Homework #2

Due: Tuesday, October 5 @ 6pm

Problem 1:

Suppose that a disease is inherited via a sex-linked mode of inheritance so that a male offspring has a 50% chance of inheriting the disease, but a female offspring has no chance of inheriting the disease. Further suppose that 51.3% of births are male.

a. Draw a probability tree and/or generate a contingency table representing the data above
b. Are the two events (inheriting the disease and being male) disjoint or non-disjoint? Explain.
c. Are the two events (inheriting the disease and being male) independent or non-independent? Explain.
d. What is the probability that a randomly chosen child will be affected by the disease? Be sure to show all work.

Problem 2:

Suppose a test is 99% accurate: it gives a positive result 99% of the time if the patient is indeed infected (i.e. a 1% false-positive rate), and a negative result 99% of the time if the patient is indeed healthy (i.e. a 1% false-positive rate). For convenience, let pos/neg denote a positive/negative test result, and let I/H denote infected/healthy.

a. Using P(A | B) notation, write down the facts described above
b. Suppose I take the test and it comes up positive. I'd like to know what that means about the chances that I'm actually infected. Using P(A|B) notation, write down the quantity that I'm interested in (not the number, just the notation for the conditional probability that corresponds to this question).
c. Supposing my test came up positive, what can you tell me about the chances that I'm actually infected? Give numbers if you can; if you can't state what else you'd need to know. [*Hint: remember that the probability of a positive result is the sum of the probabilities of getting a true positive or of getting a false positive*]
d. Suppose we administered the test in a population where the prevalence of infection (i.e., the baseline probability that a given person is infected) is 1/1000. That is, P(I) = 0.001
   - What fraction of all people would have a positive test result?
   - Of those people, what fraction of them would be truly infected?
   - What is the probability, then, that a person in this population with a positive result is truly infected? (Note: we call this the *posterior probability* or the *positive predicitve value*)
   - Should people in this population believe that they're more likely infected than not if they get a positive result?
   - Does this answer surprise you? Why or why not?
e. Suppose now we do the same thing, but in a high-risk population where the prevalence is 1/3. How do your answers to (d) change? Between this result and your answer to part (d), what kind of recommendations would you make for administering this screening test?
f. Suppose we go back to the low-risk group in (d) and re-administer the same test to those who tested positive the first time. What is the probability that someone who tests positive a second time (in addition to the first) is truly infected?

Problem 3:

The seeds of the garden pea (*Pisum sativum*) are either yellow or green. A certain cross between pea plants produced progeny in the ratio 3 yellow : 1 green.

a. Does this variable fit the assumptions for a binomial random variable? Why or why not?
b. If four randomly chosen progeny of such a cross are examined, what is the probability that one is green and three are yellow? Be sure to show ALL work.
c. Generate the probability distribution for every possible outcome given four randomly chosen progeny of such a cross are examined. [*Hint: first select which outcome will be viewed as "success" and create the probability table for that variable*]
d. What is the probability that all four randomly chosen progeny are the same color?
e. What is the expected value of green (or yellow) seeds? How does this compare to the known ratio of yellow:green seeds?
f. What is the standard deviation of green (or yellow) seeds?

Problem 4:

When red blood cells are counted using a certain electronic counter, for a certain specimen, the true value is 5,000,000 cells/mm^3 and the standard deviation is 40,000. The distribution of repeated counts is approximately normal.

a. Suppose you get a reading of 4,900,000. What is the standardized z-score for this value?
b. What is the probability that the counter would give a reading between 4,900,000 and 5,100,000?
c. If the true value of the red blood count for a certain specimen is μ, what is the probability that the counter would give a reading between 0.98μ and 1.02μ?
d. A hospital lab performs counts of many specimens every day. For what percentage of these specimens does the reported blood count differ from the correct value by 2% or more?