1. Course title and description

Course title:

STAT5107 Discrete Data Analytics
2019-2020 Semester 1

Instructor: Professor LIN, Yuanyuan (LSB 113)
Time and Venue (Lecture): Thursday 7:00pm-9:30pm, Wu Ho Man Yuen Bldg 304
Teaching Assistant: Mr. SHEN, Guohao (PhD candidate)

Description:

This course provides a practically oriented treatment of modern methods for the analysis of categorical data. Topics include analysis of two-way contingency tables, logistic regression, log-linear model, generalized linear model, and some modern topics related to classification and regression tree method etc.

2. Learning outcomes

After finishing the course, students should be able to understand topics related to the analysis of categorical data: discrete response data and frequency data, such as two-way contingency tables, logistic regression, generalized linear model, log-linear model and some modern topics in classification.

3. Course Content

Topic	**Contents/ concepts**
Distribution and inference for categorical data	Categorical response data, distribution for categorical data, Statistical inference for Binomial, Multinomial and Poisson parameters
Contingency tables	Two-way contingency tables, sampling models, test of independence, test of homogeneity, odds ratios, measures of Association for ordinal variables, Testing of independence for ordinal variables, introduction to three-way contingency tables
Generalized linear models and Logistic regression	Generalized linear models for binary data, Generalized linear models for count data, Generalized linear mixed models, Simple logistic regression, multiple logistic regression, Logit models for qualitative predictors, model selection
Loglinear models for contingency tables	Loglinear model for two-way tables, loglinear-logit connection
Multicategory logit models and some advanced topics in classification	Multicategory logit models, cumulative logit models, advanced topics on classification and clustering: linear discriminant analysis, tree-structured prediction, etc.

4. Learning activities

Lecture	
In Class	Out Class
3 hours	3 hours

5. Assessment Scheme

Type	Description	Weight
Assignments	n sets of assignments will be given, 3 <n<6 (*Remark:* Students are highly recommended to hand in their assignments to the TA during lecture in **HARD COPY**. Soft copy is acceptable subject to the instructor's approval. Penalty will be imposed on late homework. No grade if the homework is submitted after the answer has been posted.)	20%
Mid-term Examination	**October 24, 2019**	30%
Final Examination	**December 5, 2019**	50%

6. Feedback for evaluation

- Students' comments and feedback are valuable for improving the course, and students are welcome to provide comments and feedback. A mid-term course evaluation will be conducted in **Week 8 in lecture**. Term end course evaluation will be conducted by the department.

7. Course schedule

Class/ week	Topic
Week 1-2	Distribution and inference for categorical data
Weeks 2-4	Statistical inference for Contingency tables
Weeks 5-7	Generalized linear models and Logistic regression
Week 8	Midterm exam in lecture
Weeks 9-10	Loglinear models for contingency tables
Weeks 11-13	Multicategory logit models and some advanced topics in classification
	Remark: The above course schedule may be subjective to minor changes depending upon the teaching progress.

8. Teachers' or TA's contact details

Professor/ Lecturer/ Instructor	
Name:	Professor LIN, Yuanyuan
Office:	LSB 113
Telephone:	39437924
Email:	**ylin@sta.cuhk.edu.hk**

Teaching Assistant/ Tutor	
Name:	Mr. SHEN Guohao

Office Location:	LSB G26
Telephone:	39438528
Email:	ghshen@link.cuhk.edu.hk

9. A facility for posting course announcements

Course material (including Course Outline, Lecture Notes, Assignments) and other information will be posted on Blackboard:

http://blackboard.cuhk.edu.hk

10. Learning resources

- Lecture notes can be downloaded from Blackboard System

Resources for students

References:

Alan Agresti (2013). Categorical Data Analysis, 3th edition, Wiley series in Probability and Statistics.

Academic honesty and plagiarism (optional)

The Chinese University of Hong Kong places very high importance on honesty in academic work submitted by students, and adopts a policy of *zero tolerance* on cheating and plagiarism. Any related offence will lead to disciplinary action including termination of studies at the University. Relevant information can be allocated via: http://www.cuhk.edu.hk/policy/academichonesty/.

Every assignment handed in should be accompanied by a signed declaration. The form can be downloaded from: http://www.cuhk.edu.hk/policy/academichonesty/declaration_en.doc