Principles of Inferential Statistics in Medicine

Final Exam – 513–607A, December 7^{th}, 1999.

1. In women, bone mineral density (BMD) continuously increases from childhood until peak bone mass is reached, usually between the ages of 25 to 30 years old. After a period of relative stability, BMD begins to decline at approximately 50 years of age. In order to investigate this phenomenon, a researcher collects data on 250 women between the ages of 10 to 70 years old. He then calculates the least squares regression line between age (x) and BMD (y), and finds that it is

$$\text{BMD } = 1 + 0 \times age$$

(a) What is your interpretation of the zero value for the slope?

(b) What you might suggest to improve on the statistical analysis for these data?

2. Animal studies have suggested that honey in a diet may raise hemoglobin levels. To test this hypothesis in humans, a study was conducted on six sets of twins. Within each twin pair, one twin was given milk with honey every night, while the other twin was given milk without honey. Suppose that the following increases in hemoglobin were recorded:

Twin Pair	1	2	3	4	5	6
Honey	16	12	6	17	28	9
No Honey	15	10	3	5	22	14

Perform the appropriate two-sided NONPARAMETRIC test to examine if there is a difference in hemoglobin increases between those given honey and those not given honey. State clearly the null and alternative hypotheses, calculate an approximate p-value, and state your conclusion.

3. An experiment is carried out to determine the rate of success for a particular type of surgery. The results are 75 successful surgeries in 100 cases. Two investigators analyze these data. Investigator 1 states that his posterior distribution is a beta(100, 50) distribution, while the other states that his posterior distribution is a beta(165, 60) distribution.

(a) What was the prior distribution for Investigator 1? What was the prior

density for Investigator 2?

(b) State whether you think these two investigators would closely agree on the probability of a successful surgery after seeing this data, and explain why.

4. The following data have been collected on the development of breast cancer and age at the birth of a mother's first child:

	Age at first child's birth		
Cancer	less than 20	20-29	30 or greater
Yes	200	300	200
No	600	1000	700

Carry out an appropriate test to determine whether there is evidence of a relationship between breast cancer and age at first birth. State the null and alternative hypotheses, carry out the test, and provide a conclusion.

5. Alzheimer's Disease is characterized by progressive cognitive impairment, with an average duration of approximately seven years. One way to measure cognitive impairment is through the Mini-Mental State Exam (MMSE). The MMSE ranges from a high of 30 points (no cognitive impairment) to a low of 0 (very severe cognitive impairment). The following data have been collected on a patient with Alzheimer's Disease who was followed over time.

age in years (x)	MMSE score (y)
62	25
65	13
68	12
73	10

(a) Calculate the intercept and slope of the least squares regression line that best fits these data.

(b) Give a careful interpretation to numbers you just calculated for the intercept and slope.

6. Five hundred investigators each set out to test a different null hypothesis. Unknown to them, all 500 of the null hypotheses happen in fact to be true.

(a) How many out of the 500 p-values they calculate do you expect to be significant, if significance is defined by $p \leq 0.01$?

(b) What proportion of their p-values would you expect to fall in between $p = 0.01$ and $p = 0.05$?

7. In a simple random sample ($n = 100$) of patients selected from a large family medicine unit, the average age was found to be 60 years with a standard error of 2 years. State whether each of the following statements are true or false, and explain why.

(a) Assuming that the ages are approximately normally distributed, it is estimated that about 95% of the ages of patients in that practice are within the range of $60 - 2 = 58$ years and $60 + 2 = 62$ years.

(b) An approximate 95% confidence interval for the mean age of patients from this family medicine unit is (56 , 64) years.

8. The information below summarizes the probability function of a discrete random variable X.

$$\begin{aligned} Pr\{X = -7\} &= 0.4 \\ Pr\{X = 5\} &= 0.6 \\ Pr\{X = \text{any other number}\} &= 0 \end{aligned}$$

(a) Based on the information provided above, calculate the mean and variance of the variable X.

(b) Define another variable by $Y = \frac{X+6}{3}$, that is, to form Y, add 6 to X, then divide by 3. Calculate the mean and variance of the new variable Y.

9. Five hundred randomly selected subjects participated in a survey on exercise patterns. Suppose that 150 out of the 500 persons who participated in the survey reported being members of a health club.

(a) Based on the above data, provide an approximate 99% confidence interval for the proportion of persons in that community who are members of a health club.

(b) Suppose that the exact underlying rate in the community is in fact 30%. If data from a new sample of 1000 subjects are collected, what is the approximate probability that between 200 and 300 (including the values of 200 and 300) of them will say that they are members of a health club?

10. Find the following probabilities:

(a) What is the probability that a chi-squared distributed variable with 10 degrees of freedom will be equal to or greater than 13.44?

(b) If a random variable $X \sim N(0, 1)$, calculate $\Pr\{X \geq 1.25\}$.