



DEDMAN COLLEGE  
OF HUMANITIES AND SCIENCES

GRADUATE PROGRAMS ADDENDUM  
SOUTHERN METHODIST UNIVERSITY

2014-2015

# STATISTICAL SCIENCE

---

## MASTER OF SCIENCE IN DATA SCIENCE

**Associate Professor** Monnie McGee, **Co-Executive Director**

**Associate Professor** Daniel Engels, **Co-Executive Director**

**Professors:** Ira Greenberg, Lynne Stokes. **Associate Professors:** Daniel Engels, Monnie McGee. **Assistant Professors:** Eric Larson, Tyler Moore.

The M.S. in data science is an interdisciplinary degree that combines courses from SMU's Dedman College, Lyle School of Engineering and Meadows School of the Arts. The program is delivered entirely online to provide access to working professionals and students located around the world. The program gives students who are prospective data scientists an enriched background in computer science, statistics, strategic behavior and data visualization. Graduates will be able to form relevant questions, collect and analyze appropriate data, and make informed decisions regarding the question of interest. Course instruction includes oral and written communication skills and the basics of database structure, including building, maintaining and securing databases. Additional information is available online at [www.datascience.smu.edu](http://www.datascience.smu.edu), and the academic calendar is at [www.smu.edu/registrar](http://www.smu.edu/registrar) ("Academic Calendars and Course Catalogs" link).

### Admission Requirements

Applicants to the online M.S. in data science program must hold a bachelor's degree in statistics, applied mathematics, computer science, engineering or other quantitative discipline. An application fee is required. Additional requirements are as follows:

- A basic understanding of a programming language (R, JAVA, C++, Python or similar programming language).
- Calculus I and II. Business calculus does not meet this requirement.
- A completed online application, with the following:
  - Three letters of recommendation.
  - A statement of purpose.
  - All official academic transcripts.
  - An official GRE graduate school admission test score, which may be waived if a prospective student has five or more years of industry-related experience or has earned a master's degree in a related field.
- A satisfactory TOEFL English language proficiency test score is required for an applicant whose native language is not English.

### Degree Requirements

To qualify for the online M.S. in data science, the student must successfully complete at least 31 hours of credit acceptable to departmental faculty; at least 27 hours of credit must be for 6000-level courses or above. A one-credit immersion course on the SMU campus is also required. There may be additional expenses related to the on-campus immersion course.

<b>MSDS 61xx</b> Immersion on Campus	1
<b>MSDS 6306</b> Introduction to Data Science	3
<b>MSDS 6370</b> Statistical Sampling	3
<b>MSDS 6371, 6372</b> Experimental Statistics I, II	6
<b>MSDS 6383</b> New Approaches to Managerial Economics	3
<b>MSDS 6390</b> Visualization of Information	3
<b>MSDS 7330</b> File Organization and Database Management	3
<b>MSDS 7331</b> Data Mining	3
<b>MSDS 7333</b> Quantifying the World	3
<b>MSDS 7349</b> Data and Network Security	3
	<b>31</b>

### The Courses (MSDS)

**MSDS 61XX (1). IMMERSION ON CAMPUS.** Offers students the chance to meet in-person with classmates and faculty for collaborative, hands-on workshops and informational sessions. Emphasizes group work, networking, and relationship building. Takes place on the SMU campus in Dallas, Texas. Reserved for students in the M.S.D.S. program.

**MSDS 6306 (3). INTRODUCTION TO DATA SCIENCE.** An introduction to methods, concepts, and current practice in the growing field of data science, including statistical inference, algorithms, financial modeling, data visualization, social networks, and data engineering. Reserved for students in the M.S.D.S. program.

**MSDS 6370 (3). STATISTICAL SAMPLING.** Covers principles of planning and conducting surveys: simple random sampling; stratified and systematic subsampling; means, variances, and confidence limits; finite population correction; sampling from binomial populations; and margin of error and sample-size determination. Reserved for students in the M.S.D.S. program.

**MSDS 6371 (3). EXPERIMENTAL STATISTICS I.** Noncalculus development of fundamental statistical techniques, including hypothesis testing for population means and proportions, analysis of variance, factorial designs, and linear regression. Covers obtaining sample sizes during the planning stages of research studies and emphasizes interpretation of results from analysis with SAS statistical software. Reserved for students in the M.S.D.S. program.

**MSDS 6372 (3). EXPERIMENTAL STATISTICS II.** Extension of techniques in MSDS 6371 to multivariate data. Multiple linear regression, multivariate analysis of variance, canonical regression, and principal components analysis. Emphasizes interpretation of results from analysis with SAS. Reserved for students in the M.S.D.S. program.

**MSDS 6383 (3). NEW APPROACHES TO MANAGERIAL ECONOMICS.** Presents recent developments in information theory, multiproduct analysis, and principal-agent theory in order to show how these developments can be usefully applied in management decision-making. Reserved for students in the M.S.D.S. program.

**MSDS 6390 (3). VISUALIZATION OF INFORMATION.** Introduces data visualization and creative coding utilizing the Processing programming language. Explores visual and information design principles, primarily through hands-on programming exercises. Includes assignments and exams that incorporate 2-D and 3-D computer graphics, interactivity, and data input. Covers procedural and object-oriented programming approaches to data visualization and provides an overview of leading-edge data visualization libraries and application program interfaces, including Web-based approaches. Reserved for students in the M.S.D.S. program.

**MSDS 7330 (3). FILE ORGANIZATION AND DATABASE MANAGEMENT.** A survey of current database approaches and systems, and the principles of design and use of these systems. Covers query language design and implementation constraints, and applications of large databases. Includes a survey of file structures and access techniques. Also, the use of a relational database management system to implement a database design project. Reserved for students in the M.S.D.S. program.

**MSDS 7331 (3). DATA MINING.** Introduces data mining topics, with an emphasis on understanding concepts through an applied, hands-on approach. Includes other related topics such as data warehousing and dimensional modeling. All material covered is reinforced through hands-on implementation exercises. Reserved for students in the M.S.D.S. program.

**MSDS 7333 (3). QUANTIFYING THE WORLD.** In the global information age, data can be leveraged to rapidly answer previously unanswerable questions. Students explore how to make sense of the large amounts of data frequently available, from hypothesis formation and data collection to methods of analysis and visualization. Includes ways to set up Internet-level measurements and formulate testable hypotheses; ways to automatically gather, store, and query large datasets; and ways to apply statistical methods (descriptive and predictive) and information visualization to collected datasets. Students learn to use Python and R programming languages to carry out data collection, analysis, and visualization. Culminates in a final project using real data of the students' choosing. Reserved for students in the M.S.D.S. program.

**MSDS 7349 (3). DATA AND NETWORK SECURITY.** Covers conventional and state-of-the-art methods for achieving data and network security. Private key and public key encryption approaches are discussed in detail, with coverage of popular algorithms such as DES, Blowfish, and RSA. In the network security area, the course covers authentication protocols, IP security, Web security, and system-level security. Reserved for students in the M.S.D.S. program.