

Newton Kim

Economics 140

03 May 2020

What Demographics Characteristics Determine Success?

Introduction:

With September fast approaching many Canadian colleges are ready to open their doors to new and returning students. These students will be composed of a wide range of diverse backgrounds. We must collect data to ensure that students of every background have a good shot at living up to their educational potential. For most students, the goal of attending college is to earn a bachelor's degree by graduating in the standard four-year time frame. How does a student's demographic characteristics influence the on-time graduation rate of four years? It is important to consider these factors that may affect academic performance to understand how to better help students who are struggling.

Data:

The cross-sectional dataset is from a large Canadian university with one large main campus and two smaller branch campuses. The dataset I will be analyzing list 34,979 students across a Canadian university and its two satellite campuses. Each one of the 34,979 observations in the dataset counts as one student randomly sampled from the university system. Included in the summary results are variables containing GPA of student in year 1 or 2, total credit attempted by the student in year 1 or 2, sex, native language(mtongue), age at entry, how many years it took to graduate(gradin4,5,6), GPA cutoff, probation in year 1 or 2, birthplace, high school grade percentile ranking, and what campus they are attending. On table 1 I summarized the important variables I will be using in my analysis of the data.

Regression:

The question I wanted answered was how does demographic characteristic effect on time graduation rate of the student. The dependent variable of whether the student graduates in four years is represented by the dummy variable gradin4. In addition, the independent variables are age at entry, birthplace, gender, and mother tongue in this regression model. I converted the variables of sex, mother tongue, and birthplace into dummy variables. Sex became female, mtounge became English and French, and birthplace became BICanada, BIAasia, and BIAmerica. So, the first regression had gradin4 regressed by age at entry, female, English, French, BICanada, BIAmerica, and BIAasia. In the second regression I added an interaction term to see if English was picking up the effect of being born on Canada or America on the four-year graduation outcome of the student. I also added an interaction term to account for the French speaking population in the same respect as English on BICanada. To correct for homoscedasticity in my linear prediction models I used the robust command when running my regressions. I also exclude the Blother variable and mother tongue other to prevent prefect collinearity. On table 1 I summarized the important variables I will be using in my analysis of the data.

TABLE1:

Summary of Data

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
gradin4	22,262	.5444255	.4980337	0	1
age_at_entry	34,979	18.63504	.7347842	17	21
female	34,979	.6260042	.4838694	0	1
English	34,979	.7170588	.4504345	0	1
French	34,979	.005003	.0705558	0	1
-----+-----					
BICanada	34,979	.8783556	.3268793	0	1
BIAmerica	34,979	.0018583	.0430681	0	1
BIAasia	34,979	.0770176	.2666232	0	1

Results:

Regression #1

	gradin4	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
age_at_entry							
18		-.0942734	.0186551	-5.05	0.000	-.1308387	-.057708
19		-.1240246	.0181644	-6.83	0.000	-.1596282	-.088421
20		-.1976772	.0218702	-9.04	0.000	-.2405444	-.15481
21		-.2081963	.0324102	-6.42	0.000	-.2717227	-.14467
female		.1129942	.0068716	16.44	0.000	.0995254	.126463
English		-.02402	.0082463	-2.91	0.004	-.0401833	-.0078566
French		-.0683713	.04786	-1.43	0.153	-.1621803	.0254377
BICanada		.0731796	.0167474	4.37	0.000	.0403536	.1060057
BIAmerica		.0640792	.0790454	0.81	0.418	-.0908554	.2190138
BIAsia		.1299364	.0197683	6.57	0.000	.091189	.1686837
_cons		.5348646	.023635	22.63	0.000	.4885384	.5811908

Equation 1

$$\widehat{\text{gradin4}} = .535 - .094\beta_{18} - .124\beta_{19} - .198\beta_{20} - .208\beta_{21} + .113\beta_{\text{female}} - .024\beta_{\text{English}} \\ - .068\beta_{\text{French}} + .073\beta_{\text{BICanada}} + .064\beta_{\text{BIAmerica}} + .130\beta_{\text{BIAsia}} + U$$

In the first regression we see that holding all other variables constant, a student with English as their mother tongue will have a 0.024 lower rate of graduating in four years than those who speak other languages as their mother tongue on average. In addition, those who speak French as their mother tongue will have a 0.068 lower rate of graduating in four years than those who speak other languages as their mother tongue on average. The English variable has a statistically significant P-value at 0.05 while the French variable does not. A female student is 0.113 more likely to graduate in four years than males holding all other variables constant. When it comes to the age at entry variable, we see that attending this college at eighteen years of age is associated with a negative coefficient. Every year after eighteen there is an increase in the magnitude of failing to graduate in four years for the student. Indicating that the older a student at age of entry the harder the time they will have to graduate in four years. Both the variables female and age_at_entry is statistically significant at the 0.05 level.

For a student that is born in Canada they have a 0.073 higher graduation rate on average than those who are born in other places that do not include America and Asia. Students that are born in America have a 0.067 higher graduation rate on average in four years than those in other places not including Canada and Asia. For BIAmerica they have a 0.130 of graduating on time from this Canadian university compared to other areas on average. The variables born in Canada and Asia have statistically significant P-values at a 0.05 level while BIAmerica does not.

A surprising result from the regression is that those who speak English have a negative effect on graduating in four years. English is the most spoken language in Canada you would think that those who have English as their native language have a better chance of graduating on time. Furthermore, students who are born in Asia have a better chance of graduating than those who are born in Canada. Which is another surprising result since you would think those who were born in Asia might have a harder time adjusting to Canadian educational system but end up doing better than the native Canadians. Additionally, you see at every age there is a negative effect in graduating in four years. Pointing to the fact that those who enter college at any age are unlikely to graduate on time.

In the second regression model, I will be adding interaction terms to see if having English as a student's mother tongue combined with being born in Canada or America influences graduating in four years. In addition, I added the interaction effect of having French as a student's mother tongue on being born in Canada. I added these interaction terms because the most spoken languages in Canada are French and English. In addition, I used the same reasoning for adding in the interaction term for BIAmerica and English. In Equation two you can see the added interaction terms used in the regression output.

gradin4	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age_at_entry	-.0476023	.0047773	-9.96	0.000	-.0569662	-.0382383
female	.113271	.0068695	16.49	0.000	.0998063	.1267357
1.French	-.3733466	.1536631	-2.43	0.015	-.6745372	-.072156
1.BICanada	.0372485	.0192454	1.94	0.053	-.0004739	.0749708
French#BICanada						
1 1	.3287891	.1613786	2.04	0.042	.0124756	.6451026
1.English	-.099584	.0222406	-4.48	0.000	-.1431772	-.0559907
English#BICanada						
1 1	.0886407	.0239531	3.70	0.000	.0416911	.1355904
BIAAsia	.1190567	.0199659	5.96	0.000	.0799222	.1581913
1.BIAmerica	.1635985	.1446122	1.13	0.258	-.1198517	.4470487
English#BIAmerica						
1 1	-.0877221	.1706313	-0.51	0.607	-.4221715	.2467273
_cons	1.3348	.0921215	14.49	0.000	1.154236	1.515365

EQUATION 2

$$\begin{aligned}
 \widehat{gradin4} = & 1.33 - .048\beta_{age_{at_{entry}}} + .113\beta_{female} - .373\beta_{Frenc} + .037\beta_{BICanada} \\
 & + .329\beta_{FrenchBICan} - .100\beta_{English} + .089\beta_{EnglishBICa} + .119\beta_{BIAAsia} \\
 & - .088\beta_{EnglishBIAmerica} + U
 \end{aligned}$$

For the second regression model those who spoke English as their mother tongue, being born in America decreased the four-year graduation rate by an average of 0.188 holding all other variables constant. Speaking English as their mother tongue, being born in Canada results in a 0.025 increase in the four-year graduation rate holding all other variables constant. For those who speak French as their mother tongue, being born in Canada will decrease their four-year graduation rate by 0.007 units holding all other variables constant. The interaction terms for English and BIAmerica are not statistically significant. While, the interaction term for French and BICanada, English and BICanada are statistically significant at a level of 0.05.

Having your mother tongue be English and being born in Canada has a positive effect on graduating in four years. While speaking French and being born in Canada has a negative effect in graduating in four years. Being born in Asia has the highest effect of graduating in four years

as it did in the first regression. I performed a third regression to see the interaction effect of being born in Asia and mother language as other on graduating in four years. From the regression those who spoke other languages as their mother tongue, being born in Asia increased the four-year graduation rate by an average of 0.128 holding all other variables constant. The coefficient BIAAsia Other is statistically significant at 0.05 level. Compared to the interaction effect of English on BICanada of 0.025 this value is much bigger effect on four-year graduation.

REGRESSION 3

gradin4	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
age_at_entry	-.0477555	.0047803	-9.99	0.000	-.0571252	-.0383858
female	.1131738	.0068693	16.48	0.000	.0997094	.1266382
French	-.0450473	.0474928	-0.95	0.343	-.1381365	.0480419
BICanada	.006055	.07718	0.08	0.937	-.1452233	.1573332
1.BIOther	-.0886757	.0817985	-1.08	0.278	-.2490065	.0716551
1.Other	.0112892	.0088649	1.27	0.203	-.0060866	.028665
BIOther#Other						
1 1	.0399757	.0346282	1.15	0.248	-.0278981	.1078495
1.BIAAsia	-.0328994	.0815116	-0.40	0.686	-.192668	.1268691
BIAAsia#Other						
1 1	.128759	.0307845	4.18	0.000	.0684191	.1890988
_cons	1.357975	.117267	11.58	0.000	1.128123	1.587827

This makes sense because the Born in Asia variable has the biggest effect on graduating in four years in the first regression. Intuitively we can say that those students who are not born in Canada or America are better prepared to succeed at this university. Since they are attending a Canadian college, of course they are going to have the ability to speak English. In addition to English being such a prominent secondary language throughout the world. Even though their mtongue is a language other than English they may be able to speak as fluently as a native Canadian. Any effect of an advantage I thought a student had as having their mother tongue being English is nonexistent. Another factor could be that they came to Canada before they were

enrolled at the university. Making it easier to transition into this Canadian College environment. If this is the case more information is needed to complete this analysis.

Conclusion:

In the regression models we see that those who enter college at all ages are at a disadvantage in graduating in four years especially if those students are older. We see that female students more than male students graduate in four years. Those who were born in Asia do better than those born in Canada and America. Those who speak other languages beside French and English as their mother tongue do better in terms of four-year graduation rates. When adding in interaction terms we see that those who are born in Canada and speak English as their mother tongue have a positive effect in graduating in four years. The effect however is not as big as those who speak another language and are born in Asia. As we seen in the first regression those who are born outside of Canada do better in graduating at this university. Knowing these results, we can create policy at the university to help those who do not graduate on time.

At every age of entry there is a negative association with graduating in the span of four years. A potential solution maybe to help a student even before they are to enter the college. It makes sense to invest in students who are entering into the college straight from high school by better preparing them. Maybe outreach programs that enable a smooth transition into the college. For older students, the problem becomes more complicated because of a gap in education that one must have experienced. It would be hard for them to graduate if they had to enroll in remedial classes and relearned what they lost. A potential solution could be shorter remedial classes backed by tutoring and mentorship.

From the data those who are born in Canada four-year graduation rates are worse compared to those born in Asia. Being born in Canada the problem is Canadian student are more

unprepared to enter this university than their Asian peers. There is not much the college can do to help with this situation because it would be a problem at the national level. Maybe enacting policies that mimic Asian educational standard would help in reducing the gap in education outcomes.

Another effect we learn from the regression is that women do better in graduation rates than men. There could be a wide variety of factors that determine why women on average do better than men in terms of four-year graduation rates. The university should gather more information about why that is.

STATA OUTPUT

```
*** LOAD IN DATASET****
```



```

use "C:\Users\Newton\Desktop\canadiancollege.dta"

**Analyze Data**
sum GPA_year1 GPA_year2 sex mtongue age_at_entry birthplace

** Will change the sex variable into male and female first**
tab sex, gen(gender)
sum gender2 //Dummy variable//
sum gender1
tab sex

** Finding the correct variables for each**
rename gender2 male
rename gender1 female
sum female //Check to see if good//

** Repeat with the mtongue variable**
tab mtongue, gen(mtungue)
sum mtongue*
rename mtongue1 English
rename mtongue2 French
rename mtongue3 Other
sum
tab English //CHECK//

**Repeat w/ birthplace**
tab birthplace, gen(birthplace)
sum birthplace*
rename birthplace3 BICanada
rename birthplace2 BIAAsia
rename birthplace1 BIAmerica
rename birthplace4 BIOther
sum

**Regression 1***
reg gradin4 i.age_at_entry female English French BICanada BIAmerica BIAAsia, robust

*** ADDING IN INTERACTION TERMS***
regress gradin4 c.age_at_entry female i.French###i.BICanada i.English###i.BICanada BIAAsia
i.English###i.BIAmerica, robust

```