BMI 713: Computational Statistics for Biomedical Sciences

Assignment 5

October 14, 2010 (due October 21)

Contingency Tables

1. (Taken from [1]) A statistical analysis that combines information from several studies is called a meta-analysis. A meta-analysis compared aspirin with placebo on incidence of heart attack and of stroke, seperately for men and from women (*J. Am. Med. Assoc.*, **295**: 306-313, 2006). For the Women's Health Study, heart attacks were reported for 198 of 19,934 taking aspirin and for 193 of 19,942 taking placebo. We are interested in whether aspirin was helpful for reducing the risk of heart attack.

   (a) State the null hypothesis and the alternative hypothesis.

   (b) Construct the 2×2 table that cross classifies the treatment (aspirin, placebo) with whether a heart attack was reported (yes, no).

   (c) Perform the chi-square test. Report the test statistic, the degree of the freedom and the P-value. What conclusion can you draw from this test?

2. Sir Ronald Fisher, a statistician and geneticist, described a tea tasting experiment in his book *The design of Experiments* to illustrate his test – now known as Fisher's exact test. A colleague of Fisher claimed that she could distinguish whether milk or tea was added to the cup first. To test her claim, Fisher designed an experiment in which she tasted eight cups of tea. Four cups had milk added first, and the other four had tea added first. She was told there were four cups of each type and she should try to select the four that had milk added first. The cups were presented to here in random order. Table 1 shows a possible result of this experiment.

Table 1: Fiser's Tea Tasting Experiment

	Guess Added First		
Added First	Milk	Tea	Total
Milk	3	1	4
Tea	1	3	4
Total	4	4	

   (a) State the null hypothesis and the alternative hypothesis.

   (b) Perform the Fisher's exact test. What is the P-value. What conclusion can you get?

3. (Simpson's paradox). The result that a marginal association can have different direction from the conditional associations is called Simpsons paradox. This result applies to quantitative as well as categorical variables. To illustrate Simpson's paradox, here we use an example in medical study [2, 3] comparing the success rates of two treatments for kidney stones. The two treatments are open surgery (treatment A) and percutaneous nephrolithotomy (treatment B). The patients can be classified into two groups according to the kidney stone size, small stone group and large stone group. Table 2 shows the surgery results of 700 patients under the two treatments.

   (a) What is the overall success rates of treatment A and treatment B? Based on this result, which treatment is better? Perform a proper test. What is the P-value?

   (b) For the small stone group, calculate the success rates of treatment A and treatment B. Which treatment if better? Perform a proper statistic test and report the P-value.

   (c) Repeat the above analysis for the large stone group.

Table 2: Success rates for different groups of stone size.

		Treatment	
Group	Treatment result	A	B
Small stone	Success	81	234
	Failure	6	36
Large stone	Success	192	55
	Failure	71	25
Both	Success	273	289
	Failure	77	61

(d) Are the conclusion from (a), (b) and (c) consistent? If not, can you explain why?

4. **For extra credit** (Conservativeness of Fisher's exact test). For small samples, because of the discreteness of the exact distribution used in Fisher's exact test, Fisher's exact test tends to be conservative, i.e. the real type I error rate is smaller than the nominal significance level. Here we use simulation to study this phenomenon.

   (a) Generate two random numbers `n11`, `n21` from the Binomial distribution Binom$(10, 0.5)$.

   (b) Use the following command to construct a matrix `A`, `A = matrix(c(n11,10-n11,n21,10-n21),nrow=2)`

   (c) Use the matrix `A` as the input table and perform Fisher's exact test. What is the P-value? At the significance level 0.05, do you reject the null hypothesis?

   (d) Repeat the above steps 1000 times. How many times do you reject the null hypothesis at the significance level 0.05? How many times do you expect to reject the null? Are these two numbers close?

References

[1] Agresti, Alan (2007) An Introduction to Categorical Data Analysis. Wiley

[2] Charig, C. R. D. R. Webb, S. R. Payne, and O. E. Wickham (1986). Comparison of treatment of renal calculi by operative surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy. In *British Medical Journal*. 292:897-882.

[3] Julious, S. A., and Mullee, M. A. (1994). Confounding and Simpson's paradox In *British Medical Journal*. 309:1480-1481