|  |  |
| --- | --- |
| 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 | Nominal |
| Level of Agreement | ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| 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?

Sol:- set of total events={HHH,HHT,THH,HTH,TTT,TTH,THT,TTH}

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 2and 3

|  |
| --- |
| sol:- set of total events=36 |
| a) P(equal to 1) = 0/36 |
|  |
| b)P(less than or equal to 4)=3/36 |
|  |
| c)P(sum is divisble by 2 and 3)=4/36 |

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?

|  |
| --- |
| sol:- Total outcomes for drawn two balls=7C2=21 |
| P(selecting two balls)=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

Sol:-   
Expected number of candies for a randomly selected child

=  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

=  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**

Sol:-

|  |  |  |  |
| --- | --- | --- | --- |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variance | 0.285881 | 0.957379 | 3.193166 |
| Standard deviation | 0.534679 | 0.978457 | 1.786943 |
| Range | 2.17 | 3.911 | 8.4 |

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?

Sol:- Expected Value  =  ∑ ( probability  \* Value )

 ∑ P(x).E(x)

There are 9 Patients

Probability of selecting each patient = 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

=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

Sol:-

|  |  |  |
| --- | --- | --- |
|  | Speed | Distance |
| skewness | -0.11751 | 0.806895 |
| Kurtosis | -0.50899 | 0.405053 |

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| Skewness | 1.61145 | -0.61475 |
| kurtosis | 2.977329 | 0.950291 |

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



Sol:- Right side skewed or positively skewed



Sol:- The interface for this box plot is positively skewed

**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?

Sol - We are working with a 94 % confidence level

Find standard error. The standard error (SE) of the mean is:

SE = s

√ n = 30

√ 2000 = 0.670

= Compute alpha (α): α = 1 - (confidence level / 100) = 0.9933

= Find the critical probability (p\*): p\* = 1 - α/2 = 1 – 0.9933/2 = 0.4966

= find the degree of freedom (df): df = n-1 = 2000 – 1 = 1999

= The critical value is the t score having 1999 degrees of freedom and a

probability equal to 0.4966

= critical value is -0.009

= critical value \* standard error = - 0.009 \* 0.94 = -0.00846

= - 0.009 \* 0.98 = -0.00882

= -0.009 \* 0.96 = - 0.00846

**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?

Sol:- The scores are in uniformly distribution data in Ascending order

Mean 41

Median 40.5

Variance 25.52

Standard deviation 5.05

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

Sol:- Normalised Skewness

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

Sol:- Right Skewed

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

Sol:- left skewed

Q16) What does positive kurtosis value indicates for adata ?

Sol:- Sharp peak in the plot. less gap between tails to x-axis

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

Sol:- Border peak under the curve and more gap between tails and x-axis

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



What can we say about the distribution of the data?

Sol:- The data is distributed in De-assigned format

What is nature of skewness of the data?

Sol:- Left side skewed

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

Sol:- Q3-Q1

= 18-10

= 8 is IQR

Q19) Comment on the below Boxplot visualizations?



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

Sol:- The box plot 1 designed with range = 3 , The second one range is = 1.5

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)

Sol:- P(MPG>38)=33/81

* 1. P(MPG<40)

Sol:- P(MPG<40)=61/81

c. P (20<MPG<50)

Sol:- P(20<MPG<50)=69/81

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Sol:- we just calculate the below parameters to find normal distribution

|  |  |
| --- | --- |
| Mean | 34.42208 |
| Median | 35.15273 |
| Mode | 29.62994 |
| Skew | -0.17795 |
| kurtosis | -0.61168 |

Its clear that MPG of cars is normal distribution with left skewness.

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

Dataset: wc-at.csv

Sol:- for At mean> median, right whisker is larger than left whisker, data is positively skewed.

For waist mean> median, both the whisker are of same lenght, median is slightly shifted towards left. Data is fairly symetrically distributed.

|  |  |  |
| --- | --- | --- |
|  | waist | At |
| mean | 91.90183 | 101.894 |
| median | 90.8 | 96.54 |
| mode | 94.5 | 121 |
| skew | 0.134056 | 0.584869 |
| kurtosis | -1.10267 | -0.28558 |

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

Sol:- For 90% confidence interval:

We have the significance level at 5 % ( as it is a two tailed test) that is:

α = 5 % = 0.05

z at α = 0.05 = 1.645. (from the z table)

For 94 % confidence interval:

We have the significance level at 3 % ( as it is a two tailed test)

α = 3 % = 0.03

z at α = 0.03 = 1.555( from the z table)

For 60 % confidence interval

We have the significance level at 20 % ( as it is a two tailed test)

α =20 % = 0.2=0.253

Therefore, we get that the z score at 90 % confidence interval is 1.645, at 94 % confidence interval is 1.555 and at 60 % confidence interval is 0.253.

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

Sol:- To compute the 95% confidence interval, start by computing the mean and standard error: M = (2 + 3 + 5 + 6 + 9)/5 = 5. σM = = 1.118. Z.95 can be found using the normal distribution calculator and specifying that the shaded area is 0.95 and indicating that you want the area to be between the cutoff points

Confidence Level z

0.90 1.645

0.92 1.75

0.95 1.96

0.96 2.05

With a 90 percent confidence interval, you have a 10 percent chance of being wrong. A 99 percent confidence interval would be wider than a 95 percent confidence interval (for example, plus or minus 4.5 percent instead of 3.5 percent).

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

Sol:-

t - statistics for the data is given as follows:

https://tex.z-dn.net/?f=t%3D%5Cdfrac%7Bx-%5Cmu%7D%7B%5Cfrac%7Bs%7D%7B%5Csqrt%20n%7D%7D

x = mean of the sample of bulbs =  260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

https://tex.z-dn.net/?f=t%3D%5Cdfrac%7B260-270%7D%7B%5Cfrac%7B90%7D%7B%5Csqrt%2018%7D%7D

https://tex.z-dn.net/?f=t%20%3D%20%5Cdfrac%7B-10%7D%7B%5Cfrac%7B90%7D%7B3%5Csqrt%202%7D%7D

https://tex.z-dn.net/?f=t%20%3D%20%5Cdfrac%7B-10%7D%7B%5Cfrac%7B30%7D%7B%5Csqrt%202%7D%7D

https://tex.z-dn.net/?f=t%20%3D%20%5Cdfrac%7B-1%20%5Ctimes%20%5Csqrt%202%7D%7B3%7D

t = - 0.471

For probability calculations, the number of degrees of freedom is n - 1, so here you need the t-distribution with 17 degrees of freedom.

The probability that **t < - 0.471 with 17 degrees of freedom** assuming the population mean is true, the t-value is less than the t-value obtained With 17 degrees of freedom and a t score of - 0.471, the probability of the bulbs lasting less than 260 days on average of **0.3218** assuming the mean life of the bulbs is 300 days.