sf2930 Regression analysis

Exercise session 1 - Ch 2: Simple linear regression

In class:

1. Rawlings et al., 1.1. Use the least squares criterion to derive the normal equations

$$n\hat{\beta}_0 + (\sum_{i=1}^{n} x_i)\hat{\beta}_1 = \sum_{i=1}^{n} y_i$$
$$(\sum_{i=1}^{n} x_i)\hat{\beta}_0 + (\sum_{i=1}^{n} x_i^2)\hat{\beta}_1 = \sum_{i=1}^{n} x_i y_i.$$

   for the linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ \epsilon_i \sim N(0, \sigma^2), i = 1, \ldots, n.$$

2. Rawlings et al., 1.2. Solve the normal equations to obtain the estimates of β_0 and β_1.

3. Rawlings et al., 1.3. Use the statistical model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ \epsilon_i \sim N(0, \sigma^2).$$

   to show that $\epsilon_i \sim NID(0, \sigma^2)$ implies each of the following:

   (a) $E(y_i) = \beta_0 + \beta_1 x_i$,

   (b) $Var(y_i) = \sigma^2$, and

   (c) $Cov(y_i, y_j) = 0, i \neq j$

   For Parts b) and c), use the following definitions of variance and covariance.

$$Var(y_i) = E[[y_i - E(y_i)]^2]$$
$$Cov(y_i, y_j) = E[[y_i - E(y_i)][y_j - E(y_j)]].$$

4. Montgomery et al., 2.25. Consider the simple linear regression model $y = \beta_0 + \beta_1 x + \epsilon, E[\epsilon] = 0, Var(\epsilon) = \sigma^2$ and ϵ uncorrelated.

   (a) Show that $Cov(\hat{\beta}_0, \hat{\beta}_1) = -\bar{x}\sigma^2/S_{xx}$

   (b) Show that $Cov(\bar{y}, \hat{\beta}_1) = 0$.

5. Montgomery et al., 2.33. Consider the least-squares residuals $e_i = y_i - \hat{y}_i, i = 1, \ldots, n$, form the simple linear regression model. Find the variance of the residuals $Var(e_i)$. Is the variance of the residuals constant?

6. Montgomery et al., Example 2.1-2.7. A rocket motor is manufactured by bonding an igniter propellant and a sustainer propellant together inside a metal housing. The shear strength of the bond between the two types of propellant is an important quality characteristic. It is suspected that shear strength is related to the age in weeks of the batch of sustainer propellant. Twenty observations on shear strength and the age of the corresponding batch of propellant have been collected. The following model is suggested

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ \epsilon_i \sim N(0, \sigma^2).$$

   2.1) Estimate the coefficients β_0, β_1. Is this model reasonable?

   2.2) Estimate the variance σ^2.

   2.3) Construct t-test for the hypothesis $H_0 : \beta_1 = 0$.

   2.4) Construct the ANOVA table.

   2.5) Confidence interval for β_0, β_1 and σ^2.

   2.6) Confidence interval for mean of a new observation.

   2.7) Prediction interval for a new observation.

7. Montgomery et al., 2.23. Consider the simple linear regression model $y = 50 + 10x + \epsilon$ where ϵ is $NID(0, 16)$. Suppose that $n = 20$ pairs of observations are used to fit this model. Generate 500 samples for 20 observations, drawing one observation for each level of $x = 1, 1.5, 2, \ldots, 10$ for each sample.

   (a) For each sample compute the least-squares estimates of the slope and intercept. Construct histograms of the sample values of β_0 and β_1. Discuss the shape of these histograms.

   (b) For each sample, compute the estimate $E[y \mid x = 5]$. Construct histograms for each estimate you obtained. Discuss the shape of the histogram.

   (c) For each sample, compute a 95% CI on the slope. How many of these intervals contain the true values $\beta_1 = 10$? Is this what you would expect?

   (d) For each estimate of $E[y \mid x = 5]$ in part b), compute the 95% CI. How many of these intervals contain the true value $E[y \mid x = 5] = 100$? Is this what you would expect?

Recommended exercises:

Book	Theory	Implementation
Rawlings et al.	1.16	
Montgomery et al.	2.25, 2.26, 2.27, 2.28, 2.29, 2.31, 2.33	2.1, 2.7

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.

J.O. Rawlings, S.G. Pantula, and D.A. Dickey. *Applied Regression Analysis: A Research Tool*. Springer Texts in Statistics. Springer New York, 2001. ISBN 9780387984544.