# Week 3

Refer to the course map

This week we begin our brief exploration of *Probability Theory*, the main mathematical tool used in statistical modelling. We will introduce the basic concepts, and start familiarizing with tools that we will be using all Quarter long.

## Objectives for this week

* Understand general probability arguments
  + Recognize meaningful vs. meaningless statements.
  + Perform simple probability calculations.
* Examples of distributions: recognize basic facts about Bernoulli, Binomial, Exponential, Normal, and distributions related to the normal (chi square, Student)
* Perform basic calculations on normal and Student distributions
  + Additional distributions (optional): Multinomial, Weibull, Cauchy

## Reading Material

* A short introduction to this week's material is in the probability file (HTLM/MathML) (there is also a PDF version).
* As a complement, you might want to read this recap of the notion of random variable (PDF only), its features, and some examples.
* Once you have checked the general framework, it's time for a more extensive discussion, mainly on *finite random variables* in the Online Stat book. While finite random variables are of minimal practical use in statistics, they are much easier to understand, and to play with. Please, do go through this material: it should help you gain a better intuition of this topic.
* As an example of how familiarity with finite random variables, and the very important subject of *conditional probability* (which applies to much more general situations, but is fairly easy to understand in this case) works in practice, you may want to check out these examples.
* We then go on to a more extended discussion of the Normal Distribution from the Online Stat Book
* **Optional, "advanced" work** consider the discussion about the so-called "truncated normal" distributions (XHTML version) (Here is a link to the PDF version). While we have no specific assignment on this issue, you might want to consider it when discussing data sets with only positive entries, for a critical comparison with other modeling options. As an optional assignment for week 5, you could compare the estimates for the exponential data set, with what you would get by presuming the data to be "truncated normal".

We will also need to refer to two tables, which will be with us for most of our course. The first is a table of the "Standard Normal Distribution", and the second is a table of the "Student 't' Distribution(s)". For example, you can find them (and others) at [the Statsoft site](http://www.statsoft.com/textbook/distribution-tables/?button=3). The text shows how you can use them for quick estimations. A public domain table for the standard normal distribution is enclosed with this course.

There are other tables that are common. One is the table for the "chi square" distribution (this is used in estimating variances of normal samples, and, surprisingly, in estimating means of exponential distributions), and another is the table for the "F distribution" ("F" stands for Fisher), used in the so-called *Analysis of Variance*, commonly shortened to ANOVA .

## Assignments

Since this is a theoretical chapter, we will work on a few "pencil and paper" problems

**Please, turn in your work by Monday, Week 4**

### Quizzes

You can get credit for participation by taking the

1. *quizzes from 16 to 25* from the quizzes on the [WAMAP](http://www.wamap.org) site: this group is about finite probabilities
2. *quizzes from 26 to 35* of quizzes on the [WAMAP](http://www.wamap.org) site: this group is about the normal and, occasionally, continuous distributions. Note that many problems use the normal distribution, motivated by the Central Limit Theorem.

Note that these will stay as our quizzes for week 4 as well