# A picture containing text Description automatically generated

**J.E. CAIRNES SCHOOL OF BUSINESS & ECONOMICS**

**INDIVIDUAL ASSIGNMENT COVER PAGE**

**Module Name and Code: Applied Customer Analytics MS5108**

**Student Name and ID: Dilip Venkatesan Sankar 22225743**

In submitting this assignment, I am aware that it is my responsibility to adhere to the submission guidelines. Please tick (double click… or Yes/No) for the following:

|  |  |  |
| --- | --- | --- |
|  | **Yes** | **No** |
| I am aware of what the NUI Galway plagiarism policy entails. |  |  |
| I have named the assignment file (MS Word docx , doc, or jar ) to contain my student ID and the module code and assignment number (e.g., 1187404\_MS220\_A1.docx or 1187404\_MS220\_A1.jar). |  |  |

**Declaration for this Assignment Submission*:***

*In submitting this work, I confirm that it is entirely my own. I acknowledge that I may be invited to interview if there is any concern in relation to the integrity, and I am aware that any breach will be subject to the University’s Procedures for dealing with plagiarism (*[*http://www.universityofgalway.ie/plagiarism*](http://www.universityofgalway.ie/plagiarism) *).*

## **Question 1:**

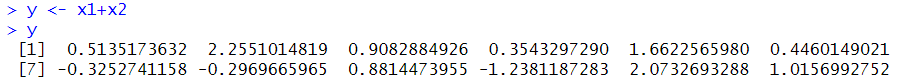
a) The two vectors x1 and x2 are created using Normal Distribution (rnorm) and Exponential Distribution (rexp) respectively,







b) Another vector y is created using the linear combination of x1 and x2,



c) All three vectors are converted into a data frame and a linear regression model is created lm(),

Text

Description automatically generated

Text, letter

Description automatically generated

The Residuals which is the distance between the actual data point and the regression line which has the below values,

1. Min = -2.198e-15 and Max = 3.709e-15 almost reflect each other

2. 1st Quartile = -1.278e-16 and 3rd Quartile = 4.800e-17 do not reflect each other

3. Median = -4.450e-17 which is very much less than zero

So, in this model, the residuals are not normally distributed which is also checked using a histogram.

The coefficient shows y intercept = 3.140e-17 which is the point where the regression line crosses the y-axis and the slope(x1 and x2) = 1.000e+00 based on which we can create the linear model equation,

y = m.x +b

Where y is the Predicted value, m is the slope, x is the independent value (x1+x2) and b is the intercept.

The p-value of intercept(y) = 0.808, slope(x1 and x2) < 2e-16 which shows they are lower than 0.05 supporting the Alternate hypothesis and having a significant difference.

Residual Standard error denotes the average distance that the observed values fall from the regression line = 6.812e-16 with 47 degrees of freedom

Multiple and adjacent R-squared values show how much variance the dependent variable can be accounted for by the independent variable which shows 100% of the variance in y is accounted for by x1 and x2 measures.

F- Statistic = 1.185e+32 and p-value < 2.2e-16 which is lesser than 0.05 which means x1 or x2 is significantly related to y.

## **Question 2:**

Text

Description automatically generated with low confidencea) The dataset is converted into a data frame and the BMI is calculated,

Text

Description automatically generated

b) An object called a sample is created with the condition height is ≥1.70 and the weight is <70,

Text, letter

Description automatically generated

c) The mean and SD of df and the sample for height, weight, and BMI are shown below,

Text

Description automatically generatedText

Description automatically generated Values of df object Values of the sample object

Based on the above comparison we can incur that the mean and SD for height, weight and BMI in the df object are greater compared to the values in sample object