

中級ミクロデータサイエンス

Problem Set 2: Report

花澤 楓

学籍番号：2125242

2024 年 2 月 5 日

CONTENTS

1	はじめに	2
2	記述統計	2
2.1	全体の記述統計量の確認	2
2.2	2 学期制と 4 学期制の間で、目的変数である 4 年卒業率を比較	3
2.3	2 学期制を導入している大学の割合の推移を確認	4
2.4	4 年卒業率とその他の関係	4
2.5	semester 制の導入と共変量の関係を分析	6
3	回帰分析	7
3.1	シンプルに	7
3.2	回帰式の改善	8

1 はじめに

本レポートでは、大学を4年（及び6年）で卒業する割合が、2学期制と4学期制の間でどのように変化するかを分析した Bostwick, Fischer, and Lang (2022) の部分的なレプリケーションを目的としている。

2 記述統計

以下の手順でデータの基本的な性質を確認する。

1. 全体の記述統計量の確認
2. 2学期制と4学期制の間で記述統計量を比較
3. 2学期制と4学期制の間で、目的変数である4年卒業率を比較
4. 2学期制を導入している大学の割合の推移を確認

2.1 全体の記述統計量の確認

表1 記述統計量

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max
semester	2	0	0.9	0.3	0.0	1.0	1.0
quarter	2	0	0.1	0.3	0.0	0.0	1.0
totcohortsize	3374	0	1099.4	1183.0	18.0	653.0	8533.0
w_cohortsize	2292	0	599.5	628.7	0.0	367.0	4556.0
m_cohortsize	2079	0	499.9	571.1	0.0	282.0	4411.0
tot4yrgrads	1859	0	398.4	563.9	1.0	211.0	5243.0
m_4yrgrads	1065	0	154.8	238.0	0.0	73.0	2460.0
w_4yrgrads	1377	0	243.7	333.1	0.0	135.0	2847.0
women_gradrate_4yr	6632	0	0.4	0.2	0.0	0.4	1.0
gradrate4yr	920	0	0.4	0.2	0.0	0.3	1.0
men_gradrate_4yr	918	0	0.3	0.2	0.0	0.3	1.0
instatetuition	7031	0	11088.5	9181.6	0.0	8562.0	42600.0
costs	13874	0	192.1	400.5	0.0	63.6	5988.4
faculty	1799	0	340.0	382.6	18.0	188.0	3438.0
white_cohortsize	2805	0	777.1	914.8	0.0	444.0	6792.0
per_white_cohort	960	0	0.7	0.2	0.0	0.8	1.0
per_women_cohort	714	0	0.6	0.1	0.0	0.6	1.0

ここで、2学期制と4学期制の大学で比較すると、

表 2 Balance Table

	0 / Mean	0 / Std. Dev.	1 / Mean	1 / Std. Dev.	Diff. in Means	Std. Error
Four-year graduation rate	0.34	0.23	0.37	0.23	0.04***	0.01
Four-year men graduation rate	0.29	0.23	0.32	0.23	0.03***	0.01
Four-year women graduation rate	0.38	0.24	0.41	0.23	0.04***	0.01
Total expenditures (\$/million)	368.05	714.70	179.08	363.43	-188.97***	23.32
Cohort size	1500.97	1361.36	1069.73	1163.32	-431.24***	45.18

0 が 4 学期制、1 が 2 学期制を表す。また、6 年で卒業のデータがないため、Bostwick, V., Fischer, S., & Lang, M. (2022) Table1 と異なる部分がある。

2 学期制と 4 学期制の間で、卒業率、授業料、生徒の数のどれも統計的に有意に異なっていて、かつ 4 学期制の大学の方が卒業率が高いことがわかる。

2.2 2 学期制と 4 学期制の間で、目的変数である 4 年卒業率を比較

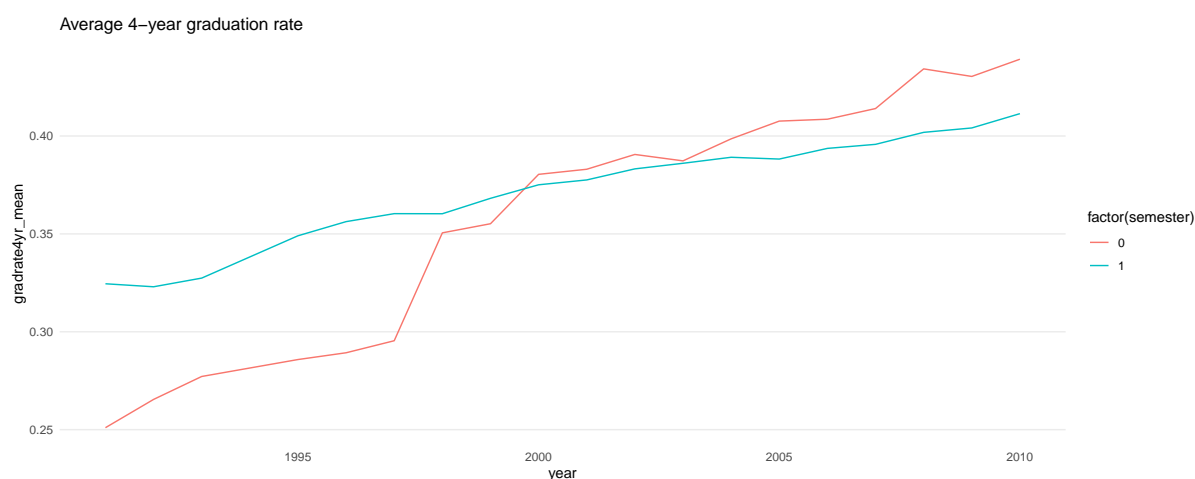


図 1 4 年卒業率：推移

2.3 2 学期制を導入している大学の割合の推移を確認

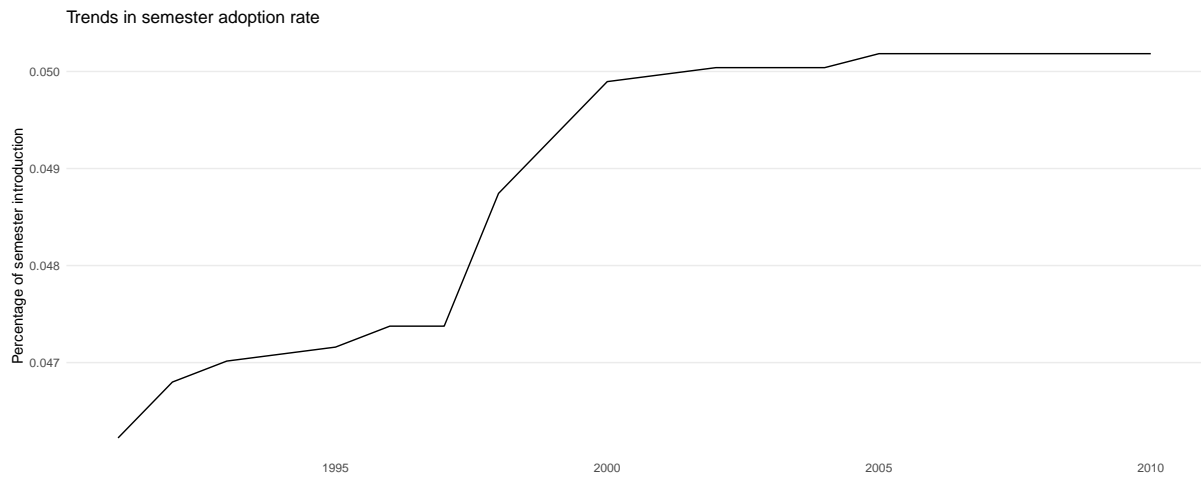


図2 2 学期制を導入している大学の割合の推移

2.4 4 年卒業率とその他の関係

以下の組み合わせについて、散布図を示す。

- 4 年卒業率と男子学生比率
- 4 年卒業率と女子学生比率
- 4 年卒業率と白人学生割合
- 4 年卒業率と年間運営コスト (costs)
- 4 年卒業率と学費 (instatetuition)

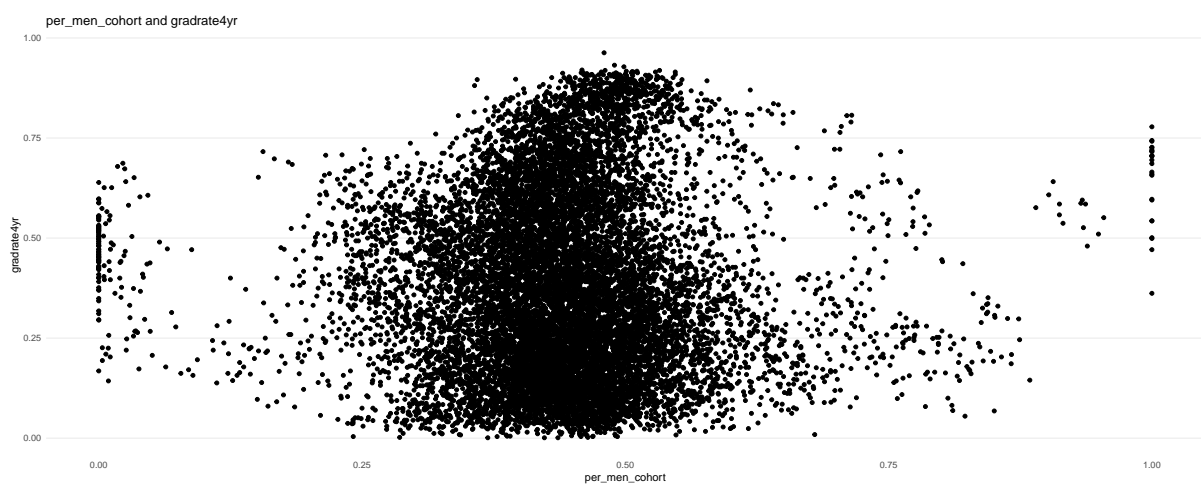


図3 4 年卒業率と男子学生比率

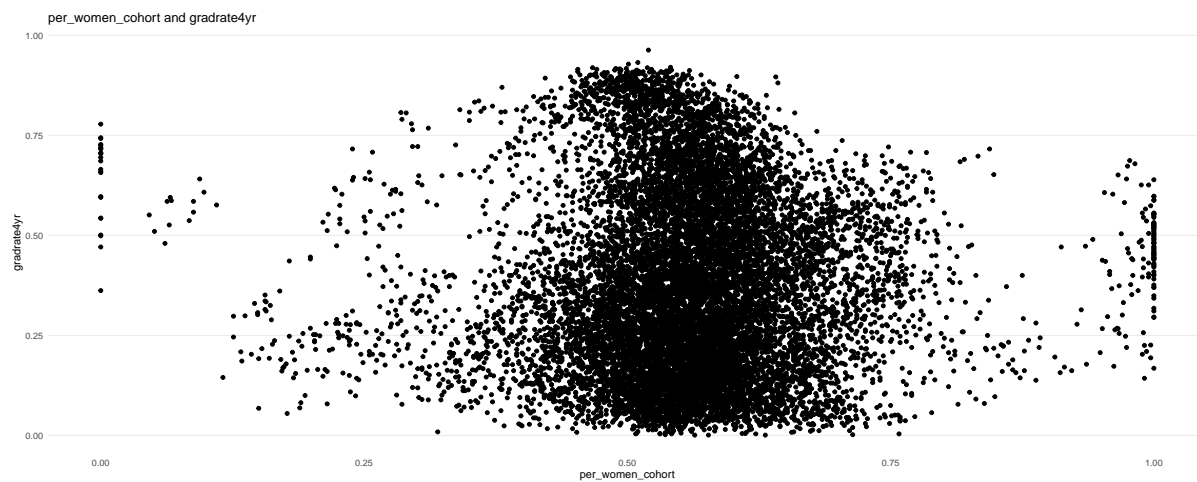


図4 4年卒業率と女子学生比率

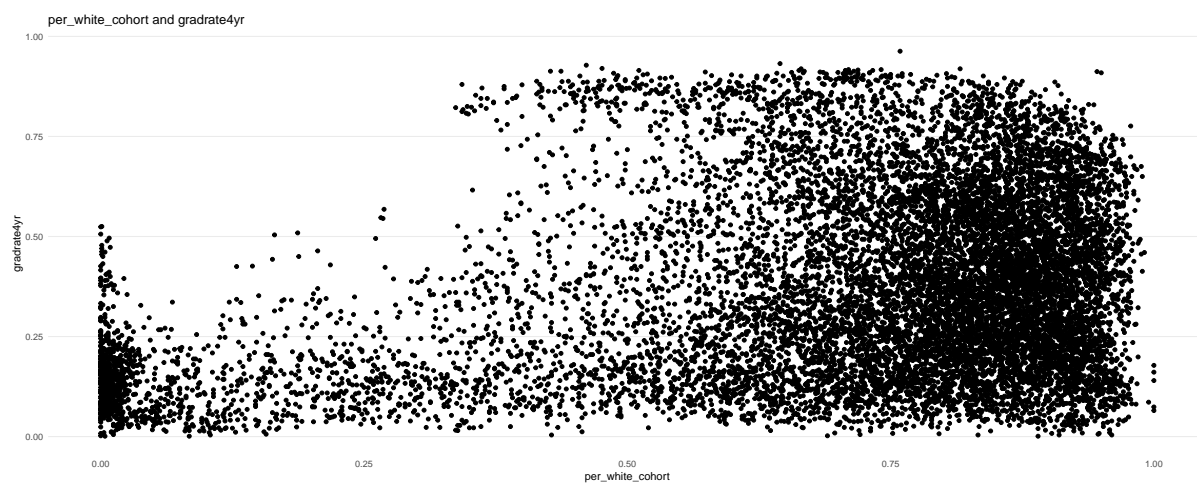


図5 4年卒業率と白人学生割合

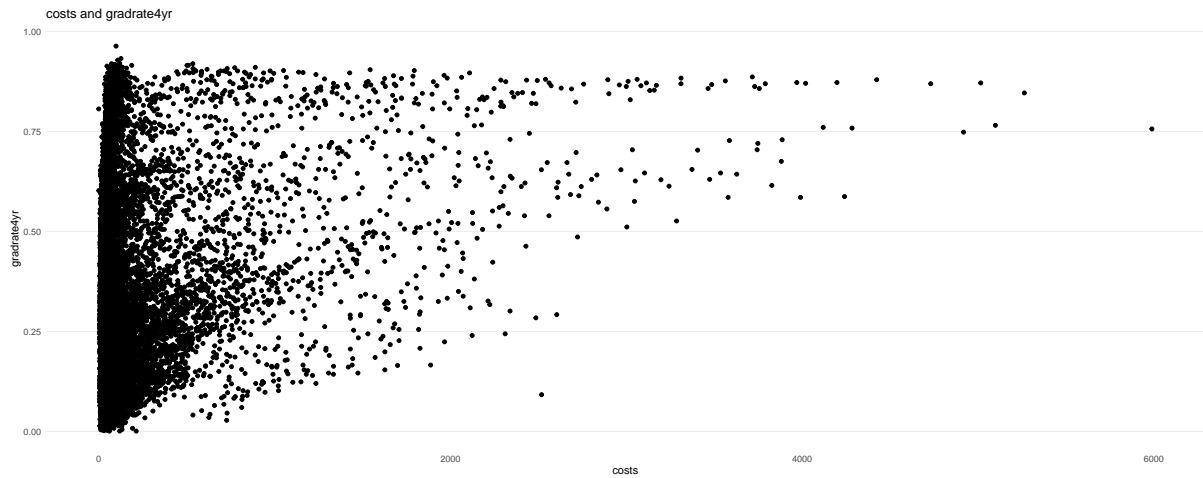


図6 4年卒業率と年間運営コスト (costs)

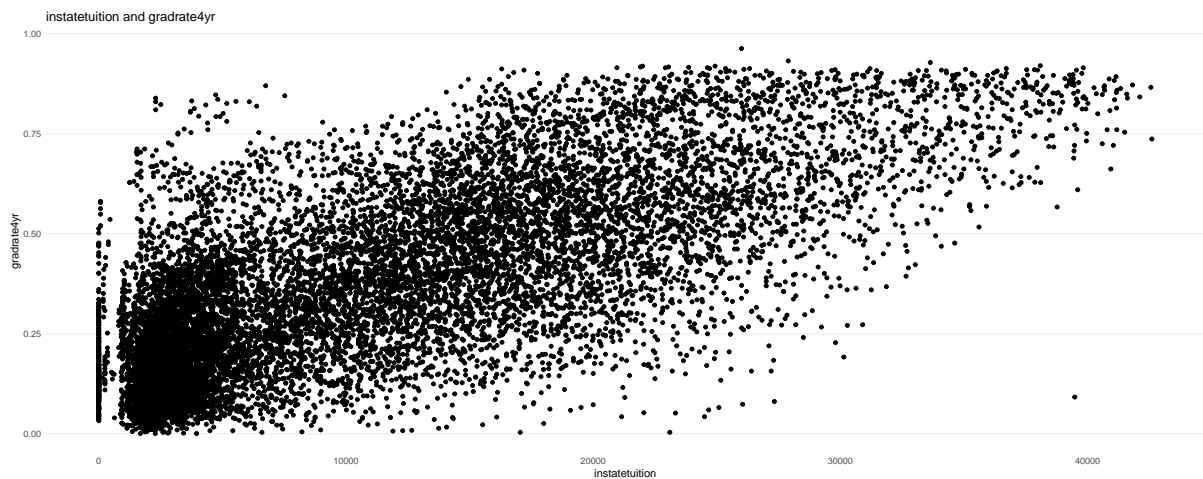


図7 4年卒業率と学費 (instatetuition)

2.5 semester 制の導入と共変量の関係を分析

2学期制に変更することの、各共変量への効果を分析する。各共変量を目的変数に、2学期制に変更したかどうかを表すダミー変数と、時間固定効果・大学固定効果を含めた回帰モデルを推定した。つまり、

$$\text{Covariate} = \beta_1 \text{treated} + \gamma_s + \phi_t + \text{誤差項} \quad (2.1)$$

を推定した。以下に推定結果を示す。

	FTE_faculty	Costs	In.state.Tuition	Cohort.size	Percent.white	Percent.female
treated	7.381 (9.434)	30.662 (30.087)	-1572.596** (539.851)	31.417 (54.476)	0.000 (0.008)	-0.004 (0.004)
Num.Obs.	13 243	13 243	13 243	13 243	13 243	13 243
R2	0.974	0.883	0.922	0.957	0.958	0.875
R2 Within	0.000	0.001	0.005	0.000	0.000	0.000
Std.Errors	by: unitid	by: unitid	by: unitid	by: unitid	by: unitid	by: unitid

表3 semester 制の導入と共変量の関係

3 回帰分析

3.1 シンプルに

以下の式を推定する。ここで、 s : 大学、 k : 相対年数、 Y_{sk} : 4 年卒業率である。

$$Y_{sk} = \beta_0 + \beta_1 \text{treated} + \varepsilon_{sk} \quad (3.1)$$

表4 シンプル回帰分析

	(1)
(Intercept)	0.25*** (0.01)
treated	0.12*** (0.01)
Num.Obs.	13 243
R2	0.007

$p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001$

treated の推定値が統計的に有意ではあるが、この回帰式 (3.1) では、以下の問題があることから正しく treated の係数を推定できていない可能性がある。

- 脱落変数バイアス: treated が誤差項 ε_{sk} に含まれる共変量と相関している場合、推定値にはバイアスが生じる。例えば、授業料が高い大学では財政に余裕があるため、比較的 semester 制へと移行しやすい（移行のコストを厭わない）可能性がある。そうした可能性を考慮して、回帰式にコントロール変数を含む必要がある。
- 時間トレンドによるバイアス: ある年特有の理由（パンデミックや戦争など）で、treated の値に関わらず、4 年卒業率が影響を受けている可能性がある。そうした時間固定効果を考慮に入れるため、固定効果モデルを採用する

- 大学固有のバイアス: 大学が先進的である場合や保守的である場合など、その大学特有の性質に4年卒業率が影響を受けている可能性がある。そうした可能性を考慮に入れて、大学特有の固定効果を捉える項を回帰式に含める必要がある。

3.2 回帰式の改善

以下の通り回帰式を修正し、いくつかの回帰式を設定する。

$$Y_{sk} = \beta_0 + \beta_1 \text{treated} + \varepsilon_{sk} \quad \text{OLS} \quad (3.2)$$

$$Y_{sk} = \beta_0 + \beta_1 \text{treated} + \text{Controls} + \varepsilon_{sk} \quad \text{Multiple OLS} \quad (3.3)$$

$$Y_{sk} = \beta_0 + \beta_1 \text{treated} + \log(\text{Controls}) + \varepsilon_{sk} \quad \text{Multiple OLS with log} \quad (3.4)$$

$$Y_{sk} = \beta_0 + \beta_1 \text{treated} + \phi_t + \gamma_s + \text{Controls} + \varepsilon_{sk} \quad \text{Two way fixed effects} \quad (3.5)$$

$$Y_{sk} = \beta_0 + \beta_1 \text{treated} + \phi_t + \gamma_s + \log(\text{Controls}) + \varepsilon_{sk} \quad \text{Two way fixed effects with log} \quad (3.6)$$

ここで、 ϕ_t : 時間固定効果、 γ_s : 大学固定効果である。以下に回帰式を推定した結果を示す。

表5 各モデルの推定値

	OLS	Multi.OLS	Multi.ln	TFE	TFE.ln
(Intercept)	0.251*** (0.013)	0.111*** (0.008)	-1.408*** (0.025)		
treated	0.122*** (0.013)	0.010 (0.008)	0.003 (0.009)	-0.004 (0.008)	-0.002 (0.009)
totcohortsize		0.000*** (0.000)		0.000*** (0.000)	
costs		0.000*** (0.000)		0.000*** (0.000)	
instatetuition		0.000*** (0.000)		0.000 (0.000)	
log(totcohortsize)			0.011** (0.003)		-0.015* (0.007)
log(costs)			0.038*** (0.002)		0.025** (0.008)
log(instatetuition)			0.173*** (0.002)		0.009 (0.008)
Num.Obs.	13 243	13 243	13 075	13 243	13 075
R2	0.007	0.579	0.550	0.951	0.950
R2 Within				0.025	0.007
Std.Errors				by: unitid	by: unitid

推定結果より、単回帰モデル以外の全てのモデルにおいて treated の推定値が統計的に有意な結果にならなかった。これは、本レポートで検討した範囲内では2学期制に移行することは4年卒業率に影響を与えないことを示唆する。

References

Bostwick, Valerie, Stefanie Fischer, and Matthew Lang. 2022. “Semesters or quarters? The effect of the academic calendar on postsecondary student outcomes.” *American Economic Journal: Economic Policy* 14 (1): 40–80.