Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| 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 | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Nominal |

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) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

**Ans**: when three coins are tossed, total possible outcomes are 8 ( 2^3 ).

The Sample space is : { HHH, HHT,HTT,HTH,THT,TTH,THH,TTT}

Number of favorable outcomes are 3 (HHT,HTH,THH)

So, the probability of obtaining two heads and one tail is:

Number of favorable outcomes / Total number of possible outcomes

= 3 / 8 = 0.375

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 2 and 3

**Ans** : When two dice are rolled , the total possible outcomes are 36 (6^2 ).

1. The number of favorable outcomes are 0

So, probability that sum is equal to 1 is

Number of favorable outcomes / Total number of possible outcomes

= 0 / 36

= 0

1. The number of favorable outcomes are 6 { (1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1) }

So, probability that sum is equal to less than or equal to 4 is

Number of favorable outcomes / Total number of possible outcomes

= 6 / 36

= 1 / 6

1. To get a sum divisible by both 2 and 3, the sum must be 6 or 12.

The favorable outcomes are: (1, 5), (2, 4), (3, 3), (4, 2), (5, 1), (6, 6).

So, the number of favorable outcomes are 6.

Probability = number of favorable outcomes/total number of outcomes

= 6 / 36

= 1 / 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?

**Ans**: Total number of balls are 2 + 3 + 2 = 7

Two balls are drawn at random

So, total number of possible cases are 7C2 =7! / (2! 5! ) = 21

Now, the number of none of the balls drawn is blue are 2 red and 3 green

So, non-blue balls are 5

And two balls drawn at random, so possible cases are 5C2 = 5! /2! 3! = 10

The probability that none of the balls drawn is blue = 10/21 = 0.4762

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**: The Expected value = E(X)=∑(x⋅P(x))

E = (1 \* 0.015) + (4 \* 0.20) + (3 \* 0.65) + (5 \* 0.005) + (6 \* 0.01) + (2 \* 0.120)

E = 0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.24

E = 3.09

The expected number of candies for a randomly selected child is approximately 3.09.

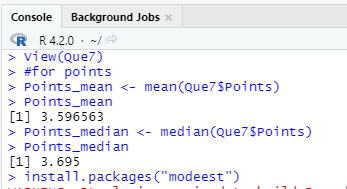
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

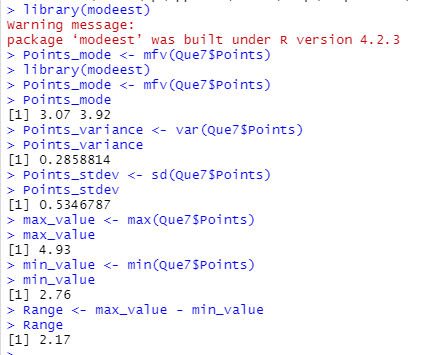
* For Points,Score,Weigh

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

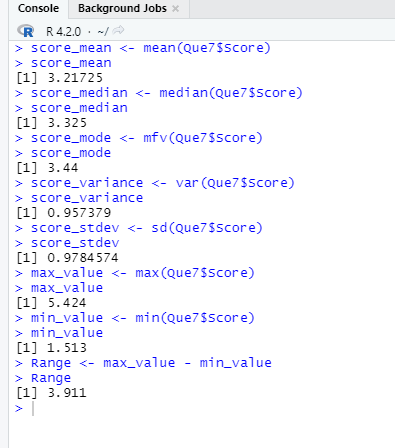
**Use Q7.csv file**

**Ans:** For points:

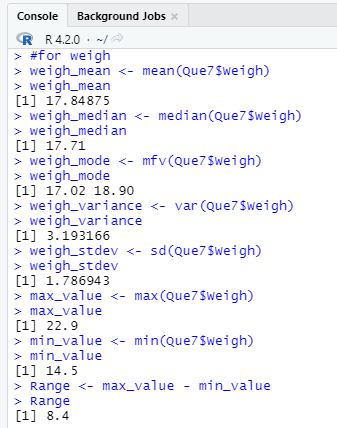




For score:



For weigh:



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** : The Expected value = E(X)=∑(x⋅P(x))

Since there are 9 patients, each patient has an equal probability of being chosen, which is 1/9.

|  |  |  |
| --- | --- | --- |
| x | P(x) | x\* p(x) |
| 108 | 1/9 | 108/9 |
| 110 | 1/9 | 110/9 |
| 123 | 1/9 | 123/9 |
| 134 | 1/9 | 134/9 |
| 135 | 1/9 | 135/9 |
| 145 | 1/9 | 145/9 |
| 167 | 1/9 | 167/9 |
| 187 | 1/9 | 187/9 |
| 199 | 1/9 | 199/9 |
| Total |  | 1308/9 |

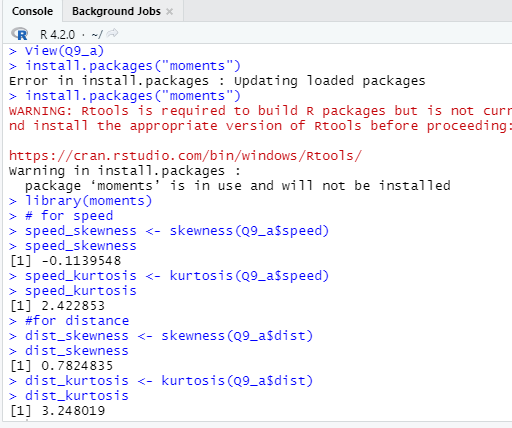
E(X) = 1308 /9 ≅ 145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**

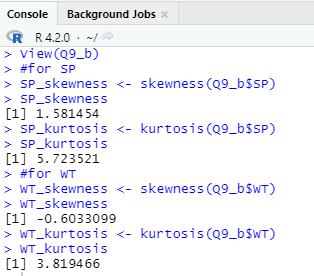
**Ans:**



**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:**



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



Ans : Inference drawn from the histogram : Histogram shape looks like positively skewed as more data is on the right side of the data. The most chick weight lies between 50-100.

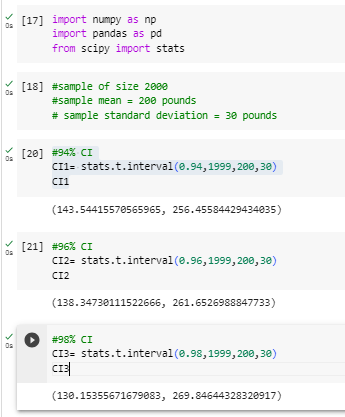


**Ans**: Inference drawn from the boxplot : There are 7 values which are greater than Upper fence.

Upper fence is Q3 + (1.5\* IQR), So, outliers are present. Also, median is close to the quartile 1, this means data is more on the right side. Hence right 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?

Ans:

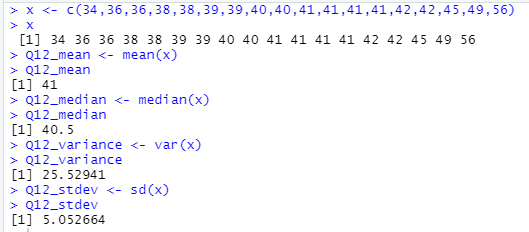


**Q12)** Below are the scores obtained by a student in tests

**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?

**Ans**: **1)**



1. The mean of the student marks is 41 and median is 40.5

From this we can observe that, the median is slightly less than the mean. This suggests that the scores might be slightly skewed to the right.

The variance is approximately 25.5294 and standard deviation is 5.0527. This indicates that there is some variability in the student's scores.

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

**Ans**: When the mean and median of data are equal, it indicates that the data is approximately symmetrically distributed. In a symmetric distribution, the values are evenly distributed around the center point, which is the mean and the median. There is no skewness in the data.

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

**Ans**: When the mean is greater than the median, it indicates that the data is right-skewed (positively skewed). In a right-skewed distribution, the tail on the right side is longer and most of the data is concentrated on the left side.

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

**Ans**: When the median is greater than the mean, it indicates that the data is left-skewed (negatively skewed). In a left-skewed distribution, the tail on the left side is longer and most of the data is concentrated on the right side.

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

**Ans**: Positive kurtosis indicates that the data has more extreme values and is more peaked or concentrated around the mean compared to a normal distribution. This type of distribution is called a leptokurtic distribution.

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

**Ans**: Negative kurtosis indicates that the data has fewer extreme values and is more spread out and less peaked around the mean compared to a normal distribution. . This type of distribution is called a platykurtic distribution.

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



From boxplot, we can observe that min value = 1 , max value = 19 , Q1 = 10 , median = 15.1 , Q3 = 18.2

What can we say about the distribution of the data?

**Ans** : In this case, with Q1 at 10 and Q3 at 18.2, and the median at 15.1, the data seems to be approximately symmetrically distributed.

What is nature of skewness of the data?

**Ans**: The whisker is more on the left side. So, more data is concentrated on right side. This means it is left skewed (negatively skewed).

What will be the IQR of the data (approximately)?   
**Ans**: IQR is Interquartile range which is Q3 – Q1

Here Q1 = 10 and Q3 = 18.2 approximately

So, IQR = Q3 – Q1 = 18.2 – 10 = 8.2 (approximately)

Q19) Comment on the below Boxplot visualizations?



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

**Ans**:

In both the box plots, the median value is same i.e. approx. 263

In box plot 1 , the Q1 is approx. 260 , Q3 is approx. 280 , min value is 240 and max value is 287.5

In box plot 2, the Q1 is approx. 225, Q3 is approx. 306.25, min value is 198 and max value is 338 approx.

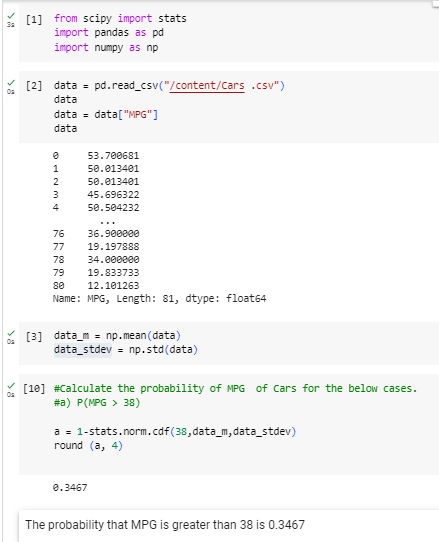
Q 20) Calculate probability from the given dataset for the below cases

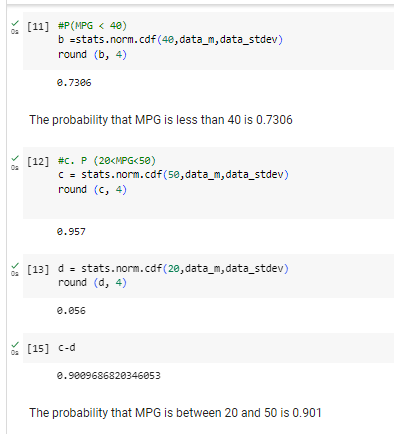
Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)



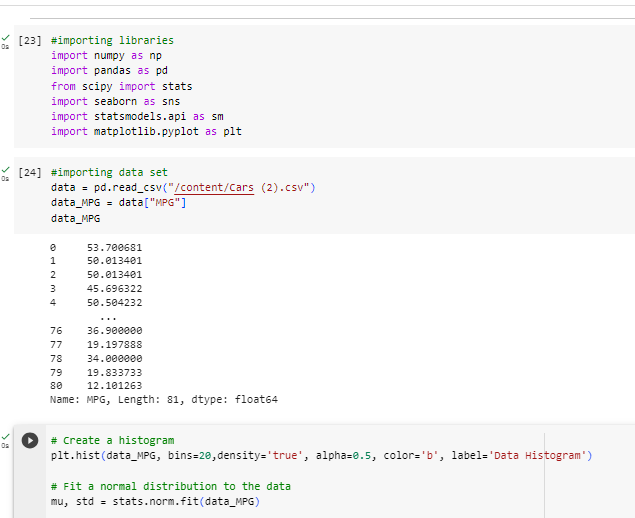


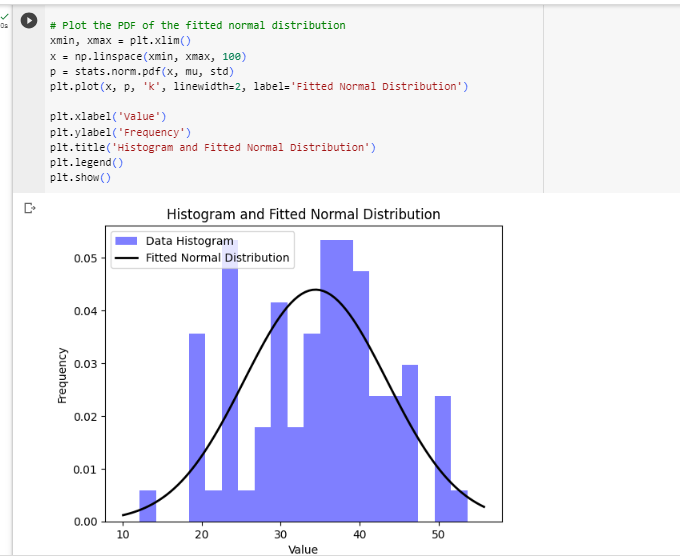
Q 21) Check whether the data follows normal distribution

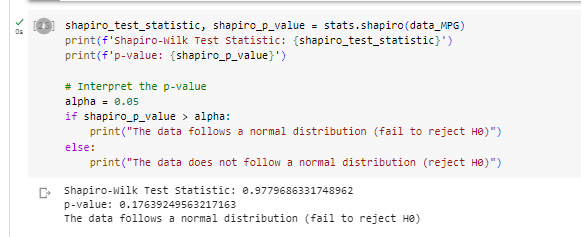
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:



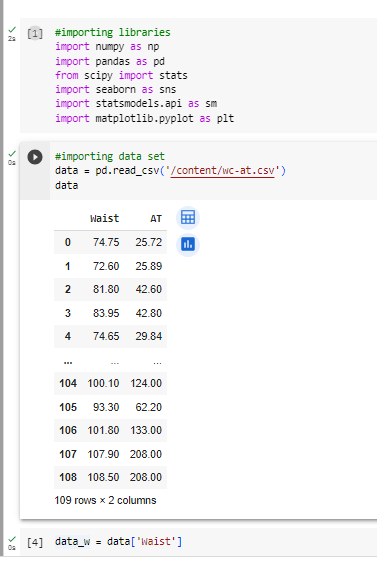


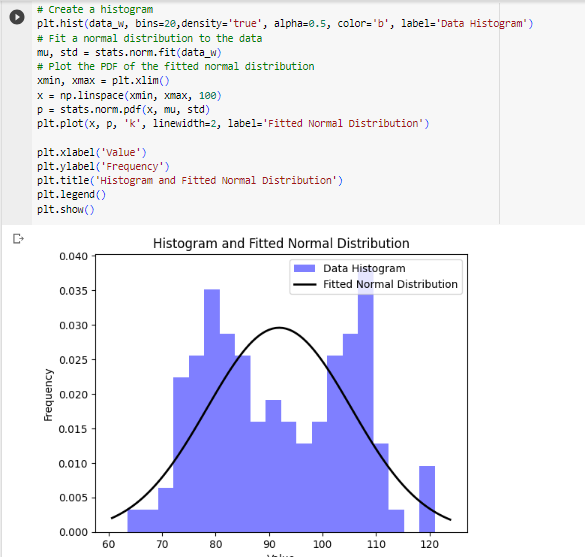


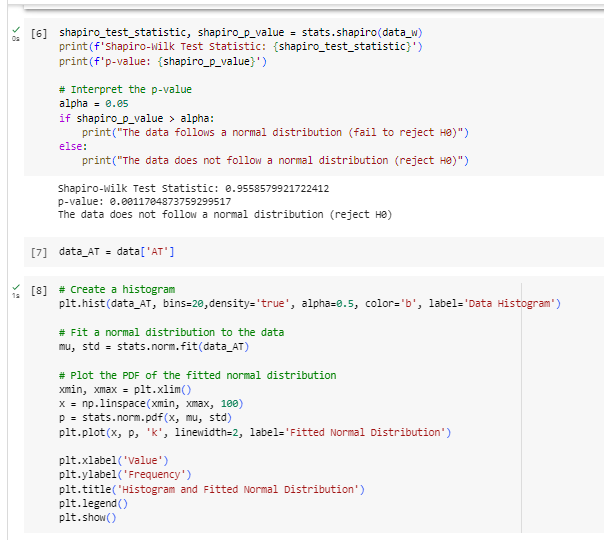
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

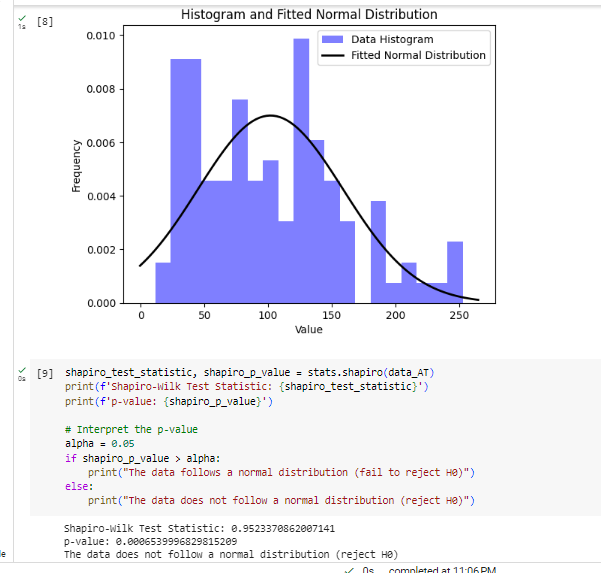
Dataset: wc-at.csv

Ans:



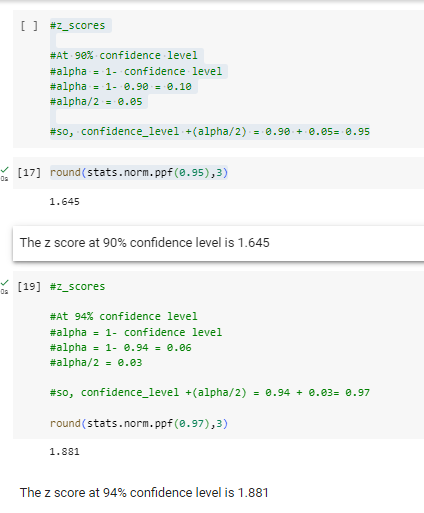


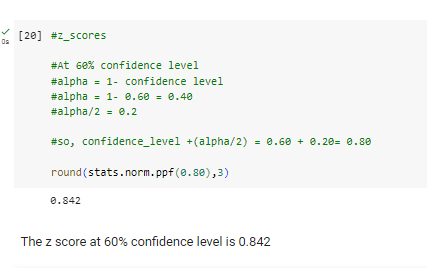




Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

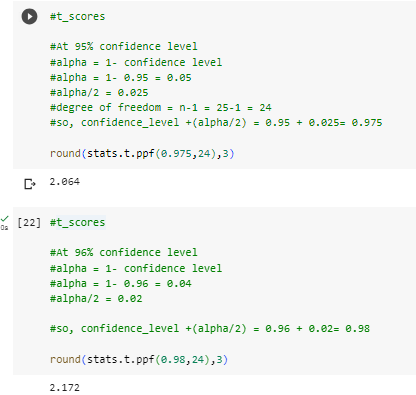
Ans:





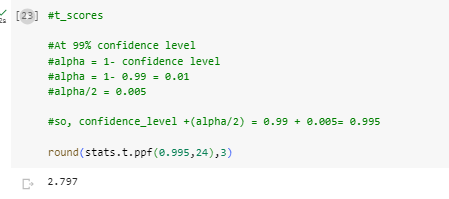
Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

Ans:



The t score at 95% confidence level is 2.064

The t score at 96% confidence level is 2.172



The t score at 99% confidence level is 2.797

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

Ans:

