Bayes' Rule and LOTP

1. During the 2020-21 academic year, Reed College administered approximately 60,000 COVID-19 saliva RT-PCR tests to students, staff and faculty. The test used has an estimated *sensitivity* of 83% and an estimated *specificity* of 99.2% (sensitivity is the rate of detection when the virus is present, while specificity is the rate of rejection when it is not)[1]. Let p be the average prevalence of COVID-19 in the general Reed population during this time.

   (a) What are some consequences to receiving a positive diagnosis when a community member is actually COVID-19 negative? Conversely, what are some consequences to receiving a neagative diagnosis when a community member is actually COVID-19 positive?

   (b) Assume a prevalence of $p = 0.05\%$. How many positive test results would you anticipate? Reed reported 35 positive results during this period. What might this suggest about the true prevalence rate, sensitivity, and/or specificity?

   (c) Suppose a Reed community member has a positive test result. Find a formula (in terms of p) for the posterior probability that that the community member has COVID-19. Then evaluate for $p = 0.01\%, .05\%, .1\%, .5\%$ (a plausible range of values for the prevalence, based on existing data). For which values would you be comfortable concluding the community member has COVID-19?

   (d) Conversely, suppose the community member receives a negative test. Express the posterior probability that the community member does not have COVID-19 in terms of p. What do you think is an acceptable threshold to conclude that the individual does not have COVID? Evaluate for $p = 0.01\%, .05\%, .1\%, .5\%$.

   (e) What is the moral of the story? Would you recommend that all Reed community members undergo weekly COVID-19 surveillance testing?

   (f) Calculating posterior probabilities of infection is only possible if we know the prevalence p. But what are some fundamental challenges to obtaining a good estimate for p?

Independence and Simpson's Paradox

2. (*) Simpson's Paradox says that it is possible to have events A, B, C such that $P(A|B, C) < P(A|B^c, C)$ and $P(A|B, C^c) < P(A|B^c, C^c)$, yet $P(A|B) > P(A|B^c)$.

   (a) Can Simpson's paradox occur if A and B are independent?

   (b) Can Simpson's paradox occur if A and C are independent?

   (c) Can Simpson's paradox occur if B and C are independent?

[1] Estimates based on Butler-Laporte et. all "Comparison of Saliva and Nasopharyngeal Swab Nucleic Acid Amplification Testing for Detection of SARS-CoV-2" (2021)