MATH3714/MATH5714M Linear Regression Practical

This practical uses personalised data files for each student. Your data file can be downloaded from:

- https://seehuhn.github.io/practical/ .

Do not share your data file with others; every student's data file is different.

The deadline for submitting your answers is **Monday, 8th December, 12noon**. Submission will be via Minerva, once the submission link is set up. Questions with numeric answers will be automatically graded. Note: You do not need to justify your answers.

1. Load data and inspect

Load your data file and inspect it.

1.1 What is n (the number of observations)? (number)

1.2 What is p (the number of predictors, excluding the intercept)? (number)

2. Fit a model to the full data set with intercept and all predictors

Fit a linear regression model using all available predictors.

2.1 What is the estimate for the intercept? (number)

2.2 What is the estimated error variance $\hat{\sigma}^2$? (number)

2.3 What are the lower and upper bounds of the 95% confidence interval for β_1? (two numbers)

2.4 What is the predicted value when the categorical variable is "classic" and all numeric predictors are set to 1? (number)

2.5 What is h_{11} (the top-left element of the hat matrix)? (number)

2.6 What is the value of the condition number $\kappa(X)$? (number)

2.7 Comment on what the condition number tells you about multicollinearity. (1 or 2 sentences)

3. Investigate the model fit

Conduct diagnostic checks on your fitted model.

3.1 Create a residual plot. (image)

3.2 Interpret the residual plot. (1 or 2 sentences)

3.3 Create a Q-Q plot. (image)

3.4 Interpret the Q-Q plot. (1 or 2 sentences)

3.5 What is R^2? (number)

3.6 What is adjusted R^2? (number)

3.7 What is the PRESS value? (number)

3.8 Determine the 5 largest Cook's D_i values. (5 pairs of i and D_i)

3.9 What do these Cook's D_i values say about the data? (1 or 2 sentences)

4. Fit a reduced model using regsubsets()

Use the `regsubsets()` function to find the best model according to adjusted R^2.

4.1 Which predictors are included in the reduced model? (list of predictor names)

4.2 What is the adjusted R^2 of the reduced model? (number)

4.3 Determine the 5 largest Cook's D_i values for the reduced model. (5 pairs of i and D_i)

4.4 What is the value of the condition number $\kappa(X)$ for the reduced model? (number)

4.5 Comment on notable properties of the reduced model. Discuss how the reduced model compares to the full model and whether it provides a satisfactory fit to the data. (3-5 sentences)