

An A/B testing study on different schooling methods based on fMRI data

Which schooling method is better ? Montessori Education or Traditional Education ?

Qingwen Wang

Study Description:

The study I am interested in is “An fMRI study of error monitoring in Montessori and traditionally-schooled children”. [<https://rossier.usc.edu/files/2020/07/fmri-study-error-monitoring.pdf>]. It discusses which kind of students learn better over time, the students who are focused on monitoring their own learning process, or the students who are focused on getting right answers. The unit in this study is 8-12 year-old students from Montessori schools and students educated in traditional schools. Students in the study were asked to solve math problems while an fMRI tracking brain activity. The researchers evaluates potential learning outcomes by both brain activity and students' quiz answer. Only Montessori students showed coherent changes in brain activity following errors, suggesting that they were engaging with the errors strategically to learn. Traditionally-schooled students, by contrast, showed coherent activity only after correct answers, and the activity pattern suggested that they were trying to memorize that event. Though both groups got the same number of problems right, the Montessori students skipped far fewer and got more wrong, making them learn the task more efficiently by the end. This study gives the conclusion that Montessori-schooled students will learn better over time.

Y^1 : the potential learning outcomes for students from a Montessori school

Y^0 : the potential learning outcomes for students from a traditional school

D : taking schooling method of Montessori pedagogy

Possible bias:

Family's economic status can be a factor causing selection bias. Since Montessori school usually has a higher tuition fee than the tradition public school so that family

with a worse economic status are less likely to send their children to a Montessori school. So this selection bias will cause the potential learning outcomes to be better, NATE to be higher. [NATE = ATE + Selection bias(positive in this study) + differential bias]. Besides this selection bias, there also could have other problems that may cause bias, such as: children's intelligence and anxiety, or home pedagogical environment, etc.

Simple simulation:

The original study evaluate both math quiz outcomes, and brain activity tracked by fMRI. To simplify, here I only simulate the math quiz outcomes. In the original study, there are total 32 participates from different family, and these children are educated in a Montessori school or a traditional school; for math quiz performance, the overall efficiency was higher in Montessori students than in traditionally-schooled students, $M_m = 3.26$, $M_t = 2.46$. Each student has a different family economic status, and family with higher economic status will have higher probability to send their children to a Montessori school.

So, here I set sample size 3200 (for a more intuitive graphing of bias and ate); ATE as 0.8; family economic status as a variable causing the selection bias; classify different 'econ_status' into different groups: 1)<25%: low, 2)25%~75% :medium, and 3) >75%: high.

Y_{i0} =total CVD events in the control group (no exercise)

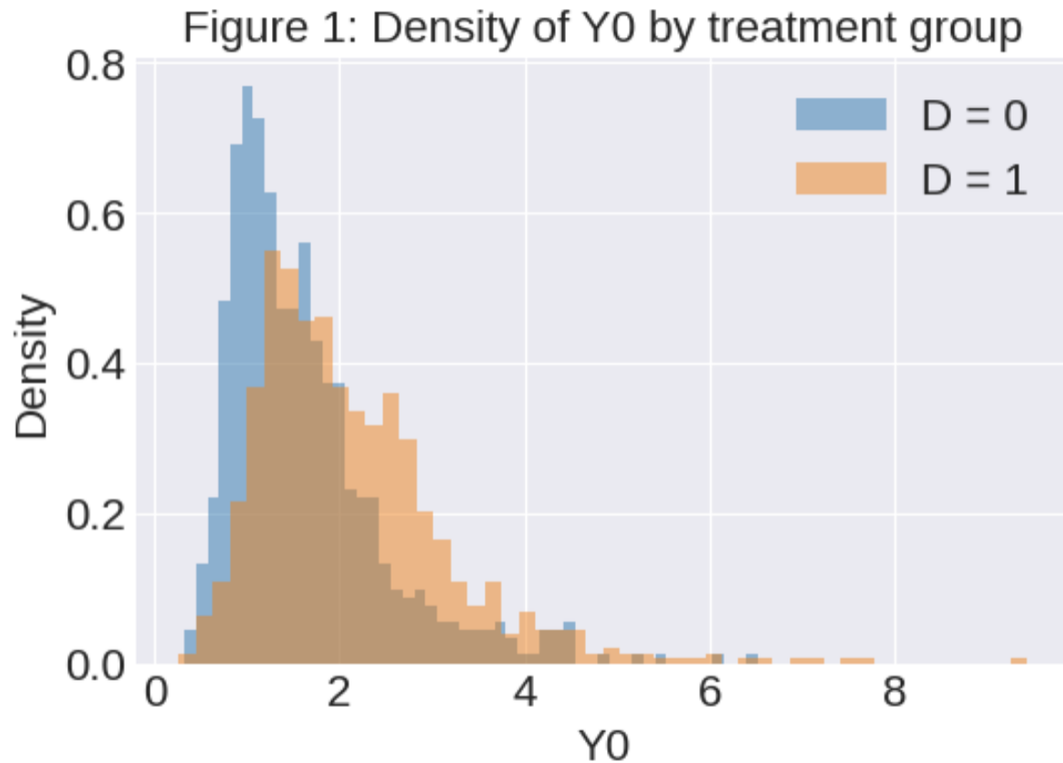
Y_{i1} = total CVD events in the treatment group (do exercise)

	D	Y0	Y1	Y	y1-y0	econ_status	econ_status_label
0	0	2.229693	3.029693	2.229693	0.8	0.276648	medium
1	0	0.737154	1.537154	0.737154	0.8	-0.724951	low
2	1	2.658408	3.458408	3.458408	0.8	0.501950	high
3	0	1.600921	2.400921	1.600921	0.8	0.029863	medium
4	0	1.222128	2.022128	1.222128	0.8	-0.283706	medium
...
1595	1	1.726591	2.526591	2.526591	0.8	0.065007	medium
1596	1	2.061011	2.861011	2.861011	0.8	0.162726	medium
1597	1	2.477428	3.277428	3.277428	0.8	0.326658	high
1598	1	1.553691	2.353691	2.353691	0.8	-0.020610	medium
1599	1	1.897107	2.697107	2.697107	0.8	0.126847	medium

1600 rows x 7 columns

Simulate selection bias

In the picture, we can observe that the blue shade is not totally overlapped by the yellow shade, which means the density of Y^0 when $D=1$ is not same with the density of Y^0 when $D=0$; so that these two groups do not have the same probability to donate even there is no treatment, selection bias exists.



Also, we can find the selection bias from a mathematic way:

$$NATE = E[Y^1 | D = 1] - E[Y^0 | D = 0] = 1.31;$$

$$SelectionBias = E[Y^0 | D = 1] - E[Y^0 | D = 0] = 0.51;$$

$$ATE = 0.8;$$

$$Differentialbias = 0$$

So by using formula, we can get the the selection bias is 0.346, derived from the different economic status.

Simulate an experiment

Perfect stratification is a good method to remove bias. This means that we will stratify the data into several subgroups (strata). "Perfect" means that we can put each person in the experiment into exactly the right group such that all selection bias is accounted for. With the data split into several strata, we will get a separate estimate of the ATE within each one. Then we will average together the estimates of each of the strata. Besides perfect stratification, regression can also be a convenient tool to implement a stratification. In this case, we can get the correct result by including a dummy variable for each economic status label.

a) Regression (y ~ d) result without stratification:

ATE = 1.3

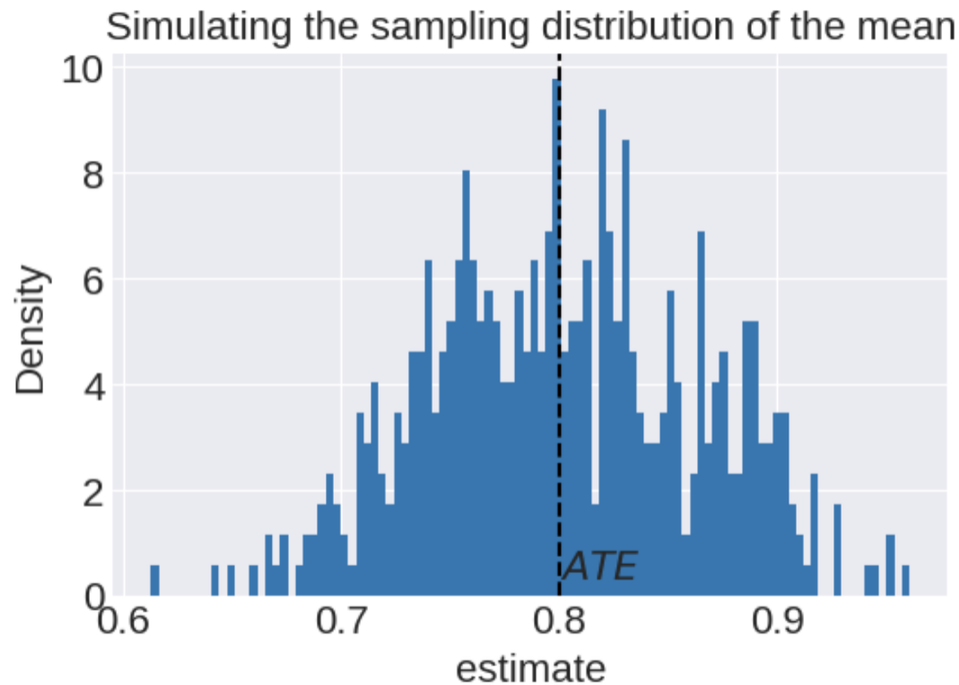
	coef	std err	z	P> z	[0.025	0.975]
Intercept	1.6158	0.032	51.105	0.000	1.554	1.678
D	1.3162	0.048	27.441	0.000	1.222	1.410
Omnibus:	615.406		Durbin-Watson:		2.024	
Prob(Omnibus):	0.000		Jarque-Bera (JB):		2933.042	
Skew:	1.774		Prob(JB):		0.00	
Kurtosis:	8.605		Cond. No.		2.72	

b) Regression (y ~ d + econ_status_label) result with stratification:

ATE = 0.8

	coef	std err	z	P> z	[0.025	0.975]
Intercept	3.1371	0.049	63.804	0.000	3.041	3.233
C(econ_status_label)[T.low]	-2.2046	0.048	-45.642	0.000	-2.299	-2.110
C(econ_status_label)[T.medium]	-1.4712	0.048	-30.721	0.000	-1.565	-1.377
D	0.7839	0.028	27.973	0.000	0.729	0.839
Omnibus:	1088.429		Durbin-Watson:		2.026	
Prob(Omnibus):	0.000		Jarque-Bera (JB):		25736.941	
Skew:	2.833		Prob(JB):		0.00	
Kurtosis:	21.814		Cond. No.		5.35	

Thus, we identify our ATE by removing bias through regression method.



Simulating hypothesis tests

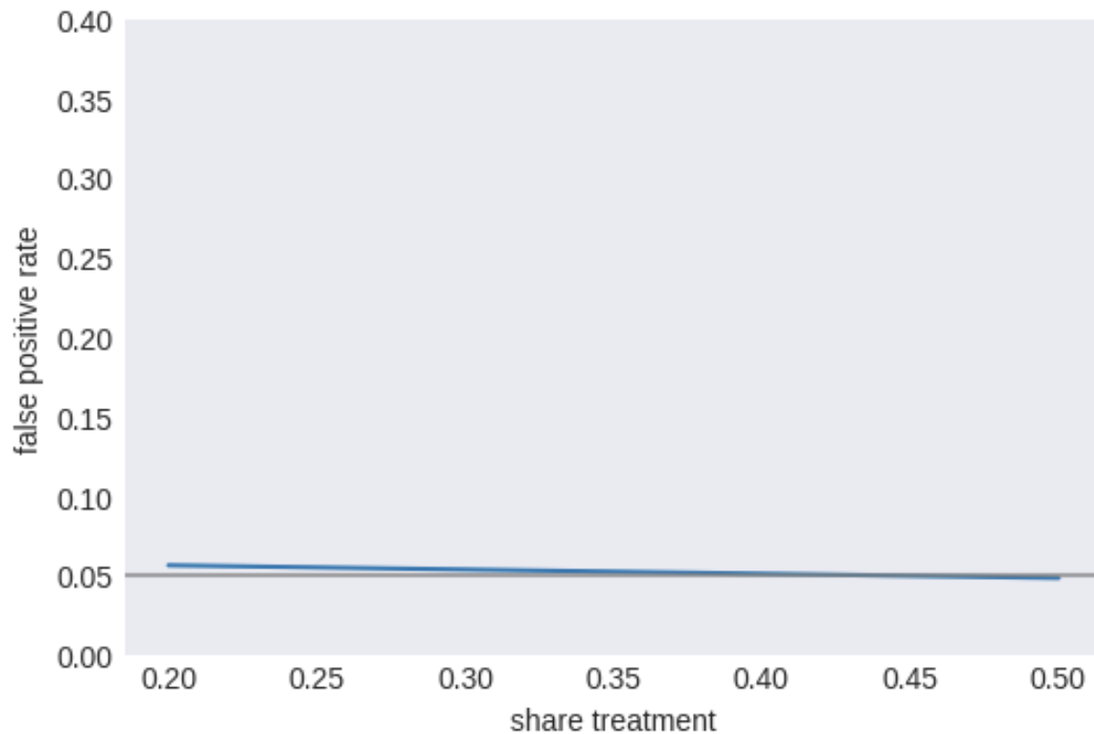
Two sampling method:

- 1) split treatment and control share by 50/50;
- 2) split treatment and control share by 20/80;

From the table below, both 'p_reject' are around 0.05(alpha is set to be 0.05), so the false positive rate is correctly controlled.

simulation_results

	diff_mean	diff_std	t_mean	t_std	alpha	p_reject	B	N	N_treatment	share_treatment	expectation_Y0	sigma	ATE
0	-0.017185	0.245712	-0.068882	0.997174	0.05	0.056	500	100.0	20.0	0.2	0.5	1	0.0
1	-0.010480	0.200923	-0.049364	1.009630	0.05	0.048	500	100.0	50.0	0.5	0.5	1	0.0



Compare two simulation results:

- 1) treatment_share = 0.2 , N = 100, 500, 1000, power increase from 0.224 to 0.967;
- 2) treatment_share = 0.5 , N = 100, 500, 1000, power increase from 0.348 to 0.998;

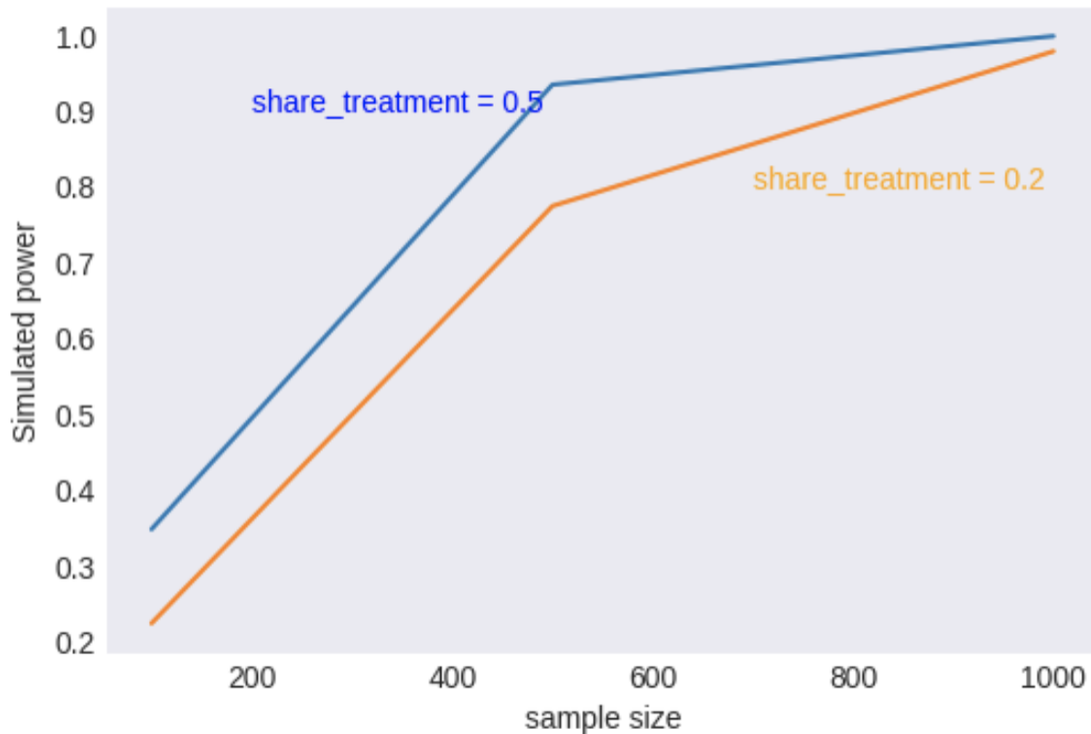
And from the plot, we can observe that when treatment share = 0.5, the power is always much greater. So 50/50 is a good sampling to method to both control false positive rate and increase power.

[71] ts02

	diff_mean	diff_std	t_mean	t_std	alpha	p_reject	B	N	N_treatment	share_treatment	expectation_Y0
0	0.295173	0.248321	1.177067	0.988166	0.05	0.224	500	100.0	20.0	0.2	0.5
1	0.305356	0.110763	2.740153	1.004796	0.05	0.774	500	500.0	100.0	0.2	0.5
2	0.306107	0.075509	3.882549	0.967261	0.05	0.978	500	1000.0	200.0	0.2	0.5

ts05

	diff_mean	diff_std	t_mean	t_std	alpha	p_reject	B	N	N_treatment	share_treatment	expectation_Y0
0	0.306980	0.195423	1.546582	0.984097	0.05	0.348	500	100.0	50.0	0.5	0.5
1	0.299047	0.086878	3.349471	0.980499	0.05	0.934	500	500.0	250.0	0.5	0.5
2	0.304737	0.065517	4.822116	1.040998	0.05	0.998	500	1000.0	500.0	0.5	0.5



Compare two sampling methods:

Method 1, splitting the treatment and control for 20/80.

The advantage is that we can have less participants in the treatment group, that we can save experiment cost. And if the treatment effect is uncertain, maybe negative, we can lower our risk as much as possible. The disadvantage is that, our experiment will play less power so that we will have less confidence in our results if there is truly treatment effect.

Method 2, splitting the treatment and control for 50/50.

The advantage is that we can get higher power in experiment, and we can be more confidence to make a final decision. However, the cost is that we need more people to take treatment, which leads to higher cost.

Power calculation

Power calculation result:

Rejection rate from power calculator when treatment share is 50/50= 0.9173801660338057

So the result matches with our previous experiment result.

Four combination of alpha and power: alpha = 0.05 or 0.1, power = 0.8 or 0.9.

Usually, the power can increase at the cost of the higher false positive rate. Each combination has its own advantage and disadvantage. To take which combination depends on if you care higher power more or lower false positive rate. Also, how much sample size you can get can also determines which combination is more feasible for your experiment.

If we take ATE = 0.8, alpha = 0.05, power = 0.8, we can choose sample size N as 52(26 in each group) that our power can reach 0.8.

When we want to increase power or decrease false positive rate we may need to increase sample size.

Run the experiment

Estimate of ATE = 0.78391

OLS Regression Results						
=====						
Dep. Variable:	Conversion	R-squared:	0.753			
Model:	OLS	Adj. R-squared:	0.753			
Method:	Least Squares	F-statistic:	1968.			
Date:	Mon, 19 Oct 2020	Prob (F-statistic):	0.00			
Time:	17:23:57	Log-Likelihood:	-1404.7			
No. Observations:	1600	AIC:	2817.			
Df Residuals:	1596	BIC:	2839.			
Df Model:	3					
Covariance Type:	HC1					
=====						
	coef	std err	z	P> z	[0.025	0.975]

Intercept	3.1371	0.049	63.804	0.000	3.041	3.233
C(econ_status_label)[T.low]	-2.2046	0.048	-45.642	0.000	-2.299	-2.110
C(econ_status_label)[T.medium]	-1.4712	0.048	-30.721	0.000	-1.565	-1.377
D	0.7839	0.028	27.973	0.000	0.729	0.839
=====						
Omnibus:	1088.429	Durbin-Watson:	2.026			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	25736.941			
Skew:	2.833	Prob(JB):	0.00			
Kurtosis:	21.814	Cond. No.	5.35			

From the regression result, the coefficient of 'D' is 0.78 which is near to ATE = 0.8 . The standard error is 0.028, which means the variation is small. P-value is 0, it is significantly less than alpha = 0.05, we can be confident to say that there is treatment effect in our experiment. And the confidence intervals is 95% which

indicates that the true treatment effect lies have 95% probability of lying between 0.729 and 0.839.

We do this experiment to verify whether Montessori students truly performance better in academic than traditionally-schooled students. So we design this experiment by adding possible covariance into it to check if bias really exists, does Montessori teaching method will still have such high effect? Thus, we do regression on different economic status(low, medium, and high). From the regression result, we can see that when family economic status takes accounts in the study, p-value is less than 0.05, the ATE is 0.78, a little bit lower than 0.8, so we can reject the null, and conclude that there is treatment effect.

My experiment so lucky that it is a true positive one, since we correctly reject the null hypothesis that no effect exist, but there is actually treatment effect with ATE very close to 0.8. But for this study, the possible flaw is that the sample size is too small, if we can have more participates in, the result will be more reliable.

Rerun the experiment with ATE = 0

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Conversion    R-squared:                0.665
Model:                  OLS          Adj. R-squared:           0.664
Method:                 Least Squares  F-statistic:             1176.
Date:                   Mon, 19 Oct 2020  Prob (F-statistic):       0.00
Time:                   21:31:56      Log-Likelihood:          -1404.7
No. Observations:      1600          AIC:                    2817.
Df Residuals:          1596          BIC:                    2839.
Df Model:               3
Covariance Type:       HC1
=====
                        coef      std err          z      P>|z|      [0.025      0.975]
-----
Intercept              3.1371      0.049       63.804     0.000       3.041       3.233
C(econ_status_label)[T.low] -2.2046      0.048     -45.642     0.000     -2.299     -2.110
C(econ_status_label)[T.medium] -1.4712      0.048     -30.721     0.000     -1.565     -1.377
D                      -0.0161      0.028     -0.574     0.566     -0.071      0.039
=====
Omnibus:               1088.429    Durbin-Watson:           2.026
Prob(Omnibus):         0.000      Jarque-Bera (JB):        25736.941
Skew:                  2.833      Prob(JB):                 0.00
Kurtosis:              21.814      Cond. No.                 5.35
=====

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC1)

```

From the regression result, the coefficient of 'D' is 0.01 which is near to ATE = 0 . The standard error is 0.028, which means the variation is small. P-value is 0.566, it is larger alpha = 0.05, so that we cannot reject the null hypothesis which states that there is no treatment effect in our experiment. And the confidence intervals is 95%

which indicates that the true treatment effect lies have 95% probability of lying between -0.071 and 0.039. The conclusion is that we accept the null hypothesis that there is no treatment effect.

We do this experiment to verify whether Montessori students truly performance better in academic than traditionally-schooled students. So we design this experiment by adding possible covariance into it to check if bias really exists, does Montessori teaching method will still have such high effect? Thus, we do regression on different economic status(low, medium, and high). From the regression result, p-value is larger than 0.05, and coefficient of 'D' is close to zero, so that we cannot reject the null hypothesis.

This experiment is unlucky because we cannot reject the null hypothesis that no effect exist, while we want to find Montessori schooling method makes effect.