

STEPHEN OGBODO (CAnD3 Fellow, 2023/2024)

RESEARCH REPLICABILITY ASSIGNMENT

PROGRAM

Title: Immigrant Status and Commuting times in Canada: A Cross-Sectional Study

Study Aim: To determine if immigrants in Canada commute for longer to work than non-immigrants.

Data: Canadian 2016 population census.

Task 1: Clean the dataset

- Select the seven needed variables: age (AGEGRP, *factor*), total income (TotInc, *integer*), education (HDGREE, *factor*), immigration status (IMMSTAT, *factor*), commuting time (PWDUR, *integer*), employment status (POWST, *factor*), province (PR, *factor*)
- Remove participants with age = 88 ("not available")
- Remove participants with net income = 88888888 or 99999999 ("not available/applicable")
- Remove participants with education = 88 or 99 ("not available/applicable")
- Remove participants with commute time = 8 or 9 ("not available/applicable")
- Since the commute time is provided as an ordinal variable (1 to 5) representing 15-minute increments, multiply each category by 15 to produce commuting time in minutes. This should yield commuting times ranging between 15 and 75 minutes, enabling the use of a linear regression for this continuous outcome.
- Remove all persons without employment (POWST = 8 or 9) and those who worked from home (POWST=1) or work outside Canada (POWST=3)
- Remove participants without available immigrant status information (IMMSTAT=8)
- Create a new immigration status variable from IMMSTAT, with level 1 representing "Non-immigrant" (IMMSTAT=1) and level 2 representing all immigrants (including both *immigrants* (IMMSTAT=2) and *non-permanent residents* (IMMSTAT=3)).
- Name the provinces in the PR variable (for tabulation in subsequent steps).
- Delete redundant variables and proceed to analysis.

Task 2: Descriptive statistics

- Compute the percentages of individuals in each province/territory that are immigrants and present this in a table.
- Compute the average commute time for *immigrants* and *non-immigrants* for the entire dataset.

Task 3: Regression analysis without weights

- Regress *immigration status* on *commuting time*, adjusting for age, income, education, province, and employment status (i.e. POWST).
- Present the output in a table.
- Note: given that multiple participants were dropped from the dataset, it is unclear if the provided survey weights are still applicable. As such, I did not include them in the regression. In addition, the weights are essentially the same for all participants (37.0), which seems very odd.

My results are presented below.

RESULTS

Table 1: Percentage of immigrant workers in each province/territory

PROVINCE/TERRITORY	% IMMIGRANTS
Alberta	22.3
British Columbia	27.8
Manitoba	18.1
New Brunswick	2.3
Newfoundland and Labrador	1.6
Northern Canada	2.6
Nova Scotia	3.9
Ontario	29.7
Prince Edward Island	2.1
Quebec	14.0
National Average	21.9

Table 2: Average commute time by immigration status

Immigration status	Average Commute Time (Minutes)
Immigrant	40.1
Non-immigrant	33.8

Interpretation: On average, immigrants spend 7 more minutes commuting to work than non-immigrants, although this crude estimate may be confounded by educational attainment, income level, province of residence, etc. These are accounted for, using the multivariate linear regression below.

Table 3: Multivariate Linear Regression Output

Variable	Estimate	Standard Error	P-val
Immigration status: Non-immigrant (ref)	-	-	-
Immigration status: Immigrant	5.35	0.06	<0.0001
Age X 15 levels	-	-	<0.0001
Income (continuous)	0.00	0.00	<0.0001
Education X 11 levels	-	-	<0.0001
Province X 10 levels	-	-	<0.0001
Employment status X 4 levels	-	-	<0.0001

Interpretation: Controlling for other factors, immigrants in Canada commute for 5 minutes more than non-immigrants on average, and this is statistically significant.

The End.