Principles of Inferential Statistics in Medicine

Final Exam – 513–607A, December 7^{th}, 2000.

1. It is suspected that carriers of the sickle cell gene may be less susceptible to malaria than non-carriers. The following table of data is collected:

		Malaria		
		+	–	
Sickle Cell Gene	+	10	140	150
	–	130	20	150
		140	160	300

(a) Test the null hypothesis of no association between malaria and carrying the sickle cell gene versus a two sided alternative. State the null and alternative hypotheses, carry out the test, and state your conclusion.
(b) Calculate a 95% confidence interval for the difference in rates of malaria between carriers and non-carriers of the sickle cell gene.
(c) What would you conclude from (a) and (b)?

2. Suppose that blood pressure values are approximately normally distributed in a given population. A sample of 100 individuals is randomly selected from this population, and the blood pressure (in $mm\ Hg$) of each subject is measured. The results were:

$$\begin{aligned} \overline{x} &= 80 \\ s^2 &= 100 \end{aligned}$$

(a) Provide a 95% confidence interval for the average blood pressure in that population.
(b) State whether this statement is true or false, and explain why: We expect that approximately 95% of the blood pressures belonging to members of this population fall between 60 and 100.

3. Two surgeons are arguing about the value of a new surgical technique. Surgeon 1 is enthusiastic, and therefore has a beta(70,30) prior probability distribution on the success rate of the technique. Surgeon 2 is more pessimistic, and therefore states that his prior probability distribution is beta(40,60).

(a) Suppose that the two surgeons agree to collect some data to settle the issue. They observe 300 patients, with 180 of these having successful surgeries. What

is the posterior distribution for the success rate for Surgeon 1? What is the posterior distribution for the success rate for Surgeon 2?
(b) Did Surgeon 1 or Surgeon 2 have a prior mean value which came closer to the mean success rate actually observed in the data given in part (a)?

4. A regression analysis is carried out to examine the relationship between drug dosage (x) and pain score (y) for arthritis patients. Drug dosage is measured in milligrams, while pain score is measured on a 0 to 100 scale, with higher numbers indicating more pain. The slope is found to be –15 points on the scale per milligram increase in dosage. The average dosage in the experiment was 1 milligram with a standard deviation of 0.5 milligrams, and the average pain score was 50 points, with a standard deviation of 10 points.

(a) Calculate the intercept of the regression line.
(b) Calculate the correlation coefficient between dosage and pain relief.

5. The following data were observed on a mouse:

age in weeks	weight (grams)
1	100
2	150
6	400

(a) Calculate the slope of the best fitting least squares line.
(b) Provide an interpretation for the slope you calculated.

6. An investigator is planning a study to estimate the proportion of children in a given city that have received appropriate vaccinations. He expects that 70% of children in this city will have had appropriate vaccinations, and wants the 95% confidence interval at the end of his study to have total length of 15%. What sample size does his study require, assuming his expected rate of 70% is correct?

7. Twenty investigators each set out to test a different null hypothesis. Unknown to them, all 20 of the null hypotheses happen in fact to be true.

(a) How many out of the 20 p-values they calculate do you expect to be significant, if significance is defined by $p \leq 0.05$?
(b) What proportion of their p-values would you expect to fall above the value $p = 0.5$? Explain your answer.

8. There is a new class of drugs that may be able to increase bone mineral density (BMD) in patients with osteoporosis. To test whether the drug works, the following data are collected on ten individuals, all of whom were given the drug for 12 weeks:

Patient #	BMD at baseline	BMD 12 after weeks on the drug
1	0.9	1.0
2	0.85	0.7
3	0.65	0.9
4	0.6	1.0
5	0.7	1.2
6	0.8	0.85
7	0.9	0.7
8	1.0	0.7
9	0.65	1.0
10	0.8	1.4

Perform the appropriate two-sided NONPARAMETRIC test to examine if there is a difference in BMD before and after the drug is given for 12 weeks. State the null and alternative hypotheses, show your calculations, and state your conclusion.

9. There are two ways to measure blood glucose level. Method 1 (x) is both easier to perform and is relatively inexpensive compared to Method 2 *(y)*. However, Method 2 is known to give more accurate results. In order to test if Method 1 gives results close enough to the Method 2 so that it can be used regularly, both tests are carried out on a group of 102 persons.

(a) Suppose that the correlation is found to be 0.999. Is this sufficient evidence to conclude that the two methods give near identical results? Explain your answer.
(b) Suppose that the regression equation calculated from the above data is

$$y = 0.01 + .92x,$$

and that the standard deviation of the estimate of β is 0.002. Calculate the p-value for the test of the hypotheses

$$\begin{aligned} H_0 &: \beta = 1 \\ H_A &: \beta \neq 1. \end{aligned}$$

What can you conclude about the equivalence of the two methods? What would your recommendation be concerning the use of the new test?

10. An investigator performs three diagnostic tests on each subject in a group of 100 subjects. One of the tests is a perfect gold standard (sensitivity = specificity = 100%) for the detection of the disease under study, and the other two tests are newly developed and under investigation. Assume that

Test 1 and Test 2 results are independent within each true disease state. The data are given below:

Truth (from Gold Standard)	Test 1 Result	Test 2 Result	Number of Subjects
+	+	+	25
+	+	−	5
+	−	+	10
+	−	−	3
−	+	+	1
−	+	−	5
−	−	+	1
−	−	−	50

(a) What is the sensitivity of Test 1, and what is the specificity of Test 2, according to the data above?
(b) Calculate a 95% confidence interval for the sensitivity of Test 1.