**Q1) Identify the Data type for the Following:**

|  |  |
| --- | --- |
| **Activity** | **Data Type** |
| Number of beatings from Wife | Discrete data |
| Results of rolling a dice | Discrete data |
| 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 data |
| Number of kids | Discrete data |
| Number of tickets in Indian railways | Discrete data |
| Number of times married | Discrete data |
| Gender (Male or Female) | Discrete data |

**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 scale |
| Weight | Ratio scale |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval scale |
| Height | Ratio scale |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval scale |
| Sales Figures | Ratio scale |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval scale |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio scale |
| SAT Scores | Interval scale |
| Years of Education | Ordinal |

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

**Answer-**

Three coins are tossed; there are 2\*2\*2 = 8 possible outcomes.

S= (HHH, HHT, HTH, HTT, THH, THT, TTH, TTT)

We get 3 result out of 8 with two heads and a tail (HHT, HTH, THH)

Probability of getting two heads and a tail is 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

**Answer –**

When 2 dice are rolled, number of sample space are 62 = 36

{(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)}

n(S) = 36

Then

1. Equal to1 = P(sum is equal to1)/ n(S) = 0/36 = 0
2. Less than or equal to 4 = P(Less than or equal to 4)/ n(S) = 6/36 = 0.1666
3. Sum is divisible by 2 and 3 = P(Sum is divisible by 2 and 3)/ n(S) = 6/36 = 0.1388

**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?**

**Answer-**

1. If we drown 2 balls randomly out of 2 red, 3 green and 2 blue = n (S)

n(S) = 7!/(2!\*(7-2)!) =7C 2 =21

1. None of the balls drawn is blue = n(A) = 5!/(2!(5-2)!) = 5C2 = 10
2. Probability that none of the balls drawn is blue = n(A)/n(S) = 10/21 = 0.4761

**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

**Answer-**

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

probabilities of count of candies for children 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.

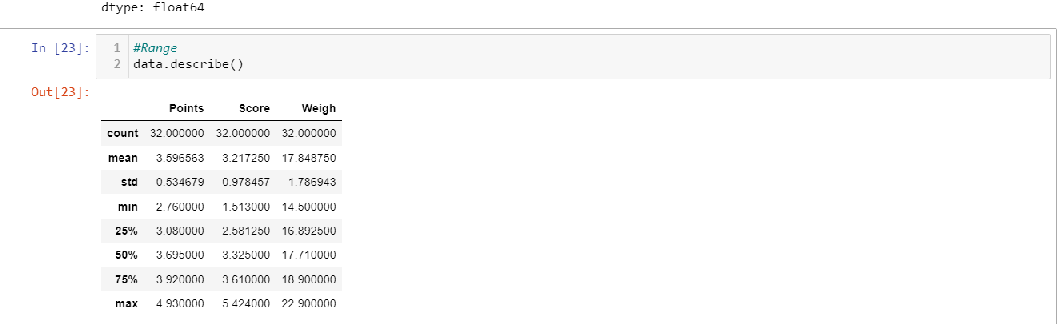
**Use Q7.csv file**

**Answer-**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| **Mean** | 3.597 | 3.217 | 17.849 |
| **Median** | 3.695 | 3.325 | 17.710 |
| **Mode** | 3.920 | 3.440 | 17.020 |
| **Variance** | 0.286 | 0.957 | 3.193 |
| **Standard Deviation** | 0.535 | 0.978 | 1.787 |
| **Range** | 2.76- 4.93 | 1.513 - 5.424 | 14.5 - 22.9 |

****

****

****

**Observation-** As we go through dataset their is non-symmetric data and mean, median and mode are not same so data is positively skewed**.**

**Q8) Calculate Expected Value for the problem below**

1. **The weights (X) of patients at a clinic (in pounds), are 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?

**Answer –**

Expected Value of the Weight of that patient= (108+110+123+134+135+145+167+187+199) / 9 = 1309/9 = 145.33

**OR**

We can calculate mean using R

I/p - wt<-c(108,110,123,134,135,145,167,187,199)

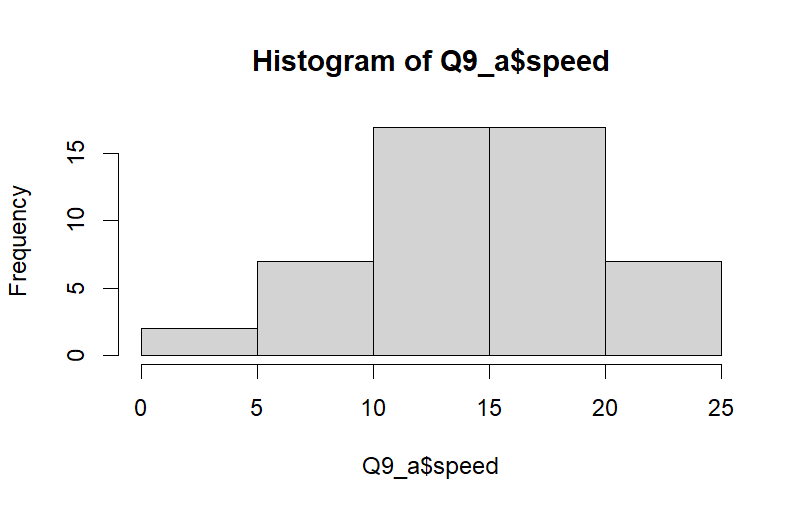
mean(wt)

O/p - 145.3333

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data,Cars speed and distance - Use Q9\_a.csv**

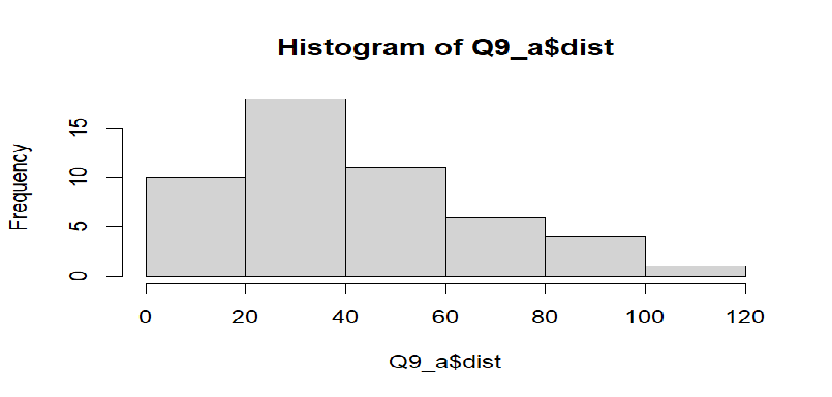
**Answer-**

|  |  |
| --- | --- |
| **Skewness of speed** | |
| **I/p** | **O/p** |
| Q9\_a<-read.csv("Q9\_a.csv") | > skewness(Q9\_a$speed) |
| Q9\_a | [1] -0.1139548 |
| skewness(Q9\_a$speed) |  |
| **Kurtosis of speed** | |
| **I/p** | **O/p** |
| kurtosis(Q9\_a$speed) | > kurtosis(Q9\_a$speed) |
|  | [1] 2.422853 |

****

**Observation** – As we can see data is left skewed means negatively skewed data and non-symmetric. There are 5 bins and data is widely spread so negative kurtosis we can see.

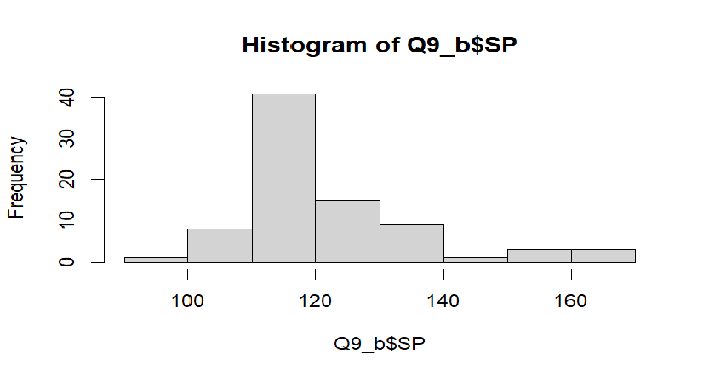
|  |  |
| --- | --- |
| **Skewness of distances** | |
| **I/p** | **O/p** |
| skewness(Q9\_a$dist) | > skewness(Q9\_a$dist) |
|  | [1] 0.7824835 |
| **Kurtosis of distances** | |
| **I/p** | **O/p** |
| kurtosis(Q9\_a$dist) | > kurtosis(Q9\_a$dist) |
|  | [1] 3.248019 |

****

**Observation** – As we can see data is right skewed means positively skewed data and data is symmetric. There are 6 bins and data having pick and tail at right side so positive kurtosis we can see.

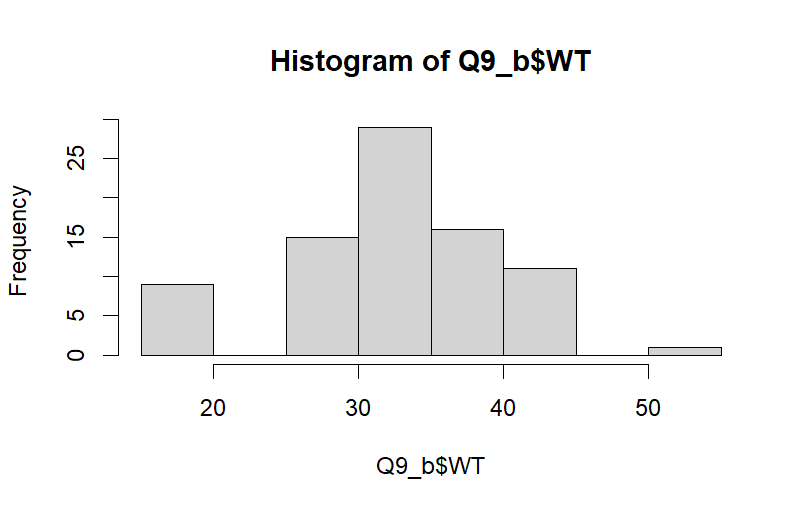
**SP and Weight(WT) - Use Q9\_b.csv**

|  |  |
| --- | --- |
| **Skewness of SP** | |
| **I/p** | **O/p** |
| Q9\_b<-read.csv("Q9\_b.csv") | > skewness(Q9\_b$SP) |
| Q9\_b | [1] 1.581454 |
| skewness(Q9\_b$SP) |  |
| **Kurtosis of SP** | |
| **I/p** | **O/p** |
| kurtosis(Q9\_b$SP) | > kurtosis(Q9\_b$SP) |
|  | [1] 5.723521 |

****

**Observation** – As we can see data is right skewed means positively skewed and symmetric data. There are 8 bins and data is sharp pick so positive kurtosis we can see.

|  |  |
| --- | --- |
| **Skewness of WT** | |
| **I/p** | **O/p** |
| skewness(Q9\_b$WT) | > skewness(Q9\_b$WT) |
|  | [1] -0.6033099 |
| **Kurtosis of WT** | |
| **I/p** | **O/p** |
| kurtosis(Q9\_b$WT) | > kurtosis(Q9\_b$WT) |
|  | [1] 3.819466 |



**Observation** – As we can see data is normally distributed. Negative skewness and it shows positive kurtosis that indicates heavy pick in data.

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

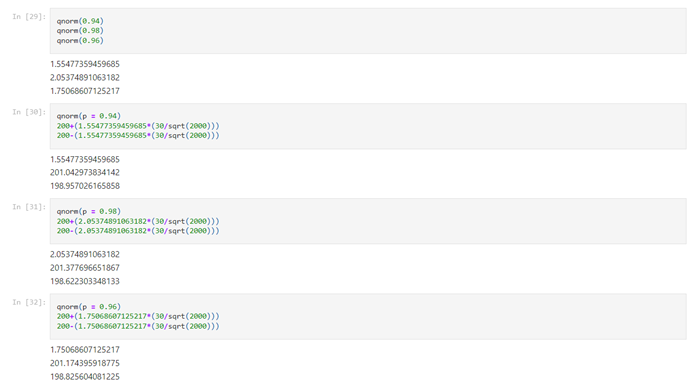


**Q11) Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?**

|  |  |  |  |
| --- | --- | --- | --- |
|  | 94% | 98% | 96% |
| Upper | 201.04 | 201.38 | 201.17 |
| Lower | 198.96 | 198.62 | 198.83 |

**Answer-**

**Python Code Screenshot-**



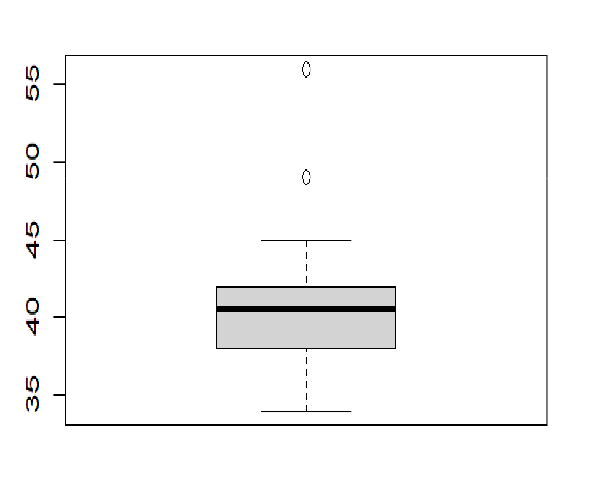
**Q12) Below are the scores obtained by a student in tests**

**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.**

|  |  |
| --- | --- |
| Code | Output |
| marks.mean() | 41 |
| marks.median() | 40.5 |
| marks.var() | 25.52941176 |
| marks.std() | 5.052663829 |

**Answer-**



1. **By using plots, we say about the student marks-**

**Answer-**

1. Data is not normally distributed
2. Data has outlier.
3. Majority of the students scored between 35 – 45 Marks

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

**Answer-**

If the mean, median and skewness is equal, then data distribution is symmetric and the mean equals the median and the skewness of the distribution is zero.

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

**Answer-**

If the median is less than the mean, then distribution of data is skewed to the right. If the distribution of data is on right, means data is positively skewed.

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

**Answer-**

If the mean is less than the median, then distribution of data is skewed to the left. If the distribution of data is on left, means data is negatively skewed.

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

**Answer-**

Positive values of kurtosis indicate that distribution is peaked and possesses thick tails.

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

**Answer-**

A negative kurtosis means that your distribution is flatter than a normal curve with the same mean and standard deviation.

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



1. **What can we say about the distribution of the data?**

**Answer-**

The data is not symmetric. Data is more concentrated towards right side and positively skewed.

1. **What is nature of skewness of the data?**

**Answer-**

The distribution of data is on left, means data is negatively skewed.

1. **What will be the IQR of the data (approximately)?**

**Answer-**

IQR data is 8 (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.**

**Answer-**

1. Data is Normally Distributed. No Outliers. Center around 262.5. Comparatively, first graph has less range
2. Data is Normally Distributed. No Outliers. Center around 262.5

Comparatively, second graph has more range

**Q20) 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

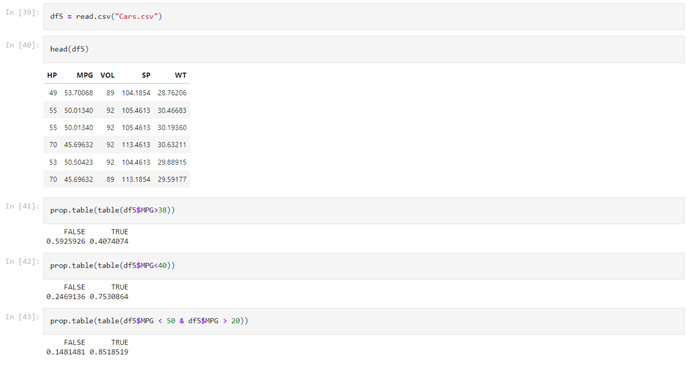
MPG <- Cars$MPG

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

**Answer-**

1. P(MPG>38) - probability is 0.4074074
2. P(MPG<40) - probability is 0.7530864
3. P (20<MPG<50) - probability is 0.8518519

**Python code screenshot-**



**Q21) Check whether the data follows normal distribution**

* + 1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

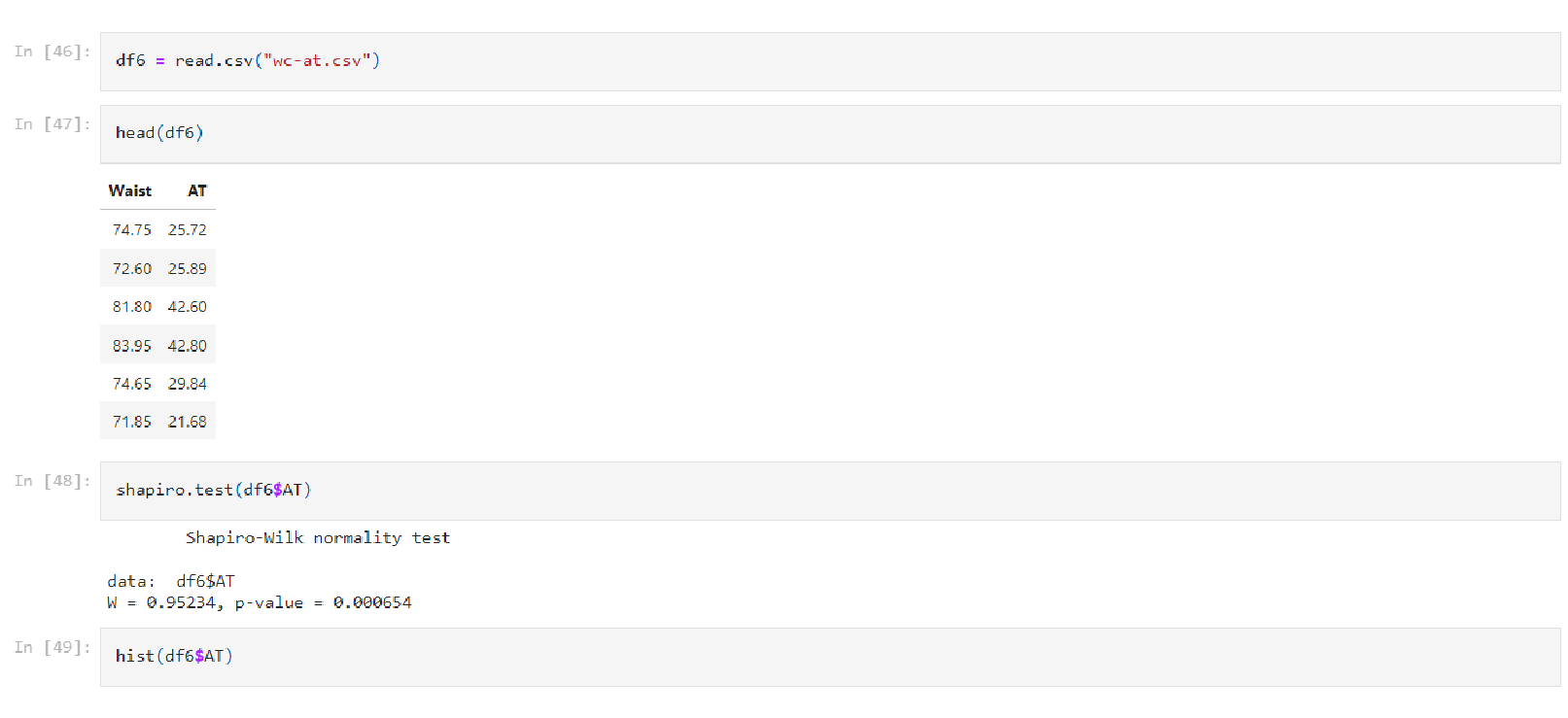
**Answer-** MPG is Normal Distribution

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

Dataset: wc-at.csv

**Answer**- Both AT and Waist doesn’t follow Normal Distribution.

**Python code screenshot-**

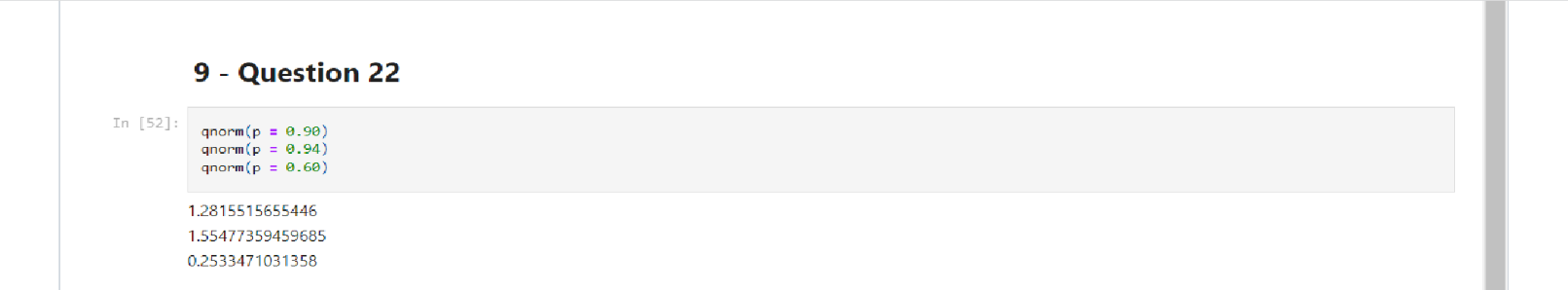


**Q22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval.**

**Answer-**

|  |  |
| --- | --- |
| 90% | ± 1.711 |
| 94% | ± 1.828 |
| 60% | ± 2.492 |

**Python code screenshot-**

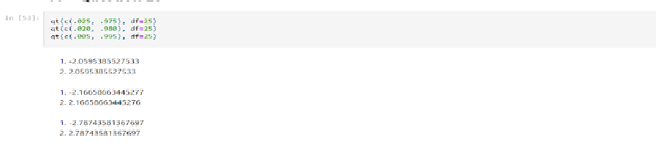


**Q23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25**

|  |  |
| --- | --- |
| 95% | ± 2.060 |
| 96% | ± 2.167 |
| 99% | ± 2.787 |

**Answer-**

**Python Code screenshot-**



**Q24) 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

**Answer-**

t-score = -0.4714,  
Degree of freedom = 17   
P(t) = 0.3 tscore 216725

