Principles of Inferential Statistics in Medicine

Final Exam – 513–607A, December 3^{rd}, 1998.

1. A study was carried out to assess the average weight gain in adults aged 50 years and over. A sample of 100 adults aged between 50 and 80 years old was collected, and a regression line was calculated. Weight was measured in Kg, and age in years. Suppose that the best fitting regression line was:

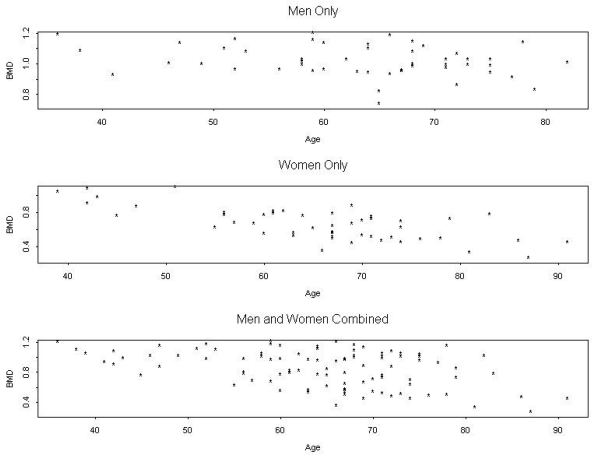
$$\text{weight} = 75 + 3 \times \text{ age}$$

(a) Provide an interpretation for the slope of the regression line that was found here.
(b) Do you think that this linear regression line would remain approximately valid for subjects in the range of 5 to 10 years old? Explain why or why not.
(c) Twenty of the 100 subjects were between 70 and 80 years old in this sample. If a regression line is calculated using these 20 subjects only, the best fitting line is found to be:

$$\text{weight} = 70 - 2 \times \text{ age}$$

State two possible reasons why the slope may have changed in direction compared to the slope that was calculated using all of the subjects.

2. A researcher is deciding whether to combine the data for men and women in investigating the decline of bone mineral density (BMD) with age. He looks at the following scatter plots:



State whether you would recommend that this data be combined or not by discussing the advantages and disadvantages of combining the data into a single analysis verses treating them separately. Assume that a simple linear regression analysis will be performed.

3. Two bone specialists are debating the value of a new technique for diagnosing osteoporosis. Specialist 1 is enthusiastic about the technique, and therefore has a beta(80,20) prior probability distribution on the sensitivity of the test. Specialist 2 is more pessimistic, and therefore states that his prior probability distribution is beta(5,5).

(a) Suppose that the two specialists agree to collect some data to help to settle the issue. They observe 1000 patients with osteoporosis, and the test correctly identifies 700 of these subjects as osteoporotic. What is the posterior distribution for the sensitivity according to the prior density of Specialist 1? What is the posterior distribution for the sensitivity according to the prior distribution of specialist 2?

(b) Do you think that the opinions of these two specialists will remain different or become quite similar after observing the data from part (a)?

4. There is a new class of drugs that may be able to increase forced expiratory volume (FEV) in patients with asthma. To test whether a new drug works, the following data are collected on 12 individuals, all of whom were given the drug for 3 months:

Patient #	FEV at baseline	FEV after 3 months on the drug
1	15	22
2	20	32
3	18	15
4	34	38
5	25	30
6	14	24
7	27	36
8	29	21
9	31	25
10	19	32
11	24	35
12	22	24

Perform the appropriate two-sided NONPARAMETRIC test to examine if there is a difference in FEV before and after the drug is given for 3 months. State the null and alternative hypotheses, show your calculations, and state your conclusion.

5. The following data from three different subjects were observed:

HDL cholesterol (X)	Total cholesterol (Y)
40	150
50	250
60	280

(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 aged 6 to 8 years old who, on a given day, are absent from school. She expects that 7% of children will be absent, and would like the 95% confidence interval at the end of this study to have total length of 3%.

(a) What sample size does this study require, assuming the expected rate of 7% is correct?

(b) For logistical reasons, taking a random sample of all school children in the city is impossible, so the investigator plans to select a random sample of classrooms within a random number of schools, and survey each child in the selected classrooms. Do you see any statistical complications arising from this sampling plan?

7. At a busy emergency ward, data are collected on the number of patients seen per day. Data are collected on 60 randomly selected days throughout the year. On average, $\overline{x} = 225$ patients were seen each day, with an observed standard deviation of $s = 25$ patients per day.

State whether each statement is true or false, and explain why:

(a) An approximate 95% confidence interval for the average number of visits per day to this emergency ward is (176, 274).

(b) In order to calculate a 95% confidence interval for the average number of visits per day, one either has to assume that the number of visits are normally distributed throughout the year, or that a sample size of 60 days is sufficient for the Central Limit Theorem to apply to this problem.

8. A regression analysis is carried out to examine the relationship between drug dosage and weight loss for patients taking a drug to induce weight loss. Drug dosage is measured in milligrams, while weight loss is measure in Kg. The regression slope is found to be –1 Kg of weight lost per milligram increase in dosage. The average dosage in the experiment was 2 milligrams with a standard deviation of 1 milligram, and the average weight loss was 5 Kg, with a standard deviation of 0.75 Kg.

(a) Calculate the intercept of the regression line.

(b) Calculate the correlation coefficient between dosage and weight loss.

9. In a study of a new drug versus a placebo, a hypothesis test has just resulted in a p-value of exactly 0.05. Suppose that the investigators now want to carry out another identical trial for the same comparison. That is, they want to replicate the study. Assume that the means and standard deviations observed in the first trial are in fact, by coincidence, exactly the correct true values. If they also use the same sample sizes as in the first study, what is the probability that they will observe a p-value that is equal to or less than 0.05 in the second study?

10. A pilot study is conducted in 10 patients to collect preliminary information about the safety of a new drug. The drug will be considered as safe if no adverse reactions are seen within three days following administration of the drug. Suppose that out of the 10 patients, only one has an adverse reaction within three days. Test the null hypothesis that the rate of adverse reactions is equal to or less than 5% versus the one-sided alternative that the rate is greater than 5%. State your null and alternative hypotheses, calculate an EXACT p-value, and provide your conclusion.