Stat 61S Fall 2015 Assignment 8 - due Monday, Nov. 9 by 2pm

Readings: Rice 9.1, 9.2, 9.3

1. A test for Lyme disease is positive with probability 0.9 for a person who has Lyme, and is negative with probability 0.95 for a person who does not have Lyme. You want to test H_o: "no Lyme" vs. H_1 : "Lyme", and you will reject H_o if the test is positive. What are the significance level and the power of this test?

2. Suppose the lifetimes of a certain component follow independent $\text{Exp}(1/\mu)$ distributions (so the expected value of the lifetimes is μ). For a random sample of $n = 60$ components, the measured lifetimes $X_1, \ldots, X_n$ have an average lifetime $\bar{X} = 120$ time units. On your last assignment you saw that the sampling distribution of $\bar{X}$ is Gamma.

   a) **Revised:** Find the distribution of $Y = 2n\bar{X}/\mu$ and use this as a pivot to construct an exact 95% confidence interval for μ (this is very similar to problem 4 from the Class 24 Likelihood Practice problems, and you will need to use Table 3).

   b) Compare your interval in part a to the 95% CI you obtain using a Normal approximation to the distribution of $\bar{X}$ (using your estimate $\hat{\mu}$ to obtain a standard error for $\bar{X}$).

3. Suppose $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Unif}(0, \theta)$. For example, the X_i values might represent the ages of n fossil finds for a particular organism, with θ marking the time of extinction. We found that, for this model, $Y = \max(X_1, \ldots, X_n)$ is a sufficient statistic for θ. The pdf for Y is as follows (see section 3.7 if you want to know how to derive this pdf):

$$f_y(y) = \frac{ny^{n-1}}{\theta^n}, \qquad 0 < y \leq \theta.$$

   a) Find the mean of Y and use this to identify an unbiased estimate for θ.

   b) **Revised:** Find the CDF for the random variable Y and use this to solve for the value c such that $P(Y > \theta/c) = 0.95$. Show that $P(Y < \theta < cY) = 0.95$, and explain why this implies the interval (Y, cY) defines a 95% confidence interval for the parameter θ.

   c) Evaluate the unbiased estimate and the interval bounds assuming you observe $X_1 = 45$, $X_2 = 100$, $X_3 = 25$, and $X_4 = 72$.

4. A handwriting expert claims he can identify a psychotic subject based on a writing sample. Five pairs of writing samples are prepared, each pair consisting of one sample from a person judged to be psychotic, and one sample from a subject judged not to be psychotic. The expert identifies the psychotic subject in 4 out of the 5 trials. State hypotheses for testing whether the expert can do better than guessing (i.e., use a 1-sided alternative) and compute the p-value for the test. Would you say there is strong evidence that the expert is doing better than guessing? Explain.

5. You observe $X_1, \ldots, X_{25} \stackrel{\text{i.i.d.}}{\sim} N(\mu, 10^2)$ and want to test $H_o : \mu = 0$ vs. $H_1 : \mu = 5$.

    a) What data values would lead you to reject H_o at $\alpha = 0.05$? At $\alpha = 0.01$?

    b) If you work at the $\alpha = 0.05$ significance level, what is the power of the test?

    c) Find the smallest sample size you would need to have power of at least 0.9 when working at $\alpha = 0.05$.

6. In a study of human body temperatures, Fahrenheit body temperatures were measured for a representative random sample of $n = 64$ healthy adult female subjects. The average temperature was $\bar{x} = 98.36°F$ with a sample standard deviation of $s = 0.68°F$. Assume that the distribution of temperatures in the population is close to Normal.

    a) Suppose for now that the true standard deviation of temperatures for women is known to be $\sigma = 0.7°F$. Find 95% and 99% confidence intervals for μ, the true mean temperature for all women.

    b) It is commonly assumed that "normal" human body temperature is $98.6°F$. State hypotheses for testing whether the true mean temperature for healthy adult women is $98.6°F$.

    c) Assuming $\sigma = 0.7$, compute the p-value for the test, having observed an average temperature $\bar{x} = 98.36$ for a random sample of $n = 64$ women. Explain precisely what the p-value represents in this example.

7. When you spin a coin, it will land heads-up with probability θ and tails-up with probability $1 - \theta$. To test $H_o : \theta = 0.5$ vs. $H_1 : \theta \neq 0.5$, you spin a coin $n = 100$ times independently and record X, the number of times the coin lands heads-up. Rather than using a set value α to determine the rejection region, you decide you will reject H_o if $|X - 50| > 10$.

    a) Use a Normal approximation with the continuity correction to find the approximate significance level α implied by this decision rule.

    b) Evaluate the power of this test for $\theta = 0.6$ and for $\theta = 0.7$ (note that, due to symmetry, these values will be the same as the power for $\theta = 0.4$ and for $\theta = 0.3$). Sketch a graph of the power as a function of θ.

8. What is the proportion θ of the globe that is covered by water? One way to estimate θ would be to choose n points independently and uniformly on the globe and calculate the sample proportion of points that are in water. In Stat 1 students simulate this process by tossing an inflated globe from person to person and noting for each catch whether the person's right index finger is touching water or land. We assume the trials are independent and that there is probability θ that a particular touch is a water touch. In one class we had 40 trials with 30 water touches and 10 land touches. Based on these data, compute a 95% Binomial CI for θ using (a) the unconservative method (using sample estimates to get a standard error for $\hat{\theta}$) and (b) the conservative method (using the maximum possible standard deviation for $\hat{\theta}$).