plotted at the “Median”

The variance is also “very high” ,compared to the boxplot(1)

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)
  3. P (20<MPG<50)

(IN PYTHON FILE Q20)

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 scores of 90% confidence interval,94% confidence interval, 60% confidence interval

A) We need to calculate the Z-score for 90%confidence interval from the Z-score table

90% confidence interval =1.645 (from Z-score table we will get these values)

60% confidence interval =0.842 (from Z-score table we will get these values)

94% confidence interval for that we need to find Area

Area= 1+ confidence interval/2 =(1 + 94 /100) /2

= 1 + 0.94 / 2

=0.9700

We need to find the 0.9700 in the Z- score table

=1.8+0.08

OUTPUT = 1.88 (approximately)

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

A) Sample size(n)= 25

Degree of freedom =sample size -1

=24

From the t- score table ,we need to find the 95% confidence interval

95 % confidence interval =2.064

99 % confidence interval=2.797

96 % confidence interval=2.171

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

To find the probability that 18 randomly selected bulbs would have an average life of no more than 260 days if the CEO's claim were true, we can use a one-sample t-test.

Given:

* Sample size (n) = 18
* Sample mean = 260 days
* Population mean= 270 days (CEO's claim)
* Standard deviation (s) = 90 days

Then, we find the degrees of freedom (df), which is equal to *n*−1.

Finally, we find the probability using the cumulative distribution function of the t-distribution.

Let's calculate:

t=260−270/90/√18

t= −10/90/√18

t= −10/90/4.2426

t= −10/21.2132

t≈−0.4714

Now, degrees of freedom (df) = 18 - 1 = 17. 17

Using the R code provided, we can find the probability:

Pt (-0.4714,17)7)

The above R code will give us the probability that a t-distribution with 17 degrees of freedom is less than -0.4714.

This will give us the probability that 18 randomly selected bulbs would have an average life of no more than 260 days if the CEO's claim were true.

Top of Form