Lab4_2

2023-02-24

Question 1 It's important to randomly assign both students and teachers to determine any causal effect of class size to prevent selection bias influencing potential given coefficients. It could be that really good teachers could improve test scores, just as really motivated students could improve their test scores as well, confounding the causal effect of class size on test scores.

```
#Question 2: Average class size comparison between small vs. regular kindergarten classes
#Group data set by whether class size was categorized as small or not then summarize mean class size fo

class_size <- star |>
  group_by(small) |>
  summarise(mean = mean(class_size))
class_size
```

```
## # A tibble: 2 x 2
##   small  mean
##   <dbl> <dbl>
## 1     0  22.5
## 2     1  15.1
```

The average class size for small classes was roughly 15 students per class; for regular classes, about 22-23 students per class.

```
#Question 3: Converting different test scores into a standardized unit

#3A: Creating standardized test scores
#creating control variables
control <- star |>
          filter(small == 0) |>
          select(read, math, listen, wordskill)

#Math
math_ctrl_mean <- mean(control$math, na.rm = TRUE)
math_ctrl_sd <- sd(control$math, na.rm = TRUE)

math_stdrd <- (star$math - math_ctrl_mean)/math_ctrl_sd

#Reading
read_ctrl_mean <- mean(control$read, na.rm = TRUE)
read_ctrl_sd <- sd(control$read, na.rm = TRUE)

read_stdrd <- (star$read - read_ctrl_mean)/read_ctrl_sd

#Wordskill
wordskill_ctrl_mean <- mean(control$wordskill, na.rm = TRUE)
wordskill_ctrl_sd <- sd(control$wordskill, na.rm = TRUE)

wordskill_stdrd <- (star$wordskill - wordskill_ctrl_mean)/wordskill_ctrl_sd
```

```r
#Listening
listen_ctrl_mean <- mean(control$listen, na.rm = TRUE)
listen_ctrl_sd <- sd(control$listen, na.rm = TRUE)

listen_stdrd <- (star$listen - listen_ctrl_mean)/listen_ctrl_sd

#Reporting summary statistics for four standardized test score variables

summary(math_stdrd)
```
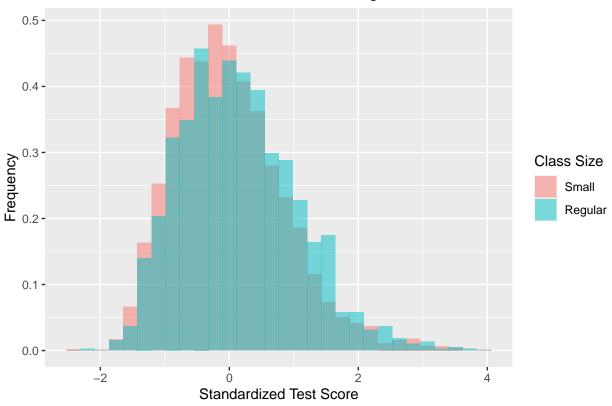
```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -3.49539 -0.62800  0.01395  0.05194  0.63451  3.05253
```

```r
sd(math_stdrd)
```

```
## [1] 1.021684
```

```r
summary(read_stdrd)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -2.47023 -0.68020 -0.07287  0.05089  0.56643  6.12831
```

```r
sd(read_stdrd)
```

```
## [1] 1.0152
```

```r
summary(wordskill_stdrd)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -2.79641 -0.76367 -0.15934  0.04772  0.69222  4.40060
```

```r
sd(wordskill_stdrd)
```

```
## [1] 1.012056
```

```r
summary(listen_stdrd)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -4.22024 -0.61906 -0.01382  0.03368  0.71247  4.07155
```

```r
sd(listen_stdrd)
```

```
## [1] 1.00281
```

```r
#Creating standardized test score across all scores

sat_index <- (math_stdrd + read_stdrd + wordskill_stdrd + listen_stdrd)/4
summary(sat_index)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -2.43578 -0.59900 -0.04005  0.04606  0.56693  3.92905
```

```r
#3C: Histogram of sat_index for small and regular kindergarten classes

histosat <- star |>
ggplot(aes(x = sat_index,
          fill = factor(small, labels=c("Small", "Regular")),
          y = ..density..)) +
          geom_histogram(alpha = 0.5, position = "identity") +
          labs(x = "Standardized Test Score",
```

```
                y = "Frequency",
                fill = "Class Size",
                title = "Standardized SAT Scores for Small vs. Regular Classrooms")
histosat
```

```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```


Standardized SAT Scores for Small vs. Regular Classrooms

```
ggsave("histostat.png")
```

```
## Saving 6.5 x 4.5 in image
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

There is some overlap between the standardized test scores for small and regular classrooms.