**Q 1)** 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)

Ans: To Calculate the Probability of the given data set, first we need to check whether it is following the Normal distribution or not.

By using the **shapiro.test**, we have calculated the p value ( 0.1764). As per the Hypothesis Testing P high H0 fly, we fail to reject the Null Hypothesis. That means it is distributed normally.

By looking the Q-QPlot also we can find the whether it is normal or not. I am adding the Visual diagram in the question.

**Mean = 34.42208**

**Standard deviation = 9.131445**

1. **P(MPG>38)** : Find the Probability greater than 38, so we have to calculate from MPG=39 to till end.

= ( 1- pnorm(39,34.42,913) ) = 0.307 = **30.7%**

1. **P(MPG<40)** : Find the Probability less than 39, so we have to calculate till MPG=39

= pnorm(39,34.42,913) = 0.692 = **69.2%**

1. **P(20<MPG<50)** : Find the Probability less than 20 and less than 50 , so we have to calculate from 21 to 49

= ( pnorm(49,34.42,913) - pnorm(21,34.42,913) ) = 0.8740 = **87.4%**

**Q 2)** Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Business Problem:**

1. **Objective:** MPG of Cars dataset follows Normal Distribution or not.

**Data Pre-processing:**

1. We have the good quality of data.

**Model Building:**

1. Hypothesis Testing:

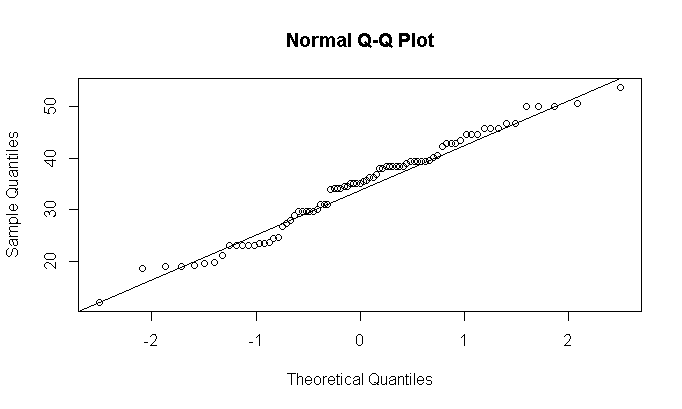
**H0 :** Assume MPG of Cars follows the Normal distribution.

**Ha:** Assume MPG of Cars NOT follows the Normal distribution

By using the **shapiro.test**, we have calculated the **p value** ( 0.1764). As per the Hypothesis Testing

P high H0 fly,

We fail to reject the Null Hypothesis. That means it is distributed normally.

1. By looking the Q-QPlot also we can find the whether it is normal or not. Anderson Darling test. By looking the Q-Q Plot , line is not covering the most of the data points. Still the P value is more than 0.05. 

**Conclusion:** MPG of Cars follows Normal Distribution.

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

Dataset: wc-at.csv

**Business Problem:**

1. Objective:Adipose Tissue (AT) and Waist Circumference(Waist) of wc-at dataset follows Normal Distribution or not.

**Data Pre-processing:**

1. We have the good quality of data for Adipose Tissue (AT) and Waist Circumference(Waist).

**Model Building:**

1. Hypothesis Testing: for Adipose Tissue (AT).

**H0 :** Assume for Adipose Tissue of wcat data follows the Normal distribution.

**Ha:** Assume for Adipose Tissue of wcat data NOT follows the Normal distribution

By using the **shapiro.test**, we have calculated the **p value** ( 0.000654). As per the Hypothesis Testing

P Low H0 Go,

We reject the Null Hypothesis. That means it is **not** distributed normally.

1. Hypothesis Testing: for Waist Circumference(Waist).

**H0 :** Assume for Waist Circumference of wcat data follows the Normal distribution.

**Ha:** Assume for Waist Circumference of wcat data NOT follows the Normal distribution

By using the **shapiro.test**, we have calculated the **p value** ( 0.00117). As per the Hypothesis Testing

P Low H0 Go,

We reject the Null Hypothesis. That means it is **not** distributed normally.

**Q 3)** Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**Ans :** Calculate the Z score by taking input as probability then we will use qnorm function in R.

|  |  |
| --- | --- |
| **Probability** | **Z score** |
| 94% | 1.880794 |
| 90% | 1.644854 |
| 60% | 0.841621 |

**Q 4)** Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

**Ans :** Calculate the t score by taking input as probability then we will use qt(probability , n-1) function in R.

|  |  |  |
| --- | --- | --- |
| **Probability** | **n=25** | **t score** |
| 95% | n-1=24 | 2.063899 |
| 96% | n-1=24 | 2.171545 |
| 99% | n-1=24 | 2.79694 |

Q 5**)**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:

 μ = 270 days

X̅ = 260 days

n = 18 bulbs

S = 90 days.

T score = ( X̅ - μ ) / ( S √n )

= 260 -270 / 90 / √18 = -10/90/4.2426 = **-0.4714**

Degree of freedom n-1 = 17

Probability is = pt(-0.4714,17) = 0.3216741 = **32.16%**

Q 6) The time required for servicing transmissions is normally distributed with μ = 45 minutes and σ = 8 minutes. The service manager plans to have work begin on the transmission of a customer’s car 10 minutes after the car is dropped off and the customer is told that the car will be ready within 1 hour from drop-off. What is the probability that the service manager cannot meet his commitment?

1. 0.3875
2. 0.2676
3. 0.5
4. 0.6987

**Ans :**

μ = 45 minutes

σ = 8 minutes

x = 60 minutes – 10 minutes = 50 minutes.

Z = x-μ / σ = 50-45 / 8 = 5/8

Probability that the service manager meet his commitment = Pnorm(5/8) =0 .7340 = 73.40

Q = 1- P = 1 - .7340 = **.2659**

**Answer is B**

Q 7) The current age (in years) of 400 clerical employees at an insurance claims processing center is normally distributed with mean μ = 38 and Standard deviation σ =6. For each statement below, please specify True/False. If false, briefly explain why.

**Ans :** σ =6 and μ = 38

Z = X - μ / σ

The data was distributed normally, with the above values, calculate the Z score.

1. More employees at the processing center are older than 44 than between 38 and 44.

Z = 38-38 / 6 = 0 => pnorm(0) = 0.5 = 50%

Z = 44-38 / 6 = 0 => pnorm(1) = 1= 100%

Between 38 and 44 there are 50% and less than 38 are 50% are Processing center. So the given statement “More employees at the processing center are older than 44 than between 38 and 44” is **FALSE.**

1. A training program for employees under the age of 30 at the center would be expected to attract about 36 employees.

X = 30

Z = 30-38 / 6 = 0 => pnorm(-8/6) = 0.09121 = 9.1%

Total ages of employees given as 400. As we convert into expected number of employees are = 400\*9.1 / 100 = 36 approximately.

The above given statement was **TRUE.**

Q 8) If X1 ~ N(μ, σ2) and X2 ~ N(μ, σ2) are iid normal random variables, then what is the difference between 2 X1 and X1 + X2? Discuss both their distributions and parameters.

Ans: Given statement X1 ~ N(μ, σ2) and X2 ~ N(μ, σ2) and both are **Independent Identically Normal random variables**. Two or more random variable are said to be i.i.d , if they are **mutually independent** and each variable has same **probability distribution**(Uniform distribution).

Consider a simple example.

x1 ,x2,x3….xn are i.i.d sequence which designate the result of the 1st and 2nd and so on toss of the coin.

They are Independent, since every time you flip a coin; the previous result doesn’t influence the current result.

They are identically distributed, since every time you flip a coin, the chance of getting head or tail is identical (Probability distribution is identical over the time)

Example :

**Flipping a coin.**

X1 = ½ and X2 = ½

Since we want to check the 2X1 and X1+X2.

2X1 = 2(1/2) = 1

X1+X2 = ½ +1/2 =1

Rolling a Dice.

X1=1/6 and X2 = 1/6

2X1 = 2(1/6) = 1/3

X1+X2 = 1/6 +1/6 = 1/3

**Conclusion :** If X1 and X2 are i.i.d random variables then 2X1 is equal to X1+X2 and result of the 2X1 and X1+X2 probabilities are getting doubled. So they are **not fall under i.i.d** because if result of probability is not same like X1 and X2.

Q 9) Let X ~ N(100, 202). Find two values, a and b, symmetric about the mean, such that the probability of the random variable taking a value between them is 0.99.

1. 90.5, 105.9
2. 80.2, 119.8
3. 22, 78
4. 48.5, 151.5
5. 90.1, 109.

Ans : N(100,202) = > in this statement, mean =100 and Stadard deviation 202 then it is giving outbox limits. So I have changed into 20^2 then I am getting the below given **answer(D).**

If I consider 202 is Standard devation, Obviously it was effect of Outlier.

Given statement changed to X ~ N(100, 20^2) and it was distributed normally and symmetrically.

X = 100 and σ = 20

Confidence level = 0.99 probability of taking two random variables a and b.

Lower Probability is =0.005

Upper Probability is = 0.995

a= qnorm(0.995,100,20) = **48.43**

b= qnorm(0.005,100,20) = **151.51**

Q 10) Consider a company that has two different divisions. The annual profits from the two divisions are independent and have distributions Profit1 ~ N(5, 32) and Profit2 ~ N(7, 42) respectively. Both the profits are in $ Million. Answer the following questions about the total profit of the company in Rupees. Assume that $1 = Rs. 45

Ans : There are two divisions are given Prionfi1 and Profit2. Both are distributed normally**. I have changed them similar like above question.**

Profit1 ~ N(5, 3^2) = >μ = 5 , σ = 3

Profit2 ~ N(7, 4^2) = >μ = 7 , σ = 4

$1 = Rs. 45 and both profits are in $ Million.

1. Specify a Rupee range (centered on the mean) such that it contains 95% probability for the annual profit of the company.

Profit1 Lower = qrorm(0.975,5,3) = 10.879

Profit1 Upper = qrorm(0.025,5,3) = -0.8798

Profit 1 Range = [-0.8798 $ Million , 10.879 $ Million]

In dollars = [-879800 $ ,10879000 $]

**Profit1 in Rupees = [-39591000, 489555000]**

Profit2 Lower = qrorm(0.975,7,4) = 14.839

Profit2 Upper = qrorm(0.025,7,4) = - 0.8398

Profit2 Range = [-0.8398 $ Million , 14.839 $ Million ]

In dollars = [ -839800 $ , 14839000 $]

**Profit2 in Rupees = [-37791000 ,667755000]**

1. Specify the 5th percentile of profit (in Rupees) for the company

5th percentile of profit means 1/5 of Profit.

**Profit1 in Rupees = [-7918200, 97911000]**

**Profit2 in Rupees = [-7558200 ,133551000]**

1. Which of the two divisions has a larger probability of making a loss in a given year?

Ans: To find the probability first find out range for Profit1 &Profit2. Profit1 = 10 $ Million

Profit2 = 14 $ Million

Then find the probability using the pnorm(range, μ,σ).

Profit1 = pnorm(10,5,3) =0 .9522 = 95.22%

Profit2 = pnorm(14,7,4) =0 .9599 = 95.99%

**Conclusion:** There is larger probability of making a loss in a given year to **Profit2.**