东北大学秦皇岛分校本科课程教学大纲

一、课程基本信息

<table>
<tr><td>开课单位</td><td colspan="3">悉尼智能科技学院</td><td>课程代码</td><td colspan="5">3100313032</td></tr>
<tr><td>中文名称</td><td colspan="8">回归分析与线性模型</td></tr>
<tr><td>英文名称</td><td colspan="8">Regression and Linear Models</td></tr>
<tr><td>总学时</td><td>72</td><td rowspan="2">理论学时</td><td rowspan="2">48</td><td rowspan="2">实验学时</td><td rowspan="2">24</td><td rowspan="2">上机学时</td><td rowspan="2">0</td><td rowspan="2">设计学时</td><td rowspan="2">0</td><td rowspan="2">课外学时</td><td rowspan="2">0</td></tr>
<tr><td>学　分</td><td>4</td></tr>
<tr><td>课程类型</td><td colspan="11">☐理论课程　☑理论（含实验）课程　☐实践类课程</td></tr>
<tr><td>课程类别</td><td colspan="11">☐通识类　☐学科基础类　☑专业方向类　☐实践类</td></tr>
<tr><td>适用专业</td><td colspan="11">应用统计学</td></tr>
<tr><td>先修课程</td><td colspan="11">数学分析与建模（一）、线性代数、统计学导论</td></tr>
<tr><td>建议开课学期</td><td colspan="11">☐1-1 ☐1-2 ☑2-1 ☐2-2 ☐3-1 ☐3-2
☐4-1 ☐4-2 ☐5-1 ☐5-2 ☐*（代表每学期都开设）</td></tr>
<tr><td>考核方式</td><td colspan="3">☑考试 ☐考查</td><td>成绩分制</td><td colspan="6">☑百分制 ☐五级制 ☐二级制</td></tr>
<tr><td>使用教材</td><td colspan="11">1、 无</td></tr>
<tr><td>参考书及教学资源</td><td colspan="11">1、 N.R. Draper, H. Smith., Applied Aggression Analysis(3rd ed), A Wiley-Interscience Publication JOHN WILEY & SONS, INC. New York,1998.</td></tr>
<tr><td>其他说明</td><td colspan="11">无</td></tr>
</table>

二、课程简介

中文描述

本课程为应用统计学专业重要的专业基础课程之一。本课程主要包括一元和多元线性回归方程的参数估计、显著性检验及其应用，违背回归模型基本假设的异方差、自相关等问题的诊断和处理方法，回归变量选择，逐步回归方法及多重共线性，加权最小二乘法等方法处理不满足方差齐性，逻辑回归模型、检验以及应用等内容。本门课使学生掌握统计学的基本思想、理论和方法的主要课程，以及培养学生熟练应用计算机软件处理统计数据和建模的能力。

English Description

Applied regression analysis is one of the important professional basic courses of students majoring in Statistics. This course mainly includes parameter estimation, significance test and application of simple and multivariate linear regression equations, diagnosis and treatment methods of heteroscedasticity, autocorrelation and other problems that violate the basic assumptions of regression model, selection of regression variables, stepwise regression method and multi-collinearity. Methods such as weighted least squares are used to address non homogeneity of variance, logistic regression models, tests, and applications. This course enables students to master the basic ideas, theories and methods of statistics, and cultivate students' ability to skillfully use computer software to process statistical data.

三、课程目标

1、知识目标：

1-1：了解回归分析的历史和发展过程。

1-2：掌握随机变量的定义、性质以及描述方法。

1-3：掌握简单线性回归模型的建立与检验。

1-4：掌握多元线性回归模型的建立与检验。

1-5：掌握“非线性”线性回归的方法。

1-6：掌握广义和加权最小二乘法。

1-7：掌握分类随机变量分析。

1-8：掌握简单和多元逻辑回归模型。

2、能力目标：

2-1：会使用定义和性质描述随机变量。

2-2：能建立简单线性回归模型并对模型检验。

2-3：能建立多元线性回归模型并对模型检验。

2-4：能对非线性关系的变量建立模型。

2-5：会使用加权法处理违背方差齐性的数据。

2-6：会用交叉表分析分类数据。

2-7：会对类别因变量数据建立逻辑回归模型并检验。

3、德育/素养目标：

3-1：培养具有不畏困难、锲而不舍、敢于尝试、不惧失败、迎难而上的精神，并在学习过程中培养自己的细心和耐心的勇气和精神。

3-2：培养服务意识，具有 “以人为本” 的服务精神。

3-3：培养遵守法律、懂规则、讲规则的新时代公民。

表 1 课程目标与毕业要求的对应关系

毕业要求	指标点	课程目标	关联 (H/M/L)
1. 理学知识：具有扎实的数学基础，能够将数学、自然科学和专业知识用于解决复杂实际问题。	1.2 掌握统计调查、统计数据处理、统计分析、计算机与统计软件使用等应用统计学的基本理论、知识与方法，具备采集、处理、分析数据的能力，熟悉预研报告、可行性分析报告、研究方案等文档的撰写规范。	1-1 到 1-7、2-1 到 2-7、3-1 到 3-3	H
2. **问题分析**：能够借助应用统计学的基本原理、方法和手段，识别、表达、并通过文献研究分析复杂实际问题，以获得有效结论。	2.2 能够借助应用统计学的基本原理、方法和手段，针对复杂实际问题设计针对性的方案，并综合运用文献、科学理论和技术手段予以解决。	1-1 到 1-7、2-1 到 2-7、3-1 到 3-3	M
4. **研究**：能够基于科学原理并采用科学方法对复杂实际问题进行研究，包括设计实验、分析与解释数据、并通过信息综合得到合理有效的结论。	4.1：能够基于科学原理并采用科学方法，在本专业相关理论指导下对复杂实际问题设计实验进行研究。	1-1 到 1-7、2-1 到 2-7、3-1 到 3-3	M
	4.2：能够结合本专业知识对实验数据进行分析与解释，设计并优化实验方案，并通过信息综合得到合理有效的结论。	1-1 到 1-7、2-1 到 2-7、3-1 到 3-3	H

四、主要教学内容

（一）理论教学

第 1 章：Introduction（支撑教学目标 1-1、3-1 至 3-3）

【1】知识点：

(1) A brief history of Regression Analysis and its development.

(2) The aim of Regression Analysis.

(3) Basic ideas of Regression Analysis.

【2】学习目标：

理解：The aim of Regression Analysis.

掌握：Basic ideas of Regression Analysis.

【3】重点内容：

(1) The aim of Regression Analysis.

(2) Basic ideas of Regression Analysis.

【4】难点内容：

(1) Basic ideas of Regression Analysis.

第 2 章：Statistics review（支撑教学目标 1-2、2-1、3-1 至 3-3）

【1】知识点：

(1) Random variable.

(2) Population statistics.

(3) Sample statistics.

(4) Hypothesis testing.

【2】学习目标：

理解：Definition, properties and description methods of RV; Basic statistics and sample statistics.

掌握：Hypothesis testing.

【3】重点内容：

(1) Basic statistics and sample statistics..

(2) Several important hypothesis tests.

【4】难点内容：

Construction of statistics of hypothesis test

第 3 章： Simple Linear Regression（支撑教学目标 1-3、2-2、3-1 至 3-3）

【1】知识点：

(1) Model fit (ANOVA and F-test; ANOVA and R2; leverage and influence).

(2) Assumptions of model, Population statistics.

(3) Checking model assumptions (residual analysis).

(4) Hypothesis testing.

【2】学习目标：

理解：Definition, properties and description methods of RV; Basic statistics and sample statistics.

掌握：Hypothesis testing.

【3】重点内容：

(1) ANOVA and F-test..

(2) Checking model assumptions (residual analysis).

【4】难点内容：

Residual analysis.

第 4 章： Multiple Linear Regression（支撑教学目标 1-4、2-3、3-1 至 3-3）

【1】知识点：

(1) Multidimensional least squares.

(2) Model assumptions and statistical properties.

(3) Model fit.

(4) Model selection (forward selection、backward elimination、stepwise regression).

(5) Model test.

(6) Model selection (selection from all possible models, forward selection, backward elimination, stepwise regression).

(7) Categorical predictors and dummy variables.

(8) Interaction effects.

(9) Partial sum squares and F-test

【2】学习目标：

理解：Multidimensional least squares.

掌握：Fit models and hypothesis testing.

【3】重点内容：

(1) Model assumptions and statistical properties.

(2) T-test, ANOVA and F-test.

(3) Categorical predictors and dummy variables.

【4】难点内容：

(1) multidimensional least squares(scalar form, matrix form).

(2) Testing of model assumptions.

(3) Interaction effects.

【5】思政案例：

Introducing the case of the consumption level of Chinese residents, since the reform and opening up, the income and consumption level of Chinese residents have rapidly increased. There are many factors that affect residents' consumption level, and a multiple linear regression model is introduced through the case of residents' consumption level. In the subsequent knowledge explanation, students will discover the similarities and differences between multiple linear regression models and univariate linear regression models, and better understand and master multiple linear regression models. The development of things from simple to complex, and from low-level to high-level, follows a spiral upward trend. By discovering both similarities and differences between multiple linear regression models and univariate linear regression models, students can better understand and master the basic assumptions, solving methods, and significance tests of multiple linear regression models.

第 5 章： “Non-linear” linear regression（支撑教学目标 1-5、2-6）

【1】知识点：

(1) Transforming independent variables.

(2) Transforming dependent variable.

(3) Quadratic models and polynomial models.

【2】学习目标：

理解：Model assumptions; Model selection.

掌握：Fit models and hypothesis testing.

【3】重点内容：

(1) Fitting and testing of non-linear models..

(2) Model selection.

【4】难点内容：

Model selection.

第 6 章： Generalized Least Squares（支撑教学目标 1-6、2-5、3-1 至 3-3）

【1】知识点：

(1) Generalized Least Squares (GLS).

(2) Weighted least squares (WLS).

(3) Model fit.

(4) Model test.

【2】学习目标：

理解：Definition of Generalized Least Squares.

掌握：Essence of GLS and WLS.

【3】重点内容：

(1) Fitting and testing of models with WLS..

(2) Fitting and testing of models with GLS.

【4】难点内容：

Model fit with WLS, and model selection.

第 7 章： Analysis of Categorical RVs（支撑教学目标 1-7、2-6）

【1】知识点：

(1) Multinomial distribution.

(2) Chi-square goodness-of-fit test.

(3) Two-way tables.

(4) Chi-square independence test

(5) Relative risks.

(6) Odds and odds ratios.

【2】学习目标：

理解：Multinomial distribution.

掌握：Chi-square goodness-of-fit test and independence test.

【3】重点内容：

(1) Chi-square independence test.

(2) Chi-square goodness-of-fit test.

【4】难点内容：

Odds and odds ratios.

第 8 章： Logistic Regression（支撑教学目标 1-8、2-7）

【1】知识点：

(1) Binary response variable.

(2) Link functions.

(3) Simple logistic regression (continuous predictor, categorical predictor, confidence intervals, hypothesis tests).

(4) Multiple logistic regression (log-likelihood function).

(5) Model fit.

(6) Wald test(upper-tail rejection, two-tail rejection, lower-tail rejection).

(7) Omnibus test(null hypothesis rejection).

(8) Hosmer-Lemeshow test.

(9) Pseudo-R2 statistics.

(10) model selection.

【2】学习目标：

理解：Binary response variable and link functions.

掌握：Fit a simple (multiple) logistic regression model.

【3】重点内容：

(1) Fit logistic regression model.

(2) goodness-of-fit test and independence test.

【4】难点内容：

(1) Chi-square goodness-of-fit test.

(2) The meaning of coefficients.

（二）实验教学

表 2　实验教学信息表

序号	实验项目名称	实验学时	每组人数	实验类型	实验性质
1	Introduction to R	2	1	Verify	Compulsory
2	Simple Linear Regression I	2	1	Comprehensive	Compulsory
3	Simple Linear Regression II	2	1	Comprehensive	Compulsory
4	Multiple Linear Regression I	2	1	Comprehensive	Compulsory
5	Multiple Linear Regression II	2	1	Comprehensive	Compulsory
6	Multiple Linear Regression III	2	1	Comprehensive	Compulsory
7	Non-linear Regression	2	1	Comprehensive	Compulsory
8	Weighted Least Squares Regression	2	1	Comprehensive	Compulsory
9	Analysis of Categorical RVs	2	1	Comprehensive	Compulsory
10	Simple Logistic Regression	2	1	Comprehensive	Compulsory
11	Multiple Logistic Regression	2	1	Comprehensive	Compulsory
12	Collect Data and Build Regression Models	2	1	Verify	Compulsory

实验项目一：　Introduction to R（支撑教学目标 1-1 至 1-7，2-1 至 2-7，3-1 至 3-3）

【1】实验内容：

(1) Creates various data objects.

(2) Import and open data tables.

(3) Basic calculation functions.

(4) Various data visualization charts.

【2】实验目的：

(1)　Familiar to create various data objects.

(2)　Master the method of data visualization by R.

【3】教学要求：

(1)　Master the basic methods of analyzing data.

【4】实验方式： Programming.
【5】实验场地与设备： Laboratory, Computer.

实验项目二： Simple Linear Regression I（支撑教学目标 1-3，2-2）
【1】 实验内容：
(1) Find the linear relation between variables.
(2) Establish simple linear regression model.
【2】实验目的：
(1) Familiar to testing the linear relation between variables.
(2) Master the method of Simple Linear Regression by R.
【3】教学要求：
(1) Given data, learn and master the steps to solve the problem in simple linear regression..
【4】实验方式： Programming.
【5】实验场地与设备： Laboratory, Computer.

实验项目三： Simple Linear Regression II（支撑教学目标 1-3，2-2）
【1】实验内容：
A simple linear regression equation is established for the given data
【2】实验目的：
Master the method of Simple Linear Regression by R
【3】教学要求：
learn and master the steps to solve the problem in simple linear regression, and master the method of testing the coefficients and model.
【4】实验方式： Programming.
【5】实验场地与设备： Laboratory, Computer.

实验项目四： Multiple Linear Regression I （支撑教学目标 1-4，2-3）
【1】实验内容：
A multiple regression equation is established for the given data
【2】实验目的：
Master the method of multiple Linear Regression by R
【3】教学要求：
learn and master the steps to solve the problem in multiple linear regression, and master the method of testing the coefficients and model.
【4】实验方式： Programming.
【5】实验场地与设备： Laboratory, Computer.

实验项目五： Multiple Linear Regression II（支撑教学目标 1-4，2-3）
【1】实验内容：
(1) Based on correlations, chose two predictors as candidate predictors for a multiple regression.
(2) Write down the model selected at each step of the forward(backward) procedure Basic calculation functions.
(3) Describe two ways in which the final model of the backward procedure is superior to that of the forward model.
【2】实验目的：

(1) Learn the forward method and backward method in multiple regression.

【3】教学要求：

For the given data, skillfully use R to do multiple regression

【4】实验方式： Programming.

【5】实验场地与设备： Laboratory, Computer.

实验项目六： Multiple Linear Regression III（支撑教学目标 1-4，2-3）

【1】实验内容：

(1) Write down the estimated regression equation and provide interpretations of the estimated beta coefficients.

(2) Determine if there is any statistical evidence of serial correlation which would violate the assumption of independence.

(3) Determine if there is any statistical evidence of multicollinearity.

(4) Perform a visual analysis of the residuals for compliance with the normality, independence and constant variance assumptions.

(5) Using 0.05 significance level, perform a hypothesis test as to the normality of the residuals.

【2】实验目的：

Master multicollinearity and residual analysis

【3】教学要求：

Learn to use R for multicollinearity and residual analysis of multiple regression model

【4】实验方式： Programming.

【5】实验场地与设备： Laboratory, Computer.

实验项目七： "Non-linear" Linear Regression（支撑教学目标 1-5，2-4，3-1 至 3-3）

【1】实验内容：

(1) Describe the direction, type and strength of the relationship between wall and weight.

(2) If a straight-line model was fitted to the data, two problems we would see with the residuals.

(3) Write down the estimated regression model and provide interpretations of the estimated beta coefficients.

(4) Write down the estimated regression model in log-unitsand provide interpretations of the estimated beta coefficients.

【2】实验目的：

Master the method of choosing log-units model instead of linear model.

【3】教学要求：

Understand the benefits of log- units.

【4】实验方式： Programming.

【5】实验场地与设备： Laboratory, Computer.

实验项目八：Weighted Least Squares Regression（支撑教学目标 1-6，2-5，3-1 至 3-3）

【1】实验内容：

(1) Obtain scatter plots of photo versus obs1 and of photo versus obs2 (a matrix scatter plot will do)..

(2) If you carried out a simple linear regression of photo versus obs1 or photo versus obs2, what problems would you expect?

(3) What would the weights be for the WLS regression of photo versus obs1?

(4) For both WLS models, analyze the Student-T version weighted residuals using scatter plots

involving the independent variables.

【2】实验目的：

Learn the essence and key steps of WTS model

【3】教学要求：

Master the method of obtaining WTS model by R

【4】实验方式： Programming.

【5】实验场地与设备： Laboratory, Computer.

实验项目九：Analysis of Categorical RVs（支撑教学目标 1-7，2-6，3-1 至 3-3）

【1】实验内容：

(1) Based on the test output, conclude if field goal success linked to game time quarter.

(2) Using the count data (not the row probabilities), calculate the odds of successful field goal attempts ($\mathrm{good} = 1$) in the first quarter of game time ($\mathrm{qtr} = 1$).

(3) Based on the test output, conclude if field goal success linked to kicking distance quartile.

【2】实验目的：

Learn the essence and key steps of analysis of Categorical RVs.

【3】教学要求：

For specific problems, analyze and classify random variables.

【4】实验方式： Programming.

【5】实验场地与设备： Laboratory, Computer.

实验项目十：Simple Logistic Regression（支撑教学目标 1-8，2-7，3-1 至 3-3）

【1】实验内容：

(1) Write down the fitted logistic regression model in log-odds scale, odds scale and probability scale.

(2) Using the regression model, carry-out the calculations to show that the odds of successful field goal in quarter 1 and in quarter 4 almost match those calculated in Lab 8.

(3) Interpret the impact of $\mathrm{distance}$ on the log-odds scale and the odds scale.

【2】实验目的：

Master the establishment method of logistic regression model.

【3】教学要求：

For specific data, an appropriate logistic regression model is established.

【4】实验方式： Programming.

【5】实验场地与设备： Laboratory, Computer.

实验项目十一：Multiple Logistic Regression（支撑教学目标 1-8， 2-7，3-1 至 3-3）

【1】实验内容：

(1) Write down the fitted logistic regression model in log-odds scale, odds scale and probability scale.

(2) Using 0.05 significance level, test if the regression is significant. Write down the null and alternative hypotheses , the test statistic and p-value, the result of the test and a conclusion in non-mathematical language.

(3) Using 0.05 significance level, test if the model provides an adequate fit to the data.

【2】实验目的：

Master the establishment method of logistic regression model.

【3】教学要求：

Appropriate logistic regression model is established, and model tests are done.

【4】实验方式： Programming.

【5】实验场地与设备： Laboratory, Computer.

实验项目十二：Collect Data and Build Regression Models（支撑教学目标 1-1 至 1-8，2-1 至 2-7，3-1 至 3-3）

【1】实验内容：

(1) Collect certain economy or social data from web.

(2) Try various models and find the most suitable one.

(3) Conduct various tests on the model and explain it reasonably.

【2】实验目的：

Master the establishment method of various regression models.

【3】教学要求：

Find suitable models for real data.

【4】实验方式： Programming.

【5】实验场地与设备： Laboratory, Computer.

【6】思政案例：

Enable students to truly experience the significant effectiveness of mathematical models in solving practical problems, and cultivate their ability to learn and apply flexibly. At the same time, students should be guided to delve deeper into the cultural connotations behind the cases. To integrate elements such as historical perspective, cultural perspective, national sentiment, and confidence into case teaching, promote Chinese culture, tell Chinese stories well, guide students to establish cultural confidence, cultivate students' sense of responsibility, and strengthen the educational function of ideological and political education in the curriculum.

五、教学安排与学时分配

表 3 学时分配表

讲课内容		学时				
		讲课	实验	上机	设计	课外
第 1 章	Introduction to Regression Analysis	2	0	0	0	0
第 2 章	Statistics review	4	2	0	0	0
第 3 章	Simple Linear Regression	8	4	0	0	0
第 4 章	Multiple Linear Regression	12	6	0	0	0
第 5 章	“Non-linear” linear regression	4	2	0	0	0
第 6 章	Generalized Least Squares	4	2	0	0	0
第 7 章	Analysis of Categorical RVs	4	2	0	0	0
第 8 章	Logistic Regression	8	6	0	0	0
	Review	2	0	0	0	0
小计		48	24	0	0	0
总计	72					

六、教学方法

1、Multimedia Teaching: Classroom Teaching Based on Information Technology Equipment.

2、Combining theory with industrial practical problems

3、Knowledge delivery with ethic education

七、成绩评定

1、Behavior : The full score is 100 points.5 points are counted for each attendance, and no score is given for absence. And 5 points are counted for each assignment, no score for plagiarism, plagiarism for others or no assignment. The final total score is not more than 100 points, not less than 0 points

2、Experiment : The full score is 100 points. There are 12 computer experiment classes in total. Each class needs to submit a report.

3、Final: The full score is 100, and the students' scores are given by marking the final examination papers.

八、评价及持续改进

1、课程负责人根据课程教学内容与人才培养目标组织课程团队讨论并修改教学大纲，报分管教学工作副院长审核后由执行院长批准。

The subject coordinator shall be responsible for the syllabus discussion and improvement, and the revised version shall be submitted to deputy dean (teaching affairs) for reviewing then to executive dean for approval.

2、课程负责人根据课程教学内容、课堂教学效果以及成绩分布，对课程教学方法和成绩评定环节进行改进，并同步优化评定办法。

The subject coordinator shall revise the syllabus based on the teaching content, effect and result distribution while optimize the assessment measures.

….

撰写： 薛昌涛

审核： 韩鹏

批准： 史文博

日期： 2023.07.05