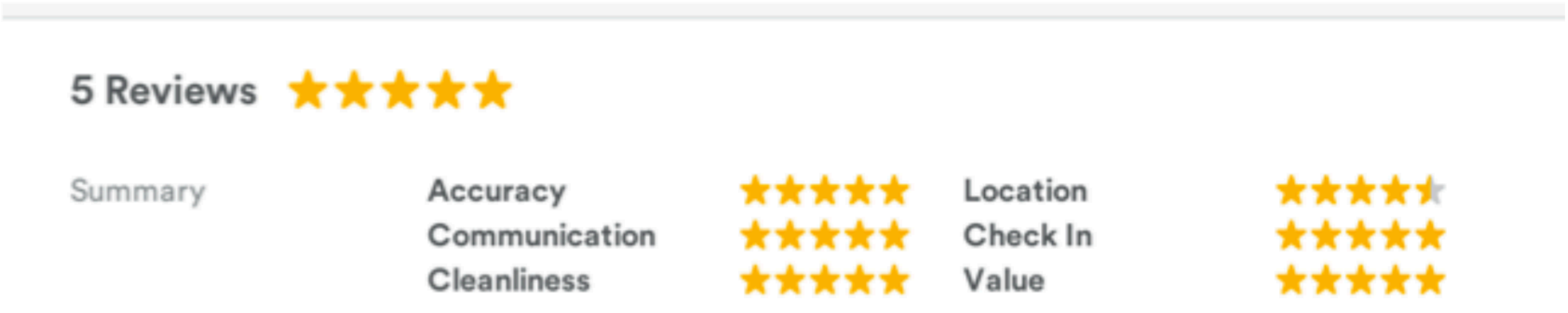
Business Experimentation and Causal Methods

Prof. Fradkin
Topic: Experimenting in Practice



Last Time

1. **How to design an experiment.**

    - Problem to be solved?

    - Types of treatment arms.

    - Number of treatment arms.

    - Unit of randomization.

2. How to analyze an experiment after.

    - Collect appropriate data.

    - Check for proper randomization.

    - Obtaining estimates.

    - Thinking about effect sizes.

This Time

1. How to design an experiment.

   - Problem to be solved?

   - Types of treatment arms.

   - Number of treatment arms.

   - Unit of randomization.

2. **How to analyze an experiment after.**

   - Collect appropriate data.

   - Check for proper randomization.

   - Obtaining estimates.

   - Thinking about effect sizes.

Checking for proper randomization

- Step 1: Check if the proportion of units treated is what you would expect.

  The test below shows a test that 5% of people were randomized if 5 out of 83 people are treated. We see that the p value is > .05 so that we can't reject the null (that is good in this case).

```
>>> count = 5
>>> nobs = 83
>>> value = .05
>>> stat, pval = proportions_ztest(count, nobs, value)
>>> print('{0:0.3f}'.format(pval))
0.695
```

Checking for proper randomization

- Step 2: Check if treatment and control have similar pre-experiment characteristics.
- In practice: do a t.test or a regression where the outcome is the pre-experiment characteristic.
- Below is a table from the analysis of the reviews paper where we do several of these checks.
- Note, that if we have proper randomization, we expect most differences to be insignificant (e.g. the p-value should be higher than 5% and the absolute difference between treatment and control should be small.

Variable	Difference	Mean Treatment	Mean Control	P-Value	Stars
Total Bookings by Guest	-0.024	2.999	3.024	0.270	
US Guest	-0.002	0.285	0.286	0.558	
Guest Tenure (Days)	-2.065	268.966	271.032	0.271	
Host Listings	0.015	1.858	1.843	0.566	
Listing Reviews	-0.039	10.662	10.700	0.715	

Getting Estimates

- Run a regression of the outcome on the treatment.
- If we need more precision, add pre-treatment covariates to the regression.

	Guest			Host		
	Control Mean	*Treatment Mean*	*Effect*	*Control Mean*	*Treatment Mean*	*Effect*
Submits Review	0.68	0.69	0.01 ***	0.72	0.79	0.07 ***
Recommends	0.97	0.97	0.00	0.99	0.99	0.00 *
Overall Rating = 5	0.74	0.73	-0.01 ***			

Reading regression tables:

Different Outcomes

	Dependent variable:				
	Log(Nights in Exp.)	Log(Price in Exp.)	Log(Rev. in Exp.)	Log(Bookings by 2015)	Active in 2015
	(1)	(2)	(3)	(4)	(5)
Treatment	−0.010	−0.005	−0.025	0.002	−0.003
	(0.006)	(0.005)	(0.019)	(0.004)	(0.002)
Controls Included	Yes	Yes	Yes	Yes	Yes
Observations	119,550	73,234	119,550	119,550	119,550
R^2	0.262	0.411	0.219	0.631	0.078

Note: :0.05; ***p<0.01

This table displays the treatment effects on listing outcomes after the first transaction in the experiment. Controls are included for greater than median effective positive percentage (EPP), whether the EPP is calculable, log of prior bookings, log of the first price, and whether the guest submitted a customer service complaint. Columns with 'In Exp' in the name refer to outcome calculated only through June 12, 2014, the end of the experimental period. There are fewer observations for the price variable, because we can't measure transaction prices for hosts who did not transact after the initial transaction in the experiment.

Reading regression tables:

Estimate and standard error in parentheses
Notice, not significant.

	Dependent variable:				
	Log(Nights in Exp.)	Log(Price in Exp.)	Log(Rev. in Exp.)	Log(Bookings by 2015)	Active in 2015
	(1)	(2)	(3)	(4)	(5)
Treatment	−0.010	−0.005	−0.025	0.002	−0.003
	(0.006)	(0.005)	(0.019)	(0.004)	(0.002)
Controls Included	Yes	Yes	Yes	Yes	Yes
Observations	119,550	73,234	119,550	119,550	119,550
R^2	0.262	0.411	0.219	0.631	0.078

Note: :0.05; ***p<0.01

This table displays the treatment effects on listing outcomes after the first transaction in the experiment. Controls are included for greater than median effective positive percentage (EPP), whether the EPP is calculable, log of prior bookings, log of the first price, and whether the guest submitted a customer service complaint. Columns with 'In Exp' in the name refer to outcome calculated only through June 12, 2014, the end of the experimental period. There are fewer observations for the price variable, because we can't measure transaction prices for hosts who did not transact after the initial transaction in the experiment.

Reading regression tables:

	Dependent variable:				
	Log(Nights in Exp.)	Log(Price in Exp.)	Log(Rev. in Exp.)	Log(Bookings by 2015)	Active in 2015
	(1)	(2)	(3)	(4)	(5)
Treatment	−0.010	−0.005	−0.025	0.002	−0.003
	(0.006)	(0.005)	(0.019)	(0.004)	(0.002)
Controls Included	Yes	Yes	Yes	Yes	Yes
Observations	119,550	73,234	119,550	119,550	119,550
R^2	0.262	0.411	0.219	0.631	0.078
Note:					$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Were controls included, observations, R-squared

Reading regression tables:

Details about the table usually written below in small text.

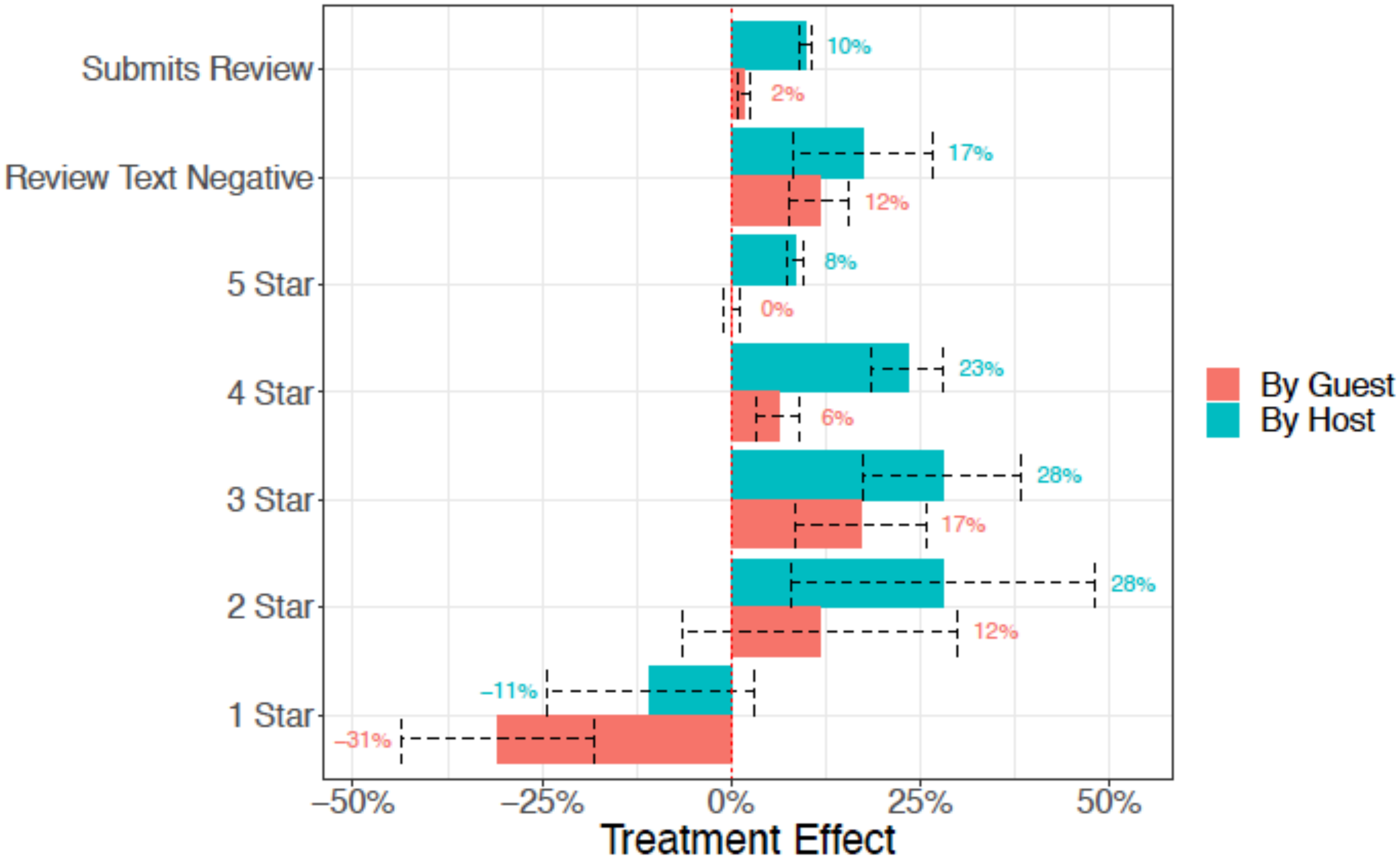
	Dependent variable:				
	Log(Nights in Exp.)	Log(Price in Exp.)	Log(Rev. in Exp.)	Log(Bookings by 2015)	Active in 2015
	(1)	(2)	(3)	(4)	(5)
Treatment	−0.010	−0.005	−0.025	0.002	−0.003
	(0.006)	(0.005)	(0.019)	(0.004)	(0.002)
Controls Included	Yes	Yes	Yes	Yes	Yes
Observations	119,550	73,234	119,550	119,550	119,550
R^2	0.262	0.411	0.219	0.631	0.078

Note: :0.05; ***p<0.01

This table displays the treatment effects on listing outcomes after the first transaction in the experiment. Controls are included for greater than median effective positive percentage (EPP), whether the EPP is calculable, log of prior bookings, log of the first price, and whether the guest submitted a customer service complaint. Columns with 'In Exp' in the name refer to outcome calculated only through June 12, 2014, the end of the experimental period. There are fewer observations for the price variable, because we can't measure transaction prices for hosts who did not transact after the initial transaction in the experiment.

Plotting effects and 95% confidence intervals

Makes it easier to see everything.
Treatment increases reviews, and those reviews are more likely to be 2 - 4 stars.



How to interpret the effect?

- In surprise to the experimenters, review rates actually increased!
- It helps to put effects into percent terms. Guest reviews increased by 2% (or from .68 to .69 of transactions). Small but probably positive.
- Criteria for launching the treatment were met.
- My research showed that this increase in ratings is due to curiosity by users about what the other person has said.
- This is an unintended consequence of the treatment! Exactly why we run experiments.

Recap

1. How to design an experiment.

   - Problem to be solved?

   - Types of treatment arms.

   - Number of treatment arms.

   - Unit of randomization.

2. **How to analyze an experiment after.**

   - Collect appropriate data.

   - Check for proper randomization.

   - Obtaining estimates.

   - Thinking about effect sizes.