# Week 6

Refer to the course map

This week we look at the first part of the structure and methodology to handle *statistical tests*, one of the main (if not **the** main) areas of statistical activity.

## Objectives for this week

* Recognize the meaning of the "Null Hypothesis", and (perhaps more importantly) the "Alternate Hypothesis": identify them in examples.
* Understand the logic of tests: discuss what it means to "reject", or not to "reject" the Null Hypothesis, and apply this to simple tests
* **Optional "advanced" work**: Consider the consequence of applying a "standard" test based on the normal or *t* distribution) to a sample from a different distribution (e.g., exponential, "fat tailed", etc.)

## Reading Material

We start to learn about a very common tool in statistics: *tests*. This is possibly the most important topic, but it lends itself to misunderstandings. Hence, we want to go first of all over the logic and the meaning of this tool. This week, we look at the basics (look for the PDF version), what a test actually tests, so to speak, and how to interpret the result of a test.

And now, we can go into a longer, more practical, and detailed discussion of the most common tests (as suggested in the objectives and work for 211, the logic, if not the exact same tools, carry over to situations other than Gaussian distributions). As we browse over a number of situations (for example, testing for the mean of a random variable, and testing for the difference of two means from independent samples), you may notice that practically all cases are normally distributed, or assumed to be close to normal. This is not the only case you may encounter, but it is the most common (and, just to reiterate, it is sometimes assumed, more out of habit than out of solid grounds).

One last issue we should consider more closely is the choice between using a normal distribution or a Student distribution in interval estimation or in tests. This file provides some information about this question, and you might want to check out these spreadsheet files (depending on your preferred format) for a relevant example of a one-tailed test:

* Gnumeric format
* Open Document (\*.ods) format
* Excel XP (\*.xls) format

We will also have a *sample exam*, to prepare for the proctored Midterm next week (**not a graded assignment**). WHile the exam will, obviously, be different, the format and the type of question will be similar.

## Assignments

* Apply the theory to the simulated data you downloaded in week 1: test for different values of the mean (try different numbers, as, for example, the one that is listed as the one used by the simulation, a number fairly close, and a number a bit far off), check the *p* - values, and discuss what hypotheses you should reject and which you should not.
* Look up this PDF file containing a data set, and questions for this, the previous, and the next week. Answer the questions under "Week 6". **Do not type the data into a spreadsheet: that is both tedious, and terribly prone to errors**. Rather, use the summary data in the file to do the computations you need (you may want to use a calculator, but the cheapest one will do)

**Optional, "advanced" assignment**: test the "advanced" data sets, comparing how a "naive" application of normality-based tests performs in non normal cases. If possible, repeat for the data you simulated in Week 4.

**Please, turn in your work by Monday, Week 7**

### Quizzes

You can get credit for participation by taking the quizzes 51 to 65 on the [WAMAP](http://www.wamap.org) site: this group is about tests for normal (or approximated by normal) variables.