**Lab 5 – Conditional Probability**

**To submit before your next lab: answers to all numbered questions. Also submit all commands and/or functions you used to generate your output, and submit a single .R file containing all of the scripts you wrote for this lab. If a question asks you to submit the output of one of your functions, submit that as well.**

In our last lab, we simulated various simple experiments: flipping a fair coin, rolling a fair die, and drawing cards from a deck. We are now going to simulate a real-life situation. Simulations are often conducted in order to help policymakers make better decisions.

# Application: COVID-19 antibody test

If a person has recently been sick with COVID-19, their blood contains antibodies that make the person immune to contracting the virus a second time. Serology testing can be used to determine the presence of antibodies. This has the potential to be very useful, since many people who contract COVID-19 have no symptoms.

1. Suppose that it is known that 5% of people have recently been sick with COVID-19 and therefore have COVID-19 antibodies. In a new script file called **Lab5.R**, write a function, **hasAntibodies(n)** that simulates checking **n** people and returns the number with antibodies. Assume people being checked are independent and that person has a 5% chance of having antibodies. Run your function for n=100000 and give your output. (You should be able to tell whether your function returns a reasonable result.)

The serology test, of course, is not perfect. A particular serology test has a **sensitivity** of 90%; that is, 90% of people who have recently been sick with COVID-19 “test positive” (ie, the test detects the presence of antibodies). The sensitivity is also known rate of **true positives**: that is, the proportion of the time a person with COVID-19 antibodies “tests positive”. We can express the sensitivity in the language of conditional probability: the sensitivity is the probability that a person tests positive, given that they have COVID-19 antibodies.

1. Write a function called **truePositives(n)** that simulates testing **n** people who have COVID-19 antibodies, and returns the number that test positive. Run your function for n=100000 and give your output.

The **specificity** of the test is 92%; that is, 92% of people who have not recently been sick with COVID-19 “test negative” (ie, the test does not detect the presence of antibodies).

1. In a sentence, express the rate of **false positives** (that is, the proportion of the time that a person **without** COVID-19 antibodies tests positive) in terms of conditional probabilities.
2. Write a function called **falsePositives**(n) that simulates testing **n** people who do not have COVID-19 antibodies, and returns the number who test **positive** (ie, your function returns the number of people without COVID-19 antibodies out of **n** that the test **incorrectly** identifies as having these antibodies). Run your function for n=100000 and give your output.

Now, in real life, we don’t know for sure whether a person has COVID-19 antibodies – all we know is the result of our test. The sensitivity tells us the probability that a person tests positive, given that they have COVID-19 antibodies.

We want to know the opposite: what is the probability that a person has COVID-19 antibodies, given that they test positive? In other words…how reliable is the test? We will guess this probability and then run a simulation to get a precise answer.

1. Suppose our test tells us that a person does not have COVID-19 antibodies. What do you think is the probability that this person actually does not have COVID-19 antibodies? (There is no right or wrong answer to this question – we’ll get an answer with a simulation in Question 7.)
2. Now suppose our test tells us that a person has COVID-19 antibodies. What do you think is the probability that our test is correct – ie, the person actually has COVID-19 antibodies?

Now we will run some simulations to see how accurate our guesses in Questions 5 and 6 were.

1. In this question, we will create a simulation that uses the relative frequency approach to answer Question 5: *Suppose our test is negative – that is, the test tells us that the person does not have COVID-19 antibodies. What is the probability that the person actually doesn’t have COVID-19 antibodies?*

To do this, you will write a function called **probNegativeGivenTestsNegative(n)** that simulates testing **n** samples and returns the following proportion:  
  
 .   
  
Note that the denominator will be a sum, because there are two types of negative results: false negatives, which arise from people who actually have COVID-19 antibodies; and true negatives, which arise from people who do not have COVID-19 antibodies.

Run your function for n=100000 and give your output. (Hint: your output should be above 0.99) Note: By calling the functions you wrote in Question 1, 2, and 3, you can write this function in just a few lines – my **probNegativeGivenTestsNegative(n)** function was six lines long. But you will need to be very careful with your function calls. Hint: start by finding the number of those **n** people who do not have antibodies.

1. Now write a function called **probPositiveGivenTestsPositive(n)** to get an answer to the following question: *suppose our test is positive – that is, our test tells us that the person has COVID-19 antibodies. What is the probability that our test is correct – that is, the person actually has COVID-19 antibodies?*

Run your function for n=100000 and give your output.

1. How do your outputs for questions 7 and 8 compare to your guesses for questions 5 and 6? If you worked at the lab that tested samples, would you consider a negative result to be fairly accurate? How about a positive result?
2. Now let’s see what happens if the incidence of COVID-19 were different. Suppose that in a different region, 10% of people have recently been sick with COVID-19 and therefore have antibodies. Now what are your results for Questions 7 and 8? What if a 50% of people in a region have recently had COVID-19 and have antibodies? Give your results, clearly labeled.