# Microeconometrics Task - Problem Set 2

Ricardo Semião e Castro

2024-08-27

## 1 Setup

```
library(tidyverse)
library(glue)
library(broom)

library(knitr)
library(kableExtra)
```

Loading data:

```
data_lot <- haven::read_dta("ps2/base_lotteries.dta")
data_exam <- haven::read_dta("ps2/voluntary_exam.dta")
```

## 2 Question 1

First, we standardize the variables.

```
cols <- expand.grid(c("mt", "lp"), 7:9) %>%
  apply(1, \(x) paste0(c("score", x), collapse = "_"))

data_lot <- data_lot %>%
  mutate(
    across(all_of(cols), function(col) {
      control <- col[won_lotteryA == 0]
      (col - mean(control)) / sd(control)
    }),
    across(sex:ever_failed_school, as.factor)
  ) %>%
  na.omit()
```

### 2.1 Item a)

The grades of the test are largely defined by the difficulty of the test. Without standardization, we would be capturing this effect, which would pollute the effect of the treatment between grades. When we standardize the scores, we are capturing the relative position of students, clensed from the difficulty effect.

### 2.2 Item b)

The averages and standard deviations are expected to change with the treatment, such that we only want to consider the changes in difficulty of the conterfactual, that is, we want to standardize using the averages and standard deviation of the control group.

# 3 Question 2.

## 3.1 Items a), and c)

First, lets create dummies for each variable.

```
labels <- list(
  sex = c("male", "female"),
  race = c("white", "brown", "black", "yellow", "indigenous"),
  mother = c("l.t. 5", "5 to 9", "9 to HS", "HS to college", "college", "more", "unsure"),
  father = c("l.t. 5", "5 to 9", "9 to HS", "HS to college", "college", "more", "unsure"),
  failed = c("never", "once", "more")
) %>%
  imap(~ paste(.y, .x)) %>%
  reduce(c)

data_dummies <- fastDummies::dummy_cols(data_lot[2:6], remove_selected_columns = TRUE) %>%
  set_names(labels) %>%
  cbind(data_lot[-c(1:6)])
```

Now, we can create a function that makes the comparisons.

```
comparate <- function(data, treat_ind, control_ind = !treat_ind) {
  imap_dfr(data, function(col, name) {
    control <- col[treat_ind]
    treat <- col[control_ind]

    test <- tryCatch(t.test(control, treat), error = \(e) list(p.value = NA))
    n <- c(length(control), length(treat))

    results <- c(
      mean(control),
      sd(control),
      mean(treat) - mean(control),
      sqrt((sd(control)^2 / n[1]) + (sd(treat)^2 / n[2])),
      test$p.value
    ) %>%
      round(3)

    tibble(
      Variable = name,
      Control = glue("{results[1]} ({results[2]})"),
      Diff = glue("{results[3]} ({results[4]})"),
      `P-value` = results[5],
      `Ajd. P-value` = p.adjust(`P-value`, method = "holm")
    )
  })
}
```

The results can be seen at the end of the question.

## 3.2 Items b) and c)

As with everything in inference, there is a chance of any unbalancedness to arise randomly, which wouldn't indicate a problem with the lottery. With lots of control variables, the probability of such random deviation is relevant.

But, if the difference in groups was not random, it can be an indication that the individuals are self selecting, or that the lottery is not fair. This would put in check the hypothesis of the exogeneity of the treatment.

To get further evidence, we could test if any unbalancedness is correlated with the treatment, with methods/tests such as Randomization Inference, Bonferroni, and Holm.

At the end, I present the results with the Hold p-value adjustment.

We can see that some variables present statistically significant differences, meaning that the evidence for balanced groups and lottery fairness is not good.

## 3.3 Item d)

Then, the hypothesis of the exogeneity of the treatment would be hard to defend, once that we have an indication that the individuals are self selecting, or that the lottery is not fair.

## 3.4 Item e)

Then, the comparison talked above would be valid, since the lottery would similarly divide the groups.

## 3.5 Results

Below I present the results from this question. They wouldn't fit in a horizontally doubled table with both comparisons, so I presented one table for each comparison.

```
comparison_treated <- select(data_dummies, `sex male`:`failed more`) %>%
  comparate(data_dummies$won_lotteryA + data_dummies$won_lotteryB >= 1)

comparison_treated %>% kable() %>% kable_styling(latex_options = "HOLD_position")
```

| Variable | Control | Diff | P-value | Ajd. P-value |
|---|---|---|---|---|
| sex male | 0.405 (0.492) | -0.001 (0.046) | 0.989 | 0.989 |
| sex female | 0.595 (0.492) | 0.001 (0.046) | 0.989 | 0.989 |
| race white | 0.265 (0.442) | -0.013 (0.041) | 0.748 | 0.748 |
| race brown | 0.502 (0.501) | 0.05 (0.046) | 0.286 | 0.286 |
| race black | 0.191 (0.394) | -0.039 (0.035) | 0.272 | 0.272 |
| race yellow | 0.019 (0.135) | 0.013 (0.014) | 0.355 | 0.355 |
| race indigenous | 0.023 (0.151) | -0.011 (0.012) | 0.365 | 0.365 |
| mother l.t. 5 | 0.07 (0.255) | 0.038 (0.026) | 0.146 | 0.146 |
| mother 5 to 9 | 0.214 (0.411) | 0.026 (0.039) | 0.504 | 0.504 |
| mother 9 to HS | 0.153 (0.361) | -0.049 (0.031) | 0.115 | 0.115 |
| mother HS to college | 0.298 (0.458) | 0.026 (0.043) | 0.542 | 0.542 |
| mother college | 0.107 (0.31) | -0.023 (0.027) | 0.404 | 0.404 |
| mother more | 0.07 (0.255) | -0.014 (0.023) | 0.545 | 0.545 |
| mother unsure | 0.088 (0.284) | -0.004 (0.026) | 0.867 | 0.867 |
| father l.t. 5 | 0.112 (0.316) | 0 (0.029) | 0.990 | 0.990 |
| father 5 to 9 | 0.214 (0.411) | 0.026 (0.039) | 0.504 | 0.504 |
| father 9 to HS | 0.14 (0.347) | -0.048 (0.03) | 0.113 | 0.113 |
| father HS to college | 0.251 (0.435) | 0.017 (0.041) | 0.680 | 0.680 |
| father college | 0.056 (0.23) | 0.016 (0.023) | 0.476 | 0.476 |
| father more | 0.019 (0.135) | -0.015 (0.01) | 0.148 | 0.148 |
| father unsure | 0.209 (0.408) | 0.003 (0.038) | 0.943 | 0.943 |
| failed never | 0.712 (0.454) | 0.012 (0.042) | 0.768 | 0.768 |
| failed once | 0.2 (0.401) | 0.02 (0.038) | 0.598 | 0.598 |

| | | | | |
|---|---|---|---|---|
| failed more | 0.088 (0.284) | -0.032 (0.024) | 0.183 | 0.183 |

```r
comparison_ab <- select(data_dummies, `sex male`:`failed more`) %>%
  comparate(data_dummies$won_lotteryA == 1, data_dummies$won_lotteryB == 1)

comparison_ab %>% kable() %>% kable_styling(latex_options = "HOLD_position")
```

| Variable | Control | Diff | P-value | Ajd. P-value |
|---|---|---|---|---|
| sex male | 0.432 (0.497) | -0.084 (0.071) | 0.240 | 0.240 |
| sex female | 0.568 (0.497) | 0.084 (0.071) | 0.240 | 0.240 |
| race white | 0.24 (0.428) | 0.079 (0.067) | 0.238 | 0.238 |
| race brown | 0.514 (0.502) | -0.035 (0.073) | 0.630 | 0.630 |
| race black | 0.192 (0.395) | -0.003 (0.058) | 0.953 | 0.953 |
| race yellow | 0.021 (0.142) | -0.006 (0.019) | 0.746 | 0.746 |
| race indigenous | 0.034 (0.182) | -0.034 (0.015) | 0.025 | 0.025 |
| mother l.t. 5 | 0.062 (0.241) | 0.025 (0.04) | 0.524 | 0.524 |
| mother 5 to 9 | 0.212 (0.41) | 0.005 (0.06) | 0.933 | 0.933 |
| mother 9 to HS | 0.151 (0.359) | 0.009 (0.053) | 0.870 | 0.870 |
| mother HS to college | 0.308 (0.463) | -0.033 (0.066) | 0.621 | 0.621 |
| mother college | 0.116 (0.322) | -0.029 (0.043) | 0.497 | 0.497 |
| mother more | 0.068 (0.253) | 0.004 (0.038) | 0.916 | 0.916 |
| mother unsure | 0.082 (0.276) | 0.019 (0.043) | 0.656 | 0.656 |
| father l.t. 5 | 0.11 (0.313) | 0.006 (0.047) | 0.892 | 0.892 |
| father 5 to 9 | 0.247 (0.433) | -0.102 (0.056) | 0.070 | 0.070 |
| father 9 to HS | 0.151 (0.359) | -0.035 (0.049) | 0.478 | 0.478 |
| father HS to college | 0.226 (0.42) | 0.078 (0.066) | 0.236 | 0.236 |
| father college | 0.048 (0.214) | 0.025 (0.036) | 0.498 | 0.498 |
| father more | 0.021 (0.142) | -0.006 (0.019) | 0.746 | 0.746 |
| father unsure | 0.199 (0.4) | 0.033 (0.061) | 0.586 | 0.586 |
| failed never | 0.678 (0.469) | 0.105 (0.063) | 0.101 | 0.101 |
| failed once | 0.212 (0.41) | -0.038 (0.057) | 0.503 | 0.503 |
| failed more | 0.11 (0.313) | -0.066 (0.036) | 0.067 | 0.067 |

# 4 Question 3

```r
form_base <- score_mt_8 ~ won_lotteryA
mod_base <- lm(form_base, data_lot)

form_controls <- update.formula(form_base,
  . ~ . + sex + race + mother_schooling + father_schooling + ever_failed_school
)
mod_controls <- lm(form_controls, data_lot)
```

```r
test_permutation <- function(f, data, n = 1000) {
  coef_true <- abs(coef(lm(f, data)))
  n_coef <- length(coef_true)

  coef_perm <- map_dfr(1:n, function(i) {
    data$score_mt_8 <- sample(data$score_mt_8)
```

```r
    abs(coef(lm(f, data)))
  })

  sapply(1:n_coef, \(i) mean(coef_perm[,i] >= coef_true[i]))
}
```

## 4.1 Item c)

Using averages differences does not account for the different variances in the potential outcomes, such that the test, while asymptotically valid, can underreject. That is why we used the adjustments.

## 4.2 Item d)

```r
test_wildboot <- function(f, data, n = 1000) {
  mod <- lm(f, data)
  res <- resid(mod)
  pred <- predict(mod)
  coef_true <- coef(mod)
  n_coef <- length(coef_true)

  coef_perm <- map_dfr(1:n, function(i) {
    y <- pred + res * sample(c(-1, 1), length(residuals), replace = TRUE)
    data$score_mt_8 <- y
    coef(lm(f, data))
  })

  sapply(1:n_coef, \(i) mean(coef_perm[,i] >= coef_true[i]))
}
```

## 4.3 Results

```r
summary(mod_base)$coef[,c(1,4)] %>%
  as_tibble(rownames = "Coefficient") %>%
  mutate(
    `Pr Perm` = test_permutation(form_base, data_lot),
    `Pr Wild` = test_wildboot(form_base, data_lot)
  ) %>%
  kable() %>%
  kable_styling(latex_options = "HOLD_position")
```

| Coefficient | Estimate | Pr(>|t|) | Pr Perm | Pr Wild |
|---|---|---|---|---|
| (Intercept) | 0.0008760 | 0.9873211 | 0.998 | 1.000 |
| won_lotteryA | 0.2389622 | 0.0154668 | 0.020 | 0.513 |

```r
summary(mod_controls)$coef[,c(1,4)] %>%
  as_tibble(rownames = "Coefficient") %>%
  mutate(
    `Pr Perm` = test_permutation(form_controls, data_lot),
    `Pr Wild` = test_wildboot(form_controls, data_lot)
  ) %>%
```

```
  kable() %>%
  kable_styling(latex_options = "HOLD_position")
```

| Coefficient | Estimate | Pr(>|t|) | Pr Perm | Pr Wild |
|---|---|---|---|---|
| (Intercept) | 0.4046692 | 0.0444820 | 0.059 | 1.000 |
| won_lotteryA | 0.2479112 | 0.0123628 | 0.015 | 0.000 |
| sex2 | -0.0446908 | 0.6380350 | 0.656 | 0.000 |
| race2 | -0.0408518 | 0.7126682 | 0.726 | 1.000 |
| race3 | -0.2117029 | 0.1454232 | 0.181 | 0.000 |
| race4 | 0.7114344 | 0.0170654 | 0.022 | 1.000 |
| race5 | 0.0135123 | 0.9704270 | 0.966 | 0.000 |
| mother_schooling2 | -0.2296147 | 0.2294087 | 0.226 | 0.000 |
| mother_schooling3 | -0.0790496 | 0.7056870 | 0.710 | 1.000 |
| mother_schooling4 | -0.1592331 | 0.3976212 | 0.410 | 1.000 |
| mother_schooling5 | -0.0054524 | 0.9815636 | 0.980 | 1.000 |
| mother_schooling6 | -0.0452888 | 0.8596504 | 0.866 | 1.000 |
| mother_schooling7 | -0.3237400 | 0.1856243 | 0.205 | 1.000 |
| father_schooling2 | -0.0593496 | 0.7393442 | 0.752 | 1.000 |
| father_schooling3 | -0.1890603 | 0.3524776 | 0.364 | 0.000 |
| father_schooling4 | -0.0505090 | 0.7786279 | 0.775 | 0.000 |
| father_schooling5 | -0.1814210 | 0.4622589 | 0.464 | 0.000 |
| father_schooling6 | -0.0420376 | 0.9306210 | 0.946 | 0.506 |
| father_schooling7 | -0.1408316 | 0.4568370 | 0.474 | 0.000 |
| ever_failed_school2 | -0.5098885 | 0.0000103 | 0.000 | 1.000 |
| ever_failed_school3 | -0.0036747 | 0.9836599 | 0.975 | 1.000 |

# 5 Question 4

## 5.1 Item a)

Two options can be:

- Define a new, combined, dependent variable. For example, any weighted sum of the outcomes. In the most simple way, just the sum or mean of both.
- Do separate regressions, but jointly test the p-values using Bonferroni or Holm methods.

## 5.2 Item b)

I did both approaches from above. The results are in the next section.

```
mod_sum <- lm(
  score_mt_8 + score_lp_8 ~ won_lotteryA + sex + race + mother_schooling +
    father_schooling + ever_failed_school,
  data_lot
)
```

```
mod_mt <- lm(
  score_mt_8 ~ won_lotteryA + sex + race + mother_schooling +
    father_schooling + ever_failed_school,
  data_lot
)
mod_lp <- lm(
```

```r
  score_lp_8 ~ won_lotteryA + sex + race + mother_schooling +
    father_schooling + ever_failed_school,
  data_lot
)

adj_pvalues <- p.adjust(
  c(summary(mod_mt)$coefficients[, 4], summary(mod_lp)$coefficients[, 4]),
  method = "holm"
)
```

### 5.3 Results

Again, all in one table wouldn't be very easy to see.

```r
summary(mod_sum)$coef[,c(1,4)] %>%
  as_tibble(rownames = "Coefficient") %>%
  kable() %>%
  kable_styling(latex_options = "HOLD_position")
```

| Coefficient | Estimate | Pr(>|t|) |
|---|---:|---:|
| (Intercept) | 0.3411289 | 0.2367740 |
| won_lotteryA | 0.4055460 | 0.0043610 |
| sex2 | 0.1257289 | 0.3561922 |
| race2 | -0.2444985 | 0.1247445 |
| race3 | -0.3436771 | 0.0994232 |
| race4 | 0.4217956 | 0.3227434 |
| race5 | -0.3195906 | 0.5409690 |
| mother_schooling2 | -0.2172668 | 0.4275198 |
| mother_schooling3 | -0.1260763 | 0.6744632 |
| mother_schooling4 | -0.3208463 | 0.2347856 |
| mother_schooling5 | -0.0718971 | 0.8317221 |
| mother_schooling6 | -0.1179993 | 0.7480111 |
| mother_schooling7 | -0.8283897 | 0.0184300 |
| father_schooling2 | 0.2774653 | 0.2783181 |
| father_schooling3 | -0.0813376 | 0.7801825 |
| father_schooling4 | 0.1855781 | 0.4714731 |
| father_schooling5 | 0.0077334 | 0.9825601 |
| father_schooling6 | -0.0188552 | 0.9782741 |
| father_schooling7 | 0.1081930 | 0.6901004 |
| ever_failed_school2 | -0.5917364 | 0.0003394 |
| ever_failed_school3 | -0.0031021 | 0.9903802 |

```r
summary(mod_mt)$coef[,c(1,4)] %>%
  as_tibble(rownames = "Coefficient") %>%
  mutate( `Pr Adj` = adj_pvalues[1:length(coef(mod_mt))]) %>%
  kable() %>%
  kable_styling(latex_options = "HOLD_position")
```

| Coefficient | Estimate | Pr(>|t|) | Pr Adj |
|---|---:|---:|---:|
| (Intercept) | 0.4046692 | 0.0444820 | 1.0000000 |

| | | | |
|---|---|---|---|
| won_lotteryA | 0.2479112 | 0.0123628 | 0.5068754 |
| sex2 | -0.0446908 | 0.6380350 | 1.0000000 |
| race2 | -0.0408518 | 0.7126682 | 1.0000000 |
| race3 | -0.2117029 | 0.1454232 | 1.0000000 |
| race4 | 0.7114344 | 0.0170654 | 0.6826161 |
| race5 | 0.0135123 | 0.9704270 | 1.0000000 |
| mother_schooling2 | -0.2296147 | 0.2294087 | 1.0000000 |
| mother_schooling3 | -0.0790496 | 0.7056870 | 1.0000000 |
| mother_schooling4 | -0.1592331 | 0.3976212 | 1.0000000 |
| mother_schooling5 | -0.0054524 | 0.9815636 | 1.0000000 |
| mother_schooling6 | -0.0452888 | 0.8596504 | 1.0000000 |
| mother_schooling7 | -0.3237400 | 0.1856243 | 1.0000000 |
| father_schooling2 | -0.0593496 | 0.7393442 | 1.0000000 |
| father_schooling3 | -0.1890603 | 0.3524776 | 1.0000000 |
| father_schooling4 | -0.0505090 | 0.7786279 | 1.0000000 |
| father_schooling5 | -0.1814210 | 0.4622589 | 1.0000000 |
| father_schooling6 | -0.0420376 | 0.9306210 | 1.0000000 |
| father_schooling7 | -0.1408316 | 0.4568370 | 1.0000000 |
| ever_failed_school2 | -0.5098885 | 0.0000103 | 0.0004324 |
| ever_failed_school3 | -0.0036747 | 0.9836599 | 1.0000000 |

```
summary(mod_lp)$coef[,c(1,4)] %>%
  as_tibble(rownames = "Coefficient") %>%
  mutate( `Pr Adj` = adj_pvalues[1:length(coef(mod_lp))]) %>%
  kable() %>%
  kable_styling(latex_options = "HOLD_position")
```

| Coefficient | Estimate | Pr(>|t|) | Pr Adj |
|---|---|---|---|
| (Intercept) | -0.0635403 | 0.7645948 | 1.0000000 |
| won_lotteryA | 0.1576348 | 0.1311373 | 0.5068754 |
| sex2 | 0.1704197 | 0.0898444 | 1.0000000 |
| race2 | -0.2036467 | 0.0826321 | 1.0000000 |
| race3 | -0.1319742 | 0.3897353 | 1.0000000 |
| race4 | -0.2896389 | 0.3564862 | 0.6826161 |
| race5 | -0.3331029 | 0.3870123 | 1.0000000 |
| mother_schooling2 | 0.0123479 | 0.9511576 | 1.0000000 |
| mother_schooling3 | -0.0470267 | 0.8315169 | 1.0000000 |
| mother_schooling4 | -0.1616132 | 0.4162259 | 1.0000000 |
| mother_schooling5 | -0.0664447 | 0.7897361 | 1.0000000 |
| mother_schooling6 | -0.0727105 | 0.7880761 | 1.0000000 |
| mother_schooling7 | -0.5046497 | 0.0509931 | 1.0000000 |
| father_schooling2 | 0.3368149 | 0.0742731 | 1.0000000 |
| father_schooling3 | 0.1077227 | 0.6157820 | 1.0000000 |
| father_schooling4 | 0.2360871 | 0.2137992 | 1.0000000 |
| father_schooling5 | 0.1891544 | 0.4679616 | 1.0000000 |
| father_schooling6 | 0.0231823 | 0.9637367 | 1.0000000 |
| father_schooling7 | 0.2490246 | 0.2130699 | 1.0000000 |
| ever_failed_school2 | -0.0818479 | 0.4979544 | 0.0004324 |

| | | | | |
|---|---|---|---|---|
| ever_failed_school3 | 0.0005726 | | 0.9975887 | 1.0000000 |

# 6 Question 5

In this question, i'll focus on the computational answers.

## 6.1 Item a)

```
data_join <- full_join(data_lot, data_exam, by = "student_code") %>%
  filter(lotteryA == 1) %>%
  mutate(
    did_test = !is.na(voluntary_exam_score),
    military = !(is.na(won_lotteryA) | won_lotteryA == 0)
  )

summary(lm(did_test ~ military, data_join))
```

```
##
## Call:
## lm(formula = did_test ~ military, data = data_join)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -3.450e-16 -3.450e-16 -3.450e-16  0.000e+00  1.098e-13
##
## Coefficients:
##                 Estimate Std. Error    t value Pr(>|t|)
## (Intercept)    1.000e+00  2.861e-16  3.495e+15   <2e-16 ***
## militaryTRUE -3.452e-16  5.106e-16 -6.760e-01    0.499
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.11e-15 on 463 degrees of freedom
## Multiple R-squared:  0.5006, Adjusted R-squared:  0.4995
## F-statistic: 464.1 on 1 and 463 DF,  p-value: < 2.2e-16
```

## 6.2 Item b)

```
data_dummies <- filter(data_join, did_test)[2:6] %>%
  fastDummies::dummy_cols(remove_selected_columns = TRUE) %>%
  set_names(labels) %>%
  cbind(data_lot[-c(1:6)]) %>%
  na.omit()

comparison_treated <- select(data_dummies, `sex male`:`failed more`) %>%
  comparate(data_dummies$won_lotteryA == 1)

comparison_treated %>% kable() %>% kable_styling(latex_options = "HOLD_position")
```

| Variable | Control | Diff | P-value | Ajd. P-value |
|---|---|---|---|---|
| sex male | 0.432 (0.497) | -0.04 (0.049) | 0.423 | 0.423 |

| | | | | |
|---|---|---|---|---|
| sex female | 0.568 (0.497) | 0.04 (0.049) | 0.423 | 0.423 |
| race white | 0.24 (0.428) | 0.027 (0.043) | 0.537 | 0.537 |
| race brown | 0.514 (0.502) | 0.022 (0.05) | 0.656 | 0.656 |
| race black | 0.192 (0.395) | -0.032 (0.039) | 0.409 | 0.409 |
| race yellow | 0.021 (0.142) | 0.008 (0.015) | 0.610 | 0.610 |
| race indigenous | 0.034 (0.182) | -0.025 (0.016) | 0.123 | 0.123 |
| mother l.t. 5 | 0.062 (0.241) | 0.042 (0.026) | 0.113 | 0.113 |
| mother 5 to 9 | 0.212 (0.41) | 0.023 (0.041) | 0.583 | 0.583 |
| mother 9 to HS | 0.151 (0.359) | -0.035 (0.035) | 0.318 | 0.318 |
| mother HS to college | 0.308 (0.463) | 0.005 (0.046) | 0.910 | 0.910 |
| mother college | 0.116 (0.322) | -0.032 (0.031) | 0.304 | 0.304 |
| mother more | 0.068 (0.253) | -0.009 (0.025) | 0.719 | 0.719 |
| mother unsure | 0.082 (0.276) | 0.006 (0.028) | 0.841 | 0.841 |
| father l.t. 5 | 0.11 (0.313) | 0.003 (0.031) | 0.917 | 0.917 |
| father 5 to 9 | 0.247 (0.433) | -0.027 (0.043) | 0.525 | 0.525 |
| father 9 to HS | 0.151 (0.359) | -0.054 (0.034) | 0.117 | 0.117 |
| father HS to college | 0.226 (0.42) | 0.05 (0.043) | 0.246 | 0.246 |
| father college | 0.048 (0.214) | 0.024 (0.023) | 0.293 | 0.293 |
| father more | 0.021 (0.142) | -0.014 (0.013) | 0.258 | 0.258 |
| father unsure | 0.199 (0.4) | 0.018 (0.04) | 0.662 | 0.662 |
| failed never | 0.678 (0.469) | 0.059 (0.046) | 0.204 | 0.204 |
| failed once | 0.212 (0.41) | -0.002 (0.041) | 0.955 | 0.955 |
| failed more | 0.11 (0.313) | -0.056 (0.029) | 0.052 | 0.052 |

```r
comparison_treated <- select(data_dummies, `sex male`:`failed more`) %>%
  comparate(data_dummies$won_lotteryB == 1)

comparison_treated %>% kable() %>% kable_styling(latex_options = "HOLD_position")
```

| Variable | Control | Diff | P-value | Ajd. P-value |
|---|---|---|---|---|
| sex male | 0.348 (0.48) | 0.066 (0.063) | 0.294 | 0.294 |
| sex female | 0.652 (0.48) | -0.066 (0.063) | 0.294 | 0.294 |
| race white | 0.319 (0.469) | -0.071 (0.061) | 0.242 | 0.242 |
| race brown | 0.478 (0.503) | 0.06 (0.066) | 0.366 | 0.366 |
| race black | 0.188 (0.394) | -0.022 (0.051) | 0.671 | 0.671 |
| race yellow | 0.014 (0.12) | 0.013 (0.017) | 0.428 | 0.428 |
| race indigenous | 0 (0) | 0.02 (0.007) | 0.005 | 0.005 |
| mother l.t. 5 | 0.087 (0.284) | 0.004 (0.037) | 0.915 | 0.915 |
| mother 5 to 9 | 0.217 (0.415) | 0.012 (0.054) | 0.820 | 0.820 |
| mother 9 to HS | 0.159 (0.369) | -0.038 (0.047) | 0.422 | 0.422 |
| mother HS to college | 0.275 (0.45) | 0.043 (0.059) | 0.470 | 0.470 |
| mother college | 0.087 (0.284) | 0.009 (0.037) | 0.810 | 0.810 |
| mother more | 0.072 (0.261) | -0.012 (0.034) | 0.725 | 0.725 |
| mother unsure | 0.101 (0.304) | -0.018 (0.039) | 0.645 | 0.645 |
| father l.t. 5 | 0.116 (0.323) | -0.005 (0.042) | 0.909 | 0.909 |
| father 5 to 9 | 0.145 (0.355) | 0.097 (0.048) | 0.044 | 0.044 |
| father 9 to HS | 0.116 (0.323) | -0.002 (0.042) | 0.956 | 0.956 |
| father HS to college | 0.304 (0.464) | -0.052 (0.06) | 0.389 | 0.389 |
| father college | 0.072 (0.261) | -0.009 (0.034) | 0.783 | 0.783 |

| | | | | |
|---|---|---|---|---|
| father more | 0.014 (0.12) | -0.004 (0.015) | 0.775 | 0.775 |
| father unsure | 0.232 (0.425) | -0.025 (0.055) | 0.653 | 0.653 |
| failed never | 0.783 (0.415) | -0.076 (0.055) | 0.173 | 0.173 |
| failed once | 0.174 (0.382) | 0.043 (0.05) | 0.393 | 0.393 |
| failed more | 0.043 (0.205) | 0.032 (0.028) | 0.253 | 0.253 |

# 7 Item e)

```
leedata <- data_join %>%
  select(military, did_test, voluntary_exam_score) %>%
  setNames(c("treat", "selection", "outcome"))

leebounds::leebounds(leedata)
```

```
## $lower_bound
## [1] 0.2242981
##
## $upper_bound
## [1] 0.2242981
##
## $p0
## [1] 1
##
## $trimmed_mean_upper
## [1] 6.876583
##
## $trimmed_mean_lower
## [1] 6.876583
##
## $mean_no_trim
## [1] 6.652285
##
## $odds
## [1] 0.4576803
##
## $yp0
##     100%
## 9.779208
##
## $y1p0
##       0%
## 1.336305
##
## $s0
## [1] 1
##
## $s1
## [1] 1
##
## $prop0
## [1] 0.6860215
##
```

```
## $prop1
## [1] 0.3139785
```