STAT 703/J703
March 29th, 2005
-Lecture 21-

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

STAT 703/J703   B.Habing   Univ. of SC

1

Today

• Exam 2 Solutions

STAT 703/J703   B.Habing   Univ. of SC

2

1-4: Many processes are observed only once a day to see if a certain event occurred.   For example, did your car fail to start the first time it was tried in the morning, did the computer network crash sometime over night, etc… One way of estimating p=chance of failure per day would be observe a single process, assume it was a binomial experiment, and observe what day it fails on first.

STAT 703/J703   B.Habing   Univ. of SC

3

---

1) Find the likelihood ratio test of H_0: p=0.3 vs. H_A: p=0.2 that has α as close to 0.05 as possible without exceeding it. Specify the statistic you are using, the rejection region, and the actual α-level. (Hint: `help(dgeom)`)

2) According to the test in (1), if the first observed failure occurs on day 7 what is the p-value?

3) Find the power of the test in (1) if p=0.2.

4) Verify or not the test you found in (1) is UMP for H_A: p<0.3 .

5) Someone could complain that we might have to wait a long time after the experiment is started before we can even conduct the test. This is because the day of the first failure can be any integer greater than zero (1, 2, 3, …1000, 1001, …). Why isn't this really a problem?

On page 267 a formula for the 100(1-a)% confidence interval for s^2 for a random sample from a normal population is given as:

$$\left(\frac{n\hat{\sigma}^2}{\chi^2_{df=n-1,\alpha/2}}, \frac{n\hat{\sigma}^2}{\chi^2_{df=n-1,1-\alpha/2}} \right)$$

where is the MLE for the variance and is the upper a^{th} percentile point for a chi-squared distribution with $n-1$ degrees of freedom.

6) Use the above formula to say how you would test H_0: $\sigma^2 = \sigma_0^2$ vs. H_A: $\sigma^2 \neq \sigma_0^2$.

7) Test H_0: $\sigma^2 = 2$ vs. H_A: $\sigma^2 \neq 2$ for the following sample that is from a normal distribution.

```
2.26    1.40    -0.64    0.31    2.88
0.65    0.00     4.37   -0.81    1.56
```

8-10: We saw in Lecture 6 that the likelihood for logistic regression could be written as:

$$lik(\alpha, \beta \mid y_1, \ldots y_n) = f(y_1, \ldots y_n \mid \alpha, \beta)$$

$$= \prod_{i=1}^{n} \left(\frac{1}{1 + e^{-(\alpha + \beta x_i)}} \right)^{y_i} \left(1 - \frac{1}{1 + e^{-(\alpha + \beta x_i)}} \right)^{1 - y_i}$$

In addition to estimating the shape of the curve we might want to actually test whether there is a non-zero relationship by testing H_0: $\beta = 0$ vs. H_A: $\beta \neq 0$.

8) Briefly explain why we can't simply use the Neyman-Pearson Lemma to conduct this test and need to use the generalized likelihood ratio test.

9) When constructing the Λ for the generalized LRT, we need to get values for both the numerator and denominator. Which one of those (the numerator or denominator) can we get simply by plugging the MLEs from the current computer code into the formula given above? Briefly say what would need to be changed in the computer code to get the other. (You don't actually need to change the code or try it out, just say what you would change).

10) How would we use the Λ to test the hypothesis at the $\alpha=0.05$ level? (e.g. How do we transform Λ, what distribution do we need to look up the critical region in, what is/are its degrees of freedom, and what is the cut-off value).

11) A random sample of size 40 supposedly comes from a geometric distribution with p=0.4. Zero was observed twelve times, 1 eight times, 2 six times, 3, 4, 5, and 6 one time each.   Use the LRT for contingency tables to test the null hypothesis that the data does indeed come from the specified distribution at the $\alpha=0.05$ level. (You may ignore the issue of some cells possibly being two small).