MORE PROBABILITY PROBLEMS

1. In the handout Basic Probability, you did a problem involving medical tests. You were given the *false positive rate* (P(positive test result | disease not present)) and the *false negative rate* (P(negative test result |disease present)) as well as the *prevalence rate* (the percent of the population that has the disease). However, it is customary to give the *sensitivity* and *specificity* instead of the false positive and false negative rates for a medical. These are defined as

   Sensitivity = P(positive test result | disease present)
   Specificity = P(negative test result | disease not present)
We'll use the following notation:
   s_e = sensitivity
   s_p = specificity
   r = prevalence rate
   + = the event "test result positive"
   - = the event "test result negative"
   D = the event "has the disease" (So D^C = the event "does not have the disease")

a. To get a feel for the terminology, consider the population as broken down into the following contingency table, with cells labeled TP for Number of True Positive, etc.

	Has Disease	Doesn't Have Disease
Test +	TP	FP
Test -	FN	TN

Express the false positive rate, the false negative rate, the sensitivity, and the specificity in terms of TP, FP, FN, and TN.

b. Express the false positive rate and the false negative rate in terms of the specificity and sensitivity (not necessarily respectively).

c. Use Bayes Theorem to express P(D| +) and P(D^C| -) in terms of s_e, s_p, and r.

->Use your formulas from part (c) to set up a spreadsheet to use in the rest of this problem to calculate P(D| +) and P(D^C| -) in terms of s_e, s_p, and r.

d. When a disease is diagnosed by a lab test (e.g., on blood), the false positive and false negative rates are typically low, as in the earlier example. But when a diagnosis is made on the basis of symptoms, the false positive and false negative rates can be higher. For example, in diagnostic procedures for major depressive disorder (MDD), sensitivity might be 0.85 and specificity 0.75 (so false positive and false negative rates are …).

Using these values for sensitivity and specificity, use your spreadsheet to find the probability that a person diagnosed as having MDD actually has MDD *and* the probability that a person diagnosed as not having MDD actually does not have MDD, under each of the following prevalence rates:

      i. Prevalence rate = 0.03 (This is the estimated prevalence rate for adolescents.)

      ii. Prevalence rate = 0.54. (This is the estimated prevalence rate for chronic pain patients.)

Comment on the difference between the two situations.

e. One particular method of diagnosing MDD is estimated to have sensitivity 0.25 and specificity 0.98. Repeat part c for this test. Compare with what you got in part c. Is this what you expect? What possible lesson might one learn from this?

f. i. For the sensitivity and specificity given in part (d), use your spreadsheet to calculate $P(D| +)$ and $P(D^C, -)$ for values of r in increments 0.1. Use these to make a graph of both $P(D| +)$ and $P(D^C, -)$ as functions of r . Using your graph, answer the following questions:

      Approximately what range of prevalence rates that will give P(has MDD|diagnosed with MDD) ≥ 0.90?

      Approximately what range of prevalence rates will give P(does not have MDD|diagnosed as not having MDD) ≥ 0.90?

      Can you achieve both of these at the same time?

      In what range of prevalence rates do you think this test would work best?

  ii. Repeat part (i) with the sensitivity and specificity in part d.

  iii. What possible lesson might one learn from this?

Optional. Investigate the questions in part (f) for other values of sensitivity and specificity.

Note: We will come back to this problem to investigate the role of sensitivity and specificity further.