677 Final Project - Hot, Cold, and COVID

Tim Hogan

5/8/2020

This Write-Up Will be Relatively Unprofessional. I want to make sure it's a little more fun! If you want to mark me down because of this, please feel free to let me know!

Introduction

As all of us know, the spread of COVID-19 has been rampant and wide. It struck suddenly and fast, hardly giving us time to process what was going on. The uncertainty at the beginning made it incredibly difficult to predict how this disease was going to behave, and how we should progress forward.

This did not, however, stop people from speculating. While medical practitioners were conducting their studies, many media outlets began to make claims and their own predictions on how COVID would behave. In this time period, they made a lot of connections to the flu, since it's a common virus in the same group. Of these claims, one of them seemed especially optimistic: that COVID=19 won't survive the summer heat, and would dissipate like the flu does.

The problem with this is twofold. On one hand, this probably isn't how the flu works*, so the reasoning for the mechanism is wrong. Secondly, we see COVID causing havoc all over the globe, including countries like Singapore and the Philippines, which are in the tropics, where it's not cold. While the statement doesn't seem to hold grounds off of this information alone, I wanted to see if it was possible to test its accuracy using statistics. Because of the number of interacting parts and unknowns behind the data, I wanted to make my approach as simple as possible. And in this manner, since we don't have too much data collected from different times, I instead looked to compare a warm country to a cold country, and to test if their spread rates differed.

In this write-up report thing, I'll be documenting how I conducted this analysis, and show the results. This will be in a few short parts.

In part 1, I'll be explaining the countries I looked at, and will show some preliminary results of the data.

In part 2, I'll describe how I pick my distribution and how each country looks.

In part 3, I'll compare the two countries using a simple t-test.

In part 4, I'll assess the problem a little further, using power analysis highlight some of the limitations with this data.

- A note on this, I don't think the mechanism behind the flu season has been confirmed. There are many simple explanations - cold weakening immune system, or people staying indoors together more, or less sun to provide vitamin D. I think the lack of understanding for the mechanism helps show why we may not apply it to COVID, since we can't really understand if the underlying mechanism applies to both cases.

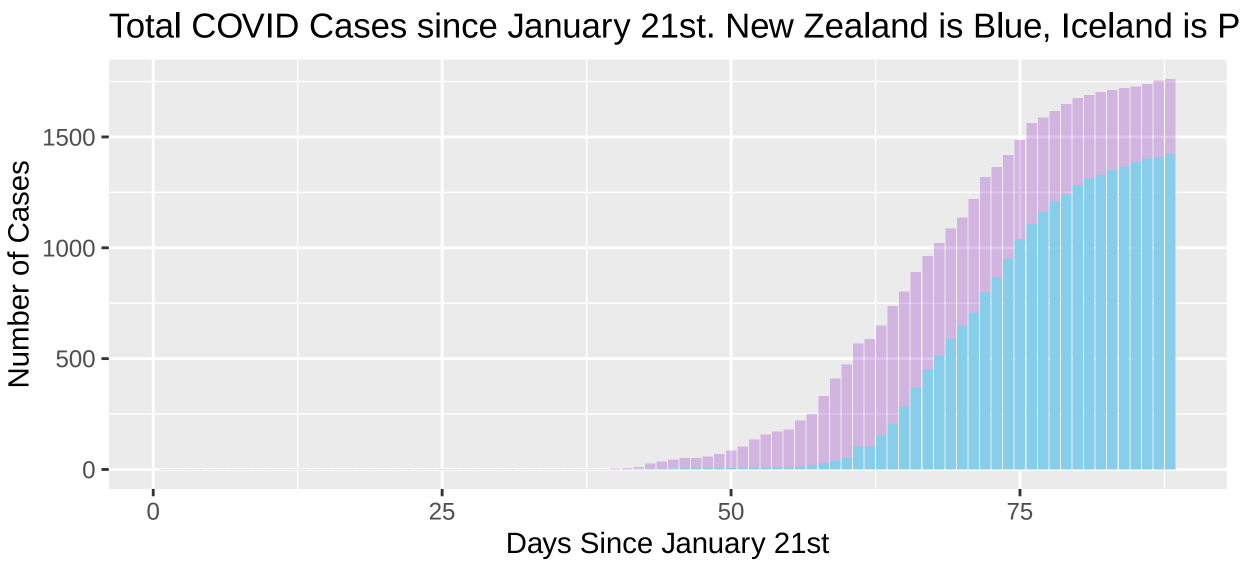
Part 1: Picking Countries

The process of selecting countries is tricky. In this time of COVID testing, there are so many different approaches countries take, leading to some areas of success and some areas of failure. I wanted to make sure I carefully picked the countries I selected, so I wouldn't accidentally capture any differences in population or geography, or simply any difference in competence. I wanted to make sure the countries I selected were similar in terms of how they were handling the test.

Ultimately, the two countries I selected were Iceland and New Zealand. In the time of the pandemic, Iceland was experiencing extremely low temperatures. In Reykjavik in march, the temperature ranged from -0.5 to 6 degrees celsius, making it an ideal cold candidate. On the other hand, many locations in New Zealand ranged from 10 to about 22 degrees celsius. While this isn't the warmest, it's significantly warmer, and can probably act as a comparison point.

Outside of temperature, these countries are also relatively similar. Firstly, these are both island-based countries, which means that the way the disease spreads would be restricted by the ocean in similar ways. Secondly, as far as I understand, these are two locations that were documented for handling testing well. While their populations are different, I wanted to use these as a starting point to investigate the difference in scale.

Next I wanted to compare their spread rates. I used the Confirmed Cases data table from the JHU COVID database, which includes the number of confirmed cases each day from January 22nd to April 18th. I constructed a simple bar chart to help illustrate their spread rates over time, which should be in the future that's nearby this. Hopefully.

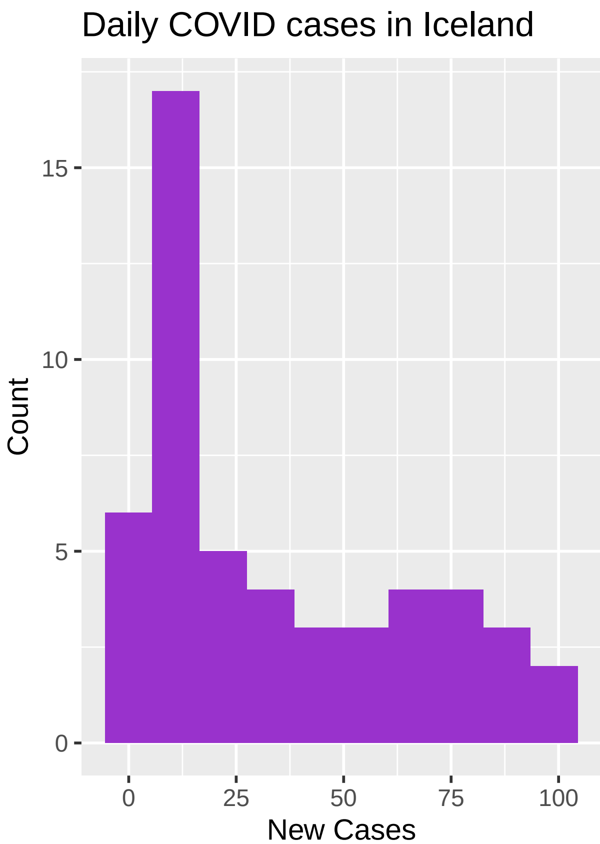


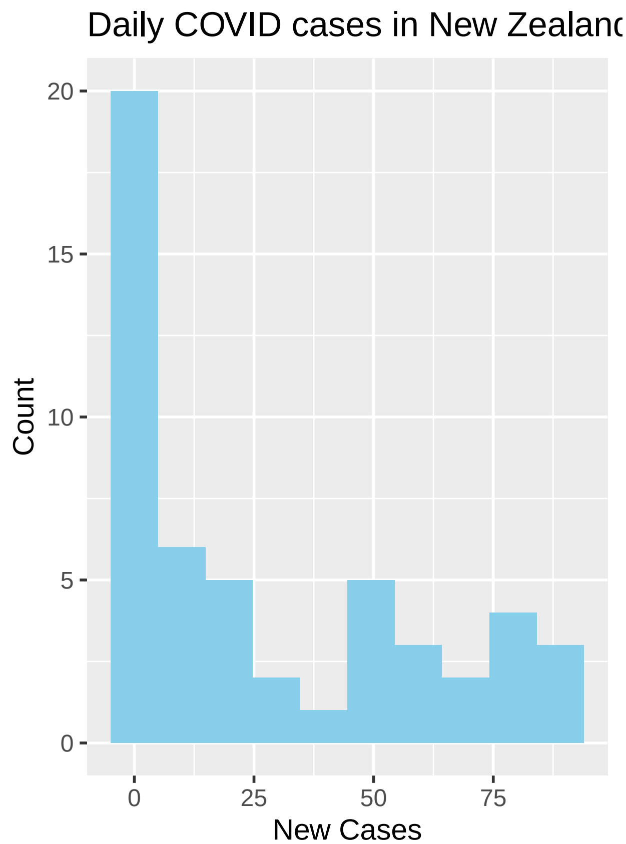
There are a few things to be noted with this. Firstly, it looks like the New Zealand cases are slightly behind the Iceland cases. Secondly, it should also be noted that these two places seem to share very similar patterns. This could be what's expected for good places that experience COVID, but this could also be unusual. If it's unusual, then this may not be the best comparison. However, I will still proceed!

Part 2: Comparing Countries.

Looking at the first figure again, what can be noted is that the pattern may be familiar. It actually appears to take shape of some kind of cumulative distribution. Looking at the problem a little more closely, we're basically counting individuals who have COVID. Based on this, the process can be described as a poisson process, and also psosibly a negative binomial process.

To look closer, I displayed the new cases of Iceland and New Zealand in the graph below. Based off of this, we can see a high initial peak closer to 0, and then lower. Secondly, as expected, these distributions also do look pretty similar. However, the New Zealand cases seem to have slightly fewer samples and seem to be more centralized around 0. This high initial peak and low number afterwards could also help support it being a Poisson distribution.



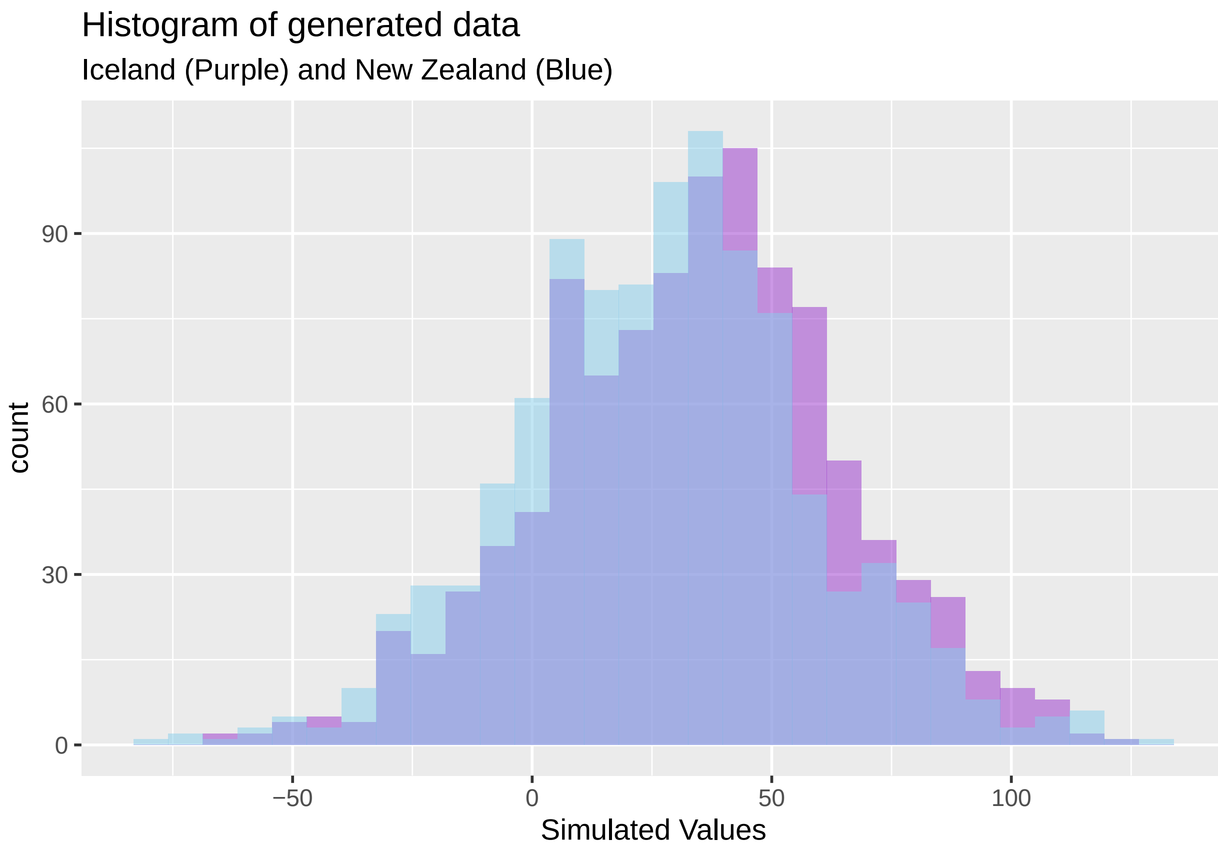


We can also check by comparing their means and variances. In a poisson distribution, the mean and variance should be about the same. Iceland's daily COVID cases has a mean of 34.51 and variance of 30.44. While not exact, they are relatively close, which may support it being a poisson distribution. Comparatively, New Zealand has a mean of 27.88 and variance of 31.13. The variance is close to Iceland's, but the mean seems much lower. New Zealand's mean also seems lower than its variance, which could be a sign of overdispersion. This does make sense - in the first week, New Zealand has very few new COVID cases. This is most likely the system counting its initial patient and waiting for symptoms to begin appearing among others. For this comparison, I will not adjust for overdispersion, but this is an additional step that can be taken. For reference, taking out the first week increases New Zealand's mean to about 31.

Comparison

If we compare them above two distributions, visually they don't appear to be very different. Also, based on their means and variances, they probably aren't very different. Below I plotted normal distributions using their above means and variances - since a poisson distribution eventually converges to a normal distribution, this is theoretically appropriate.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Basically, what we can see here is that the simulations are very hard to distinguish. In any hypothesis test we'll do will likely not distinguish these to distributions. This is reflected by the result of a simple t-test.

```
##
##  Welch Two Sample t-test
##
## data:  icedata and nzdata
## t = 1.0871, df = 99.95, p-value = 0.2796
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.467484 18.722386
## sample estimates:
## mean of x mean of y
##  34.50980  27.88235
```

The t-test results support this claim. Based off this p-value, we don't have enough evidence to reject the null hypothesis. Therefore, given the COVID counts provided, our data does not seem to support the claim that New Zealand and Iceland have different spread rates.

Discussion

It should be reiterated that this test is pretty simple and straightforward, and isn't free of problems. This could be capturing some kind of underlying similarity between New Zealand and Iceland that suppresses the temperature effect. Secondarily, we may not have enough samples to actually properly assess these distributions, and more data may need to be collected for a proper comparison to be made.

Additionally, we may need more time and some more comparisons to eventually see if this trend exists across other countries. We saw that New Zealand and Iceland have similar spread rates, and we can use that to try to support that temperature may not have too much of an impact, but we may see very different patterns comparing other locations.

Appendix

Posterior Analysis.

Using a