STATISTICS 210 – Fall 2009 – Homework 1

Handed out: Wednesday September 30, 2009
Due: Wednesday October 7, 2008

Reading:	Sep. 26 - Oct. 6	Two-sample methods (App A.6-A.7)
	Oct. 8 – Oct. 20	One factor ANOVA (Chap 16-18)

NOTE: There is no SAS on this homework. I believe it makes the most sense to have a SAS orientation at the next discussion before assigning SAS work.

1. **Review of one-sample methods:** We did not discuss the one sample t-test in class. The argument follows closely the one given in class for two samples. You can also find it in almost any introductory statistics text. Here's a setting where one might do a one sample test:
   A large state University is concerned about the length of time required for undergraduates to complete their degree. The average time for all students is 4.7 years. It is claimed that students who take a freshman seminar course (an experimental program at the University) graduate faster and the Office for Institutional Research (OIR) would like to check this claim. Of course it is difficult to go through all the records and identify students who have participated in the program. A random sample of 40 students who participated in the program is taken and their records reviewed. For this sample the average time to degree is 4.5 years and the standard deviation is 0.6 years.
   (a) Identify the population parameter of interest and the hypothesis that OIR would like to test.
   (b) Carry out a test of the hypothesis and report the result. Your answer should be **explicit** about the alternative hypothesis considered and p-value obtained. You should give a one sentence summary of your conclusion in addition to the p-value.
   (c) Provide a 95% confidence interval (CI) for the population parameter of interest. How does the CI relate to your test result?
   (d) Is this an experiment or an observational study? Explain. If it is an observational study, then identify a possible confounding variable and describe how you might eliminate it as a potential confounder.
   (e) Years to graduation is (sadly) a highly-skewed random variable. Does this effect your confidence in the results of parts (b) and (c)? Explain.

2. **Two-sample confidence interval:** In one weight-loss study 89 sedentary men were randomly assigned to either a special diet or exercise for a year. Forty-two men were placed on a diet and they lost an average of 7.2 kg with a standard deviation of 3.7 kg. The other 47 men were put on an exercise program and they lost an average of 5.3 kg with a standard deviation of 3.9 kg.
   (a) Are the assumptions for a two sample t procedure met in this case? Explain why or why not.
   (b) Regardless of your answer to (a) use the two sample t procedures to find a 95% confidence interval for the difference between the mean weight loss via diet and the mean weight loss via exercise.
   (c) Give a 90% CI for the difference in means? Give a 99% CI for the difference in means?
   (d) Use the CI results to tell what p-value would be obtained in a two-sample t-test. (Hint: You can't give a precise value, only an interval. You can do the test to check but should report the answer to the question that is asked!)
   (e) In studies like this investigators are not always able to monitor whether subjects complete or comply with the treatment they are given. A subject may not do the exercise program for example. In that case one might compare only those who complete the assigned treatment. Explain why this might lead to a misleading comparison.

3. **Two-sample test:** One of the ways that small businesses advertise is to put a flyer or paper on the windshield of every car in a large shopping center parking lot. (Boy is that annoying!) A study was carried out to compare the response rate to advertisements printed on different colored paper. Twelve shopping center parking lots were randomly assigned to receive orange or red ads. For each parking lot the response variable is the number of phone calls received in the next 24 hours divided by the number of flyers placed. We'll call this the response rate (it's not exactly a percentage because it's possible though highly unlikely that you could get more calls than flyers if people shared the flyer). The response rates are summarized below:

	Orange	Red
n	6	6
mean	3.1	2.7
s.d.	0.8	0.6

   (a) Is this study an experiment? Explain.

(b) Define the population of interest, the experimental units, the treatment and the response.

(c) An alternative (easier) design might use the individual car as a unit. Then you could just use each color in one parking lot and take the response from each car as 0 (if nobody called) or 1 (if somebody called) – you'd have to identify each flyer with a code or something. Explain any concerns you might have about this design.

(d) The **effect size** is a measure of the difference in two population means (i.e., it measures the effect of the treatment). There are various definitions for effect size depending on the study. For a two sample study the effect size is ususally defined as $\mu_2 - \mu_1$ (if the units are meaningful) or $(\mu_2 - \mu_1)/\sigma$. These can be estimated by plugging in sample means and the pooled s.d. Using the second definition what is the effect size for this study? (Note: Interpretation of effect sizes depends on the field. Note that 1.0 would be a very large effect, shifting the entire population by 1.0 standard deviations)

(e) Show that the usual pooled two-sample t-test statistic can be written as the effect size times some function of sample size.

(f) Carry out a test of the null hypothesis that the two colors are equally effective. (Report results as described above in (1b))

(g) Suppose the study is repeated with four times as many parking lots (n=24 in each group). You get EXACTLY the same mean and s.d. Would the effect size estimate change? Would the p-value change? Interpret these answers for the investigator (e.g., did the second study work better?)

4. **Theory – Power calculation I:** A key advantage of the model-based approach to inference is that the model can be used to plan studies. This problem revisits the two-sample power calculation discussed briefly in class. Suppose that we will be collecting data from two populations and are willing to assume that the samples $Y_{11}, \ldots, Y_{1n}$ are iid $N(\mu_1, \sigma^2)$ and the samples $Y_{21}, \ldots, Y_{2n}$ are iid $N(\mu_2, \sigma^2)$ as required by the model. To keep things simple here let's assume σ is known and that therefore we can use normal (rather than t) tables in carrying out our tests.

   (a) We wish to test the null hypothesis that $\mu_2 = \mu_1$ versus the alternative hypothesis that $\mu_2 > \mu_1$. We are going to use a size α test. Show that the usual z-test rejects H_o if $\bar{Y}_2 - \bar{Y}_1 > c$ and identify c (it should involve σ, n and $z_{1-\alpha}$).

   (b) To plan the study (i.e., determine an appropriate sample size) we need to specify the alternative that we believe may be true. Assume that we hope/expect to find $\mu_2 - \mu_1 = \delta > 0$. If this alternative is true, then find an expression for the probability that the null hypothesis is rejected in the test you developed above. This probability is known as the power of the test for the alternative δ.

   (c) Sketch the power as a function of δ. Put in as much detail as you can – e.g., what is power when $\delta = 0$? what happens as δ increases?

   (d) Suppose we would like to have power $1 - \beta$ (where β is the type II error rate). Find the formula for the sample size required to obtain power $1 - \beta$ for the alternative δ using a size α test. You are basically being asked to derive the sample size formula given on page 27 of the slides (except for a one-sided test).

5. **Theory (?) - Power calculation II:** We next consider how the idea behind the power calculation generalizes to other settings. When data are collected regarding a categorical variable, the chi-squared test is often used to test a null hypothesis about the probabilities of the categories. Suppose that we consider the number of male children in the population of households with two children. There are three possible outcomes 0, 1 or 2. Let p_o, p_1, p_2 denote the proportion of the population with 0, 1, 2 male children.

   (a) A natural null hypothesis is $H_o : p_o = 0.25, p_1 = 0.50, p_2 = 0.25$. Explain why this is the natural null hypothesis.

   (b) This can be tested by using the chi-squared test statistic. If n_i is the number of sample households with i males and $N = n_o + n_1 + n_2$ is the total number of sample households. Then $\chi^2 = \sum_{i=0}^{2} (n_i - Np_i)^2/(Np_i)$ is computed and compared to the chi-square distribution with two degrees of freedom. One biological theory holds that in fact two male and two female child families occur more than the null would suggest. A local researcher would like to develop a test of the theory. Explain why $H_a : p_o = .25 + \delta/2, p_1 = 0.50 - \delta, p_2 = 0.25 + \delta_2$ makes sense as a one parameter alternative to the null hypothesis. Are there any conditions on δ? Explain.

   (c) HARD - Suppose we decide to use a .05 test. This test will reject if $\chi^2 > 5.99$. If $\delta = .05$, then how large a sample N is required to achieve power .80. (Hint: This is not trivial. I'd like you to explore but it is not a research project in which I want you to go learn a bunch of new statistical theory. Effort counts. You can try to work with the χ^2 statistic to get some ideas as to how large it would be for different N. An alternative is to use simulation (try a few N's and see what power you achieve) – this can be done in Excel or R or any other computing tool you have access to.)