Assignment 6 — Due October 24, 2003

1. The purpose of this problem is to illustrate the meaning of the probability of type I error (the probability of rejecting the null hypothesis when the null hypothesis is true) through simulation. Imagine performing a hypothesis test (using the T-test) with a type I error rate (i.e. α) of 0.05. This means that out of every 100 tests performed when the null hypothesis is true, the null hypothesis will be rejected, *on average* 5 times.

   (a) In this problem you will simulate 200 samples of size 15 from a $N(62, 9^2)$ distribution. If you perform a test of $H_0 : \mu = 62$ versus the two-sided alternative for each sample at the level $\alpha = 0.10$, *on average* how many times will H_0 be rejected? What if $\alpha = 0.05$? What if $\alpha = 0.01$? (Note that this is a "theoretical" question that should be answered prior to performing the simulations.)

   (b) Perform the simulations. By comparing the p-values (calculated by R) with each level of α, determine how many times H_0 is actually rejected for your simulated data at $\alpha = 0.10$, $\alpha = 0.05$, and $\alpha = 0.01$. Compare these realized results with the *on average* results from part (a). Comment briefly on the comparison. [Appendix 6.9.2 has some R code to conduct such simulations.]

2. Reconsider the data on stem volume propagated from healthy buds (Assignment #5 Problem 2). Test $H_0 : \sigma^2 = 40000$ versus the two-sided alternative. *From this point on, when you are asked to test some hypothesis, you will be required to provide the null hypothesis, alternative hypothesis, test statistic, p-value, and an interpretation of the p-value, even though sometimes the question will not specifically indicate this requirement.*

3. Lengths of 9 randomly sampled oak seedlings from a given plantation are listed below:
   2.58 2.43 1.98 2.62 2.40 2.96 2.36 2.77 2.54 Assume that the population of oak seedling lengths follows a normal distribution; let μ be the mean length for oak seedlings from this plantation and let σ^2 be the variance.

   (a) Construct 90% and 95% confidence intervals for μ and interpret the confidence intervals.

   (b) Construct 90% and 95% confidence intervals for σ^2 and interpret the confidence intervals.

   (c) Suppose you obtained data on 36 seedlings. Suppose that the sample mean and variance are exactly the same as in (a). Construct a 95% confidence interval for the mean lengths of oak seedlings in that case. How does it compare to your answer for part (a)?

4. (a) Consider the experiment testing a new drug on sheep from Assignment #5 Problem 4. Let p be the "true" effective rate of the drug. Using the data from part (a), find a 90% CI for p.

   (b) Using the data from part (b) of Assignment #5 Problem 4, find a 90% CI for p. Compare with (a).

5. (a) Suppose we are sampling from a N(μ, 16) distribution. How large must n be so that a 90% CI for μ has length equal to 0.5?

   (b) Suppose you have a random sample from a N(μ,σ^2) distribution with σ^2 unknown. Let $n = 10$. Consider testing $H_0 : \mu = 22$ versus $H_A : \mu \neq 22$. Suppose you observe $\bar{x} = 20.7$ and $s^2 = 4.17$. Consider testing this hypothesis by using confidence intervals. Do you reject H_0 at α =0.10?, at α =0.05?, at α =0.01?

   (c) i. Using the data in part (b) of this problem, perform the T-test in the usual fashion. Use the `pt` command to find the exact p-value. Is this consistent with your results in part (b)?

   ii. Using the data in part (b) of this problem, use the `qt` command to find a 99.5% confidence interval for μ. [See the R commands `pt` and `qt` described in Appendix 6.9.1.]

Readings:

- Week 7: Course Notes Chapter 7