



SDSU Data Science Certificate Information

Contents

- Admissions Criteria
- Data Science Certificate Course Descriptions
- Preparatory Course Descriptions (not part of the Certificate Program)
- Example Programs of Study
- Cost to Attend
- Course Topic Lists

Admissions Criteria

Those wishing to pursue the Data Science Certificate should have the following background.

- An introductory course in statistics equivalent to SDSU course STAT 281 Introduction to Statistics. For those needing such a course, STAT 281 is available online every fall, spring, and summer.
- Prior statistical programming experience is strongly recommended, including some experience with SAS, R, and SQL. SDSU online courses STAT 510 SAS Programming I (3 credits) and STAT 514 R Programming (1 credit) provide the necessary background. It is also possible to self-teach most of this material using available online resources.

Data Science Certificate Course Descriptions

- **STAT 541 Statistical Methods II** - Analysis of variance, various types of regression, and other statistical techniques and distributions. Sections offered in the areas of Biological Science and Social Science. Prerequisites: STAT 281, MATH 381, or STAT 381, STAT 210 or STAT 410. Credit not given for both STAT 541 and STAT 582.
- **STAT 700 Statistical Programming** Fundamentals of statistical programming languages including descriptive and visual analytics in R and SAS, and programming fundamentals in R and SAS including logic, loops, macros, and functions.
- **STAT 701 Modern Applied Statistics I** Topics include statistical graphics, modern statistical computing languages, nonparametric and semiparametric statistical methods, longitudinal and repeated measures, meta-analysis, and large-scale inference. Prerequisite: STAT 700, STAT 541 or equivalent.
- **STAT 702 Modern Applied Statistics II** Topics include data mining techniques for multivariate data, including principal component analysis, multidimensional scaling, and cluster analysis; supervised learning methods and pattern recognition; and an overview of statistical prediction analysis relevant to business intelligence and analytics. Prerequisite: STAT 701

Preparatory Course Descriptions (not part of the Certificate Program)

- **STAT 281 Introduction to Statistics** - A study of descriptive statistics including graphs, measures of central tendency and variability and an introduction to probability theory, sampling and techniques of statistical inference with an emphasis on statistical applications.
- **STAT 510 SAS Programming I** - The Base SAS programming language for data reading and manipulation, data display, summarization, and graphing. Introduction to statistical procedures, high resolution graphics, the Output Delivery System, and some menu-driven interfaces. Includes basic SQL techniques
- **STAT 514 R Programming (1 credit)** – introduction to the R programming language. Topics will include the R programming language and environment, preparation and summarization of data, presentation of data, programming basics, and additional selected advanced topics.

Example Programs of Study

Base plan of study, for those having all necessary background

Spring I	STAT 541, 3 credits	Summer	STAT 700, 3 credits	Fall	STAT 701, 3 credits
Spring II	STAT 702, 3 credits				

Alternate plan of study, for those needing statistical programming background

Fall I	STAT 510, 1 credit	Spring I	STAT 541, 3 credits STAT 514, 1 credit	Summer	STAT 700, 3 credits
Fall II	STAT 701, 3 credits	Spring II	STAT 702, 3 credits		

Cost to Attend

For the 2016/17 academic year, cost per credit hour will be \$387.05 for state residents, and \$655.60 for non-residents.

Course Topic Lists

<p>STAT 510 SAS Programming I</p> <ul style="list-style-type: none"> • Prerequisite: familiarity with computer basics • Reading external data files into SAS • Creating permanent SAS data sets • Formatting and labeling variables • Using Excel through SAS • Conditional processing; Do loops • Subsetting and combining data sets • Using numeric and character functions; Arrays • Displaying data and customizing reports • Summarizing results; Frequencies; Tables; Graphing 	<p>STAT 514 R Programming</p> <ul style="list-style-type: none"> • Prerequisite: familiarity with computer basics • R installation and environment • Data structure, management, and manipulation • Logic statements (if else) • Basic loops (for, repeat, while, do until) • Use of in-built functions • Basic custom function writing • Installation of extensions to R • Intermediate programming techniques <ul style="list-style-type: none"> ○ apply, lapply, supply
<p>STAT 541 Statistical Methods II</p> <ul style="list-style-type: none"> • Prerequisite: STAT 281 or equivalent • Check here to see if you are ready for STAT 541. • Simple and multiple linear regressions • ANOVA for one or multiple factors • Design of experiments • Linear models with categorical data • Models with categorical response variable 	<p>STAT 700 Statistical Programming</p> <ul style="list-style-type: none"> • Prerequisite: STAT 541 • R programming <ul style="list-style-type: none"> ○ Formatting; Aggregation; Loops ○ Dynamic report generation (Sweave/Knitr) ○ External data sources • SAS programming <ul style="list-style-type: none"> ○ Defining and using macros ○ PROC SQL in Macros ○ Graphing in SAS
<p>STAT 701 Modern Applied Statistics I</p> <ul style="list-style-type: none"> • Prerequisite: STAT 541; STAT 700 or STAT 514 • Introduction to Statistical Graphics and GGplot • Logistic Regression I • Generalized Linear Models • Density Estimation • Recursive Partitioning • Generalized Additive Models and Spline Models • Survival Analysis • Longitudinal Data Analysis and Mixed Models • Multiple Comparisons; False Discovery Rates • Simultaneous Inference; Meta-Analysis 	<p>STAT 702 Modern Applied Statistics II</p> <ul style="list-style-type: none"> • Prerequisite: STAT 701 • Introduction to Statistical Learning • Introduction to Classification • Resampling Methods • Model Selection • "Moving Beyond Linearity" • Tree- Based Methods • Support Vector Machines • ROC curves • Clustering/Unsupervised Learning