main_regression

Nadia Avianti

2024-12-10

Set up

```
# Import library
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.4.2
```

```
library(estimatr)
library(modelsummary)
```

```
## Warning: package 'modelsummary' was built under R version 4.4.2
```

```
## `modelsummary` 2.0.0 now uses `tinytable` as its default table-drawing
##   backend. Learn more at: https://vincentarelbundock.github.io/tinytable/
##
## Revert to `kableExtra` for one session:
##
##   options(modelsummary_factory_default = 'kableExtra')
##   options(modelsummary_factory_latex = 'kableExtra')
##   options(modelsummary_factory_html = 'kableExtra')
##
## Silence this message forever:
##
##   config_modelsummary(startup_message = FALSE)
```

```r
# Import data
setwd("C:/Users/nadia/OneDrive/Documents/GitHub/SPI507")
data <- read_excel("District_Data.xlsx")
head(data)
```

```
## # A tibble: 6 x 21
##   district_id state_id county_fips commute_zone pass_math_2019 pass_math_2021
##         <dbl>    <dbl>       <dbl>        <dbl>          <dbl>          <dbl>
## 1          10        1        8001            5          0.191          0.119
## 2          20        1        8001            5          0.350          0.283
## 3          30        1        8001            5          0.136          0.04
## 4          40        1        8001            5          0.279          0.186
## 5          50        1        8001            5          0.243          0.107
## 6          60        1        8001            5          0.362          0.242
## # i 15 more variables: pass_ela_2019 <dbl>, pass_ela_2021 <dbl>,
## #   share_black <dbl>, share_hisp <dbl>, share_white <dbl>, share_other <dbl>,
## #   share_lunch <dbl>, share_ell <dbl>, share_inperson <dbl>,
## #   share_virtual <dbl>, share_hybrid <dbl>, participation_math_2019 <dbl>,
## #   participation_math_2021 <dbl>, participation_ela_2019 <dbl>,
## #   participation_ela_2021 <dbl>
```

Prep data

```r
# Calculate diff
data$pass_math_diff <- data$pass_math_2021 - data$pass_math_2019
data$pass_ela_diff <- data$pass_ela_2021 - data$pass_ela_2019
data$pass_avg_diff <- (data$pass_math_diff + data$pass_ela_diff)/2

# Interaction variables
demographics <- c("black", "hisp", "white", "other", "lunch", "ell")

for (demographic in demographics) {
  str_demo <-
  data[[paste0("int_virtual_", demographic)]] <- data$share_virtual * data[[paste0("share_", demographi
  data[[paste0("int_hybrid_", demographic)]] <- data$share_hybrid * data[[paste0("share_", demographic)]
}
```

Run models

```r
# Create list to store results
results <- list()

# Run regressions
subjects <- c("math", "ela", "avg")
for (subject in subjects) {
  # Set model name
  name_reg1 <- paste0(subject, "_reg1")
  name_reg2 <- paste0(subject, "_reg2")
```

```
  name_reg3 <- paste0(subject, "_reg3")

  dep_var <- paste0("pass_", subject, "_diff")

  # Controlling state_id
  results[[name_reg1]] <- lm_robust(data[[dep_var]] ~ share_virtual + share_hybrid
                                    + state_id
                                    + int_virtual_black + int_hybrid_black
                                    + int_virtual_hisp + int_hybrid_hisp
                                    + int_virtual_lunch + int_hybrid_lunch
                                    , data = data, se_type = "stata")

  # Controlling state_id + county_fips
  results[[name_reg2]] <- lm_robust(data[[dep_var]] ~ share_virtual + share_hybrid
                                    + state_id + county_fips
                                    + int_virtual_black + int_hybrid_black
                                    + int_virtual_hisp + int_hybrid_hisp
                                    + int_virtual_lunch + int_hybrid_lunch
                                    , data = data, se_type = "stata")

  # Controlling state_id + commute_zone
  results[[name_reg3]] <- lm_robust(data[[dep_var]] ~ share_virtual + share_hybrid
                                    + state_id + commute_zone
                                    + int_virtual_black + int_hybrid_black
                                    + int_virtual_hisp + int_hybrid_hisp
                                    + int_virtual_lunch + int_hybrid_lunch
                                    , data = data, se_type = "stata")
}
```

Export results

```
# Create list
models <- list(
  "(1)" = results[["math_reg1"]],
  "(2)" = results[["math_reg2"]],
  "(3)" = results[["math_reg3"]],
  "(4)" = results[["ela_reg1"]],
  "(5)" = results[["ela_reg2"]],
  "(6)" = results[["ela_reg3"]],
  "(7)" = results[["avg_reg1"]],
  "(8)" = results[["avg_reg2"]],
  "(9)" = results[["avg_reg3"]]
)

# Rename coefficient labels
renamed_labels <- c(
  "share_virtual" = "% virtual",
  "share_hybrid" = "% hybrid",
  "int_virtual_black" = "% Black x % virtual",
  "int_hybrid_black" = "% Black x % hybrid",
  "int_virtual_hisp" = "% Hispanic x % virtual",
```

```
  "int_hybrid_hisp" = "% Hispanic x % hybrid",
  "int_virtual_lunch" = "% Lunch x % virtual",
  "int_hybrid_lunch" = "% Lunch x % hybrid"
)

# Extra rows
extra_rows <- tribble(
  ~name, ~`(1)`, ~`(2)`, ~`(3)`, ~`(4)`, ~`(5)`, ~`(6)`, ~`(7)`, ~`(8)`, ~`(9)`,
  "Control: State", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes",
  "Control: County", "No", "Yes", "No", "No", "Yes", "No", "No", "Yes", "No",
  "Control: Commuting Zone", "No", "No", "Yes", "No", "No", "Yes", "No", "No", "Yes"
)

# Export table
modelsummary(models,
             coef_map = renamed_labels,
             statistic = "std.error",
             stars = TRUE,
             add_rows = extra_rows,
             output = "main_regressions.docx",
             fmt = 3,
             gof_map = c("nobs", "r.squared")
)
```