

INFO251 – Applied Machine Learning

Lab 2
Suraj R. Nair

Announcements

- **Problem Set 1** solutions are posted
 - **Problem Set 2** due Monday February 6, 8AM
-

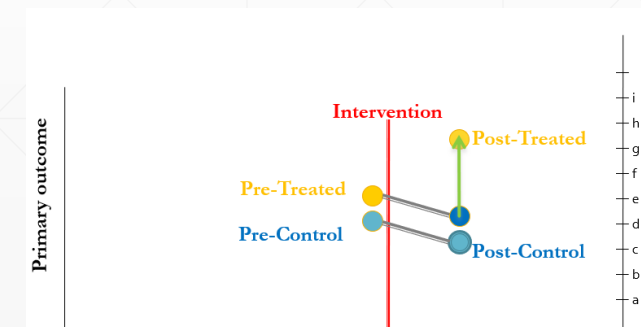
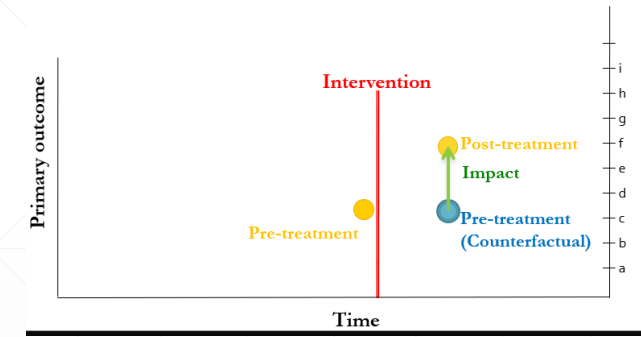
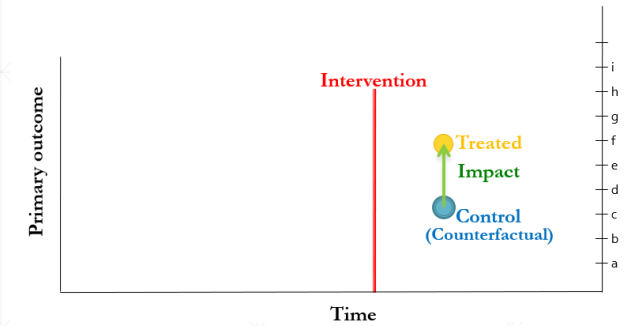
Today: Regression and Impact Evaluation

- Linear Regression
 - Dummy variables
- Difference-in-differences

Today's tools: `pandas` and `statsmodels`

Quick Review / Group Exercise

Design	Key Identifying Assumption	Confounds?
Randomized experiment (T v/s C)	?	?
Pre v/s Post	?	?
Double Difference	?	?

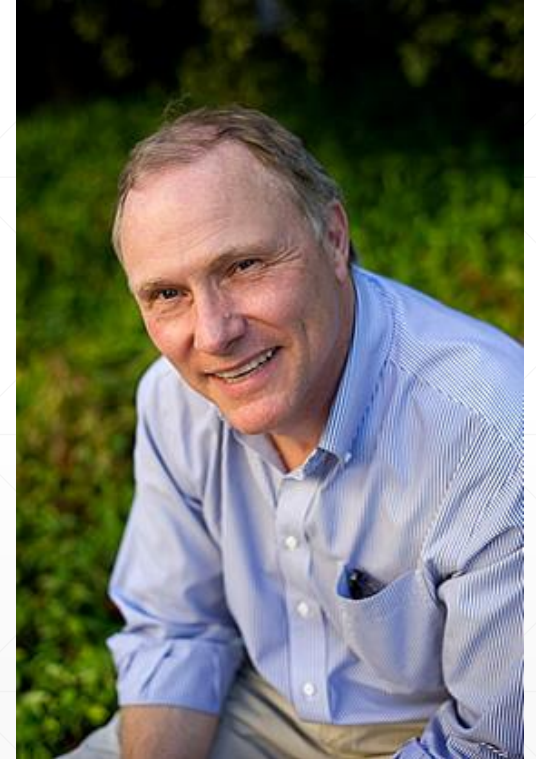


Quick Review

Design	Key Identifying Assumption	Confounds?
Randomized experiment (T v/s C)	<i>Outcomes</i> in Treatment and Control <i>would have</i> been the same in the absence of treatment	Differential selection / non-compliance / attrition
Pre v/s Post	<i>Outcomes</i> pre- and post-treatment <i>would have</i> been the same in the absence of treatment	Temporal Trends, Seasonality, etc.
Double Difference	<i>Trends (or changes over time)</i> in treatment group and control group <i>would have</i> been the same in the absence of treatment	Differential trends

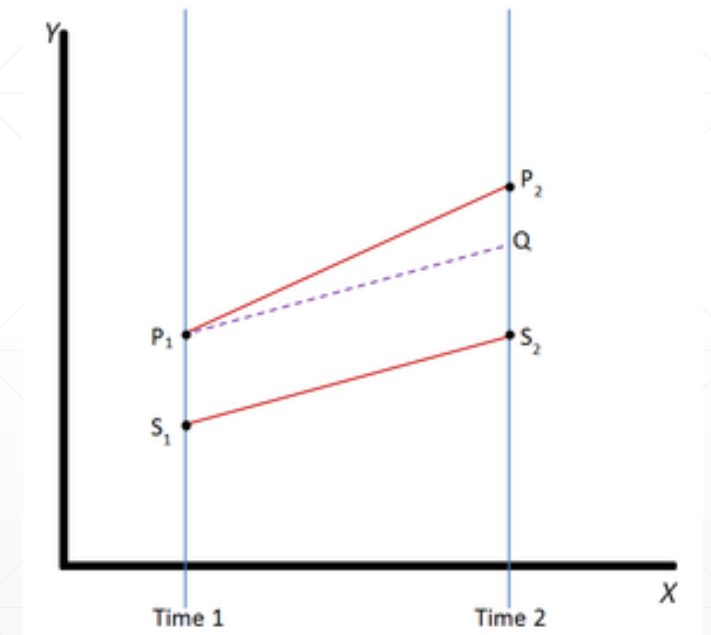
Today's Data: Minimum Wage and Employment

- **Source:** [Card and Krueger \(1994\)](#). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review* Vol. 90 No. 5.
- April 1992: NJ raises the minimum wage from \$4.25/hour to \$5.05/hour.
- Card and Krueger conducted a survey with 410 fast food restaurants in NJ and Eastern PA, once in March 1992 and once in December 1992.
 - Number of employees
 - Region, chain, hours of operation, number of registers, other covariates



Difference-in-differences

- Compare **changes in outcomes** in the treatment group and the control group
- Key assumptions:
 - “Parallel trends”: Trends in pre-treatment outcomes are the same in treatment and control groups
 - Composition of treatment and control groups stable across time
 - No spillover effects



Difference-in-differences

- $Y = B_0 + B_1 \cdot \text{Time} + B_2 \cdot \text{Treatment} + B_3 \cdot (\text{Time} \cdot \text{Treatment}) + B_4 \cdot \text{Covariates}$

