STAT 2620: Introduction to Data Analysis and Statistical Modeling

2023 Summer Session	
Total Class Sessions: 25 Class Sessions Per Week: 5 Total Weeks: 5 Class Session Length (Minutes): 145 Credit Hours: 4	Instructor: Staff Classroom: TBA Office Hours: TBA Language: English

Course Description:

This course introduces the method of data analysis and statistical modeling based on the introductory statistical knowledge. This course also provides an introduction to the Statistical Analysis System (SAS) software, and students will become proficient in the use of SAS for managing and analyzing complex data sets.

Topics discussed include the various types of data, visual display of data through graphics, linear and nonlinear regression models, analysis of ANOVA and ANCOVA . It specifically introduces survival time analysis and time series analysis, which includes appropriate censoring mechanisms and independent assumptions.

Prerequisite: Introduction to Statistics.

Learning objectives:

Upon successful completion of this course, the student will be able to do the following:
•Develop understanding on the different types of data, apply different techniques to display example dataset
•Demonstrate ability on analyzing complex dataset in the use of SAS
•Be able to choose the appropriate statistical model for given data and situations
•Appreciate the role of statistics in empirical research and scientific study
•Perform simple and multiple regression, logistic regression, survival data analysis and interpret the statistical output from SAS to make proper inference
•Read the scientific literature and comprehend the use of survival analysis methodologies and time series analysis methodologies
•Develop the skills of statistical computation, report writing, and oral presentations

Course Materials:

There is no required textbook, course materials will select from the following reference materials
Reference materials:
1.Categorical Data Analysis, 3rd edtion, Alan Agresti, 2012
2.STAT2: Modeling with Regression and ANOVA, 2nd Edition, by Ann R.Cannon, George W. Cobb, Bradley A. Hartlaub, 2018
3.The little SAS Book: A Primer, 5 th ed., Delwich & Slaughter, 2012, SAS Institute

Course Format and Requirements:

The primary format of this course is lecture, problem solving and discussion. Familiarizing with the course material before class, you will gain a better understanding the information presented during lecture. Each student will be responsible for learning as much as possible. Students are strongly encouraged to ask questions on things you did not understand.

Attendance:

Attendance will not be taken but all quizzes will be the taken at the beginning in class. Arriving late may cause you to miss a quiz, impacting your performance assessment. There is no made-up quiz.

Course Assignments:

Assignments: There will be weekly assignments that are all individual works. They will be randomly assigned to certain topics. Pay attention that the skills and knowledge would enhance students understanding about the topic and would serve as a part of the exams. **ALL assignments require skills learned during the labs.**
Project: A project will focus on the application of data analysis and statistical modeling on certain case. The project aims to help students deepen theoretical knowledge and technical skills acquired in the class.
Exams: Midterm Exam There will be one midterm exam in this course. The midterm exam will be based on concepts covered in class. They will be in-class, close-book and non-cumulative. Final Exam The final will be cumulative and close-book. Note that the final will not be taken during the normal class times. Exact time and location for final will be announced later.
Computer lab: A significant laboratory component will be integrated and will complement the classroom. Student will be required to analyze data with SPSS. The mastery of SPSS will benefit in completing the project and future application.

Course Assessment:

Assignments	15%
Project	15%
Computer lab engagement	20%
Midterm Exams	20%
Final Exam	25%
Attendance	5%
Total	**100%**

Grading Scale (percentage):

A+	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F
98-100	93-97	90-92	88-89	83-87	80-82	78-79	73-77	70-72	68-69	63-67	60-62	<60

Academic Integrity:

Students are encouraged to study together, and to discuss lecture topics with one another, but all other work should be completed independently.

Students are expected to adhere to the standards of academic honesty and integrity that are described in the Chengdu University of Technology's *Academic Conduct Code.* Any work suspected of violating the standards of the *Academic Conduct Code* will be reported to the Dean's Office. Penalties for violating the *Academic Conduct Code* may include dismissal from the program. All students have an individual responsibility to know and understand the provisions of the *Academic Conduct Code*.

Special Needs or Assistance:

Please contact the Administrative Office immediately if you have a learning disability, a medical issue, or any other type of problem that prevents professors from seeing you have learned the course material. Our goal is to help you learn, not to penalize you for issues which mask your learning.

Course Schedule:

Week	Topics	Assignments
Week One (Class 1~5)	• Introduction to SAS, SAA data step, SAS syntax • Different types of statistical date, example datasets, techniques for the visual display of different types of statistical data through graphics, goals of data analysis, types of studies, types of variables, • Review on Basic statistics, random variables and distributions, estimation, hypothesis testing • Introduction to linear regression, the linear model, finding the best fitting line, interpreting intercept and slope estimates, testing hypotheses and forming confidence intervals, assessing model assumptions, programming concepts • **Computer Lab** 1, 2, 3: graphical methods for presenting data, categorical data, contingency tables	• Quiz 1 • Assignment 1

Week Two (Class 6~10)	• Interpreting parameter estimates from transformed data, computing predicted values, the correlation coefficient • Categorical independent variables, splines • The Simple and multiple regression model, hypothesis testing, evaluating model assumptions • PROC GLM vs. PROC REG, multiple and partial correlations • Interaction and confounding • Computer Lab 4, 5, 6: linear regression models	• Quiz 2 • Assignment 2
Week Three (Class 11~15)	• Regression diagnostics, residual analysis, detection of outliers, collinearity, scaling X variables, model selection • Analysis of variance (ANOVA), 1-way design, the F-test, multiple comparisons, factorial designs, repeated measures, split plot analysis using mixed models • Analysis of covariance (ANCOVA), general approaches to nonparametric analysis • Computer Lab 7, 8, 9: Regression diagnostics, analysis of ANOVA and ANCOVA	• Quiz 3 • Midterm • Assignment 3
Week Four (Class 16~20)	• Nonlinear Regression Models, least squares estimation in nonlinear regression, logistical regression, fitting logistic model, test of goodness of fit, logistic regression diagnostics, Poisson regression model, Bayesian methods/ Cure models • Survival time analysis, exponential and Weibull Distribution, Survival function, hazard function, mean and median life, principal parametric model • Computer Lab 10, 11, 12: Nonlinear Regression Model (Logistic and Poisson Regression Model), survival time analysis	• Quiz 4 • Assignment 4
Week Five	• Different types on censoring, left and interval censoring, Building	• Quiz 5

(Class 21~25)	the likelihood function, left truncation, Kaplan-Meier estimator for the survival function, Nelson-Aalen estimator for the cumulative risk function, asymptotic properties, confidence intervals and confidence bands ● Time series analysis, characteristics of time series data, Stationarity, time series regression, detrending and de-seasoning, smoothing, estimation and forecasting, basics of ARIMA models, basics of GARCH models, intro to spectral analysis ● Computer Lab 13, 14, 15: survival time analysis and time series analysis	• Assignment 5 • Project • Final exam (cumulative) TBA