

**The Chinese University of Hong Kong**  
**Department of Linguistics and Modern Languages**  
**First Term, 2025-26**

**1–3. Course code, English title and Chinese title**

<p><b>Course Code:</b> LING3403  <b>Title in English:</b> Quantitative Methods for Linguistics  <b>Title in Chinese:</b> 語言學的定量方法</p>
---

**4. Course description**

This course serves as an introduction to the fundamental statistical concepts and analytical techniques frequently employed in the fields of linguistics and applied linguistics. It provides students with a solid understanding of both descriptive and inferential statistics, equipping them with the necessary tools to analyze and interpret data effectively. Through a combination of engaging lectures and hands-on tutorials, students will not only acquire a foundational understanding of statistical principles but also gain practical experience in conducting data analyses. This course is particularly beneficial for students who aspire to engage in quantitative research for their theses, as it will provide them with the necessary skills and knowledge to undertake robust investigations. No prior knowledge in statistics is assumed, making this course accessible to students from various academic backgrounds. By the end of the course, students will have developed proficiency in visualizing, analyzing, and summarizing quantitative data using the R programming language. This practical application of statistical techniques will enable students to confidently explore and derive insights from real-world linguistic data sets.

這門課程旨在介紹語言學和應用語言學領域常用的基本統計概念和分析技巧。它使學生對描述性統計和推論統計都有扎實的理解，並為他們提供分析和解釋數據的必要工具。通過生動有趣的講座和實踐性的教學，學生不僅將建立統計原理的基礎理解，還將獲得實際操作數據分析的經驗。這門課對於有意在論文中從事量化研究的學生特別有益，因為它將為他們提供進行堅實調查所需的技能和知識。本課程不要求學生具備統計學的先備知識，因此對來自不同學術背景的學生來說都是可接受的。在課程結束時，學生將能夠熟練使用 R 程序語言來視覺化、分析和總結定量數據。這種統計技巧的實際應用將使學生能夠自信地探索和從真實語言數據集中獲得洞察。

**5. Learning outcomes**

- Students will learn how to interpret statistical results in experimental reports on language and related areas.
- Students will have a good understanding of experimental designs and their implications for statistical analyses.
- Students will develop critical quantitative thinking in their empirical research.
- Students will have conceptual understanding of common statistical techniques in language research and the actual skills to apply them to collected data.
- Students will be able to use R to conduct simple data analyses.

**6. Course syllabus**

<b>Topic</b>	<b>Contents/fundamental concepts</b>
Research design in experimental linguistics	Correlational design, cause-effect design, cross-sectional design, within/between-subject design
R as a statistical platform	R for handling data, R for statistical analyses, R for making graphs
Descriptive statistics	Mean, mode, median, z-score, standard deviation
Correlation	The relation between two datasets
t-test	How to use parametric test to compare two means
Non-parametric tests	How to use non-parametric tests to compare means
ANOVA	How to compare multiple groups

Regression	How to model a set of observations using independent predictors
------------	---

### 7. Course components

Lecture (hr) in /out class		Interactive tutorial (hr) in /out class		Lab (hr) in /out class		Discussion of case (hr) in /out class		Field-trip (hr) in /out class		Projects (hr) in /out class		Web-based teaching (hr) in /out class		Other (hr) in /out class	
2		0.75					0.5				0.5		1		
M	NA	M	NA	NA	NA	NA	M	NA	NA	NA	M	NA	O	NA	NA

M: Mandatory activity in the course; O: Optional activity; NA: Not applicable

### 8. Assessment type, percentage and rubrics

Task nature	Description	Weight
Participation	Participation in lectures/tutorials, discussion, and presentation	10%
Assignment	Weekly assignment	25%
Group project	Report of an experiment	25%
Exam	In-class data analyses and reports	40%
		Total: 100%

### 9. Required and recommended readings

#### *Introductory textbooks*

Levshina, N. (2015). *How to do linguistics with R. Data Exploration and Statistical Analysis*. Amsterdam-Philadelphia.

#### *More advanced textbooks*

Field, A., Miles, J., Field, Z. (2012). *Discovering statistics using R*. Sage Publications. [Reference book]

De Vries, A., & Meys, J. (2015). *R for Dummies*. John Wiley & Sons.

Crawley, M. J. (2012). *The R book*. John Wiley & Sons. [Reference book for R]

Kerns, G. J. (2010). *Introduction to probability and statistics using R*. Downloadable from <https://cran.r-project.org/web/packages/IPSUR/vignettes/IPSUR.pdf>. [Textbook for R and statistics]

### 10. Feedback for evaluation

Students are encouraged to give feedback or comments on course contents and teaching materials throughout the course. Students can contact either the lecturer or tutor directly. See contact details below. Mandatory term-end evaluation for teacher's reflection will also be conducted.

### 11. Course schedule

Week	Date	Lecture
1	4-Sept	Introduction and R
2	11-Sept	Descriptive statistics
3	18-Sept	T-tests 1
4	25-Sept	T-tests 2
5	2-Oct	ANOVA 1
6	9-Oct	ANOVA 2
7	16-Oct	Non-parametric tests
8	23-Oct	Binomial and chi-square test
9	30-Oct	Correlation
10	6-Nov	No class (make up class to be arranged)
11	13-Nov	Linear regression 1

12	20-Nov	Linear regression 2
13	27-Nov	<b>Exam</b>

## 12. Contact details for teacher(s) or TA(s)

<b>Professor/Lecturer/Instructor:</b>	
Name:	Prof. Zhenguang Cai
Office Location:	Leung Kau Kui Building G5 Thursday 14:00 – 15:00, or by appointment
Telephone:	3943 7909
Email:	<a href="mailto:zhenguangcai@cuhk.edu.hk">zhenguangcai@cuhk.edu.hk</a>
Teaching Venue:	Lee Shau Kee Building 308 (Thursday 9:30-11:15)
Website:	<a href="http://www.cuhk.edu.hk/lin/people/">http://www.cuhk.edu.hk/lin/people/</a>
Other information:	

<b>Teaching Assistant/Tutor:</b>	
Name:	Xiaohui RAO, Yicheng LI
Office Location:	
Telephone:	
Email:	Xiaohui RAO: <a href="mailto:1155197282@link.cuhk.edu.hk">1155197282@link.cuhk.edu.hk</a> Yicheng LI: <a href="mailto:1155248207@link.cuhk.edu.hk">1155248207@link.cuhk.edu.hk</a>
Teaching Venue:	
Website:	
Other information:	

## 13. Details of course website

All teaching materials will be uploaded (either before or after lecture) onto Blackboard ( <a href="https://blackboard.cuhk.edu.hk">https://blackboard.cuhk.edu.hk</a> ) or via email.
--

## 14. Academic honesty and plagiarism

Attention is drawn to University policy and regulations on honesty in academic work, and to the disciplinary guidelines and procedures applicable to breaches of such policy and regulations. Details may be found at <http://www.cuhk.edu.hk/policy/academichonesty/>.

With each assignment, students will be required to submit a signed **declaration** that they are aware of these policies, regulations, guidelines and procedures.

- In the case of group projects, all members of the group should be asked to sign the declaration, each of whom is responsible and liable to disciplinary actions, irrespective of whether he/she has signed the declaration and whether he/she has contributed, directly or indirectly, to the problematic contents.
- For assignments in the form of a computer-generated document that is principally text-based and submitted via VeriGuide, the statement, in the form of a receipt, will be issued by the system upon students' uploading of the soft copy of the assignment.
- Students are fully aware that their work may be investigated by AI content detection software to determine originality.
- Students are fully aware of the AI approach(es) adopted in the course. In the case where some AI tools are allowed, students have made proper acknowledgment and citations as suggested by the course teacher.

Assignments without a properly signed declaration will not be graded by teachers.

Only the final version of the assignment should be submitted via VeriGuide.

The submission of a piece of work, or a part of a piece of work, for more than one purpose (e.g. to satisfy the requirements in two different courses) without declaration to this effect shall be regarded as having committed undeclared multiple submissions. It is common and acceptable to reuse a turn of phrase or a sentence or two from one's own work; but wholesale reuse is problematic. In any case, agreement from the course teacher(s) concerned should be obtained prior to the submission of the piece of work.

The copyright of the teaching materials, including lecture notes, assignments and examination questions, etc., produced by staff members/ teachers of The Chinese University of Hong Kong (CUHK) belongs to CUHK. Students may download the teaching materials produced by the staff members/ teachers from the Learning Management Systems, e.g. Blackboard, adopted by CUHK for their own educational use, but shall not distribute/ share/ copy the materials to a third-party without seeking prior permission from the staff members/ teachers concerned.

### 15. Use of Generative Artificial Intelligence (AI) Tools in Teaching, Learning and Assessment

Approach 2 – Use only with prior permission

AI is allowed for grammar and spelling checks but not for content generation in learning activities or assessments unless the teacher explicitly allows this in a learning activity. The use of AI, if at all, will contribute to no more than 30% of the final grade.

**16. Use of AI tools for grading and provision of feedback on student work** (include this Section where applicable):

(If Approaches 2 to 4 are adopted)

No AI tools supported will be used to grade and provide feedback on the following assignment(s) and assessment task(s)

### 17. Guideline on sharing lecture recordings

Attention is drawn to the following:

- The copyright of any lecture recordings shared in the course, whether they are produced by teachers, students, or peer note-takers, belongs to the University.
- Students should not share these recordings with others without obtaining prior written consent from the teacher(s).

### 18. Grade Descriptors

Grade	Descriptors
A	<ul style="list-style-type: none"><li>• Excellent understanding of common statistical tests and always apply them appropriately in answering a wide range of research questions</li><li>• Excellent ability to visualize, analyze, and summarize many different types of data using R with very high presentation standards</li><li>• Excellent ability to interpret and evaluate quantitative findings in linguistics with an extremely high degree of accuracy and aptness</li></ul>
A-	<ul style="list-style-type: none"><li>• Good understanding of common statistical tests and almost always apply them appropriately in answering a wide range of research questions</li><li>• Good ability to visualize, analyze, and summarize many different types of data using R with high presentation standards</li><li>• Good ability to interpret and evaluate quantitative findings in linguistics with a high degree of accuracy and aptness</li></ul>

B	<ul style="list-style-type: none"> <li>• Fair understanding of common statistical tests and often apply them appropriately in answering some research questions</li> <li>• Fair ability to visualize, analyze, and summarize some types of data using R with satisfactory presentation standards</li> <li>• Fair ability to interpret and evaluate quantitative findings in linguistics with a satisfactory degree of accuracy and aptness</li> </ul>
C	<ul style="list-style-type: none"> <li>• Poor understanding of common statistical tests and only sometimes apply them appropriately in answering some research questions</li> <li>• Poor ability to visualize, analyze, and summarize some types of data using R with poor presentation standards</li> <li>• Poor ability to interpret and evaluate quantitative findings in linguistics with only some degree of accuracy and aptness</li> </ul>
D	<ul style="list-style-type: none"> <li>• Very poor understanding of common statistical tests and rarely apply them appropriately in answering research questions</li> <li>• Very poor ability to visualize, analyze, and summarize limited types of data using R with poor presentation standards</li> <li>• Very poor ability to interpret and evaluate quantitative findings in linguistics with a barely satisfactory degree of accuracy and aptness</li> </ul>
F	<ul style="list-style-type: none"> <li>• Almost no understanding of common statistical tests and almost never apply them appropriately in answering research questions</li> <li>• Almost no ability to visualize, analyze, and summarize very limited types of data using R with extremely poor presentation standards</li> <li>• Almost no ability to interpret and evaluate quantitative findings in linguistics inaccurately and inaptly</li> </ul>

--- End ---