

CAnD3

Fellow name: Nicole Antunes Rezende

Assignment: Research Replicability and Workflow Management (RRWM)

Research question:

How does the region of origin influence the educational attainment of immigrants, controlling for sex, province of residence, and age at immigration?

INSTRUCTIONS

1. Open the dataset using the package 'readr'

2. Select only the variables of our study (PPSORT, HDGREE, AGEIMM, POB, PR, Sex) using the package 'dplyr'

(Optional) have an idea of the variables using the function 'glimpse' of the package 'tibble'

3. Recode Sex: Female from 1 to 0 and Male from 2 to 1.

 # 3.1 Transform the variable Sex into a categorical variable using the function 'factor'

4. Recode POB, creating a new variable 'originregion', knowing that:

1 = "Born in Canada",

2 = "Born in USA",

3 to 6 = "Born in Latin America",

7 to 15 = "Born in Europe",

16 to 18 = "Born in Africa",

19 to 31 = "Born in Asia",

32= "Born in Oceania and others",

88= "Missing value"

4.1 Add the previous labels to the new variable 'originregion':

5. Recode HDGREE creating a new variable 'educlevel', knowing that:

1 = "No certificate, diploma or degree" ///

2 = "Secondary (high) school diploma or equivalency certificate" ///

3 and 4 = "Trades or Apprenticeship certificate/diploma" ///

5 to 8 = "College, CEGEP or other non-university certificate or diploma;
University certificate or diploma below bachelor level" ///

9 = "Bachelor degree" ///

10 to 13 = "University certificate, diploma or degree above bachelor level"

88 or 99 = NA

6. Add the labels to the variable PR, following the information of the Census codebook (pdf) available with the Census dataset

7. Add the labels to the variable AGEIMM, following the information of the Census codebook (pdf) available with the Census dataset:

8. Create one frequency table for the variables originregion and PR

8.1 Export to a csv document using the function 'write.csv'

9. Perform a linear regression, using as dependent variable: educ level. The main independent variable is originregion. Control for Sex, PR and AGEIMM

9.1 Format the results using the function 'tidy' of the package named 'broom', before exporting it to a cvs document using the function 'write.csv'

10. Format the two output cvs documents (convert the text of the two csv documents into separate columns, using the path: Data --> Text to Columns --> Delimited --> Delimiter: comma --> Finish)

