Statistical Rethinking: Chapter 5 - Multivariate Linear Models

Chris Hayduk

January 6, 2019

1 Easy

Problem 5E1.
Which of the linear models below are multiple linear regressions?

1. $\mu_i = \alpha + \beta x_i$

2. $\mu_i = \beta_x x_i + \beta_z z_i$

3. $\mu_i = \alpha + \beta(x_i - z_i)$

4. $\mu_i = \alpha + \beta_x x_i + \beta_z z_i$

  Linear models 2 and 4 are multiple linear regressions.

Problem 5E2.
Write down a multiple regression to evaluate the claim: *Animal diversity is linearly related to latitude, but only after controlling for plant diversity.* You just need to write down the model definition.

$$\text{animal diversity}_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \beta_{\text{latitude}}\text{latitude}_i + \beta_{\text{diversity}}\text{diversity}_i$$
$$\beta_{\text{latitude}} \sim \text{Normal}(0, 10)$$
$$\beta_{\text{diversity}} \sim \text{Normal}(0, 10)$$
$$\sigma \sim \text{Uniform}(0, 10)$$

Problem 5E3.
Write down a multiple regression to evaluate the claim: *Neither amount of funding nor size of laboratory is by itself a good predictor of time to PhD degree; but together these variables are both positively associated with time to degree.* Write down the model definition and indicate which side of zero each slope parameter should be on.

$$\text{time}_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \beta_{\text{lab size}}\text{lab size}_i + \beta_{\text{funding}}\text{funding}_i$$
$$\beta_{\text{lab size}} \sim \text{Normal}(0,\ 10)$$
$$\beta_{\text{funding}} \sim \text{Normal}(0,\ 10)$$
$$\sigma \sim \text{Uniform}(0,\ 10)$$

Both parameters should have slopes greater than zero since the problem specifies that "together the variables are both positively associated with time to degree".

Problem 5E4.
Suppose you have a single categorical predictor with 4 levels (unique values), labeled A, B, C, and D. Let A_i be an indicator variable that is 1 where case i is in category A. Also suppose B_i, C_i, and D_i for the other categories. Now which of the following linear models are inferentially equivalent ways to include the categorical variable in a regression? Models are inferentially equivalent when it's possible to compute one posterior distribution from the posterior distribution of another model.

1. $\mu_i = \alpha + \beta_A A_i + \beta_B B_i + \beta_D D_i$

2. $\mu_i = \alpha + \beta_A A_i + \beta_B B_i + \beta_C C_i + \beta_D D_i$

3. $\mu_i = \alpha + \beta_B B_i + \beta_C C_i + \beta_D D_i$

4. $\mu_i = \alpha_A A_i + \alpha_B B_i + \alpha_C C_i + \alpha_D D_i$

5. $\mu_i = \alpha_i(1 - B_i - C_i - D_i) + \alpha_B B_i + \alpha_C C_i + \alpha_D D_i$

Models 1, 3, 4, and 5 are all inferentially equivalent.

2 Medium