Elements of Data Science

PEIMS Code: N1110035

Abbreviation: ELEMDSCI

Grade Level(s): 11-12

Award of Credit: 1.0

Approved Innovative Course

- Districts must have local board approval to implement innovative courses.
- In accordance with Texas Administrative Code (TAC) §74.27, school districts must provide instruction in all essential knowledge and skills identified in this innovative course.
- Innovative courses may only satisfy elective credit toward graduation requirements.
- Please refer to TAC §74.13 for guidance on endorsements.

Course Description:

The purpose of the Elements of Data Science course is to introduce students to statistical modeling and analysis considerably beyond the scope of Statistics/AP Statistics. In Elements of Data Science, students will learn to manipulate large datasets containing multiple explanatory variables, learn techniques for modeling, analysis, and visualization, and combine these skills with fundamental statistical principles to propose solutions to real-world problems. This course will empower students to grow in intuition as well as skillset and mature as analysts. Those who wish to pursue data science, or another STEM field, will find themselves prepared for the next level of their chosen educational or professional path.

Essential Knowledge and Skills:

General Requirements. This course is recommended for students in grades 11-12. Recommended prerequisites: *AP Statistics or Statistics.* Students shall be awarded one credit for successful completion of this course.

(b) Introduction.

(1) In Elements of Data Science, students will build on the knowledge and skills acquired in Statistics or AP Statistics, such as descriptive and inferential statistics, and introductory modeling. In this course, students will study predictive statistics and advanced modeling using multiple predictors for linear, analysis of variance (ANOVA), and logistic models. Students will use models to solve academic as well as real-world problems and use technology as a tool for statistical analysis.

(2) The purpose of the Elements of Data Science course is to introduce students to statistical modeling and analysis considerably beyond the scope of Statistics/AP Statistics. In Elements of Data Science, students will learn to manipulate large datasets containing multiple explanatory variables, learn techniques for modeling, analysis, and visualization, and combine these skills with fundamental statistical principles to propose solutions to real-world problems. This course will empower students to grow in intuition

as well as skillset and mature as analysts. Those who wish to pursue data science, or another STEM field, will be prepared for the next level of their chosen educational or professional path.

(c) Knowledge and skills.

(1) Mathematical process standards. The student uses mathematical processes to acquire and demonstrate mathematical understanding. The student is expected to:

(A) apply mathematics to problems arising in everyday life, society, and the workplace;

(B) use a problem-solving model that incorporates analyzing given information, formulating a plan or strategy, determining a solution, justifying the solution, and evaluating the problem-solving process and the reasonableness of the solution;

(C) select tools, including real objects, manipulatives, paper and pencil, and technology as appropriate, and techniques, including mental math, estimation, and number sense as appropriate, to solve problems;

(D) communicate mathematical ideas, reasoning, and their implications using multiple representations, including symbols, diagrams, graphs, and language as appropriate;

(E) create and use representations to organize, record, and communicate mathematical ideas;

(F) analyze mathematical relationships to connect and communicate mathematical ideas; and

(G) display, explain, or justify mathematical ideas and arguments using precise mathematical language in written or oral communication.

(2) Data Processing. The student uses the mathematical process standards to develop facility with spreadsheets and statistical software. The student is expected to:

(A) use the process of data cleaning in a spreadsheet to create data that is more useable and fits a preferred format;

(B) collect data from multiple sources and aggregate in a spreadsheet;

(C) import data sets into statistical software for analysis;

(D) perform row and column operations on data sets using statistical software;

(E) use statistical software to represent distributions with graphical displays; and

(F) use statistical software to generate summary statistics for distributions.

(3) Modeling and Regression. The student uses the mathematical process standards to build and assess models. The student is expected to:

(A) compare statistical models and their components including discussions of parameters and degrees of freedom;

(B) construct distributions and analyze statistics with the use of technology;

(C) develop and use a process for statistical modeling that includes, choosing a form for the model, fitting the model to data, determining how well the model fits data, and using the model to analyze mathematical and real-world situations;

(D) create and analyze an ANOVA table to partition variability in the context of simple linear regression; and

(E) construct prediction intervals based on linear models.

(4) Multiple Regression. The student uses the mathematical process standards to build and assess models with multiple explanatory variables. The student is expected to:

(A) analyze multiple simple linear regression models for the same response variable and predict how each explanatory variable affects the multiple regression model;

(B) compare two regression lines using a multiple regression model;

(C) create and assess regression models using functions of predictors, interactions, and polynomials;

(D) compare statistics, including adjusted R^2 and the VIF, for choosing predictors to include in a multiple regression model;

(E) describe issues of multicollinearity with correlated predictors;

(F) construct and interpret an added variable plot;

(G) use computer simulation methods to perform randomization tests for regression parameters; and

(H) use computer simulation methods to compute bootstrap confidence intervals for regression parameters.

(5) ANOVA and Randomized Experiments. The student uses the mathematical process standards to create and interpret ANOVA models. The student is expected to:

(A) identify a one-way ANOVA model and quantify within-group and between-group variation;

(B) assess the conditions for experimental design in an ANOVA setting;

(C) explain how the process of data collection affects the scope of conclusions;

(D) use transformations to select and assess an ANOVA model;

(E) use an ANOVA model to asses if groups have significantly different means and create interval estimates of group differences;

(F) identify a randomized complete block design;

(G) create and assess a main effects ANOVA model with two explanatory factors and use the model to make predictions;

(H) create and interpret interaction plots to determine if there is any interaction between groups;

(I) create and assess a two-way interaction model and use the model to make predictions;

(J) use computer simulation methods to perform inference in ANOVA settings;

(K) use ANOVA-based tests for homogeneity of variances;

(L) analyze group contributions to *F* statistics to determine which group(s) are different than the others; and

(M) identify and use an analysis of covariance (ANCOVA) model to determine the significance in the difference of the linear relationships of the levels of a categorical explanatory variable.

(6) Logistic Regression. The student uses the mathematical process standards to build and assess models using logistic regression. The student is expected to:

(A) describe the logistic transformation, the idea of odds, the odds ratio, and the logit function;

(B) identify and create a logistic regression model;

(C) use formal inference in the setting of logistic regression;

(D) create logistic regression models with two or more predictors;

(E) justify the appropriateness of the logistic regression model and perform transformations if necessary;

(F) analyze two-way tables using logistic regression; and

(G) use computer simulation methods to perform inference procedures for logistic regression parameters.

Recommended Resources and Materials:

- Microsoft Excel, Google sheets, or other spreadsheet program
- R and R Studio
- R Markdown

Recommended Course Activities:

Students will work individually or in small groups to choose, fit, assess, and use models for description, inference, and prediction.

Suggested methods for evaluating student outcomes:

Students participate in a rigorously paced lecture class. Students will be assessed by their performance on exams and quizzes. Students will apply content knowledge and modeling techniques to complete projects, based on real-world scenarios, to demonstrate mastery of the advanced applications of the sections for each grading period.

Teacher qualifications:

One of the following certificates is required:

- Grades 6-12 Mathematics or Grades 9-12 Mathematics
- Junior High School (Grades 9-10 only) or High School – Mathematics
- Junior High School (Grades 9-10 only) or High School – Mathematical Science, Composite
- Legacy Master Mathematics Teachers (Grades 8-12)
- Mathematics: Grades 7-12
- Mathematics: Grades 8-12
- Mathematics/Physical Science/Engineering: Grades 6-12
- Mathematics/Physical Science/Engineering: Grades 8-12
- Physics/Mathematics: Grades 7-12
- Physics/Mathematics: Grades 8-12
- Secondary Mathematics (Grades 6-12)

Recommended: Experience teaching an introductory statistics course such as AP Statistics or Statistics

Additional information:

Not applicable