|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | D |
| Results of rolling a dice | D |
| Weight of a person | C |
| Weight of Gold | C |
| Distance between two places | C |
| Length of a leaf | C |
| Dog's weight | C |
| Blue Color | D |
| Number of kids | D |
| Number of tickets in Indian railways | D |
| Number of times married | D |
| Gender (Male or Female) | D |

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 | N |
| High School Class Ranking | O |
| Celsius Temperature | I |
| Weight | I |
| Hair Color | N |
| Socioeconomic Status | O |
| Fahrenheit Temperature | I |
| Height | I |
| Type of living accommodation | O |
| Level of Agreement | O |
| IQ(Intelligence Scale) | O |
| Sales Figures | I |
| Blood Group | N |
| Time Of Day | R |
| Time on a Clock with Hands | O |
| Number of Children | O |
| Religious Preference | R |
| Barometer Pressure | I |
| SAT Scores | O |
| Years of Education | O |

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

Ans: - 3/8

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

1. Equal to 1
2. 0
3. Less than or equal to 4
4. 6/36
5. Sum is divisible by 2 and 3
6. 24/36= 2/3

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?

1. 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:- 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.

|  |  |  |  |
| --- | --- | --- | --- |
| particulars | points | score | weigh |
| MEAN | 3.596563 | 3.217250 | 17.848750 |
| MEADIAN | 3.695 | 3.325 | 17.710 |
| MODE  AMC Javelin  Cadillac Fleetwood | 3.07  3.92 | 3.44  NaN | 17.02  18.90 |
| VARIANCE | 0.285881 | 0.957379 | 3.193166 |
| STANDARD DEVIATION | 0.534679 | 0.978457 | 1.786943 |
| RANGE | 2.76-4.93 | 1.513-5.424 | 14.5-22.9 |



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:- 145.33

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

**a) Cars speed and distance b) SP and Weight(WT)**

|  |  |  |
| --- | --- | --- |
| **DATA** | **SKEWNESS** | **KURTOSIS** |
| 1. **SPEED** 2. **DIST** | **-0.117510**  **0.806895** | **-0.508994**  **0.405053** |
| 1. **SP**   **B)WT** | **1.611450**  **-0.614753** | **2.977329**  **0.950291** |

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



Ans:-

It shows the frequency distribution of values of a single variable and how many observations take the value within a certain interval

Divide the values into bins and show a bar plot of the no.of object in each bin

The hight of each bar indiacates the np.of observations.



Ans:-

It a graph that gives you a good indications of how the values in the data are spread out.

Median( Q2/50th percentage) the middle values of the dataset

First Quartile (Q1/25th percentage)

**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 ?

94% :- (198.75 -201.25)

98% :- (198.63 -201.37)

99%:- (198.44-201.56)

**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?
3. MEAN -41
4. MEDIAN – 40.5
5. VARIANCE – 25.52
6. STANDARD DEVIATION- 5.05

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

1. Zero skewness

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

1. Positive skewness

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

1. Negative skewness

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

1. Peakness of the data is high closed to mean and variance is very less

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

1. The data is spreaded . the variance is high from the means

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

1. Asymmetrically data has distributed

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

1. Inter quartile range ( Q3-Q1)  
     
     
     
   Q19) Comment on the below Boxplot visualizations?



Ans:- Boxplot 1:

In the given data has distributed in a boxplot. The mean of the data is 262.5. the IQR ranges between the 255-280. In the boxplot there is no out layers

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

Ans:-

The mean of the given data is 260. The IQR9inter quartile range) from 225-305. There is no out layers in the given boxplot.

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)

Ans:- pnorm(mpg>38)

[1] 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447

[9] 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447

[17] 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447

[25] 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.5000000 0.8413447 0.8413447

[33] 0.5000000 0.5000000 0.8413447 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000

[41] 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000

[49] 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000

[57] 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000

[65] 0.5000000 0.8413447 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000

[73] 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000

[81] 0.5000000

* 1. P(MPG<40)

Ans:- pnorm(mpg<40)

[1] 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000

[9] 0.5000000 0.5000000 0.5000000 0.8413447 0.8413447 0.5000000 0.5000000 0.5000000

[17] 0.5000000 0.8413447 0.5000000 0.8413447 0.8413447 0.5000000 0.8413447 0.5000000

[25] 0.8413447 0.8413447 0.8413447 0.8413447 0.5000000 0.8413447 0.8413447 0.8413447

[33] 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447

[41] 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447

[49] 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447

[57] 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447

[65] 0.8413447 0.5000000 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447

[73] 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447 0.8413447

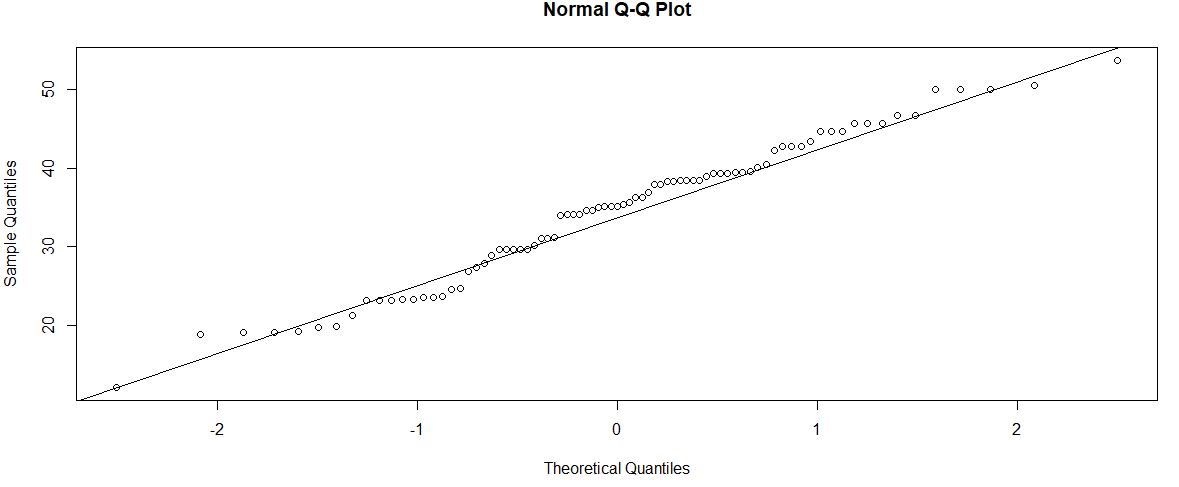
[81] 0.8413447

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

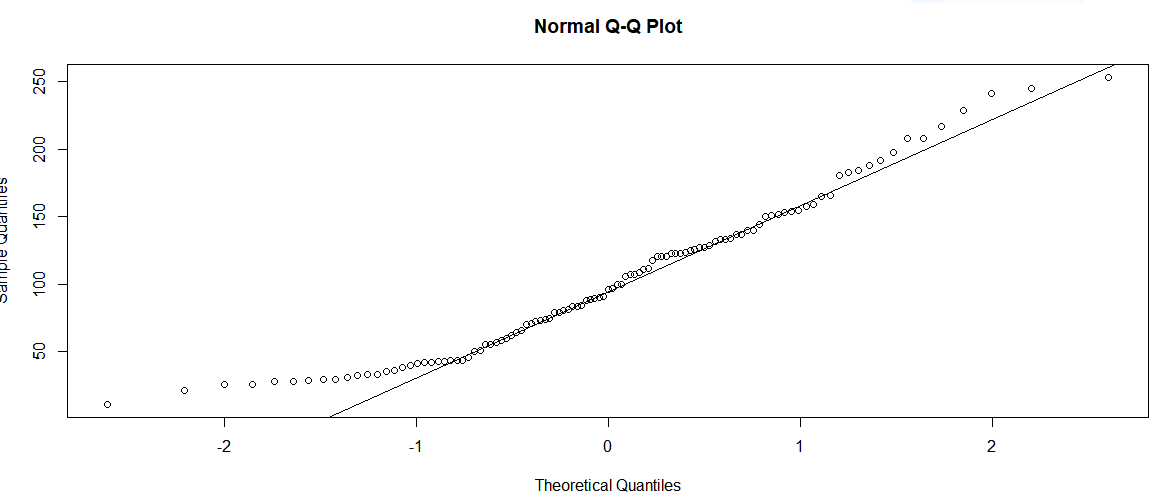
Ans: the given data set MPG is following the normal distribution



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

Dataset: wc-at.csv

Ans:- the given data set (AT)following the normal distribution



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

Ans:- Z scores at 94%=1.88

90%=1.64

60%=0.84

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

Ans:- T scores at 95%= 2.06

96%=2.17

99%=2.79

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

Ans:-