PS 1: My amazing solutions

2023-09-15

1. Tenn STAR, an experiment

a) Looking at the standard errors

The standard errors did/did not change after accounting for clustering....

b) Extend your model

```
# some code goes here!
```

Small class size did/did not have an impact on test scores....

c) Further control for classroom characteristics

```
# some code goes here!
```

Explain what happened when you included fixed effects for classroom...

d) Testing a hypothesis about teachers' years of experience and small class sizes

```
# some code goes here!
```

2. Aggregation practice (coding)

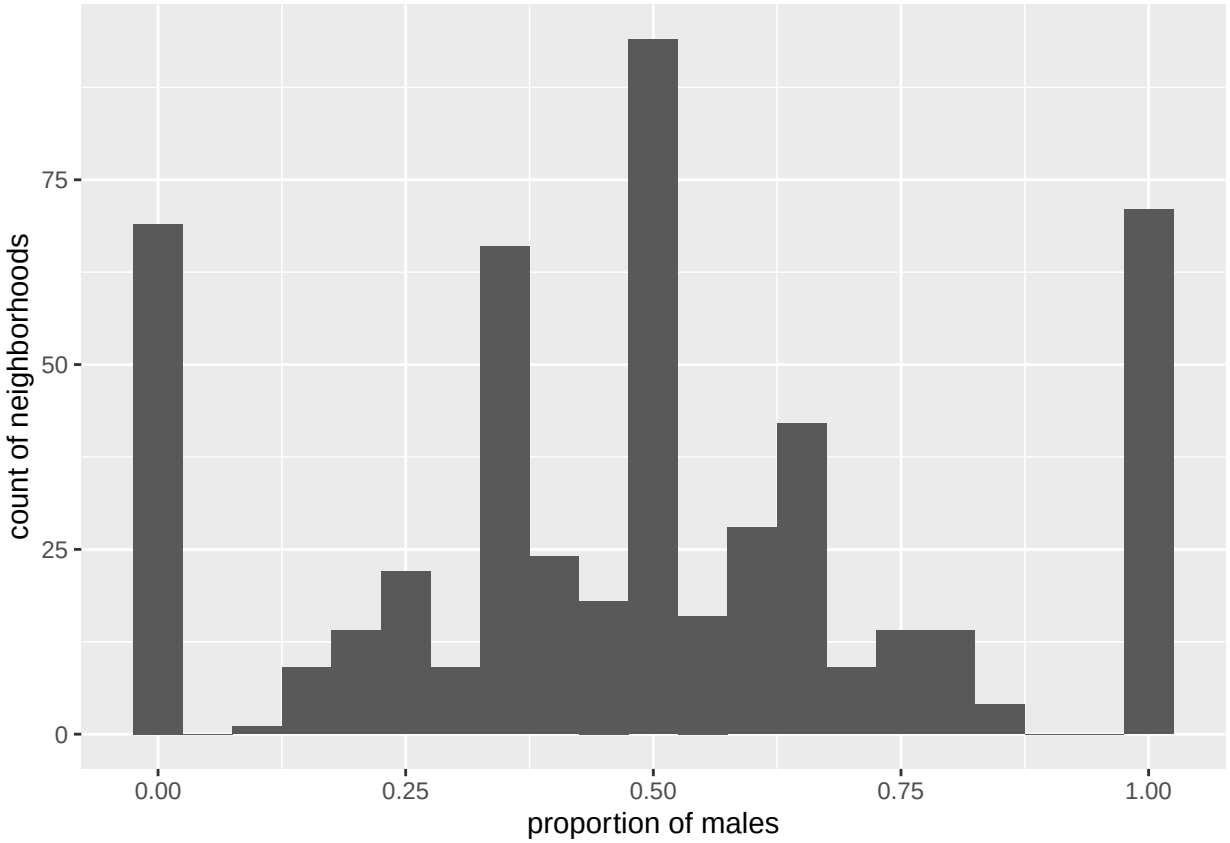
a) Make a summarized dataframe with the average attainment, the proportion male, and the number of observations for each neighborhood in the dataset.

```
# load the data and summarize it
dat <- read.dta13("neighborhood.dta")

# summarize the data
dat_sum <- dat %>%
  group_by(neighid) %>%
  summarise(
    mean_attain = mean(attain),
    prop_male = mean(male),
    obs = n())

# add a variable for deprivation
dat_sum$deprive <- dat$deprive[match(dat_sum$neighid, dat$neighid)]

head(dat_sum)
```

```
## # A tibble: 6 x 5
##   neighid mean_attain prop_male   obs  deprive
##     <int>       <dbl>     <dbl> <int>    <dbl>
## 1      26      0.188      0.4       5 -0.551
## 2      27     -0.132      0         1  0.147
## 3      29     -0.300      0.444     9 -0.0830
## 4      30      0.693      1         2 -0.274
## 5      31      0.544      1         2  0.00700
## 6      32      0.0563     0.5       6 -0.195
```

b) Make a histogram of the proportion male. Is this what you would expect? Is anything odd with this histogram?

```
# make a histogram of the prop_male variable
ggplot(dat_sum, aes(x = prop_male)) +
  geom_histogram(binwidth = 0.05) +
  labs(x = "proportion of males", y = "count of neighborhoods")
```



This histogram is odd because there are many neighborhoods that have 0 males and some neighborhoods only have males. This is not what we would expect.

c) calculate the average, standard deviation, and range of the number of observations in a neighborhood. What does this tell you about (b)?

```
# calculate the average, standard deviation, and range of the number of observations in a neighborhood
mean(dat_sum$obs)
```

```
## [1] 4.408397
```

```
sd(dat_sum$obs)
```

```
## [1] 2.785033
```

```
range(dat_sum$obs)
```

```
## [1]  1 16
```

```r
# install.packages("gtsummary")
library(gtsummary)

# Present the results in a table
dat_sum %>%
  select(obs) %>%
  tbl_summary(
    type = all_continuous() ~ "continuous2",
    statistic = all_continuous() ~ c("{min}, {max}", "{mean} ({sd})"),
    missing = "no",
    digits = all_continuous() ~ 4
  )
```

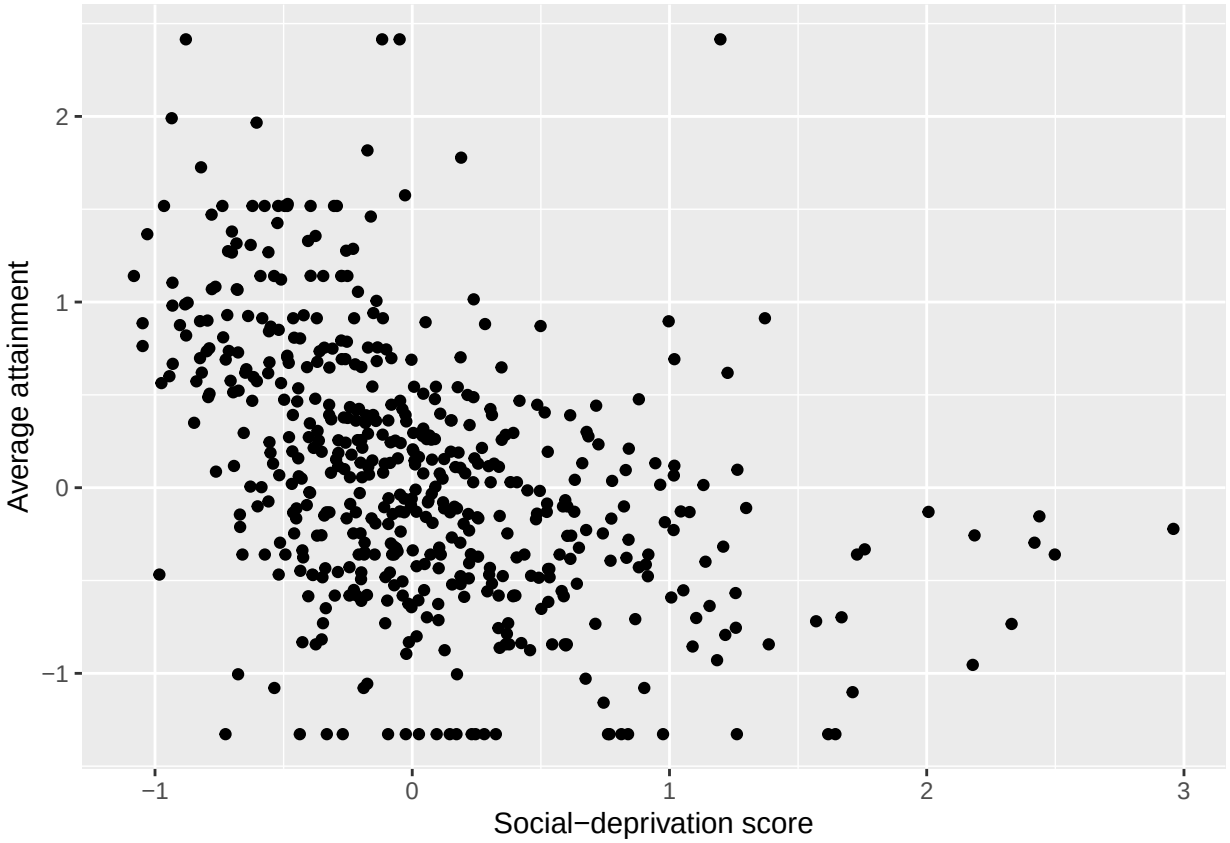
```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**N = 524**
obs	
Range	1.0000, 16.0000
Mean (SD)	4.4084 (2.7850)

We see that the average number of observations is 4.4, the standard deviation is 2.8, and the range is 1 to 16. This tells us that there are some neighborhoods that only have 1 observation. This is why we see the odd histogram in (b).

d) Scatterplot of the social-deprivation score (deprive) and average attainment. What can you learn, if anything, from this plot? What things are distorted, and how?

```r
# make a scatterplot of the social-deprivation score (deprive) and average attainment
ggplot(dat_sum, aes(x = deprive, y = mean_attain)) +
  geom_point() +
  labs(x = "Social-deprivation score", y = "Average attainment")
```




```
  mean(dat_sum$deprive) # average social-deprivation score is 0.37
```

```
## [1] 0.03714885
```

```
  range(dat_sum$deprive) # range of social-deprivation score is -1.082  2.959
```

```
## [1] -1.082  2.959
```

```
  mean(dat_sum$mean_attain) # average attainment of the neighborhood averages is 0.08
```

```
## [1] 0.08087043
```

```
  range(dat_sum$mean_attain) # average attainment of the neighborhood averages is -1.3276, 2.4151
```

```
## [1] -1.3276  2.4151
```

Neighborhoods with higher social deprivation score tend to have lower average attainment. However, there are a few neighborhoods with high social-deprivation scores (above 1) that have high average attainment. Neighborhoods with the lowest average attainment of -1.3276 have a wide range of social-deprivation scores, which is odd. The plot is distorted partially because there are some neighborhoods that only have 1 observation, so the average attainment is the same as the attainment of that observation.