Data Fest- Unemployed persons by industry, class, and sex

Brandon Zhao

5/13/2020

Introduction to the Data

Throughout the COVID-19 Crisis, it can be clearly seen that unemployment is rising in general, but it may not be entirely apparent clear at first whether unemployment for certain industries is rising faster than others. Here, we will analyze which industries are suffering the most unemployment in the United States by comparing current unemployment levels by industry to unemployment levels in the previous year.

Stay at home orders began at different times depending on the state, but generally began at the end of March/beginning of April (Source: https://www.washingtonpost.com/health/2020/04/06/coronavirus-stay-at-home-by-state/ (https://www.washingtonpost.com/health/2020/04/06/coronavirus-stay-at-home-by-state/)). Therefore, we are assuming COVID-19 began impacting unemployment rates starting in April 2020, so we will be comparing unemployment data by industry between April 2020 and April 2019. The following dataset was taken from the US Bureau of Labor Statistics and formatted into a csv file. The first six entries are shown below. (Source: https://www.bls.gov/web/empsit/cpseea31.htm (https://www.bls.gov/web/empsit/cpseea31.htm))

```
unemployment <- data.frame(read.csv("Data Fest Dataset.csv", stringsAsFactors = FALSE))
colnames(unemployment) = c("Industry of Worker",
                           "Total Unemployed April 2019 (in thousands)",
                           "Total Unemployed April 2020 (in thousands)",
                           "Total Unemployment Rate April 2019",
                           "Total Unemployment Rate April 2020")
head(unemployment)
```

```
##                      Industry of Worker
## 1                          Construction
## 2           Nonmetallic mineral products
## 3 Primary and fabricated metal products
## 4                Machinery manufacturing
## 5       Computer and electronic products
## 6   Electrical equipment and appliances
##   Total Unemployed April 2019 (in thousands)
## 1                                        439
## 2                                         26
## 3                                         55
## 4                                         35
## 5                                          8
## 6                                          1
##   Total Unemployed April 2020 (in thousands) Total Unemployment Rate April 2019
## 1                                      1,531                                4.7
## 2                                         38                                5.1
## 3                                        198                                3.1
## 4                                        166                                2.7
## 5                                         35                                0.8
## 6                                         34                                0.3
##   Total Unemployment Rate April 2020
## 1                               16.6
## 2                               10.0
## 3                               11.3
## 4                               13.3
## 5                                3.5
## 6                                8.2
```

Analysis

In order to compare the unemployment between different industries, we will be using their unemployment rate, as each industry's total unemployed varies heavily based on the total number of people in each industry. First we will calculate the increase in unemployment rate of each industry between April 2019 and April 2020, and then we will sort the industries by increasing change in unemployment rate. We will display the six industries with the lowest incrase in unemployment rate, and the six industries with the largest increase in unemployment rate.

```
unemployment$Rate_Change = unemployment$`Total Unemployment Rate April 2020` - unemployment$`Total Unemployment Ra
te April 2019`
unemployment <- unemployment[order(unemployment$Rate_Change),]
head(unemployment$`Industry of Worker`)
```

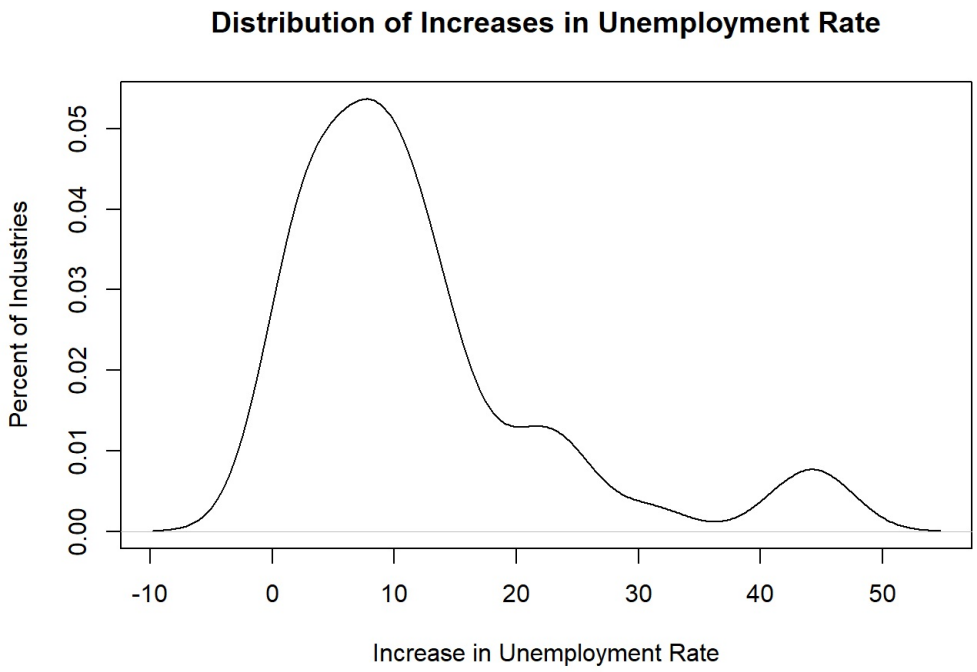
```
## [1] "Publishing, except Internet"    "Finance"
## [3] "Utilities"                      "Broadcasting (except internet)"
## [5] "Insurance"                      "Hospitals"
```

```
tail(unemployment$`Industry of Worker`)
```

```
## [1] "Textile, apparel, and leather"       "Rental and leasing services"
## [3] "Food services and drinking places"   "Arts, entertainment, and recreation"
## [5] "Accommodation"                       "Personal and laundry services"
```

Now, we will show the distribution of increases in unemployment rates, as well as a summary of the increases in unemployment rates.

```
plot(density(unemployment$Rate_Change),
     xlab = "Increase in Unemployment Rate",
     ylab = "Percent of Industries",
     main = "Distribution of Increases in Unemployment Rate")
```




```
summary(unemployment$Rate_Change)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.600   4.075   8.350  11.943  13.900  45.500
```

On average, the the unemployment rate of each industry increased by 11.943%. The distribution of increases in unemployment rate is centered approximately at the median, 8.35, and is skewed right. This means that there are many more industries with a much higher increase in unemployment rate than the median than there are industries with a much lower increase in unemployment rate than the median.

Now, in order to find industries with much less and much more increase in unemployment rate, we will be finding those industries whose increase in unemployment rate is more than one standard deviations away from the mean, as well as seeing if any are more than two standard deviations away from the mean.

```
lower_change = unemployment[unemployment$Rate_Change < (mean(unemployment$Rate_Change) - sd(unemployment$Rate_Chan
ge)),]
lower_change$`Industry of Worker`
```

```
## [1] "Publishing, except Internet"
```

```
greater_change = unemployment[unemployment$Rate_Change > (mean(unemployment$Rate_Change) + sd(unemployment$Rate_Ch
ange)),]
greater_change$`Industry of Worker`
```

```
## [1] "Textile, apparel, and leather"       "Rental and leasing services"
## [3] "Food services and drinking places"   "Arts, entertainment, and recreation"
## [5] "Accommodation"                       "Personal and laundry services"
```

```
greatest_changes = unemployment[unemployment$Rate_Change > (mean(unemployment$Rate_Change) + 2 * sd(unemployment$Rate_Change)),]
greatest_changes$`Industry of Worker`
```

```
## [1] "Arts, entertainment, and recreation" "Accommodation"
## [3] "Personal and laundry services"
```

Out of all industries, only the Non-Internet Publishing Industry, a subset of the Information Industry, had an increase in unemployment rate lower than one standard deviation below the mean increase in unemployment rate. In fact, it had actually had a decrease in unemployment rate.

Out of all industries, six industies have increases in unemployment rate greater than one standard deviation above the mean increase in unemployment rate. They are Textile, apparel, and leather; Rental and leasing services; Food services and drinking places; Arts, entertainment, and recreation; Accommodation; and Personal and laundry services. Out of those, three industries have increase in unemployment rate greater than two standard devations above the mean increase in unemployment rate. They are Arts, entertainment, and recreation; Accommodation; and Personal and laundry services.

Summary of Findings

Unemployment rates have increased in April 2020 because of the stay-at-home-orders that resulted COVID-19 Pandemic. On average, unemployment rates increased by approximately 12 percent in every industry when comparing between April 2019 and April 2020. However, the unemployment for some industries did not change as much as others. Specifically, the unemployment rate for the Non-Internet Publishing industry actually saw a decrease in unemployment rate between April 2019 and April 2020. Also, even more industries saw unemployment rate increases that were even greater than the average increase. The Textile, apparel, and leather; Rental and leasing services; and Food services and drinking places industries saw much greater increases in unemployment rates, and the Arts, entertainment, and recreation; Accommodation; and Personal and laundry services industries saw the greatest increases in unemployment rates.