

# MATH-UA 234: Mathematical Statistics

<b>Instructor</b>	Zsolt Pajor-Gyulai	<b>Lecture</b>	Tue-Thurs 12:30-1:45PM
<b>Email</b>	zsolt@cims.nyu.edu	<b>Classroom</b>	CIWW 517
<b>Office</b>	WWH 1105A	<b>Office hours</b>	TBA,
<b>TA:</b>	TBA	<b>TA's email</b>	TBA
<b>Recitation</b>	Fri 2:00-3:15, CIWW201		
<b>Course Page</b>	<a href="http://www.cims.nyu.edu/~zsolt/math-ua234.htm">www.cims.nyu.edu/~zsolt/math-ua234.htm</a>		

## 1 Short description

This course is intended as a thorough mathematical introduction to the theory of statistics.. Topics covered in this class will include: sampling theory, hypothesis testing, point (parameter) estimation, regression and linear least squares, tests of significance, likelihood methods, and Bayesian statistics. Topics in computational statistics will be covered using Python and Pandas.

## 2 Prerequisites

This class is intended to be taken after sufficiency in probability is obtained at the level of MATH-UA 233 Theory of Probability with a grade of C or better and/or the equivalent. However, students who received a C and those who are not confident enough in their knowledge are strongly encouraged to review the corresponding chapters of the textbooks.

## 3 Course goals

Statistical literacy is a fundamental for anybody who wishes to understand and work with data. The steadily growing demand for skilled data scientist in both industry and academia attracts more and more people to this fascinating field. Being well versed in the mathematical foundations is certainly a way to stick out in the competitive job market. While this is not a machine learning course (although, if time permits, we will cover a few simple classification methods), it will provide solid foundations for anyone who wishes to continue in that direction.

As this is a math class, the main focus will be on the theory underlying statistical concepts. Be advised that we will not simply learn recipes. You will be expected to understand the theory and to be able to write your own logical explanations or proofs.

Since this is a course about data, you will work with actual examples using Python Pandas with Jupyter Notebooks. While previous familiarity with Python is not expected, it is certainly helpful. We will provide references to tutorials combined with plenty of sample notebooks and you will be expected to pick things up on the fly using these aids. We recommend that you start preparing ahead of time, check back later for references..

## 4 Textbook

### Mandatory

- *All of Statistics* by Wasserman freely available through SpringerLink:

<http://www.springer.com/statistics/statistical+theory+and+methods/book/978-0-387-40272-7>

- *Probability and Statistics* 4th edition, by DeGroot and Schervish, available at the bookstore. Older editions may work provided you determine the mapping between sections.

## Recommended

- *Statistical Inference* by Casella and Berger, available at your favorite online bookstore

Each week, readings will be assigned from the textbooks, and posted on the website. It is strongly suggested that you read these before coming to class.

## 5 Recitations

You are each registered for a recitation. This is an opportunity to practice solving more problems, and to go over solutions to the homework problems. Computational help will also be given. It is not mandatory but you are strongly encouraged to attend.

## 6 Homework

Homework will be due every week and will be based on the previous week's lectures. The homework is meant to help you practice working with key concepts and to give you a chance to practice *neatly* and *clearly* writing up your solutions.

### 6.1 Written homework

Math is something that must be learned gradually and consistently, and if you fall behind, it is hard to catch up. Therefore solving homework problems is essential to keep up with the material. On the other hand, learning math requires solving many more problems than what will be assigned on the homework. You are strongly encouraged to seek out other problems; for example the textbook has lots of examples, both with and without solutions.

Written homework is due at every recitation. Late homework will not be accepted. (Homework is considered late if it is received after the end of recitation.) *There only ways to hand in written homework is to either bring it to class or typeset and submit it through NYUClasses.* In particular, homework left in a mailbox or scanned handwritten online submissions *will not be accepted*. If you cannot attend the class, you must find someone else to hand it in for you.

The lowest *two* homework grades will be dropped.

### 6.2 Computer homework

Statistics is the science of extracting useful information from data. Consequently, the theory is merely a shell without hands on experience applying the concepts to actual data. Therefore there will be frequent computational exercises using Python and Pandas through Jupyter Notebooks.

Understanding what the above words mean is not a prerequisite, although some previous knowledge is certainly useful. The TA will provide tutorials into how to use these and a good amount of sample notebooks will be provided. The emphasis will be on doing statistics and not on Python programming. The computer homework is to be submitted on NYUClasses in the form of completed Jupyter Notebooks. They are due the same day as the written homework by midnight.

The following books and tutorials are useful:

- *Python for Data Analysis* by McKinney
- <http://www.codecademy.com/tracks/python>
- <https://developers.google.com/edu/python/>
- <http://www.nyu.edu/lynda>
- <http://pandas.pydata.org/>

## 7 Collaboration

You are strongly encouraged to work with others – for example to solve homework problems or to prepare for exams. However, you must make sure to *write up your own* homework solutions, and to *acknowledge* those you worked with. Submitting two very similar sets of solutions is a violation of academic integrity and will be disciplined by the university. The best way to ensure this doesn't happen is to write your solutions separately.

## 8 Communication

The online program Piazza will be used to communicate information about the course, and to communicate between students in the course. You will all have access to the course page, and you will be able to post messages either privately to myself, or publicly so that everyone can see.

If you have questions about the homework, or a general question about the lecture material, please **post this publicly**. This way everyone in the class will benefit from the answer as it is likely that several people have a similar question. Mistakes/typos on the homework problems will also be posted here.

If you see a question for which you know or have an idea about the answer, please do **post to respond** – another goal of the Piazza is to generate discussion, as there may be several different (yet equally valuable) interpretations of the course material. I will monitor the posts frequently and indicate whether they are correct. Even if they aren't, it's ok! It's much better to have an incorrect idea and find out soon, than to keep thinking the incorrect idea and make mistakes on the test, or further along in life, etc.

The NYU Classes website will probably be used to post homework assignments, and may also be used to communicate this information. If you hand in homework electronically (encouraged for code used to produce homework), you can do that here.

## 9 Participation

A small percentage of your grade is listed as “participation.” This will be graded by your participation on the Piazza platform. You will get full marks if you make at least 3 well-thought-out posts: either questions, or responses to questions. (They do not have to be correct, but they do have to show that thought was put into them, and are not just there to get grades!)

## 10 Exams

There will be an in-class midterm, and a final exam set by the registrar.

Midterm	Thursday, April 5
Final Exam	TBA

## 11 Grades

			93	A
			90	A-
			87	B+
			83	B
			80	B-
			75	C+
			65	C
			50	D
Final Exam	35%	Letter grades will be assigned as follows.		
Midterm	30%			
Written homework	20%			
Computer homework	10%			
Participation	5%			

This grading scheme may be changed, but only to act in your favour – the grades can only be higher, not lower.