

Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness

Kellie Ottoboni

Department of Statistics, UC Berkeley
Berkeley Institute for Data Science

October 1, 2015

Student evaluations of teachers (SET) are used to

- ▶ Quantify teaching effectiveness
- ▶ Compare instructors across courses
- ▶ Make hiring, firing, and promotion decisions

Are SET a valid measure of teaching effectiveness?

No!

- ▶ SET measure “customer satisfaction”
- ▶ Ratings are biased against female instructors
- ▶ Biases are inconsistent across universities and disciplines; impossible to “adjust” SET

some stuff about Anne's data

some stuff about Lilly's data

Permutation tests

- ▶ Model-free, distribution-free test
- ▶ Give exact level of test without crazy assumptions, only assume the randomization of the treatments
- ▶ Python package: `permute`
<https://github.com/statlab/permute>

- Anne Boring. Gender biases in student evaluations of teachers. Document de travail OFCE 13, OFCE, April 2015.
- L. MacNell, A. Driscoll, and A. N. Hunt. Whats in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, pages 1–13, 2014.