PHP 2500 Introduction to Biostatistics

Problem Set Three

Due: Tuesday October 23rd, 2007 at the beginning of class

Please show all your work; that is, write the question in terms of a probability statement and solve. Useful Stata commands for calculating various probabilities are included at the end of this problem set. Additionally there are tables of Binomial, Poisson, and Standard Normal probabilities in the back of Pagano (Appendix A, pA-9) that will help with your calculations. You are encouraged to use Stata and/or these tables to complete probability calculations, but be sure that you can solve each problem on your own if you had to.

1. Pagano #10, skip part (a), (Chapter 7, p192)

2. Pagano #11, skip part (a), (Chapter 7, p192)

3. According to the Behavioral Risk Factor Surveillance Systems, 58% of all Americans adhere to a sedentary lifestyle (<u>do not</u> exercise regularly).

   (a) If you selected a random sample of twelve Americans, what would be the chance of observing two or fewer Americans who <u>do not</u> exercise regularly?

   (b) If you selected a random sample of twelve Americans, what would be the chance of observing two or fewer Americans who <u>exercise regularly</u>?

   (c) Let F represent an American who has a sedentary life style and S represent and active American. What is the probability of observing the sample SFFSSSFFFSFF? How about FFFFFFFSSSSS?

   (d) What is the probability of observing 10 individuals with a sedentary lifestyle (out of the sample of twelve), given that three sedentary individuals were observed earlier?

4. Pagano #14, skip (d), (Chapter 7, p193)

5. Pagano #15, (Chapter 7, p193)

   (d) What is the probability that no suicides will be reported during a <u>four-month</u> period?

   (e) What is the probability that at most four suicides will be reported during a <u>four-month</u> period?

   (f) What is the probability that at most 16 suicides will be reported during a <u>four-month</u> period?

6. In the United States, the probability that a child dies during his or her first year of life is 0.00075.

   (a) In a group of 2000 newly born infants, what is the probability that at most two infants die in their first year of life? (use a binomial model)

   (b) Suppose in some small city only 2000 infants are born a year. What is the probability that the city will have 10 consecutive years where two or fewer infants die each year?

   (c) For that same city, what is the probability that the city will have exactly 7 years out of the next 15, where two or fewer infants die each year?

   (d) Use a Poisson model to approximate the answer to (a). Assess the accuracy of this approximation. (Remember that $\lambda = n\theta$, the average number of infant deaths in the first year)

7. There are 80 students who will receive grades in Biostatistics 701, and 7 of them are doctoral students. The grades will be curved so that 5% (4 students) get A's. Suppose that grades are assigned by selecting four students at random and giving them A's.

   (a) If X is the number of doctoral students who receive A's, what is the name of the probability distribution of X?

   (b) What is the probability that all four A's will go to doctoral students?

   (c) What is the probability that one or more of the A's will go to doctoral students?

8. Select a simple random sample of size n without replacement from a class of size N. (Your answers to this question will be in terms of n and N.)

   (a) What is the probability that one specified student (e.g. you) will be in the selected sample?

   (b) What is the probability that two specified students (e.g., you and your best friend) will both be selected?

   (c) What is the probability that neither of the two (part (b)) will be selected?

9. You are going to do a job that requires ten electrical connectors. Each connector has probability 0.04 of being defective.

   (a) If you have ten connectors, what is the probability that all of them are good (no defectives)?

   (b) If you have eleven, what is the probability that you have at least ten good ones?

   (c) How many connectors do you need in order to be at least 95% certain that you have at least ten good ones?

10. From the table of the standard normal (z) distribution, evaluate the following probabilities and sketch the calculated areas:

    (a) $P(Z \geq 1.96)$
    (b) $P(Z \geq -1.645)$
    (c) $P(|Z| \geq 1.96)$
    (d) $P(Z \geq q)=0.80$ (Find q)

11. Pagano #17, (Chapter 7, p193)

12. Pagano #18, (Chapter 7, p194)

13. Pagano #19, (Chapter 7, p194)

14. A new Chemo-radiation regimen is developed for advanced, non-resectable stage 4 breast cancer and is tested in 20 volunteers. Currently, the probability of surviving 1 year with this disease is thought to be 28%, but if this regimen works it will increase the probability of 1-year survival to 40%. (We will consider only these two scientific hypotheses, H_1: $\theta=0.28$ and H_2: $\theta=0.4$, in this problem). Out of the 20 Volunteers, 8 survived at least one year.

    (a) The inventors, not surprisingly, believed before hand that that probability that the regimen increases 1-year survival was 90%. After observing these data, what does this probability change to?

    [Hint: Find, $P(H_2 \mid 8 \text{ successes})$, the probability that the regimen works given 8 successes in 20 tries. We know that before hand P(regimen works)=$P(H_2)=0.9$, so after seeing these data we ask what is P(regimen works | 8 success in 20 tries)? To solve this, use Bayes theorem to flip the conditionals and then use the Binomial distribution to calculate the individual probabilities.]

    (b) A more circumspect group of experts believes that the regimen has only a 35% chance to increase 1-year survival. After observing these data, what does this group of experts think the probability of 1-year survival is?

(c) Calculate the prior odds that the drug works [i.e., P(works)/1-P(works)], the posterior odds that the regimen works [i.e., P(works |data)/(1-P(works |data))], and their odds ratio [i.e., posterior odds/prior odds] for parts (a) and (b) above.

(d) Calculate the Likelihood ratio comparing the hypothesis that the regimen works versus the hypothesis that the regimen does not work [i.e., LR=P(8 success | regimen works) / P(8 success | regimen does not work)]. [The likelihood ratio should be equal to the odds ratio in part (c)].

(e) What do parts (c) and (d) imply about what should be reported at the end of a study, if you want to communicate the strength of evidence in these data? That is, does reporting the posterior probability at the end of the study suffice for describing what the data say? Or it is more important to report how the data change this probability? Justify your answer.

(f) Repeat parts (a)-(d) using the following data from a much larger study where 200 participants were enrolled and 80 survived. Notice that the estimated probability of 1-year survival was found to be the same as in the smaller study. Would the inventors and the experts agree now? What is the strength of evidence in the larger study and how much greater is it than the smaller study?

Useful *Stata* Commands

Stata provides cumulative probabilities such as $P(Y \geq k)$.
Note that $P(Y = k)= P(Y \geq k)- P(Y \geq k+1)$.

<u>Binomial distribution</u>
$Y \sim Bin(n,\theta)$

$P(Y \geq k)$ is given by the *Stata* command:
"display Binomial(n,k,θ)"

$P(Y = k)$ is given by the *Stata* command:
"display Binomial(n,k,θ)-Binomial(n,k+1,θ)"

<u>Poisson distribution</u>
$Y \sim Poiss(\lambda)$

You have two options:

(1) Use epoisson and cpoisson as described in lab 6

(2) Use the following:

$P(Y \geq k)$ is given by the *Stata* command:
"display gammap(k,λ)"

$P(Y = k)$ is given by the *Stata* command:
"display gammap(k,λ)- gammap(k+1,λ)"

<u>Standard Normal distribution</u>
$Z \sim N(0,1)$

$P(Z \leq k)$ is given by the *Stata* command:
"display normal(k)"

$P(Z \leq q)=a$ (where $0<a<1$). Now q is given by the *Stata* command:
"display invnorm(a)"