Sample Final PH5450, Fall 2001

1. (10 points) Suppose a health care maintenance firm in the past received an average payment from its customers of 100 dollars a month. Now suppose the firm raises its prices by 20 dollars for one quarter of its customers and 40 dollars for the rest. Assume no one joins or leaves the plan after the price change.

   A. If you have enough information to compute the answer, what is the average payment made by customers of this company after the price change? If you think not enough information is provided, what other information would you need?

2. (5 points) According to the 1990 U.S. census, states which have an above average number of people who fail to complete high school tend to have an above average number of infant deaths. What is a plausible explanation for this association?

3. (10 points) Suppose we recruit 12 subjects with chronic fatigue syndrome (8 men and 4 women) for a randomized, placebo controlled experiment to determine if shark fin as a dietary supplement increases the energy level of the patients.

   A. If we think gender will be an important factor in determining the effect of shark fin we should use a randomized block design. Explain how we would conduct such an experiment.

   B. Explain why the randomized block design is useful if we think the effect of the dietary supplement depends on gender.

   C. If one of the researchers conducting the experiment is to evaluate changes in the energy level of the patients, what experimental procedure should we employ to prevent any biases of the researcher from impacting the results?

4. (5 points) Suppose a couple is equally likely to have a boy or a girl, and the couple has 3 children.

   A. What is the probability that 2 of the children are of the same sex?

   B. If we look at 1,000 families with 3 children, how many would you guess have 2 children of the same sex?

   C. What theoretical result have you used to answer part B of this question?

5. (5 points) We know that, for large n (under certain conditions), $\overline{x}$ has a sampling distribution which is approximately normal with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. Use this to find the approximate sampling distribution of the sample total $\sum_{i=1}^{n} x_i$.

6. (5 points) Suppose we take a simple random sample and obtain a confidence interval for the mean of the population with a margin of error of 0.27. By what factor must we increase the sample size to cut the margin of error in half?

7. (5 points) Researchers are studying the reduction in blood pressure, due to a certain drug, in 2 different patient populations. To summarize their findings, the researchers construct 2 95% confidence intervals for the reduction in blood pressure, one for each patient group. What is the probability that at least one of the confidence intervals contains the true mean reduction in blood pressure due to the drug for that patient population?

8. (15 points) Suppose there are 17 hospitals in some city, and for each hospital we have the average number of years that patients survive after heart surgery, for patients having the operation done in the last 10 years. Furthermore, many (i.e. at least 50) such operations were performed at each of these hospitals over the last 10 years.

   A. For each hospital, we have the average number of years of survival. Do you think these 17 numbers are well modeled as normal variables? Explain your answer.

   B. If we average over all hospitals we find that the average number of years is 5.4 and we find a standard deviation of 1.3. Compute the standard error of the mean.

   C. Test the hypotheses
   $$H_0 : \mu = 5.0$$
   versus the 2-sided alternative. Can we reject H_0 with an α-level of 0.05?

9. (10 points) A researcher wanted to compare 2 exercise routines with the objective being to determine which resulted in the highest heart rate. A group of 50 volunteers is randomly allocated into 2 groups of equal size, group 1 and group 2. The mean heart rate in group 1 was 90 beats per minute (bpm) with a standard deviation of 9 bpm, while the mean in group 2 was 96 bpm with a standard deviation of 12 bpm.

   A. Construct a 90% confidence interval for the difference in the means of the 2 groups with a 2-sided alternative. The researchers feel 9 bpm is really different from 12 bpm, so do not pool the variances from the groups.

   B. Test the hypothesis that the mean is the same in both groups with an α-level of 0.05. Once again, do not pool the variances.

10. (20 points) A study was done in 1963 to measure the effectiveness of early screening for detecting breast cancer. The subjects were 62000 women aged 40 to 64. These women were divided into two equal sized groups. In the treatment group, women were encouraged to come in for annual screening, including examination by a doctor and x-rays. Of the 31000 women in the treatment group, 10800 of the women refused to participate. The placebo group received standard health care (for 1963). The subjects were then followed until death. Results for the first five years (summarized as the number of deaths due to breast cancer and the number of deaths due to any other cause, for each group) are presented in the following table.

	Cause of Death	
	breast cancer	all other
treatment group		
Examined	23	428
Refused	16	409
control	63	879

A. How would you use this data to argue that screening is or is not beneficial?

B. Find the relative risk of developing breast cancer when one gets screening done, and interpret your result. (Since screening is a "protective exposure" the relative risk would be less than one if screening really does lower the risk of breast cancer.)

C. Compute a 95% confidence interval for the odds ratio for the effect of screening. Can we reject the hypothesis $H_0 : OR = 1$ based on this confidence interval?

D. Compare those who were assigned to the treatment group but refused treatment to the control group. Use a statistical test to see if those who refused treatment have a higher death rate from "all other causes"?

E. Use your answer from part D to try to explain why those who refused treatment had a lower incidence of breast cancer than the women in the control group.

11. (10 points) Some researchers were interested in studying what factors influence birth weight. To this end, they obtained a sample of 680 mothers and found out how many weeks the mother was pregnant prior to delivery, x_1, about how many cigarettes the mother smokes a week, x_2 and the weight of the baby (in kg) at delivery, y.

A. If we regress y on x_1, so we have the model

$$y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i,$$

we get an estimate of β_1 of 0.248. Explain what this regression coefficient means as thoroughly as you can.

B. What is wrong with the argument: if β_1 is about 0.25 then the baby gains a kilo every month, so β_1 can't be 0.25 because newborns never weigh 9 kilos.

C. Why did the researchers collect data on x_1 if they are really interested in the relationship between x_2 and y?