This is a preview of this best practice. The full document is restricted and is available only to those with a .mil email address.

To request this best practice, contact the STAT COE at COE@afit.edu using your .mil email account.

Using Regression Trees

Authored by: Francisco Ortiz, PhD

Updated: 29 August 2017

Revised 1 November 2018

The goal of the STAT COE is to assist in developing rigorous, defensible test strategies to more effectively quantify and characterize system performance and provide information that reduces risk. This and other COE products are available at www.afit.edu/STAT.

STAT Center of Excellence
2950 Hobson Way – Wright-Patterson AFB, OH 45433

Executive Summary

This paper introduces regression trees and provides a tutorial on how to build them using JMP, MATLAB, and R statistical software. The tutorial uses a digital simulation example to demonstrate how regression trees can be used to help plan for costlier follow-on tests.

Keywords: Regression Trees, Decision Trees, Classification Trees, Data Mining, Simulation, Prediction, Modeling

Introduction

Data is usually a rare and expensive commodity to obtain in the test and evaluation (T&E) community. However, digital simulations that are computationally inexpensive can produce an extensive database of information. Digital simulations tend to explore rather large test spaces where the response (i.e., performance measure) of interest can exhibit very complex nonlinear behavior. Due to this, traditional regression analysis may fail to provide adequate predictions and fail to identify where certain inputs affect the response. When large amounts of data are available, data mining approaches such as regression and decision trees become a viable option. Regression and decision trees can yield accurate prediction models that are easy to interpret. This best practice serves as a brief introduction to regression trees and provides a tutorial on how to create them using JMP, MATLAB, and R statistical software. This paper demonstrates how to use regression trees to model digital simulation results and how they can help plan for costlier hardware-in-the-loop and field tests.

This is a preview of this best practice. The full document is restricted and is available only to those with a .mil email address.

To request this best practice, contact the STAT COE at COE@afit.edu using your .mil email account.