COVID-19 in Latin America

UNC Gillings School of Global Public Health

Jesus E. Vazquez, PhD Student Biostatistics

3/18/2020

Introduction

The purpose of this analysis is to determine the extent at which COVID-19, better known as Coronavirus, has progressed in Latin America. Not only is monitoring COVID-19 important to help local and federal Latin America governments make better informed decisions, but it is neccesary due great amount of social disparities that will tremendously worsen the impact of the virus in the health of these communities. In this report we will present initial statistics and graphical illustrations of the current development of COVID-19 and future projections in Latin America.

Methods

Scatterplots with the counts of cases will be use to illustrate the progression of the virus, with a comparison to a poisson distribution derived by the current trends. Indexes regarding **airline prices, availability and cancelations, food & gas availability/prices** as well as the own progression of the virus will be used to predict and forecast the spread of the virus using time-series analysis and multiple linear regression.

Results

Importing, Cleaning Data, and Generating Plots

Data was provided by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE), ESRI Living Atlas Team and the Johns Hopkins University Applied Physics Lab (JHU APL) and update these data daily from information provided by the WHO, US and China CDC, among other worldwide organizations. github

```r
library(tidyverse)
library(reshape2)

# Reading and subsetting data
mydfcases <- read.csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_da
  # subset(Country.Region == c('Peru')) %>%
  melt(id.vars=c("Province.State", "Country.Region", "Lat", "Long")) %>%
  rename("Area"= "Province.State", "Country"="Country.Region", "Count" = 'value', 'Date' = 'variable')
  mutate(Date = sub('X', '', Date)) %>%
  mutate(Date = as.Date(Date,"%m.%d.%Y"))
```

```r
library(ggplot2)

# Creating Timeseries Plot
mydfcases %>% subset(Country %in% c("Mexico", "Peru")) %>%
  ggplot(aes(x=Date, colour = Country, y=Count, label = Count)) +
  geom_line() +
```

```r
geom_text(aes(label=ifelse(Count>50,as.character(Count),'')),hjust=0,vjust=0) +
theme_minimal() +
labs(y="Number of Cases", x='', title = 'Number of COVID-19 Reported Cases, Latin America')
```

Number of COVID−19 Reported Cases, Latin America
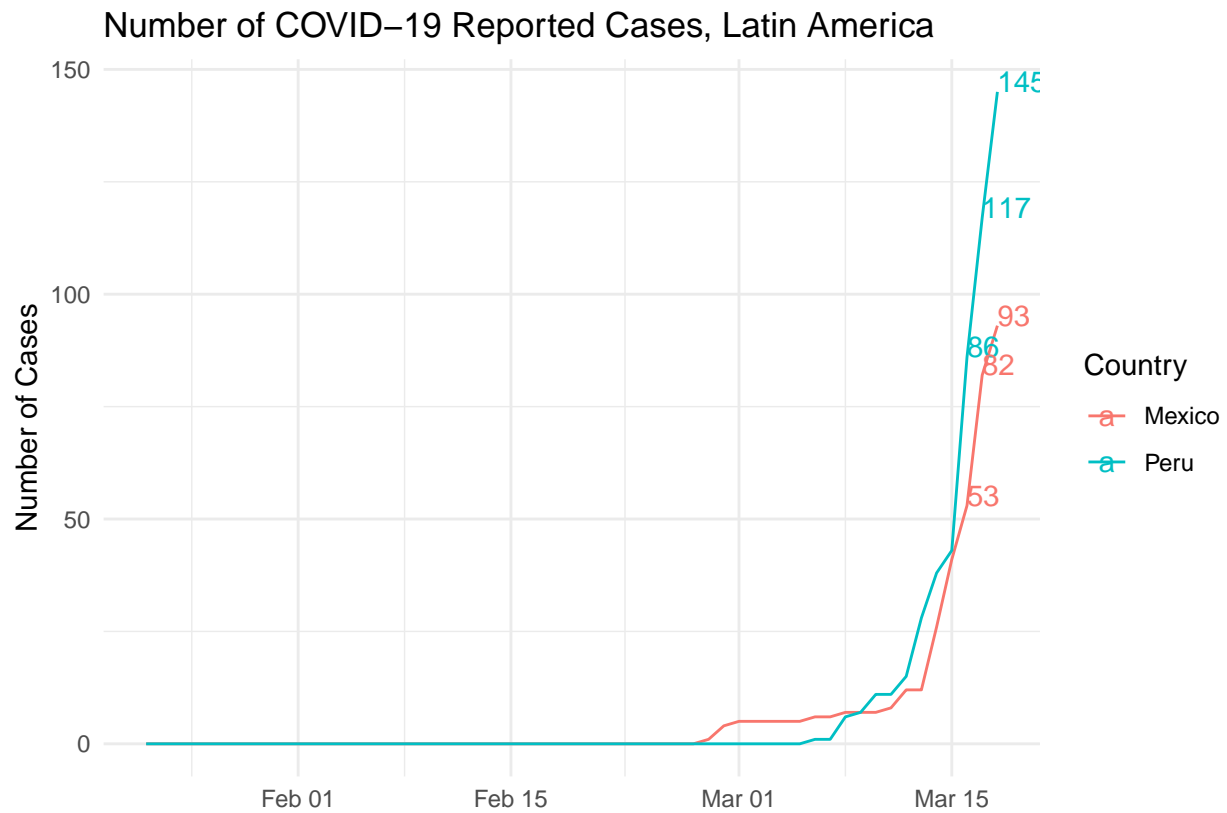

```r
# Comparing againts Poisson Random Variable
peru = mydfcases %>% subset(Country %in% c("Peru") & Count > 0)
num_days = as.numeric(max(peru$Date) - min(peru$Date))
num_total_cases = max(peru$Count)
lamba_hat = num_total_cases/num_days
peru$expected = ppois(0:num_days, lambda = lamba_hat)*num_total_cases

# Plotting Agaist Distribution
ggplot(peru, aes(x=Date)) +  theme_minimal() +
  geom_text(aes(x = as.Date("20-03-18"), y = 120, label = "Actual")) +
  geom_text(aes(x = as.Date("20-03-18"), y = 70, label = "Expected")) +
  geom_line(aes(y=expected), colour="red") +
  geom_line(aes(y=Count), colour="green")  +
  labs(title = 'Expected vs Actual COVID-19 Cases, Peru', subtitle = 'Approximation with Poisson Model'
```
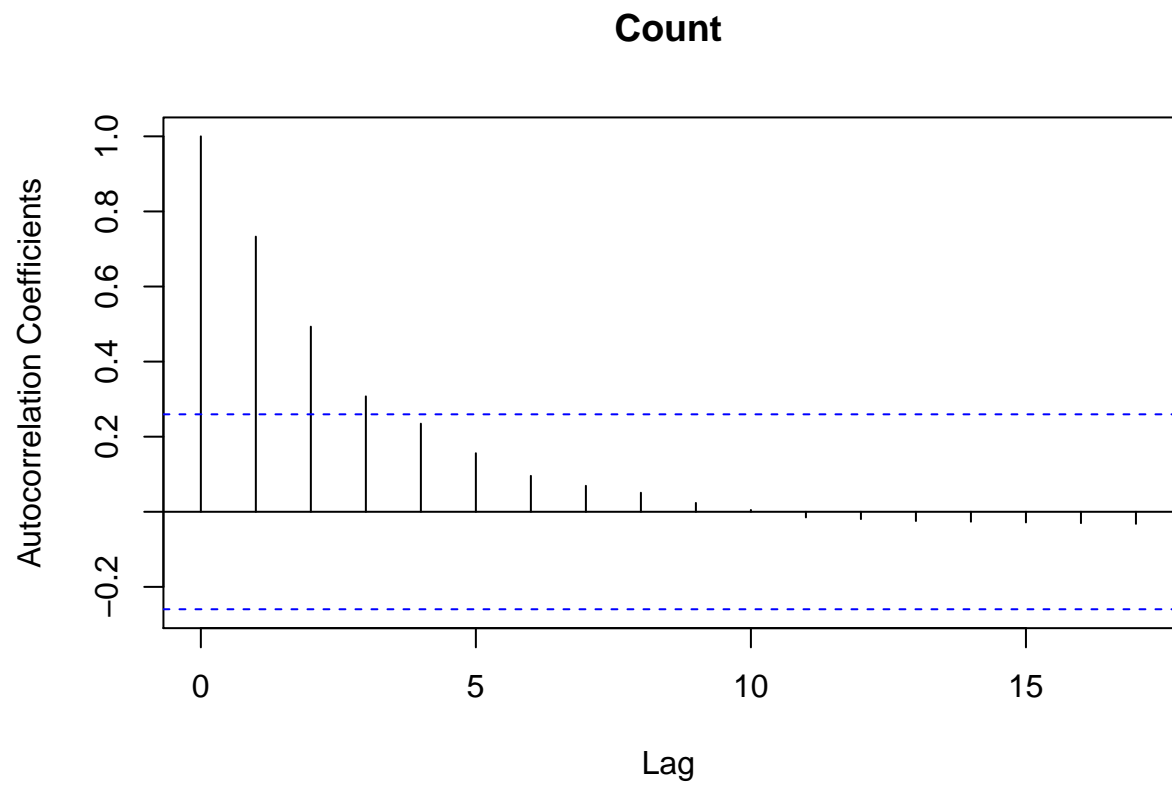
Expected vs Actual COVID−19 Cases, Peru
Approximation with Poisson Model
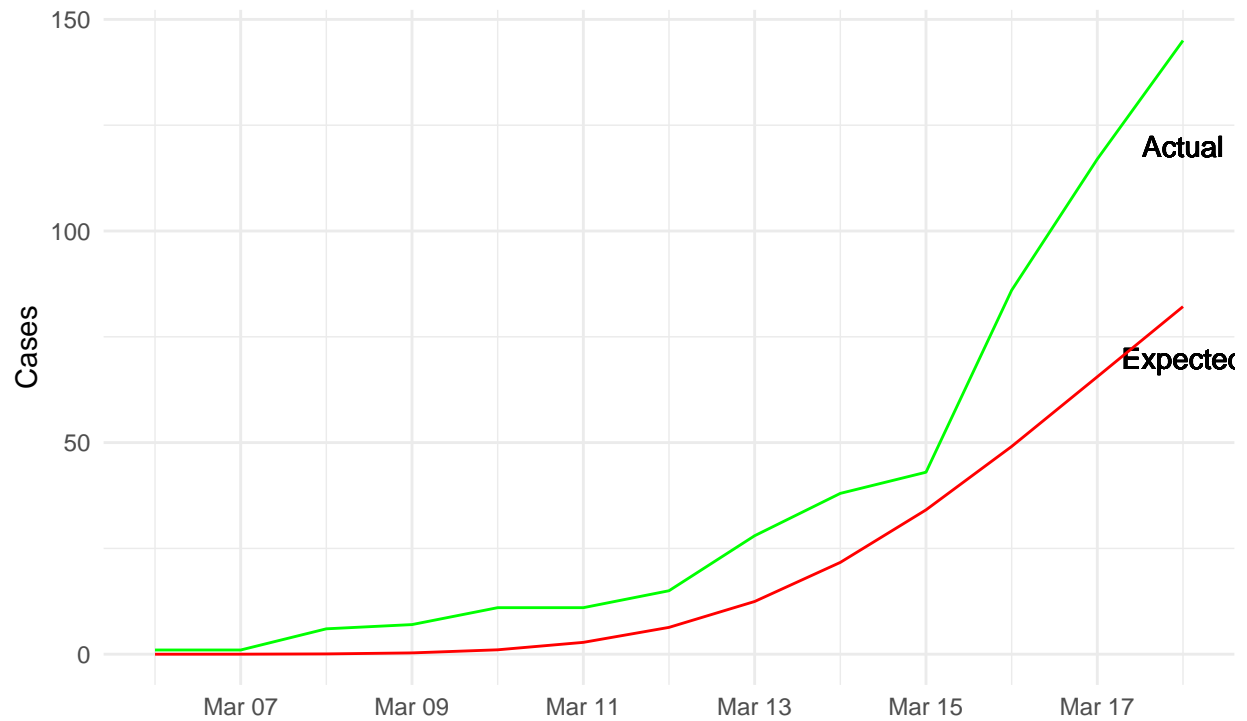
Model Training and Prediction (in progress)

```r
# Timeseries
library(forecast)
mydfcases %>% subset(Country == "Peru", select = 'Count' ) %>% acf(ylab = 'Autocorrelation Coefficients
```

Count


```r
mydfcases %>% subset(Country == "Peru", select = 'Count' ) %>% pacf(ylab = 'Partial Autocorrelation Coe
```

Series .