Miami Dade College

Course Description

CAP3330 | Programming R for Statistics | 4.00 credits

This upper division course is for students majoring in data analytics. Students will learn the R programming language and use it to perform intermediate-level statistical analysis. Techniques used in data analysis, such as analysis of variance and regression, will be emphasized. Prerequisite: STA2023.

Course Competencies:

Competency 1: The student will perform statistical computations using the R programming language by:

1. Creating a vector and understanding the difference between a vector, a list, and a matrix
2. Computing different operations that involve manipulation of matrices, vectors, and lists
3. Handling bivariate data (categorical vs numerical)
4. Managing data frames

Competency 2: The student will list and describe data characteristics (focusing on exploratory data analysis) using the R programming language by:

1. Interpreting and assessing data displayed using visual graphic presentation methods
2. Interpreting and analyzing the five-number summary
3. Interpreting and analyzing box plots plot diagrams
4. Identifying the most convenient graph to display a certain data set

Competency 3: The student will compute and interpret measures of central tendency and variance using the R programming language by:

1. Comparing and contrasting the common standard methods for gathering sample data
2. Determining the value of the mean, median, and the mode of both grouped and ungrouped data
3. Identifying the relationships among the three measures of central tendency for symmetrical and skewed distributions
4. Understanding the advantages and disadvantages of the three measures

Competency 4: The student will choose, compute, and interpret statistical tests (parametric and non-parametric) using the R programming language by:

1. Interpreting the output obtained from a statistical software package applied to tests of variances
2. Obtaining a confidence interval for the ratio of two population standard deviations when the variable under consideration is normally usually distributed on both populations
3. Solving problems applying different statistical tests for means and proportions
4. Setting up and solving problems by applying statistical tests for means and proportions with two samples of data
5. Differentiating between parametric and nonparametric statistics
6. Performing hypothesis testing using different nonparametric methods, such as the Sign Test, Wilcoxon Signed-Rank Test, and Mann-Whitney Test

Competency 5: The student will apply, compute, and interpret analysis of variance and analysis of covariance using the R programming language by:

1. Stating the hypothesis and assumptions for a one-way analysis of variance (ANOVA)
2. Conducting a one-way and a two-way ANOVA test
3. Interpreting the results from a one-way and a two-way ANOVA test
4. Stating the hypothesis for a one-way and two-way ANOVA
5. Conducting a two-way ANOVA test by testing for an interaction between the two factors and, if necessary, testing for the effect from the row factor and the column factor
6. Interpreting the results from the ANOVA test
7. Comparing the relationships between quantitative variables and at least one categorical variable

Competency 6: The student will develop, compute, and interpret appropriate regression models using the R programming language by:

1. Comparing and contrasting the assumptions of the linear regression model
2. Measuring the y-intercept and the slope of a line
3. Evaluating and interpreting the coefficient of determination
4. Designing a test for the significance of a correlation
5. Designing an F-test
6. Evaluating and interpreting residuals

Course Competency 7: The student will perform the model selection process by:

1. Selecting the appropriate method for using R software to conduct a multiple regression analysis
2. Evaluating the use of linear and multiple regression, describing its limitations, assumptions, and error measures
3. Understanding the difference between time series and linear regression models
4. Identifying and understanding non-linear trends

Learning Outcomes:

- Use quantitative analytical skills to evaluate and process numerical data
- Solve problems using critical and creative thinking and scientific reasoning
- Use computer and emerging technologies effectively