

Problem Set2

Yamato Igarashi (2125701)

GitHub URL: https://github.com/yamato5810/MicroDataScience_Intermediate_ProblemSet2

(a) 記述統計

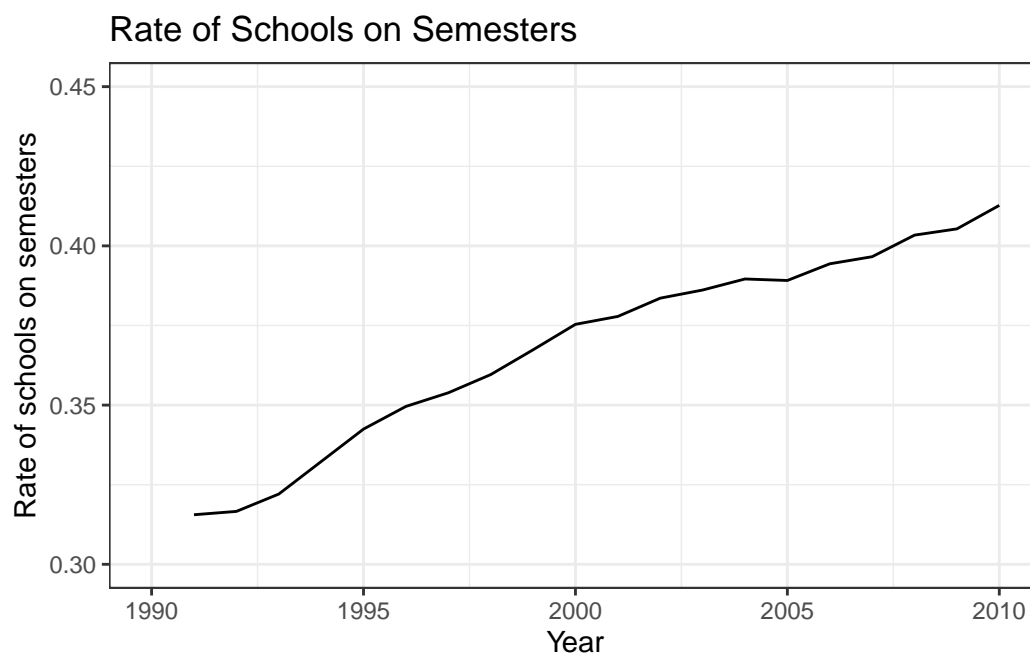
1. 記述統計

	all	never_switchers	switchers
semester	0.93	0.95	0.70
gradrate4yr	0.37	0.38	0.27
women_gradrate_4yr	0.41	0.42	0.32
men_gradrate_4yr	0.32	0.33	0.22
faculty	340.00	335.03	401.04
totcohortsize	1099.45	1084.86	1278.78
instatetuition	11088.47	11375.81	7556.80
costs	192.10	188.67	234.23

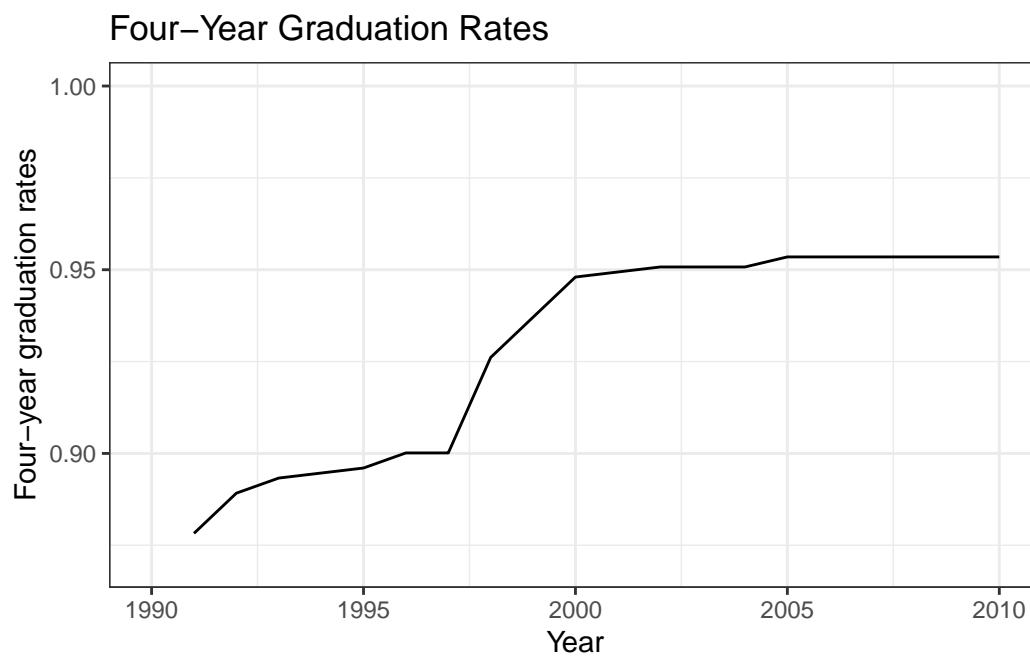
上記の記述統計表より、quarter から semester に変更した大学は、それ以外の大学と比べて、男女にかかわらず 4 年卒業率が低く、学費が安いことがわかる。一方で、cohort size や faculty の人数、大学全体の支出に関しては、semester に変更した大学のほうが平均して多いことも読み取れる。

また、どの集団にも共通して言えることとしては、女性は男性より 4 年卒業率が高くなっている。

2. 4 年卒業率の平均推移

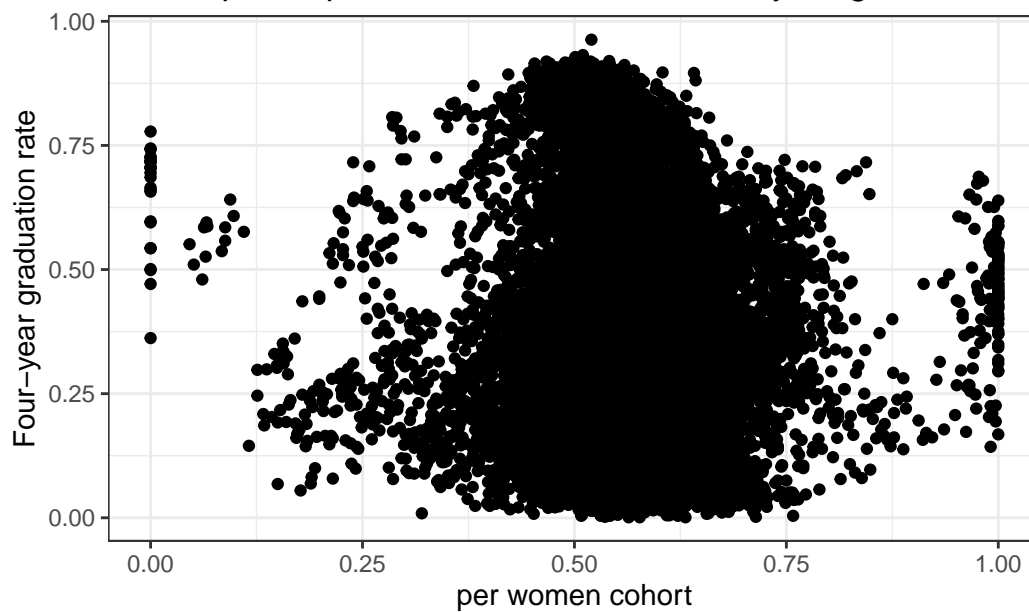


3. semester 導入率の推移

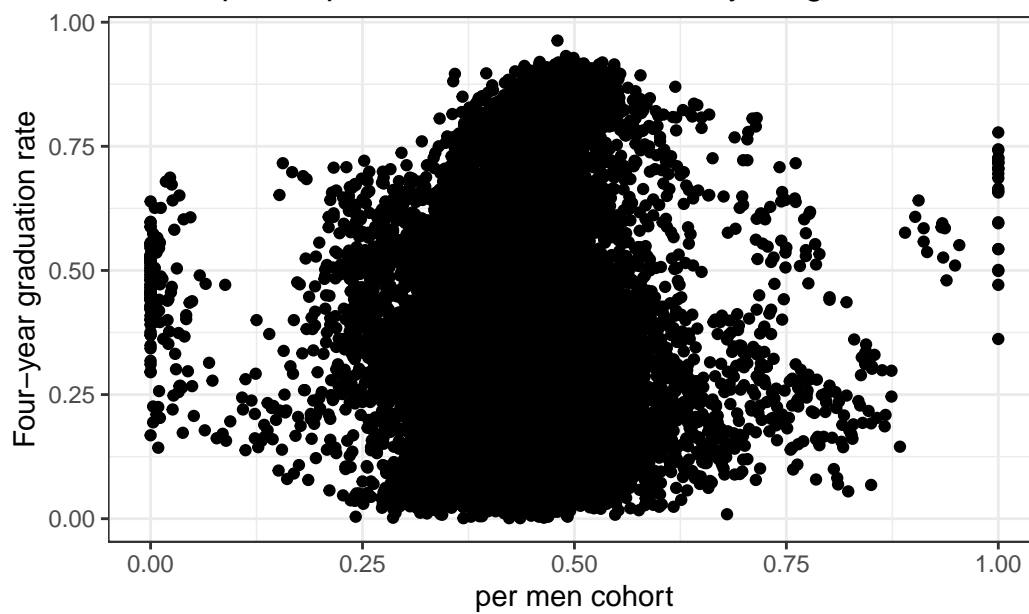


4. 散布图

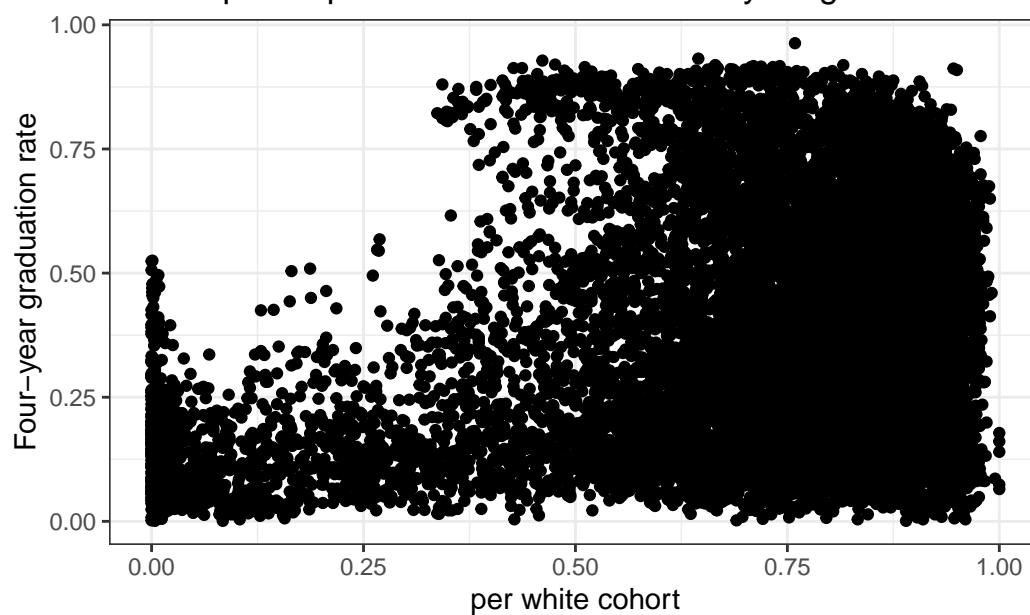
scatter plot of per women cohort and Four-year graduation rate



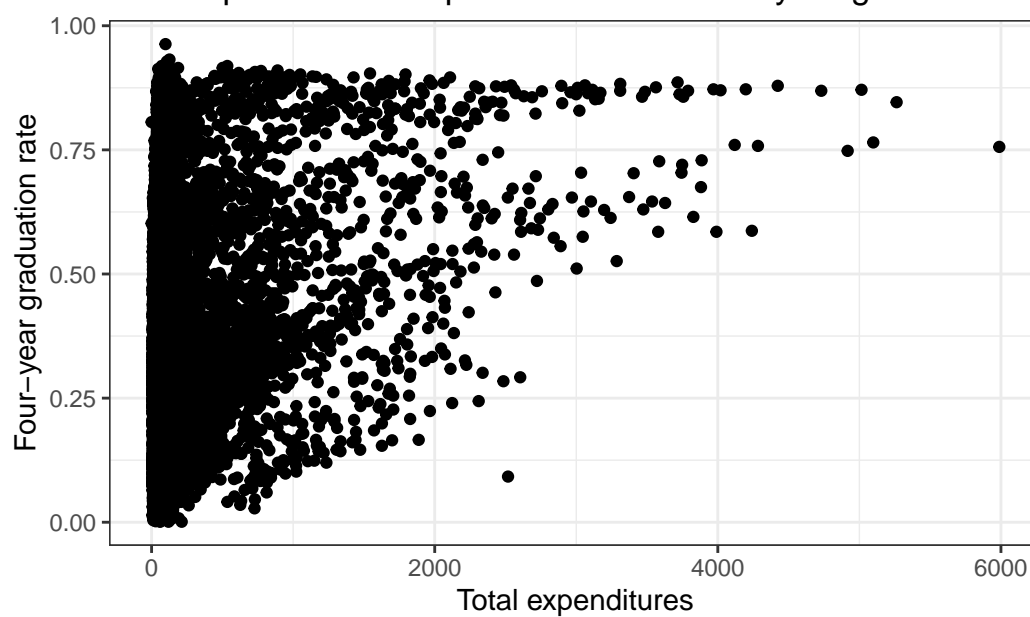
scatter plot of per men cohort and Four-year graduation rate

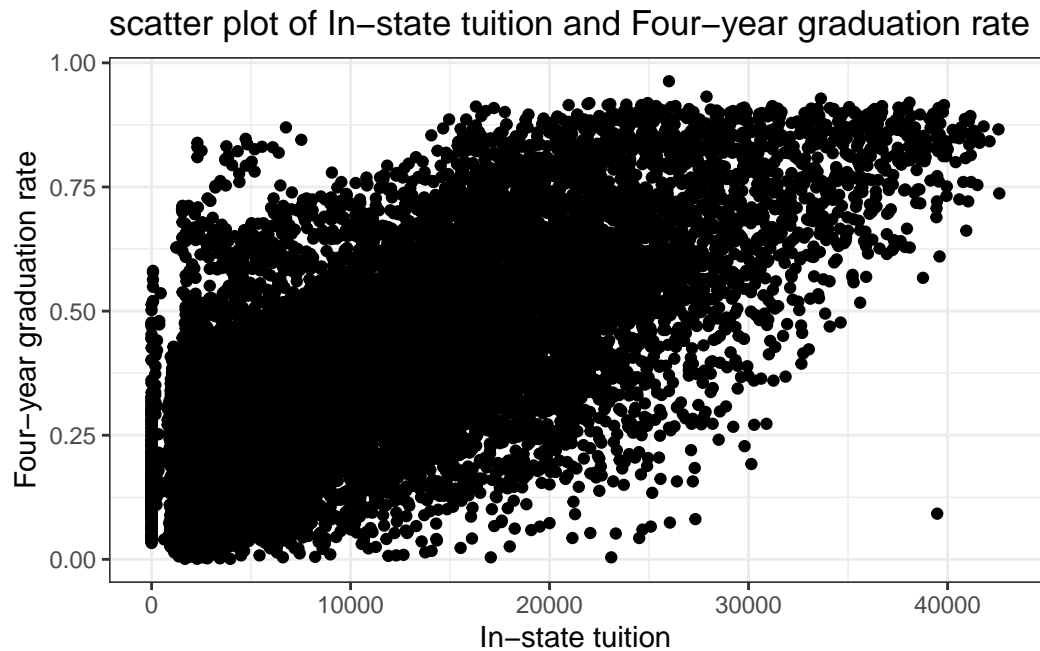


scatter plot of per white cohort and Four-year graduation rate



scatter plot of Total expenditures and Four-year graduation rate





3つ目の散布図より、白人の学生数が極端に少ない場合、4年卒業率が高い大学が存在しないことがわかる。また、4つ目の散布図より、大学の運営コストが極端に大きい場合、4年卒業率が高くなる。最後に5つ目の散布図より、学費が高い大学ほど4年卒業率も高い。この関係性は他の4つの変数より顕著で、相関は0.736542である。

(b) 回帰分析

2. 式の推定と解釈

$Y_{sk} = \beta_0 + \beta_1 \text{treated}_{sk} + \epsilon_{sk} \dots (1)$ について推定を行うと、

	Estimate	Std. Error	t value	Pr(> t)	CI Lower	CI Upper
(Intercept)	0.33675549	0.007506939	44.859225	0.000000e+00	0.3220409	0.35147010
treated	0.03633584	0.007763718	4.680211	2.892878e-06	0.0211179	0.05155377
DF						
(Intercept)	13887					
treated	13887					

上記のような結果が得られた。すなわち、 $\beta_0 = 0.33675549$, $\beta_1 = 0.03633584$ ということである。したがって、quarter 制の大学は $\text{treated} = 0$ であるので平均的な4年卒業率が $0.33675549 = \beta_0$ 、semester 制の大学は平均的な4年卒業率が $0.37309133 = \beta_0 + \beta_1$ ということがわかる。

3. 数式 (1) の問題点

上記の数式 (1) の問題点は主に以下の 3 点である。

- ・ quarter から semester になった大学を処置群、quarter のままの大学を対照群と仮定した場合、観測以前から semester であった大学はそのどちらでもないで、推定の外に出すか、それに対応した項を設けなければいけない点。
- ・ semester 制が導入されたことによる 4 年卒業率への影響は、導入された年に完結するわけではない。それにもかかわらず、10 年以上前にすでに変更している場合や 1 年前に変更した場合などのすべての場合について、同一の関連性 (same β_1) があると仮定してしまっている点。
- ・ 例えば、1991 年に semester 制が導入された場合と 2001 年に導入された場合では、景気やその他の交絡因子の影響を受けて、同一の関係性 (same β_1) があるとはいいがたい。すなわち、処置のタイミングを考慮していない点。

4. 改善が加えられている回帰式

$$Y_{sk} = \beta_0 + \beta_1 1(k=1) + \beta_2 1(k=2) + \beta_3 1(k=3) + \beta_4 1(k=4) + \beta_5 1(k=5,6) + \beta_6 1(k>6) + \epsilon_{sk}$$

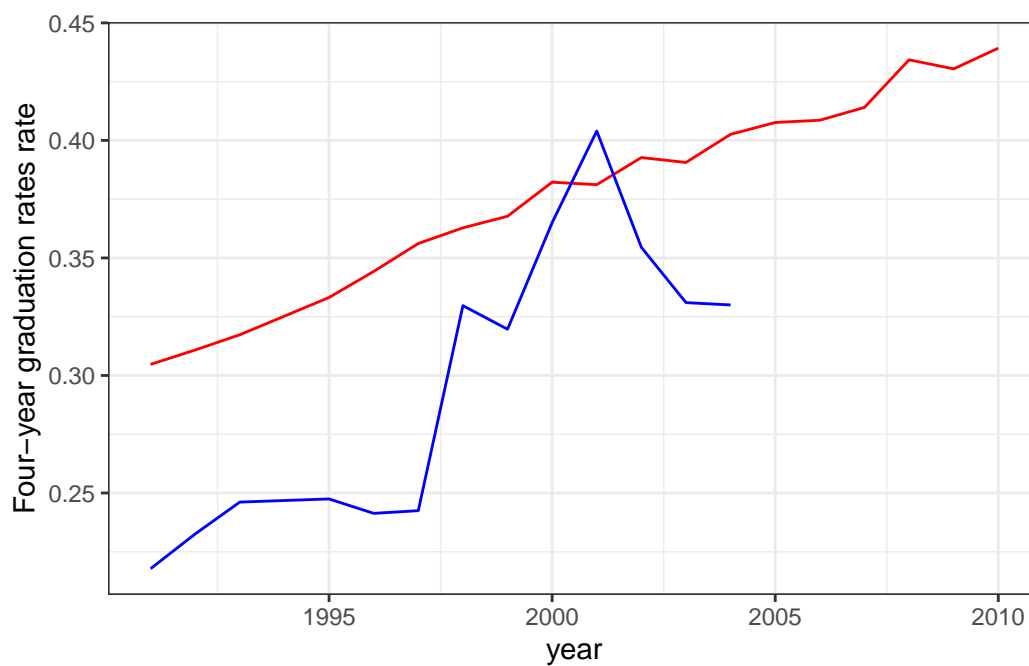
1(.) を k (相対年数 (yearstosem)) の値に応じた指示関数とする。上記のように、処置の時点からの経過年数ごとにダミー変数を定めることで、Partially treated の期間などをそれぞれ別に考えることができる。

加えて、以下の推定では、観測された期間中で少なくとも 1 年は、quarter であった大学のみに対して分析した。(観測以前から semester 制であった大学を排除した。)

	Estimate	Std. Error	t value	Pr(> t)	CI Lower
(Intercept)	0.33675549	0.007506939	44.859225	8.622072e-290	0.32203157
sem1yr	-0.09626458	0.024462213	-3.935236	8.649543e-05	-0.14424412
sem2yrs	-0.07779395	0.025278651	-3.077456	2.121411e-03	-0.12737483
sem3yrs	-0.06071293	0.027147740	-2.236390	2.545674e-02	-0.11395980
sem4yrs	-0.07730094	0.024891288	-3.105542	1.930969e-03	-0.12612206
sem5_6yrs	-0.05856458	0.018985599	-3.084684	2.070819e-03	-0.09580243
sem_above6	-0.04050729	0.012051663	-3.361137	7.935848e-04	-0.06414511
	CI Upper	DF			
(Intercept)	0.351479399	1684			
sem1yr	-0.048285037	1684			
sem2yrs	-0.028213066	1684			
sem3yrs	-0.007466069	1684			
sem4yrs	-0.028479822	1684			
sem5_6yrs	-0.021326722	1684			
sem_above6	-0.016869478	1684			

結果として、2. での結果と異なり、quarter から semester にすると、4 年卒業率が低下することがわかる。

5. 平行トレンド



赤線が対照群、青線が介入前の処置群の 4 年卒業率の推移を示している。介入前の処置群の数が時間とともに減っているため、1997 以降は平行トレンドになっていないが、それ以前は平行トレンドの仮定を満たしていると考えられる。