**Biostatistical Methods I (BIOS 5710)**
**Breheny**

Assignment 8
Due: Wednesday, November 5

1. You have developed a new, cholesterol-lowering drug as a treatment for patients with high cholesterol. From past studies, you know that the standard deviation of cholesterol levels for patients in the population you're looking at is about 30 mg/dl. You assume that your new drug will reduce cholesterol levels by about 10 mg/dl on average, and that the placebo will have no effect on patients' cholesterol levels.

   (a) Suppose you carried out a two-sample study, randomizing separate patients to drug and placebo. What is the power of a two-sample $t$-test with $n = 100$ subjects in each group given the assumptions above?

   (b) Suppose instead that you designed a crossover trial to test the efficacy of the drug. If a patient's cholesterol level on drug and a patient's cholesterol level on placebo were completely independent, what is the standard deviation of the difference between a patient's level on each treatment?

   (c) Using the standard deviation you calculated in (b), what is the power of a paired $t$-test with $n = 100$ subjects?

   (d) Comment on the difference between the power you calculated in (a) and in (c). In particular, which was the more powerful test, and what can you conclude from this comparison?

   (e) Do you feel the assumption of independence in (b) is believable? Why or why not?

2. A researcher has developed a preventative therapy that she expects to cut the risk of diabetes in half. In the population she is researching, the current prevalence of diabetes is 20%.

   (a) Suppose she carries out this test with $n = 100$ subjects in the treatment group and $n = 100$ subjects in the control group. What is her power to detect a difference?

   (b) How many subjects does she need in each group in order to achieve at least 80% power?

3. Consider representing observations from a binomial distribution as a $1 \times 2$ contingency table, where one cell counts the success and the other cell counts the failures. Suppose we carried out a $\chi^2$ test of $H_0 : \pi = \pi_0$ on this data by calculating

$$X^2 = \sum_{i=1}^{2} \frac{(O_i - E_i)^2}{E_i}$$

   and comparing it to a $\chi^2_1$ distribution. Show that this test is identical to the $z$ test we discussed in class (Oct. 6), in the sense that the two tests will always produce identical $p$-values.

4. Consider testing for a difference in proportions by treating proportions as means and using a central-limit theorem-based approach. In particular, let $\hat{\pi}_1$ denote the sample proportion in group 1, $\hat{\pi}_2$ the sample proportion in group 2, and $n_1$ and $n_2$ the respective sample sizes. Then

$$\frac{\hat{\pi}_1 - \hat{\pi}_2 - \Delta}{\text{SE}} \sim \text{N}(0, 1)$$

   for some $\Delta$ and SE.

(a) What is $\Delta$? What does $\Delta$ equal under the null hypothesis of no difference between groups?

(b) What is SE?

(c) Consider estimating SE under the null by calculating $\hat{\pi}$, the overall proportion obtained by pooling both groups together, then plugging that value in for $\pi_1$ and $\pi_2$ in your expression for SE. Carry out this test for the Lister experiment we discussed in class.

(d) Compare your answer to the result from the $\chi^2$-test we carried out in class and (briefly) discuss their relationship.