

We have measured the expression of Gene 1 (g1) and Gene 2 (g2) in multiple patient samples. The data is provided in the CSV file.

Data files:

[data1.csv](#): Use this data file if the last digit of your roll number is 0 to 4.

[data2.csv](#): Use this data file if the last digit of your roll number is 5 to 9.

Use R to perform linear regression to find the linear relation between the expression of these two genes.

Important:** Based on the statistical test results for the linear regression, you may have to choose between $y = mx + c$ and $y = mx$. **In that case, you must show the statistical test results in your report to justify your chosen model.

Submit the following using the Teams Assignment:

a) **Your R script:** use the following naming convention – your_roll.R

b) **A report in MS Word.** Use the following naming convention – your_roll.docx. The report should have the following:

- 1) Your name
- 2) Your roll number
- 3) The data set used
- 4) A graph of the data with the regression line. The axes of the graph must be labeled appropriately. The color of the regression line should differ from the color of the data points.
- 5) Write the regressed equation of the fitted line in terms of the variable names (NOT x and y or something else). The variable names in the equation should match those on the graph.
- 6) R-squared value.
- 7) p-values for each coefficient/parameter
- 8) If you perform the regression with the reduced model ($y = mx$), show the statistical test results for the whole model ($y = mx + c$) also. Add an explanation for choosing the reduced model.