Lingnan University
Department of Philosophy

Course Title	:	Data Analysis for the Humanities
Course Code	:	PAI3001
Recommended Study Term	:	Term 2, Third Year
No. of Credits/Term	:	3
Mode of Tuition	:	Sectional approach
Class Contact Hours	:	3 hours per week
Category	:	Required course
Prerequisite(s)	:	PAI3003 Programming for the Humanities
Co-requisite(s)	:	NIL
Exclusion(s)	:	NIL
Exemption Requirement(s)	:	NIL

Brief Course Description

This course is intended for humanities students who are interested in learning about the fundamentals of programming, data analysis, and artificial intelligence. It provides an introduction to various types of mathematical functions, different kinds of data, and methods of data analysis, as well as the theoretical underpinnings of statistical analysis. The course places particular emphasis on the philosophy of probability and statistics, covering key topics such as interpretations of probability, the debate between Bayesianism and Frequentism, and the role of non-epistemic values in hypothesis testing and model selection. Through this course, students will gain a comprehensive understanding of how these concepts apply to real-world scenarios, enhancing their analytical skills.

Aims

The first half of this course will cover computer programming and concepts required for understanding the foundations of data analysis. The second half of the course will cover applications of data analysis with real-life applications. Along the way, introductory concepts in artificial intelligence will be taught.

Learning Outcomes (LOs)

Upon completion of this course, students will be able to:

1. Explain the key concepts in artificial intelligence and statistical learning theory.
2. Illustrate the foundations of data analysis and machine learning.

3. Apply the knowledge of algorithm design in the programming language Python.
4. Analyze real life issues via quantitative reasoning.

Indicative Content

- Mathematical Functions (Domain, Range, Types of Functions)
- Philosophy of Probability & Statistics
- Statistical Learning: Hypothesis Space, Optimization Algorithm, Loss Functions, etc.
- Evaluation strategy (e.g. Confusion Matrix, Accuracy rate/ error rate)
- Introduction to Pandas Data Reader in Python
- Data visualization
- Application: Display trends in novel readers
- Binary Classification/Linear Regression

Teaching Method

The course will be delivered by lectures and in-class assignments. Students are expected to do readings from the core textbooks ahead of time and to bring their laptops to class to perform in-class programming exercises. Students will use some free cloud-based Python interpreter with minimal installation requirements to run their code.

Measurement of Learning Outcomes

Learning Outcomes	Assessment Methods		
	In-class assignments	Midterm examination	Final examination
LO1: Explain the key concepts in artificial intelligence and statistical learning theory.	✓	✓	✓
LO2: Illustrate the foundations of data analysis and machine learning.	✓	✓	✓
LO3: Apply the knowledge of algorithm	✓	✓	✓

design in the programming language Python.			
LO4: Analyze real life issues via quantitative reasoning.		✓	✓

Assessment

There are three assessment components:

In-class assignments : 20%

Midterm examination : 30%

Final examination : 50%

Required/Essential Readings

Suggested (selection from):

Heinold, Brian. 2012. *A Practical Introduction to Python Programming*. URL: https://www.brianheinold.net/python/A_Practical_Introduction_to_Python_Programming_Heinold.pdf. Accessed August 24, 2024.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York: springer.

Karsdorp, F., Kestemont, M., & Riddell, A. (2021). *Humanities data analysis: case studies with Python*. Princeton University Press.

McKinney, W. (2022). *Python for data analysis*. O'Reilly Media. Inc.

Rowbottom, Darrell. 2015. *Probability*. Wiley.

Russell, S. J., & Norvig, P. (2022). *Artificial Intelligence: A Modern Approach*, 4th, Global ed.

And some online resources, which be made available on Moodle.

Supplementary/Optional Readings

Suggested:

Desai, J., Watson, D., Wang, V., Taddeo, M., & Floridi, L. (2022). The epistemological foundations of data science: a critical review. *Synthese, 200*(6), 469.

And selection from:

Suárez, M. (2020). *Philosophy of probability and statistical modelling*. Cambridge University Press.

Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.

And some online resources, which be made available on Moodle.

Important Notes

1. Students are expected to spend a total of 9 hours (i.e. 3 hours of class contact and 6 hours of personal study) per week to achieve the course learning outcomes.
2. Students shall be aware of the University regulations about dishonest practice in coursework, tests and examinations, and the possible consequences as stipulated in the Regulations Governing University Examinations and Course Work. In particular, plagiarism, being a kind of dishonest practice, is "the presentation of another person's work without proper acknowledgement of the source, including exact phrases, or summarised ideas, or even footnotes/citations, whether protected by copyright or not, as the student's own work". Students are required to strictly follow university regulations governing academic integrity and honesty.
3. Students are required to submit writing assignment(s) using Turnitin.
4. To enhance students' understanding of plagiarism, a mini-course "Online Tutorial on Plagiarism Awareness" is available on https://pla.ln.edu.hk/.
5. Students must adhere to the University's guidelines and practices when using Generative Artificial Intelligence (AI) tools. The official documents "Guidelines for Using GAI Tools at Lingnan University" and "Best Practices for Ethical and Responsible Use of GAI Tools in Course Assessments" can be found here: https://www.ln.edu.hk/tlc/generative-artificial-intelligence/gai-guidelines-best-practices.

Data Analysis for the Humanities

In-Class Assignments Rubric (20%)

Assessment Categories	Excellent (A)	Good (B)	Pass (C)	Fail (F)
Quality	All assignments were completed. Evidence of consistent good	Most assignments were completed. Evidence of	Several assignments were completed with good effort.	Very few assignments were completed with good effort.

	effort.	good effort in those completed.		

Data Analysis for the Humanities

Midterm Examination Rubric (30%)

Assessment Categories	Excellent (A)	Good (B)	Pass (C)	Fail (F)
Problem	A very good answer. The student finds the correct answer in the correct way or very nearly the correct way. Any errors are minor and does not reflect any inability to correctly use the relevant method.	An incorrect but good answer using the correct method. The correct method has been almost entirely applied in the correct way.	The correct method has been applied. The application is mostly or at least partly correct.	Either: (1) There is no answer. or (2) The wrong method was applied or the right method was applied wholly inadequately.

Data Analysis for the Humanities

Final Examination Rubric (50%)

Assessment Categories	Excellent (A)	Good (B)	Pass (C)	Fail (F)
Problem	A very good answer. The student finds the correct answer in the correct way or very nearly the correct way. Any errors are minor and does not reflect any inability to correctly use the relevant method.	An incorrect but good answer using the correct method. The correct method has been almost entirely applied in the correct way.	The correct method has been applied. The application is mostly or at least partly correct.	Either: (1) There is no answer. or (2) The wrong method was applied or the right method was applied wholly inadequately.