

Problem Set 4

Tessie Dong, Derek Li, Andi Liu

Due Jan 26th, 2024

Part 1: Describe the Data (10 p)

1. Fill Table 1's columns 5 and 6 using, respectively, the data in `nswpsid.csv` and in `nswcps.csv`. Notes: You want to limit attention to observations with `treat=0`. You filled columns 3 and 4 in PSet 3.

```
# Load data
nswpsid <- read_csv("starter-files/nswpsid.csv")
nswcps <- read_csv("starter-files/nswcps.csv")
nswpsid_treat0 <- nswpsid %>% filter(treat == 0)
nswcps_treat0 <- nswcps %>% filter(treat == 0)

summary_cps <- nswcps_treat0 %>% dplyr::summarise_all(list(mean))
summary_psid <- nswpsid_treat0 %>% dplyr::summarise_all(list(mean))
```

Variable	Definition	NSW		PSID-1	CPS-1
		Treated	Control	Control	Control
[1]	[2]	[3]	[4]	[5]	[6]
age	Age in years	25.82	25.05	34.85	33.22
edu	Education in years	10.35	10.09	12.12	12.03
nodegree	1 if education < 12	0.71	0.83	0.31	0.30
black	1 if Black	0.84	0.83	0.25	0.07
hisp	1 if Hispanic	0.06	0.11	0.03	0.07
married	1 if married	0.19	0.15	0.87	0.71
u74	1 if unemployed in '74	0.71	0.75	0.09	0.12
u75	1 if unemployed in '75	0.60	0.68	0.10	0.11
re74	Real earnings in '74 (in '82 \$)	2,096	2,107	19429	14017
re75	Real earnings in '75 (in '82 \$)	1,532	1,267	19063	13631
re78	Real earnings in '78 (in '82 \$)	6,349	4,555	21554	14847
treat	1 if received offer of training	1	0	0	0
Sample Size		185	260	2,490	15,992

Table 1: Sample averages for the NSW data (treated and control groups), PSID-1 data, and CPI-1 data.

2. Briefly comment on the completed Table 1. Hint: Are the PSID-1 and CPS-1 samples “good” control groups?
3. Why do you think that Dehajia and Wahba constructed their “observational datasets” by pulling together the treated sample from NSW and a sample of individuals drawn from either the PSID or the CPS data? Hint: Both PSID and CPS include information on whether an individual enrolled in a training course during the previous 12 months. Thus, Dehajia and Wahba could have exploited exclusively observational variation in whether an individual enrolled in a training program. Why do you think that they chose not to follow this approach? \end{enumerate}