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

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

Ans – HHT+HTH+THH

=1/8+1/8+1/8

=**3/8**

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

1. Equal to 1

= **0** , Because non of the outcome of rolling two dice become a one

1. Less than or equal to 4

={(1,3),(2,2),(3,1),(1,1),(1,2),(2,1)}

=6/36 **= 1/6**

1. Sum is divisible by 2 and 3

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

=**5/36**

When we solve it has separately then P{sum}=1/2+1/3 = 5/6

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?

n(S)=No of ways of drawing 2 balls out of 7

Total No. of ball =(2+3+2)=7

=7\*6/2\*1

=21

n(E)=Drawing 2 balls , non of which is blue

remaining balls is = 2+3

=5\*4/2\* =10

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 | 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-

Child A:Probability=0.015

Total event =Interested event/Expected candies =?

So formula:

Probability = No. of interested events/Total no of events

No. of interested events= Probability\* Total no of events

 =0.015\*1=**0.015**

Similarly, for Child B

 =0.20\*4=**0.80**

Similarly other= 1 \* 0.015 + 4\*0.20 + 3 \*0.65 + 5\*0.005 + 6 \*0.01 + 2 \* 0.12

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

**= 3.090**

Expected No. of candies for child is **3**

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=https://github.com/VAIBHAVGHUGARE01/Assignments\_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?

Answer-

Expected Value  =  ∑ ( probability  \* Value )

So,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

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)

= 143.331

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**

**Answer= https://github.com/VAIBHAVGHUGARE01/Assignments\_file**

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



* The maximum of the data points is present in the range of 50-100 with Frequency 200. Weight=400 having minimum frequency around 0-10.Its Heavily right skewed



* Median is less than mean.we have outliers on upper side of boxplot shows in the form of hollow circles 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- Using the **t-distribution**, it is found that:

* The **94%** confidence interval is **(198.73, 201.27).**
* The **96%** confidence interval is **(198.61, 201.39).**
* The **98%** confidence interval is **(198.43, 201.57).**

We are given the standard deviation for the sample, which is why the **t-distribution** is used to solve this question.

The **information given** is:

* Sample **mean**of \overline{x} = 200.
* Sample **standard deviation** of s = 30.
* Sample **size**of n = 2000.

The **interval** is:

\overline{x} \pm t(s)/(√(n))

* 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:

\overline{x} - t(s)/(√(n)) = 200 - 1.8916(30)/(√(2000)) = 198.73

\overline{x} + t(s)/(√(n)) = 200 + 1.8916(30)/(√(2000)) = 201.27

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:

\overline{x} - t(s)/(√(n)) = 200 - 2.0673(30)/(√(2000)) = 198.61

\overline{x} + t(s)/(√(n)) = 200 + 2.0673(30)/(√(2000)) = 201.39

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:

\overline{x} - t(s)/(√(n)) = 200 - 2.3452(30)/(√(2000)) = 198.43

\overline{x} + t(s)/(√(n)) = 200 + 2.3452(30)/(√(2000)) = 201.57

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?

Answer

1)mean: 41

Median: 40.5

Variance: 25.23

Standard deviation: 5.052

2)Maximum no.of students got 35-40 range score.

No one get 50-55 range score.

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

Answer-Zero skewness or Normalized Skewness

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

Ans- Right skewed

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

Ans- Left skewness

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

Answer-

Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. Extremely positive kurtosis indicates a distribution where more numbers are located in the tails of the distribution instead of around the mean

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

Answer-

A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution. For example, data that follow a beta distribution with first and second shape parameters equal to 2 have a negative kurtosis value.

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



What can we say about the distribution of the data?

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

What is nature of skewness of the data?

Answer- 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)?

Ans-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 250 to 275 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 of Cars for the below cases.

MPG <- Cars$MPG

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

c. P (20<MPG<50)

Answer**- https://github.com/VAIBHAVGHUGARE01/Assignments\_file**

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

Answer-

**To calculate the Z-score for a 90%** confidence interval, we need to find the critical Z-value that corresponds to the level of confidence. In this case, we want to find the Z-score that leaves 5% of the distribution in the tails (because 90% of the distribution is within the interval, and 100% - 90% = 10% is split equally between the two tails).

Here are the steps to calculate the Z-score for a 90% confidence interval:

C = 90%.

* Calculate the level of significance (α), which is the complement of the confidence level

α = 100% - C = 100% - 90% = 10%

* Divide α by 2 because we have two tails (one in each direction):

=α/2 = 10% / 2

= 5%

Find the Z-score corresponding to the α/2 level of significance. we can look this up in a standard normal distribution table or use a calculator or software that provides this information.

he Z-score for a 90% confidence interval is approximately 1.645. This means that 90% of the data falls within ±1.645 standard deviations from the mean when assuming a normal distribution.

Top of Form

**For a 94% confidence interval**, the Z-score corresponding to the 3% level of significance is approximately 1.8808.

So, the Z-score for a 94% confidence interval is approximately 1.8808. This means that 94% of the data falls within ±1.8808 standard deviations from the mean when assuming a normal distribution.

**For a 60% confidence interval**, the Z-score corresponding to the 20% level of significance is approximately 0.8416.

So, the Z-score for a 60% confidence interval is approximately 0.8416. This means that 60% of the data falls within ±0.8416 standard deviations from the mean when assuming a normal distribution.

Top of Form

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 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

Answer=**https://github.com/VAIBHAVGHUGARE01/Assignments\_file**