

empirical_proj_2

Lindsey Greenhill

4/9/2022

Question 1

The fundamental problem of causal inference is that you cannot see the outcome for a child who both was in a small class and not in a small class. In other words, it is impossible to measure the dual outcomes when a child both is in a small class and not in a small class, because the child can only either be in a small class or not in a small class.

Question 2

The following variables have missing values:

- towncode: 318 missing
- math: 58 missing
- verb: 466 missing
- ses_index: 318 missing
- boy: 1037 missing

```
##      student_id      towncode      schlcode      class_id
## Min.      :    1  Min.      : 166  Min.      :11005  Min.      :110051
## 1st Qu.:13600  1st Qu.:26104  1st Qu.:31049  1st Qu.:310491
## Median :27412  Median :62000  Median :41284  Median :412841
## Mean   :27395  Mean   :50434  Mean   :39776  Mean   :397762
## 3rd Qu.:41158  3rd Qu.:79004  3rd Qu.:51247  3rd Qu.:512471
## Max.   :54860  Max.   :98004  Max.   :61365  Max.   :613651
##                NA's      :318
## school_enrollment  class_size      math      verb
## Min.      : 9.00  Min.      : 7.00  Min.      : 1.0  Min.      : 1.00
## 1st Qu.: 57.00  1st Qu.:28.00  1st Qu.:23.0  1st Qu.:25.00
## Median : 75.00  Median :32.00  Median :27.0  Median :27.00
## Mean   : 80.41  Mean   :31.58  Mean   :25.3  Mean   :25.97
## 3rd Qu.:105.00  3rd Qu.:36.00  3rd Qu.:29.0  3rd Qu.:29.00
## Max.   :188.00  Max.   :44.00  Max.   :30.0  Max.   :30.00
##                NA's      :58  NA's      :466
##      ses_index      boy      born_isr      religious
## Min.      : 0.00  Min.      :0.0000  Min.      :0.0000  Min.      :0.0000
## 1st Qu.: 4.00  1st Qu.:0.0000  1st Qu.:1.0000  1st Qu.:0.0000
## Median : 9.00  Median :1.0000  Median :1.0000  Median :0.0000
## Mean   :13.17  Mean   :0.5078  Mean   :0.9608  Mean   :0.2242
## 3rd Qu.:18.00  3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:0.0000
## Max.   :76.00  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000
## NA's      :318  NA's      :1037
```

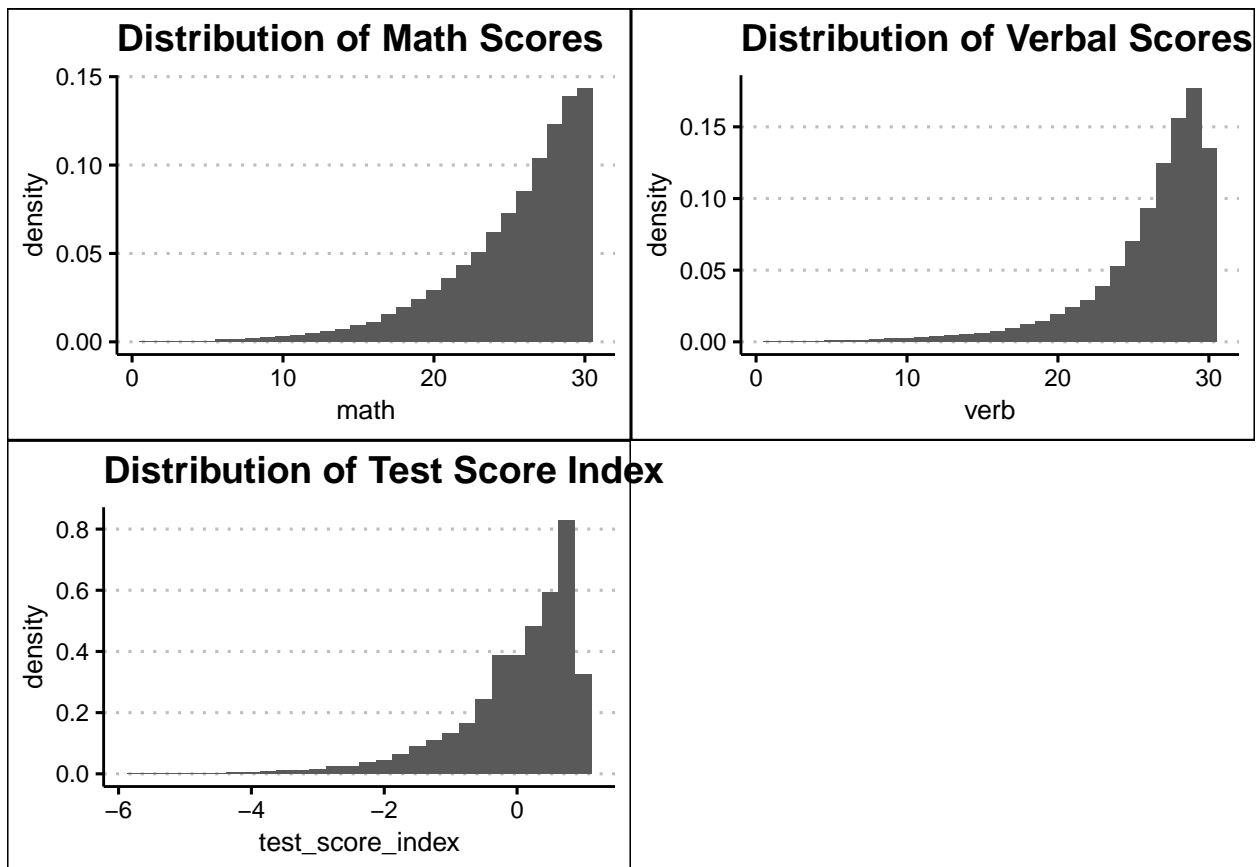
Questions 3 and 4

Table 1: Summary Statistics

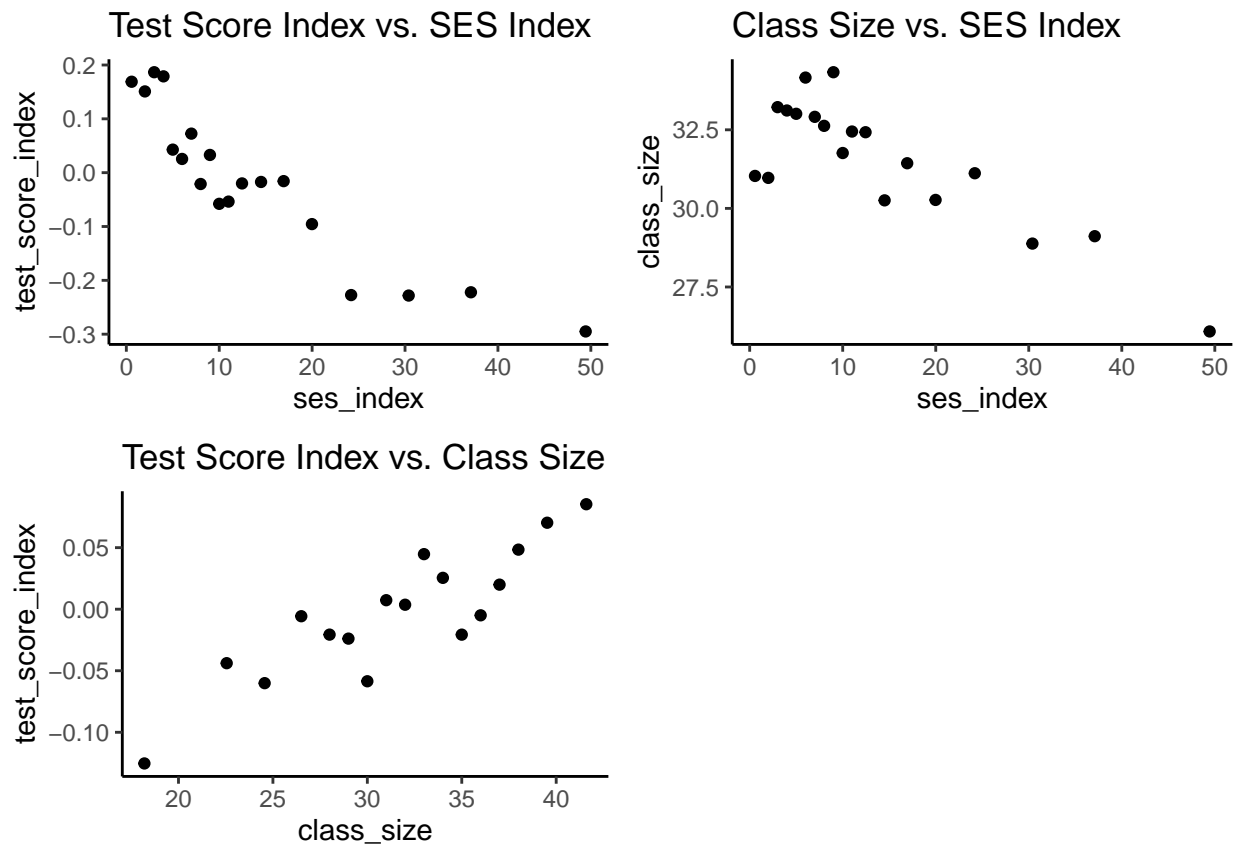
Variable	Mean	Sd	Min	Max
math	25.3	4.527	1	30
verb	26	4.279	1	30
test_score_index	0	0.903	-5.835	1.038

Question 5

All of the histograms appear to not be symmetrical and all of them have a long left tail.



Question 6



Question 7