Assignment 3

Walter Verwer & Bas Machielsen

1/17/2021

Question 1

1.1

```
data <- tribble(
    ~ color, ~ no_treated, ~ no_contr, ~ avg_treated, ~avg_control,
    "purple", 100, 100, 9, 7,
    "blue", 75, 25, 13, 8,
    "green", 25, 75, 10, 9
)

data %>%
    group_by(color) %>%
    summarize(treatment_effect = avg_treated - avg_control)
```

```
## # A tibble: 3 x 2
##   color  treatment_effect
##   <chr>             <dbl>
## 1 blue                  5
## 2 green                 1
## 3 purple                2
```

1.2

The ATE is defined as $\mathbb{E}[\delta] = \mathbb{E}[Y_1^*] - \mathbb{E}[Y_0^*]$ which are the expectations of the potential outcomes. In general, these two variables are not observed. Under the *random assignment* assumption, we assume that $\mathbb{E}[Y_1^*] = \mathbb{E}[Y_1^*|D = 1]$ and $\mathbb{E}[Y_0^*] = \mathbb{E}[Y_0^*|D = 0]$, which can be estimated by their sample by their sample equivalents:

```
data %>%
    summarize(
        e_y1_d_is_1 = (9*100 + 13 * 75 + 10 * 25) / sum(no_treated),
        e_y0_d_is_0 = (7 * 100 + 8 * 25 + 9 * 75) / sum(no_contr)) %>%
    summarize(e_y1_d_is_1 - e_y0_d_is_0)
```

```
## # A tibble: 1 x 1
##   `e_y1_d_is_1 - e_y0_d_is_0`
##                         <dbl>
## 1                        2.75
```

1.3

The ATT is defined as $\mathbb{E}[\delta|D=1] = \mathbb{E}[Y^1|D=1] - \mathbb{E}[Y^0|D=1]$. The first term is readily observable. The second term is estimated by us as $\hat{\mathbb{E}}[Y^0|D=1] = \mathbb{E}[Y^0|D=0]$. Hence:

```
data_ate <- data %>%
    mutate(n = no_treated + no_contr) %>%
    summarize(e_y1_d_is_1 = (9 * 100 + 13 * 75 + 10 * 25) / sum(no_treated),
              e_y0_d_is_0 = (7 * 100 + 8 * 25 + 9 * 75) / sum(no_contr))

data_ate
```

```
## # A tibble: 1 x 2
##   e_y1_d_is_1 e_y0_d_is_0
##         <dbl>       <dbl>
## 1        10.6        7.88
```

```
data_ate %>%
    summarize(att = e_y1_d_is_1 - e_y0_d_is_0)
```

```
## # A tibble: 1 x 1
##     att
##   <dbl>
## 1  2.75
```

So the $ATE = ATT$ (because of randomization).

Question 2

2.1

Compute the fraction of students in all three groups (control, low-reward and high-reward) that complete all first-year courses before the start of the second academic year. Show within a table that background characteristics are balanced over the treatment groups.

```
bonus_clean <- bonus %>%
    pivot_longer(cols = c(bonus0, bonus500, bonus1500),
                 names_to = "kind_treatment",
                 values_to = "treatment") %>%
    filter(treatment == 1)

bonus_clean %>%
    group_by(kind_treatment) %>%
    summarize(fraction_pass = sum(pass)/n()) %>%
    kable(caption = "Fraction passed per treatment")
```

Table 1: Fraction passed per treatment

kind_treatment	fraction_pass
bonus0	0.1951220
bonus1500	0.2409639
bonus500	0.2023810

```r
#Drop the outcome variables
bonus_clean %>%
    group_by(kind_treatment) %>%
    select(-c(pass, stp2001, stp2004, dropout)) %>%
    summarize(across(p0:math,
                     list(mean = ~ mean(., na.rm = TRUE),
                          sd = ~ sd(., na.rm = TRUE)))) %>%
    pivot_longer(-kind_treatment, names_to = "variable") %>%
    separate(variable, into = c("var", "statistic"), sep = "_") %>%
    pivot_wider(names_from = c(kind_treatment, statistic)) %>%
    kable(caption = "Means and SDs according to treatment", digits = 3)
```

Table 2: Means and SDs according to treatment

var	bonus0_mean	bonus0_sd	bonus1500_mean	bonus1500_sd	bonus500_mean	bonus500_sd
p0	0.553	0.265	0.573	0.248	0.530	0.251
job	0.760	0.430	0.805	0.399	0.829	0.379
myeduc	12.293	3.041	12.590	2.992	12.119	3.316
fyeduc	13.378	3.416	13.422	3.596	13.524	3.273
effort	19.549	9.460	18.303	10.592	18.477	10.475
math	5.476	1.468	5.388	1.258	5.386	1.360

2.2

Use the linear probability model to regress the dummy variable for completing all courses on the assignment of the three treatment groups. Interpret the treatment effects. Next include as additional regressors father's education, high-school math score and the subjective assessment about the pass probability.

```r
# First attach data set for easy access:
attach(bonus)

# We need to regress pass on the treatment group assignment. This is
# a simple linear probability model. Model is denoted by prob_1.
# Note, we need to ommit bonus0, cause of a dummy variable trap.
prob_1 <- lm(pass ~ bonus500 + bonus1500, data=bonus)

# Same as before, but now with some extra variables
prob_2 <- lm(pass ~ bonus500 + bonus1500 + fyeduc + math + p0, data=bonus)

# for q2.3:
prob_3 <- lm(pass ~ bonus500 + bonus1500 + fyeduc + math + p0 + job + effort, data=bonus)
```

```
# I think we should take the largest model. BAS???:
prob_4 <- lm(pass ~ bonus500 + bonus1500 + fyeduc + math + p0 + job + effort + dropout + stp2001 + stp20

# Create table:
stargazer(prob_1, prob_2, prob_3, prob_4, header=FALSE, style='aer')
```

Table 3:

	pass			
	(1)	(2)	(3)	(4)
bonus500	0.007	0.015	0.023	0.047
	(0.064)	(0.057)	(0.058)	(0.048)
bonus1500	0.046	0.048	0.056	0.071
	(0.064)	(0.058)	(0.059)	(0.049)
fyeduc		−0.001	0.000	0.004
		(0.007)	(0.007)	(0.006)
math		0.119^{***}	0.124^{***}	0.050^{***}
		(0.018)	(0.018)	(0.017)
p0		0.249^{***}	0.170^{*}	0.060
		(0.095)	(0.098)	(0.083)
job			−0.062	-0.085^{*}
			(0.060)	(0.050)
effort			0.008^{***}	−0.003
			(0.002)	(0.002)
dropout				0.306^{***}
				(0.069)
stp2001				0.007^{***}
				(0.002)
stp2004				0.003^{***}
				(0.001)
Constant	0.195^{***}	-0.576^{***}	-0.678^{***}	-0.679^{***}
	(0.045)	(0.128)	(0.144)	(0.133)
Observations	249	245	230	230
R^2	0.002	0.209	0.264	0.503
Adjusted R^2	−0.006	0.192	0.240	0.480
Residual Std. Error	0.411 (df = 246)	0.366 (df = 239)	0.358 (df = 222)	0.296 (df = 219)
F Statistic	0.297 (df = 2; 246)	12.629^{***} (df = 5; 239)	11.359^{***} (df = 7; 222)	22.148^{***} (df = 10; 219)
Notes:	***Significant at the 1 percent level.			
	**Significant at the 5 percent level.			
	*Significant at the 10 percent level.			

2.3

Next also include as regressors in your model whether a student has a job and the amount of study effort. Comment on this approach. Do you consider this an improvement over (ii)?

2.4

Use your preferred model specifcation to estimate the effects of the financial incentives on some other outcomes: dropping out and credit points collected (in the first year and after three years).