|  |  |
| --- | --- |
| 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) | Ratio |
| 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 | Ordinal |

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

=When coins are tossed together the total number of out come is 8 and The possible out come for two head and one tail are 3. Hence probability is 3/8.

===================================================================

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

1. Equal to 1

= The probability of getting a sum 1 is zero. Probability is 0

1. Less than or equal to 4 = N (Event (Sum is less than or equal to

4)) / N (Event (Two dice rolled))

= 6 / 36 = 1/6 = 0.166 = 16.66%

1. Sum is divisible by 2 and 3 = N (Event (Sum is divisible by 2 and 3)) / N

(Event (Two dice rolled))

= 6 / 36 = 1/6 = 0.16 = 16.66%

===================================================================

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?

= Probability = favorable outcome/total outcome

No of balls = 2+3+2=7

7C2 =(7\*6)/(2\*1)=42/2 = 21

The no of balls other than blue is 5

===================================================================

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

= 1\*0.015+4\*0.20+3\*80.65+5\*0.005+6\*0.01+2\*0.120

the Expected number of candies for a randomly selected child is = 3.090

===================================================================

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**

* **POINTS**

**🡪** Mean for Points

Sum of all data points / Total no of data points = 115.09/33 = 3.59

* Median for points

(n+1/2) = 3.695

* Mode for points = Which repeats more no of times = 3.92
* Standard deviation = S = = 0.526
* Variance for point = s=1n−1n∑i=1(xi−¯x)2 = 0.285
* Range for point = Maximum-Minimum = 4.93 – 2.76 = 2.17
* **Score**

**🡪** Mean for Score

Sum of all data Score / Total no of data points = 3.212

* Median for Score

(n+1/2) = 3.325

* Mode for Score = Which repeats more no of times = 3.44
* Standard deviation = S = = 0.963
* Variance for Score = s=1n−1n∑i=1(xi−¯x)2 = 0.957
* Range for Score = Maximum-Minimum = 5.424 – 1.513 = 3.911
* **Weight**

**🡪** Mean for weight

Sum of all data weight / Total no of data points = 17.848

* Median for weight

(n+1/2) = 17.71

* Mode for weight = Which repeats more no of times = 17.02
* Standard deviation = S = = 1.758
* Variance for weight = s=1n−1n∑i=1(xi−¯x)2 = 3.193
* Range for Sc weight = Maximum-Minimum = 14.5 – 22.9 = 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?

= Sum (x\* probability of x) = (1/9)(108)+ (1/9)(110)+ (1/9)(123)+ (1/9)(134)…… (1/9)(199) =143.33

The expected value of the weight of the patient is 143.33

===================================================================

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

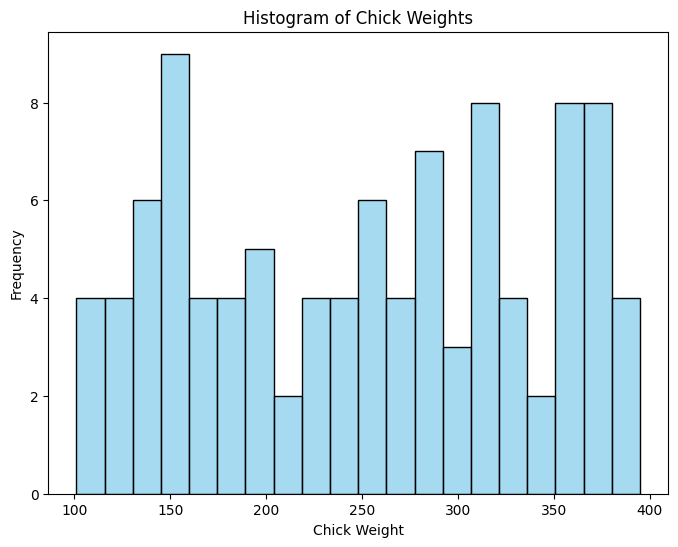
**Cars speed and distance**

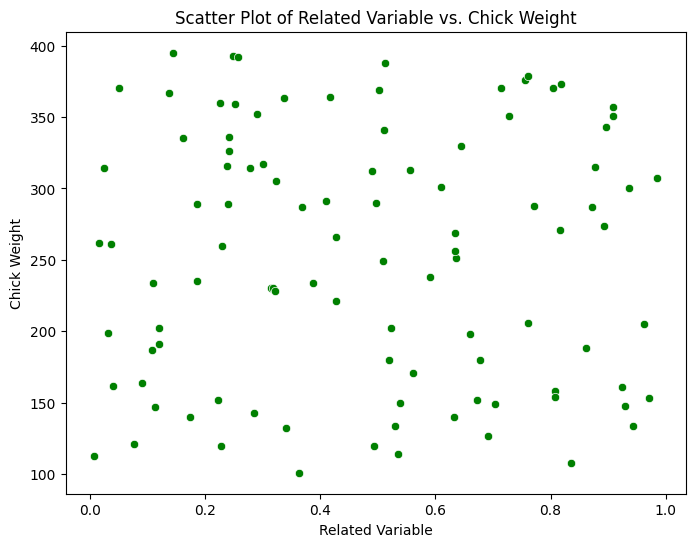
* Skewness for cars is 0.00 🡪 Kurtosis for cars is -1.200
* Skewness for speed is -0.117 🡪 Kurtosis for speed is -0.577
* Skewness for distance is 0.806 🡪 Kurtosis for distance is 0.248

===================================================================

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







==================================================================

**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?

== Confidence interval = x



* For a 94% confidence interval



* For 98% confidence interval



* For 96% confidence interval



===================================================================

**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 34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56 = **41**

18

= Median 34, 36, 36, 38, 38, 39, 39, 40, 40, 41, 41, 41, 41, 42, 42, 45, 49, 56 Median is **40**

= Variance is (1/18) \* [(34-41.11)² + (36-41.11)² + … + (56-41.11)²] = **67.7**

= Standard Deviation = √67.76 = **8.23**

===================================================================

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

* Positive Skewness

==================================================================

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

* Negatively skewed

===================================================================

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



🡪 Positive Kurtosis indicates that the distribution has a higher peak than a normal distribution.

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

🡪 Negative Kurtosis indicates that the distribution has a Lower peak than a normal distribution.

===================================================================

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



What can we say about the distribution of the data?

🡪 Not normally distributed

What is nature of skewness of the data?

* Negative skewness

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

* IQR = [Q3 – Q1] = [18 – 10] = 8
* The IQR data is about 8

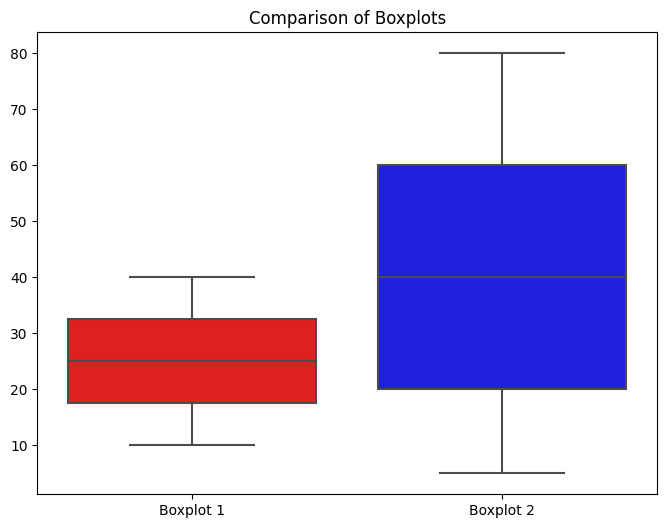
================================================================

Q19) Comment on the below Boxplot visualizations?:

* The red boxplot (1) has a smaller range (from 250 to 275) compared to the blue boxplot (2) (from 225 to 300). This suggests that the data in the red boxplot is less spread out.
* The blue boxplot (2) is taller than the red one, indicating a larger interquartile range (IQR), which is the range within which the middle 50% of your data falls.
* It’s hard to comment on the median (the line inside the box), outliers, and other specifics without more detailed markings on these boxplots.



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

* 

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.

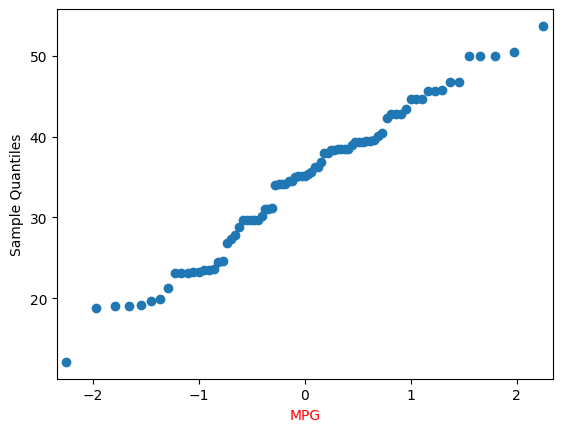
MPG <- Cars$MPG

* 1. P(MPG>38) 🡪 0.347
  2. P(MPG<40) 🡪 0.729
  3. P (20<MPG<50) 🡪 1.243

===================================================================

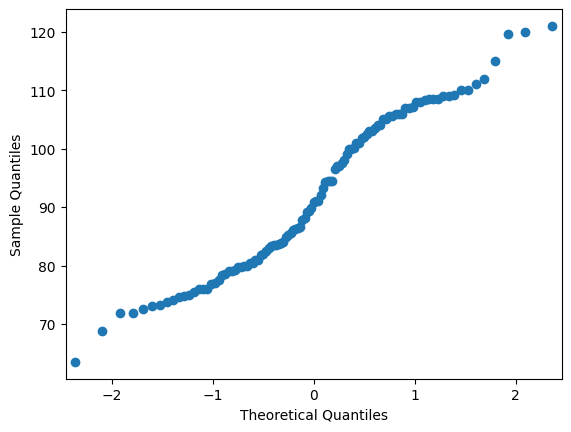
Q 21) Check whether the data follows normal distribution

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

* To find Z scores of a confidence interval

Z = X – μ / sigma

For the 90% confidence interval the critical value of Z is 1.645.

For the 94% confidence interval the critical value of Z is 1881.

For the 60% confidence interval the critical value of Z is 0.842.

X = μ Zsigma

Consider if have a population mean of 50 and a standard deviation of 10 we want to find Z score of 90 %

X = 50 1.645 \* 10

= 66.45

Therefore the Z score of a 90% confidence interval is

Z = 66.45 – 50 = 1.645

10

===================================================================

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

* To find t scores of a confidence interval

t = X – μ

s/

To find t score of a confidence interval, we need find X

X = μ t s/

Consider if have a population mean of 50 and a standard deviation of 10 we want to find t score of 95 % with the sample size of 25

X = 50 2.064 10/

= 54.13

Therefore the t score of a 95% confidence interval is

t = 54.13 - 50 = 2.064

10/ ===================================================================

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

* 0.321

===================================================================