

Adding Other Variables

David Gerard

2018-12-07

Objectives

- Understand why we add other variables in our model.

Reason 1

- You are actually interested in the effects of both variables
- E.g. pressure and temperature on espresso foam index.
- If no interactions are present, then can have same power with fewer observations than by running two separate studies (one studying pressure and one studying temperature).
 - Every observation contributes to estimating both effects, rather than having each observation contribute to estimating only one effect.

Reason 2

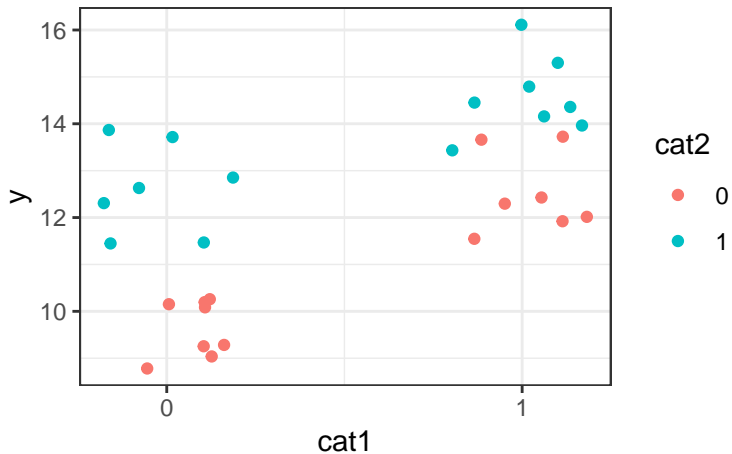
- You suspect there is an interaction between the treatment of interest and some other variable.
- Let's you look at the treatment effect at different levels of the nuisance variable.
- Or if no interactions are found, gives you an idea of the universality of the result.

Reason 3

- Reduces variance, so easier to detect an effect.

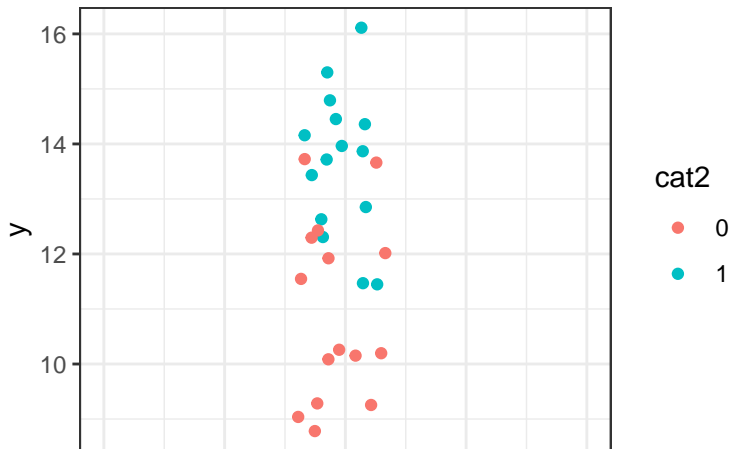
Variance reduction

- Separation of groups is distinct



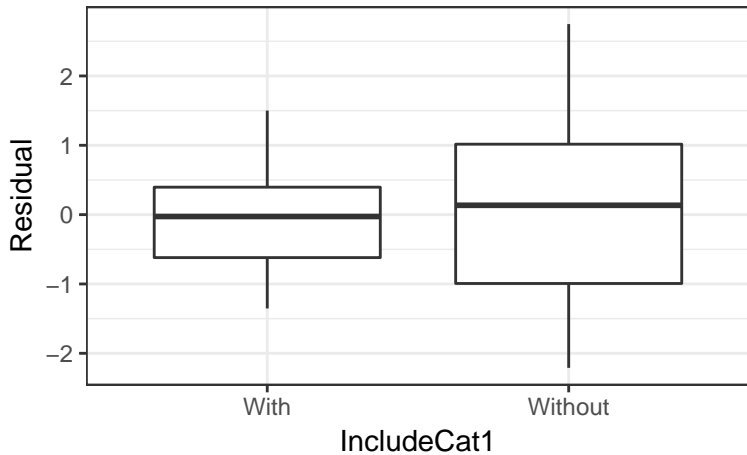
Variance reduction

- Much more overlap



- When you include a variable in an *experiment* just to reduce variability, this is called **grouping** and that variable is called the **grouping variable**.
- When you apply grouping, you include the variable in the model whether or not it is significantly associated with the response.

Residuals

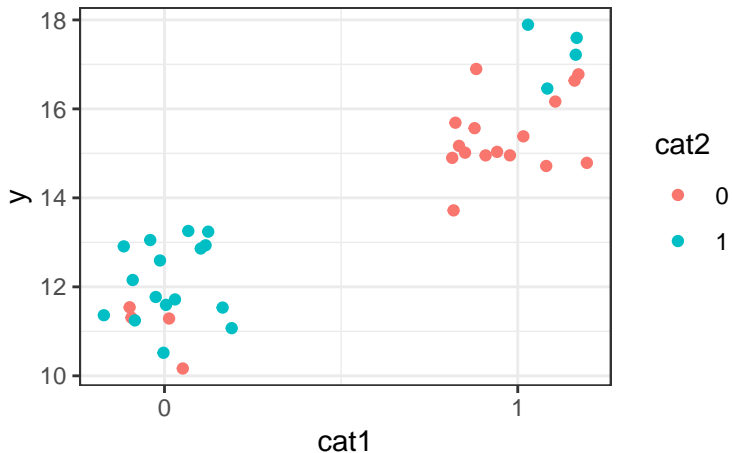


Reason 4

- In **observational studies**, allows you to control for Simpson's paradox.
- Simpson's Paradox: Strength or direction of an effect changes when you control for another variable.
- In 2-way ANOVA, occurs because of unequal numbers of units in each group (unbalanced studies).

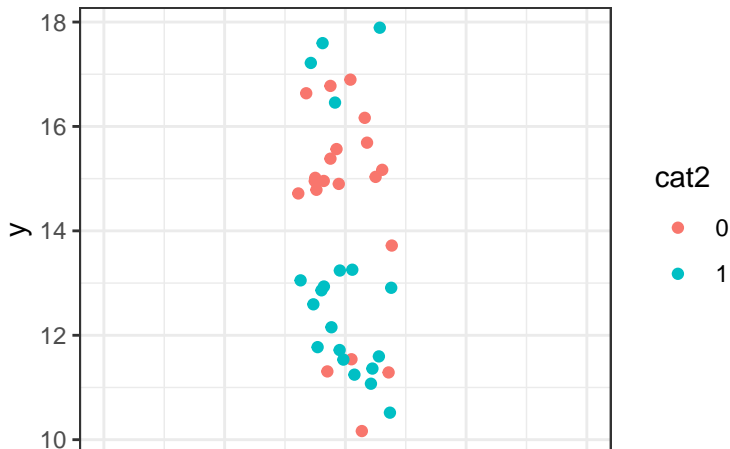
Simpson's Paradox

- Blue is better at both levels of category 1



Simpson's Paradox

- But when you aggregate it looks like Red is mostly better



Simpson's Paradox

- Red is overrepresented in the good level of category 1, and underrepresented in the bad level of category 1.
- Another example: United has worse delay times than other airlines, but that's because they have a hub at O'Hare.

Simpson's Paradox

- Not usually an issue in randomized experiments.
- The randomization makes sure that nothing is over/underrepresented