

## STA305/1004 - Class 9

February 8, 2016

- Design of observational studies
- propensity score
- Balancing
- ignorable treatment assign. and propensity scores.

## The design phase of an observational study

Good observational studies are designed.

*An observational study should be conceptualized as a broken randomized experiment . . . in an observational study we view the observed data as having arisen from a hypothetical complex randomized experiment with a lost rule for the [assignment mechanism], whose values we will try to reconstruct. Rubin (2007)*

## The design phase of an observational study

*Of critical importance, in randomized experiments the design phase takes place prior to seeing any outcome data. And this critical feature of randomized experiments can be duplicated in observational studies, for example, using propensity score methods, and we should objectively approximate, or attempt to replicate, a randomized experiment when designing an observational study. Rubin (2007)*



## Observational versus Randomized Studies

- ▶ In observational studies the researcher does not randomly allocate the treatments.
- ▶ Randomization ensures subjects receiving different treatments are comparable.

## Observational Studies

- ▶ We have an outcome measured on two groups of subjects (treated and control).
- ▶ We want to make a fair comparison between the treated group and the control group in terms of the outcome.
- ▶ We can obtain covariates that describe the subjects before they received treatments, but we can't ensure that the groups will be comparable in terms of the covariates.

## Importance of Randomization

- pre treatment ←
- ▶ Randomization tends to produce relatively comparable or “balanced” treatment groups in large experiments.
  - ▶ The covariates aren’t used in assigning treatments in an experiment.
  - ▶ There’s no deliberate balancing of the covariates – it’s just a nice feature of randomization.
  - ▶ We have some reason to hope and expect that other (unmeasured) variables will be balanced, as well.

+add page smoking cessation & weight gain

# MAIMONIDES' RULE



- Angrist and Lavy (1999) published an unusual study of the effects of class size on academic achievement.
- Causal effects of class size on pupil achievement is difficult to measure.
- The twelfth century Rabbinic scholar Maimonides interpreted the the Talmud's discussion of class size as:

*“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.”*



# MAIMONIDES' RULE



- Since 1969 it has been used to determine the division of enrolment cohorts into classes in Israeli public schools.
- Class size usually determined by affluence or poverty of a community, enthusiasm or skepticism about the value of education, special needs of students, etc.
- If adherence to Maimonides's rule were perfectly rigid then with a single class size of 40 what would separate the same school with two classes whose average size is 20.5 is one student.

# MAIMONIDES' RULE



- Maimonides' rule has the largest impact on a school with about 40 students in a grade cohort.
- With cohorts of size 40, 80, and 120 students, the steps down in average class size required by Maimonides' rule when an additional student enrolls are, respectively, from 40 to 20.5, from 40 to 27, and from 40 to 30.25.

Original Cohort Size	40	80	120
Class Size with 1 additional student	$41/2$ 20.5	$81/3$ 27	$121/4$ 30.25

# MAIMONIDES' RULE

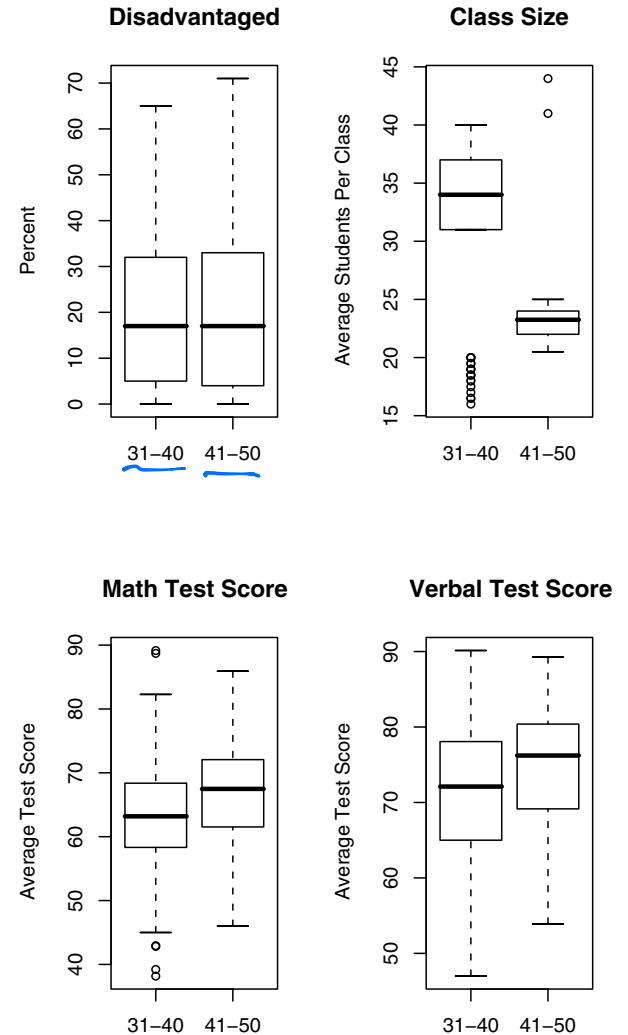


- Angrist and Lavy looked at schools with fifth grade cohorts in 1991 with between 31 and 50 students, where average class sizes might be cut in half by Maimonides' rule.
- There were 211 such schools, with 86 of these schools having between 31 and 40 students in fifth grade, and 125 schools having between 41 and 50 students in the fifth grade.

# MAIMONIDES' RULE

- Among the 211 schools with between 31 and 50 students in fifth grade, the percentage disadvantaged has a slightly negative Kendall's correlation of  $-0.10$ .
- Strong negative correlations of  $-0.42$  and  $-0.55$ , respectively, with performance on verbal and mathematics test scores.
- For this reason, 86 matched pairs of two schools were formed, matching to minimize the total absolute difference in percentage disadvantaged.

# MAIMONIDES' RULE



**Fig. 1.1** Eighty-six pairs of two Israeli schools, one with between 31 and 40 students in the fifth grade, the other with between 41 and 50 students in the fifth grade, matched for percentage of students in the school classified as disadvantaged. The figure shows that the percentage of disadvantaged students is balanced, that imperfect adherence to Maimonides' rule has yielded substantially different average class sizes, and test scores were higher in the group of schools with predominantly smaller class sizes.

# MAIMONIDES' RULE

86 pairs of Israeli schools:

*Treatment is  
class size*

1. One with between 31 and 40 students in the fifth grade ( $T = 0$ ),
  2. One with between 41 and 50 students in the fifth grade ( $T = 1$ ),
- the pairs being matched for a covariate  $x$  = percentage of disadvantaged students.

# MAIMONIDES' RULE

- Strict adherence to Maimonides' rule would require the slightly larger fifth grade cohorts to be taught in two small classes, and the slightly smaller fifth grade cohorts to be taught in one large class.
- Adherence to Maimonides' rule was imperfect but strict enough to produce a wide separation in typical class size.

# MAIMONIDES' RULE

- In their study,  $(Y(1), Y(0))$  are average test scores in fifth grade if the fifth grade cohort was a little larger ( $T = 1$ ) or a little smaller ( $T = 0$ ).
- What separates a school with a slightly smaller fifth grade cohort ( $T = 0$ , 31–40 students) and a school with a slightly larger fifth grade cohort ( $T = 1$ , 41–50 students)?
- A handful of grade 5 students.



# MAIMONIDES' RULE

- Seems plausible that whether or not a few more students enrol in the fifth grade is a haphazard event.
- An event not strongly related to the average test performance ( $Y(1)$ ,  $Y(0)$ ) that the fifth grade would exhibit with a larger or smaller cohort.
- Building a study in the way that Angrist and Lavy did, it seems plausible that the propensity of a larger/smaller class are fairly close in the 86 matched pairs of two schools.

# MAIMONIDES' RULE

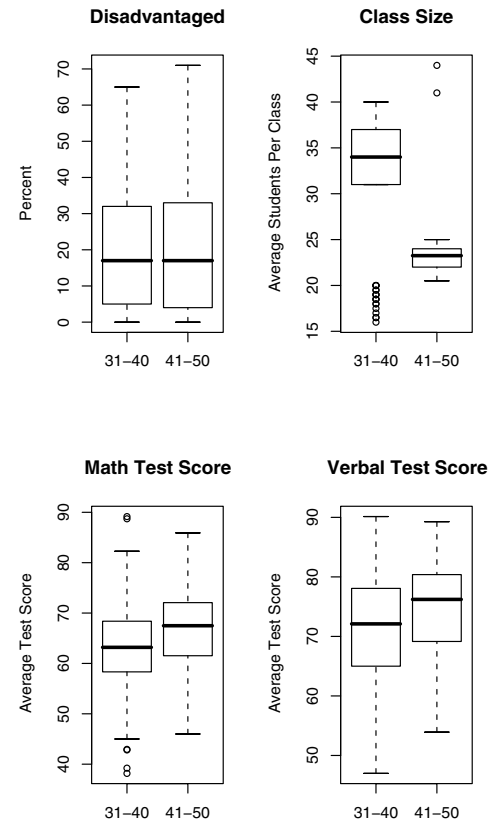
- A 'natural experiment' is an attempt to find in the world some rare circumstance such that a consequential treatment was handed to some people and denied to others for no particularly good reason at all, that is, haphazardly.

# MAIMONIDES' RULE

- It does seem reasonable to think that probability of a larger class are fairly close for the 86 paired Israeli schools.
- This might be implausible in some other context, say in a national survey in Canada where schools are funded by local governments, so that class size might be predicted by the wealth or poverty of the local community.

# MAIMONIDES' RULE

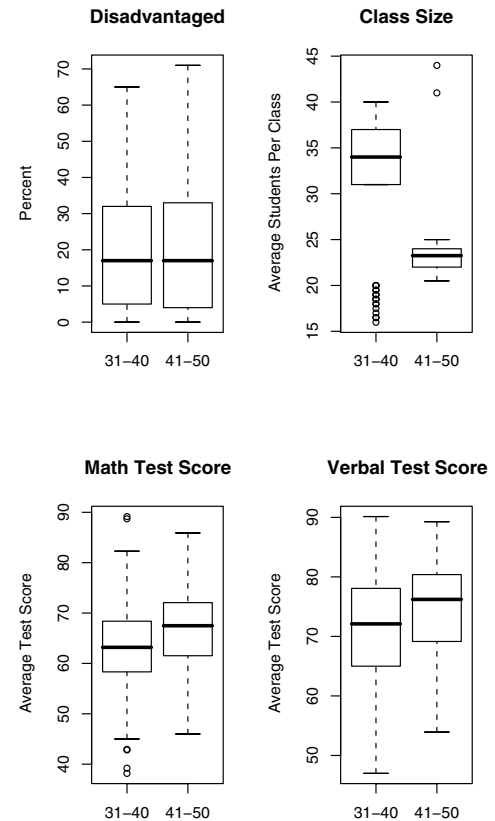
Even in the Israeli schools, the claim that  $P(T)$  for two schools are plausibly close depended upon: (i) matching for the percentage of students who were disadvantaged,  $x_k$ , and (ii) defining  $T_k$  in terms of cohort size rather than class size.



**Fig. 1.1** Eighty-six pairs of two Israeli schools, one with between 31 and 40 students in the fifth grade, the other with between 41 and 50 students in the fifth grade, matched for percentage of students in the school classified as disadvantaged. The figure shows that the percentage of disadvantaged students is balanced, that imperfect adherence to Maimonides' rule has yielded substantially different average class sizes, and test scores were higher in the group of schools with predominantly smaller class sizes.

# MAIMONIDES' RULE

Adherence to Maimonides' rule, except for a few deviations: two large fifth grade cohorts that strict adherence to the rule would have divided, and several small cohorts that were taught in smaller classes than strict adherence to the rule would have permitted.

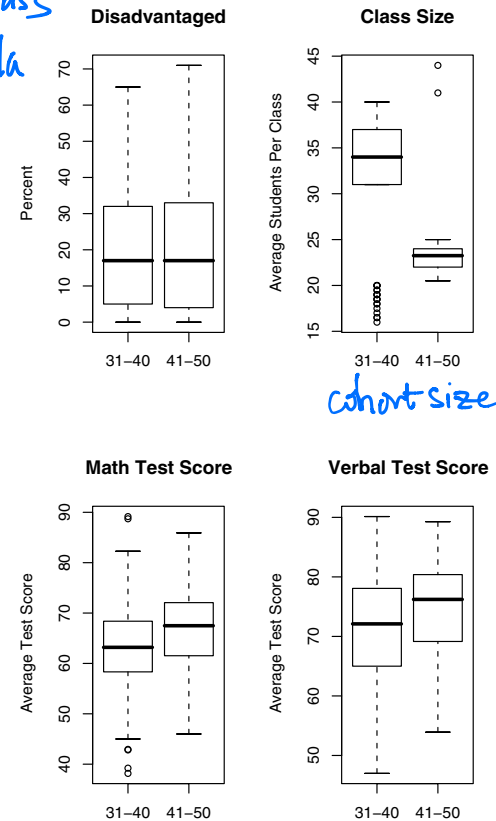


**Fig. 1.1** Eighty-six pairs of two Israeli schools, one with between 31 and 40 students in the fifth grade, the other with between 41 and 50 students in the fifth grade, matched for percentage of students in the school classified as disadvantaged. The figure shows that the percentage of disadvantaged students is balanced, that imperfect adherence to Maimonides' rule has yielded substantially different average class sizes, and test scores were higher in the group of schools with predominantly smaller class sizes.

# MAIMONIDES' RULE

The figure shows that  $T_k$  reflects the haphazard element, namely susceptibility to Maimonides' rule based on cohort size, rather than realized class size. Defining  $T_k$  in this way is analogous to an 'intention-to-treat' analysis in a randomized experiment.

=  $\begin{cases} 1 & \text{small class} \\ 0 & \text{large class} \end{cases}$




**Fig. 1.1** Eighty-six pairs of two Israeli schools, one with between 31 and 40 students in the fifth grade, the other with between 41 and 50 students in the fifth grade, matched for percentage of students in the school classified as disadvantaged. The figure shows that the percentage of disadvantaged students is balanced, that imperfect adherence to Maimonides' rule has yielded substantially different average class sizes, and test scores were higher in the group of schools with predominantly smaller class sizes.

cohort = # of children in grade 5

## Smoking cessation and weight gain

What is the effect of smoking on weight gain?

- ▶ Data was used from The National Health and Nutrition Examination Survey Data I Epidemiological Follow-up Study (NHESFS) survey to assess this question.
  - ▶ The NHESFS has information on sex, race, weight, height, education, alcohol use, and intensity of smoking at baseline (1971-75) and follow-up (1982) visits.
  - ▶ The survey was designed to investigate the relationships between clinical, nutritional, and behavioural factors.
  - ▶ A cohort of persons 25-74 who completed a medical exam in 1971-75 followed by a series of follow-up studies.
  - ▶ Would it be possible to conduct a randomized study to evaluate the treatment effect?
- 

## Smoking cessation and weight gain

	Cessation (A=1)	No cessation (A=0)
age, years <i>mean</i>	46.2	42.8
men, %	54.6	46.6
white, %	91.1	85.4
university, %	15.4	9.9
weight, kg	72.4	70.3
Cigarettes/day	18.6	21.2
year smoking	26.0	24.1
little/no exercise, %	40.7	37.9
inactive daily life, %	11.2	8.9

*Are the covariates balanced?*



## The propensity score

- ▶ Covariates are pre-treatment variables and take the same value for each unit no matter which treatment is applied.
- ▶ For example, pre-treatment blood pressure or pre-test reading level are not influenced by a treatment that would alter blood pressure or reading level.

The propensity score is

$$e(\mathbf{x}) = P(T = 1 | \mathbf{x}),$$

where  $\mathbf{x}$  are observed covariates.

The  $i^{th}$  propensity score is the probability that a unit receives treatment given all the information, recorded as covariates, that is observed before the treatment.

## The propensity score

In experiments the propensity scores are known. In observational studies they can be estimated using models such as logistic regression where the outcome is the treatment indicator and the predictors are all the confounding covariates.

## The propensity score

Consider a completely randomized design with  $n = 2$  units and one unit is assigned treatment. The treatment assignment for the  $i^{th}$  subject is:

$T_1$	$T_2$	$P(T_1)$	$P(T_2)$
0	0	0	0
0	1	0	0.5
1	0	0.5	0
1	1	0	0

Each unit's propensity score is ?.

$$P(T_i=1 | \mathbf{x})$$

$$1/2$$

## The propensity score

Q Q Q Q Q Q Q Q  
C T T T C C C C

Consider a completely randomized design with  $n = 8$  units and three units are assigned treatment. - Each unit's propensity score is ?.

- The probability of an particular treatment assignment is ?.

$$P(T_i=1|x_i)$$
$$\frac{3}{8}$$

$$\frac{1}{\binom{8}{3}} = \frac{1}{56}$$

## The propensity score

	1st Trt	Assig	2nd Trt	Assig	3rd Trt	Assig
[1,]		1		1		1
[2,]		2		2		2
[3,]		3		4		5

Each column corresponds to the units that will be treated; so the units that will not be treated are not included in the column. For example in the first treatment assignment units 1, 2, 3 will be treated and units 4, 5, 6, 7, 8 will be given control.

4, 5, 6, 7, 8  
Control

3, 5, 6, 7, 8  
Control

...

## The propensity score

- ▶ Consider a completely randomized design with  $n$  units and  $m$  units are assigned treatment.
- ▶ Each unit has probability  $\frac{m}{n}$  of receiving treatment (and  $1 - \frac{m}{n}$  of receiving control).
- ▶ Thus, each person's propensity score is  $m/n$ .
- ▶ The probability of a particular treatment assignment is  $\frac{1}{\binom{n}{m}}$ .

$\therefore$  we know the propensity score we can use tests such as randomization test

$$H_0: Y(0) = Y(1) = Y^{obs}$$

- can only use randomization test if we know propensity score  
hence treatment ...

## The propensity score

- ▶ Consider a study that plans to use a doctor's medical records to compare two treatments ( $T = 0$  and  $T = 1$ ) given for a certain condition.
- ▶ Treatments were not assigned to patients randomly, but were based on various measured and unmeasured patient factors.

prop score is not known

## Logistic Regression

- ▶ The logistic regression model with one covariate  $x$  is:

$$\log(P(T_i = 1)/P(T_i = 0)) = \beta_0 + \beta_1 x_i$$

- ▶ The logistic regression model with  $k$  covariates  $x_1, x_2, \dots, x_k$  is

$$\log(P(T_i = 1)/P(T_i = 0)) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

$$\log \left( \frac{P(T_i=1)}{1 - P(T_i=1)} \right)$$



## Parameter Estimates from Logistic Regression

- ▶ In a logistic model with one binary covariate the parameter estimate of  $\beta_1$  is:

$$\frac{(P(T = 1|x = 1)/P(T = 0|x = 1))}{(P(T = 1|x = 0)/P(T = 0|x = 0))} = \exp(\beta_1)$$

odds ratio

- ▶  $\exp(\beta_1)$  is the odds ratio comparing those with  $x = 1$  to those with  $x = 0$ .

$$\log\left(\frac{p_i}{1 - p_i}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3},$$

where  $p_i = P(T_i = 1)$ .

## Predicted probabilities from Logistic Regression

- In a logistic model with one binary covariate the predicted probabilities can be calculated using the fitted model:

$$\hat{p}_i = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1})}$$

## The propensity score

- ▶ The patient factors that were measured are age ( $x_1$ ), sex ( $x_2$ ), and health status before treatment ( $x_3$ ).
- ▶ The propensity score can be estimated for each patient by fitting a logistic regression model with treatment as the dependent variable and  $x_1, x_2, x_3$  as the predictor variables.

## The propensity score

- The predicted probabilities from the above equation are estimates of the propensity score for each patient.

$$\hat{p}_i = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3})}$$

## The propensity score in Smoking Cessation Study

The propensity score for each subject in smoking and weight gain study can be estimated by fitting a logistic regression model.

```
prop.model <- glm(qsmk ~ as.factor(sex) + as.factor(race) +  
                  age + as.factor(education.code) + smokeintensity +  
                  smokeyrs + as.factor(exercise) + as.factor(active) +  
                  wt71, family = binomial(), data = nhefshwdat)  
  
#Summary of propensity score model  
summary(prop.model)
```

## The propensity score in Smoking Cessation Study

	Estimate	Std. Error	z value
(Intercept)	-2.401228039	0.484016356	-4.9610473
as.factor(sex)1	-0.499080121	0.146530691	-3.4059767
as.factor(race)1	-0.778222994	0.207031619	-3.7589572
age	0.046207220	0.009889326	4.6724338
as.factor(education.code)2	-0.065716379	0.196122828	-0.3350777
as.factor(education.code)3	0.052634524	0.175523000	0.2998725
as.factor(education.code)4	0.108653058	0.269190883	0.4036283
as.factor(education.code)5	0.466164550	0.224105901	2.0801083
smokeintensity	-0.026527450	0.005664293	-4.6832762
smokeyrs	-0.028491730	0.010008629	-2.8467165
as.factor(exercise)1	0.359556747	0.178603430	2.0131570
as.factor(exercise)2	0.422771538	0.185656969	2.2771649
as.factor(active)1	0.044927909	0.131555137	0.3415139
as.factor(active)2	0.158150602	0.213435405	0.7409764
wt71	0.006099273	0.004368231	1.3962800

Pr(>|z|)

(Intercept)	7.011411e-07
as.factor(sex)1	6.592780e-04
as.factor(race)1	1.706230e-04
age	2.976515e-06
as.factor(education.code)2	7.375665e-01
as.factor(education.code)3	7.642744e-01

## How do we build a propensity score model?

- ▶ Usual tool is logistic regression model for the treatment allocation decision – We therefore want to consider including any variables that have a relationship to the treatment decision (i.e. precede it in time, and are relevant) – No information is included on the actual treatment received, or on the outcome(s).

## Ten commandments of Propensity Model Development

1. Thou shalt value parsimony.
2. Thou shalt examine thy predictors for collinearity.
3. Thou shalt test all thy predictors for statistical significance.
4. Thou shalt have ten times as many predictors as subjects.
5. Thou shalt examine thy regression coefficients
6. Thou shalt perform bootstrap analyses to assess shrinkage.
7. Thou shalt perform regression diagnostics and examine residuals with care.
8. Thou shalt hold out a sample of thy data for cross-validation.
9. Thou shalt perform external validation on a new sample of data.
10. Thou shalt ignore commandments 1 through 9 and instead ensure that the model adequately balances covariates.



## Propensity model development

1. Diagnostics for the successful prediction of probabilities and parameter estimates underlying those probabilities
2. Diagnostics for the successful design of observational studies based on estimated propensity scores.
3. is important 1. is not important in PS development.

## Propensity model development

- ▶ All covariates that subject matter experts (and subjects) judge important when selecting treatments.
- ▶ All covariates that relate to treatment and outcome, including any covariate that improves prediction (of exposure group).
- ▶ As much “signal” as possible.