Advanced Causal Inference for Product Analytics in R

By Joanne Rodrigues

About me

- Master’s degrees in Mathematics, Political Science and Demography
- Enterprise Data Scientist and Manager
- Author of Product Analytics: Applied Data Science Techniques for Actionable Consumer Insights
- Founder of ClinicPriceCheck.com and SlidingScaleHealth.org
- R developer for over 10+ years

Outline of Talk

I. **Introduction**
   A. Causal Inference from Observational Data
   B. Causal Inference vs Prediction
   C. Conceptualization, Operationalization and Metrics
II. **Advanced Techniques**
   A. Regression Discontinuity
   B. Statistical Matching
III. **Observable Patterns?**
   A. Heuristics for Causal Inference
   B. Questions?

Causal inference from observational data

- Causal inference is when we use randomization to isolate and quantify the impact of treatment on an outcome or multiple outcomes
- The gold standard of causal inference is experimentation
- Observational data is essentially mired in selection bias, or non-random selection into certain behavioral pattern or groups
- Can we infer causation from observational data?
  - **Maybe, and if ever only with a well-thought out design**
  - Very few out-of-the-box solutions

	Prediction	**Causal Inference**
Internal/ External Validity	Internal: Not Granted External: Validation on Test Sets	Internal: Granted by Design External: Possible, Much Harder
More Data	Improves with data	Generally does not improve
Generalizable	More, dependant on representativeness/size of the sample	Less, dependant on representativeness/size of the sample
Core Application	Predicting human behavior	Causes of human behavior
Discriminatory	Can be discriminatory; black box; confusing results for non-predictive results	Not easily discriminatory
When does it fail?	Failure to predict aberrant behavior; limits to prediction of human behavior	Failure to quantify the treatment effect for outliers
Product Applications	Future resources, recommendations, risk or fraud	Triggering behavior; behavior change; motivation; product creation/strategy

Conceptualization, Operationalization, and Metrics

- **Concept**: Abstract ideas used to explain phenomenon.
- **Operationalization**: Taking a concept and determining how it can be measured
- **Metrics**: Aggregated measure representing one data point or value, generally masking the distribution.
- More important with causal inference; Force us to cover the feature space
- How can we operationalize our concepts?
  - **Step 1:** Concept - Definition/Multiple Definitions, with all its parts
  - **Step 2:** List ‘perfect’ variables to cover or measure every part of definition, note what is immeasurable
  - **Step 3:** List all variables that currently exist in our product, could in any way be related these concepts
  - **Step 4:** Find the overlapping variables and note what is not covered.

How do we apply causal inference techniques to web products?

II. Regression Discontinuity

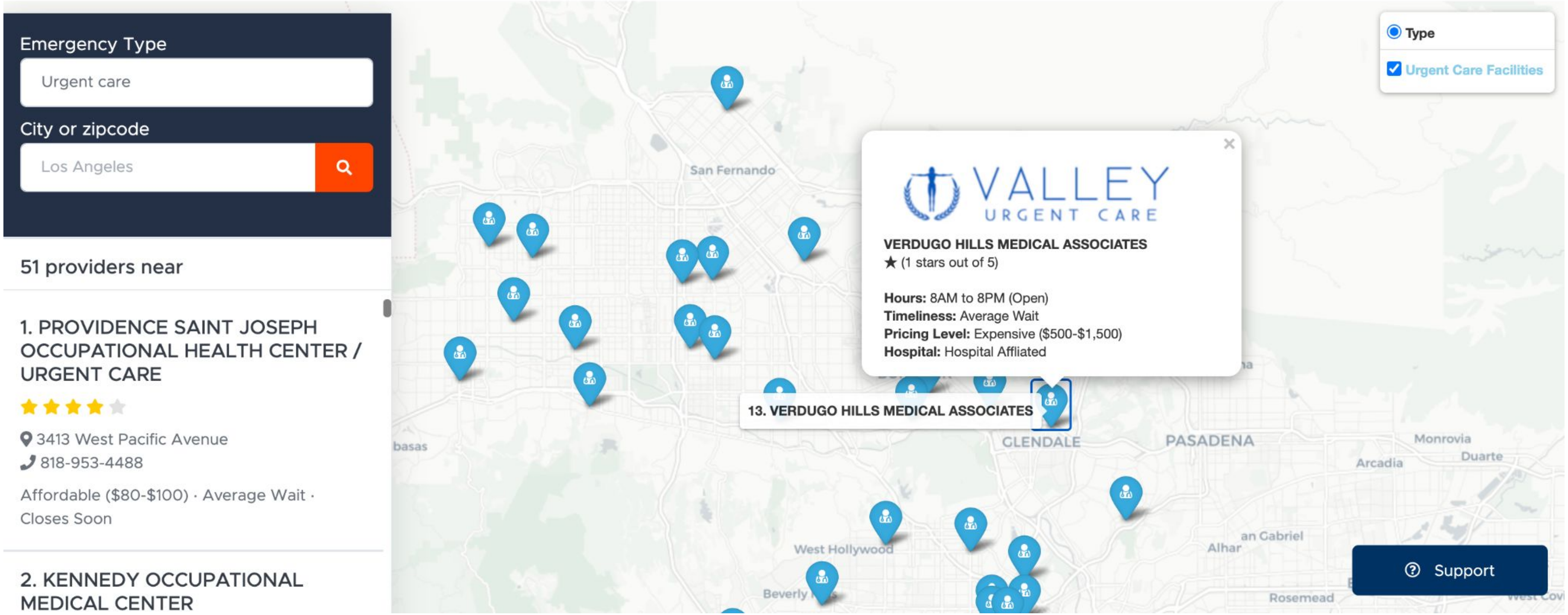
- A causal inference technique to operationalize a break in the treatment variable (relatively common in real-life given time or geographic break points)
- The idea is that as you get closer and closer to the break point, those observations at the cut point could randomly be on either side, mimicking the randomness of an experiment
- **LATE** (Local Average Treatment Effects) - RD is only defined in the limit at the break point
- **Pitfall:** Selection at the cut point, i.e. richer students more likely to pass national exam; richer candidates more likely to win elections

RD Example Design

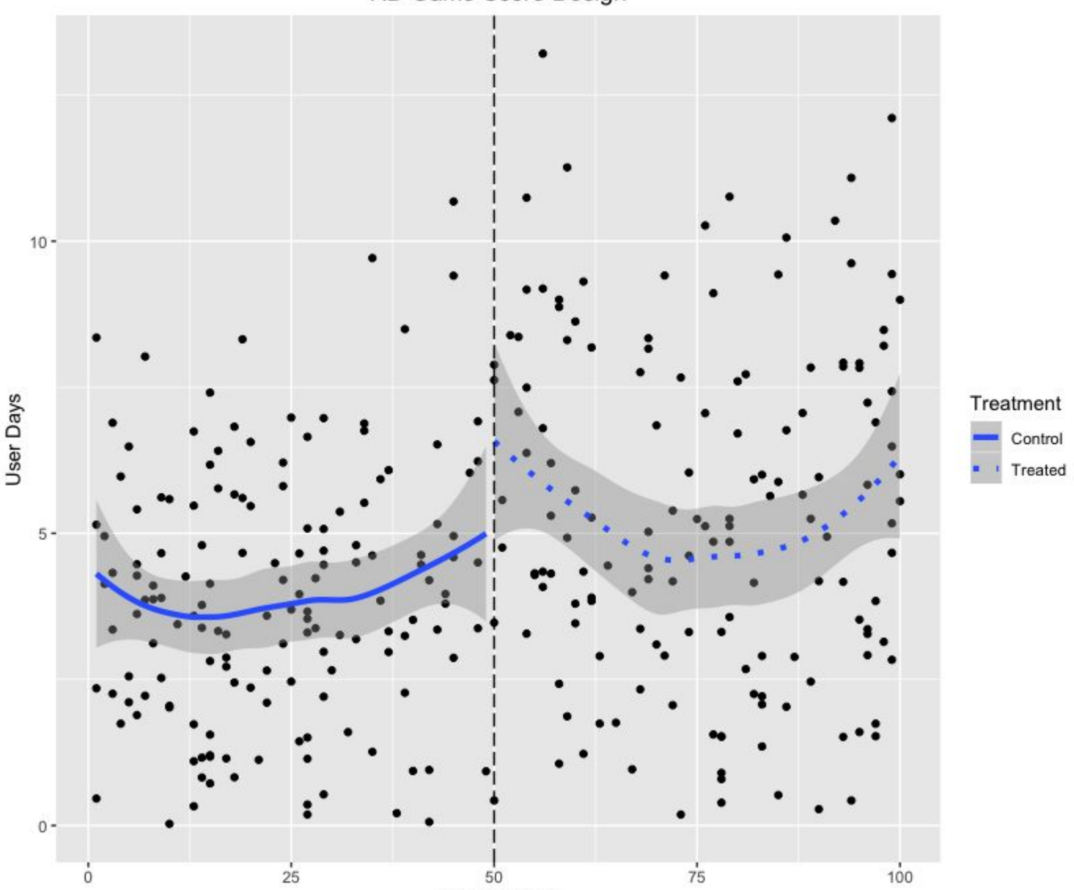
- Scenario: We want to incentivize users to ‘crowdsource’ information on hospital pricing, quality and wait times.
- We decide to use a badge system, where a certain number of reviews, likes and upvoting of prices leads to a gaining a ‘expert member’ badge.
- The points needed are 50.
- User cannot see any points.

ClinicPriceCheck

Upvoting and downvoting data

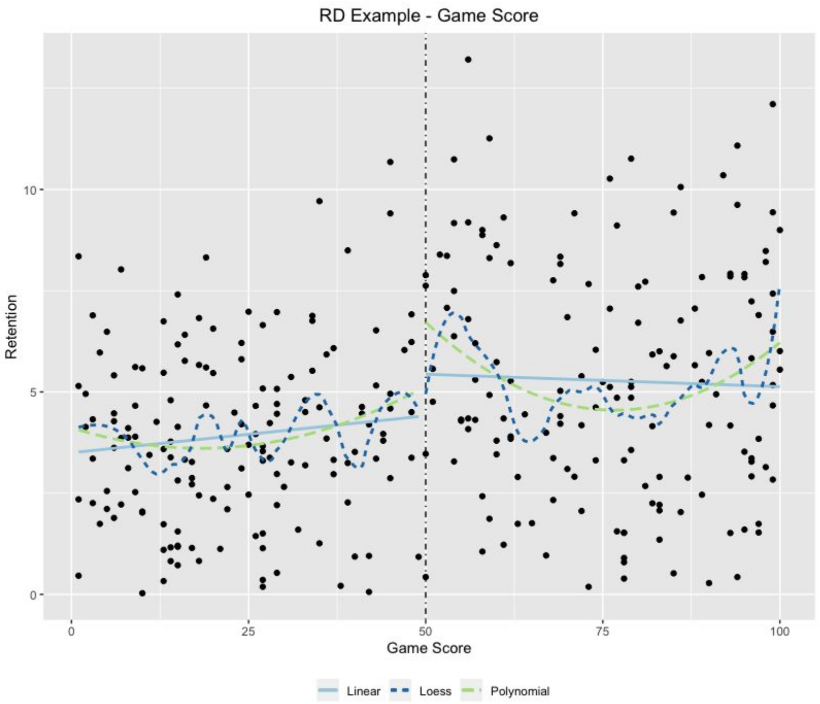


Retention
(outcome)



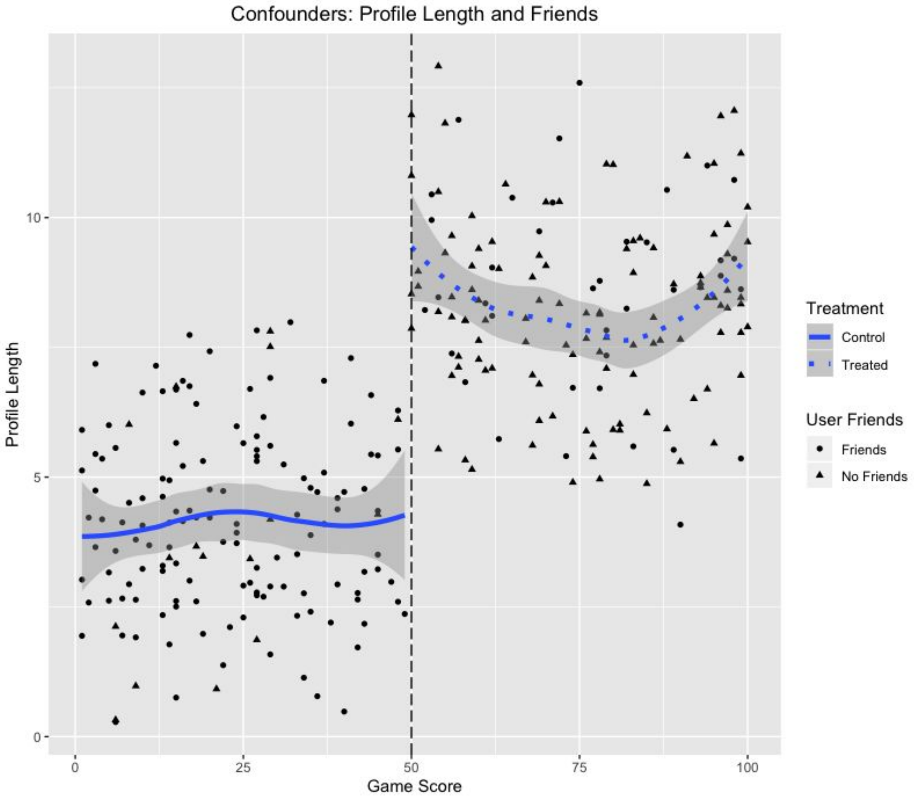
Score (treatment)

R Code - RD Design



Model	LATE Estimate
OLS model	0.98 user days
Quadratic model	1.65 user days
Loess model	.58 user days

Selection at the cut point

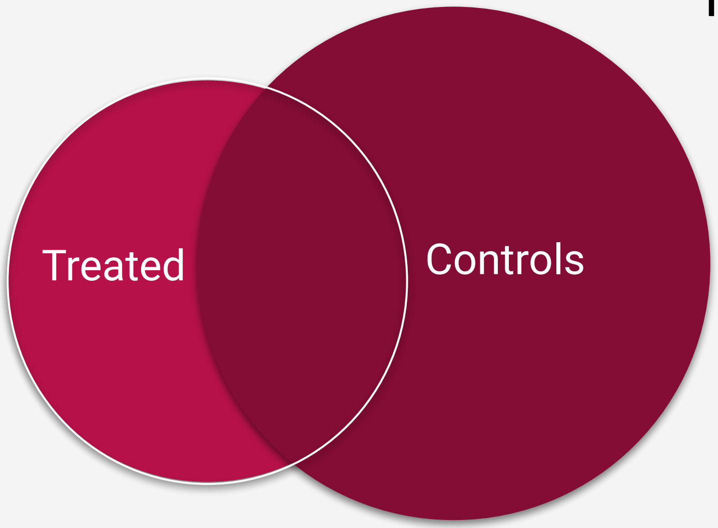


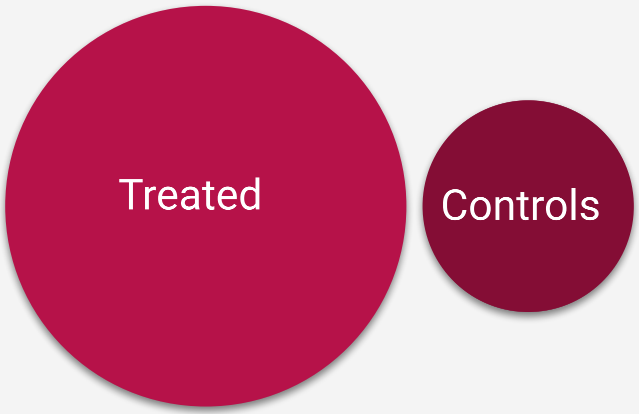
III. Statistical Matching

- Match user that look similar on other features to create a treatment and control sets that look ‘alike’ on confounder variables
- **ATE** (Average Treatment Effect), **ATC** (Average Treatment Effect on the Controls), **ATT** (Average Treatment Effect on the Treated)
- **ATT - ATC = ATE**
- How ‘alike’ are the groups? Achieve match balance, where the confounders are statistically non-differentiable in treatment and control.
- **Pitfalls:** No match balance and we have not achieved coverage of the confounders.
- Algorithms: Propensity Score, GenMatch

Matching - Visualization

Feature Space





Average Matching Situation

Matching Fails!

Matching Example

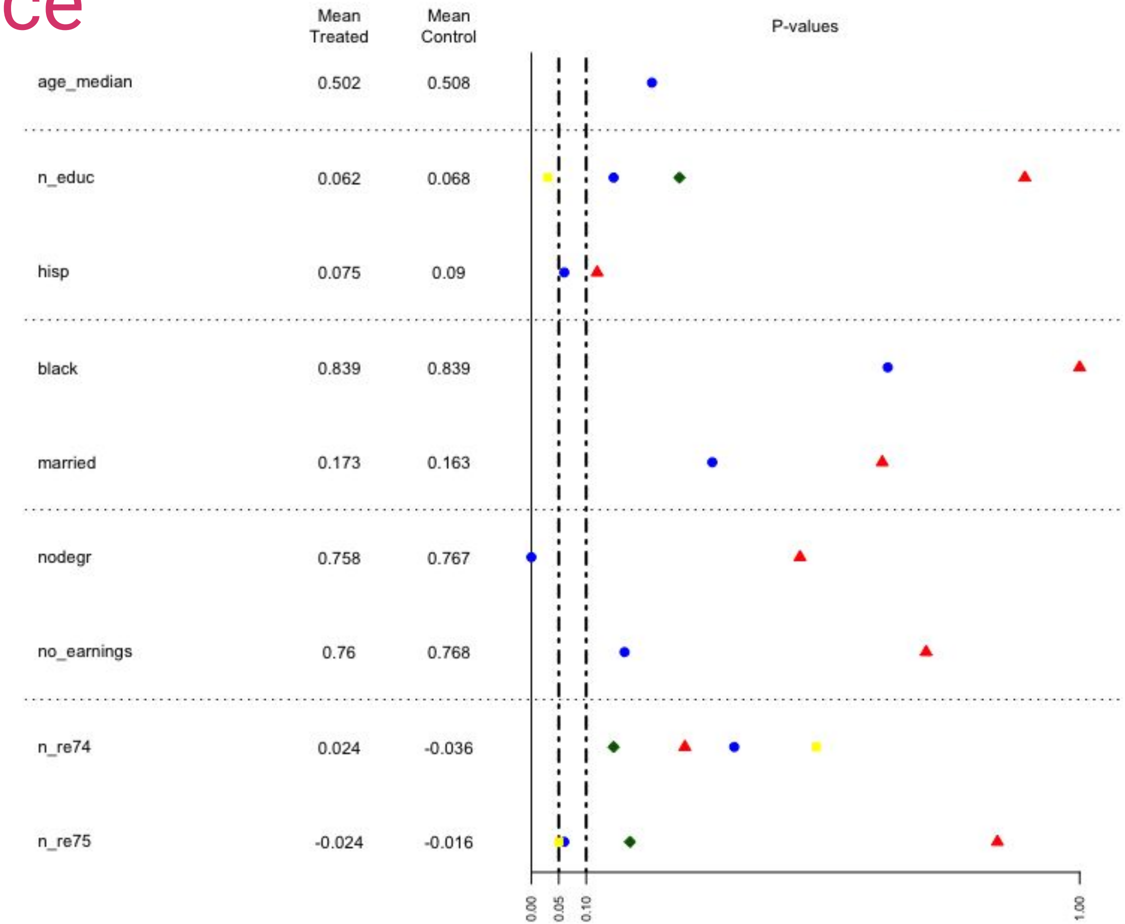
Scenario: We have a user funnel where we are trying to get users to fill out and submit their hospital financial assistance applications. We're seeing a large amount of drop-off at Step 4. **What types of treatments should we be testing to lower drop-off?**

Demo Web Example

User Funnel:

1. Calculator > 2. Advanced Calculator > 3. Sign-up > 4. Print and Complete Application > 5. Upload Documents > 6. Payment

Match Balance



Observable Patterns?

Hill's Causality Conditions

Ideally, you should see more than one of these conditions.

1. **Strength of the Effect:** Large proportional effects (odds ratio; only in extreme situations)
2. **Consistency:** High correlations in different places
3. **Specificity of the Association:** Linkages in the association
4. **Temporality:** Lagged effects
5. **Dosage Effects:** Strength of effect increase (dosage models)
6. **Plausality:** The smell test
7. **Coherence:** Does it mesh with your theory
8. **Experiment**: Use an experiment
9. **Analogy**: Similar comparison

Take Home Exercise

Link to Google Document:
https://docs.google.com/document/d/1c8WH6ZI340VIE7aluWFq1KdsutI-twLVMq2P6AixZCk/edit?usp=sharing

Bibliography

Dehejia, Rajeev, and Wahba, Sadek. "Causal Effects in Non-Experimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, no. 448 (December 1999): 1053-1062.

Dehejia, Rajeev, and Wahba, Sadek. "Propensity Score Matching Methods for Non-Experimental Causal Studies," *Review of Economics and Statistics*, 84 (February 2002): 151-161.

McCrary, Justin. "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*. 142, no. 2 (2008), 698-714.

Rosenbaum, Paul, and Rubin, Donald. “Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome.” *Journal of Royal Statistical Society*, B45 (1983): 212-218.

Rosenbaum, Paul, and Rubin, Donald. “The central role of the propensity score in observational studies for causal effects.” *Biometrika,* 70 (1983): 41-55.

Rosenbaum, Paul, and Rubin, Donald. “Reducing bias in observational studies using subclassification on the propensity score.” *Journal of the American Statistical Association*, 79 (1984): 516-524.

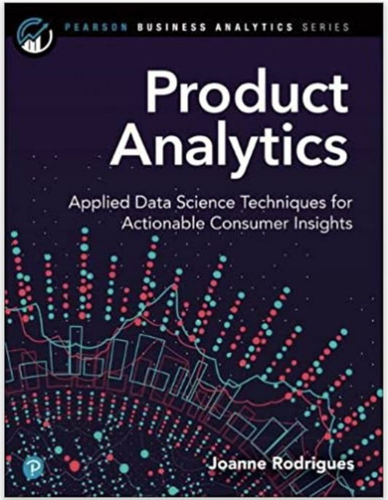
Rubin, Donald. “Estimating causal effects of treatments in randomized and nonrandomized studies.” *Journal of Educational Psychology*, 66, no. 5 (1974): 688-701.

Rubin, Donald. *Matched Sampling for Causal Effects*. Cambridge: Cambridge University Press, 2006.

Sekhon, Jasdeep."Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R." *Journal of Statistical Software,* 42, no. 7(2011): 1-52.

Hill, Bradford. The environment and disease: association or causation? *Proc R Soc Med.* 1965; 58:295–300

Contact me



joannecrodrigues@gmail.com