

## Data Analysis: test September 2021

The data in the file “test\_2109.RData” concern 100 tumor patients. For each patient, the expression of two genes is given (columns “gene1” and “gene2”) and a binary variable describing whether the tumor is metastatic (0 = not metastatic; 1 = metastatic).

Use regression methods to investigate the relationships among these variables. Discuss the meaning of your results and when appropriate provide graphs. Discuss your results in terms of overfitting by dividing the data into training and testing sets and comparing the respective predictive power of your models.

*The solution must be provided as an Rmarkdown file containing all code and answers, and successfully compiling to an html or pdf file*