14.750 Problem Set 4

Problem 1.a

```
library(tidyverse)
library(lfe)
library(stargazer)

sports <- read.csv("SportsNationBuilding.csv") %>%
  mutate(ethnicid = as.numeric(ethnic_sentiment %in% c("Ethnic id only", "Ethnic id more than
national")))
```

Problem 1.b

```
sports <- sports %>%
  mutate(post = as.numeric(dist_match < 0))
```

Problem 1.c

```
m1 <- felm(
  data = sports,
  formula =
    ethnicid ~
    post + post_victory + male + age + age_sq + unemployed + rural + education |
    country_match_fe + language_year_id + dayweek + month + day |
    0 |
    country_year_fe
  )
summary(m1)
```

```
Call:
   felm(formula = ethnicid ~ post + post_victory + male + age +      age_sq + unemployed + ru
ral + education | country_match_fe +      language_year_id + dayweek + month + day | 0 | coun
try_year_fe,      data = sports)

Residuals:
    Min      1Q  Median      3Q     Max
-0.9313 -0.1647 -0.0978 -0.0212  1.0921

Coefficients:
               Estimate Cluster s.e. t value Pr(>|t|)
post          -6.816e-04    1.629e-02  -0.042 0.966762
post_victory  -5.221e-02    1.912e-02  -2.731 0.008392 **
male          -1.593e-02    6.386e-03  -2.494 0.015549 *
age           -2.488e-03    6.760e-04  -3.680 0.000519 ***
age_sq         2.686e-05    7.731e-06   3.474 0.000986 ***
unemployed     6.947e-03    4.029e-03   1.724 0.090071 .
rural          1.340e-02    7.004e-03   1.913 0.060725 .
education     -1.463e-02    1.665e-03  -8.785 3.48e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3338 on 36301 degrees of freedom
Multiple R-squared(full model): 0.1055   Adjusted R-squared: 0.08495
Multiple R-squared(proj model): 0.009611   Adjusted R-squared: -0.01309
F-statistic(full model, *iid*):5.144 on 832 and 36301 DF, p-value: < 2.2e-16
F-statistic(proj model): 11.08 on 8 and 57 DF, p-value: 2.804e-09
*** Standard errors may be too high due to more than 2 groups and exactDOF=FALSE
```

Hide

```
stargazer(m1, type = "latex", out = "PS4-1c.tex")
```

Problem 1.d

Hide

```
for (i in seq(1, 10)) {
  first <- 3*(i - 1) + 1
  last <- 3*i
  sports <- sports %>%
    mutate("dist_match_{i}" := as.numeric(dist_match %in% sort(unique(dist_match), decreasing
= TRUE)[first : last]))
}
sports <- sports %>% select(-dist_match_5)

for (i in seq(1, 4)) {
  sports <- sports %>%
    mutate("dist_match_{i}_win" := .data[[paste("dist_match_", i, sep = "")]]*future_victory)
}
for (i in seq(6, 10)) {
  sports <- sports %>%
    mutate("dist_match_{i}_win" := .data[[paste("dist_match_", i, sep = "")]]*post_victory)
}

equation <- paste(
  "ethnicid ~",
  paste("dist_match_", 1:4, "_win", sep = "", collapse = " + "),
  "+",
  paste("dist_match_", 6:10, "_win", sep = "", collapse = " + "),
  "+",
  "male + age + age_sq + unemployed + rural + education |",
  "country_match_fe + language_year_id + dayweek + month + day |",
  "0 |",
  "country_year_fe"
)
m2 <- felm(
  data = sports,
  formula = as.formula(equation)
  )
summary(m2)
```

```
Call:
    felm(formula = as.formula(equation), data = sports)

Residuals:
     Min       1Q   Median       3Q      Max
-0.93135 -0.16406 -0.09787 -0.02095  1.09022

Coefficients:
                    Estimate Cluster s.e. t value Pr(>|t|)
dist_match_1_win  -3.389e-02    1.838e-02  -1.844 0.070381 .
dist_match_2_win  -1.117e-02    1.424e-02  -0.784 0.436190
dist_match_3_win  -2.665e-02    1.200e-02  -2.221 0.030306 *
dist_match_4_win  -1.673e-02    1.379e-02  -1.214 0.229895
dist_match_6_win  -6.324e-02    1.685e-02  -3.753 0.000412 ***
dist_match_7_win  -4.416e-02    2.695e-02  -1.639 0.106822
dist_match_8_win  -6.419e-02    3.098e-02  -2.072 0.042799 *
dist_match_9_win  -7.680e-02    3.429e-02  -2.240 0.029022 *
dist_match_10_win -8.033e-02    3.784e-02  -2.123 0.038098 *
male              -1.590e-02    6.377e-03  -2.493 0.015597 *
age               -2.485e-03    6.747e-04  -3.683 0.000515 ***
age_sq             2.683e-05    7.717e-06   3.477 0.000978 ***
unemployed         6.967e-03    4.058e-03   1.717 0.091442 .
rural              1.304e-02    6.860e-03   1.900 0.062469 .
education         -1.464e-02    1.672e-03  -8.758 3.84e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3338 on 36294 degrees of freedom
Multiple R-squared(full model): 0.1057   Adjusted R-squared: 0.08503
Multiple R-squared(proj model): 0.009885   Adjusted R-squared: -0.013
F-statistic(full model, *iid*):5.113 on 839 and 36294 DF, p-value: < 2.2e-16
F-statistic(proj model): 12.26 on 15 and 57 DF, p-value: 9.06e-13
*** Standard errors may be too high due to more than 2 groups and exactDOF=FALSE
```

Hide

```
stargazer(m2, type = "latex", out = "PS4-1d.tex")
```

Hide

```
chart <- data.frame(coef = coef(m2)[1:9], se = m2$cse[1:9]) %>%
  rownames_to_column(var = "var") %>%
  rbind(c("dist_match_5_win", 0, 0)) %>%
  mutate(x = str_remove(var, "dist_match_")) %>%
  mutate(x = as.numeric(str_remove(x, "_win"))) %>%
  mutate(
    coef = as.numeric(coef),
    se = as.numeric(se),
    lb = coef - se*1.96,
    ub = coef + se*1.96
  )

ggplot(chart, aes(x = x, y = coef, ymin = lb, ymax = ub)) +
  geom_pointrange(color = "darkblue") +
  geom_hline(yintercept = 0, color = "darkred") +
  geom_vline(xintercept = 5.5, color = "darkblue", linetype = "dashed") +
  scale_x_continuous(breaks = seq(1, 10), minor_breaks = NULL, labels = c(15, 12, 9, 6, 3, -3
, -6, -9, -12, -15)) +
  scale_y_continuous(limits = c(NA, 0.05), minor_breaks = NULL) +
  xlab("Distance to the match") +
  ylab("Impact on ethnic identification")

ggsave("PS4-1d.png")
```

```
Saving 7.29 x 4.5 in image
```