Worksheet 23 - multicollinearity

Monday, April 27, 2026

Math 58B - Jo Hardin

Name: _____

Names of people you worked with: _____

How many people in the room can you name? Try it.

Task: Consider a task where we are trying to predict the price of a home (in ln units) using the square feet of the home (in ln units) and the number of bedrooms. Below are three different models. For each model, provide the sign (either positive or negative) of the coefficient with a one sentence explanation.

Model 1: Is β_1 positive or negative, why?

$$\ln(\texttt{price}) = \beta_0 + \beta_1 \ln(\texttt{square feet})$$

Model 2: Is β_1 positive or negative, why?

$$\ln(\texttt{price}) = \beta_0 + \beta_1 \texttt{bedrooms}$$

Model 3: Is β_1 positive or negative, why? Is β_2 positive or negative, why?

$$\ln(\texttt{price}) = \beta_0 + \beta_1 \ln(\texttt{square feet}) + \beta_2 \texttt{bedrooms}$$

Solution:

Model 1: β_1 is positive because the larger the home, the more expensive it is.

Model 2: β_1 is positive because the more bedrooms the home has, the more expensive it is.

Model 3: β_1 is positive because, for a fixed number of bedrooms, the larger the home, the more expensive it is.

β_2 is **negative!!!** because, for a fixed square footage, the more bedrooms actually makes the house wonky and *less* expensive.

```
library(openintro)
data(LAhomes)
```

Model 1:

```
lm(log(price) ~ log(sqft), data = LAhomes) |>
  tidy()
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)     2.70    0.144       18.8 1.97e-71
2 log(sqft)       1.44    0.0195      73.8 0
```

Model 2:

```
lm(log(price) ~ bed, data = LAhomes) |>
  tidy()
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic   p.value
  <chr>          <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)   11.8      0.0436     271.   0
2 bed            0.532    0.0142      37.3 9.77e-220
```

Model 3:

```
lm(log(price) ~ log(sqft) + bed, data = LAhomes) |>
  tidy()
```

```
# A tibble: 3 x 5
  term        estimate std.error statistic   p.value
  <chr>          <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)    1.47     0.218       6.73 2.28e- 11
2 log(sqft)      1.66     0.0346     47.8  2.60e-310
3 bed           -0.123    0.0164     -7.46 1.46e- 13
```

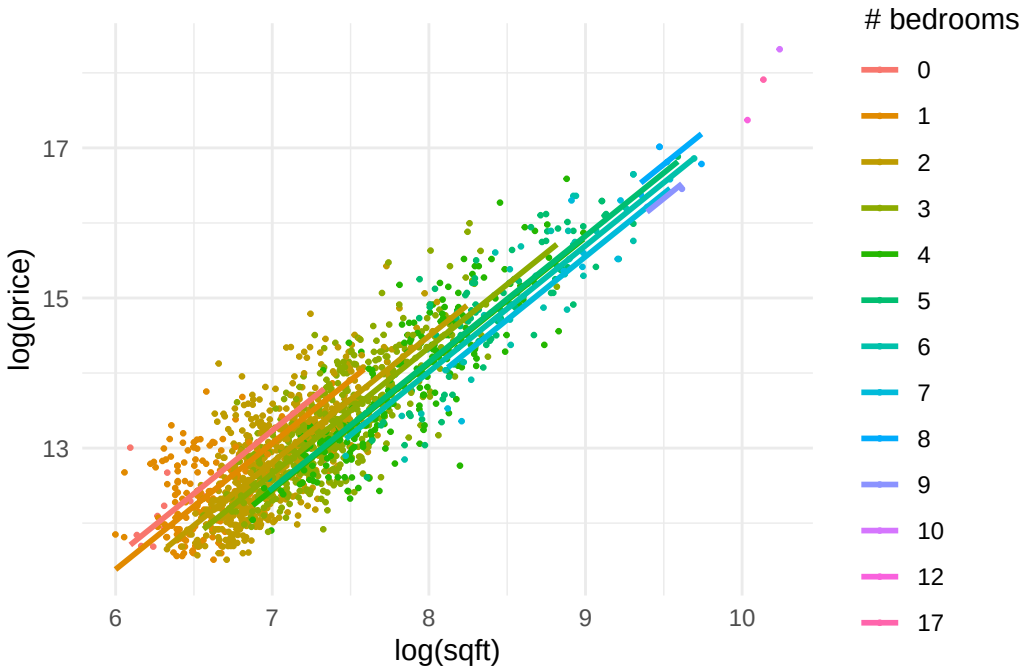

Figure 1: For a given (log) sqft, the more bedrooms, the lower the predicted log(price).

```
praise()
```

```
[1] "You are luminous!"
```