SF2930 Regression analysis

Exercise session 3 - Ch 4: Model adequasi, Ch 5: Transformations and weighting to correct model inadequacies, Ch 15.4: re-sampling (bootstrap)

In class:

1. Montgomery et al., 4.2. Consider the simple regression model fit to the National Football League team performance data in Problem 2.1.
   (a) Construct a normal probability plot of the residuals. Does there seem to be any problem with the normality assumption?
   (b) Construct and interpret a plot of the residuals versus the predicted response.
   (c) Plote the residuals versus team passing yardage x_2. Does this plot indicate that the model will be improved by adding x_2 to the model?
2. Montgomery et al., 4.2. Consider the multiple regression model fit to the National Football League team performance data in Problem 3.1.
   (a) Construct a normal probability plot of the residuals. Does there seem to be any problem with the normality assumption?
   (b) Construct and interpret a plot of the residuals versus the predicted response.
   (c) Construct plots of the residuals versus each of the regressor variables. Do these plots imply that the regressor is correctly specified?
   (d) Construct the partial regression plots for these model. Compare the plots with the plots of residuals versus regressors from part c above. Discuss the type of information provided by these plots.
   (e) Compute the studentized residuals and then R-student residuals for this model. What information is conveyed by these scaled residuals?
3. Montgomery et al., 5.8 Consider the tree models
   (a) $y = \beta_0 + \beta_1(1/x) + \epsilon$
   (b) $1/y = \beta_0 + \beta_1 x + \epsilon$
   (c) $y = x/(\beta_0 + \beta_1 x) + \epsilon$

   All of these models can be linearized by reciprocal transformations. Sketch the behavior of y as function of x. What observed characteristic in the scatter diagram would lead you to choose one of these models?

4. Montgomery et al., 5.16 Consider the tree model

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$$

where $E[\boldsymbol{\epsilon}] = \mathbf{0}$ and $Var(\boldsymbol{\epsilon}) = \boldsymbol{\sigma}^{\mathbf{2}}\mathbf{V}$. Assume that $\boldsymbol{\sigma}^{\mathbf{2}}$ and $\mathbf{V}$ are known. Derive an appropriate test statistic for the hypothesis

$$H_{:}\boldsymbol{\beta}_2 = \mathbf{0},\ H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}.$$

Give the distribution under both the null and the alternative hypotheses.

5. Montgomery et al., 5.10. Consider the soft drink data in Example 3.1. Find the bootstrap estimate of the standard deviation of $\hat{\beta}_1$ using the following number of bootstrap samples: $m = 100, 200, 300, 400, 500$. Can you draw any conclusions about how many samples are necessary to obtain a reliable estimate of the precision of estimation for $\hat{\beta}_1$?

Recommended exercises:

Book	Theory	Implementation
James et al.:		lab 5.3.4
Montgomery et al.:	4.14, 5.1, 5.2, 5.3, 5.5 5.14 5.15,	4.1, 4.4, 4.6, 5.12, 15.9

References

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer New York, 2013. ISBN 9781461471387.

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.