Causal Inference: Weighting Methods

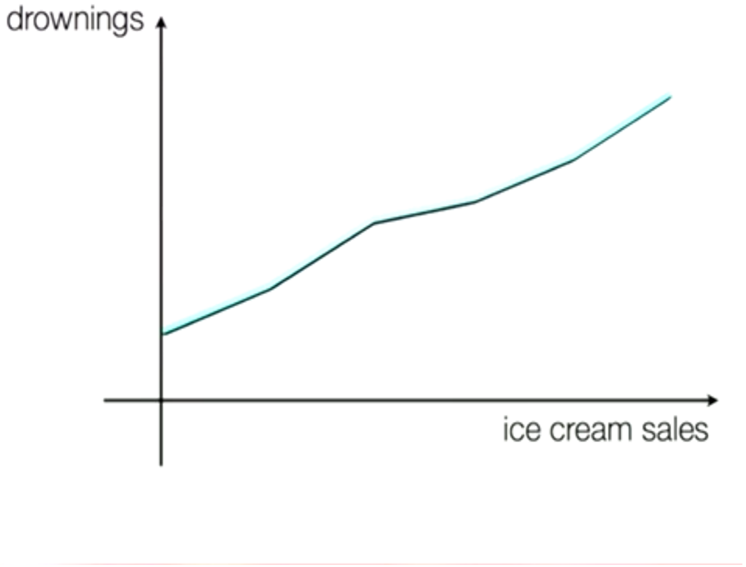
Lizzie Irlbacher and Guan Wang

Outline

1. What is the weighting method used for?
2. The comparison between different methods
3. Propensity score
4. Inverse-propensity weighting
5. Things to keep in mind
6. Application

1. What is the weighting method used for?

- Examples:



Does vocabulary training improve students’ vocabulary achievement score?

From Shadish, William R., Margaret H. Clark, and Peter M. Steiner. "Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments." *Journal of the American statistical association* 103.484 (2008): 1334-1344.

1. What is the weighting method used for?

Vocabulary training example: The school offers two training classes: math training and vocabulary training. Students chose which training they wanted on their own. After training, all participants were assessed on both mathematics and vocabulary outcomes.

Question: Does the vocabulary training help to improve students' vocabulary posttest score?

From Shadish, William R., Margaret H. Clark, and Peter M. Steiner. "Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments." *Journal of the American statistical association* 103.484 (2008): 1334-1344.

1. What is the weighting method used for?

- Is this a randomized or nonrandomized experiment?
- How would students make their decision? What does the literature say?
- Can we just regress Y on T (without any adjustments)?

	Y	T
1	15	1
2	20	1
3	14	0
4	12	0

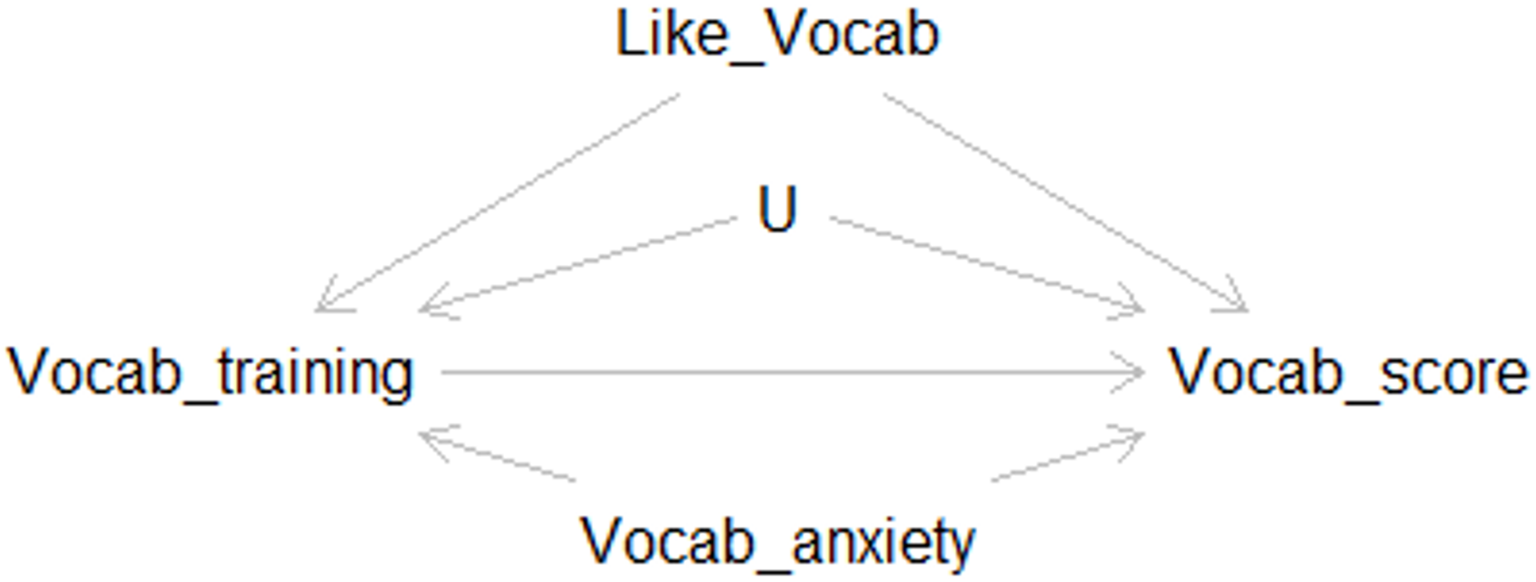
1. What is the weighting method used for?

- Confounder and confounding bias
    - Confounder: the common cause of the independent and dependent variable
    - Confounding bias arises from failure to condition on a common cause.

- How should we remove confounding bias?

1. What is the weighting method used for?

- Graphical models (Directed Acyclic Graphs)



1. What is the weighting method used for?

- Solutions

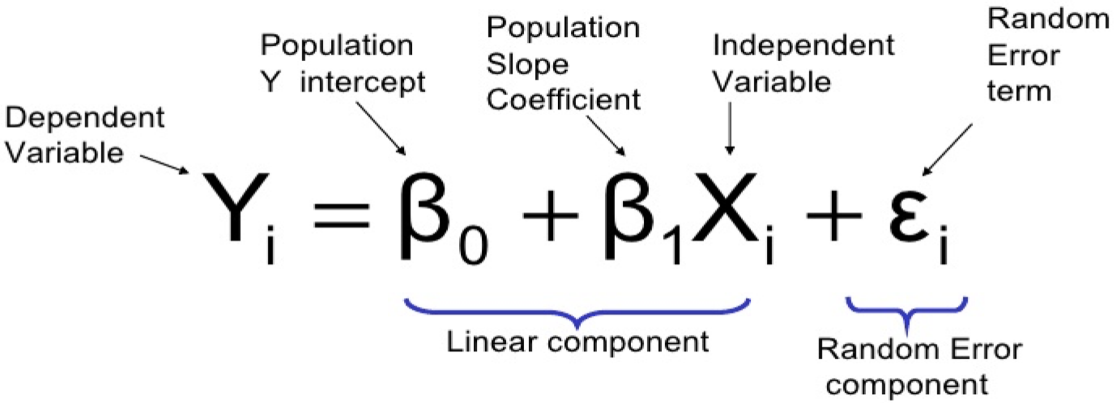
“All else equal”

Observational studies→randomized experiments

If we know that all the students who like vocabulary chose to attend the vocabulary training (treatment group). And then we observe that students in the treatment group have a higher posttest score than students in the control group. Should we attribute the difference to the training or how much the student likes vocabulary?

2. The Comparison of Different Methods

- Regressions
  - The problems of running regressions
  - The more assumptions we have, the less credible the results are.



2. The Comparison of Different Methods

- Propensity score
  - 1) Weighting
    - Advantages: More flexible
    - Ease of implementation (no need to install new R packages)
  - 2) Matching
  - 3) Stratification
  - 4) Regressions

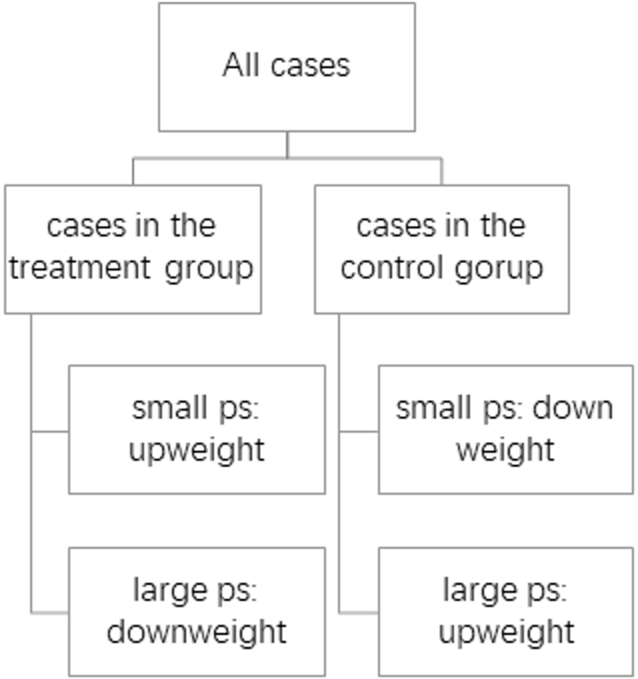
3. Propensity score

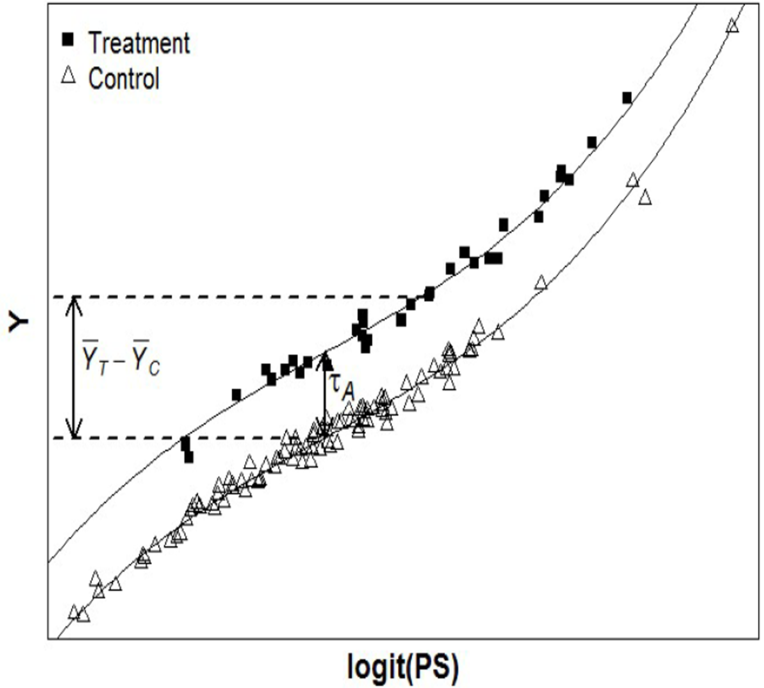
- Definition: the conditional probability that a subject belongs to the treatment group given the observed covariates.

- Upweight the under-represented cases and downweight the over-represented cases.

	Y	R	like_ vcab	voca b_an xiety
1		1		
2		0		
3		1		
4		0		

3. Propensity score





Source: Peter Steiner, EDMS647 Causal Inference & Evaluation Methods, Spring 2021, “Non-equivalent Control Group Designs: Matching & Propensity Scores”

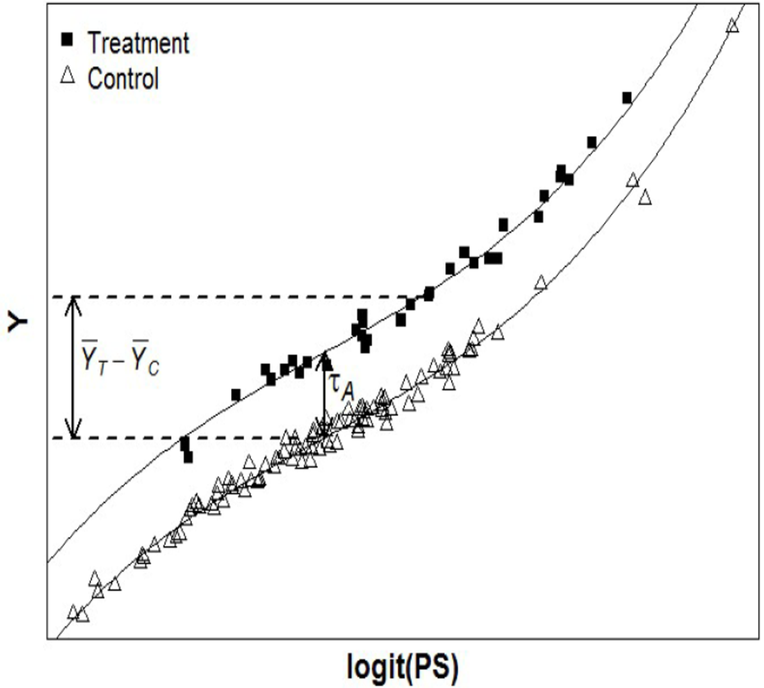
3. Propensity Score

- Estimation
  - Key questions to ask yourself:
    - What variables are potential confounders?
    - Can those confounders be observed?
    - This is a theory-driven process. The “kitchen-sink” approach is not necessarily a bad approach, but it should not be prioritized.

	Y	R	like_vocab	vocab_ anxiety
1		1		
2		0		
3		1		
4		0		

4. Weighting

- AKA Inverse Probability of Treatment Weighting (IPTW)
  - Weights for average treatment effect (ATE):
  - Wi=1/e(x) for the treated (Z=1)
  - Wi=1/(1-e(x)) for the untreated (z=0)



Source: Peter Steiner, EDMS647 Causal Inference & Evaluation Methods, Spring 2021, “Non-equivalent Control Group Designs: Matching & Propensity Scores”

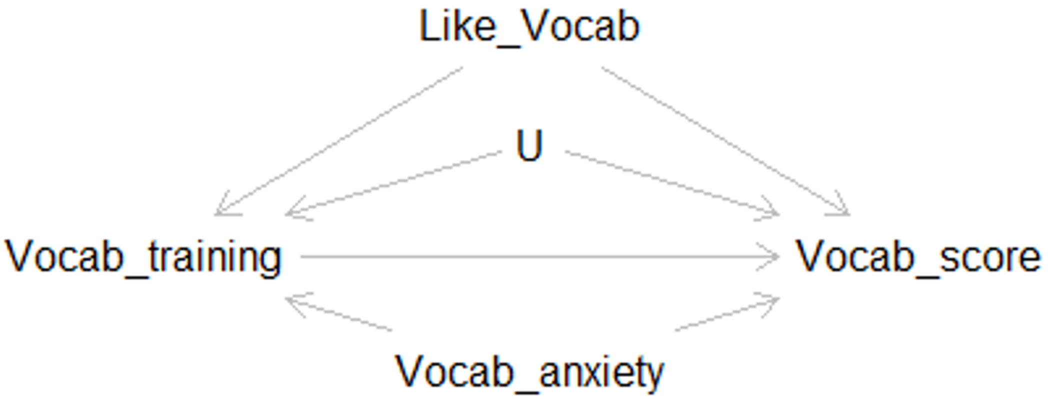
4. Weighting

- Treatment effect:
  - Weighted least squares regression without covariates;
  - With additional covariates: a second chance to remove confounding bias.

5. Things to keep in mind

- Does the weighting method help to remove all confounding bias?

No. The first thing to consider is to find out all the confounders (based on theory). Next, consider if these confounders can be reliably measured. Third, choose a method to estimate the causal effect. Think about the data-generating process, not how many variables we can easily collect.



6. Application

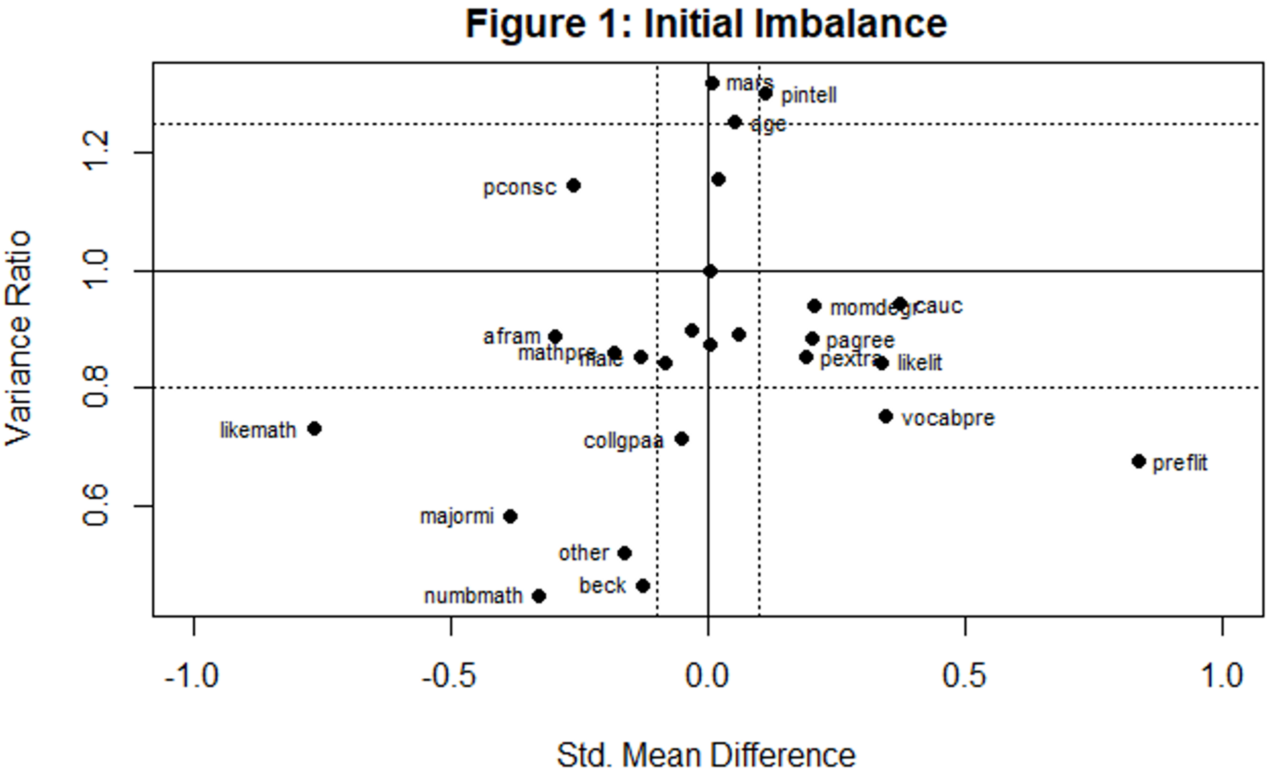
Example being used from:

Shadish, Clark & Steiner. (2008). “Can Nonrandomized Experiments Yield Accurate Answers? A Randomized Experiment Comparing Random and Nonrandom Assignments” *Journal of American Statistical Association 103(484): 1334-1344.* https://doi.org/10.1198/016214508000000733

Steps to carry out Inverse Weighting

1. Choice of treatment effect
2. Identification: Assessing strong ignorability/ adjustment criterion
3. Estimation of the propensity score & balance checks
4. Estimation of treatment effects
5. Sensitivity analysis

Assessing Imbalance



3. Propensity Score

3.2 Estimation

Regress R on X1 and X2, using logistic models (most frequently used), or:

Probit regression;

Linear probability model,

Machine learning techniques, such as decision trees (e.g., random forests);

...

3. Propensity Score

3.2 Estimation

$logit(Z)=\alpha+\beta1*X1+\beta2*X2+\varepsilon$

Or

$logit(Z)=\alpha+\beta1*(X1)^2+\beta2*X2+\varepsilon$

Or

$logit(Z)=\alpha+\beta1*X1+\beta2*X2+\beta3*X1*X2+\varepsilon$

...

Estimation of Propensity scores- code

````
```{r}
#propensty score examination
# use covariates of adjustment set (adj.set) only!
mdl1 <- as.formula(vm ~ mathpre + vocabpre + actcomp + hsgpaar + collgpaa +
                        numbmath + likemath + likelit + preflit + majormi +
                        mars + cauc + afram + male)

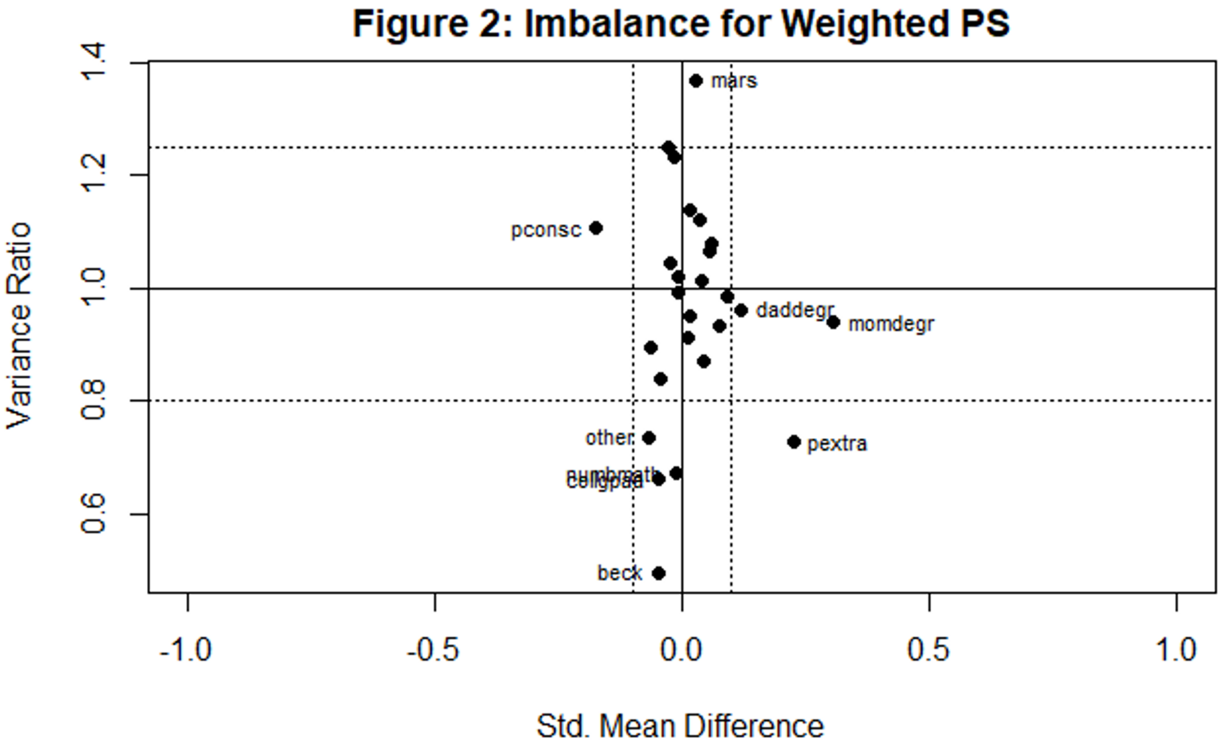
out1 <- glm(mdl1, data = scs, family = 'binomial')
summary(out1)

# ::::: get PS and PS-logit of initial model :::::
scs$ps <- out1$fitted                          # fitted values are the PS
scs$lps <- log(scs$ps / (1 - scs$ps))          # PS-logit = log(PS/(1-PS))
```
````

````
Inverse propensity weighting
```{r}
scs$z <- ifelse(scs$vm == 'vocabulary', 1, 0)     # dummy variable for treatment indicator
scs$iptw <- with(scs, z / ps + (1 - z) / (1 - ps)) # computation of IPTWs

# ::::: compute weights for overlapping cases only :::::
scs$iptwo <- scs$iptw
scs$iptwo[del.ind] <- 0
```
````

Balance Check



Estimation of Treatment Effects-code

```
Basic models
```{r}
scs$vocaball <- scs.out$vocaball

# ::::: prima facie effect, i.e., without any PS or covariance adjustment (biased estimate) ::::
testmodel <- lm(vocaball ~ vm, data = scs)
summary(testmodel)

# ::::: ATE with PS adjustment: inverse-propensity weighting :::::
inversebase <- lm(vocaball ~ vm, data = scs, weights = iptw)  # with all cases
summary(lm(vocaball ~ vm, data = scs, weights = iptwo))  # with overlapping cases
```


Double robust
```{r}
# ::::: ATE with additional covariance adjustment (doubly robust or mixed method) :::::
# create formula for outcome model (use same covariates as in PS model)
inversePSmdl <- as.formula(vocaball ~ vm + mathpre + vocabpre + actcomp + hsgpaar + collgpaa + numbmath + likemath + likelit + preflit + majormi + mars + cauc + afram + male)
inversedouble <- lm(inversePSmdl, data = scs, weights = iptw)
```
```

Table 1: Vocabulary Treatment Effect on Post-Test Scores: Double Robust

	Model 1	Model 2	Model 3
Vocabulary Treatment	9.001***	8.259***	8.266***
	(0.509)	(0.494)	(0.402)
Math Pre-Test			0.143
			(0.104)
Vocabulary Pre-Test			0.240***
			(0.059)
ACT			0.187***
			(0.068)
High School GPA			−0.599
			(0.455)
College GPA			−0.371
			(0.341)
Number Math Classes			0.333
			(0.230)
Like Math			0.148
			(0.144)
Like Literature			−0.103
			(0.123)
Prefered Literature			0.965**
			(0.416)
Math Major			0.067
			(0.555)
Math Anxiety Score			0.017
			(0.011)
White			0.143
			(1.037)
Black			−0.317
			(1.052)
Male			0.515

Sensitivity Analysis

```
ORDINARY NONPARAMETRIC BOOTSTRAP


Call:
boot(data = scs, statistic = PSboot, R = 1000, stype = "i")


Bootstrap Statistics :
    original      bias    std. error
t1* 8.259489 0.02083612   0.6333705
t2* 8.266146 0.01163381   0.5175514
         [,1]     [,2]
[1,] 6.986363 9.575998
[2,] 7.263972 9.316237
```

Conclusion

- Take causal inference class in someway
- EDMS 647 with Peter Steiner is being offered again in Spring semester