Problem Set 1

Sample Solutions

10/23/2018

Q1

We illustrate how randomization affects OVB with the following example from lecture. Consider the following estimation equation:

$$y_i = a + bT_i + cX_i + e_i$$

where y_i is the average test score in school i, T_i is a dummy variable indicating whether school i received treatment, X_i is a measure of average student ability in school i that is unobserved by the researcher, and e_i is white noise error (with mean zero, $E(e_i) = 0$).

We can compute a naive estimator by taking the difference in expected outcomes, conditional on treatment assignment:

$$\begin{aligned}
E[y_i|T=1] - E[y_i|T=0] =& E[a + bT_i + cX_i + e_i|T=1] - E[a + bT_i + cX_i + e_i|T=0] \\
=& E[a|T=1] + E[bT_i|T=1] + E[cX_i|T=1] + E[e_i|T=1] \\
& - (E[a|T=0] + E[bT_i|T=0] + E[cX_i|T=0] + E[e_i|T=0]) \\
=& a + b + cE[X_i|T=1] + 0 - a + 0 - cE[X_i|T=0] - 0 \\
=& b + c\{E[X_i|T=1] - E[X_i|T=0]\}
\end{aligned}$$

With randomization of treatment, we expect students in treated and non-treated schools not to differ in their average ability, so $E[X_i|T=1] = E[X_i|T=0] = E[X_i]$. Thus, the bias term in the equation above drops out and our estimator becomes:

$$E[y_i|T=1] - E[y_i|T=0] = b$$

Loading the data and presenting descriptive statistics

```r
worms <- read.csv("psets/Econ-172_PSet1-S17-data.csv")
worms2 <- read.csv("psets/Econ-172_PSet1-Spring2018-data.csv")

tab1 <- worms %>%
  select(-"schid") %>%
  descr(stats = c("mean", "sd", "min", "med", "max"),
        transpose = TRUE, omit.headings = TRUE)
```

This data set has 74 observations and 8 variables. Table 1 presents descriptive statistics for all the variables excluding the identifier (`schid`).

Table 1: Table 1: Descriptive Statistics

	Mean	Std.Dev	Min	Median	Max
group1	0.34	0.48	0.00	0.00	1.00
group2	0.34	0.48	0.00	0.00	1.00

	Mean	Std.Dev	Min	Median	Max
group3	0.32	0.47	0.00	0.00	1.00
part98	0.80	0.16	0.39	0.81	1.00
female	0.38	0.08	0.28	0.38	0.98
yob	1986.75	0.70	1985.35	1986.67	1988.82
sap	0.36	0.48	0.00	0.00	1.00

Assesing balance of covariates

Answer:

An R script to produce for the analysis for this and the following questions can be found at the end of the problem set.

To start, we want to estimate the following equation:

$$Y = a + bX + e$$

where Y is the characteristic of interest (`female, yob, sap`), and X is group1 or group2. In this equation, b describes the average difference between treatment and comparison schools, and a gives us the average value of the variable in comparison schools (the group left out, which is group1==0 in the first regression, or group2==0 in the second).

```
m1.1=lm(female ~ group1, data=worms)
m1.2=lm(yob ~ group1, data=worms)
m1.3=lm(sap ~ group1, data=worms)

m1.4=lm(female ~ group2, data=worms[worms$group1==0,])
m1.5=lm(yob ~ group2, data=worms[worms$group1==0,])
m1.6=lm(sap ~ group2, data=worms[worms$group1==0,])

m1.7=lm(female ~ group2, data=subset(worms,group1==0))
```

```
stargazer::stargazer(m1.1,m1.2,m1.3,m1.4,m1.5,m1.6, type = "latex",
          title="Table 2: Covariate Analysis", align = TRUE,
          omit.stat=c("LL","ser","f","adj.rsq", "rsq"), no.space=TRUE,
          header=FALSE, column.sep.width = "0pt")
```

Notice that this is equivalent to running a set of t-test comparisons of means of the three different outcomes between the two corresponding groups (group 1 v. groups 2+3 and group 2 v. group 3). For example the code below computes the difference between group 1 and groups 2+3 for the outcome variable `female`

```
t1 <- t.test(female ~ group1, data=worms, var.equal = TRUE)
est_t1 <- diff(t1$estimate)
st_t1 <- t1$statistic
```

This gives us a estimated difference of -0.0066 and a t-statistic of 0.3391, wich correspond to colum 1 and row 1 of table 2 (the t-statistic is the estimated difference divided by the standard error: -0.0066 / 0.0195 = -0.3391)

Table 2: Table 2: Covariate Analysis

	Dependent variable:					
	female	yob	sap	female	yob	sap
	(1)	(2)	(3)	(4)	(5)	(6)
group1	-0.007	0.005	-0.128			
	(0.020)	(0.173)	(0.119)			
group2				0.036	0.779^{***}	0.147
				(0.027)	(0.190)	(0.142)
Constant	0.386^{***}	$1,986.746^{***}$	0.408^{***}	0.368^{***}	$1,986.348^{***}$	0.333^{***}
	(0.011)	(0.100)	(0.069)	(0.019)	(0.136)	(0.101)
Observations	74	74	74	49	49	49

Note: *p<0.1; **p<0.05; ***p<0.01