sf2930 Regression analysis

Exercise session 2 - Ch 3: Multiple linear regression

In class:

1. Montgomery et al., 3.27 Show that
$$Var(\hat{\mathbf{y}}) = \sigma^2 \mathbf{H}.$$

2. Montgomery et al., 3.29 For the *simple* linear regression model, show that the elements of the hat matrix $\mathbf{H}$ are
$$h_{ij} = \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}}$$
and
$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}.$$
Discuss the behaviour of these quantities as x_i moves farther from $\bar{x}$.

3. Montgomery et al., 3.37. Suppose we fit the model $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}$ when the true model is actually given by $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$. For both models assume that $E[\boldsymbol{\epsilon}] = 0$ and $Var[\boldsymbol{\epsilon}] = \sigma^2\mathbf{I}$. Find the expected value and variance of the ordinary least-squares estimate, $\boldsymbol{\beta}_1$. Under what conditions is this estimate unbiased?

4. Montgomery et al., 3.26 Suppose that we have two independent samples say $(y_1, x_1), \ldots, (y_{n_1}, x_{n_1})$ and $(y_{n_1+1}, x_{n_1+1}), \ldots, (y_{n_1+n_2}, x_{n_1+n_2})$. Two models can be fit to these samples,
$$\begin{aligned} y_i =& \beta_0 + \beta_i x_i + \epsilon_i, i = 1, \ldots, n_1 \\ y_i =& \beta_0 + \beta_i x_i + \epsilon_i, i = n_1 + 1, \ldots, n_1 + n_2. \end{aligned}$$

   a) Show how these two separate models can be written as a single model.
   b) Using the result in part a), show how the general linear hypothesis can be used to test equality of the slopes β_1 and γ_1.
   c) Using the result in part a), show how the general linear hypothesis can be used to test equality of the regression lines.
   d) Using the result in part a), show how the general linear hypothesis can be used to test that both slopes are equal to a constant c.

5. Montgomery et al., 3.1. Consider th National Football League data in Table B.1.

   (a) Fit a multiple linear regression model relating the number of games won to the teams passing yardage (x_2), the percentage of rushing plays (x_7), and the opponents' yards rushing (x_8).

   (b) Construct the ANOVA table and test for significance of the regression.

   (c) Calculate t statistic for the hypotheses $H_0 : \beta_2 = 0$, $H_0 : \beta_7 = 0$ and $H_0 : \beta_8 = 0$. What conclusions can you draw about the roles the variables x_2, x_7 and x_8 play in the model?

   (d) Calculate R^2 and R^2_{Adj} for this model.

   (e) Using the partial F test, determine the contribution of x_7 to the model. How is this partial F statistic related to the t test for β_7 in c)?

Recommended exercises:

Book	Theory	Implementation
Montgomery et al.:	3.24, 3.25, 3.30, 3.31, 3.32, 3.33, 3.36, 3.38, 3.39, 8.12	3.2, 3.5, 3.7, 3.21

References

D.C. Montgomery, E.A. Peck, and G.G. Vining. *Introduction to Linear Regression Analysis*. Wiley Series in Probability and Statistics. Wiley, 2012. ISBN 9780470542811.