# Week 9

Refer to the course map

## Objectives For This Week

Draw all we have been discussing over the quarter together, and apply it to real situations. Set up a real experiment on your own computer environment, and look at experiment results that have been made public looking for possible flaws in the design (a simple example is suggested). Our main goal is to become very aware of the many "potholes" that lie everywhere and can lead us to turn out results that are without base, even though we worked hard on collecting data to support our work.

## Reading Material

We have a discussion on how to translate the theory we have studied into a a real life practical study. The emphasis is on how it is not easy at all to set up a statistical study that lives up to the lofty ideal of the theory we have studied. As usual, we also have a PDF version.

## Assignments

Consider two real examples that you can perform yourself.

1. This is an experiment on your computer network (if you have absolutely no means to perform it, get in touch with me, and I will give you data from my own setup, but try not to: a college computer will also do). Choose a target (if you have more than one computer on your network you can use a different machine than yours, and if you are a one-computer network pick a target, like www.google.com, www.yahoo.com, www.amazon.com, www.comcast.net, you name it, from the world wide web). PING this target a reasonably large number of times (you will need either a graphic PING tool or, much simpler, bring up a command line - CMD in Windows, your favorite terminal emulation in Linux, make sure you have terminal access in OS X - and "ping -c*yy* www.xxx.xxx > experiment.txt". This should create a text file called experiment.txt (the name of the file is totally up to you) with a list of *yy* numbers (for example *yy* = 100 will give you a sample of size 100), representing how long it took for a request to travel to and from your target, www.xxx.xxx. Most likely, your PING program will be nice enough to tell you the average, and the standard deviation of your measurements. Since it is not unreasonable to assume that the fluctuations in your PING responses were due to small, completely random, mean zero disturbances, you may assume the experiment resulted in a Gaussian sample. Repeating this for two different targets (two computers in your network if you have that many, or two different obvious sites in the world wide web), you may estimate or test
   * What the mean response time was for each of your targets (with proper interval estimate)
   * Whether there is a significant difference between the two response times
2. You may appreciate that this study is providing you with quantitative information about your connection speed with the targets you chose. In fact, this is an experiment that gets pretty close to an ideal physical measurement, at least given what we know about the causes that may affect its outcomes.
3. On the other hand, as a very bad sampling example, you should choose an item which has many reviews on a Web site (a product, a restaurant, a travel destination, ...). Try for one that has a significant number of reviews - five opinions do not cut it. It is almost a given that the reviews will be scored with some number of stars. Consider the collection of reviews as a poll, and use your statistics tools to summarize the result of the poll. Next, discuss why this analysis is meaningless from a statistical point of view.

Please, turn in you report by **Monday Week 10**

## Quizzes

You can get credit for participation by taking the *quizzes from 76 to 80* from the quizzes on the [WAMAP](http://www.wamap.org/) site: this group is about sampling methods.