# Multiple Linear Regression Analysis Sample Problem

This document contains a sample dataset and step-by-step instructions for performing a multiple linear regression analysis. The dataset includes a target variable ('Target') and three predictor variables ('Predictor1', 'Predictor2', and 'Predictor3'). This example will guide you through analyzing the R² and adjusted R² values, assessing the confidence intervals for coefficients, and detecting multicollinearity using the Variance Inflation Factor (VIF) method.

## Sample Data

mtcars.csv found in last chapter of the textbook.

Create a model to predict MPG’ as a function of wt, qsec, cyl, hp. The model should yield interpretable coefficients (meaning we need to pay attention to multicollinearity)

Look for and eliminate variables with high collinearity.

Build and evaluate various models looking at the R^2 and adjusted R^2 metrics.

Perform a diagnostic analysis on the residuals (graphical depictions like in section 8.2)

## Tools

Use 9.8.1: Multiple regression using two predictor variables and modify accordingly. Use Excel when necessary.

## Step-by-Step Analysis Instructions

### 1. Analyzing R² and Adjusted R² Values

After fitting the multiple linear regression model, evaluate the R² value and adjusted R² value:  
- \*\*R² Value:\*\* This represents the proportion of variance in the target variable explained by the predictors. A higher R² value indicates a better fit.  
- \*\*Adjusted R² Value:\*\* This value adjusts the R² for the number of predictors. It is especially useful when comparing models with different numbers of predictors, as it penalizes the addition of predictors that do not improve the model.

### 2. Analyzing Confidence Intervals for Coefficients

For each predictor, analyze the confidence interval of its coefficient:  
- Use a 95% confidence level to evaluate the range within which the true coefficient values are likely to fall.  
- If the confidence interval for a coefficient includes zero, it suggests that the predictor may not significantly contribute to explaining the variance in the target variable.

### 3. Assessing Multicollinearity Using VIF

To detect multicollinearity among predictors, calculate the Variance Inflation Factor (VIF) for each predictor:  
- \*\*VIF\*\*: A VIF value above 5 (or 10, depending on the context) indicates high multicollinearity. High multicollinearity implies that the predictor is highly correlated with other predictors, making coefficient estimates unstable.  
- \*\*Procedure\*\*: For each predictor, calculate its VIF by regressing it on all other predictors and finding the R² for this regression. The VIF is calculated as VIF = 1 / (1 - R²).