# Final Project of Non-parametric Statistics

Due on the class time on June 3, 2024 on the class. You must work on your own.

Consider the data-set of riboflavin production by Bacillus subtilis containing n=71 observations of p=4088 predictors (gene expressions) and a one-dimensional response (riboflavin production). You can access the data via <https://rdrr.io/rforge/hdi/man/riboflavin.html>

Apparently, this is so called large p small n problem, variable selection seems applicable. You can choose methods as many as possible, however you must include at least two methods learned from our courses in your project. Find the most appropriate model in your mind and explain it. A printed report is required.

考虑枯草芽孢杆菌生产核黄素的数据集，其中包含n=71个观察p=4088个预测因子(基因表达)和一维响应(核黄素生产)。

显然，这就是所谓的大p小n问题，变量选择似乎适用。你可以选择尽可能多的方法，但是你必须在你的项目中包括至少两种从我们的课程中学到的方法。在你的脑海中找到最合适的模型并解释它。打印的报告是必需的。