Introduction to Biostatistics (171:161)
Breheny

Lab #6

In today's lab, we will focus more on practicing probability problems rather than programming.

1. The table below contains information on employment status and impairments.

Employment Status	Population	Impairments
Currently employed	98,917	552
Currently unemployed	7,462	27
Not in labor force	56,778	368
Total	163,157	947

(a) Find the probability for being in each status of employment.

$$P[\text{Currently Employed}] = \frac{98917}{163157} = \boxed{0.606}$$

$$P[\text{Currently Unemployed}] = \frac{7462}{163157} = \boxed{0.046}$$

$$P[\text{Not in Labor Force}] = \frac{56778}{163157} = \boxed{0.348}$$

(b) Find the probability that an individual has a hearing impairment due to injury.

$$P[\text{Hearing Impairment}] = \frac{947}{163157} = \boxed{0.006}$$

(c) Find the probability that an individual has an impairment from injury, given they are not in the labor force.

$$P[\text{Impairment}|\text{Not in Labor Force}] = \frac{368}{56778} = \boxed{0.006}$$

(d) Find the probability that an individual is currently employed, given they have an impairment.

$$P[\text{Currently Employed}|\text{Impairment}] = \frac{552}{947} = \boxed{0.583}$$

2. The National Institute for Occupational Safety and Health has developed a case definition of carpal tunnel syndrome- an affliction of the wrist-that incorporates three criteria: symptoms of nerve involvement, a history of occupational risk factors, and the presence of physical exam findings. The sensitivity of this definition as a test for carpal tunnel syndrome is 0.67; its specificity is 0.58. (Pagano & Gauvreau, Problem 14, Page 157)

   (a) In a population in which the prevalence of carpal tunnel syndrome is estimated to be 15%, what is the predictive value of a positive test result?

$$\begin{aligned} PVP &= \frac{P[T^+|D^+]P[D^+]}{P[T^+|D^+]P[D^+] + P[T^+|D^-]P[D^-]} \\ &= \frac{(0.67)(0.15)}{(0.67)(0.15) + (1-0.58)(1-0.15)} \\ &= \boxed{0.22} \end{aligned}$$

   (b) How does this predictive value change if the prevalence is only 10%? If it is 5%?

$$\begin{aligned} PVP &= \frac{P[T^+|D^+]P[D^+]}{P[T^+|D^+]P[D^+] + P[T^+|D^-]P[D^-]} \\ &= \frac{(0.67)(0.10)}{(0.67)(0.10) + (1-0.58)(1-0.10)} \\ &= \boxed{0.15} \end{aligned}$$

$$\begin{aligned} PVP &= \frac{P[T^+|D^+]P[D^+]}{P[T^+|D^+]P[D^+] + P[T^+|D^-]P[D^-]} \\ &= \frac{(0.67)(0.05)}{(0.67)(0.05) + (1-0.58)(1-0.05)} \\ &= \boxed{0.08} \end{aligned}$$

   (c) What is the probability of having a positive test when the prevalence is 15%?

$$\begin{aligned} P[T^+] &= (0.15)(0.67) + (0.85)(0.42) \\ &= \boxed{0.46} \end{aligned}$$

3. Comstock and Patridge show data giving an association between church attendance and health. Compute the relative risk of a death from cirrhosis in the three-year follow-up period if one usually attends church less than once a week as compared to once a week or more. (van Belle, Exercise 11, page 197)

Church	Cirrhosis Death	
Attendance	Yes	No
≥ 1/wk	5	24,240
< 1/wk	25	30,578

$$\begin{aligned} RR &= \frac{P[\text{death}|< 1/\text{wk}]}{P[\text{death}|\geq 1/\text{wk}]} \\ &= \frac{25/30603}{5/24245} = \boxed{3.96121} \end{aligned}$$

4. The ELISA (Enzyme-Linked Immunosorbent Assay) test is used to screen donated blood for HIV, the virus which causes AIDS. ELISA checks for the presence of an antibody produced when an individual is exposed to HIV. To evaluate the sensitivity and the specificity of the ELISA test, suppose that the test is administered to a group of 500 individuals known to be HIV positive and to a group of 500 individuals known to be HIV negative. Assume the following results are obtained:

HIV Status	Test Result		Total
	Positive	Negative	
Positive	487	13	500
Negative	17	483	500

(a) Obtain estimates of the sensitivity and specificity of the ELISA test.

$$\begin{aligned} Sensitivity &= P[T^+|D^+] \\ &= \frac{487}{500} = \boxed{0.974} \end{aligned}$$

$$\begin{aligned} Specificity &= P[T^-|D^-] \\ &= \frac{487}{500} = \boxed{0.966} \end{aligned}$$

(b) Assume that the prevalence of HIV in a population of interest is estimated as 0.004. Find an estimate for the predictive value positive (PVP) of the ELISA test for this population.

$$\begin{aligned} PVP &= \frac{P[T^+|D^+]P[D^+]}{P[T^+|D^+]P[D^+] + P[T^+|D^-]P[D^-]} \\ &= \frac{(0.974)(0.004)}{(0.974)(0.004) + (1-0.966)(1-0.004)} \\ &= \boxed{0.1032} \end{aligned}$$

(c) Suppose the ELISA test is used to screen a large collection of donated blood samples for the presence of HIV. What are the implications of the results in part (b) regarding the efficacy of the test?

With a predictive value positive of 10.32%, many of the positive tests for that disease will be incorrect (false positives). Thus, in the case of testing donated blood for HIV, many blood samples which do not have HIV may come back positive for the presence of HIV.

5. A medical research team wishes to assess the usefulness of a certain symptom (call it S) in the diagnosis of a particular disease. In a random sample of 775 patients with the disease, 744 reported having the symptom. In an independent random sample of 1380 subjects without the disease, 21 reported that they had the symptom

   (a) In the context of this problem, what is a false positive?

   A subject having the symptom (S) and not having the disease.

   (b) What is a false negative?

   A subject not having the symptom (S), but having the disease.

   (c) Compute the sensitivity of the symptom.

$$\begin{aligned} Sensitivity &= P[S^+|D^+] \\ &= \frac{744}{775} \\ &= \boxed{0.960} \end{aligned}$$

   (d) Compute the specificity of the symptom.

$$\begin{aligned} Specificity &= P[S^-|D^-] \\ &= \frac{1380 - 21}{1380} \\ &= \boxed{0.985} \end{aligned}$$

   (e) Suppose it is known that the rate of the disease in the general population is 0.001. What is the predictive value positive of the symptom?

$$\begin{aligned} PVP &= \frac{P[S^+|D^+]P[D^+]}{P[S^+|D^+]P[D^+] + P[S^+|D^-]P[D^-]} \\ &= \frac{\left(\frac{744}{775}\right)(0.001)}{\left(\frac{744}{775}\right)(0.001) + \left(1 - \frac{1380-21}{1380}\right)(1 - 0.001)} \\ &= \boxed{0.0594} \end{aligned}$$

   (f) What is the predictive value negative of the symptom?

$$\begin{aligned} PVN &= P[D^-|S^-] \\ &= \frac{P[S^-|D^-]P[D^-]}{P[S^-|D^-]P[D^-] + P[S^-|D^+]P[D^+]} \\ &= \frac{\left(\frac{1380-21}{1380}\right)(1 - 0.001)}{\left(\frac{1380-21}{1380}\right)(1 - 0.001) + \left(1 - \frac{744}{775}\right)(0.001)} \\ &= \boxed{0.99996} \end{aligned}$$

(g) Find the predictive value positive and the predictive value negative for the symptom for the following hypothetical disease rates: 0.0001, 0.01, and 0.10.

$$
\begin{aligned}
PVP &= \frac{P[S^+|D^+]P[D^+]}{P[S^+|D^+]P[D^+] + P[S^+|D^-]P[D^-]} \\
&= \frac{\left(\frac{744}{775}\right)(0.0001)}{\left(\frac{744}{775}\right)(0.0001) + \left(1 - \frac{1380-21}{1380}\right)(1 - 0.0001)} \\
&= \boxed{0.0063}
\end{aligned}
$$

$$
\begin{aligned}
PVP &= \frac{\left(\frac{744}{775}\right)(0.01)}{\left(\frac{744}{775}\right)(0.01) + \left(1 - \frac{1380-21}{1380}\right)(1 - 0.01)} \\
&= \boxed{0.3892}
\end{aligned}
$$

$$
\begin{aligned}
PVP &= \frac{\left(\frac{744}{775}\right)(0.10)}{\left(\frac{744}{775}\right)(0.10) + \left(1 - \frac{1380-21}{1380}\right)(1 - 0.10)} \\
&= \boxed{0.8751}
\end{aligned}
$$

$$
\begin{aligned}
PVN &= P[D^-|S^-] \\
&= \frac{P[S^-|D^-]P[D^-]}{P[S^-|D^-]P[D^-] + P[S^-|D^+]P[D^+]} \\
&= \frac{\left(\frac{1380-21}{1380}\right)(1 - 0.0001)}{\left(\frac{1380-21}{1380}\right)(1 - 0.0001) + \left(1 - \frac{744}{775}\right)(0.0001)} \\
&= \boxed{0.999996}
\end{aligned}
$$

$$
\begin{aligned}
PVN &= \frac{\left(\frac{1380-21}{1380}\right)(1 - 0.01)}{\left(\frac{1380-21}{1380}\right)(1 - 0.01) + \left(1 - \frac{744}{775}\right)(0.01)} \\
&= \boxed{0.9996}
\end{aligned}
$$

$$
\begin{aligned}
PVN &= \frac{\left(\frac{1380-21}{1380}\right)(1 - 0.10)}{\left(\frac{1380-21}{1380}\right)(1 - 0.10) + \left(1 - \frac{744}{775}\right)(0.10)} \\
&= \boxed{0.9955}
\end{aligned}
$$

(h) What do you conclude about the predictive value of the symptom on the basis of the results obtained in part (g)?

Prevalence	PVP	PVN
0.1	0.8751	0.9955
0.01	0.3892	0.9996
0.001	0.0594	0.99996
0.0001	0.0063	0.999996

As the prevalence decreases, so does the Predictive Value Positive. As the prevalence increases, the Predictive Value Negative increases.

6. Consider the following natality statistics for the U.S. population in 1992. According to these data, the probabilities that a randomly selected woman who gave birth in 1992 was in each of the following age groups are as follows: (Pagano & Gauvreau, Problem 9, Page 156)

Age	Probability
< 15	0.003
15-19	0.124
20-24	0.263
25-29	0.290
30-34	0.220
35-39	0.085
40-44	0.014
45-49	0.001
Total	1.00

(a) What is the probability that a woman who gave birth in 1992 was 24 years of age or younger?

$$\begin{aligned} P[\leq 24] &= P[< 15] + P[15-19] + P[20-24] \\ &= 0.003 + 0.124 + 0.263 = \boxed{0.39} \end{aligned}$$

(b) What is the probability that she was 40 or older?

$$\begin{aligned} P[\geq 40] &= P[40-44] + P[44-49] \\ &= 0.014 + 0.0001 = \boxed{0.015} \end{aligned}$$

(c) Given that the mother of a particular child was under 30 years of age, what is the probability that she was not yet 20?

$$\begin{aligned} P[< 20 | < 30] &= \frac{P[< 20 \cap < 30]}{P[< 30]} = \frac{P[< 20]}{P[< 30]} \\ &= \frac{0.003 + 0.124}{0.003 + 0.124 + 0.263 + 0.290} \\ &= \frac{0.127}{0.68} = \boxed{0.0187} \end{aligned}$$

(d) Given that the mother was 35 years of age or older, what is the probability that she was under 40?

$$\begin{aligned} P[< 40 | \geq 35] &= \frac{P[< 40 \cap \geq 35]}{P[\geq 35]} \\ &= \frac{0.085}{0.085 + 0.014 + 0.001} \\ &= \frac{0.085}{0.100} = \boxed{0.85} \end{aligned}$$

This concludes lab #6.