STAT 703/J703
March 20th, 2007
-Lecture 17-

Instructor: Brian Habing
Department of Statistics
LeConte 203
Telephone: 803-777-3578
E-mail: habing@stat.sc.edu

Today

- Example 1: Test for a Multinomial Distribution
- Example 2: Two Sample Tests for Location

Example 1: Consider a data set that could have come from a binomial distribution with n=5, but may also have come from a hypergeometric or some other distribution.

X	0	1	2	3	4	5
#obs	9	21	16	10	4	0

Test H_0: X is binomial vs. H_A: it isn't

So in general we get

$$G^2 = -2\log\Lambda = 2\sum_{i=1}^{n} O_i \log\left(\frac{O_i}{E_i}\right)$$

Page 342-343 demonstrates that this is asymptotically the same as the classic

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

Theorem A pg.310: Under smoothness conditions on the pdf, the null distribution of $-2\ln\Lambda$ has an approximate chi-square distribution with d.f.=$\dim\Omega - \dim\omega_0$ for large n.

Note: You must be careful what estimates you use for the parameters!!!

Example 2 Let $X_1, \ldots X_n$ be an independent random sample from a population that is normal with mean μ_x and variance σ^2, and let $Y_1, \ldots Y_m$ be an independent random sample from a population that is normal with mean μ_Y and variance σ^2, such that both samples are independent.

Consider testing about the means.

The likelihood ratio statistic statistic says to reject for large values of

$$1+\frac{mn}{m+n}\left(\frac{(\bar{x}-\bar{y})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2-\sum_{i=1}^{n}(y_i-\bar{y})^2}\right)$$

What if the data is the data is not normal?

What if we don't know the distribution of the data?

In this case a new paradigm must be used.

Many procedures in statistics are based on replacing the data with ranks.

In the Mann-Whitney-Wilcoxon test all of the data in both groups are replaced by their ranks from smallest to largest.

The population that tends to have larger values will have more large ranks.

Another way to look at it is by considering if the probability that a randomly chosen member of one group is less than a randomly chosen member of another group.

You can show that the number of pairs where $X<Y$ is equal to the sum of the ranks of $Y - m(m+1)/2$.