GPH-GU 2338: Project proposal

Sooyoung Kim

Jessica Randazzo

**Problem statement**

The World Health Organization (WHO) declared COVID-19 as a Public Health Emergency of International Concern (PHEIC) on January 30, 2020 (World Health Organization, 2020). Until now, the COVID-19 pandemic continues to cause significant global disruption across sectors ranging from healthcare to education to the economy. Several studies from different countries suggested that the burden of the pandemic is disproportionately affecting the vulnerable population both in economic and health perspective (Ataguba, 2020; Grooms;, Ortega;, & Rubalcaba, 2020; Mottaleb, Mainuddin, & Sonobe, 2020; Qian & Fan, 2020), however there is still a research gap in the United States to identify these population timely and clearly, and to quantify the economic burden or loss they’re suffering. Using the close-to-real-time data, we aim to identify the vulnerable sub-population that are significantly affected by COVID-19 pandemic with the aim of better allocating the resources to aid the economic recovery.

**Description of data**

We will use the US Census Bureau’s weekly Household Pulse Survey from Wave 3 (Oct 2020 – Mar 2021)(US Census Bureau, 2021). This weekly survey is conducted by the US Census Bureau in collaboration with other federal agencies to identify the economic and social impact of the COVID-19 pandemic on American household. The survey is currently in its third wave and contains survey weights to make the results nationally representable. We believe it will capture those suffering the most from the pandemic due to the recent lack of government relief and assistance. The ongoing and close-to-real-time dissemination of this survey until March 2021 will enable us to reflect the most recent social and economic issues related to the pandemic, as well as to include longitudinal predictors if needed.

**Proposed approach and methods**

We will focus on the phase 3 data of the survey, starting from the waves in January 2021, as they contain variables that explores several aspect of the economic impact, namely food insecurity, rent/mortgage payment, and the stimulus check usage, We will recode the survey responses to create binary response variables corresponding to each of the aspect. We will then use one of the classification methods we are taught during the class to predict the outcomes based on the socio-demographical characteristics of the respondents.

**Table 1. Example of binary response variables**

|  |  |
| --- | --- |
| Aspect of economic vulnerability | Variable description |
| Food insecurity | Often not had enough food to eat (Y/N) |
| Housing vulnerability | Household currently not caught up on rent/mortgage payment (Y/N) |
| Stimulus check usage | Spent most of the stimulus check on essential items (i.e. food, household supplies or personal care products, rent, utilities and telecommunications, credit card, student loans or other debts) (Y/N) |

**Evaluation plan**

* Based on the knowledge acquired from the literature review, data will be examined and recoded to create a set of response variables and to choose the relevant predictor variables.
* We will randomly divide the dataset into training set, validation set, and the test set in approximately 50-25-25 ratio.
* Using the various classification methods, including logistic regression, linear/quadratic discriminant analysis, KNN and etc, we will fit the model on the training set with sociodemographic characteristics variables as our predictors, and abovementioned binary variables as responses.
* We will evaluate the model using various methods, including the confusion matrix, to measure training error rate.
* We will also use cross-validation to choose our final model using the validation set which yields the least validation error rate.
* We will then test the final model(s) to the test set to evaluate its performance.

**Timeline**

Total estimated time needed is 4 – 6 weeks. However, we expect a significant overlap in timelines for many of the tasks below, which will result in shorter timeline than 6 weeks.

* Literature review & draft proposal – 1~ 2 weeks
* Data cleaning & preparation – 1 week
* Perform ML analysis – 1~1.5 weeks
* Write-up of final reports – 1~1.5 weeks

**What each group member will do**

* Both will equally contribute to the entire process as suggested in the timeline.

**References**

Ataguba, J. E. (2020). COVID-19 Pandemic, a War to be Won: Understanding its Economic Implications for Africa. *Appl Health Econ Health Policy, 18*(3), 325-328. doi:10.1007/s40258-020-00580-x

Grooms;, J., Ortega;, A., & Rubalcaba, J. A.-A. (2020). The COVID-19 public health and economic crises leave vulnerable populations exposed. Retrieved from <https://www.brookings.edu/blog/up-front/2020/08/13/the-covid-19-public-health-and-economic-crises-leave-vulnerable-populations-exposed/>

Mottaleb, K. A., Mainuddin, M., & Sonobe, T. (2020). COVID-19 induced economic loss and ensuring food security for vulnerable groups: Policy implications from Bangladesh. *PLoS One, 15*(10), e0240709. doi:10.1371/journal.pone.0240709

Qian, Y., & Fan, W. (2020). Who loses income during the COVID-19 outbreak? Evidence from China. *Research in Social Stratification and Mobility, 68*. doi:10.1016/j.rssm.2020.100522

US Census Bureau. (2021). *Measuring Household Experiences during the Coronavirus Pandemic*. Retrieved from: <https://www.census.gov/data/experimental-data-products/household-pulse-survey.html>

World Health Organization. (2020). COVID 19 Public Health Emergency of International Concern (PHEIC)

Global research and innovation forum: towards a research roadmap. Retrieved from <https://www.who.int/publications/m/item/covid-19-public-health-emergency-of-international-concern-(pheic)-global-research-and-innovation-forum>