Biost 517: Applied Biostatistics I
Biost 514: Biostatistics I
Autumn 2019

Homework #4
Due: Monday, October 28, 2019 by 9:00 AM

Written problems: To be submitted as a pdf or MS-Word compatible file via the canvas course website.

*On this (as all homeworks) R code and unedited R output is **TOTALLY** unacceptable. Instead, prepare a table of statistics gleaned from the R output. The table should be appropriate for inclusion in a scientific report, with all statistics rounded to a reasonable number of significant digits. (I am interested in how statistics are used to answer the scientific question.)*

Questions:

1. For this problem, we will use the same data used in the previous homework assignments on a subset of information that was collected to examine magnetic resonance imaging (MRI) changes in the brain in a sample of generally healthy elderly subjects in four U.S. communities. We are interested in 5-year all-cause mortality and creatinine levels (*crt*) at baseline (i.e., at the start of the study). We will use a threshold of 1.2 mg/dl to define "high" and "low" creatine levels.
    a. Provide a 2 x 2 contingency table with the counts of each combination of the dichotomous variable for vital status at 5 years (died within 5 year / survived at least 5 years) and the dichotomous creatinine level variable (high creatinine / low creatinine).
    b. Provide a point estimate, standard error, and 95% confidence interval for the probability that a random individual would survive at least 5 years.
    c. Provide a point estimate, standard error, and 95% confidence interval for the probability that a random individual has low creatine level.
    d. Provide a point estimate, standard error, and 95% confidence interval for the probability that a random individual that dies within five years would have low creatine level.
    e. Provide a point estimate, standard error, and 95% confidence interval for the probability that a random individual with high creatine would survive at least 5 years.

2. The ELISA test was approved by the U.S. government in the mid-1980s to screen donated blood for the presence of HIV, the virus that causes AIDS. The test works by detecting antibodies, substances that the body produces when the virus is present, but it makes some mistakes. Assume the ELISA test has sensitivity 99% and specificity 96%. Suppose the prevalence of HIV-infection in the population of people who donate blood is 1%.
    a. Compute the predictive value of a positive test (PPV). That is, for an individual sampled from this population whose test turns up positive, what is the probability that person is infected with the AIDS virus?

b. Compute the predictive value of a negative test (NPV). That is, for an individual sampled from this population whose test turns up negative, what is the probability that person is not infected with the AIDS virus?

c. Suppose a writer in the New York Times criticizes American blood banks for using the ELISA test to screen for blood contaminated by the AIDS virus. The writer states that the test is unacceptable because it is "wrong up to 4% of the time." (Presumably, the writer is noticing that the specificity is "only" 96%.) Write a brief response that explains why the ELISA test is in fact a very effective test for keeping the American blood supply safe.

d. For a test that is used as a tool to keep HIV out of the blood supply of American blood banks, is it more important for the test to have high sensitivity or high specificity? Explain.

e. There are currently effective treatments for individuals infected with HIV. These treatments extend life and also may make an infected person less likely to transmit the virus to others, although the treatments have very serious side effects and are expensive. Consider the use of the ELISA test in a clinic that serves a population with a large number of intravenous drug users. Assume that 16% of the population served by the clinic is infected with HIV. It is of interest to assess the operating characteristics of ELISA test for use in this clinic, assuming the test has sensitivity 99% and specificity 96%. What are the PPV and NPV in this population?

f. Now assume it is of interest to assess the operating characteristics of ELISA test for use in screening entering college students for HIV. Assuming the test has sensitivity 99% and specificity 96% and the prevalence of HIV is 0.01% in this population, what are the PPV and NPV in this population?

3. A biased coin is tossed 5 times, with the probability of a head being equal to 0.65 for a toss (and obviously the probability of a tail being equal to 0.35). Let X denote the total number of heads on the first toss. Let Y denote the total number of heads in the 5 tosses of the coin.

   a. What is the probability that $X = 1$ and $Y = 1$?

   b. What is the probability that $Y \geq 3$?

   c. What is the probability that $X = 1$ given that $Y = 1$?

   d. What is the expected value of Y?