Factors in Vaccination Rates - Deliverable 1

Albert Fung
Jeffrey Ng

Contents:

1. Questions Relevant to Project Proposal (that we examined in our *preliminary analysis*)

1. Are vaccination rates influenced by health expenditures in the United States? We examine the effects of state health expenditures on state vaccination rates. *See below for preliminary analysis results.*
2. Does having 'looser' state exemptions (more leeway for individuals to opt out of vaccinations) actually have an impact on state vaccination rates? *See below for preliminary analysis results.*

2. Preliminary Analysis:

(2.11) Introduction/Data (health expenditures)

For our preliminary analysis, we examined the effects of health expenditures (by state) on the vaccination rates (also by state). To do this, we ran a regression (linear) in R[1] (for our purposes, this was easier to do than its counterpart in Python - especially for data cleaning). We used the coefficients to investigate whether health expenditures had a significant effect on the vaccination rates in different states.

To do this, we imported multiple datasets (uploaded to github under "data") and merged them based on states (since our analysis was focused on differences in states health expenditures. The dataset that was used for health expenditures aggregated the state by state health spending under a variable named "Total Health Spending", which is defined as, "spending for all privately and publicly funded personal health care services and products (hospital care, physician services, nursing home care, prescription drugs, etc.) by state of residence. Hospital spending is included and reflects the total net revenue (gross charges less contractual adjustments, bad debts, and charity care)"[2]

[1] For the regression and R code, see the pull request for the file labelled "deliverable_1.Rmd"
[2] Note that costs such as " insurance program administration, research, and construction expenses are not included in this total"

Furthermore, as part of our data-wrangling process, we noticed that an analysis using "Total Health Spending" could potentially be skewed by population since different states have vastly different populations, which affects the spending. To account for this, we added a population csv (2019 estimate) and added a new variable for the state by state health spending *by state capita*. With this, we ran a linear model in R (again, see the file in the Deliverable 1 folder named "deliverable_1.Rmd" for our model and the model results) using health spending by capita as our independent variable (x) and vaccination rates as our dependent variable (y).

(2.12) Preliminary Analysis - Results (health expenditures)

We found that ($p < .05$, which indicates significance), but interestingly, the estimate for the coefficient was so low (.0065) that we believe the null hypothesis (that health spending per capita does not have a major impact on vaccination rates) *cannot be rejected*. Combining the coefficient value and p-value, we essentially found that *it was significant that health care spending did not have any serious influences on vaccination rates.* While this may sound counterintuitive at first, there are a lot of reasons why this can be the case. For example, one potential reason is simply that vaccinations are mandated and covered by most insurances (or perhaps even cheap without insurances), so changing the spending won't change the vaccination rate.

This definitely warrants a deeper look, which is outside the scope of this preliminary analysis. We will likely do hypothesis testing to confirm that the null hypothesis cannot be rejected (rather than just "eyeballing" the p-values and coefficients estimates of the regression.

(2.21) Introduction/Data (health expenditures)

There are three typical categories when it comes to vaccination exemptions: medical exemptions, religious exemptions, and personal belief exemptions (which may overlap with religious exemptions). Our goal was to examine whether these exemptions had an impact on vaccination rates. Specifically, do states that allow for more exemptions have lower vaccination rates (more people claiming exemptions simply because they can).

This was a fairly simple (not deep) examination. We used datasets that classified the type of exemption allowed in each state (categorical data) and examined the differences in vaccination rates by comparing the rates in each category with each other. A box plot was created (also in the .Rmd file) for the states in each category and the resulting graph was examined as part of our analysis.

(2.22) Preliminary Analysis - Results (health expenditures)

By convention, box plots use medians. In our resulting box plot (***which is also available in the html file in our Deliverable 1*****),** *we noticed that the median vaccination rates were the lowest for states that allowed for personal belief exemptions*. On the other hand, vaccination rates *were highest for states in the category of just medical exemptions*.

These results fall in-line with our expectations. States that only allow for medical exemptions are the most "strict" in terms of who can opt-out of vaccinations, whereas states that allow for personal exemptions are the least "strict" since the idea of personal belief can be widely interpreted. So intuitively, this makes sense - laxer rules would mean less vaccinations. This could potentially have widespread implications and may help shape future vaccination policies (e.g. states with lower vaccination rates may want to consider examining and changing their current exemption policies to lower the potential for future outbreaks as well as lower virus cases in general).

These results are from a precursory assessment of the data and we may want to research this on a more granular level following this preliminary inspection.

(2.3) Later in the semester:

As we move further through the project, we will bring in more datasets (as listed below in the **"Data"** section) and either examine them independently (their impacts on vaccination rates) or aggregate them with other analysis (for example, maybe using multivariate regressions) depending on the type of data and their relationships with each other.

3. Data[3][4]

All of the data we have collected is contained within the "data" folder.

All of the data presented here is in its raw form, except for the "population_by_state.csv" and "vaccinationrate_and_exemption.csv" files.

Also included within the "data" folder is another folder called "data_for_introduction", which either contains data for our write-up or contains data/information that may be used a reference, but won't be a part of any in-depth analysis.

---

[3] While this contains all the data that we may believe to be relevant, it is very possible that we'll bring in more data as the project progresses throughout the semester. Conversely, we may not use every single dataset in the folder.

[4] For data sources, please visit our Deliverable 0. It includes a list of all our sources (which we will update as more are added)