STAT22000, Autumn 2013 Homework 4

All page, section, and exercise numbers below are for the course text (Moore, McCabe and Craig, Introduction to the Practice of Statistics, 7th edition).

Reading: Section 3.3, 4.1, 4.2, 4.5
Problems for Self-Study: (Do Not Turn In. Solutions are at the end of the textbook.)

1. Exercise 3.83 on p.211
2. Exercise 3.84 on p.211
3. Exercise 4.19 on p.245
4. Exercise 4.43 on p.248
5. Exercise 4.102 on p.290

Problems to Turn In: due Wednesday, Oct. 30, in class

1. Exercise 4.20 on p.245

2. Exercise 4.32 on p.247

3. Exercise 4.122 on p.292

4. For two events A and B in a sample space S, the event A is called independent of the event B if $P(A|B) = P(A)$. Assuming A is independent of the event B, answer the following questions by naming and providing the definition/set relation/set notation that justifies the equality:

   (a) Why is it valid to write $P(A \cap B) = P(A \mid B)P(B)$?
   (b) Why is it valid to write $P(A \mid B)P(B) = P(A)P(B)$?
   (c) Why is B independent of the event A, that is, why $P(B|A) = P(B)$?
   (d) Why $P(A \cup B) = P(A) + P(B) - P(A)P(B)$?

5. The accuracy of a medical diagnostic test, in which a positive result indicates the presence of a disease, is often stated in terms of its **sensitivity**, the proportion of diseased people that the test is positive or $P(+|\text{Disease})$, and its **specificity**, the proportion of people without the disease who test negative or $P(-|\text{No Disease})$. Suppose that 10% of the population has the disease (called the **prevalence rate**). A diagnostic test for the disease has 99% sensitivity and 98% specificity. Therefore

$$P(+|\text{ Disease}) = 0.99, \quad P(-|\text{ No Disease}) = 0.98$$
$$P(-|\text{ Disease}) = 0.01, \quad P(+|\text{ No Disease}) = 0.02$$

   (a) A person's test result is positive. What is the probability that the person actually has the disease?

   (b) A person's test result is negative. What is the probability that the person actually does not have the disease? Considering this result and the result from (a), would you say that this diagnostic test is reliable? Why or why not?

   (c) Now suppose that the disease is rare with a prevalence rate of 0.1%. Using the same diagnostic test, what is the probability that the person who tests positive actually has the disease?

(d) The results from (a) and (c) are based on the same diagnostic test applied to populations with very different prevalence rates. Does this suggest any reason why mass screening programs should not be recommended for a rare disease? Explain.