Lab #1 – how accurate is testing?

Questions to answer

1. You will have noted that the retest was quite effective. Intuitively, why was this? Hint – do you think the retest numbers would have look so good if we had retested *everybody* rather than just those people who originally tested positive?

   If we had retested everybody, then the retest would have still started with a universe where only 1% of the testees have Covid, and would have simply duplicated the same results as before. I.e., it would still have resulted in 7900 positives. Of course, they may not have been the *same* 7900 people.

   By only retesting the positives, we now have a universe where a full 25% of the people have Covid. Now our numbers are very different; all of a sudden there are far fewer false positives, simply because the percentage of uninfected people has dropped from 99% to 75%. The second test is essentially much more accurate!

2. Explain how Chapter 9 of Calling Bullshit (specifically the section on the "Prosecutor's fallacy") is really the same issue as this lab.
   If you look at the table that they fill out in Calling Bullshit, it is essentially exactly the same as ours. Their *Guilty* and *Innocent* correspond to our *sick* and *healthy*; their *match* and *no-match* correspond to our *positive* and *negative* test results. With that name change, everything else lines up perfectly.

Challenge problems

Each lab will include a few ungraded challenge problems intended to stretch your coding and analytical skills. If you found this lab easy, take some time to work on these!

1. Given the sensitivity and specificity defined above, how many tests do you need in order to be 99.99% sure that you don't have Covid?
   Start with P people.
   Round 1: 99% of the people are healthy and 1% are sick. Thus, true positives are $.01P*.98=.0098P$ and false positives are $.99P*(1-.97)=.0297P$, and so true positives are $.0098/(.0098+.0297)=.25$
   Round 2: true positives are now $(.0098P)*.98 = .009604P$; false positives are $(.0297P)*.03=.000891P$, and now 91% of our positives are true.
   Round 3: true positives are now $(.009604P)*.98 = .00941192P$; false positives are $(.000891P)*.03=.00002673P$, and now 99.7% of our positives are true. That's enough numbers – clearly one more round will do it!

2. Suppose you get tested every three days, and the test has 98% sensitivity. After a few negative tests, you'll be very sure you don't have Covid. But suppose that each day you also have a 1% chance of getting infected without knowing it. What is your limit of certainty? That is, on any given day, how sure can you be that you do not have the illness?

   You can solve this analytically or with code; we'd love to see both!