



**Biostatistics, Epidemiology, and Research Design Key Component Activity**

***OSCTR Biostatistics, Epidemiology and Research Design Core***

**WORKSHOP**

**Introduction to Classification and Regression Trees (CART) and Random Forests**

**Justin D. Dvorak, Ph.D.**

Senior Research Biostatistician

Department of Biostatistics & Epidemiology

The University of Oklahoma Health Sciences

**Wednesday, March 5th, 2025**

**12:00 PM – 2:00 PM**

|  |  |  |
| --- | --- | --- |
| **Physical:** | |  | | --- | | Hudson College of Public Health Auditorium (CHB 320) | |
| **Virtual:** | |  |  |  | | --- | --- | --- | | [Zoom](https://oklahoma.zoom.us/j/95569185971?pwd=MAkNzeq0ykUgTmnzyvpDoBr713ojau.1) | **Meeting ID:** | 955 6918 5971 | | **Passcode:** | 30139218 | |

**Box lunches will be provided for the first 15 attendees.**

[**Registration**](https://osctr.ouhsc.edu/short-course) **required in advance for this meeting.**

After registration, you will receive a confirmation email containing the Zoom link and information about the workshop materials.

**DATE:** March 5th, 2025

**TIME:** 12:00 PM – 2:00 PM

**LOCATION:** Hudson College of Public Health Auditorium (CHB 320)

**FORMAT:** Lecture with in-class, hands-on practice exercises

**SOFTWARE:**

* Prior to the workshop, please install the following software on your laptop.
  + R (<https://www.r-project.org/>)
  + RStudio Desktop: (<https://posit.co/download/rstudio-desktop/>)
* Please bring your laptop to the workshop so you can complete the in-class exercises.

**PREREQUISITES:**

* Basic R programming (variables, functions, loops, etc.)
* Knowledge of statistical methods such as linear regression and logistic regression.

**DESCRIPTION:**

* Decision trees are commonly used in many areas including patient care and predictive modeling.
* CART can construct decision trees for categorical or continuous predictors and outcomes.
* Random forests address overfitting issues in CART by generating multiple decision trees and constructing an ensemble prediction (average value for regression trees; majority vote for classification trees).
* In this workshop, we will discuss some commonly used R packages to generate decision trees and fit random forests to different types of data.

**WORKSHOP CONTENT:**

1. Introduction to decision trees.
2. Introduction to supervised-learning classifiers and performance metrics.
3. Classification and regression trees.
4. Validation methods for machine-learning algorithms (e.g., 80/20 split, k-fold, leave-one-out).
5. Random forests.

**COURSE MATERIALS:**

* Lecture slides, example code, and practice datasets may be downloaded prior to attendance, printed, and saved for personal use.
* Access materials at <https://osctr.ouhsc.edu/short-course>.

**END OF WORKSHOP EVALUATION SURVEY:**

* Please complete the survey at the following link: <https://bbmc.ouhsc.edu/redcap/surveys/?s=W3RXNL99M748F3RM>
* You will also receive the link by email after the workshop.

**REGISTRATION:**

* Registration is required by 6:00 PM on March 4th.
* Registration can be completed at <https://osctr.ouhsc.edu/short-course>.

**SPONSOR ACKNOWLEDGEMENT:**

* Sponsored by the Clinical Epidemiology Unit of the Biostatistics and Epidemiology Research Design (BERD) Core of the Oklahoma Shared Clinical and Translational Resources (OSCTR)
* National Institutes of Health, National Institute of General Medical Sciences Grant U54GM104938

**PARKING:**

* The College of Public Health Building is located on the northeast corner of 13th Street and Phillips Avenue. Parking is available on the north side of the building.
* If you are driving north on Phillips Avenue, you will see a sign for **Lot 14E** on the east side of the street. Pull into this lot.
* If the gate is down, please press the button on the speaker box and indicate that you are attending a workshop in the College of Public Health Building.

**FACULTY BIOGRAPHICAL SUMMARY:**

* Dr. Dvorak is a Senior Research Biostatistician in the Department of Biostatistics and Epidemiology at the Hudson College of Public Health. He has over 10 years of experience in data management, statistical consulting, model-building, and automation. He has worked with investigators in various content areas including ophthalmology, endocrinology, oncology, orthopedic surgery, cardiology, internal medicine, tobacco, and speech-language pathology. He has taught undergraduate and graduate courses in statistics and programming, and has delivered numerous seminars and presentations at the national and international levels.