

# Modelling Unemployment and Vaccination Rates

Damneet Thiara, 11170388 - STAT 447C

## Introduction

By April of 2020, the overall unemployment rate in Canada had hit 14.1%, an 8.4 percentage point increase from February of 2020 (Source 1 - StatsCan). This unprecedented increase was just one month after COVID-19 was declared a pandemic by the World Health Organization (WHO). Since then, deaths have been unprecedented, alongside a dramatic downturn in economic activity. By December 9th, 2020, Health Canada had authorized the first COVID-19 vaccine: the Pfizer-BioNTech vaccine (Source 3). Health Canada also reported that the Pfizer-BioNTech Comirnaty COVID vaccine had shown to have been “95% effective in protecting trial participants from COVID-19 for those 16 years and older” (Source 4).

With consumer demand and economic activity spiralling downwards amidst a succession of COVID-19 lockdowns and increasing case/death counts, it begs the question on how COVID-19 vaccines have affected these economic characteristics. Specifically, this paper aims to analyze the relationship between vaccination rates and unemployment rates in Canada from 2020 to 2023. We test the difference between a logarithmic Frequentist model versus a Bayesian model to predict Canadian unemployment rates based on vaccination rates, using time series data.

## Literature Review

Previous studies on employing a Bayesian model to unemployment rates have been conducted by Source 5. By building a hierarchical Bayesian model, Source 5 uses cross-sectional and time series data to estimate US state unemployment rates. Additionally, Source 7 uses a Bayesian analysis to estimate the effects of COVID-19 on key economic indicators, and finds that unemployment increased, interest rates decreased, and prices fell after the first lockdown measures. These effects were then dampened or counteracted by stimulus packages. The increase in unemployment rates here ties in directly with the purpose of this project.

Source 8 conducts a Bayesian analysis on how vaccine misinformation affects vaccination rates, reporting that misinformation can decline intent to vaccinate by up to 6.4 percentage points.

Finally, researchers have explored the relationship between vaccination rates and unemployment rates in many ways already. Source 9 uses an OLS model to find the relationship between vaccination rates and unemployment rates in the US. When analyzing domestic unemployment rates, they find a negative relationship between the two; an increase in vaccination rates is seen alongside a decrease in unemployment rates. However, when analyzing through a state-by-state view, they find that states with higher vaccination rates are more likely to have higher unemployment rates. Although this may seem counter-intuitive, the authors explain this by showing how states with higher vaccination rates experience a higher change in unemployment rates, indeed highlighting the benefits that higher vaccination rates have when it comes to unemployment. Indeed, source 3 also finds the same correlation between states with high unemployment rates and high vaccination rates. Similarly, Source 2 finds that there is also a negative relationship between vaccination and unemployment rates, specifically using instrumental variable analysis to isolate the effects of vaccines by using pharmacy density as a proxy.

The causal relationship can also be reversed. It is shown that a 5 percentage point increase in unemployment rate, a characteristic increase during a recessionary period, decreases the likelihood of a flu shot by 2.3 percentage points (Source 4).

Notably, there is little to no literature on comparing a Frequentist and Bayesian methods to estimate the relationship between vaccination rates and unemployment rates.

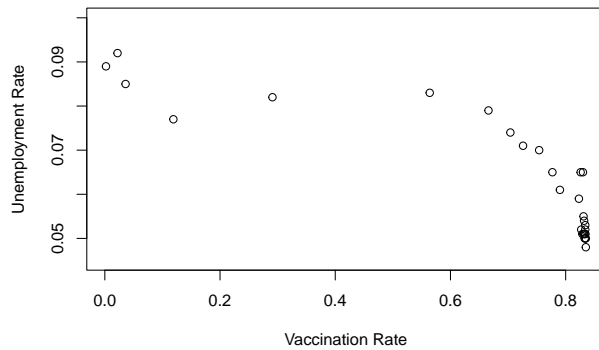
## Analysis

### Data

Based on the literature review, although higher vaccination rates are associated with greater drops in unemployment, this does not necessarily mean that states or provinces with higher vaccination rates will have lower unemployment rates. Since the purpose of this project is to test the differences between a Frequentist and Bayesian model, for simplicity, national time series data for Canadian unemployment rates and vaccination rates is used, rather than dividing it between multiple provinces.

The data for vaccination rates was retrieved from the Government of Canada Health Information Base (Source 6). Previous literature uses one dosage of a COVID-19 vaccine as the benchmark for vaccination rates, so the same standard is used here. Canadian unemployment rates were retrieved from Statistics Canada databases (Source 7). Unemployment rates are reported for an entire month, so the corresponding vaccination rate for each month is the last available vaccination rate available for the end of each month. Some months in 2023 are omitted due to limited vaccination rate data during this time. The final data looks like:

**Vaccination and Unemployment Rates in Canada, 2020 to 2023**



As vaccination rates go up, the unemployment rate that same month goes down.

### Frequentist Model

To begin, we start by building a logarithmic Ordinary Least Squares model:

$$\log(U) = \beta_1 V + \beta_0$$

The fitted model is as follows:

```
##
## Call:
## lm(formula = log(ur) ~ v)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.15709 -0.09830 -0.04948  0.11060  0.23772
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.36841    0.06106 -38.789  < 2e-16 ***
## v           -0.61204    0.08370  -7.312 7.26e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1243 on 27 degrees of freedom
## Multiple R-squared:  0.6645, Adjusted R-squared:  0.652
## F-statistic: 53.47 on 1 and 27 DF, p-value: 7.265e-08
```

## Bayesian Model

Considering the restrictions on unemployment and vaccination rates, being between 0 and 1, a Beta proportion distribution is appropriate. The Beta proportion regression model can be defined as:

$$U \sim \text{BetaProportion}(\mu, \sigma)$$

$$\mu \in (0, 1)$$

$$\sigma \in \mathbb{R}^+$$

$$U \in (0, 1)$$

We define  $\mu$  using an inverse logit function, which takes all real values as inputs and values between 0 and 1 as the output.

$$\mu = \text{inv.logit}(\beta_1 V + \beta_0)$$

The first set of priors for the  $\beta_1$  slope,  $\beta_0$  intercept, and  $\sigma$  parameters are as follows:

$$\beta_1 \sim \text{Normal}(0, 10)$$

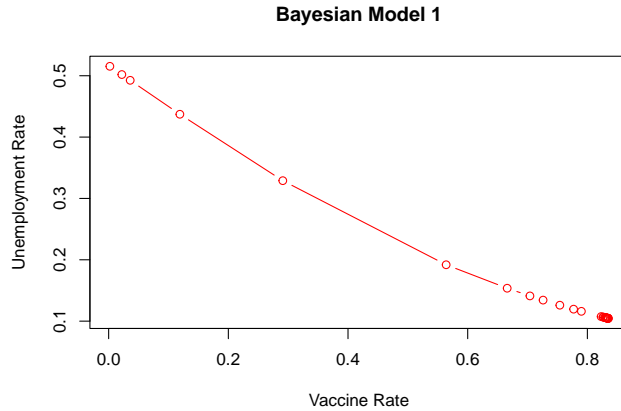
$$\beta_0 \sim \text{Exp}(0.1)$$

$$\sigma \sim \text{Exp}(0.01)$$

The slope parameter is normally distributed to allow for the possibility between positive and negative values. The intercept must be positive since we assume unemployment cannot be below 0, and  $\sigma$  must also be positive, hence the exponential distribution.

The preliminary model produces the following results:

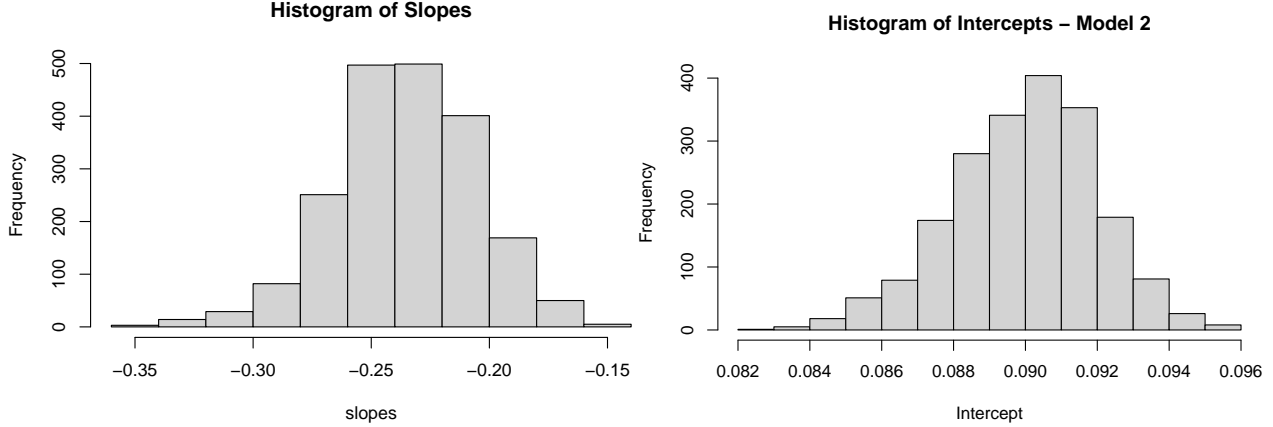
	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
slope	2.69	0.01	0.32	-3.29	-2.90	-2.69	-2.48	-2.06	1046	1
intercept	0.07	0.00	0.07	0.00	0.02	0.04	0.10	0.24	1118	1
sigma	6.15	0.05	1.75	3.25	4.93	5.96	7.18	10.13	1137	1



The unemployment rates are clearly wrong, since the observed data for unemployment falls between 5% and 9%. The issue likely seems to be with the fact that unemployment rate does not vary much and remains at low levels, typically below 10%. Employing a beta prior on this raw data might not be appropriate, since the beta distribution ranges between 0 and 1. In order to combat, the unemployment rate is scaled upwards by a random constant. In this case, scaling upwards by 10 seems appropriate since none of the data for this model falls above 10%, so scaling by 10 will ensure that the data remains within the range of 0 to 1. After doing this, the following results are produced:

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
slope	-2.36	0.01	0.30	-2.99	-2.54	-2.35	-2.15	-1.79	743	1.00
intercept	2.21	0.01	0.23	1.78	2.05	2.21	2.36	2.70	737	1.00
sigma	45.59	0.38	11.47	25.87	37.59	44.66	52.40	70.90	893	1.01

The key observations here are the intercept and the negative slope. The histograms for both are shown below.



To test different priors, I change the model priors to the following:

$$\beta_1 \sim Normal(0, 1)$$

$$\beta_0 \sim Exp(0.01)$$

$$\sigma \sim Exp(0.1)$$

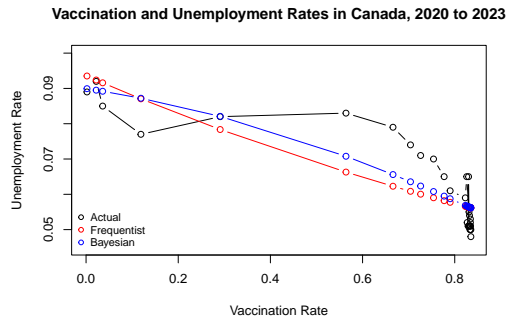
	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
slope	-2.09	0.02	0.33	-2.74	-2.31	-2.08	-1.86	-1.45	460	1
intercept	2.02	0.01	0.25	1.54	1.84	2.01	2.19	2.51	473	1
sigma	34.09	0.31	9.29	18.06	27.55	33.06	40.02	54.46	881	1

The posterior estimates are slightly different but around the same range. Notably, the effective sample size has decreased in this new model, so the first model will be kept, considering the posterior estimates are only slightly different.

## Results

### Frequentist vs Bayesian Model

The model data for both the Frequentist model and the chosen Bayesian model are plotted against the actual data below:



Both look similar, but the root mean squared error can be computed to compare the fit of both:

Frequentist	Bayesian
0.0077369	0.0066725

The RMSE for the Bayesian method is smaller, and thus fits better.

Finally, credible intervals and generated quantity predictions can be used to if the Bayesian model can predict the next missing data point. For each interval, we leave the  $i$ th observation out:

