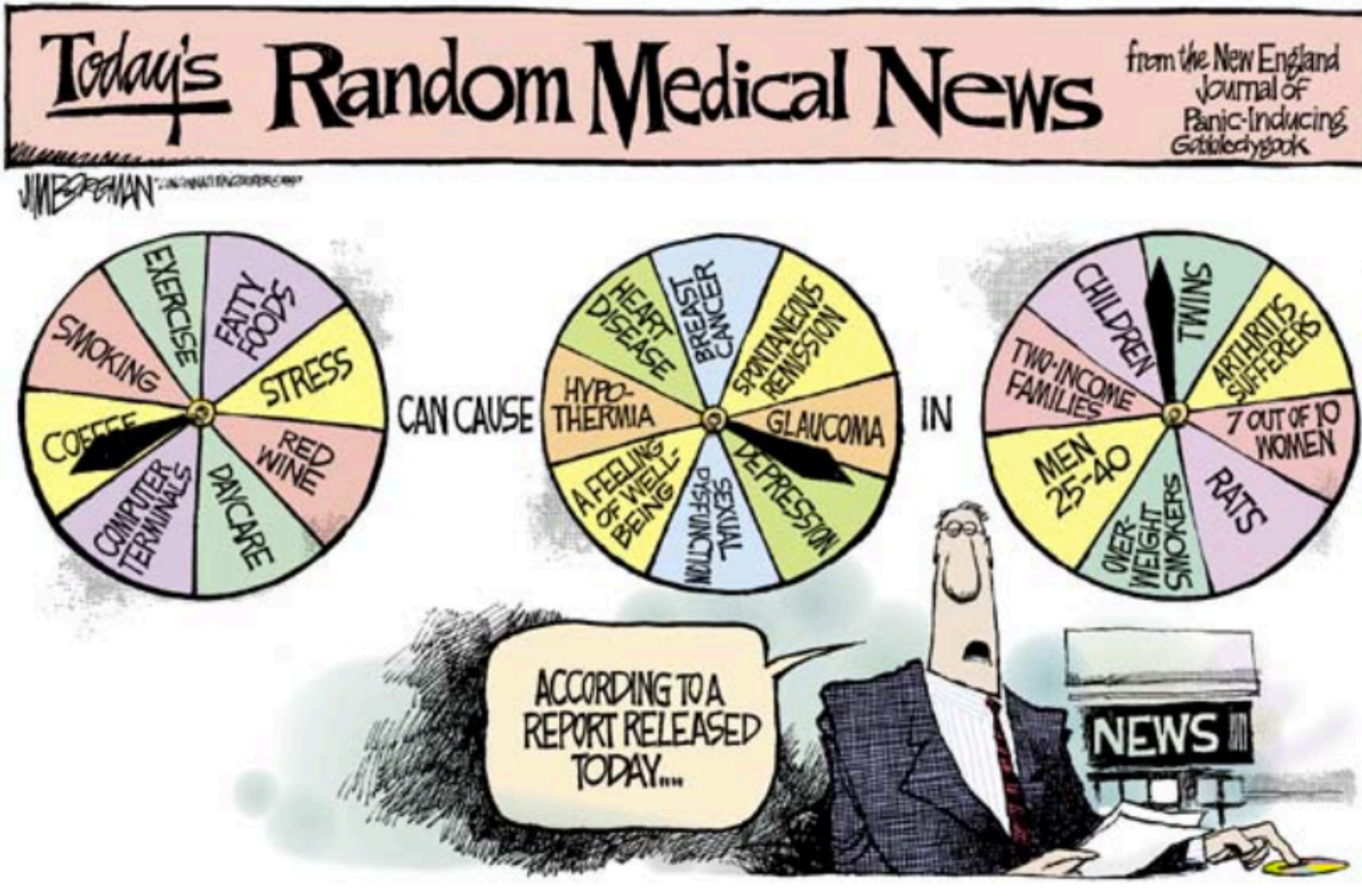
Making Policy with Data

An Introductory Course on Policy Evaluation

Lecture 4. Selection on Observables

Instructor: Prof Yiqing Xu

May 9


Today's Random Medical News
from the New England Journal of Panic-Inducing Gobbledygook
JIM BORGMAN
SMOKING
EXERCISE
FATTY FOODS
STRESS
RED WINE
DAYCARE
COMPUTER TERMINALS
COFFEE
CAN CAUSE
HEART DISEASE
BREAST CANCER
SPONTANEOUS REMISSION
GLAUCOMA
DEPRESSION
SEXUAL DYSFUNCTION
A FEELING OF WELL-BEING
HYPO-THERMIA
IN
CHILDREN
TWINS
ARTHRITIS SUFFERERS
7 OUT OF 10 WOMEN
RATS
OVER-WEIGHT SMOKERS
MEN 25-40
TWO-INCOME FAMILIES
ACCORDING TO A REPORT RELEASED TODAY....
NEWS

“Random” News

1. **Lancet 2001**: negative correlation between coronary heart disease mortality and level of vitamin C in bloodstream (controlling forage, gender, blood pressure, diabetes, and smoking)

2. **Lancet 2002**: no effect of vitamin C on mortality in controlled placebo trial (controlling for nothing)

3. **Lancet 2003**: comparing among individuals with the same age, gender, blood pressure, diabetes, and smoking, those with higher vitamin C levels have lower levels of obesity, lower levels of alcohol consumption, are less likely to grow up in working class, etc.

Observational Studies

1. Randomization forms gold standard for causal inference, because it balances **observed** and **unobserved** confounders

2. Cannot always randomize so we do observational studies, where we **adjust** for the **observed covariates** and **hope** that unobservables are balanced

3. Better than hoping: **design** observational study to approximate an experiment

> The planner of an observational study should always ask himself: How would the study be conducted if it were possible to do it b controlled experimentation. (Cochran 1965)

The Good, the Bad, and the Ugly

Treatments, Covariates, Outcomes

- Randomized Experiment: Well-defined treatment, clear distinction between covariates and outcomes, control of assignment mechanism
- Better Observational Study: Well-defined treatment, clear distinction between covariates and outcomes, precise knowledge of assignment mechanism
- Can convincingly answer the following question: Why do two units who are identical on measured covariates receive different treatments?
- Poorer Observational Study: Hard to say when treatment began or what the treatment really is. Distinction between covariates and outcomes is blurred, so problems that arise in experiments seem to be avoided but are in fact just ignored. No precise knowledge of assignment mechanism.

The Good, the Bad, and the Ugly

How were treatments assigned?

- Randomized Experiment: Random assignment
- Better Observational Study: Assignment is not random, but circumstances for the study were chosen so that treatment seems haphazard, or at least not obviously related to potential outcomes (sometimes we refer to these as natural or quasi-experiments)
- Poorer Observational Study: No attention given to assignment process, units self-select into treatment based on potential outcomes

The Good, the Bad, and the Ugly

Were treated and controls comparable?

- Randomized Experiment: Balance table for observables.
- Better Observational Study: Balance table for observables.
- Poorer Observational Study: No direct assessment of comparability is presented.

Example: The Effect of Class Size

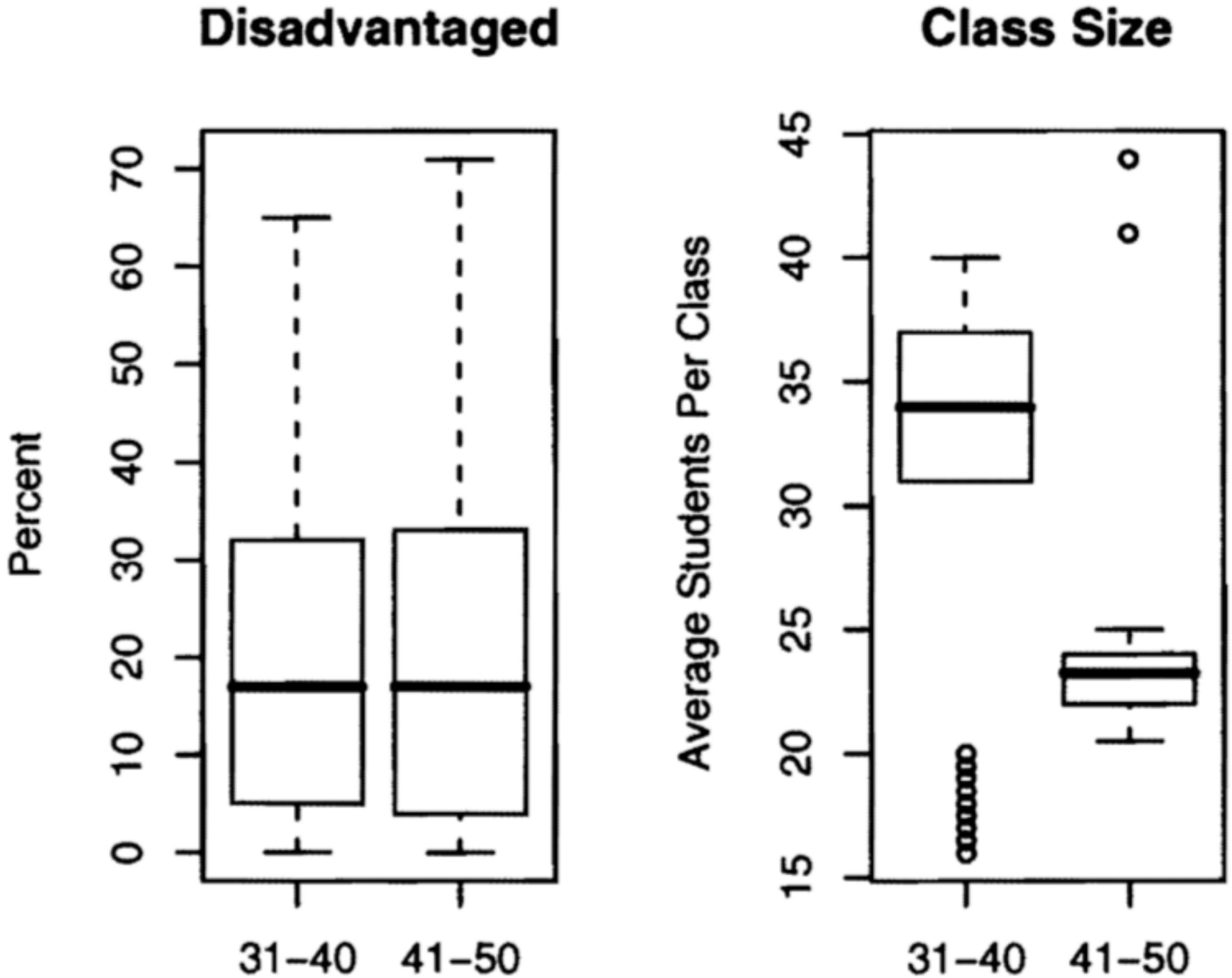
- Educators and labor economists are very interested in studying the effect of class size on learning, e.g. does smaller class size cause students to achieve higher math and verbal scores?

- Causal effects of class size on pupil achievement is difficult to measure. **Why?**

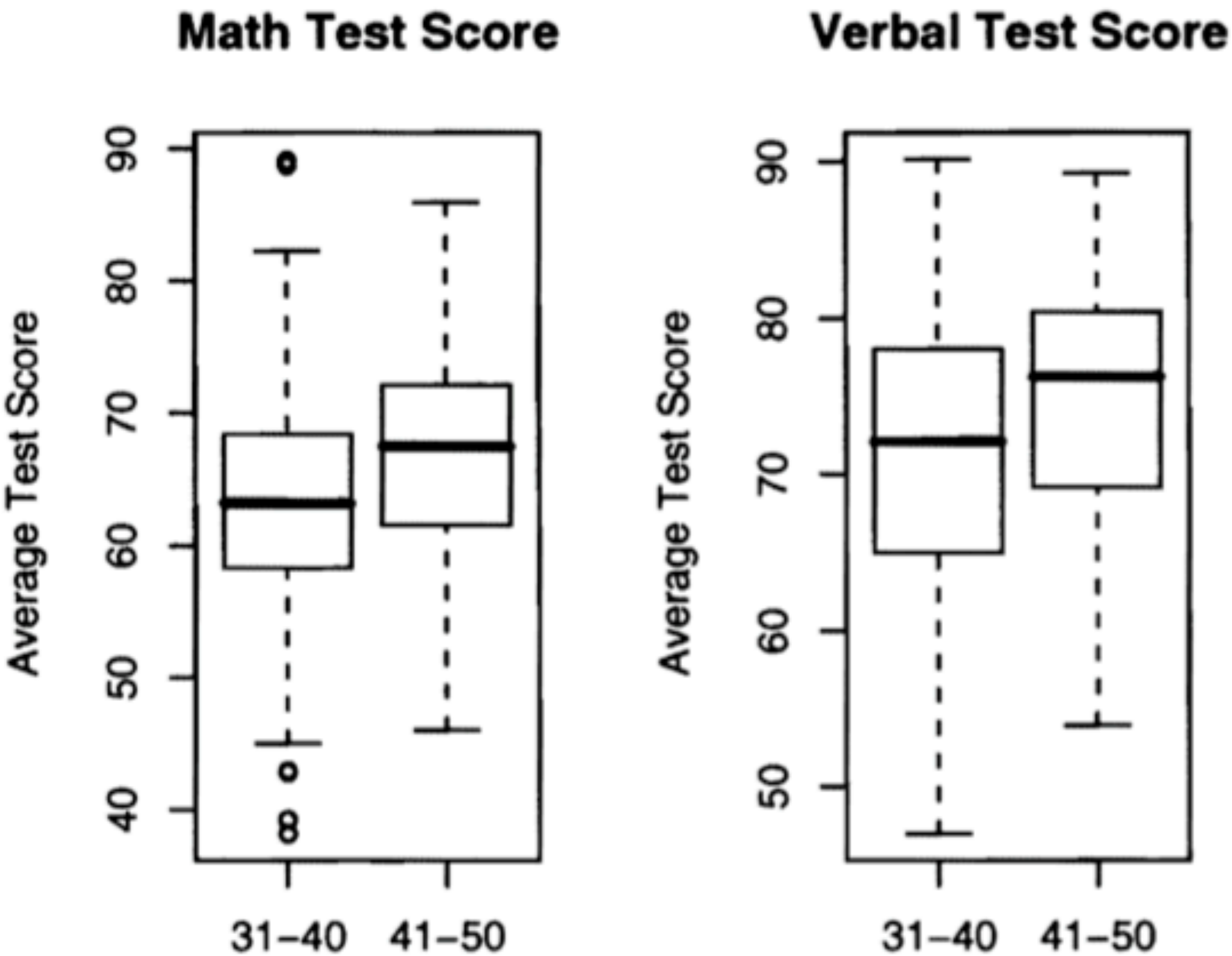
- Since 1969, Maimonides' rule has been used to determine class size in Israeli public schools.

> "Twenty-five children may be put in charge of one teacher. If the number in the class exceeds twenty-five but is not more than forty, he should have an assistant to help with instruction. If there are more than forty, two teachers must be appointed."

Angrist and Lavy (1999)



Angrist and Lavy (1999)



Plan

- Causal Inference with observational data is hard because of selection bias
- If we understand the treatment assignment mechanism very well, we can remove bias by conditioning — Selection On Observables (SOO)
- Plan
  - Stats background: conditional independence
  - Theory of identification under SOO
  - Methods of conditioning: (1) matching; (2) regression

Conditional Independence

Key to causal inference with observational data

Joint, Marginal, and Conditional!

For two discrete random variables X and Y the joint PMF $f_{X,Y}(x,y)$ gives the probability that $X = x$ and $Y = y$ for all x and y:

$$f_{X,Y}(x,y) = P(X = x \text{ and } Y = y) \qquad = P(Y = y \mid X = x)P(X = x)$$
$$= P(X = x \mid Y = y)P(Y = y)$$

Restrictions:

- $f_{X,Y}(x,y) \geq 0$ and $\sum_x \sum_y f_{X,Y}(x,y) = 1$.

Joint Probability Mass Function

- Should the U.S. allow more immigrants to come and live here?

Table: Joint Probabilities

		X: Education			
		less HS	HS	College	BA
Y: Support	oppose	0.07	0.21	0.17	0.14
	neutral	0.02	0.06	0.05	0.05
	favor	0.01	0.03	0.04	0.10

Marginal Probability Mass Function

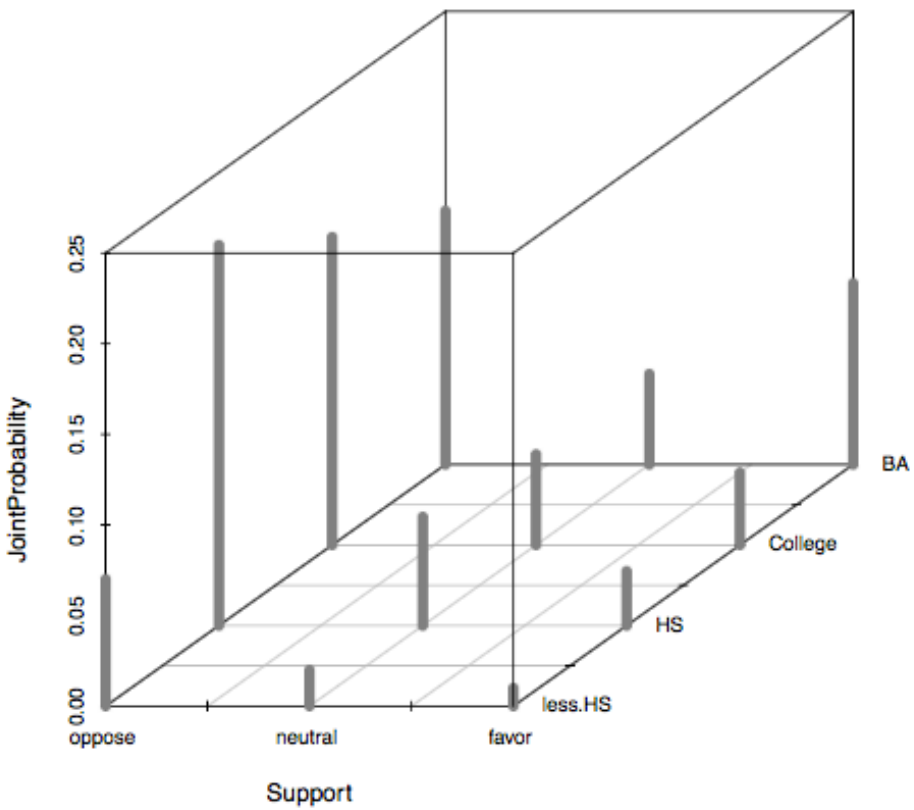
- Should the U.S. allow more immigrants to come and live here?

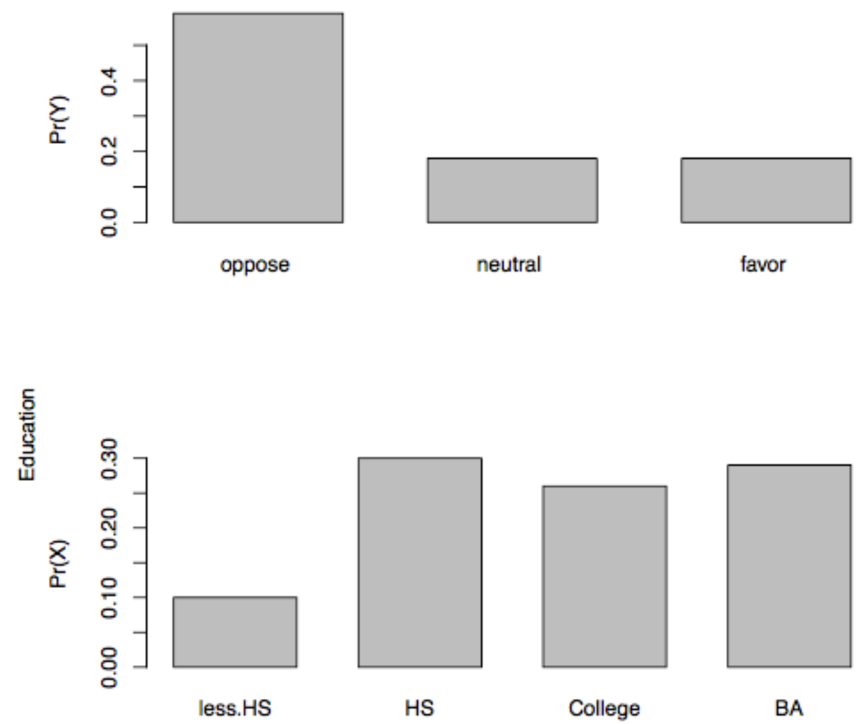
Table: Joint and Marginal Probabilities

		X: Education				
		less HS	HS	College	BA	Pr(Y)
Y: Support	oppose	0.07	0.21	0.17	0.14	0.60
	neutral	0.02	0.06	0.05	0.05	0.19
	favor	0.01	0.03	0.04	0.10	0.20

Joint and Marginal Probability Mass Function

- Should the U.S. allow more immigrants to come and live here?





Conditional PMF

- The conditional PMF of Y given X (we write Y|X) is the PMF of Y when X is known to be at a particular value X = x:

$$f_{Y|X}(y|x) = \frac{P(X = x \text{ and } Y = y)}{P(X = x)} = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X|Y}(y|x) f_Y(y)}{f_X(x)}$$

- Conditional PMFs are just like ordinary PMFs, but refer to a universe where the “conditioning event” (X = x) is known to have occurred.

- Conditional distributions are key in statistics because they informs us how the distribution of Y varies across different levels of X.

Conditional PMF

Table: Joint Probabilities f(x,y) and Marginal Probabilities f(x),f(y)

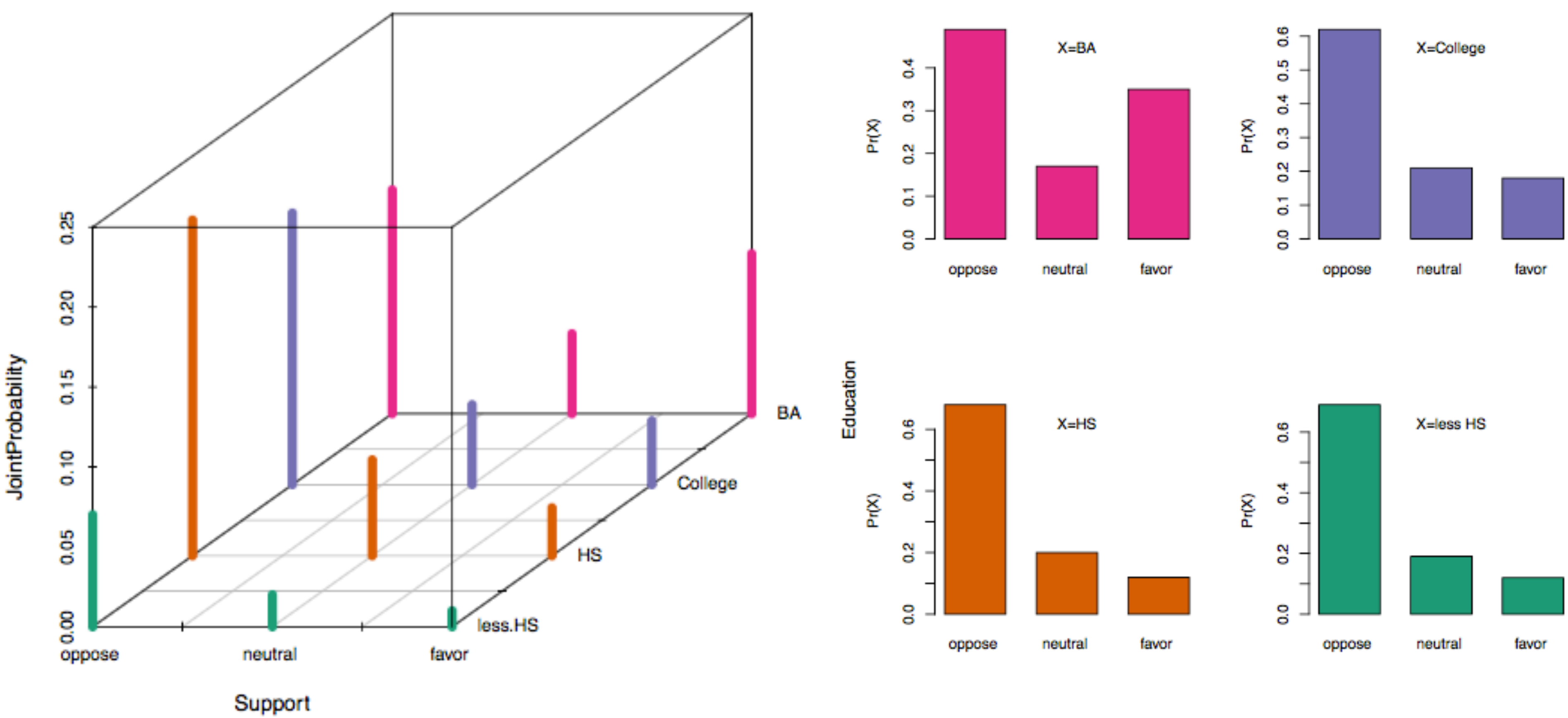
		Education				
	$f(x,y)$	less HS	HS	College	BA	$f(y)$
Support	oppose	0.07	0.21	0.17	0.14	0.60
	neutral	0.02	0.06	0.05	0.05	0.19
	favor	0.01	0.03	0.04	0.10	0.20
	$f(x)$	0.10	0.31	0.27	0.29	1.00

Table: Conditional $f(y|x)$ Probabilities

		Education				
	$f(y|x)$	less HS	HS	College	BA	
Support	oppose	0.69	0.68	0.62	0.49	0.60
	neutral	0.19	0.20	0.21	0.17	0.19
	favor	0.12	0.12	0.18	0.35	0.20
		1.00	1.00	1.00	1.00	1.00

Conditional PMF

- Should the U.S. allow more immigrants to come and live here?



Conditional Expectation

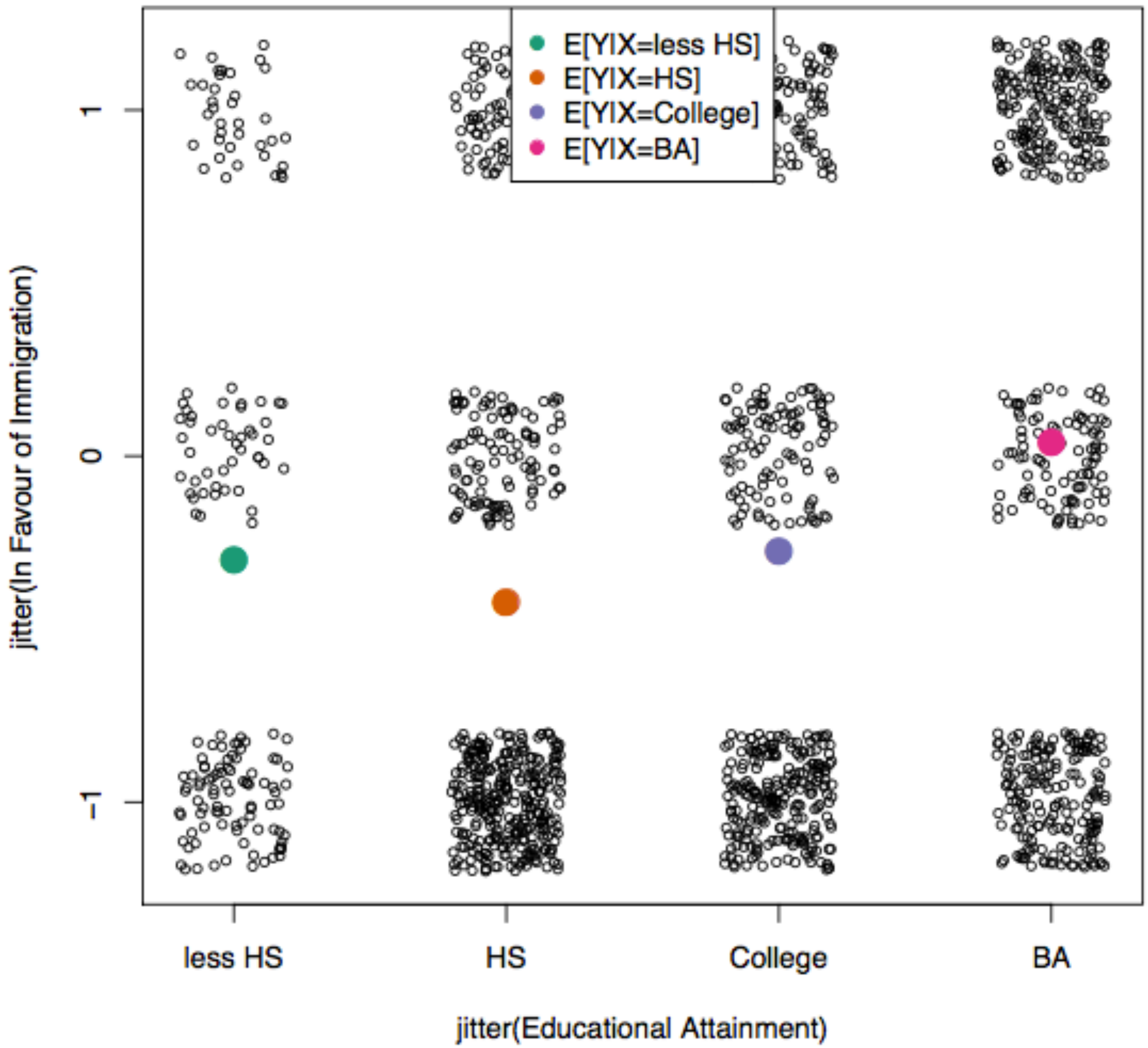
Definition (Conditional Expectation Discrete Case)

Let Y and X be discrete RVs, then the conditional expectation of Y given the event $X = x$ is given by:

$$E[Y|X = x] = \sum_y y\, P(Y = y|X = x) = \sum_y y\, f_{Y|X}(y|x)$$

Read: Given X = x, what is the average value of Y?

Conditional Expectation



Conditional Independence

Definition (Conditional Independence of Random Variables)

Random variables Y and X are conditionally independent given Z iff

$$f_{XY|Z}(x, y|z) = f_{Y|Z}(y|z) \cdot f_{X|Z}(x|z)$$

for all triplets (x, y, z).

Conditional independence implies that

$$P(Y = y|X = x, Z = z) = P(Y = y|Z = z)$$

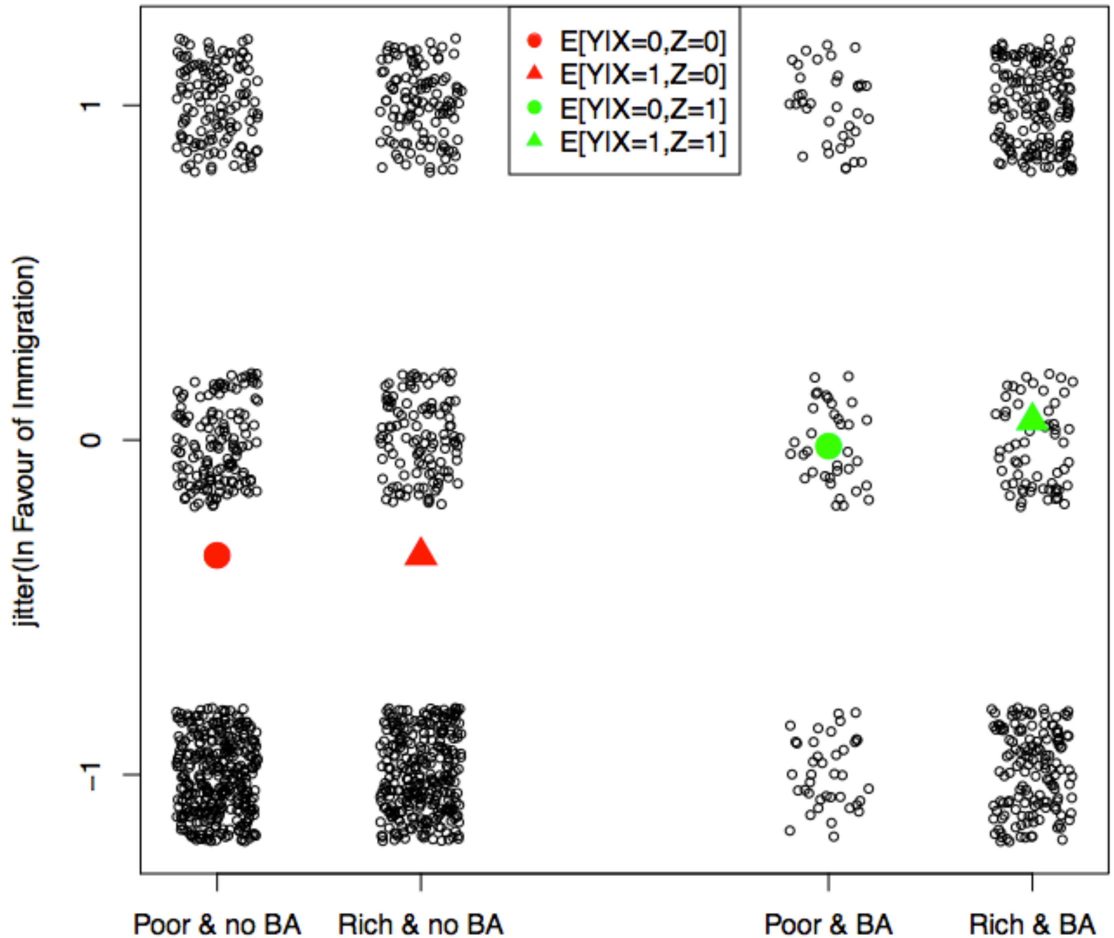
and thus

$$E[Y|X = x, Z = z] = E[Y|Z = z]$$

we usually write $Y \perp\!\!\!\perp X|Z$

Example: Y: Support X: Income Z: Education

Conditional Independence



Read: Wealth is independent of support for immigration conditional on education.

Identification under SOO

when we understand how “treatment” is assigned

Example: Private College Payoff

- All sorts of things are correlated with college attendance decisions and later earning
  - Ability, diligence, personalities, ambitions, family connections
- Controlling for such a wide range of factors seems daunting
- Stacy Dale and Alan Kruger: controlling for the characteristics of colleges to which students applied and were admitted

The College Matching Matrix

TABLE 2.1
The college matching matrix

		Private			Public			
Applicant group	Student	Ivy	Leafy	Smart	All State	Tall State	Altered State	1996 earnings
A	1		Reject	Admit		Admit		110,000
	2		Reject	Admit		Admit		100,000
	3		Reject	Admit		Admit		110,000
B	4	Admit			Admit		Admit	60,000
	5	Admit			Admit		Admit	30,000
C	6		Admit					115,000
	7		Admit					75,000
D	8	Reject			Admit	Admit		90,000
	9	Reject			Admit	Admit		60,000

Example: Private College Payoff

- Uma and Harvey:
  - Both applied to Harvard and U-Mass; both admitted
  - Harvey chooses Harvard while Uma opted for U-Mass
  - We can compare earnings of Uma and Harvey after they graduate
- What are being control for in such a comparison?
  - Lots of things, such as ability (in the eyes of the admission committees), ambition
- What are not being control for?
  - Reasons why Uma choose U-Mass instead of Harvard: a successful uncle graduated from U-Mass, a best friend who chose U-Mass, missing the deadline of a scholarship for Harvard, etc.
  - We hope that these factors are not highly correlated with earning **potentials**

Control for "Selection"

	No selection controls			Selection controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	.135	.095	.086	.007	.003	.013
	(.055)	(.052)	(.034)	(.038)	(.039)	(.025)
Own SAT score ÷ 100		.048	.016		.033	.001
		(.009)	(.007)		(.007)	(.007)
Log parental income			.219			.190
			(.022)			(.023)
Female			−.403			−.395
			(.018)			(.021)
Black			.005			−.040
			(.041)			(.042)
Hispanic			.062			.032
			(.072)			(.070)
Asian			.170			.145
			(.074)			(.068)
Other/missing race			−.074			−.079
			(.157)			(.156)
High school top 10%			.095			.082
			(.027)			(.028)
High school rank missing			.019			.015
			(.033)			(.037)
Athlete			.123			.115
			(.025)			(.027)
Selectivity-group dummies	No	No	No	Yes	Yes	Yes

Review

- Why We Often End up Doing Observational Studies
- Good and Bad Observational Studies
- Conditional Independence
- Identification under SOO — College Attendance Example

The Good, the Bad, and the Ugly

Treatments, Covariates, Outcomes

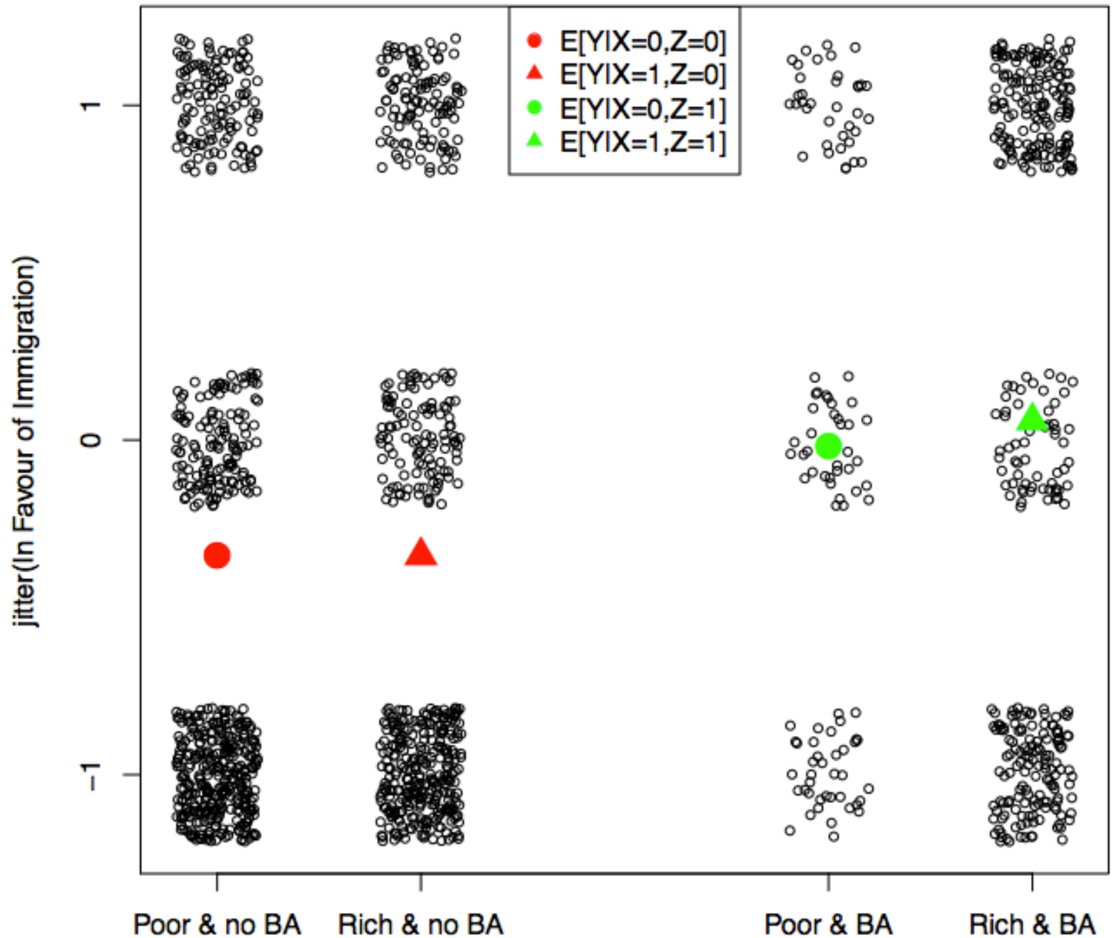
- Randomized Experiment: Well-defined treatment, clear distinction between covariates and outcomes, control of assignment mechanism

- Better Observational Study: Well-defined treatment, clear distinction between covariates and outcomes, precise knowledge of assignment mechanism

- Can convincingly answer the following question: Why do two units who are identical on measured covariates receive different treatments?

- Poorer Observational Study: Hard to say when treatment began or what the treatment really is. Distinction between covariates and outcomes is blurred, so problems that arise in experiments seem to be avoided but are in fact just ignored. No precise knowledge of assignment mechanism.

Conditional Independence



Read: Wealth is independent of support for immigration conditional on education.

The College Matching Matrix

TABLE 2.1
The college matching matrix

		Private			Public			
Applicant group	Student	Ivy	Leafy	Smart	All State	Tall State	Altered State	1996 earnings
A	1		Reject	Admit		Admit		110,000
	2		Reject	Admit		Admit		100,000
	3		Reject	Admit		Admit		110,000
B	4	Admit			Admit		Admit	60,000
	5	Admit			Admit		Admit	30,000
C	6		Admit					115,000
	7		Admit					75,000
D	8	Reject			Admit	Admit		90,000
	9	Reject			Admit	Admit		60,000

Conditional Ignorability

- Units: $i = 1, ..., n$
- Treatment: $D_i \in \{0, 1\}$
- Potential outcomes: $Y_i(d)$, where $d = 0, 1$
- Quantities of interest:

$$\begin{aligned} \text{ATE:} \quad \alpha_{ATE} &\equiv \mathbb{E}[Y_i(1) - Y_i(0)] \\ \text{ATT:} \quad \alpha_{ATT} &\equiv \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1] \end{aligned}$$

Question: Can we identify α_{ATE} and α_{ATT} when D_i is not randomized?

Conditional Ignorability

- Pre-treatment covariates: $X_i = [X_{i1}, ..., X_{iK}]^\top \in \mathcal{X}$
  - Predetermined and causally precedent with respect to D_i
  - Examples: Sex, race, age, etc.
  - X_i may be correlated with both D_i and $Y_i(d)$, thereby confounding the causal relationship
  - Excludes correlates that are potentially affected by D_i (post-treatment covariates)

Conditional Ignorability

Recall that randomized experiments work because:

$$\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp D_i$$

Assumption: Conditional Ignorability

$$\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp D_i \mid X_i = x \quad \text{for any} \quad x \in \mathcal{X}$$

(a.k.a. exogeneity, unconfoundedness, selection on observables, no omitted variables)

Read: Among units with same values of X_i, D_i is "as-if" randomly assigned.

Assumption: Common Support

$$0 < \Pr(D_i = 1 \mid X_i = x) < 1 \quad \text{for any} \quad x \in \mathcal{X}$$

Read: For any value of X_i, unit could have received treatment or control

Identifying ATE

Intuition: Within strata of X, you have an experiment

Proof: for ATE, α, we have

Part 1. Identifiability of $\alpha(X)$:

$$\begin{aligned}\mathbb{E}[Y_{1i} - Y_{0i}|X_i = x] &= \mathbb{E}[Y_{1i}|X_i = x, D_i = 1] - \mathbb{E}[Y_{0i}|X_i = x, D_i = 0] \\ &= \mathbb{E}[Y_i|X_i = x, D_i = 1] - \mathbb{E}[Y_i|X_i = x, D_i = 0] \\ &= \mathbb{E}[\hat{\alpha}|X_i]\end{aligned}$$

Part 2. Common support gets you back to α:

$$\begin{aligned}\alpha_{ATE} &= \mathbb{E}[Y_{1i} - Y_{0i}] \\ &= \mathbb{E}\big[\mathbb{E}[Y_{1i} - Y_{0i}|X_i]\big] \quad \text{Why?} \\ &= \int \Big(\mathbb{E}[Y_i|D_i = 1, X_i = x] - \mathbb{E}[Y_i|D_i = 0, X_i = x]\Big)p(x)dX \\ &= \mathbb{E}\big[\mathbb{E}[\hat{\alpha}|X_i]\big] = \mathbb{E}[\hat{\alpha}]\end{aligned}$$

Identifying ATT

By the similar logic, α_{ATT} is also identified under the conditional ignorability and common support assumptions:

$$\alpha_{ATT} \;=\; \mathbb{E}[\hat{\alpha}(X_i) \mid D_i = 1]$$

Proof is similar:

$$\begin{aligned}
\alpha_{ATT} &= \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1] \\
&= \mathbb{E}[\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i, D_i = 1] \quad \text{What is outer } \mathbb{E} \text{ over?} \\
&= \int \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x, D_i = 1] p(x \mid D_i = 1) dx \\
&= \int \{\mathbb{E}[Y_i \mid X_i = x, D_i = 1] - \mathbb{E}[Y_i \mid X_i = x, D_i = 0]\} \, p(x \mid D_i = 1) dx \\
&= \mathbb{E}[\hat{\alpha}(x) \mid D_i = 1].
\end{aligned}$$

Is $\alpha_{ATE} = \alpha_{ATT}$ when CI holds?

"As-if" randomization

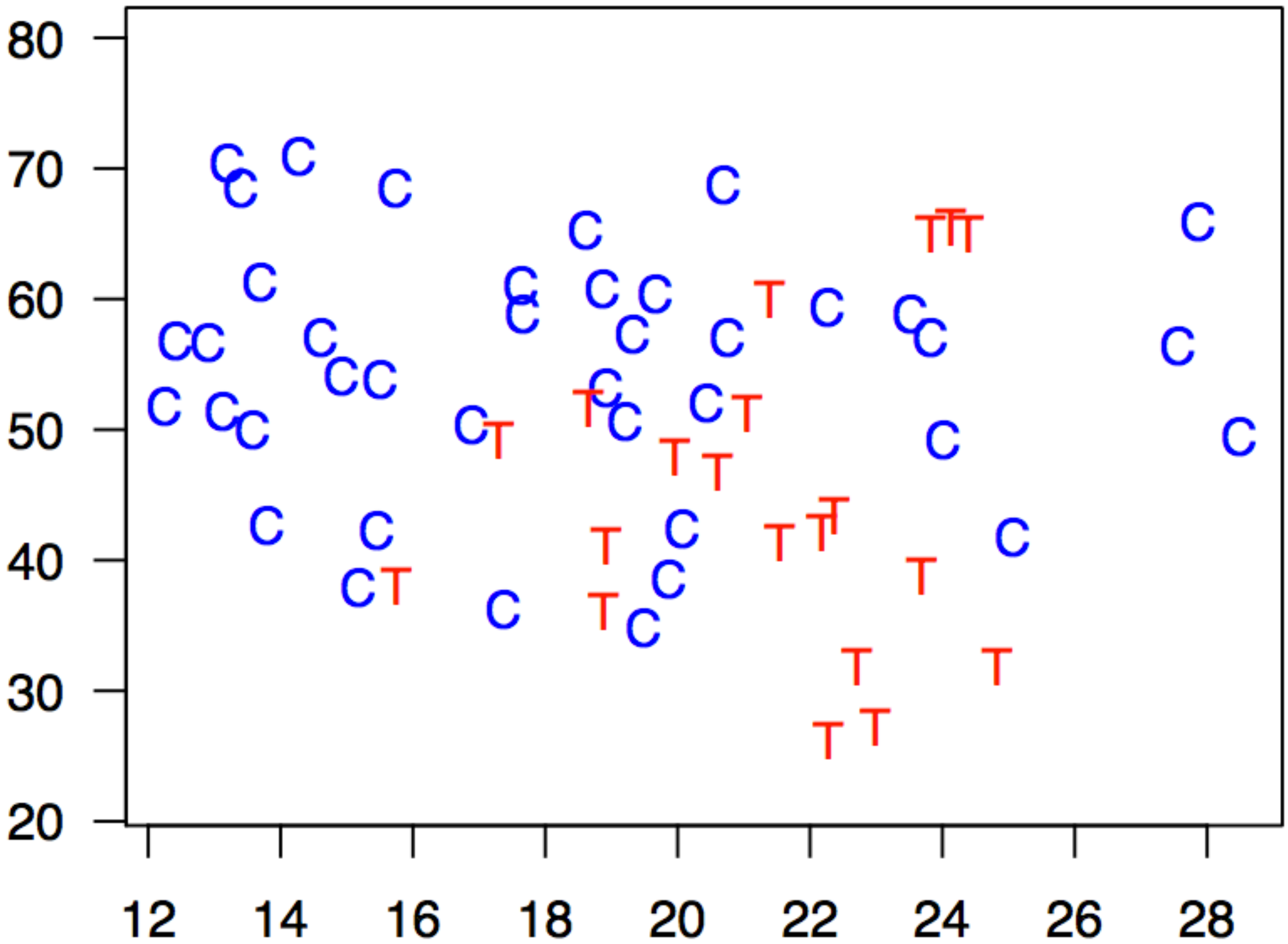
- Causal inference in observational studies often rests on this "SOO" (or CI) assumption
- A useful intuition: **"find strata of X in which you think an experiment is occurring"**
- Goal is to approximate a randomized experiment within subgroups
- Plausibility of your conditional ignorability: can argue that variation in treatment status within strata of X is random?

Methods of Conditioning

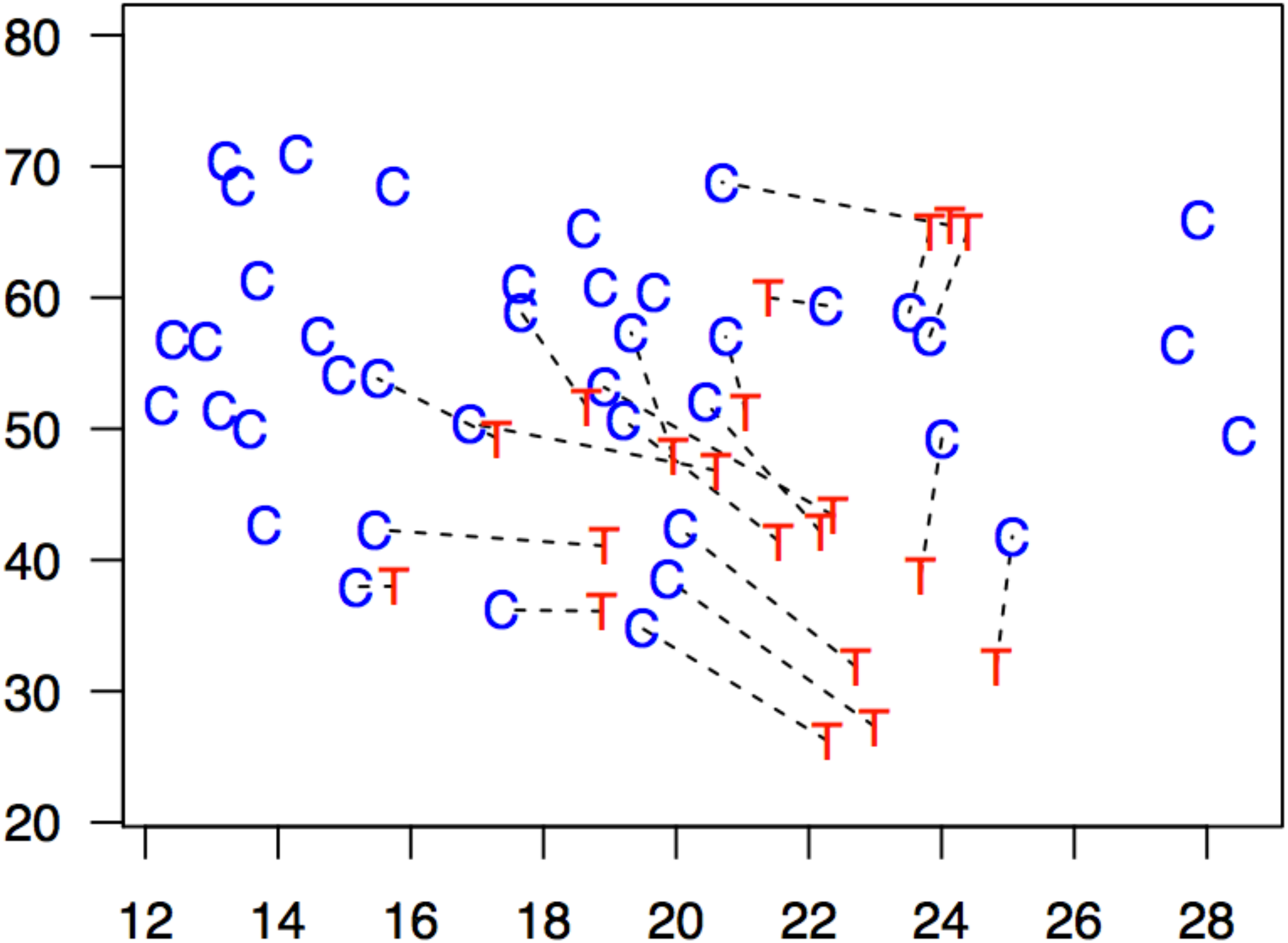
(1) Matching

Compare like with like with “as-if” random assignment

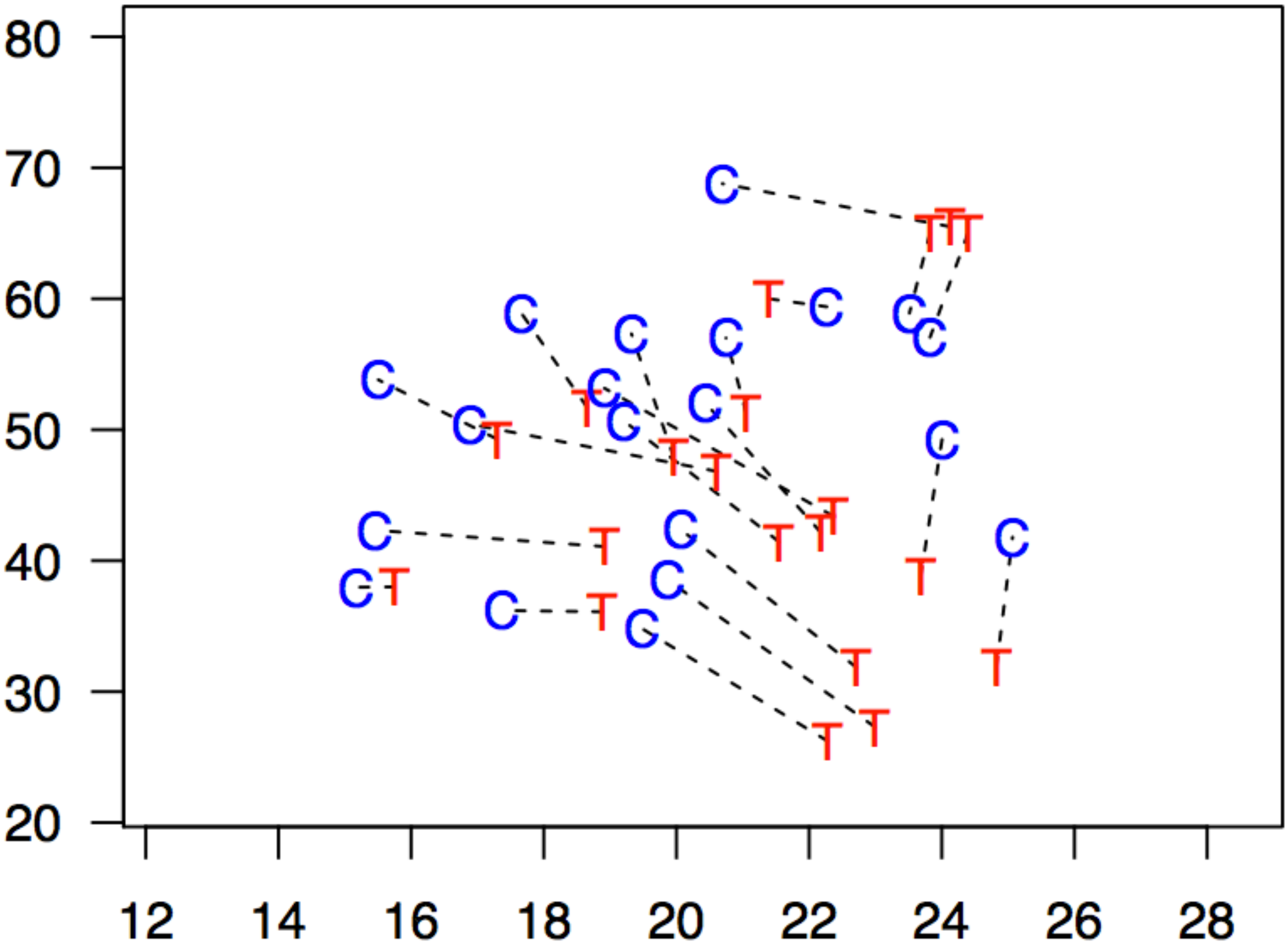
Matching



Matching



Matching



Matching

- For each treated unit i with covariates X_i, you would like to estimate $\alpha_i = Y_{1i} - Y_{0i}$.
- For treated units you observe Y_{1i}, but where to get Y_{0i}?
- Matching: borrow it from control unit with (nearly) the same X_i
- So estimator is:

$$\widehat{\alpha}_{ATT} = \frac{1}{N_1} \sum_{D_i=1} \left(Y_i - Y_{j(i)}\right)$$

where $Y_{j(i)}$ is the outcome of an untreated observation such that $X_{j(i)}$ is the closest value to X_i among the untreated observations.

We can also use the average of M closest matches:

$$\widehat{\alpha}_{ATT} = \frac{1}{N_1} \sum_{D_i=1} \left\{ Y_i - \left(\frac{1}{M} \sum_{m=1}^{M} Y_{j_m(i)}, \right) \right\}$$

Matching

unit	Potential Outcome under Treatment	Potential Outcome under Control		
i	Y_{1i}	Y_{0i}	D_i	X_i
1	6	?	1	3
2	1	?	1	1
3	0	?	1	10
4		0	0	2
5		9	0	3
6		1	0	-2
7		1	0	-4

What is $\widehat{\alpha}_{ATT} = \frac{1}{N_1} \sum_{D_i=1} \left(Y_i - Y_{j(i)} \right)$?

Matching

unit	Potential Outcome under Treatment	Potential Outcome under Control		
i	Y_{1i}	Y_{0i}	D_i	X_i
1	6	9	1	3
2	1	0	1	1
3	0	9	1	10
4		0	0	2
5		9	0	3
6		1	0	-2
7		1	0	-4

What is $\widehat{\alpha}_{ATT} = \frac{1}{N_1} \sum_{D_i=1} \left(Y_i - Y_{j(i)}\right)$?

$\widehat{\alpha}_{ATT} = 1/3 \cdot (6 - 9) + 1/3 \cdot (1 - 0) + 1/3 \cdot (0 - 9) = -3.7$

Shortcoming: Curse of Dimensionality

The volume increases exponentially when adding extra dimensions.

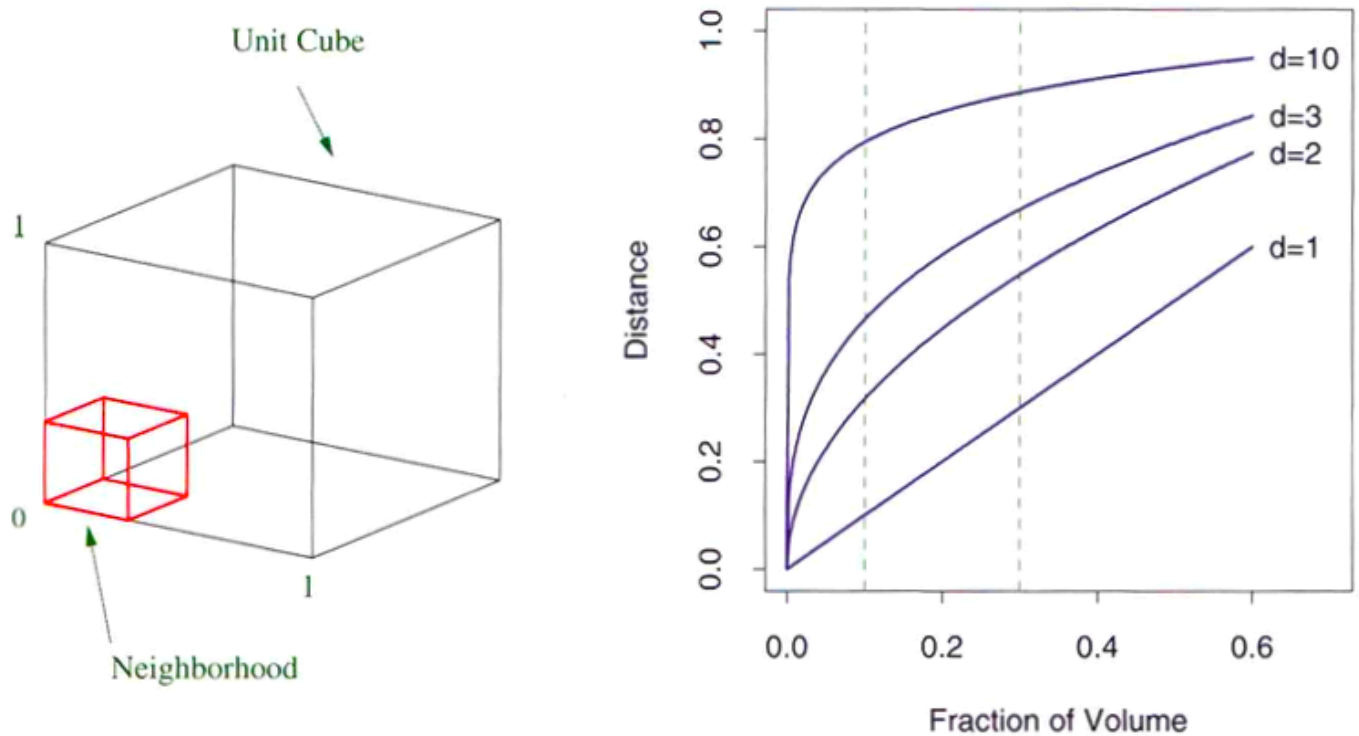


FIGURE 2.6. *The curse of dimensionality is well illustrated by a subcubical neighborhood for uniform data in a unit cube. The figure on the right shows the side-length of the subcube needed to capture a fraction r of the volume of the data, for different dimensions p. In ten dimensions we need to cover 80% of the range of each coordinate to capture 10% of the data.*

Check Balance

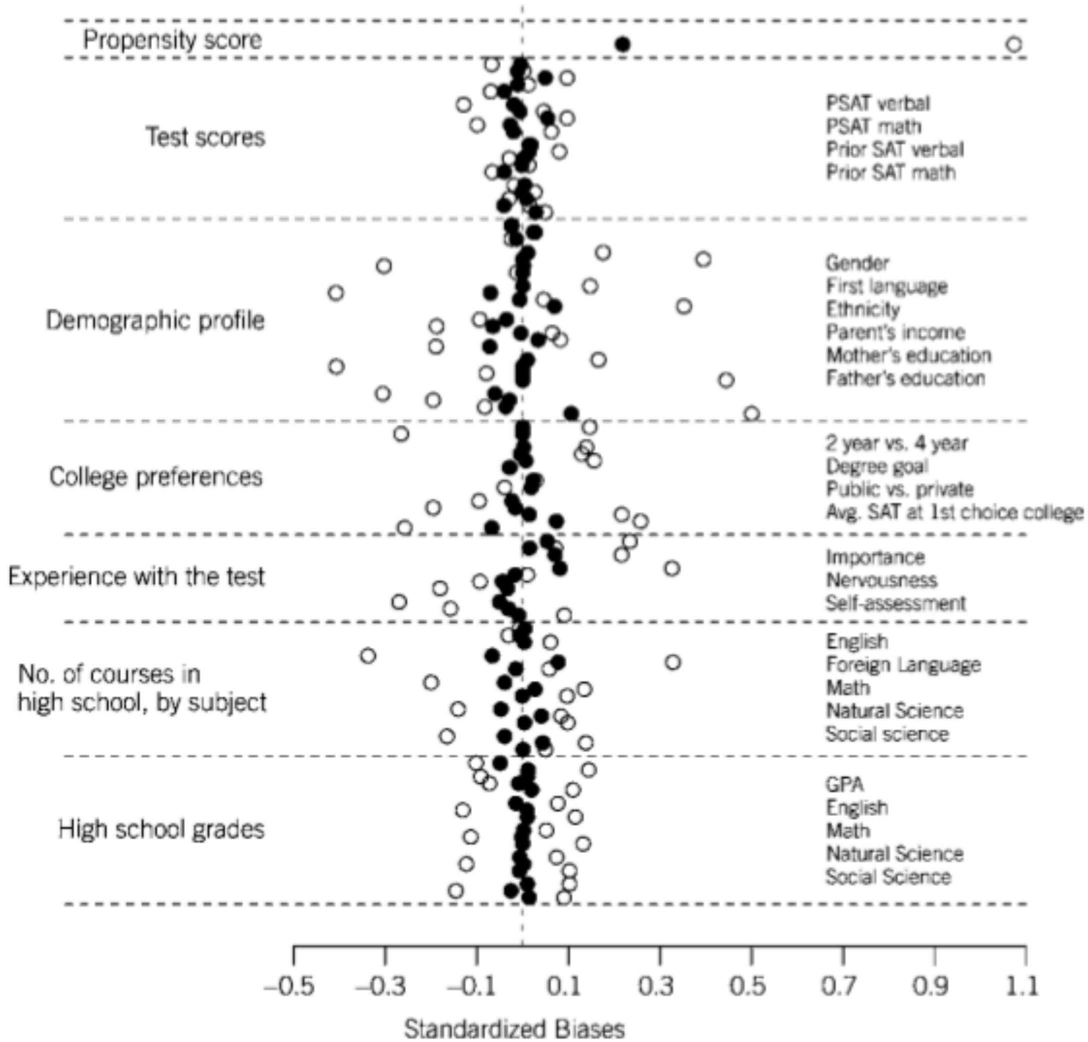


Figure 3. Standardized Biases Without Stratification or Matching, Open Circles, and Under the Optimal [.5, 2] Full Match, Shaded Circles.

Check Balance

TABLE 2. Balance Summary Statistics and Tests: Russian and Chechen Sweeps

Pretreatment Covariates	Mean Treated	Mean Control	Mean Difference	Std. Bias	Rank Sum Test	K-S Test
Demographics						
Population	8.657	8.606	0.049	0.033	0.708	0.454
Tariqa	0.076	0.048	0.028	0.104	0.331	—
Poverty	1.917	1.931	−0.016	−0.024	0.792	1.000
Spatial						
Elevation	5.078	5.233	−0.155	−0.135	0.140	0.228
Isolation	1.007	1.070	−0.063	−0.096	0.343	0.851
Groznyy	0.131	0.138	−0.007	−0.018	0.864	—
War Dynamics						
TAC	0.241	0.282	−0.041	−0.095	0.424	—
Garrison	0.379	0.414	−0.035	−0.072	0.549	—
Rebel	0.510	0.441	0.070	0.139	0.240	—
Selection						
Presweep violence	3.083	3.117	−0.034	0.009	0.454	0.292
Large-scale theft	0.034	0.055	−0.021	−0.115	0.395	—
Killing	0.117	0.090	0.027	0.084	0.443	—
Violence Inflicted						
Total abuse	0.970	0.833	0.137	0.124	0.131	0.454
Prior sweeps	1.729	1.812	−0.090	−0.089	0.394	0.367
Other						
Month	7.428	6.986	0.442	0.130	0.260	0.292
Year	2004.159	2004.110	0.049	0.043	0.889	1.000

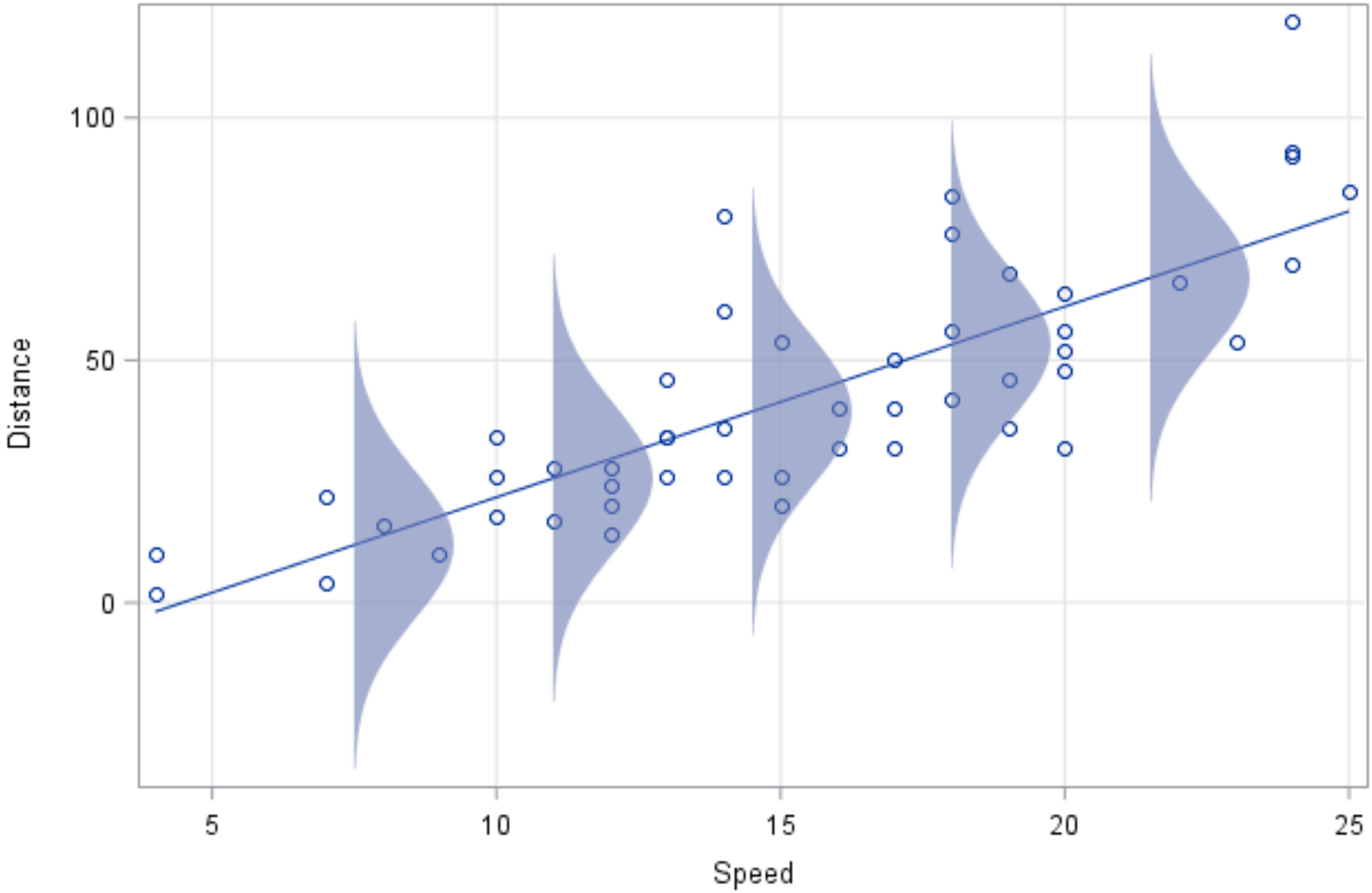
Note: 145 matched pairs. Matching with replacement.

Methods of Conditioning

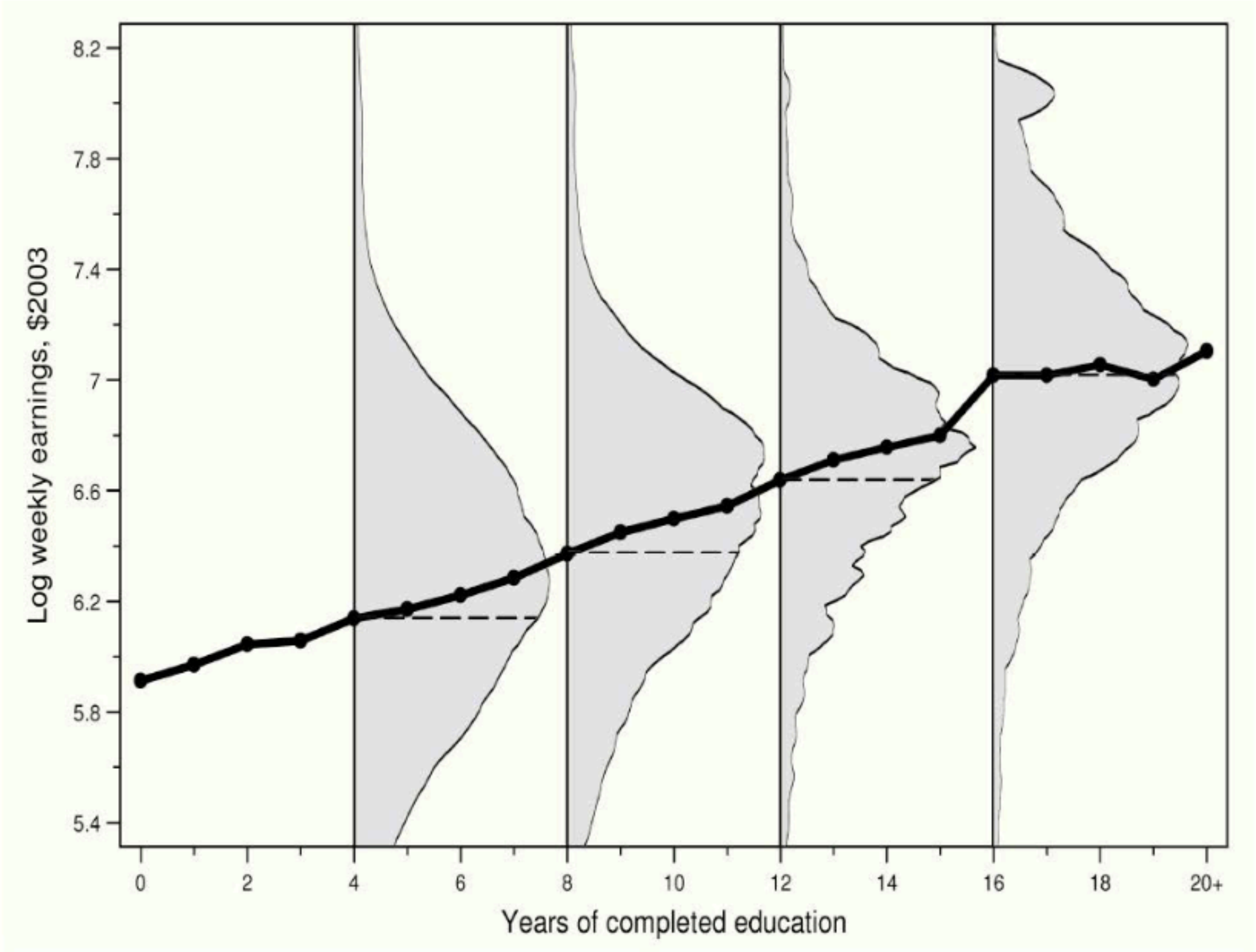
(1) Regression

Modeling the Conditional Expectation Function

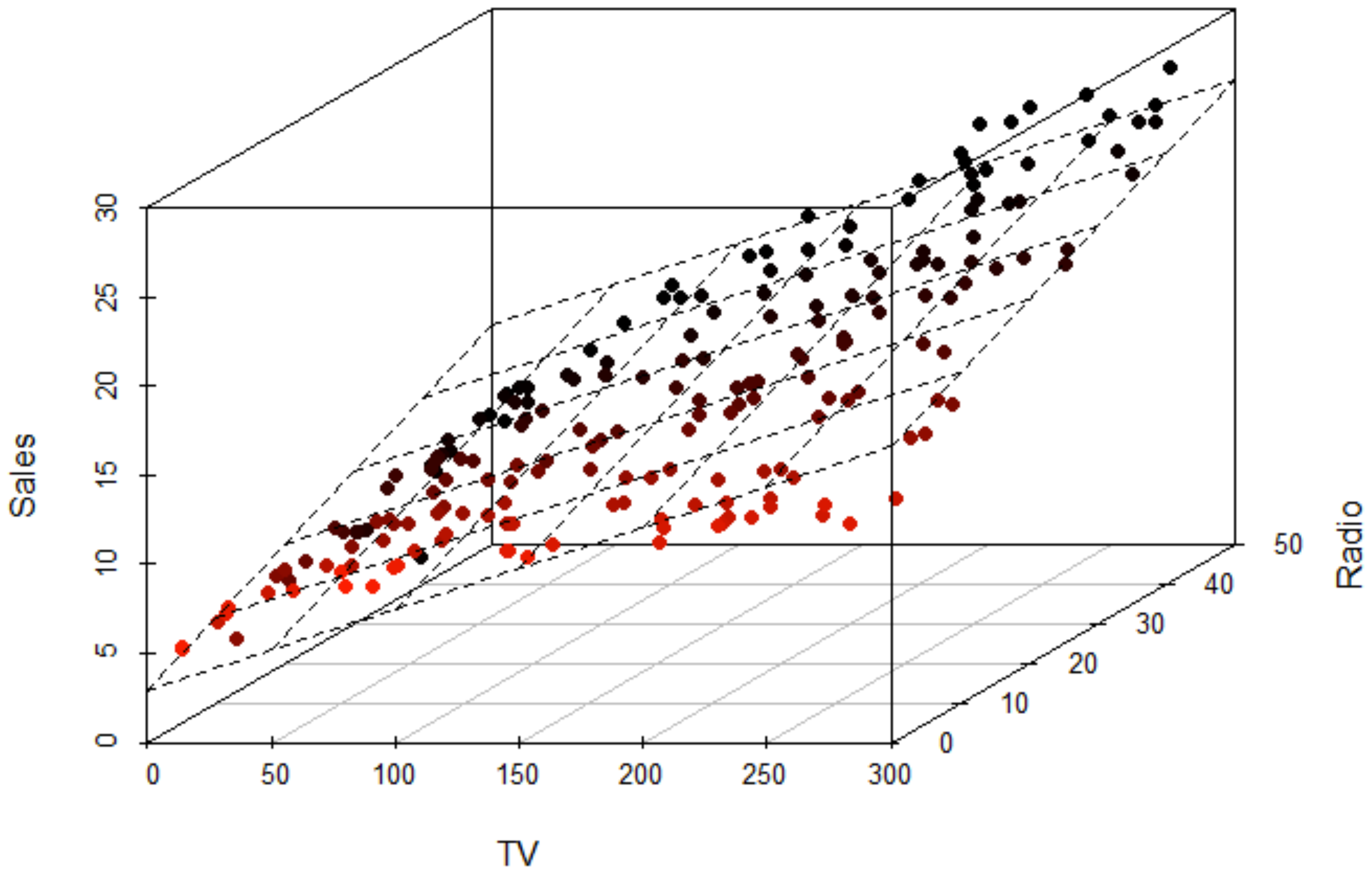
Regression is Fitting a Linear CEF



Regression is Fitting a Linear CEF



Multivariate Case



Regression as an Estimator of Casual Effects

$$\boxed{\mathbb{E}[Y_i \mid D_i, X_i] \;=\; \beta_0 + \beta_1 D_i + \gamma^\top X_i}$$

When is $\hat{\beta}_{OLS}$ a good estimator of α_{ATE}?

We slipped in two assumptions:

(1) Constant treatment effect. We assumed

- $\alpha(X_i) = \mathbb{E}[Y_{1i} - Y_{0i} | X_i]$
- ...which implies $\alpha_i = \alpha$ for all i

(2) Linearity: Between our model and the CI assumption, we asserted

$$\mathbb{E}[Y_{id} | X_i] = \beta_0 + \beta_1 d_i + \gamma^\top X_i \quad \text{for} \quad d = 0, 1$$

Equivalently,

$$Y_{id} = \beta_0 + \beta_1 d_i + \gamma^\top X_i + \varepsilon \quad \text{for} \quad d = 0, 1$$

Constant Treatment Effect w/ Linear Potential Outcomes

Result: If treatment effect is constant across units and potential outcomes are linear in X_i, then the OLS estimate of β_1 in the following regression model

$$Y_i = \beta_0 + \beta_1 D_i + \gamma^\top X_i + \varepsilon_i$$

is an unbiased and consistent estimator of α_{ATE}.

Proof (just how we got here):

$$\begin{aligned}
\mathbb{E}[\beta_1] &= \mathbb{E}[Y_i | D_i = 1, X_i] - \mathbb{E}[Y_i | D_i = 0, X_i] \quad \text{(correct specification)} \\
&= \mathbb{E}[Y_{1i} | X_i] - \mathbb{E}[Y_{0i} | X_i] \quad \text{(SOO)} \\
&= \alpha(X) \\
\mathbb{E}[\beta_1] &= \alpha \quad \text{(constant effect assumption)}
\end{aligned}$$

Note that if CI and linearity hold, ε cannot be related to D: traditional CIA assumption

Example: The Effect of Schooling on Wages

- Wages on schooling (S), controlling for ability (A)

$$Y_i = \alpha + \rho S_i + A_i'\gamma + \epsilon_i$$

- Ability is hard to measure. What if we leave it out?

- Omitted variable bias = The effect of the omitted ×
  The correlation between the omitted (A) and the included (S)

$$\frac{Cov(Y_i, S_i)}{V(S_i)} = \rho + \gamma'\delta_{AS}$$

Moreover

When the two assumptions are not satisfied

- If potential outcomes are not linear in covariates X, regression provides the best linear approximation to the population regression function.

- If the treatment effect is not constant, regression provides an unbiased estimator for conditional-variance-weighted average treatment effect, but not ATE or ATT.

$$\hat{\alpha}_{ATE} = \sum_{x \in \mathcal{X}} \{\mathbb{E}[Y_i \mid D_i = 1, X_i = x] - \mathbb{E}[Y_i \mid D_i = 0, X_i = x]\} \Pr(X_i = x)$$

$$\hat{\alpha}_{ATT} = \sum_{x \in \mathcal{X}} \{\mathbb{E}[Y_i \mid D_i = 1, X_i = x] - \mathbb{E}[Y_i \mid D_i = 0, X_i = x]\} \Pr(X_i = x \mid D_i = 1)$$

$$\hat{\beta}_{OLS} = \sum_{x \in \mathcal{X}} \{\mathbb{E}[Y_i \mid D_i = 1, X_i = x] - \mathbb{E}[Y_i \mid D_i = 0, X_i = x]\} \frac{\mathrm{Var}(D_i \mid X_i = x) \Pr(X_i = x)}{\sum_{x'} \mathrm{Var}(D_i \mid X_i = x') \Pr(X_i = x')}$$

Summary

- Matching and regression are main methods to estimate average causal effects when one can assume conditional ignorability
- These are estimation strategies; **the validity of the identification strategy (SOO) remains a first-order concern**
- Always ask yourself: what is the experiment your SOO estimation strategy is approximating?
- Welcome to the **Messy Real World!**