# 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.

$$\begin{gathered}
\text{time}_{\text{i}} \sim \text{Normal}(\mu_{\text{i}}, \sigma) \\
\mu_{\text{i}} = \beta_{\text{lab size}} \text{lab size}_{\text{i}} + \beta_{\text{funding}} \text{funding}_{\text{i}} \\
\beta_{\text{lab size}} \sim \text{Normal}(0, 10) \\
\beta_{\text{funding}} \sim \text{Normal}(0, 10) \\
\sigma \sim \text{Uniform}(0, 10)
\end{gathered}$$

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 $\text{A}_{\text{i}}$ be an indicator variable that is 1 where case $i$ is in category A. Also suppose $\text{B}_{\text{i}}$, $\text{C}_{\text{i}}$, and $\text{D}_{\text{i}}$ 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. $\mu_{\text{i}} = \alpha + \beta_{\text{A}}\text{A}_{\text{i}} + \beta_{\text{B}}\text{B}_{\text{i}} + \beta_{\text{D}}\text{D}_{\text{i}}$
2. $\mu_{\text{i}} = \alpha + \beta_{\text{A}}\text{A}_{\text{i}} + \beta_{\text{B}}\text{B}_{\text{i}} + \beta_{\text{C}}\text{C}_{\text{i}} + \beta_{\text{D}}\text{D}_{\text{i}}$
3. $\mu_{\text{i}} = \alpha + \beta_{\text{B}}\text{B}_{\text{i}} + \beta_{\text{C}}\text{C}_{\text{i}} + \beta_{\text{D}}\text{D}_{\text{i}}$
4. $\mu_{\text{i}} = \alpha_{\text{A}}\text{A}_{\text{i}} + \alpha_{\text{B}}\text{B}_{\text{i}} + \alpha_{\text{C}}\text{C}_{\text{i}} + \alpha_{\text{D}}\text{D}_{\text{i}}$
5. $\mu_{\text{i}} = \alpha_{\text{i}}(1 - \text{B}_{\text{i}} - \text{C}_{\text{i}} - \text{D}_{\text{i}}) + \alpha_{\text{B}}\text{B}_{\text{i}} + \alpha_{\text{C}}\text{C}_{\text{i}} + \alpha_{\text{D}}\text{D}_{\text{i}}$

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