|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| 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 |
| 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 | Interval |
| 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 | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| 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?

= If 3 coins are tossed, the possible probability we find is

HHH, HHT, HTH, THH, TTH, THT, HTT, TTT

The number we have two heads and one tail is

* HHT, HTH, THH
* 3

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

1. Equal to 1

* 2 dice are rolled when possibilities is

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

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

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

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

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

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

Sum = 36

Then we have sum equal to 1 = 0

1. Less than or equal to 4

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

= 6/36

= 0.17

1. Sum is divisible by 2 and 3

= {(1,5)(2,4),(3,3),(4,2).(5,2),(6,6)}

= 6/36

=0.17

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 case

2+3+2 = 7

= 7i/2i5i

7\*6/2

7\*3

= 21

And probability drawn blue balls

5i/2i3i

5\*4/2

5\*2

=10

* the probability that 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 – 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

= 3.09

Expected number of candies for a randomly selected child = 3.09

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviati\ & 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

* Mean - Points 3.596563

Score 3.217250

Weigh 17.848750

* Median - Points 3.695

Score 3.325

Weigh 17.710

* Min - Points 2.76

Score 1.513

Weigh 14.5

* Variance - Points 0.285881

Score 0.957379

Weigh 3.193166

* Std - Points 0.534679

Score 0.978457

Weigh 1.786943

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?

* Total patients is = 9

= 1/9

= (1/9) ( 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9) (1308)

=9/1308

=145

the Expected Value of the Weight of that patient = 145

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

**Cars speed and distance**

**Use Q9\_a.csv**

* **Skewness\_x=0.0**

**Y=** **0.6985354731356995**

* **Kurtosis\_x=** **0.8573388203017832**

**Y =** **0.14864276382638986**

**SP and Weight(WT)**

**Use Q9\_b.csv**

* **Skewness\_x=0.0**

**Y=0.0**

* **Kurtosis\_ x=** **8.35612089936015**

**Y=** **8.35612089936015**

**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 - confidence interval - 94%,98%,96%

standard deviation of the sample – 30

sample - 2,000

sample weighs – 200

\overline{x} \pm t\frac{s}{\sqrt{n}}

Considering a 94% confidence, 200 - 1 = 199 df, the critical value is 1.8916,

\overline{x} - t\frac{s}{\sqrt{n}} = 200 - 1.8916\frac{30}{\sqrt{2000}} = 198.73

\overline{x} + t\frac{s}{\sqrt{n}} = 200 + 1.8916\frac{30}{\sqrt{2000}} = 201.27

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

Considering a 96% confidence, 200 - 1 = 199 df, the critical value is 2.0673,

\overline{x} - t\frac{s}{\sqrt{n}} = 200 - 2.0673\frac{30}{\sqrt{2000}} = 198.61

\overline{x} + t\frac{s}{\sqrt{n}} = 200 + 2.0673\frac{30}{\sqrt{2000}} = 201.39

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

Considering a 98% confidence, 200 - 1 = 199 df, the critical value is 2.3452,

\overline{x} - t\frac{s}{\sqrt{n}} = 200 - 2.3452\frac{30}{\sqrt{2000}} = 198.43

\overline{x} + t\frac{s}{\sqrt{n}} = 200 + 2.3452\frac{30}{\sqrt{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?

* mean – 41
* median – 40.5
* variance - 25.529411764705884
* standard deviation – 5.05

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

* No skewness is present we have a perfect symmetrical distribution

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

* Skewness and tail is towards Right

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

* Skewness and tail is towards left

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

* The positive kurtosis value indicates that the distribution has heavier tails than the normal distribution.

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

* Negative Kurtosis means the curve will be flatter and broader

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



What can we say about the distribution of the data?

* 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.

Ans – the 1st here are no outliers, 2nd 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 of Cars for the below cases.

MPG <- Cars$MPG

* 1. P(MPG>38)

= 33/81

* 1. P(MPG<40)

= 61/81

c. 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

And - MPG Mean - 34.422075728024666

MPG Median - 35.15272697

MPG Mode - 29.629936

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

Dataset: wc-at.csv

Ans- df1['Waist'].mode() – 94.5

df1['Waist'].mean() - 91.90183486238533

df1['Waist'].median() – 98.0

df1['AT'].mode() – 121.0

df1['AT'].mean() - 101.89403669724771

df1['AT'].median() – 96.54

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

Ans - 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 from the z table will be:

z = 1.645.

For 94 % confidence interval, we get:

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

that is:

α = 3 % = 0.03

z at α = 0.03 from the z table will be:

z = 1.555.

For 60 % confidence interval, we get:

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

that is:

α =20 % = 0.2

z at α = 0.2 from the z table will be:

z = 0.253

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

Ans - 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 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

* 260-270/90

== -10

== 90/18

= -10/21.213203

== -0.4714045