Jieru Lin, Jiesen Zhang, Yu Xu
A16635527, A16853508, A16247845
COGS 109
June 5, 2023

Problem: During the past year, unemployment has been the topic of discussion, whether it was the waves of layoffs in the tech industry or the worry of ChatGPT taking over people's jobs. The different ways people around us were affected and the debates on the internet inspired us to relook at the issue of unemployment. By looking at the unemployment data from Canada from 1976 to the present, we hope to get a better understanding of the factors that affect unemployment and create a model that is capable of reliably predicting unemployment in the future.

Dataset: Unemployment_Canada_1976_present.csv.

Source:
https://www.kaggle.com/datasets/pienik/unemployment-in-canada-by-province-1976-present
Hypothesis: People in their 50s are more likely to become unemployed

Methods:

There are in total 13 columns in the dataset provided, in which 5 of them we believe we could use to make a model. We can conduct a set of linear regression models using a forward stepwise selection method to determine the best model. In the meantime of doing stepwise selection, we use a k-fold cross validation (e.g. 10-fold) to get the mse of each model and by comparing the mses we choose the best models out of all the models of different parameters. After we find our best model with all the possibly most significant parameters, we can check our model using bootstrap sampling to see what our confidence intervals are and determine how confident we are on our hypothesis.

Choosing models: cross-validation, bootstrap sampling, regularization…

Model estimation: mean square error, p-value, r-square, t-value.

Hypothesis:

In any country, some provinces or states can significantly influence the national economy because of their size, population, or economic output. Consequently, the unemployment rates in these areas can have a substantial impact on the national unemployment rate. With this in mind, we pose the following research question: "Which province's unemployment rate is most closely correlated with the national unemployment rate in Canada, and can we predict national unemployment trends based on this province's unemployment rate?" We expect Ontario's unemployment rate is most closely correlated with the national unemployment rate in Canada, and we can predict national unemployment trends based on its unemployment rate. This is because Ontario is the largest state, and in the biggest state' employment is not stable when people have to fight for the limited positions.