Covid positivity at U-M

Feb. 05, 2021 Youngwoo Kwon

The covid positivity rate is a percentage of coronavirus tests performed that are positive. Positivity rate helps people to estimate the covid infected rate and number of tests that are needed because a high positivity rate means there are too many positive tests or too few total numbers of tests. World Health Organization suggests countries lower the restrictions when the covid positivity rate is below 5% for two weeks.[1] In this paper, the data from the University of Michigan Covid-19 data will be used to find 1) What is the 95% confidence interval of the covid positivity rate in University of Michigan each week? 2) Does the positivity rate related to the number of people tested and the number tested of people in past two weeks? 3) Are positivity rates constant within four-time intervals but different across two of these intervals? The data is 19 weeks data consist of covid positivity rate and some people tested each week. From the interpretation, the covid positivity rate had a meaningful relationship between the number of people tested in the last two weeks, and we cannot say that positivity rates are constant within four-time intervals. Also, even if we add the intervals into two intervals, we still cannot say it was constant.

Instead of the standard interval, Wilson interval was used to avoid a risk of getting a lower coverage interval than expected. For the relationship between positivity rate and number of people tested/number of people tested in the past two weeks, linear model (positivity rate) ~ (number of people) and correlation has been used. Moreover, we filtered the data into two sets – where the number of tests was larger than 5000 and lower than 5000 and compared the mean positivity rate for each set. Also, for the intuitive interpretation, confidence interval and the relationship between two variables were plotted on the graph. Four-time intervals were made to answer the third question. Each interval contains 4~5 weeks, and two merged intervals contain 9 and 10 weeks. A likelihood ratio test was done for each interval to check whether each interval has the same positivity rate or not. It checks the null hypothesis $H_0 : p = p_{mle}$ for each week where p_{mle} is the mean value of the positivity rate in the given interval. The formula for the Wilson interval and likelihood ratio test was from the other paper.[2]

Figure 1 shows the Wilson confidence interval for each week. We can check that the positivity rate usually moves between 0.02~0.06 during 19 weeks, and the range of the positivity rate gets closer when the positivity rate gets lower.

Figure 2 represents the relationship between the number of people tested and the positivity rate for each week. We can check the negative relationship between the two variables. However, the correlation seems quite low and the distribution of points is not enough to claim there exists a true negative relationship. More specifically, the correlation of two variables n_tests, the number of people tested, and pos_rate, the positivity rate for that week were -0.184 and the slope of the linear model (positivity rate) ~ (number of people) were $-1.138 * 10^{-6}$. Therefore,

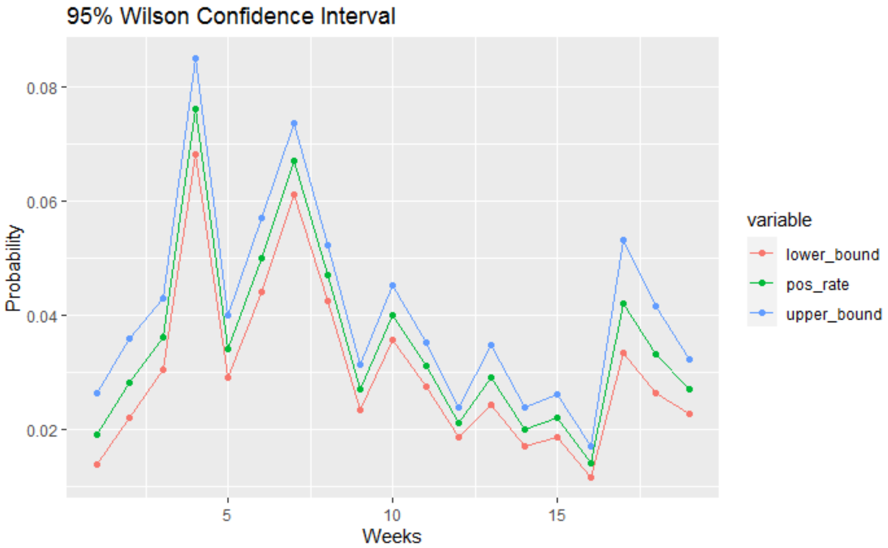

Figure 1. 95% Wilson confidence interval

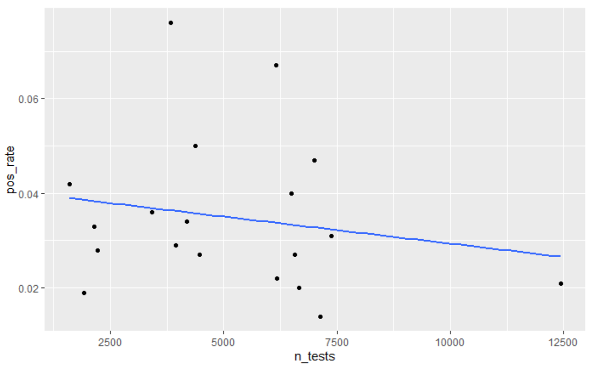

Figure 2. Number of people tested and positivity rate for each week

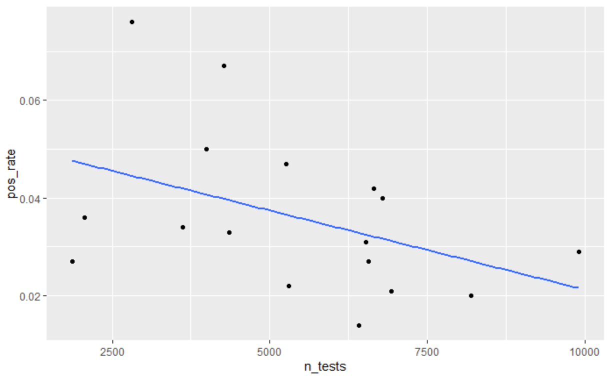

Figure 3. Number of people tested in past two weeks and positivity rate for each week

there is little relationship between n_tests and pos_rate.

World Health Organization claims that the countries should lower the restrictions when the covid positivity rate is below 5% for two weeks. Figure 3 checks the relationship between the number of people tested at the last two weeks and the positivity rate. Comparing to Figure 2, Figure 3 showed a strong negative relationship between the two variables. Two variables showed the correlation -0.427 with the slope of the model (positivity rate) ~ (number of people)$-3.238 * 10^{-6}$. Therefore, there is a negative relationship between people tested in the past two weeks and the positivity rate.

To find the relationship between the number of people tested and positivity in another way, the data was filtered into two sets- data which (n_tests) > 5000 and (n_tests) <= 5000. The mean positivity rates for (n_tests) > 5000 and (n_tests) <= 5000 were $3.74 * 10^{-2}$ and $3.21 * 10^{-2}$ where n_tests are the number of people tested this week. The mean positivity rates were $4.61 * 10^{-2}$ and $2.93 * 10^{-2}$ when n_tests are the number of people tested in the last two weeks. Both results support that higher n_tests refer to lower positivity rates. However, since the 95% confidence interval for the positivity rate was $(2.71 * 10^{-2},\ 4.27 * 10^{-2})$, we cannot reject the hypothesis ‘the positivity rate remains same when n_tests > 5000’ for the first result since its mean positivity rates was $3.74 * 10^{-2}$. However, we can reject the hypothesis for the second result since its mean positivity rate was $4.61 * 10^{-2}$, which is out of the confidence interval.

Likelihood ratio test checked the hypothesis “positivity rate is constant in the given range”. The data was first divided into four sets – week 1~5, week 6~10, week 11~15, week 16~19, and likelihood ratio tests were done for each interval. As a result, we cannot reject the hypothesis for 12 weeks, but we can reject the hypothesis for 7 weeks and there was no interval that all the weeks in the interval accepted the hypothesis. So, we cannot say that the positivity rate is constant for a given interval. The sets were then merged into two sets – week 1~10, week 11~19 and repeated the same process. As a result, we cannot reject the hypothesis for 14 weeks, but we can reject the hypothesis for 5 weeks. Both intervals contained the week that rejected the hypothesis.

Figure 1 and the likelihood ratio test for sets of intervals supports the claim that the positivity rate varied overtime at the University of Michigan. The reason for this result is because the number of people tested in each week was not large. This turned into two problems. First, since the data represented a small society, the number of infected people could increase or decrease rapidly. Second, since the positive tested people were few, the deviation intensified for each week.

The result of the second question ‘Does the positivity rate related to the number of people tested and number tested of people in past two weeks?’ was quite significant. Even if the formula of the positivity rate was (positive tests)/(total tests), we could not find a strong relationship between n_tests and pos_rate. Instead, the number of people tested in the last two weeks showed a negative relationship with positivity rate; It had a large correlation value and we can reject the hypothesis ‘the positivity rate remains same when n_tests > 5000’ by checking the confidence interval of pos_rate. The first result for this result is simply because when the number of tests increases, the number of positive tests increases. So there may not be a direct relationship between tests and the positivity rate. The second reason is, if there were many numbers tests during the last two weeks, we were able to isolate more patients by giving positive tests to them, which led us to reduce the infection rate. Since the incubation period for covid-19 is 2 weeks, this interpretation makes sense. In other words, if we want to lower the positivity rate, we can increase the number of tests that affect the positivity rate 2 weeks later.

One shortcoming that can lead the wrong interpretation is that we don’t know how the number of people in society has been changed. If the number of people tested has been increased because there were many visitors from other societies to the University of Michigan, the number of infected patients wouldn’t change much and the infection rate may remain similar. Another shortcoming for the methodology would be the mean of the positivity rate. Since the number of tests has been different for each week, it might be better to calculate the mean of the positivity rate not just summation them and divide them into n but summation the positivity rate times numbers of tests and divide it into total numbers of tests.

Citations

[1] DAVID DOWDY AND GYPSYAMBER D'SOUZA. "*COVID-19 Testing: Understanding the "Percent Positive"*. *Johns Hopkins Bloomberg School of Public Health*. August 10, 2020. https://www.jhsph.edu/covid-19/articles/covid-19-testing-understanding-the-percent-positive.html

[2] Lawrence D. Brown, Tony Cai and Anirban DasGupta (2001). *"Interval Estimation for a Binomial Proporation". Statistical Science.* Vol. 16. No. 2. 101-133