




GLOBAL WINTER PROGRAMME 2026

Statistical Learning		
<div>Instructor Name : Dr. Shew Fan Liu</div> <div>Title : Senior Lecturer of Statistics</div> <div>Email : sfliu@smu.edu.sg</div> <div>Office : School of Economics, 5044</div>		
COURSE DESCRIPTION		
<p>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 <i>R</i> programming language into most topics, providing students with hands-on experience in applying statistical techniques to real-world scenarios commonly encountered in business environments.</p> <p>Basic understanding of statistics and familiarity with programming concepts is recommended. No prior experience with <i>R</i> is required, as introductory <i>R</i> programming concepts will be covered in the course.</p>		
LEARNING OBJECTIVES		
<p>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.</p>		
PRE-REQUISITES		
<p>Basic Statistics/Econometrics/Quantitative Methods</p>		
ASSESSMENT METHODS		
Class Participation	20%	
Assignments	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 *R* 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., (2021). *An Introduction to Statistical Learning with Applications in R* (2nd 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 R for data analysis and visualization Review of Statistics and Probability <ul style="list-style-type: none"> ○ Sample vs. population ○ Statistics vs. parameters ○ Descriptive statistics vs. inferential statistics ○ Discrete vs. continuous random variables
2		Review of inferential statistics <ul style="list-style-type: none"> ○ Confidence interval estimation ○ Hypothesis testing
3		Probability modelling <ul style="list-style-type: none"> ○ Distribution modelling ○ Goodness-of-fit test
4		Review Simple Linear Regression <ul style="list-style-type: none"> ○ Least squares estimation ○ Goodness-of-fit and related inferences ○ Gauss-Markov conditions ○ Residual analysis ○ Confidence intervals and prediction intervals
5	Ch 3	Multiple Linear Regression <ul style="list-style-type: none"> ○ Least squares estimation ○ Model interpretation ○ Model assumptions and inferences
6	Ch 3	Multiple Linear Regression: Additional topics <ul style="list-style-type: none"> ○ Polynomial regression ○ Transformations ○ Interactive Effects ○ Qualitative predictors ○ Instrumental variables
7	TBC	Advanced Regression Topics <ul style="list-style-type: none"> ○ Another look at autocorrelation: Durbin-Watson test, Moran's I test ○ Time series models ○ Spatial models ○ Panel data models
8	Ch 4	Classification: Logistic Regression <ul style="list-style-type: none"> ○ Estimation, prediction, interpretation ○ Confusion matrix and ROC ○ Multiple logistic regression
9	Ch5	Resampling Methods <ul style="list-style-type: none"> ○ Bias-variance trade-off ○ Training MSE vs. test MSE ○ Cross-validation ○ The bootstrap

10	Ch 6	Linear Model Selection <ul style="list-style-type: none"> ○ Subset selection vs. shrinkage ○ Best subset selection ○ Stepwise selection
11	Ch 6	Regularisation methods <ul style="list-style-type: none"> ○ Ridge regression ○ LASSO ○ Elastic net regression Review lesson
12		Exam

Class sessions are of three-hour duration per session and twelve sessions in total. The above is a tentative lesson plan. Note that this schedule is subject to change.