

# STAR Dataset Analysis

Julieta Pappano

2024-11-23

## Introduction

The Star project was a large experiment conducted in Tennessee in the United States. Three treatments were assigned at the classroom level: small classes (13–17 students), regular classes (22–25 students), and regular classes with an aide who would work with the teacher. Krueger (1999) used the experimental data from STAR to analyze if class size mattered for learning.

The dataset contains a total of 5786 observations and 19 variables. The following table summarizes the 5 variables that are relevant for the analysis conducted in the following section

Table 1: Summary of STAR Dataset

boy	freelunch	white_asian	black	tchexper	small
Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. : 0.000	Min. :0.0000
1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.: 5.000	1st Qu.:0.0000
Median :1.0000	Median :0.0000	Median :1.0000	Median :0.0000	Median : 9.000	Median :0.0000
Mean :0.5135	Mean :0.4817	Mean :0.6766	Mean :0.3209	Mean : 9.306	Mean :0.3004
3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:13.000	3rd Qu.:1.0000
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :27.000	Max. :1.0000
NA	NA	NA	NA	NA's :20	NA

## Analyzing Small Classes

The following section focuses on students who took part of small classes ( $\text{small} = 1$ ), compared to those who were part of regular-sized classes ( $\text{small} = 0$ ). The mean test score of those in a small class was 931.9419 and the mean score of those in a regular with or without aid was 918.2013. These averages were calculated with the following code:

```
## [1] 931.9419
```

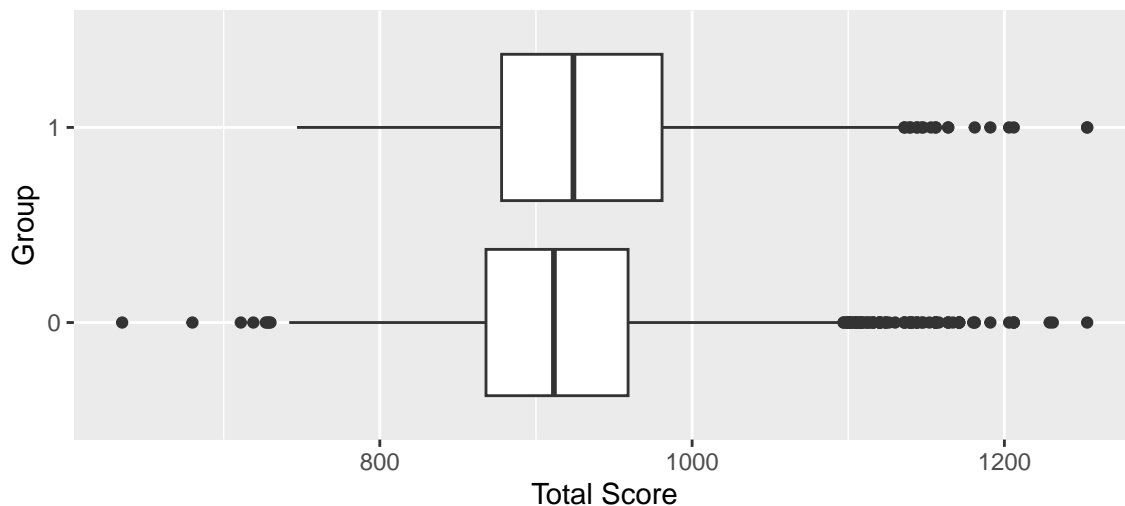
```
## [1] 918.2013
```

```
##   Group.1      x
## 1      0 918.2013
## 2      1 931.9419
```

As it can be seen, smaller groups demonstrated having better results. Figure 1 shows the distribution of the data and it can be seen that it is shifted to the right, which can be confirmed by the fact that the median is higher for small classes. Thus, it can be argued that higher grades are obtained. Also, regular-sized classes

present more outliers in both ends of the spectrum, however, the lower-end outliers disappear in the case of small classes. Overall, the graph shows that, when compared to those in regular classes, students in small classes obtain better grades.

Figure 1. Regular vs. Small Classes



## Covariate Analysis

The covariate variables found in the set are: gender, free launch, race, and teacher experience. Three methods are used to analyze whether treatment and control groups are similar and, therefore, randomization was effective. If there are any significant coefficients, then it means there is an imbalance in that variable, which signifies that treatment and control groups systematically differ on that factor.

1. Covariate balance table is generated using the vtble package, which shows the statistical significance of each variable:

Table 2: Covariate Balance

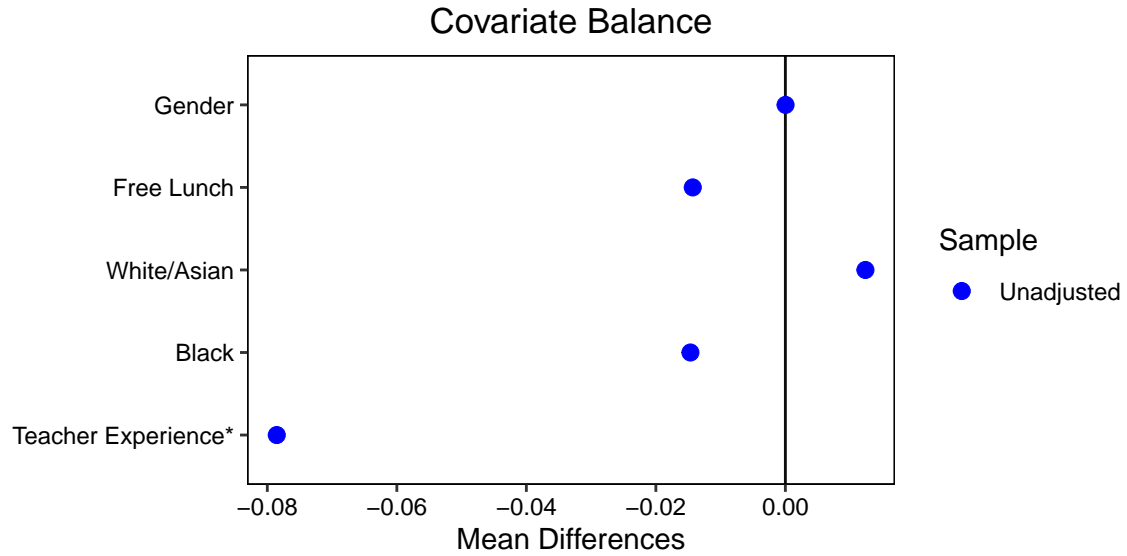
Class Size	0			1			Test
Variable	N	Mean	SD	N	Mean	SD	
Gender	4048	0.51	0.5	1738	0.51	0.5	F=0.022
Free Lunch	4048	0.49	0.5	1738	0.47	0.5	F=0.97
White/Asian	4048	0.67	0.47	1738	0.68	0.46	F=0.738
Black	4048	0.33	0.47	1738	0.31	0.46	F=1.066
Teacher Experience	4028	9.4	5.8	1738	9	5.7	F=7.244***

Statistical significance markers: \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

As the above table shows, the only significant difference between groups is teacher experience, so this variable should be incorporated as a control variable.

2. Covariate balance graph:

## Note: 's.d.denom' not specified; assuming "pooled".



### 3. Regression analysis

Similarly to the previous analysis, it shows that teacher experience is not balanced among groups.

Table 3: Covariate Analysis

	<i>Dependent variable:</i>
	Class Size
Gender	-0.0003 (0.012)
White/Asian	-0.193 (0.123)
Black	-0.208* (0.123)
Teacher Experience	-0.003*** (0.001)
Free Lunch	-0.008 (0.014)
Constant	0.530*** (0.123)
Observations	5,766
R <sup>2</sup>	0.002
Adjusted R <sup>2</sup>	0.001
Residual Std. Error	0.459 (df = 5760)
F Statistic	2.427** (df = 5; 5760)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

## Impact Analysis

In this section, the changes in the total scores are tested considering different scenarios, with different controls. As the table shows, even though there are some variations across models, the estimate of the treatment (*small*) remains fairly stable, representing an increase between 13.55 to 14.31 points in total test scores. As such, it could be argued that the treatment of having smaller class sizes is effective when it comes to improving performance.

Table 4: Covariate Analysis

	<i>Dependent variable:</i>				
	Test Score				
	(1)	(2)	(3)	(4)	(5)
Small (Treatment)	13.741*** (2.107)	14.307*** (2.099)	14.309*** (2.090)	13.832*** (2.057)	13.584*** (2.010)
Teacher Experience		1.467*** (0.167)	1.426*** (0.166)	1.179*** (0.165)	1.118*** (0.161)
Gender			-13.286*** (1.919)	-13.959*** (1.889)	-14.043*** (1.846)
White/Asian				28.336 (19.179)	27.267 (18.740)
Black				0.293 (19.217)	15.599 (18.801)
Free Lunch					-34.192*** (2.066)
Constant	918.201*** (1.155)	904.442*** (1.952)	911.645*** (2.205)	895.202*** (19.257)	908.142*** (18.832)
Observations	5,786	5,766	5,766	5,766	5,766
R <sup>2</sup>	0.007	0.020	0.028	0.060	0.102
Adjusted R <sup>2</sup>	0.007	0.020	0.028	0.059	0.101
Residual Std. Error	73.483	73.079	72.783	71.615	69.977
F Statistic	42.515***	59.767***	56.148***	73.081***	109.433***

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Race as a Determinant

As the following plots show, there is no relation between teacher experience and the total score of students, whether they are black, white, or Asian. The data points are highly dispersed and do not match the pattern of the teacher experience variable.

Figure 3. Black Students

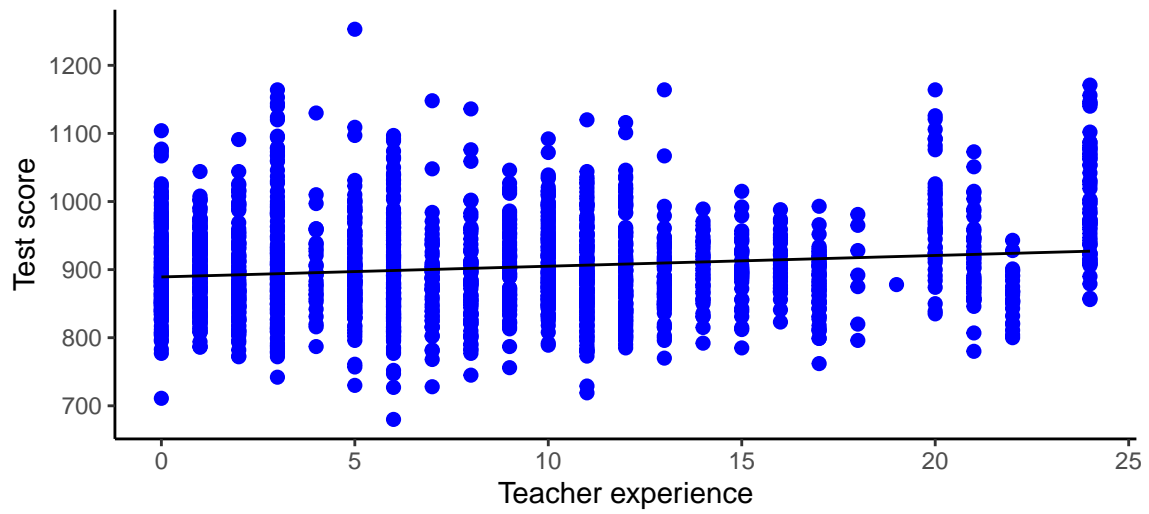
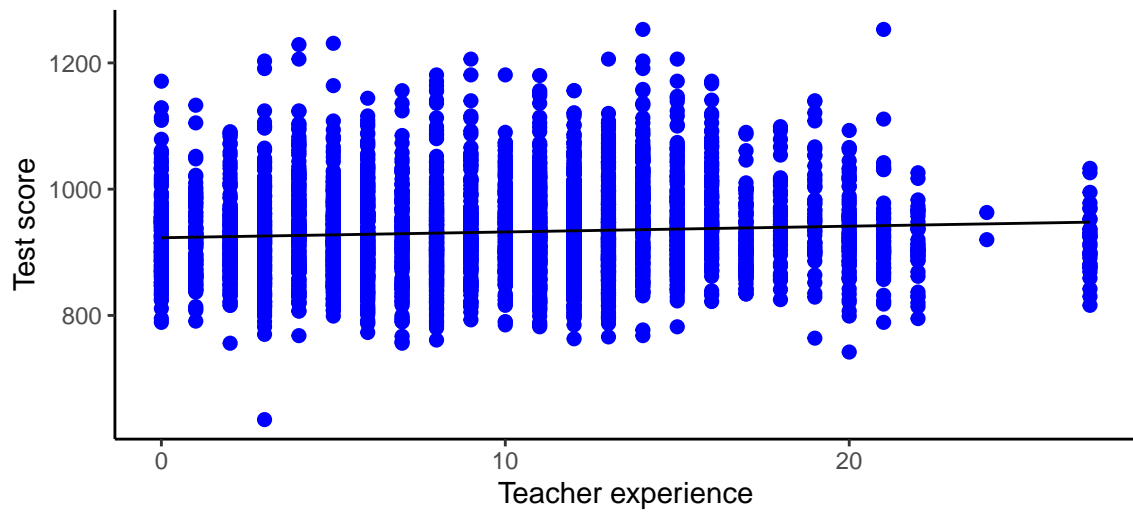


Figure 4. White or Asian Students



## References

Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*, 114 (2), 497–532. doi:10.1162/003355399556052