北京大学
汇丰商学院
Peking University HSBC Business School

Course Code
Mathematical Foundations and Data Skills I
Module 1, 2021-2022

Course Information

Instructor: Prof. Thomas Sargent, Dr. Chase Coleman, Dr. Spencer Lyon
Office: PHBS Building, Room 672

Phone: 86-755-2603-1223
Email: sargent@phbs.pku.edu.cn, cc7768@gmail.com, spencerlyon2@gmail.com
Office Hour: By Appointment (email professors)

Classes:
Lectures: 8:30AM-10:20AM, Tuesday & Friday
Venue: PHBS Building, Room TBA

Course Website:
http://www.tomsargent.com

1. Course Description

1.1 Context

Course overview: This course is the first in a sequence of four Computational Economics & Finance classes. It is designed to teach two sets of skills. On one hand, it introduces core concepts underlying computational social science, including optimization, probability, and statistical modeling. We learn how to apply these concepts to a variety of social science questions. On the other hand, this course teaches the foundational skills necessary to do modern data analytics using the Python programming language.

We assume that students have previously worked with Python. We will add to existing Python skills and teach the core scientific and data-specific libraries (numpy, scipy, matplotlib, and pandas). We will use these skills to analyze a variety of datasets and answer research and business questions.

Prerequisites:

Students should have taken at least a course in calculus and a course in linear algebra. Prior courses in probability and statistics will also be helpful.

Highly motivated graduate students lacking some prerequisites can still succeed in this course but will have to put in more time and effort. Such students should meet with us well in advance of the course so that we can recommend readings that will help close the gaps in their preparations. Also, a willingness to consult Wikepedia for math and statistics concepts will help all students.

Students should have prior experience using the Python programming language before starting this course. Throughout this course we will leverage programming skills such as control flow constructs (`if/else`, `for/while`), defining custom functions (`def`), and finding help on existing functions (`?` in Jupyter environments and `help` elsewhere).

Although a course in probability or statistics is not a prerequisite, students will find some knowledge of these topics to be helpful.

Students are required to have access to a personal computer (laptop) that can be brought with them to lectures. Computers will be used by students in each class session and those without a laptop will not get the necessary in-class practice in order to master the concepts we study.

1.2 Textbooks and Reading Materials

- **Linear Algebra Done Right, 3rd edition** by Sheldon Axler
- **All of Statistics** by Larry Wasserman
- **Convex Optimization** by Stephen Boyd
- **Probability and Information: An Integrated Approach** by David Applebaum
- **QuantEcon** Various lectures from QuantEcon available at https://quantecon.org
- **Software Carpentry** We will use materials from the Software Carpentry organization available at https://software-carpentry.org/
- **Statistical Rethinking** by Richard McElreath
- **QuantEcon Datascience** Lectures from the QuantEcon datascience sequence at https://datascience.quantecon.org
- **Python Data Science Handbook** by Jake Vanderplas https://jakevdp.github.io/PythonDataScienceHandbook/
- **Python for Data Analysis, 2nd Edition** by Wes McKinney

2. Learning Outcomes

2.1 Intended Learning Outcomes

Learning Goals	Objectives	Assessment (YES with details or NO)
1. Our graduates will be effective communicators.	1.1. Our students will produce quality business and research-oriented documents.	NO
	1.2. Students are able to professionally present their ideas and also logically explain and defend their argument.	NO
2. Our graduates will be skilled in team work and leadership.	2.1. Students will be able to lead and participate in group for projects, discussion, and presentation.	YES
	2.2. Students will be able to apply leadership theories and related skills.	NO
3. Our graduates will be trained in ethics.	3.1. In a case setting, students will use appropriate techniques to analyze business problems and identify the ethical aspects, provide a solution and defend it.	NO

	3.2. Our students will practice ethics in the duration of the program.	NO
4. Our graduates will have a global perspective.	4.1. Students will have an international exposure.	YES
5. Our graduates will be skilled in problem-solving and critical thinking.	5.1. Our students will have a good understanding of fundamental theories in their fields.	YES
	5.2. Our students will be prepared to face problems in various business settings and find solutions.	YES
	5.3. Our students will demonstrate competency in critical thinking.	YES

2.2 Course specific objectives

2.2.1 This course teaches powerful mathematical tools and topics that will

- Demonstrate how computational tools can be used to formulate and to solve problems in Economics and Finance that may not have otherwise had good answers
- Allow students themselves to answer questions in Economics and Finance using computational tools
- Give students a concise "language" (i.e., mathematics) that will allow them to learn other new tools that are not covered in this course

2.2.2 The key objective in this course is for students to use Python to do meaningful data analysis using social science data sets. Success in this course can be described as the student's ability to do the following:

- Read, write, and understand basic programs written in the Python programming language
- Import, clean, combine, and summarize datasets from a variety of sources
- Construct informative visualizations of raw data and model results
- Implement basic data engineering best practices such as optimizing organization and structure of datasets, using effective storage formats for a given task, and automating repetitive extract-transform-load (ETL) processes

2.3 Assessment/Grading Details

Assessments

A mixture of homework assignments, in-class quizzes, exams, and a final project will be used to evaluate students' learning achievements.

Homework: At the beginning of the course, homework will be assigned each week. Later, there will be less frequent homework assignments because then you will be spending substantial time on your class project. Your two lowest homework grades will be dropped.

In-class participation: Students are expected to attend lectures and be contributing members of our class community.

Exams: There will be 1 take-home exam.

Project: The class project will be structured to give you practice creatively applying the tools that you have learned to a real-world problem.

Except for quizzes and exams, we highly encourage students to work together. We have found that groups of 3-4 seem to work best. We believe that collaborative work is the best way to learn the type of material that we cover. We advise students not to rely on others to do work that you do not understand.

Grading Policy

Assignments described above will be the main inputs to the grade for the course. Assignments will be weighted evenly within groups and overall according to the following decision rule:

- Homework assignments: 30%
- In-class participation: 5%
- Test: 25%
- Project: 40%

This weighting reflects our opinion that the most important skills to be acquired in this class are communicated by one's ability successfully to apply the tools that you learn to an interesting question in Economics and Finance.

2.4 Academic Honesty and Plagiarism

It is important for a student's effort and credit to be recognized through class assessment. Credits earned for a student work due to efforts done by others are clearly unfair. Deliberate dishonesty is considered academic misconducts, which include plagiarism; cheating on assignments or examinations; engaging in unauthorized collaboration on academic work; taking, acquiring, or using test materials without faculty permission; submitting false or incomplete records of academic achievement; acting alone or in cooperation with another to falsify records or to obtain dishonestly grades, honors, awards, or professional endorsement; or altering, forging, or misusing a University academic record; or fabricating or falsifying of data, research procedures, or data analysis.

All assessments are subject to academic misconduct check. Misconduct check may include reproducing the assessment, providing a copy to another member of faculty, and/or communicate a copy of this assignment to the PHBS Discipline Committee. A suspected plagiarized document/assignment submitted to a plagiarism checking service may be kept in its database for future reference purpose.

Where violation is suspected, penalties will be implemented. The penalties for academic misconduct may include: deduction of honour points, a mark of zero on the assessment, a fail grade for the whole course, and reference of the matter to the Peking University Registrar.

For more information of plagiarism, please refer to *PHBS Student Handbook*.

3. Topics, Teaching and Assessment Schedule

Schedule and weekly learning goals

The scheduled contents are *intended to be covered over the entire Fall semester, in two modules*. Contents are presented in two parts organized by subjects: Math and Data Skills. In actual classes, however, topics from these two will be taught in a "*mixed and match*" manner, to allow students to combine theory with practice. The schedule is **tentative and subject to change**.

PART I: Mathematical Foundations

1. The calculus
Sources and tools:

- `sympy package: https://www.sympy.org/en/index.html`
- `https://python.quantecon.org/complex_and_trig.html`
- `https://python.quantecon.org/complex_and_trig.html`

Topics to be mastered:

- Derivatives
- Integrals
- Complex valued functions and their derivatives
- Fourier transforms and the convolution theorem
- The fast Fourier transform and its inverse
- First and second order conditions for extremization
- Euler's method for solving ordinary differential equations
- Runge-Kutta numerical intergration
- Gaussian quadrature integration
- Introduction to genetic algorithms and other procedures for extremizing over rugged landscapes

2. Applications of Linear algebra
Sources and tools:

- `https://python.quantecon.org/linear_algebra.html`
- `scipy.linalg`

Topics to be mastered:

- Vectors and matrices
- Elementwise operations, matrix multiplication, determinants
- Linear independence and dependence
- Linear transformations
- Matrix inverses
- Determinants and volumes
- Eigenvalues and eigenvectors
- Courant-Fischer minimax theorem and some of its applications in machine learning
- Inner product spaces
- Hilbert spaces
- Orthogonal projection
- Gram-Schmidt orthogonalization
- The Cholesky decomposition

3. Convex optimization, I
Sources and tools:

- Chapters 1-3 of Stephen Boyd's text.
- `CVXpy: https://www.cvxpy.org`

- `scipy.optimize.linprog`

Topics to be mastered:

- Linear programming problems
- Primal and dual problems
- Relationship of a “welfare theorem” and prices
- Pathologies
- Simplex algorithm
- Diet problem
- Least absolute deviations regressions

4. Convex optimization, II
Sources and tools:

- Chapters 4-5 of Stephen Boyd's text.
- `CVXpy`
- `scipy.optimize`

Topics to be mastered:

- Quadratic programming
- Nonlinear programming
- Nonlinear regression

5. Probability, I
Sources and tools:

- **Probability and Information**
- **Statistical Rethinking**
- **All of Statistics**

Topics to be mastered:

- Laws of probability
- Continuous and discrete probability distributions
- Conditional and marginal probabilities
- Stochastic processes
- Stationarity
- Ergodicity
- Concepts of statistical convergence
- Strong laws of large numbers
- Central limit theorems
- A statistical model as a probability distribution over a sequence indexed by a vector of parameters

6. Probability, II
Sources and tools:

- `https://python.quantecon.org/finite_markov.html`

- https://python.quantecon.org/ar1_processes.html
- https://python.quantecon.org/stationary_densities.html

Topics to be mastered:

- The direct problem given parameters
  - Computation
  - Random simulations of paths
- The inverse problem given data
  - Likelihood function
  - Prior over parameters
  - Maximum likelihood and Bayesian estimation
- Time series models
- Discrete state Markov chains
- Continuous state Markov processes
- First order autoregressions
- Moving average processes
- Mixed autoregressive, moving average processes
- Nonlinear models
- Illustrations of stationarity and ergodicity

7. Statistics, I
Sources and tools:

- **Statistical Rethinking**
- https://python.quantecon.org/heavy_tails.html
- https://python.quantecon.org/arma.html

Topics to be mastered:

- Linear regression
- Population and sample regressions
- Nonlinear population and sample regressions
- Mathematical expectations versus linear least squares projections
- Gram-Schmidt process as recursive projection
- Link of Gram-Schmidt to efficient computation and interpretation

8. Statistics, II
Sources and tools:

- **Statistical Rethinking**
- https://python.quantecon.org/mle.html
- https://python.quantecon.org/exchangeable.html
- quantecon Jupyter notebook on estimating a first-order a.r.

Topics to be mastered:

- Bayesian parameter estimation
- Bayesian model evaluation
- Frequentist inference

- Frequentist model evaluation
- Learning versus hypothesis testing
- Exchangeability and Definetti theorem
- David Kreps's famous “chapter 11” story about Totrep also known as “Totally Rational Economic Person”

9. Statistics, III

Sources and tools:

- **Statistical Rethinking**
- `https://python.quantecon.org/wald_friedman.html`
- `https://python.quantecon.org/odu.html`

Topics to be mastered:

- Applications of Bayesian updating to decision making
- Job search with learning
- The classic problem that stumped Milton Friedman
- How Abraham Wald solved Milton Friedman's problem

10. Statistics, IV

Sources and tools:

- **Statistical Rethinking**
- **All of Statistics**
- **Linear Models of Dynamic Stochastic Economies**

  Topics to be mastered:

  - Linear rational expectations models
  - Simulation
  - Inference
  - Connections between simulation and inference (two sides of one coin)

11. Dynamic Equilibrium Models, I

  Sources and tools:

  - `https://python.quantecon.org/cattle_cycles.html`
  - `https://python.quantecon.org/rosen_schooling_model.html`

  Topics to be mastered:

  - Cattle Cycles
  - The Rosen schooling model

12. Case studies, I

  Sources and tools:

- https://python.quantecon.org/black_litterman.html
- Cedric Villani on optimal transport https://www.youtube.com/watch?v=zo46TEp6FB8l (start at minute 28)

Topics to be mastered:

- Portfolio theory and its challenges
- The Black Litterman model and its relationship to Bayesian statistics and robust control theory
- Monge-Kantorovich transport problem and applications in social sciences

13. Case studies, II

Sources and tools:

- https://www.youtube.com/watch?v=DEHqIxX1Kq4

Topics to be mastered:

- Using Bayes to infer who wrote the Federalist papers?
- Probabilistic graphic models with the Python package pgmpy

14. Case studies, III

Sources and tools:

- https://link.springer.com/chapter/10.1007/978-94-011-2410-2_19

Topics to be mastered:

- From political science: the Law of Cubic Proportions
- Student presentations of class projects

15. Wrapping up and looking forward

Sources and tools:

- Lectures notes on duality between filtering and control

Topics to be mastered:

- Useful math and going forward
  - Why "forecasting" and "decision" are related mathematically
  - How this can be exploited
- Student presentations of class projects

PART II: Data Skills

1. Introduction to Pandas
Sources and tools:

- Class notes
- `pandas package: https://pandas.pydata.org/`
- `https://python.quantecon.org/pandas.html`
- `https://datascience.quantecon.org/pandas/intro.html`
- `pandas.DataFrame`
- `pandas.Series`
- Chapter 5 of Python for Data Analysis

Topics to be mastered:

- Pandas datatypes: DataFrame and Series
- Basic operations with DataFrames: summary statistics, aggregations, transformations, data selection
- Sorting and ranking
- Value counts
- Function application and mapping
- Duplicate labels
- Basic visualization using the `plot` method

2. Organizing Data with Pandas, I
Sources and tools:

- Class notes
- `pandas.Index`
- `https://datascience.quantecon.org/pandas/the_index.html`
- `https://datascience.quantecon.org/pandas/storage_formats.html`
- Chapter 6 of Python for Data Analysis

Topics to be mastered:

- Understanding the `Index` in pandas
- Storage formats
- Reindexing
- Stacking and melting
- Hierarchical indexing

3 Organizing Data with Pandas, II
Sources and tools:

- Class notes
- `https://datascience.quantecon.org/pandas/data_clean.html`
- `https://datascience.quantecon.org/pandas/reshape.html`
- Chapter 7 and 8 of Python for Data Analysis

Topics to be mastered:

- Cleaning, reshaping, and merging datasets

- `merge`, `join` and `combine`
- Stacking and melting
- Handling missing data
- Discretization and binning
- Random sampling
- String manipulation

4 Grouped Operations with Pandas, I
Sources and tools:

- Class notes
- `pandas.groupby`
- `https://pandas.pydata.org/pandas-docs/stable/user_guide/groupby.html#`
- `https://datascience.quantecon.org/pandas/groupby.html`
- Chapter 10 of Python for Data Analysis

Topics to be mastered:

- `groupby` method with built-in methods
- Groupby mechanics
- Custom grouped functions

5 Grouped Operations with Pandas, II
Sources and tools:

- Class notes
- `pandas.groupby`
- `https://pandas.pydata.org/pandas-docs/stable/user_guide/groupby.html#`
- `https://datascience.quantecon.org/pandas/groupby.html`
- Chapter 10 of Python for Data Analysis

Topics to be mastered:

- Aggregation with multiple function application
- General split-apply-combine
- `transform`
- `apply`

6. Grouped Operations with Pandas, III
Sources and tools:

- Class notes
- `https://pandas.pydata.org/pandas-docs/stable/user_guide/groupby.html#`
- `https://datascience.quantecon.org/pandas/groupby.html`
- Chapter 12 of Python for Data Analysis

Topics to be mastered:

- Pivot tables and cross-tabulation
- Method chaining with `pipe`
- Categorical data

7. Time Series with Pandas, I
Sources and tools:

- Class notes
- `https://pandas.pydata.org/pandas-docs/stable/user_guide/timeseries.html`
- `https://datascience.quantecon.org/pandas/timeseries.html`
- Chapter 11 of Python for Data Analysis

Topics to be mastered:

- Rolling-window operations
- Resampling frequency of observations
- Doing arithmetic with dates, date ranges, periods, and TimeDeltas

8. Time Series with Pandas, II
Sources and tools:

- Class notes
- `https://pandas.pydata.org/pandas-docs/stable/user_guide/timeseries.html`
- `https://datascience.quantecon.org/pandas/timeseries.html`
- Chapter 11 of Python for Data Analysis

Topics to be mastered:

- Upsampling and interpolation
- Downsampling
- Handling time zones

9. Data Visualization, I
Sources and tools:

- Class notes
- `https://datascience.quantecon.org/pandas/matplotlib.html`
- `https://seaborn.pydata.org/`
- `https://plot.ly/python/`
- `https://altair-viz.github.io/`
- Chapter 9 of Python for Data Analysis
- `https://datascience.quantecon.org/applications/maps.html`

Topics to be mastered:

- Intermediate `matplotlib`
- Statistical visualization with `seaborn`

- Widgets

10. Data Visualization, II
Sources and tools:

- Class notes
- Chapter 9 of Python for Data Analysis
- `https://datascience.quantecon.org/applications/maps.html`

Topics to be mastered:

- Interactive web-based visualizations, and dashboards using `plotly` and `altair` — As an example of what could be done, see Mike Waugh's webpage

11. Data Harvesting
Sources and tools:

- Class notes
- `https://scrapy.org/`
- `https://camelot-py.readthedocs.io/en/master/`
- `https://www.crummy.com/software/BeautifulSoup/bs4/doc/`

Topics to be mastered:

- Integrating with Web APIs
- Scraping data from websites without an api (`scrapy`)
- Extracting data from PDFs (`camelot`)

12. Data Engineering
Sources and tools:

- Class notes
- `https://airflow.apache.org/`
- `https://www.sqlalchemy.org/`

Topics to be mastered:

- Basic introduction to databases (using SQLite through `sqlalchemy`)
- Automation and data pipelines using Apache Airflow
- We will illustrate these tools by creating an automatically updating database on one of a few potential topics. Our choice of topic will depend on class interest.

13. Case studies, I
Sources and tools:

- `http://www.tomsargent.com/research/ReadMe_Pub.pdf`
- `https://datascience.quantecon.org/applications/`
- Chapter 14 of Python for Data Analysis
- `https://datascience.quantecon.org/applications/recidivism.html`

Topics to be mastered:

- Combine the tools learned in this class to generate automatically updated databases and visualizations, covering topics such as
  - Inequality data
  - U.S. bond data and term structure of interest rates; see Hall, Payne, Sargent bond dataset

14. Case studies, II
Sources and tools:

- `https://datascience.quantecon.org/applications/`
- Chapter 14 of Python for Data Analysis

Topics to be mastered:

- Examples from \textit{The Great Reversal} by Thomas Phillipon

15. Case studies, III
Topics to be mastered:

- Student presentations of class projects

4. Miscellaneous

Professional Behaviour

Attend class. They say "eighty percent of success is just showing up." We have found that those who show up perform systematically better.

Arrive to class on time and stay until the end of class. Chronically arriving late or leaving class early is unprofessional and disruptive to the rest of the class.

We understand that the electronic recording of notes will be important for class and so computers will be allowed in class. Please refrain from using computers for anything but activities related to the class. Phones are prohibited as they are rarely useful for anything in the course.