PROBLEM SET #1, ASTR 600

Statistical Overview

due Tuesday January 30, 2018

1. Binomial Probabilities [1 pt]

A coin that is evenly balanced between heads and tails is flipped 9 times. What are the odds of obtaining 3 heads?

2. *Bootstrap* [4 pts]

Continuing with the example in the handout, suppose you have a sample $\{x_i\} = \{1,2,4,18,19,19,40,42,43,50\}$. The mean of this sample is 23.80, and the uncertainty in the mean, calculated from

$$\sigma_{\overline{x}} = \frac{\sigma_{x_i}}{\sqrt{n}} \sim \frac{s}{\sqrt{n}} = 5.88 \tag{1}$$

Recording all of your steps from the beginning, load the data into a vector **x** in R, apply 10^4 bootstraps and determine the mean and the 95% confidence interval for the mean. How do these upper and lower limits compare to what you would get from a 95% confidence interval assuming a standard normal distribution with $\mu = 23.80$ and $\sigma = 5.88$? Print out a list of all the R commands that solved the problem, and plot a histogram of the bootstrapped means using a width of 2 for the histogram cells.

In this problem you will need to learn how to use the *function*, *boot*, and *boot.ci* commands in R. Recall I have a help page on R off my home page under the computer link, and there are extensive pages of help documentation avialable on-line.

3. *Hypothesis testing* [2 pts]

Note in this problem and in problem 4 you will need to find tables of the standard normal, and t-distributions, respectively. You may look these tables up on-line, but please calculate all the sums, differences, products, etc. by hand for now.

A real-estate broker who is anxious to sell a piece of property to a motel chain assures them that during the summer months on average, 4200 cars pass by the location each day. Being suspicious that this number might be too high, the motel does its own survey, and over 36 days obtains a mean of 4038 with a standard deviation of 512. What can they conclude at a level of confidence $\alpha = 0.05$? What about $\alpha = 0.01$?

4. *T-test for means* [2 pts]

Often one wants to compare two samples to see whether or not they were drawn from the same distribution. Comparing two means in this manner is done with the 't' test. A similar test for variances is the 'F' test.

There are several versions of the t-test, depending on what is known about the parent populations. Here we will consider two samples, with n_1 and n_2 points, the null hypothesis being that they were drawn from the same distribution, meaning that they have the same intrinsic μ and σ. The t-statistic for this case has $n_1 + n_2 - 2$ degrees of freedom, and is defined as

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $\overline{x_1}$ and $\overline{x_2}$ are the averages from samples 1 and 2, respectively, and s_1 and s_2 are the sample variances defined as

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \overline{x_1})^2$$

$$s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (x_i - \overline{x_2})^2.$$

A professor has a hypothesis that freshmen do not do as well in his class as sophomores do. The final scores in the class for the seven freshmen were 84, 83, 49, 60, 92, 70, and 78. The scores for the five sophomores were 95, 68, 79, 82, and 87. Are the higher sophomore scores statistically significant (pick your own significance level)?