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)

c. P (20<MPG<50)

Explanation: -

1. Let’s find out the probability for MPG of cars below 38.

> library(readr)

> Cars <- read\_csv("Desktop/Digi 360/Module 4/Archive/Cars.csv")

Parsed with column specification:

cols(

HP = col\_double(),

MPG = col\_double(),

VOL = col\_double(),

SP = col\_double(),

WT = col\_double()

)

> View(Cars)

> attach(Cars)

The following objects are masked from Cars (pos = 5):

HP, MPG, SP, VOL, WT

The following objects are masked from Cars (pos = 6):

HP, MPG, SP, VOL, WT

> mean(MPG)

[1] 34.42208

> sd(MPG)

[1] 9.131445

> install.packages("UsingR")

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.6/UsingR\_2.0-6.tgz'

Content type 'application/x-gzip' length 2093795 bytes (2.0 MB)

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

downloaded 2.0 MB

The downloaded binary packages are in

/var/folders/kv/w79zffc14fd2hj518gqdhnmc0000gn/T//Rtmpg5OV6x/downloaded\_packages

> pnorm(38,34.42208,9.131445)

[1] 0.6524059

So, the probability of cars with MPG greater than 38 = 1 – 0.6524059 = 0.3475941 i.e. **34.76%**

1. Let’s find out the probability for MPG of cars below 40.

> pnorm(40,34.42208,9.131445)

[1] 0.7293497

So, the probability of cars with MPG less than 40 = 0.7293497 i.e. **72.93%**

1. Let’s find out the probability for MPG of cars between 20 and 50.

> pnorm(20,34.42208,9.131445)

[1] 0.05712373

> pnorm(50,34.42208,9.131445)

[1] 0.9559926

So, the probability for MPG of cars between 20 and 50 is 0.9559- 0.0571 = 0.8989 i.e. **89.89%**

Q 2) 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

Explanation: -

1. Let’s draw QQ plot for MPG to identify the normal distribution.

> library(readr)

> Cars <- read\_csv("Desktop/Digi 360/Module 4/Archive/Cars.csv")

Parsed with column specification:

cols(

HP = col\_double(),

MPG = col\_double(),

VOL = col\_double(),

SP = col\_double(),

WT = col\_double()

)

> View(Cars)

> attach(Cars)

> attach(Cars)

The following objects are masked from Cars (pos = 3):

HP, MPG, SP, VOL, WT

> install.packages("UsingR")

> qqnorm(MPG)

> qqline(MPG)



Interpretation: - Here we can see that most of the data are almost equally scattered from the QQ line. So, we can say that the dataset is approximately normally distributed.

1. Let’s draw QQ plot for AT to identify the normal distribution.

> library(readr)

> wc\_at <- read\_csv("Desktop/Digi 360/Module 4/Archive/wc-at.csv")

Parsed with column specification:

cols(

Waist = col\_double(),

AT = col\_double()

)

> View(wc\_at)

> attach(wc\_at)

> attach(wc\_at)

The following objects are masked from wc\_at (pos = 3):

AT, Waist

> install.packages("UsingR")

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.6/UsingR\_2.0-6.tgz'

Content type 'application/x-gzip' length 2093795 bytes (2.0 MB)

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

downloaded 2.0 MB

The downloaded binary packages are in

/var/folders/kv/w79zffc14fd2hj518gqdhnmc0000gn/T//Rtmpg5OV6x/downloaded\_packages

> qqnorm(AT)

> qqline(AT)



Interpretation: - Here we can see that the data tends to form curves and moving away from the line at the ends. So, we can say that the dataset is skewed and **not** normally distributed.

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

Ans: -

Z score for 90% CI = 1.645

> qnorm(0.95)

[1] 1.644854

Z score for 94% CI = 1.881

> qnorm(0.97)

[1] 1.880794

Z score for 60% CI = 0.842

> qnorm(0.80)

[1] 0.8416212

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

Ans: -

t score for 95% CI = 2.064

> qt(0.975,24)

[1] 2.063899

t score for 96% CI = 2.172

> qt(0.98,24)

[1] 2.1715

t score for 99% CI = 2.797

> qt(0.995,24)

[1] 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

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

**Explanation: -**

Given that = 260, s = 90, μ = 270 and n = 18.

Here we don’t have σ so, we need to go for t distribution to find out the probability.

t = ()

p <- pt((260-270)/(90/sqrt(18)), 18)

> p

[1] 0.3215076

So, the probability that 18 randomly selected bulbs would have an average life of no more than 260 days is **32.15%**

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: - D

Explanation: - We have a normal distribution with μ = 45 and σ = 8. Let X be the amount of time it takes to complete the repair on a customer's car. To finish in one hour, you must have X ≤ 50 so the question is to find P (X > 50).

> pnorm(50,45,8)

[1] 0.7340145 i.e. 73.4%

Probability that the service manager will not meet his demand will be = 100-73.4 = 26.6% or 0.2676

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.

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

Ans: - Given that μ = 38 and σ = 6.

We need to find P (X > 44) which is 1 – P (X <= 44)

> pnorm(44,38,6)

[1] 0.8413447 i.e.84.13%. which is P(X<=44)

So, the probability of # of employees older than 44 is

P (X > 44) = (1 – 0.8413447) which is 0.1586 i.e. 15.86%.

So, the number of employees older than 44 are 15.86 \* 400/100 = 63 (approximately)

Now let’s find out probability for age of 38 which is P (X = 38)

> pnorm(44,38,6)

> 0.5

So, the probability of employees between age of 38 and 44 is

P (38 < X < 44) = P (X = 44) – P (X =38) = 0.8413447 – 0.5 = 0.3413447 which is 34.13%

So, the number of employees with age between 38 and 44 are 34.13 \* 400/100 = 136 (approximately)

Hence the above statement is false.

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

Ans: - Given that μ = 38 and σ = 6.

We need to find P (X < 30)

> pnorm(30,38,6)

[1] 0.09121122 i.e. 9.12%.

So, the # of employees under 30 are 9.12% of 400 which is approximately 36. Hence above statement is 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.

Explanation: -

From the question X1 ~ N (μ, σ2) and X2 ~ N (μ, σ2) are **identical** independent normal random variables.

*The mean of 2X1: -*

E(2X1) = 2 E(X1)

E(2X1) = 2μ

*The variance of 2X1: -*

V(2X1) = 4 E(X1)

V(2X1) = 4σ2

Therefor 2X1 is normal with mean 2μ and standard deviation 2σ. i.e.

2X1 ~ N (2μ, 2σ)

*The mean of X1 + X2: -*

*The mean of the sum of two random variables X and Y is the sum of their means* - That means

E (X1 + X2) = E(X1) + E(X2)

Since X1 and X2 are identical

E (X1 + X2) = 2μ

*Variance of X1 + X2: -*

*For independent random variables X and Y, the variance of their sum or difference is the sum of their variances* – That means

V (X1 + X2) = V (X1) + V (X2)

= σ2 + σ2

= 2σ2

Therefore X1 + X2 is normal with mean 2μ and standard deviation σ.

Hence X1 + X2 ~ N (2μ, σ)

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, 109.9

Ans: - D

**Explanation: -** Since we need to find out the values of a and b, which are symmetric about the mean, such that the probability of random variable taking a value between them is 0.99, we have to work out in reverse order.

The Probability of getting value between a and b should be 0.99.

So, the Probability of going wrong, or the Probability outside the a and b area is 0.01 (i.e. 1 - 0.99).

The Probability towards left from a = -0.005 (i.e. 0.01/2).

The Probability towards right from b = +0.005 (i.e. 0.01/2).

Since we have the probabilities of a and b, we need to calculate X, the random variable at a and b which has got these probabilities.

By finding the Standard Normal Variable Z (Z Value), we can calculate the X values.

Z = (X- μ) / σ --- > Z \* σ + μ = X

For Probability 0.005 the Z Value is -2.57 (from Z Table).

Z (-0.005) \*20 + 100 = -(-2.57) \*20+100 = **151.4**

Z (+0.005) \*20 + 100 = (-2.57) \*20+100 = **48.6**

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

1. Specify a Rupee range (centered on the mean) such that it contains 95% probability for the annual profit of the company.
2. Specify the 5th percentile of profit (in Rupees) for the company
3. Which of the two divisions has a larger probability of making a loss in a given year?

Ans: - Total Profit = Profit 1 + Profit 2

Profit 1 ~ N (5, 32) and Profit2 ~ N (7, 42)

Therefore, Total Profit ~ N (12, 52)

1. Let’s calculate Z value for 95% probability range

i.e. z value for 0.025 and 0.975

which are -1.96 and 1.96

so, probability for lower range will be -1.96 \* 5 + 12 = 2.2

probability for upper range will be 1.96 \* 5 + 12 = 21.8

The range of profit would be [2.2 \* 45, 21.8 \* 45] = [99, 981]

B) 5th Percentile of the company is: -

= 0.05

Z value for 0.05 = – 1.645

So, X = -1.6448545 \* 5 + 12 = 12 - 8.225 = 3.775

Therefore 5th percentile of profit = 3.775 \* 45 = 169.875 ~ **170** millions

C) Probability of making loss for 1st division is

X = 0 since we are calculating loss that means profit is zero.

= - 1.66666 = 0.4746

Probability of making loss for 2nd division is

X = 0 since we are calculating loss that means profit is zero.

= - 1.75 = 0.4006

So, **probability of making loss for 1st division is more.**