Lecture 8

STAT 109: Introductory Biostatistics

Lecture 8: Practice Problems (Lectures 6–7)

Practice Problem 1. A clinic is testing a new rapid test for COVID on 1000 people. Of these, 20 are known to have COVID and 980 do not. Among those known to have COVID, 17 test positive on the rapid test; among those known to not have COVID, 78 test positive.

Let C = "has COVID" and T^+ = "tests positive on the new rapid test"

a. Find $P(C)$ and $P(\text{not } C)$ for a randomly selected people from the 1000 people in the study.
b. Find $P(T^+ \mid C)$ (sensitivity) and $P(T^+ \mid \text{not } C)$ (false positive rate).
c. Find $P(C \text{ and } T^+)$ and $P(\text{not } C \text{ and } T^+)$.

Practice Problem 2. Using the COVID data from 1:

a. If C and T^+ were **independent**, what would $P(C \text{ and } T^+)$ equal?
b. Is it reasonable to assume that test result (T^+ or T^-) and COVID status (C or not C) are independent? Explain briefly.

Practice Problem 3. In the COVID example from 1, are C and T^+ actually independent? Explain briefly using the definition of independence.

Practice Problem 4. Using the COVID data from 1, find $P(T^+)$

Practice Problem 5. Suppose the **prevalence** of COVID-19 in a population is $P(C) = 0.02$ (2%). A rapid test has **sensitivity** $P(T^+ \mid C) = 0.85$ and **specificity** $P(T^- \mid \text{not } C) = 0.92$ (where $T^- =$ "tests negative").

a. Create tree diagram of the sample space of a person with a disease status and rapid test result, labeling each branch.
b. What is $P(T^+ \mid C)$?
c. Find $P(T^+ \mid \text{not } C)$ (the false positive rate).
d. Find $P(T^+)$ — the overall probability of a positive test.
e. Find $P(C \mid T^+)$ — the probability a person has COVID given they test positive.
f. Interpret your answer in part (d) in one sentence.
g. The method of computing $P(A|B)$ using information from items like $P(B|A) is called **Bayes Rule**. Match each piece of the calculation in e with the rules shown in Lecture 7.

On Your own:

Practice Problem 6. Recompute $P(C \mid T^{+})$ like in problem 5 but assume the prevalence were $P(C) = 0.10$ (10%) instead of 2%. How does the result change, and why?