PHP 2500 Introduction to Biostatistics

Problem Set Two

Due: Tuesday October 9, 2007 at the beginning of class

---

1. Pagano #7 (Chapter 6, p155)

2. A study population consists of persons living in five villages (villages a, b, c, d, and e). Three of the villages will be chosen for observation.

   (a) List the sample space (sets of three villages). How many points are in the sample space?
   (No repetitions. We want to observe three different villages; that is we will not observe any village twice.)

   Assign equal probabilities to all of the sample points and consider the following events (This is called simple random sampling without replacement.):
   A : Village a is selected
   B : Village b is selected
   C : Village c is selected

   (b) (i) Find P(A), P(B), and P(C).
   (ii) Find P(A and B), P(A and C), and P(B and C).
   (iii) Find P(A or B), P(A or C), and P(B or C).
   (iv) What is P(A and B and C) and P(A or B or C)?

3. A study of parasite infection in a rural population estimated that 50% of the population have parasite A, 25% have parasite B, and 40% have neither A nor B.

   Suppose we select one person at random from the population. (This means that every person has the same chance of being the one selected.) Let A represent the event "The selected person has parasite A," let A^c represent "The person does not have parasite A," etc.

   (a) Draw a Venn diagram to represent this random trial.
   (b) Compute P(A or B)
   (c) Compute P(A and B)

3. (d) Are the events A and B mutually exclusive?
   (e) Are the events A and B independent?
   (f) Are the events A and A^c mutually exclusive?
   (g) Are the events A and A^c independent?

4. A bus carrying 30 passengers has an accident. Fourteen passengers suffered fractured bones, ten passengers were burned, and ten were not injured. Consider choosing one of the 30 passengers at random.

   (a) Draw a diagram or table that represents this situation.

   (b) Calculate the probability that the selected person has both burns fractures.

   (c) What is the probability that the selected person is burned, but has no fractures?

   (d) You observe that the selected person has burns. What is the probability that he also has fractures?

   (e) Are the occurrences of fractures and burns independent in this problem? Why?

5. A certain county health department has received 25 applications for a new public health nurse job opening. Of these applications, there are ten applicants who are over 30 years of age and fifteen who are not. Seventeen hold bachelor's degrees only, and eight have master's degrees. **Of those under 30, six have master's degrees.**

   (a) What is the probability that a person over 30 years of age or a person with a master's degree will be selected?

   (b) Given that the applicant has a master's degree, what is the probability that the applicant is over 30 years of age?

6. A driver who has had a high school driver education course has an 85% chance of not having an accident during the first year of driving. This probability drops to 60% for a driver who has not had such a course. Suppose that 80% of all new drivers enroll in a driver education course while in high school.

   (a) What is the probability that a new driver will finish their first year of driving without an accident?

   (b) What is the probability that a new driver who is involved in an accident during the first year has had a high school driver education course?

7. A large group of narcotics addicts filled out questionnaires that were designed to measure anxiety. There were three questions, I , II, and III, that were of particular interest to the investigators. The results were as follows:

   20% answered “yes” to I and “no” to the others.
   16% answered “yes” to II and “no” to the others.
   14% answered “yes” to III and “no” to the others.
   8% answered “yes” to I and II and “no” to III.
   5% answered “yes” to I and III and “no” to II.
   4% answered “yes” to II and III and “no” to I.
   3% answered “yes” to all three.

   (a) What percentage answered “yes” to at least one question?

   (b) What percentage answered “no” to all three questions?

   (c) Of those who answered “yes” to at least one question, what percentage answered “yes” to both I and II?

8. Suppose I toss a quarter three times. Let *X* represent the number of "heads".

   (a) Write out the probability model for this experiment?

   (b) What is the probability that I will get no heads? ...one? ...two? ...three?

   (c) If you toss a quarter three times also (independently of me), what is the probability that we both will get two heads?

   (d) What is the probability that you and I will both get the same number of heads?

   (e) If five people all toss quarters three times, it is unlikely that all five will observe the same number of heads. Just how unlikely is it? (I.e., what is the probability that all five people will observe the same number of heads?)

9. In the year 3001, it is predicted that on average 70.8% of the population will live until they are at least 99 years old and 85.3% will live to at least 85 years old.

   (a) For someone having their 85^{th} birthday in 3001, what is their estimated probability of living to be at least 99 years old?

   (b) What is the probability of a married couple, both of whom turn 85 in 3001, will live to celebrate their 99% birthday together? (assume the couple are independent)

   (c) What are the chances that one of the partners in the married couple (both of whom turn 85 in 3001) lives to be 99 and the other does not?

10. Pagano #13 (Chapter 6, p157)

11. Pagano #15 (Chapter 6, p155)

12. Pagano #16 (Chapter 6, p158) ; Skip (b) and (c)

13. The last page presents three graphs of the Positive Predictive Value as a function of sensitivity, prevalence, and specificity. Each graph varies the prevalence (ranges from 0 to 1) and sensitivity (ranges from 5% to 100%) for a fixed value of specificity (92%, 75%, 8%).

    (a) What is your first impression after looking at these graphs?

    (b) What role does prevalence play in determining the usefulness of a diagnostic test?

    (c) Does this change your perception about the usefulness of a 90% sensitive test versus a 75% sensitive test? Why?

14. Calculate the Odds (in each group), odds ratio and relative risk (for group 1 versus groups 2) for the follow pairs of proportions:

    (a) $P_1 = 0.5$; $P_2 = 0.5$ and $P_1 = 0.3$; $P_2 = 0.4$
    (b) $P_1 = 0.05$; $P_2 = 0.1$ and $P_1 = 0.95$; $P_2 = 0.9$
    (c) $P_1 = 0.1$; $P_2 = 0.9$ and $P_1 = 0.9$; $P_2 = 0.1$
    (d) $P_1 = 0.01$; $P_2 = 0.04$ and $P_1 = 0.99$; $P_2 = 0.96$
    (e) $P_1 = 0.01$; $P_2 = 0.04$ and $P_1 = 0.96$; $P_2 = 0.99$
    (f) When is the OR a good approximation to the RR?
    (g) Compare the odds ratios and relative risks in #d and #e. Think about the underlying probability of success and notice how close each proportion is to zero or 1 (Do you see the mathematical symmetry here?)
        a. True or False: When reporting odds ratios, it does not matter if you focus on the probability of success or the probability of failure because the OR of interest will be the same.
        b. True of False: When reporting the relative risk, it does not matter if you focus on the probability of success or the probability of failure because the RR of interest will be the same.

15. In a study conducted at the University of Cincinnati Medical Center, 209 at risk Caucasian women, aged 41 to 81 years, were evaluated for presence of Osteoporosis. The most common evaluation is bone mineral density (BMD) scanning of the femur or spine, but several innovative procedures for measuring BMD in the peripheral skeleton are also available, which are less costly and less intrusive. Low BMD measurements are predictive of Osteoporosis.

    The data set ostio_roc.dta contains standardized BMD measurements from the *heel* and *finger* on each of the 209 individuals. *[WARNING: The finger and heel BMD measurements have been multiplied by -1 because STATA expects that higher test values indicate positivity for disease. Be careful to account for this when you interpret your output!]* Women were defined to have Osteoporosis if they met the 1994 World Health Organization clinical definition, which does not depend on any of the tests under evaluation. The presence or absence of Osteoporosis is recorded in the variable *ostio* (=1 yes ; = 0 no). Forty-two women in our analysis set had Osteoporosis as defined by the WHO criteria.

    Use STATA to answer the following questions:

    (a) Graph the ROC curve for each location on the same graph (Remember to use the 'aspectratio(1)' command after the comma to make the ROC plot square. Hint try 'roccomp').
    (b) What is the area under each ROC curve? Which is larger? Which curve has overall better predictive ability?
    (c) What is the sensitivity and specificity associated with a standardized finger BMD measurement of -0.6?
    (d) What is the sensitivity and specificity associated with a standardized heel BMD measurement of -0.6?
    (e) What finger BMD measurement is associated with a sensitivity of 47.62% and specificity of 88.02%?
    (f) What heel BMD measurement is associated with a sensitivity of 47.62% and specificity of 88.62%?
    (g) Compare the results of (e) and (f). What important information is hidden in the ROC graphs? (Hint: If I wanted to use both tests so they had the same sensitivity and specificity, would I apply the test in the same way, i.e., use the came cutoff value for the heel and finger?)

PPV graphs: Sensitivity Ranges from 5% to 100% (by 5%)