# GLOBAL SUMMER PROGRAMME 2027

## DS601 STATISTICAL LEARNING

Instructor Name : Dr. Shew Fan Liu
Title : Senior Lecturer of Statistics
Email : sfliu@smu.edu.sg
Office : School of Economics, 5044

### COURSE DESCRIPTION

Statistical Learning is a course designed to equip students with the fundamental concepts and methodologies of statistics for data modelling. The course integrates the use of the *Python* programming language into most topics, providing students with hands-on experience in applying statistical techniques to real-world scenarios commonly encountered in business environments.
Basic understanding of statistics and familiarity with programming concepts is recommended. No prior experience with *Python* is required, as introductory *Python* programming concepts will be covered in the course.

### LEARNING OBJECTIVES

By the end of the course students will have had a firm grasp of the key concepts and issues in basic statistical modelling and regression.

### PRE-REQUISITES

Introductory Statistics

### ASSESSMENT METHODS

| | |
|---|---|
| Class Participation | 20% |
| Problem sets and group project | 40% |
| Final exam | 40% |
| **Total** | **100%** |

## INSTRUCTIONAL METHODS AND EXPECTATIONS

**Class Participation:**
Active participation in discussions during classes is highly encouraged.
Each class will include Python labs coupled with learning activities. Students are expected to attempt these activities and submit their work before the following class. Time will be given after the lab sessions to attempt the activities. Submissions will contribute to the class participation grade.

**Assignments:**
There will be three problem sets assigned over the course of the term. These problem sets are to be completed individually.
In addition, there will be a group project that requires you to build a statistical learning model for a large dataset.

**Final Examination:**
The final examination will be a 3-hour computer-based exam, conducted in an open-book format.

## CONSULTATIONS

By appointment

## RECOMMENDED TEXT AND READINGS

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J., (2023). *An Introduction to Statistical Learning with Applications in Python* (1st ed.). Springer.

## UNIVERSITY POLICIES

**Academic Integrity**
All acts of academic dishonesty (including, but not limited to, plagiarism, cheating, fabrication, facilitation of acts of academic dishonesty by others, unauthorized possession of exam questions, or tampering with the academic work of other students) are serious offences.
All work (whether oral or written) submitted for purposes of assessment must be the student's own work. Penalties for violation of the policy range from zero marks for the component assessment to expulsion, depending on the nature of the offense.
When in doubt, students should consult the instructors of the course. Details on the SMU Code of Academic Integrity may be accessed at https://smu.sharepoint.com/sites/oasis/SitePages/DOS-WKLSWC/UCSC.aspx.

**Copyright Notice**
Please note that all course materials are meant for personal use only, namely, for the purposes of teaching, studying and research. You are strictly not permitted to make copies of or print additional copies or distribute such copies of the course materials or any parts thereof, for commercial gain or exchange.
For the full copyright notice, please visit https://researchguides.smu.edu.sg/copyright.

**Accessibility**
SMU strives to make learning experiences accessible for all. If you anticipate or experience physical or academic barriers due to disability, please let me know immediately. You are also welcome to contact the university's accessibility services team if you have questions or concerns about academic provisions: accessibility@smu.edu.sg. Please be aware that the accessible tables in our seminar room should remain available for students who require them.

| LESSON PLAN | | |
|---|---|---|
| **Lesson** | **Reading** | **Topics** |
| 1 | | Introduction to Python for data analysis |
| 2 | | Review of Statistics and Probability<br>○ Sample vs. population<br>○ Statistics vs. parameters<br>○ Descriptive statistics vs. inferential statistics<br>○ Discrete vs. continuous random variables |
| 3 | | Review of inferential statistics<br>○ Confidence interval estimation<br>○ Hypothesis testing |
| 4 | | Probability modelling<br>○ Distribution modelling<br>○ Goodness-of-fit test |
| 5 | Ch 3 | Review Simple Linear Regression<br>○ Least squares estimation<br>○ Goodness-of-fit and related inferences<br>○ Gauss-Markov conditions<br>○ Residual analysis<br>○ Confidence intervals and prediction intervals |
| 6 | Ch 3 | Multiple Linear Regression<br>○ Least squares estimation<br>○ Model interpretation<br>○ Qualitative predictors |
| 7 | Ch 3 | Extensions of the linear model<br>○ Polynomial regression<br>○ Transformations<br>○ Interactive Effects |
| 8 | | Guest lecture by invited speaker/Industrial visit (TBC) |
| 9 | Ch 4 | Logistic Regression<br>○ Estimation, prediction, interpretation<br>○ Multiple logistic regression |
| 10 | Ch5 | Resampling Methods<br>○ Bias-variance trade-off<br>○ Training MSE vs. test MSE<br>○ Cross-validation |

| | | |
|---|---|---|
| | | ○ The bootstrap |
| 11 | Ch 6 | Linear Model Selection<br>○ Subset selection vs. shrinkage<br>○ Best subset selection<br>○ Stepwise selection |
| 12 | | Exam |