University of Colorado **Boulder**

Master of Science in Data Science

DTSA 5011 Modern Regression Analysis in R

- **Specialization:** Statistical Modeling for Data Science Applications
- **Instructor:** Brian Zaharatos, Director, Professional Master's Degree in Applied Mathematics
- **Prior knowledge needed:** Basic calculus (differentiation and integration), linear algebra, probability theory

View on Coursera

Learning Outcomes

Successful completion of this course demonstrates your achievement of the following learning outcomes for the MS-DS on Coursera:

- Acquire, clean, wrangle, and manage data.
- Correctly perform exploratory data analyses in order to assist with the generation of scientific hypotheses.
- Apply principles and methods of probability theory and statistics to draw rational conclusions from data.
- Construct an appropriate statistical model in order to answer important scientific or business-related questions.
- Assess the validity of a statistical model when applied to a particular dataset.
- Be sensitive to ethical issues that are involved in dealing with data science applications arising in real world situations.
- Clearly communicate the results of a data science analysis to a non-technical audience.
- Use peer feedback, self-reflection and video analysis to improve collaboration skills.
- Create reproducible statistical workflows.
- Act ethically in the role of professional data scientist.

Syllabus

Week 1 | Introduction to Statistical Models
Duration: 7h 32m

In this module, we will introduce the basic conceptual framework for statistical modeling in general, and linear statistical models in particular.

Week 2 | Linear Regression Parameter Estimation
Duration: 8h 14m

In this module, we will learn how to fit linear regression models with least squares. We will also study the properties of least squares, and describe some goodness of fit metrics for linear regression models.

Week 3 | Inference in Linear Regression
Duration: 8h 32m

In this module, we will study the uses of linear regression modeling for justifying inferences from samples to populations.

Week 4 | Prediction and Explanation in Linear Regression Analysis
Duration: 5h 52m

In this module, we will identify how models can predict future values, as well as construct confidence intervals for those values. We will also explore the relationship between statistical modelling and causal explanations.

Week 5 | Regression Diagnostics
Duration: 7h 12m

In this module, we will learn how to diagnose issues with the fit of a linear regression model. In particular, we will use formal tests and visualizations to decide whether a linear model is appropriate for the data at hand.

Week 6 | Model Selection and Multicollinearity
Duration: 7h 36m

In this module, we will identify measures to improve our models after they have been fit to the data. In particular, we will learn when and how to apply model selection techniques such as forward selection and backward selection, criterion-based methods and multicollinearity diagnostics.

Grading

Assignment	**Percentage of Grade**
Quiz: Introduction to Statistical Modeling	1%
Quiz: The Linear Regression Model	1%
Programming Assignment: Module 1 Autograded	5%
Peer Review: Module 1 Peer Review Submission	5%
Quiz: Least Squares	1%
Quiz: Variability and Identifiability in Regression Models	1%
Programming Assignment: Module 2 Autograded Assignment	5%

Peer Review: Module 2 Peer Review Submission	6%
Quiz: Statistical Inference: Intro and T-Tests	1%
Quiz: Statistical Inference: the F-tests and Confidence Intervals	1%
Peer Review: Ethics in Statistical Practice and Communication: Five Recommendations	5%
Programming Assignment: Module 3 Autograded Assignment	5%
Peer Review: Module 3 Peer Reviewed Lab Submission	6%
Quiz: Prediction	1%
Programming Assignment: Module 4 Autograded Assignment	5%
Peer Review: Module 4 Peer Reviewed Lab Submission	6%
Quiz: Diagnostics I: Linearity and Independence	1%
Quiz: Diagnostics II: Constant Variance and Normality	1%
Programming Assignment: Module 5 Autograded Assignment	5%
Peer Review: Module 5 Peer Review Submission	5%
Quiz: Model Selection II: Criterion-based Procedures	1%
Quiz: Multicollinearity	1%
Programming Assignment: Module 6 Autograded Assignment	5%
Peer Review: Module 6 Peer Review Submission	6%
Quiz: Final Exam	20%

Letter Grade Rubric

Letter Grade	Minimum Percentage
A	90%
B	80%
C	70%
D	60%
F	50%

Note: This page is periodically updated. Course information on the Coursera platform supersedes the information on this page. Click View on Coursera *button above for the most up-to-date information.*

University of Colorado Boulder

© Regents of the University of Colorado