Linear Regression / Broom Lab

David Gerard

2019-09-19

Learning Objectives

- Practice implementing and interpreting simple linear models.
- Practice the `nest()`-`map()`-`unnest()` pipeline.

Ecosystem Decay

In Brazil, there is a requirement that 50% of the land in any development project remains in forest. This has resulted in many forest "islands". Researchers were interested in studying the effect of forest "island" size on species diversity. The data in `ex0822` from the Sleuth3 package contain the area of these islands and the number of butterfly species in each island.

```
library(Sleuth3)
data("ex0822")
```

1. Does a linear model appear appropriate for Species on Area? Why or why not? What are some possible fixes?

2. Apply any fixes you suggested in part 1 and verify that the assumptions of the linear model are satisfied.

3. Does there appear to be a relationship between area and the number of butterflies?

4. Interpret the coefficient estimates from part 2. Provide 95% confidence intervals (on the original scale).

5. Provide predictions for islands of size 500 ha and 50 ha.

Simulation Output

The data in sims.csv contain the output of some simulations I ran comparing two methods: updog and fitPoly. The variables include:

- `updog`: A measure of how well updog performed (the "mean squared error"). Smaller values are better.
- `fitPoly`: A measure of how well fitPoly performed (the "mean squared error"). Smaller values are better.
- `bias`: A parameter called the "bias" that was controlled in the simulations. Different values correspond to different simulation settings. Methods are expected to behave worse for `bias` values further from 1.
- `od`: A parameter called the "overdispersion rate" that was controlled in the simulations. Different values correspond to different simulation settings. Methods are expected to behave worse for larger `od` values.
- `seed`: The random seed used (via `set.seed()`) in that iteration.

1. Load these data into R.

2. The main goal of the simulation study is to see if `updog` has smaller MSE (Mean Squared Error) on average than `fitPoly` for each different combination of `bias` and `od`. To compare differences in means, we should use a t-test. Do we want to use a paired or two-sample t-test?

3. Plot the data to see if the assumptions of the t-test would be satisfied.

4. Whether or not it's appropriate, we'll stay on the original scale and apply a t-test. Use the `nest()`-`map()`-`unnest()` command to obtain effect size estimates and 95% confidence intervals for the effect sizes for each level of `bias` and `od`.

5. Plot effect size estimates (and their confidence levels) against bias and od. Your plot should like this: