

# Multivariable models

2022-09-20

## Contents

Homework . . . . .	1
Homework Exercises: . . . . .	1
Notes on Homework: . . . . .	1

## Homework

### Homework Exercises:

**Self-study:** Work these exercises, but do not turn them in.

- Exercises 5E1–5E4

**Turn in:** Work these exercises and turn them in.

- 5M3, 5M4, 5H3

### Notes on Homework:

Exercise 5E4 gets at a subtle point about independence of variables when you have indicators for categories. This connects to a subtle, but important point about *identifiability* in models. When you can infer the exact value of a variable from other variables, then including the exactly predictable variable in your models can create problems by making the models *non-identifiable*. A good example is if you have indicator variables *male* and *female* for biological sex (for simplicity, I am leaving out the possibility of intersex individuals). If you have a regression model  $y = \alpha + \beta_1 I_{\text{male}} + \beta_2 I_{\text{female}}$ , then the model will predict the same result if you use parameters  $\alpha' = \alpha + \delta$ ,  $\beta_1' = \beta_1 - \delta$ , and  $\beta_2' = \beta_2 - \delta$ . If you omit  $I_{\text{male}}$  or  $I_{\text{female}}$  from your model (but not both), you will have a model that works just as well (because  $I_{\text{male}} = 1 - I_{\text{female}}$ , so the model will have just as much information), but the model will now be completely *identifiable* because we can't get equivalent results by changing  $\alpha$  and  $\beta$ . This is why the kind of analysis in this exercise, to check whether a model is fully identifiable, is important.