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")
```

Table 2: Means and SDs according to treatment

var	bonus0_mean	bonus0_sd	bonus1500_mean	bonus1500_sd	bonus500_mean	bonus500_sd
p0	0.552561	0.2646769	0.5726807	0.2481354	0.5302381	0.2507145
job	0.760000	0.4299591	0.8051948	0.3986477	0.8292683	0.3785899
myeduc	12.292683	3.0408243	12.5903614	2.9920062	12.1190476	3.3162788
fyeduc	13.378049	3.4161571	13.4216867	3.5957986	13.5238095	3.2726548
effort	19.548788	9.4598240	18.3028336	10.5915504	18.4767184	10.4747985
math	5.475610	1.4675554	5.3875000	1.2578236	5.3855422	1.3599796