

**MATH 3424 Regression Analysis  
Course Outline --- Spring 2024**

**1. Instructor(s)**

*Name:* Dr. XIA, Dong (L1)

*Contact Details:* Room 3431, tel: ext-2244, email: [madxia@ust.hk](mailto:madxia@ust.hk)

If you want to make an appointment with Dr. XIA, send him an email.

**2. Teaching Assistant(s)**

*Name:* Ms. DUAN, Congyuan (T1A), email: [cduanac@connect.ust.hk](mailto:cduanac@connect.ust.hk)

Mr. GU, Yanwu (T1B), email: [ygubg@connect.ust.hk](mailto:ygubg@connect.ust.hk)

If you want to make an appointment with your TA, send him an email.

**3. Meeting Time and Venue**

Lectures:

**Date/Time:** L1: Tue. 3:00PM - 4:20PM  
Thu. 3:00PM - 4:20PM

**Venue:** Rm 4621 (lift 31-32)

Tutorials:

**Date/Time:** T1A: Tue. 6:00PM - 6:50PM; [Rm2302, lift 17-18](#)

T1B: Fri. 12:30PM - 1:20PM; [Rm4580, lift 27-28](#)

(R programming is important. Our TA's will give several tutorial sessions in the first few weeks for the basics of R programming.)

**4. Course Description**

Credit Points: 3

Pre-requisite: MATH 2411 OR MATH 2421 OR MATH 3423

Exclusion: ISOM 5520

Brief Information/synopsis:

Estimation and hypothesis testing in linear regression, residual analysis, multicollinearity, indicator variables, variable selection, non-linear regression

**5. Intended Learning Outcomes**

Upon successful completion of this course, students should be able to:

No.	ILOs
1	Get familiar with simple and multiple linear regression, the involved statistical inference, prediction, the measure of goodness-of-fit, and applications on real data examples
2	Understand the diagnostics of regression, residuals, checking linearity, leverage, influence, dealing with outliers and applications on real data examples
3	Understand how to treat categorical variables as predictors, the transformation of variables, treating heteroscedastic errors and applications on real data examples
4	Understand variable selection, the procedures and applications
5	Understand logistic regression for classification problem, the quality of fit, determination of important variables and applications on real data examples

## 6. Assessment Scheme

a. Percentage of coursework, examination, etc.:

<u>Assessment 1</u>	<u>Assessing Course ILOs</u>
15% Assignments	<u>1,2,3,4,5</u>
30% Mid-term Test	1,2,3
55% Final Examination	1,2,3,4,5
<u>Assessment 2</u>	<u>Assessing Course ILOs</u>
15% Assignments	<u>1,2,3,4,5</u>
0% Mid-term Test	1,2,3
85% Final Examination	1,2,3,4,5

The final grade= $\max\{\text{Assessment 1, Assessment 2}\}$ , both schemes are applied to every student.

## 7. Student Learning Resources

Lecture notes and slides ([on canvas](#))

Textbook: Chatterjee and Hadi, *Regression Analysis by Example*, Fifth Edition (not required)  
David Dalpiaz, *Applied Statistics with R* ([available on GitHub](#), [google it](#))

## 8. Teaching and Learning Activities

Scheduled activities: 4 hrs (lecture + tutorial)

## 9. Course Schedule

Keyword Syllabus:

- **Chap 1 Introduction and Basic Prerequisite** (~3 hours)  
(i) Why regression analysis; (ii) Selected examples and general procedures; (iii) Prerequisite I: probability distribution; (iv) Prerequisite II: confidence intervals and t-Test ; (v) Prerequisite III: elements of matrix algebra.
- **Chaps 2 and 3 Simple and Multiple Linear Regression** (~7 hours)  
(i) Covariance and correlation coefficient; (ii) Least squares estimator, test of hypothesis and confidence interval; (iii) Prediction; (iv) Measure the quality of fit; (v) Centering and scaling; (vi) Multiple correlation coefficients; (vii) Inference for individual regression coefficients; (viii) Examples using R codes
- **Chap 4 Regression Diagnostics: Detection of Model Violations** (~5 hours)  
(i) Standard regression assumptions; (ii) residuals of various types; (iii) Graphical methods for model diagnostics; (iv) Checking linearity and normality assumption; (v) Leverage, influence and detection of outliers; (vi) Measures of Influence; (vii) Examples using R codes
- **Chap 5 Qualitative Variable as Predictors** (~4 hours)  
(i) Indicator variable; (ii) Variable for interactions; (iii) Systems of regression equations; (iv) More applications of indicative variables; (v) Seasonality; (vi) Examples with R codes
- **Chaps 6 and 7 Transformation of Variables and Collinearity** (~7 hours)  
(i) Transformation to achieve linearity; (ii) Transformation to stabilize variance; (iii) Detection

and removal of heteroscedastic errors; (iv) Logarithmic and power transformation; (v) Influence of collinearity on inference; (vi) Effects of collinearity on forecasting; (vii) detection of collinearity; (viii) Reduction of collinearity and Ridge regression

- **Chap 8 Variable and Model Selection** (~3 hours)  
(i) Motivation and consequences of variable selection; (ii) Uses of regression equations; (iii) Criteria for evaluating equations; (iv) Collinearity and variable selection; (v) Variable selection procedures; (vi) Examples using R codes
- **Chap 9 Logistic Regression** (~3 hours)  
(i) Logit model; (ii) Logistic regression diagnostic; (iii) Determination of variables to retain; (iv) Measure the fit of logistic regression; (v) Multinomial Logit Model; (vi) Examples with R with R codes