# Map of the Course

This is the schedule and file map for the course.

Week 1 | Week 2 | Week 3 | Week 4 | Week 5 | Week 6 | Week 7 | Week 8 | Week 9 | Week 10 | exam information

It is subject to change, if circumstances should suggest it. For the necessary links, please refer to the appropriate weekly file. In particular, refer to the weekly file for the relevant assignment(s).

## Project

As an experiment, you are urged to choose one or more data sets from publicly available sources (presumably, via the Internet), and analyze them with or tools, as we introduce them. It would be very good if you could work as teams of 2 - 3 members, even if an on line class does not make this very easy. In any case, the idea is for you to choose data relevant to a topic that is of interest to you.

As a generic help, we may divide you in groups (as provided by Blackboard), according to the general area of interest you choose. Even if you should not team up, exchanging info and thoughts with people addressing similar issues could be useful. As a start (further suggestions for additions and/or refinements are welcome), we could consider groups looking for data in

* Economy and Finance
* Demographic Data
* Environmental Sciences
* Health Sciences
* Sports

Please, let me know what you would like to work on, so we may evaluate together what kind of data analysis should be feasible.

## Simulated Data

We will apply the tools we are studying to a few data sets. Almost each week, as we acquire new tools, we will apply them to these sets, as detailed each week. To read more about the way these sets were generated, and to download them for your work, please go to the Simulated Data page.

Back to the top

## Introduction (Week 1)

* Sources - used in this course, and additional ones
* Tools needed for this course
* Short introduction of mathematical notation that we will be using constantly, including a formula that will turn useful many times during the course
* A Brief Introduction to Statistics
* A Longer Introduction, with examples, and discussion (this is taken, with minimal adaptations from the free course developed at Rice University, as listed in the Sources file - we will make extensive use of this course - see the sources file for more details).
* Don't forget to download the simulated data, in your preferred format.

Back to the top

## Descriptive Statistics (Week 2)

This week we work on the following material

* A short introduction to descriptive statistics. You will find links to additional files, some from the Rice course, with more extensive discussions of some points in there. This file does not use formulas, hence does not need a PDF version.
* Simulated data files on which to apply the tools we learn about this week, and in the future weeks, in various formats.

#### **Participation/Quizzes**

You can get credit for participation by taking the *first 15* quizzes on the [WAMAP](http://www.wamap.org) site. These quizzes are on descriptive statistics.

Back to the top

## Probability (Week 3 and Week 4)

This topic is broken down over two weeks:

1. **Basic probability**
2. **Multiple random variables, independence, limit theorems**

### Reading Material

The first part is covered in the following items:

* A short introduction to the essential information on probability models, that we will need in the upcoming chapters.
* A discussion of *finite probability distributions* from the Online Stat Book. These distributions are easy to understand and manipulate (and provide fun ways to confound you less probability-savvy friends), but have little or no direct use in the statistics we will study next. Nonetheless, you should work through this chapter, in order to gain some intuition about the many probabilistic statements that statistics produces - the easier context should help your intuition.
* A more extended discussion of the Normal Distribution from the Online Stat Book
* We will also need to refer to at least two tables, as discussed in the text: a table of the Standard Normal Distribution, and a table of the *t* distributions (optionally, we may use tables for the chi-square distribution, and, towards the end, we will be referring to the *F* distribution). Tables like these are available in any statistics book, as well as all over the Internet (for example, [the Statsoft site, we mentioned in the sources list](http://www.statsoft.com/textbook/distribution-tables/?button=3)). Because of copyright reasons, we could only include a typical standard normal distribution table in this course. Other tables are freely available on the World Wide Web, but are not necessarily copyright free, so that their inclusion here is not possible.

To complete our discussion, for the second part, look up the file on multiple random variables

The ideas introduced there can now be applied to the study of *sampling distributions* (what you have to work with when you take a sample for your statistical work)

#### **Participation/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.

Back to the top

## Inferential Statistics (Weeks 5, 6, 7, 8

Short general introduction

We will have three "pencil and paper" statistics assignments. They are set up so you can work them without a spreadsheet or other dedicated program: a simple calculator will be more than enough. In fact, **don't try to type in the data in a computer or a calculator**: that is tedious, and most likely will lead to mistakes. The PDF file (that's why it is in PDF format only) has a few summary data that allow you to compute all you need, with simple arithmetic.

### Estimation (Week 5)

In this introductory course, we will not discuss *point estimation*, and its properties: this is a topic that is of limited interest if used naively, and more of mathematical interest, if discussed in depth. If you are curious, you can take a look at the **optional** material in the On Line Stat Book.

* We apply probability to *quantitative* "interval" estimation
* We can then look at a more detailed "how-to" for a number of common situations.

**Math 211 students** should additionally work similarly on the "advanced" data.

#### **Participation/Quizzes**

You can get credit for participation by taking the quizzes 36 to 50 on the [WAMAP](http://www.wamap.org) site: this group is about interval estimation for normal (or approximated by normal) variables.

Back to the top

### Statistical Tests (Week 6 - 7)

This topic is spread over two weeks:

1. **Basic test theory: significance**
2. **Power of a test**

We will refer to the following pages:

1. In the first week (Week 6) we look at the basics of what a test actually tests, so to speak, and how to interpret the result of a test. We can then go into a longer, more practical, and detailed discussion of the most common tests. In particular, a major example is testing for the mean of a random variable.
2. **We will also have a *sample exam*, to prepare for the proctored Midterm next week.**
3. Next, we discuss the notion of "power" (Week 7) as it illuminates what the result of a test (especially, when the result is "not significant") means. In particular, we will discuss the notion, we will informally call "resolution power" of a test. Our companion course is less expansive on this than it is on other subjects, but we can look up a couple of examples there, to help us understand this notion better. We will also have a proctored Midterm, covering all our material up to Week 6.

#### **Participation/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.

Back to the top

### Regression (Week 8)

This week we conclude our exploration with another very common application: regression analysis, and study of correlation. This is a rich source of misunderstanding, so we want to be clear and focused. The most obvious source of misunderstanding is possible confusion between "correlation" and "causation": the fact that two things seem to come in pairs, does not mean at all that one is a *cause* of the other. A less obvious problem is under what conditions a relation like the one discussed here will hold, and what it would mean (and *not* mean) when they don't hold. We don't have much time to dwell on this issue, but we can have a first good look at the various issues.

We look at the theoretical underpinnings of *linear* regression, and the study of correlations. We also take a brief visit to some *nonlinear* cases, which, however,

* reduce, in a sense, to linear (if they don't, they are not easily studied)
* do not have as solid a theory behind them, but are rather "ad hoc" models

#### **Participation/Quizzes**

You can get credit for participation by taking the quizzes 66 to 75 on the [WAMAP](http://www.wamap.org) site: this group is about regression and correlations between variables.

Back to the top

## Applications (Week 9)

We take a general look at the practice of statistical experiments (our limited time and still limited toolbox prevent us from getting into too much detail). We will have a quick overview of some of the techniques that are used in practice, as well as some of the pitfalls that they carry. We will also check out a few examples.

#### **Participation/Quizzes**

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

Back to the top

## Review (Week 10)

This week is for review:

* Bring your project to completion
* Discuss with your instructor and your fellow students any point that is not clear
* Take a look at the sample Final exam in view of next week's test
* Inquire about where to go from here to improve on your statistical prowess

Back to the top

# Exam Schedule

We will have two proctored exams during this course:

1. A Midterm Exam, to be taken any time during Week 7 (5/9 - 5/13), covering all the material up to Week 6 ("Tests: significance"). This is a one-hour exam.
2. A Final Exam , to be taken any time during Week 11 (6/8 - 6/10), covering the rest of the course (with possible reference to the first part as well). This is a two-hour exam

You will be allowed to bring a regular (non scientific) calculator (the problem data will be provided in a form that will not require any pre-programmed statistics, only basic arithmetic), and **one letter-size, double sided sheet of notes.** Distribution tables will be provided with the test. The point of this limitation is to push you to set up a useful selection of material you may need in working on your exam. This preparation is, in itself, a great way to get ready for an exam. Also, organized material makes for a much more effective reference than a pile of unorganized notes, or a book, where you may waste a lot of time hunting around for the appropriate hint. In the week before each proctored exam, you will find a *sample exam* which will show the topics that will be covered in the corresponding test. That should also provide a good indication for your "cheat sheets".

You will be able to take each in the time frame indicated above **(but watch the Announcements for possible changes)**. Dates are flexible, to allow you to take it in a way that is compatible with your constraints.

* The simplest way to take the exams is on campus. If you choose to do so, arrange day and time with me. Your exam can take place at the Math Learning Center (no reservation needed: if you let me know this is your choice, I will leave a test for you, and you can take it any time during the week, during opening hours), at the Testing Center (reserve a time with them, and let me know, so I can give them your test), or with me (let me know your time options, and we'll schedule it so I can proctor you personally).
* If coming to campus is not convenient for you, you can arrange to take it in an appropriate location, with a proctor. This could be a library, with a librarian, a college near you, with an instructor, and so on. If you choose this method, **make sure to put me in contact with your proctor early enough, so we can arrange the test**. See below for further details. Please, look also at the official College rules for proctoring - you will find them in Blackboard, under *Course Information.*

Back to the Top

**Please, Note:** This is a fully on line class, so some might be disturbed by the requirement to take two proctored tests. Fact is, this requirement is much lower than the standard elsewhere, for on line Math classes. It is practically impossible to pass a reliable judgment on anybody's capabilities in Math, relying on the type of testing that would be available on line only. This is why, for instance, University of Washington on line Math classes require at least three fully proctored exams. Being well aware that for some, or many, of you, coming to Campus could be a serious obstacle, date, time, and location of these tests is open to arrangement, to make it as painless as possible.