Problem Set 7

Program Evaluation and Causal Inference

Christian Birchler Fenqi Guo Mingrui Zhang Wenjie Tu Zunhan Zhang

Spring Semester 2021

Names are listed in alphabetical order

Matching

1. Regression as a matching estimator: theory

1(a)

$$
\begin{aligned}
\text{ATT} &= \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1] \\
&= \mathbb{E}[Y_{1i} | D_i = 1] - \mathbb{E}[Y_{0i} | D_i = 1] \\
&= \mathbb{E}(\mathbb{E}[Y_{1i} | X_i, D_i = 1] | D_i = 1) - \mathbb{E}(\mathbb{E}[Y_{0i} | X_i, D_i = 1] | D_i = 1) \quad \text{Law of Iterated Expectations} \\
&= \mathbb{E}(\mathbb{E}[Y_{1i} | X_i, D_i = 1] | D_i = 1) - \mathbb{E}(\mathbb{E}[Y_{0i} | X_i, D_i = 0] | D_i = 1) \quad \text{CIA} \\
&= \mathbb{E}(\underbrace{\mathbb{E}[Y_{1i} | X_i, D_i = 1] - \mathbb{E}[Y_{0i} | X_i, D_i = 1]}_{\equiv \delta_x} | D_i = 1) \\
&= \mathbb{E}_x[\delta_x, X_i = x | D_i = 1] \\
&= \sum_x \delta_x \cdot Pr(X = x | D = 1) \\
&= \sum_x \delta_x \cdot \frac{Pr(D = 1 | X = x) Pr(X = x)}{Pr(D = 1)} \quad \text{Bayes Rule} \\
&= \sum_x \delta_x \cdot \frac{Pr(D = 1 | X = x) Pr(X = x)}{\sum_x Pr(D = 1 | X = x) Pr(X = x)} \quad \text{Law of Total Probability}
\end{aligned}
$$

Note:

- Generalized *Law of Iterated Expectations*: $\mathbb{E}[\mathbb{E}(Y|X,Z)|X] = \mathbb{E}(Y|X)$.
- Under *Conditional Independence Assumption*: $\mathbb{E}(Y_{0i}|X_i, D_i = 1) = \mathbb{E}(Y_{0i}|X_i, D_i = 0) = \mathbb{E}(Y_{0i}|X_i)$

Matching estimator for ATT:

$$
\delta_M = \frac{\sum_x \delta_x Pr(D = 1 | X = x) Pr(X = x)}{\sum_x Pr(D = 1 | X = x) Pr(X = x)}
$$

1(b)

Note that $\tilde{D}_i$ is the residual from a regression of D on X, we therefore have:

- $\mathbb{E}[\tilde{D}_i] = 0$.
- $\mathbb{E}[\tilde{D}_i|X_i] = 0$.

$$\begin{aligned}
Cov(Y_i, \tilde{D}_i) &= \mathbb{E}[\tilde{D}_i Y_i] - \mathbb{E}[\tilde{D}_i]\mathbb{E}[Y_i] \\
&= \mathbb{E}[\tilde{D}_i Y_i] \\
&= \mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])Y_i] \\
&= \mathbb{E}\left[(D_i - \mathbb{E}[D_i|X_i])\mathbb{E}[Y_i|D_i, X_i]\right]
\end{aligned}$$

$$\begin{aligned}
\mathbb{V}(\tilde{D}_i) &= \mathbb{E}[\tilde{D}_i^2] - (\mathbb{E}[\tilde{D}_i])^2 \\
&= \mathbb{E}[\tilde{D}_i^2] \\
&= \mathbb{E}\left[(D_i - \mathbb{E}[D_i|X_i])^2\right]
\end{aligned}$$

According to the hints, the regression estimator is given by:

$$\begin{aligned}
\delta_R &= \frac{Cov(Y_i, \tilde{D}_i)}{\mathbb{V}(\tilde{D}_i)} \\
&= \frac{\mathbb{E}\left[(D_i - \mathbb{E}[D_i|X_i])\mathbb{E}[Y_i|D_i, X_i]\right]}{\mathbb{E}\left[(D_i - \mathbb{E}[D_i|X_i])^2\right]}
\end{aligned}$$

1(c)

According to the hints:

$$\mathbb{E}[Y_i|D_i, X_i] = \mathbb{E}[Y_i|D_i = 0, X_i] + \delta_x D_i$$

Plug it into the regression estimator:

$$\begin{aligned}
& \mathbb{E}\left[(D_i - \mathbb{E}[D_i|X_i])\mathbb{E}[Y_i|D_i, X_i]\right] \\
= \quad & \mathbb{E}\left[(D_i - \mathbb{E}[D_i|X_i])(\mathbb{E}[Y_i|D_i = 0, X_i] + \delta_x D_i)\right] \\
= \quad & \mathbb{E}\{(D_i - \mathbb{E}[D_i|X_i])\mathbb{E}[Y_i|D_i = 0, X_i]\} + \mathbb{E}\{(D_i - \mathbb{E}[D_i|X_i])\delta_x D_i)\} \\
& \quad \textit{Conditional Independence Assumption} \\
= \quad & \mathbb{E}\{(D_i - \mathbb{E}[D_i|X_i])\mathbb{E}[Y_i|X_i]\} + \mathbb{E}\{(D_i - \mathbb{E}[D_i|X_i])\delta_x D_i)\} \\
& \quad \because \tilde{D}_i \perp\!\!\!\perp \mathbb{E}[D_i|X_i] \\
= \quad & \underbrace{\mathbb{E}\{\tilde{D}_i \mathbb{E}[Y_i|X_i]\}}_{0} + \mathbb{E}\{(D_i - \mathbb{E}[D_i|X_i])\delta_x D_i)\} \\
= \quad & \mathbb{E}\{(D_i - \mathbb{E}[D_i|X_i])\delta_x D_i)\} \\
& \quad \because \delta_x \perp\!\!\!\perp \mathbb{E}[D_i|X_i] \\
= \quad & \mathbb{E}\{(D_i - \mathbb{E}[D_i|X_i])^2 \delta_x)\}
\end{aligned}$$

1(d)

Plug $\mathbb{E}\left[(D_i - \mathbb{E}[D_i|X_i])\delta_x\right]$ into regression estimator:

$$
\begin{aligned}
\delta_R &= \frac{\mathbb{E}\left[(D_i - \mathbb{E}[D_i|X_i])\mathbb{E}[Y_i|D_i, X_i]\right]}{\mathbb{E}\left[(D_i - \mathbb{E}[D_i|X_i])^2\right]} \\
&= \frac{\mathbb{E}\{(D_i - \mathbb{E}[D_i|X_i])^2\delta_x)\}}{\mathbb{E}\left[(D_i - \mathbb{E}[D_i|X_i])^2\right]} \\
&\text{use the Law of Iterated Expectations to obtain:} \\
&= \frac{\mathbb{E}\{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2|X_i]\delta_x\}}{\mathbb{E}\{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2|X_i]\}} \\
&= \frac{\mathbb{E}[\sigma_D^2(x)\delta_x]}{\mathbb{E}[\sigma_D^2(x)]}
\end{aligned}
$$

1(e)

D_i is a dummy variable[1] and its variance can be obtained using the formula:

$$
\sigma_D^2(x) = Pr(D_i = 1|X_i) \cdot (1 - Pr(D_i = 1|X_i))
$$

$$
\begin{aligned}
\delta_R &= \frac{\mathbb{E}[\sigma_D^2(x)\delta_x]}{\mathbb{E}[\sigma_D^2(x)]} \\
&= \frac{\mathbb{E}[Pr(D_i = 1|X_i)(1 - Pr(D_i = 1|X_i))\delta_x]}{\mathbb{E}[Pr(D_i = 1|X_i)(1 - Pr(D_i = 1|X_i))]} \\
&= \frac{\mathbb{E}_x[Pr(D_i = 1|X_i = x)(1 - Pr(D_i = 1|X_i = x))\delta_x]}{\mathbb{E}_x[Pr(D_i = 1|X_i = x)(1 - Pr(D_i = 1|X_i = x))]} \\
&\text{rewrite in summation form:} \\
&= \frac{\sum_x \delta_x[Pr(D_i = 1|X_i = x)(1 - Pr(D_i = 1|X_i = x))]Pr(X_i = x)}{\sum_x [Pr(D_i = 1|X_i = x)(1 - Pr(D_i = 1|X_i = x))]Pr(X_i = x)}
\end{aligned}
$$

2. Regression as a matching estimator: application

```r
# import library
library(readxl) # read xls file
library(stargazer) # regression output
library(knitr) # table output
library(dplyr)
library(splitstackshape) # expand rows
library(vcd) # mosaic plot
```

```r
# read data
UCBAdmissions <- read_xls("berkeley_data.xls")
```

[1] Recall **Bernoulli Distribution**, the variance is $p(1-p)$.

2(a)

Gender as a treatment variable takes two values (1 for male and 0 for female) and people cannot change their treatment status[2].

2(b)

Full independence assumption implies $(Y_1, Y_0) \perp\!\!\!\perp D$

```r
UCBAdmissions2 <- expandRows(UCBAdmissions, "Total")
UCBAdmissions2$Applicants <- factor(UCBAdmissions2$Applicants,
                                    levels=c("Women", "Men"))

## Accept=1 if admitted, Accept=0 if rejected
UCBAdmissions2$Accept <- c(rep(1,512), rep(0,313), rep(1,353), rep(0,207),
                           rep(1,120), rep(0,205), rep(1,138), rep(0,279),
                           rep(1,53), rep(0,138), rep(1,22), rep(0,351),
                           rep(1,89), rep(0,19), rep(1,17), rep(0,8),
                           rep(1,202), rep(0,391), rep(1,131), rep(0,244),
                           rep(1,94), rep(0,299), rep(1,24), rep(0,317))

model1 <- lm(Accept ~ Applicants, data = UCBAdmissions2)

stargazer(model1, header = F, dep.var.labels.include = F,
          dep.var.caption = "Accept (yes/no)",
          covariate.labels = "Male", single.row = T,
          title = "Regression in 2(b)", keep.stat = c("n", "rsq"))
```

Table 1: Regression in 2(b)

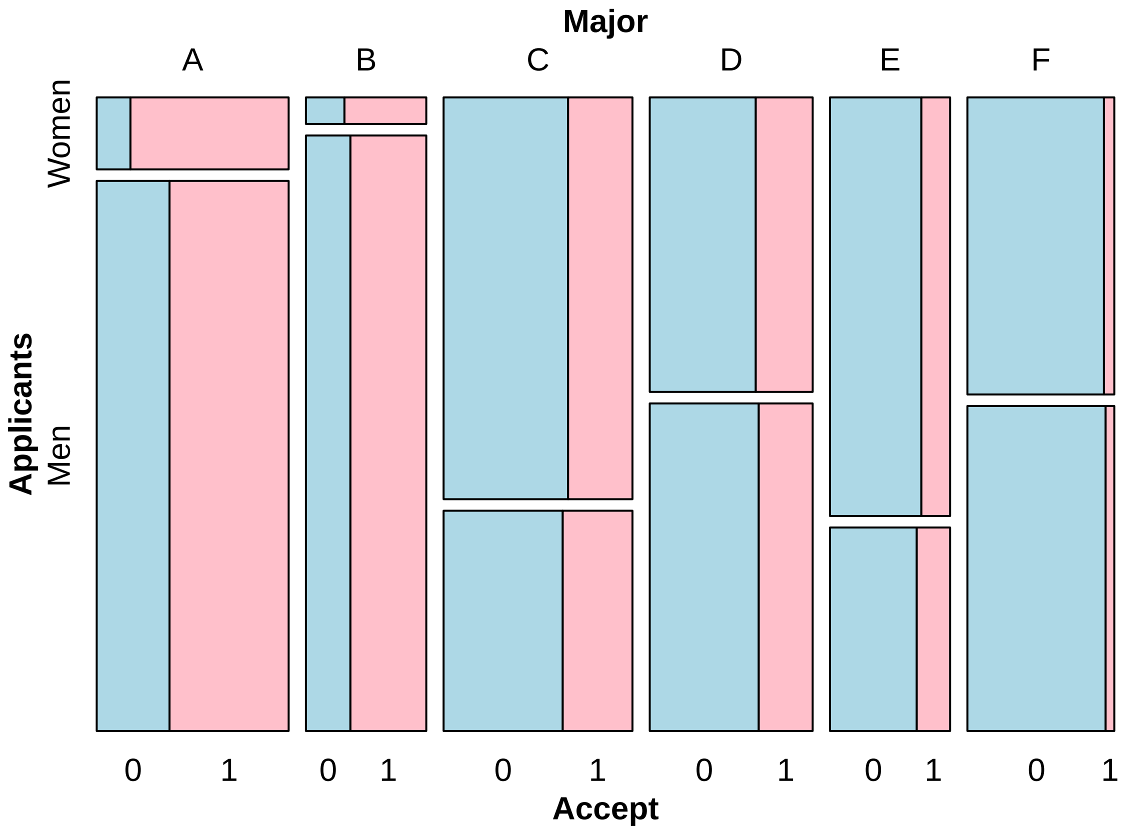
	Accept (yes/no)
Male	0.142*** (0.015)
Constant	0.304*** (0.011)
Observations	4,526
R^2	0.020
Note:	*p<0.1; **p<0.05; ***p<0.01

2(c)

Full independence between treatment and potential outcome is a very strong assumption and it is less likely to hold in general. As implied by Simpson's paradoxon, full independence assumption can lead us to a very wrong conclusion. In our case, it seems that UC Berkeley admits more men than women at the first glance but the conclusion is not convincing if we think it through. Applicants may have different preferences for majors and admission criteria may also differ from department to department. We also can see this from the *Mosaic Plot*.

[2] Sex reassignment surgery is not considered in this setting.

```r
mosaic( ~ Major + Applicants + Accept, data = UCBAdmissions2,
        highlighting = "Accept", highlighting_fill = c("lightblue", "pink"),
        direction = c("v","h","v"))
```



2(d)

Conditional independence assumption implies $(Y_1, Y_0) \perp\!\!\!\perp D|X$,

$$\begin{cases} \text{ATE} = \mathbb{E}(Y_1 - Y_0|X) \\ \text{ATT} = \mathbb{E}(Y_1 - Y_0|D = 1, X) \end{cases}$$

```r
# summarize data by major
## calculate difference in admission rates between men and women
## calculate weights in total population
sub1 <- UCBAdmissions %>% group_by(Major) %>%
  summarise(diffs=-diff(Pct_Admitted),
            weights.ATE=sum(Total)/sum(UCBAdmissions$Total))

# subset data with men
sub2 <- UCBAdmissions[UCBAdmissions$Applicants=="Men", ]

## calculate weights in male population
sub1$weights.ATT <- sub2$Total/sum(sub2$Total)
```

```r
ATE <- sum(sub1$diffs*sub1$weights.ATE)

ATT <- sum(sub1$diffs*sub1$weights.ATT)

cat(sprintf("ATE: %.3f\nATT: %.3f", ATE, ATT))
```

```
## ATE: -0.043
## ATT: -0.071
```

2(e)

- In (b), we assume that the treatment is unconditionally independent of the potential outcome. This is a very strong assumption and we ignore other confounders which affects the outcome. For example, admission criteria/procedure may differ a lot from major to major. We could not treat them as a whole and simply take the difference.
- In (d), we assume that the treatment conditional on major is independent of the potential outcome. This is a less strong assumption and we take majors into account. Therefore, the conclusion based on this assumption is more reliable.

To sum up, we should be very careful about these assumptions as they often can lead to different results and implications.

2(f)

```r
model2 <- lm(Accept ~ Applicants + factor(Major), UCBAdmissions2)

stargazer(model2, dep.var.caption = "Accept (yes/no)",
          dep.var.labels.include = F, title= "Regression in 2(f)",
          model.names = F, single.row = T, digits = 3, header = F,
          covariate.labels = c("Male", "Major B", "Major C",
                               "Major D", "Major E", "Major F"),
          keep.stat = c("n", "rsq"), table.placement = "h")
```

Table 2: Regression in 2(f)

	Accept (yes/no)
Male	-0.018 (0.015)
Major B	-0.010 (0.023)
Major C	-0.303^{***} (0.022)
Major D	-0.311^{***} (0.022)
Major E	-0.403^{***} (0.025)
Major F	-0.586^{***} (0.023)
Constant	0.660^{***} (0.020)
Observations	4,526
R^2	0.172
Note:	*p<0.1; **p<0.05; ***p<0.01

2(g)

The results between matching and OLS differ, mainly due to the point raised in c). There is a valid argument to be made about specific major decisions that differ between men and women. The effect in the matching, as well the OLS setting, share the same sign i.e that in fact not women but men were more likely to be rejected.