Biostat II: Lab 5, Hypothesis testing using R
Date: 28 April 2008

This lab is based on data from a recent study:

> Hu G, Jousilahti P, Nissinen A, Bidel S, Antikainen R, Tuomilehto J. "Coffee consumption and the incidence of antihypertensive drug treatment in Finnish men and women." *Am J Clin Nutr.* 2007 Aug;86(2):457-64

Here is an excerpt from the introduction to this study:

> "Coffee is one of the most popularly consumed beverages in the world (1). During the past decade or so, research has attempted to clarify the health benefits or harms related to coffee drinking. The association between coffee consumption and blood pressure has been studied for ≥ 70 y (2). Two recent meta-analyses have shown a relation between coffee and caffeine intakes and an increase in blood pressure (3, 4). However, the data on the relation of coffee drinking and the risk of hypertension from prospective studies are still scarce (5, 6). According to international statistics, the Finns have the highest per capita coffee consumption in the world, at 11.4 kg/y (7). On the other hand, compared with many other populations, men and women in Finland have a high prevalence of hypertension (8). Therefore, research into the potential health effects of coffee in this population is of particular interest."

Our lab will use the following data provided in the article on the baseline relation between coffee consumption (cups/day) and systolic blood pressure (BP) in a sample of the Finnish population:

		Coffee Consumption	
		low (0-1)	high (≥ 8)
systolic BP (mm Hg)	n	2681	4179
	mean ($\bar{X}$)	133.9	135.4
	SE ($s/\sqrt{n}$)	0.3	0.3
	SD (s)		

1. Fill in the SD (standard deviation) values in the table above using the values for the Standard Error (SE) and the sample sizes (n) in the respective groups.

2. Construct a 95% confidence interval for the true mean systolic BP in the population of low coffee consumers in Finland.

3. According to US standards, the optimal systolic BP is 120 mmHg. Perform a two-sided hypothesis test at the $\alpha = 0.05$ significance level to determine whether the true mean systolic BP in the population of low coffee consumers in Finland is different from 120 mmHg.

   (a) State the null hypothesis and the alternate hypothesis.

   (b) Perform the hypothesis test using a critical region approach.

   (c) Perform the hypothesis test by calculating the p-value associated with the hypothesis test.

4. Compare the conclusion you have reached about the test using the 95% CI, the critical region and the p-value approaches to the hypothesis test.

5. When you have a very large sample size, you can detect small differences with strong statistical evidence. Repeat the hypothesis test (using your preferred approach) assuming the sample size in the sample of low coffee consumers was actually 10 and the population distribution of systolic BP is normally distributed. Use the same sample standard deviation that you calculated in the table above in your calculations. How do your reduced sample size results compare to your previous results?

6. Perform a hypothesis test (at the $\alpha = 0.05$ significance level) to test for a difference in the true mean systolic BP in the population of low coffee consumers in Finland versus the population of high coffee consumers in Finland. Use the observed sample sizes in your calculation. State your null and alternate hypotheses and what you decide about whether the variances are equal in the two groups.