# Advanced regression analysis

## Introduction

At this step of the data analyse minor, you are quite familiar with exploratory data analysis technics and have been working towards developing a solid understanding of the Student Performance dataset and regression analysis. You have also made assumptions and draw conclusions regarding data distribution, missing data and outliers. In this advanced regression analysis, you are using those assumptions and conclusions as starting point for data preparation before incorporating regression analysis.

Goal of this opdracht:

* Review the basic regression analysis concepts and findings from the previous assignment (advanced verkennende analyse and basic regression analyse) and perform a more complex regression analysis on the complete Student Performance dataset, utilizing advanced techniques and models.

The advanced regression analysis you can perform by fulfilling the followings steps. The steps are not meant as an ultimate guide, but more for inspiration by the fulfillment of your analysis.

**Data Preparation and Exploration**

* Revisit your findings from the gevorderdeverkennende analyse regarding relationship between the dependent variable and the independent ones.
* Expand the preprocessing steps you did for the basis regressie analyse to include more features.

**Advanced Regression Models**

* Introduce multiple regression models, such as polynomial regression, ridge regression, lasso regression, and elastic net regression.
* Encourage exploration of non-linear relationships and interactions between variables.

**Model Selection and Evaluation**

* Use advanced model selection techniques, like stepwise regression.
* Employ rigorous model evaluation strategies, including k-fold cross-validation and learning curves.
* Discuss model interpretability and the trade-offs between model complexity and performance.

**Feature Engineering and Selection**

* Experiment with advanced feature engineering techniques (e.g., PCA for dimensionality reduction, interaction terms).
* Apply feature selection methods to identify the most predictive variables.

**Tuning and Optimization**

* Implement hyperparameter tuning using methods like Grid Search or Random Search.
* Explore how different preprocessing steps affect model performance.

**Comparative Analysis**

* Compare the performance of different regression models on the dataset.
* Discuss the strengths and weaknesses of each model in context.