

Data Analysis in Environmental Applications (SEE5211/SEE8212)

Sem B 2023/24 (15 Jan – 20 Apr)

Prof. Yiming QIN

Assistant Professor

School of Energy and Environment

City University of Hong Kong

Lecture Hours: Tue 19:00 – 21:50; Venue: YEUNG LT-13



L00: Course Information

Prof. Yiming QIN
Assistant Professor
School of Energy and Environment
City University of Hong Kong

Who we are

Instructor:
Prof. Yiming QIN

Email:
ymqin3@cityu.edu.hk

Office: YEUNG- B5428

TAs:

HAO Xinya
Xyhao5-c@my.cityu.edu.hk

ZHENG Siwen
siwezheng4-c@my.cityu.edu.hk

Course Description

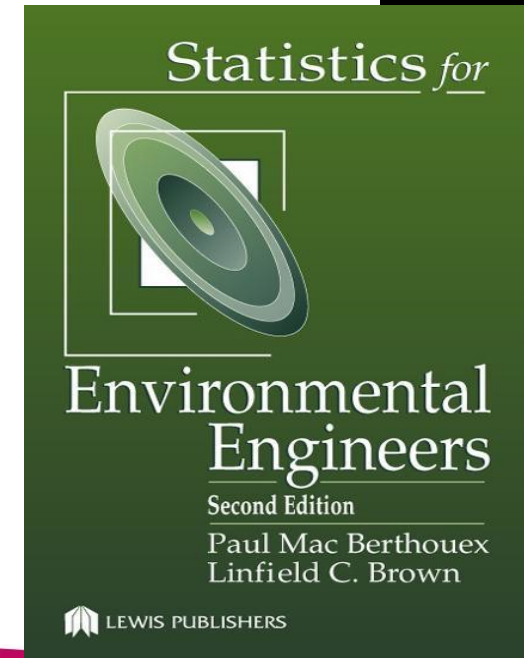
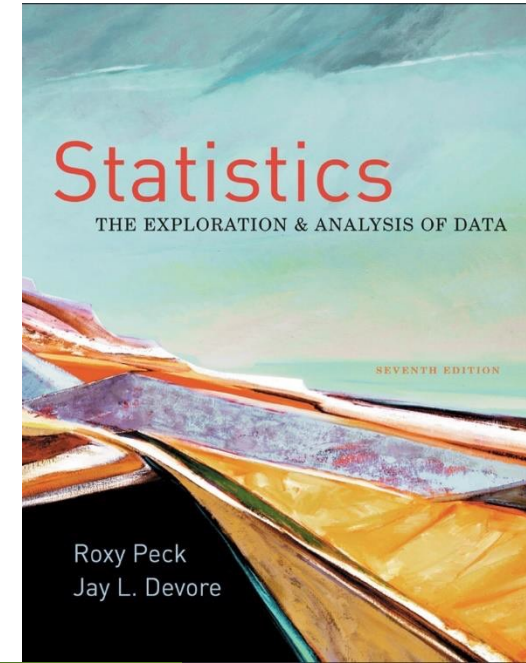
- The course will provide the **knowledge** of using **statistical methods** in **energy and environmental science**.
- Analysis methods, such as **probability**, **random variable** (discrete & continuous), **parameter estimation**, **confidence interval** and **hypothesis testing** involving one and two populations, **simple linear regression**, **analysis of variance**, and **goodness-of-fit test**, are very helpful to understand the physical processes occurring in the environment, and to work on climate prediction.
- You are required to use the knowledge learnt from this course to analyse the data with **computational tools**, such as Python.

What will You Learn from this Course?

- 1) Understand the concepts of **basic statistical methods**;
- 2) Use **probability, random variable (discrete & continuous), parameter estimation, confidence interval and hypothesis testing involving one and two populations, simple linear regression, analysis of variance, and goodness-of-fit test** to describe energy and environmental datasets and **solve energy and environmental problems** creatively;
- 3) Use correlation method to analyse energy and environmental data and discover the **linkage between the data results and energy and environmental problems**;
- 4) Apply the **statistical methods** creatively to **explain the problems in energy and environmental science**.

Textbooks

- Statistics: The exploration and analysis of data, 7th Edition, 2012. Roxy Peck Jay L DeVore.
 - Statistics for Environmental Engineers, 2nd Edition, 2002. Linfield C. Brown, Paul Mac Berthouex.
- A set of lecture notes will be provided on the web phase by phase



SEE5211/SEE8212 Class Schedule (Sem B 2023/24)

Week 1 (Jan 16)	<ul style="list-style-type: none"> L00 Course Information L01 Introduction to Statistics L02 Descriptive Statistics Python Programming Tutorial 	Week 8 (Mar 12)	<ul style="list-style-type: none"> Mid-term Discussion L08 Hypothesis Testing – Two Population
Week 2 (Jan 23)	<ul style="list-style-type: none"> L03 Probability Python Programming Tutorial 	Week 9 (Mar 19)	<ul style="list-style-type: none"> L09 Simple Linear Regression L10 Analysis of Variance In-class Quiz 4: L07-L08
Week 3 (Jan 30)	<ul style="list-style-type: none"> L04 Random Variable – Discrete Python Programming Tutorial 	Week 10 (Mar 26)	<ul style="list-style-type: none"> P01 Project L11 Goodness-of-fit Test In-class Quiz 5: L09-L10
Week 4 (Feb 6)	<ul style="list-style-type: none"> L05 Random Variable - Continuous In-class Quiz 1: L01-L03 	Week 11 (Apr 2)	<ul style="list-style-type: none"> L12 Time Series Analysis Final Exam Review In-class Quiz 6: L11
Feb 13	<ul style="list-style-type: none"> Lunar New Year Break (Feb 9-15) 	Week 12 (Apr 9)	<ul style="list-style-type: none"> L13 Principal Component Analysis P02 Python Programming Tutorial for the Project
Week 5 (Feb 20)	<ul style="list-style-type: none"> L06 Estimation and Confidence Interval Mid Term Review In-class Quiz 2: L04-L05 	Week 13 (Apr 16)	<ul style="list-style-type: none"> P03 Python Programming Tutorial for the Project
Week 6 (Feb 27)	<ul style="list-style-type: none"> L07 Hypothesis Testing – One Population In-class Quiz 3: L06 	Week 14	<ul style="list-style-type: none"> Apr 22-27 Student Revision Period
Week 7 (Mar 5)	<ul style="list-style-type: none"> Mid Term Exam (2.5 hrs); Venue: LT- 13 Up to L05 	TBC	<ul style="list-style-type: none"> Final Exam. TBD

Grading

- In-Class Quiz (30 mins, the best 4 out of 6): 20%
Open book, no electronic devices
- Mid-term Exam (2.5 hrs, Week 7, March 5): 20%
Close-book, only one page of specified cheat sheet and calculators are allowed
- Group Project (Week 14): 20%
A written report with at least 15 pages
- Final Exam (TBD): 40%
Close-book, only one page of specified cheat sheet and calculators are allowed

Guidelines for Students

- **Being in class on time.** Being attentive, polite, respectful, and no disruptive in class.
- **No academic dishonesty,** e.g., plagiarism, cheating, etc., is allowed for all assessment tasks.
- **Work hard!**

Project

Scope of work:

- Form a group of 4-5 members. Each group needs to pick up **one environmental pollutant or energy related problem** to investigate. You are required to analyze the data by **statistical methods and computational tools, such as Python, learned** from the lecture. Relevant information sources, i.e. data, should be available online (e.g. **EPD, Journal Paper, other Hong Kong government websites**, etc.). For each group, send your topic and your group members to the TA and instructor **on Week 10**.
- Report: **At least a 15-page report** needs to be written.
- Please send the report to the TA and Instructor via Canvas in PDF or Word format.

Project Objectives:

- To investigate environmental or energy-related problems by statistical methods;
- To practice the use of computational tools and learn how to analyze real data by statistical methods;
- ***To foster how to solve the problems creatively and innovatively;***
- ***To develop critical thinking and teamwork capability.***

Project

Information Collection (some references for you):

- Environmental Data available from HKEPD website:

https://www.epd.gov.hk/epd/english/resources_pub/envir_info/envir_info.html

- Energy Data available from Census and Statistics department website:

<https://www.censtatd.gov.hk/hkstat/sub/sp90.jsp?productCode=B1100002>

- Climatological Information available from HKO website:

<https://www.hko.gov.hk/en/cis/climat.htm>

- The interest of the project is not only limited to Hong Kong, for example data in US:

<https://www.epa.gov/air-trends>

Project

Suggested format of the report and assessment criteria:

- Abstract (5%)
- Introduction (10%)
- Methodology (15%)
- Results & Discussion (20%)
- Recommendations (20%)
- Conclusions (10%)
- References (5%)
- Appendix (e.g. programming code in Jupyter notebook) (5%)
- English Writing (10%)

Outline

<i>Chapter 0</i>	<i>Course Introduction</i>
	Introduction to the course and the required contribution and learning outcome
<i>Chapter 1</i>	<i>Introduction to Statistics</i>
	Concepts in statistics, common terminology and methods
<i>Chapter 2</i>	<i>Descriptive Statistics</i>
	Types of variables, data presentation (graphical and numerical)
<i>Chapter 3</i>	<i>Probability</i>
	Basic set theory, Venn diagram, properties of probability and conditional probability
<i>Chapter 4</i>	<i>Random Variable – Discrete</i>
	Discrete probability distribution, population mean and variance, Binomial and Poisson distribution
<i>Chapter 5</i>	<i>Random Variable – Continuous</i>
	Continuous probability distribution, population mean and variance, Normal distribution

Related energy and environmental issues will be discussed in each chapter

Outline

<i>Chapter 6</i>	<i>Estimation and Confidence Interval</i>
	Estimator and estimate, accuracy and precision, maximum likelihood estimator, random interval, confidence interval, <i>t</i> -distribution
<i>Chapter 7</i>	<i>Hypothesis Testing - One Population</i>
	Null hypothesis, alternative hypothesis, type I error, type II error, probability-value approach, classical approach, binomial probability of success
<i>Chapter 8</i>	<i>Hypothesis Testing - Two Population</i>
	Independent versus dependent samples, compare two populations
<i>Chapter 9</i>	<i>Simple Linear Regression</i>
	Paired data, regression coefficients, least squares approach, fitted regression line, residual sum of squares, confidence intervals and hypothesis testing for regression coefficients
<i>Chapter 10</i>	<i>Analysis of Variance (ANOVA)</i>
	Main objectives of ANOVA models, one-way ANOVA analysis
<i>Chapter 11</i>	<i>Goodness-of-fit Test</i>
	Principles and usage of Goodness-of-fit test in discrete & continuous cases
<i>Chapter 12</i>	<i>Time Series analysis</i>
<i>Chapter 13</i>	<i>Principal Component Analysis</i>

Computational Tool for Data Analysis

Energy and Environmental Problems in Python

Why Python and jupyter notebook?

- An open source
- Libraries for data analysis tasks
- Easy to learn and use

How to install it?



Free software, open standards, and web services for interactive computing across all programming languages

202402SEE5211_SEE8212 > Modules

63 Student View

2023/24 Semester B

Home
Announcements
Assignments
Discussions
Grades
Pages
Files
Syllabus
Quizzes
Modules
Collaborations

Collapse All View Progress Publish All + Module

Week 1

JupyterInstallationGuide_Mac.pdf	✓
JupyterInstallationGuide_Win.pdf	✓

Essential Python Libraries

- NumPy

Multidimensional array object, linear algebra operations, Fourier transform, and random number generation.

- Pandas

High-level data structures and functions designed to achieve fast, easy, and expressive data manipulation.

- Matplotlib

Plots and two-dimensional visualizations.

Essential Python Libraries

- SciPy

Standard continuous and discrete probability distributions, various statistical tests, and more descriptive statistics.

- Statsmodels

A statistical analysis package which contains regression models, analysis of variance, time series analysis, non-parametric methods.

Jupyter Notebook

Installation

Shell

```
pip install jupyter
```

Starting the Jupyter Notebook Server

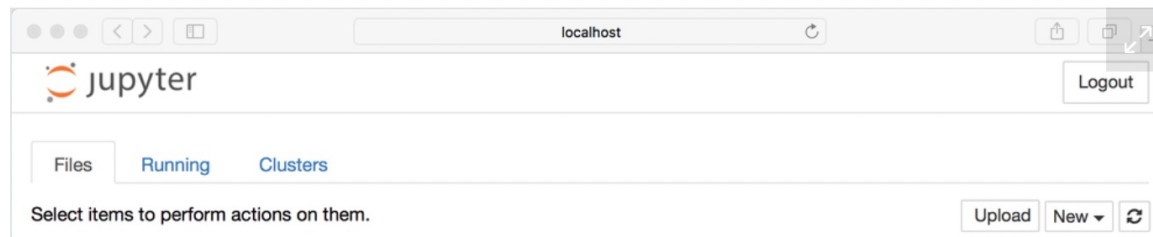
To get started, all you need to do is open up your terminal application.

Shell

```
jupyter notebook
```

This will start up Jupyter and your default browser should start (or open a new tab) to the following URL: <http://localhost:8888/tree>

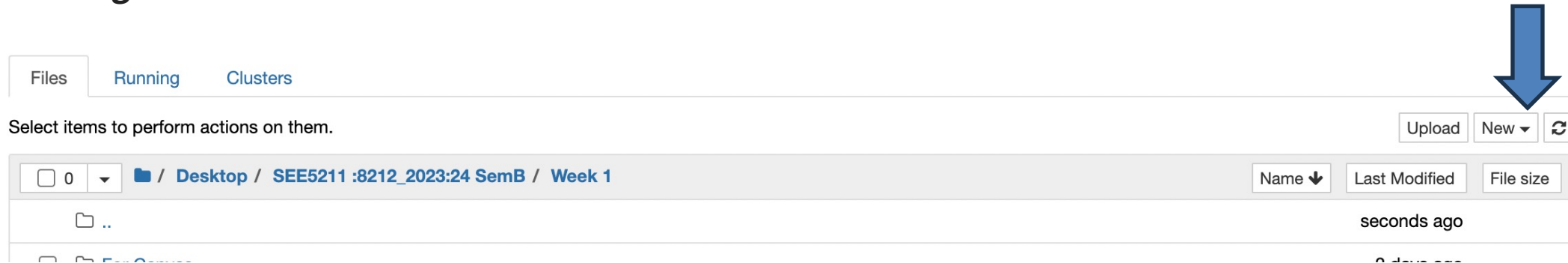
Your browser should now look something like this:



Jupyter Notebook

Creating a Notebook

All you need to do is click on the *New* button (upper right).



Your web page should now look like this:

