Evaluating the Influence of Education on Current Unemployment Trends*

Adrian Ly

March 26, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

Table of contents

*Code and data are available at: LINK

1 Introduction

2 Data

Table 1: Unemployed Dataset from 2019 to 2023 by education level

Education Level	Reference Period	Youth Unemployment	Adult Unemployment	Senior Unemployment
Total, all education levels	2019	10.7	5.1	4.6
Total, all education levels	2020	20.1	8.5	7.4
Total, all education levels	2021	13.5	6.4	6.7
Total, all education levels	2022	10.1	4.6	4.4
Total, all education levels	2023	10.8	4.9	4.2
0 to 8 years	2019	17.5	9.4	7.8

2.1 Data Measurement

2.2 Data Analysis

This paper uses the Statistics of Canada `Unemployment rate, participation rate and employment rate by educational attainment, annual` (Canada 2024) dataset. We are using the `R programming language` (R Core Team 2023) to conduct our analysis along with `readR` (Wickham, Hester, and Bryan 2024), `lubridate` (Grolemund and Wickham 2011), `tidyverse` (Wickham et al. 2019), `dplyr` (Wickham et al. 2023), `tidyr` (Wickham, Vaughan, and Girlich 2023), `knitr` (Xie 2014), `janitor` (Firke 2023), `scales` (Wickham, Pedersen, and Seidel 2023), `RColorBrewer` (Neuwirth 2022), `ggplot2` (Wickham 2016), `kableExtra` (Zhu 2024), `here` (Müller 2020), `arrow` (Richardson et al. 2024), `rstanarm` (Brilleman et al. 2018), `modelsummary` (Arel-Bundock 2022), and `lme4` (Bates et al. 2015).

3 Model

The goal of our modeling strategy, utilizing multiple linear regression, is multifaceted:

- **Understanding Relationships**: To understand the relationship between several independent variables, such as education level and age groups, and the dependent variable, which in this context is the youth unemployment rates. The model is designed to isolate the effect of each variable on unemployment rates.
- **Quantifying Impact**: To quantify the impact of each predictor. Regression coefficients provide a numerical value that represents the expected change in the dependent variable for a one-unit change in an independent variable, all else being equal.
- **Predictive Analysis**: To develop a predictive framework that can be employed to estimate or forecast the outcome variable, in this case, unemployment rates, when provided with new data for the predictors.
- **Inferential Statistics**: To conduct hypothesis testing to ascertain whether the relationships observed within the data are statistically significant and not merely due to random variation.
- **Policy Decision Support**: To offer empirical evidence that may inform policy decisions. For instance, if a correlation is found between higher education levels and increased unemployment, it might indicate the necessity for policies that better align education with employment opportunities.
- **Data-Driven Insights**: To extract insights from the data that could lead to an enhanced understanding of labor market dynamics, such as how different age groups' unemployment rates interact.
- **Modeling Assumptions Testing**: To validate the assumptions underlying the regression model, ensuring the reliability and accuracy of the model's predictions.
- **Confounding Variable Control**: To account for potential confounders that might influence the relationship between education level and unemployment rates. The inclusion of relevant variables aims to minimize estimation bias.
- **Sensitivity Analysis**: To comprehend the model's sensitivity to changes in inputs, which assists in evaluating the robustness of the model.

This strategy is designed to produce a model that is both explanatory and predictive, offering a current snapshot of the relationships in the data as well as a tool for future forecasting. Such a model can be utilized to guide decision-making processes, or pinpoint where interventions may be effective.

3.1 Model set-up

$$\begin{aligned}\text{Youth Unemployment} = \beta_0 &+ \beta_1 \cdot \text{Education Level} \\ &+ \beta_2 \cdot \text{Reference Period} \\ &+ \beta_3 \cdot \text{Adult Unemployment} \\ &+ \beta_4 \cdot \text{Senior Unemployment} + \epsilon\end{aligned}$$

Where:

- `Youth Unemployment` is the dependent variable we are trying to predict. These are generally categorized as people who are between the ages of 15-24.
- `Education Level` is an independent variable representing the level of education. (e.g., high school, undergraduate, postgraduate)
- `Reference Period` is another independent variable that represents the time frame of the data collection (e.g., year).
- `Adult Unemployment` and `Senior Unemployment` are independent variables representing unemployment rates for adult and senior age groups, with people ages ranging between 25-44 and above 45, respectively.
- β_0 is the y-intercept, representing the expected value of `Youth Unemployment` when all the independent variables are 0.
- $\beta_1, \beta_2, \beta_3, \beta_4$ are the coefficients for each independent variable, representing the change in `Youth Unemployment` for a one-unit change in the respective independent variable, holding all other variables constant.
- ϵ represents the error term, accounting for the variability in `Youth Unemployment` not explained by the model.

This model aims to understand how various factors contribute to youth unemployment. The coefficients $\beta_1, \beta_2, \beta_3, \beta_4$ provide insights into the relationship between each independent variable and youth unemployment:

- β_1 tells us how changes in the education level might affect youth unemployment, holding other factors constant. A positive coefficient suggests that higher education levels are associated with higher youth unemployment rates, while a negative coefficient suggests the opposite.

- β_2 gives insight into how changes over the reference period (e.g., years) impact youth unemployment, which can help identify trends over time.

- β_3 and β_4 help understand how unemployment rates in other age groups (adults and seniors) are related to youth unemployment. This can reveal whether higher unemployment rates in these groups are associated with higher or lower youth unemployment rates.

The intercept β_0 provides the baseline level of youth unemployment when all other variables are zero, which may not always be a practical scenario but is necessary for the model's mathematical formulation. The error term ϵ captures the model's limitations and any random factors affecting youth unemployment that aren't included in the model.

3.2 Model justification

The utilization of a multiple linear regression model for analyzing the relationship between people's unemployment status, age groups, and education level over time is substantiated by various rationales. Primarily, the model adeptly accommodates multiple independent variables such as education level and age group, each potentially exerting influence on the dependent variable, unemployment status. This capacity allows for an assessment of the individual impact of each predictor while simultaneously controlling for the influence of others. Such a model not only facilitates the quantification of the relationships between these factors and unemployment—via coefficients that express the extent of change in response per unit change in a predictor (with other variables held constant)—but also enables hypothesis testing on the parameters. This testing is instrumental in determining whether the associations between predictors and the response are statistically significant, thereby offering insights into the meaningfulness of these relationships.

4 Results

4.1 Youth Unemployment

The following formula was used to calculate the net percentage change in the unemployment rate:

$$\text{Net Percentage Change} = \left(\frac{\text{Value at End Period} - \text{Value at Start Period}}{\text{Value at Start Period}} \right) \times 100\%$$

Table 3: Percentage Change in Youth Unemployment by education level (2019-2023)

Education Level	Unemployment Rate (2019)	Unemployment Rate (2023)	Percentage Change (%)
0 to 8 years	17.5	18.6	6.29
Above bachelor's degree	10.3	9.3	-9.71
Bachelor's degree	7.7	8.5	10.39
High school graduate	10.9	11.2	2.75
Postsecondary certificate or diploma	6.7	6.4	-4.48

Table 3: Percentage Change in Youth Unemployment by education level (2019-2023)

Education Level	Unemployment Rate (2019)	Unemployment Rate (2023)	Percentage Change (%)
Some high school	16.8	16.8	0.00
Some postsecondary	9.8	9.1	-7.14
Total, all education levels	10.7	10.8	0.93
University degree	8.0	8.6	7.50

Table 4: Percentage Change in Youth Unemployment by education level (2022-2023)

Education Level	Unemployment Rate (2022)	Unemployment Rate (2023)	Percentage Change (%)
0 to 8 years	15.0	18.6	24.00
Above bachelor's degree	8.8	9.3	5.68
Bachelor's degree	7.6	8.5	11.84
High school graduate	10.6	11.2	5.66
Postsecondary certificate or diploma	6.7	6.4	-4.48
Some high school	14.9	16.8	12.75
Some postsecondary	8.5	9.1	7.06
Total, all education levels	10.1	10.8	6.93
University degree	7.7	8.6	11.69

Table 2: Regression Analysis: Impact of Education Level and Age on Youth Unemployment

	Baseline Model for Youth Unemployment
(Intercept)	328.01
	(368.38)
Above Bachelor's	5.14***
	(1.40)
Bachelor's	4.14**
	(1.37)
High School Grad	−0.41
	(1.16)
Postsec. Cert/Dip	1.24
	(1.29)
Some High School	−0.95
	(1.10)
Some Postsec.	−2.20+
	(1.28)
All Levels	4.55***
	(1.25)
University	4.27**
	(1.37)
Reference Period	−0.16
	(0.18)
Adult Unemployment	2.19***
	(0.31)
Senior Unemployment	−0.20
	(0.37)
Num.Obs.	45
R2	0.923
R2 Adj.	0.898
AIC	182.4
BIC	205.9
Log.Lik.	−78.222
RMSE	1.38

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

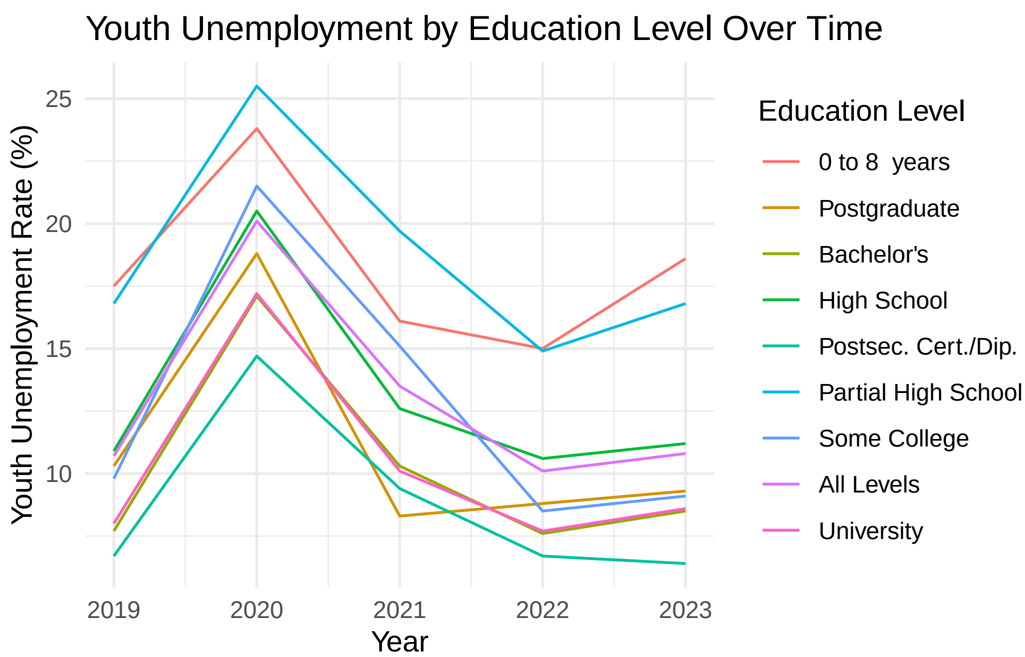

Figure 1: Youth Unemployment Rates

We then used the following formula to calculate the overall percentage change between years:

$$\text{Overall Percentage Change} = \left(\frac{\text{Sum of Rates in X Year} - \text{Sum of Rates in Y Year}}{\text{Sum of Rates in Y Year}} \right) \times 100\%$$

The overall percentage change for the youth unemployment rate is $\approx 0.92\%$ increase between 2019 and 2023 and $\approx 10.45\%$ increase between 2022 and 2023.

4.2 Adult Unemployment

Table 5: Percentage Change in Adult Unemployment by education level (2019-2023)

Education Level	Unemployment Rate (2019)	Unemployment Rate (2023)	Percentage Change (%)
0 to 8 years	9.4	11.3	20.21
Above bachelor's degree	4.4	4.0	-9.09
Bachelor's degree	4.1	4.2	2.44
High school graduate	7.1	7.1	0.00

Table 5: Percentage Change in Adult Unemployment by education level (2019-2023)

Education Level	Unemployment Rate (2019)	Unemployment Rate (2023)	Percentage Change (%)
Postsecondary certificate or diploma	4.4	4.3	-2.27
Some high school	9.9	9.6	-3.03
Some postsecondary	7.1	6.7	-5.63
Total, all education levels	5.1	4.9	-3.92
University degree	4.2	4.1	-2.38

Table 6: Percentage Change in Adult Unemployment by education level (2022-2023)

Education Level	Unemployment Rate (2022)	Unemployment Rate (2023)	Percentage Change (%)
0 to 8 years	8.5	11.3	32.94
Above bachelor's degree	3.5	4.0	14.29
Bachelor's degree	3.8	4.2	10.53
High school graduate	6.5	7.1	9.23
Postsecondary certificate or diploma	4.1	4.3	4.88
Some high school	9.9	9.6	-3.03
Some postsecondary	6.5	6.7	3.08
Total, all education levels	4.6	4.9	6.52
University degree	3.7	4.1	10.81

Again using the same formula as before, the overall percentage change for the adult unemployment rate is $\approx 0.89\%$ increase between 2019 and 2023 and $\approx 9.98\%$ increase between 2022 and 2023.

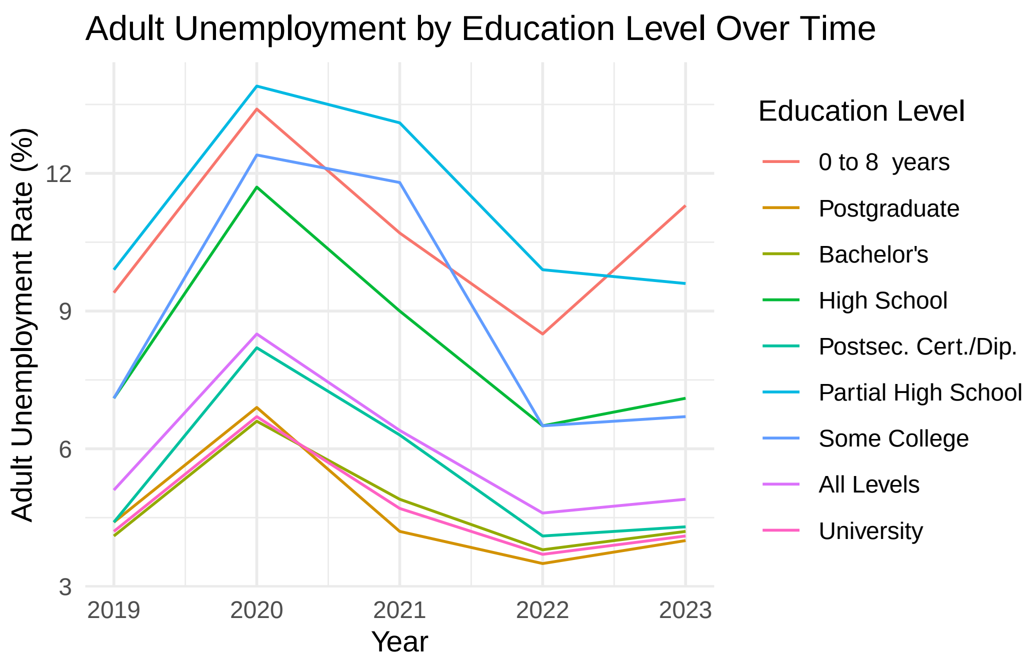

Figure 2: Adult Unemployment Rates

4.3 Senior Unemployment

Table 7: Percentage Change in Senior Unemployment by education level (2019-2023)

Education Level	Unemployment Rate (2019)	Unemployment Rate (2023)	Percentage Change (%)
0 to 8 years	7.8	7.1	-8.97
Above bachelor's degree	3.2	3.6	12.50
Bachelor's degree	3.8	3.7	-2.63
High school graduate	4.8	4.4	-8.33
Postsecondary certificate or diploma	4.6	4.0	-13.04
Some high school	7.0	6.8	-2.86
Some postsecondary	4.8	4.0	-16.67
Total, all education levels	4.6	4.2	-8.70
University degree	3.6	3.7	2.78

Table 8: Percentage Change in Senior Unemployment by education level (2022-2023)

Education Level	Unemployment Rate (2022)	Unemployment Rate (2023)	Percentage Change (%)
0 to 8 years	8.0	7.1	-11.25
Above bachelor's degree	3.3	3.6	9.09
Bachelor's degree	3.7	3.7	0.00
High school graduate	4.9	4.4	-10.20
Postsecondary certificate or diploma	4.3	4.0	-6.98
Some high school	7.6	6.8	-10.53
Some postsecondary	4.4	4.0	-9.09
Total, all education levels	4.4	4.2	-4.55
University degree	3.6	3.7	2.78

The overall percentage change for the senior unemployment rate is $\approx 13.54\%$ decrease between 2019 and 2023 and $\approx 6.11\%$ decrease between 2022 and 2023.

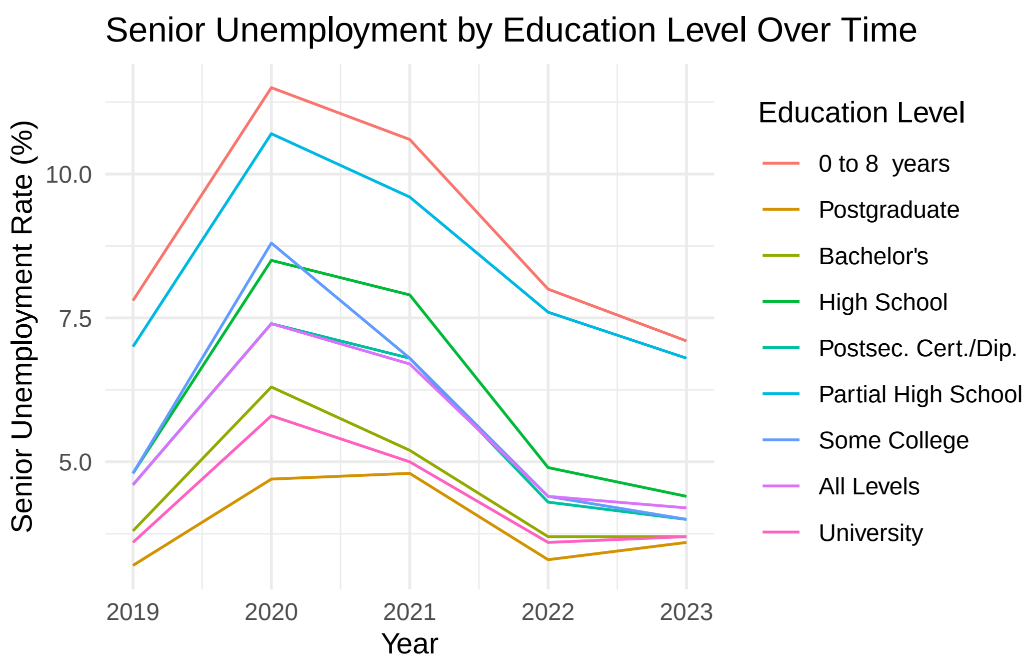

Figure 3: Senior Unemployment Rates

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Brilleman, SL, MJ Crowther, M Moreno-Betancur, J Buros Novik, and R Wolfe. 2018. "Joint Longitudinal and Time-to-Event Models via Stan." https://github.com/stan-dev/stancon_talks/.

Canada, Statistics. 2024. *Unemployment Rate, Participation Rate and Employment Rate by Educational Attainment, Annual*. https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=1410002001.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Neuwirth, Erich. 2022. *RColorBrewer: ColorBrewer Palettes*. https://CRAN.R-project.org/package=RColorBrewer.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Wickham, Hadley, Thomas Lin Pedersen, and Dana Seidel. 2023. *Scales: Scale Functions for Visualization*. https://CRAN.R-project.org/package=scales.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2023. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich

Leisch, and Roger D. Peng. Chapman; Hall/CRC.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.