

SMART Program

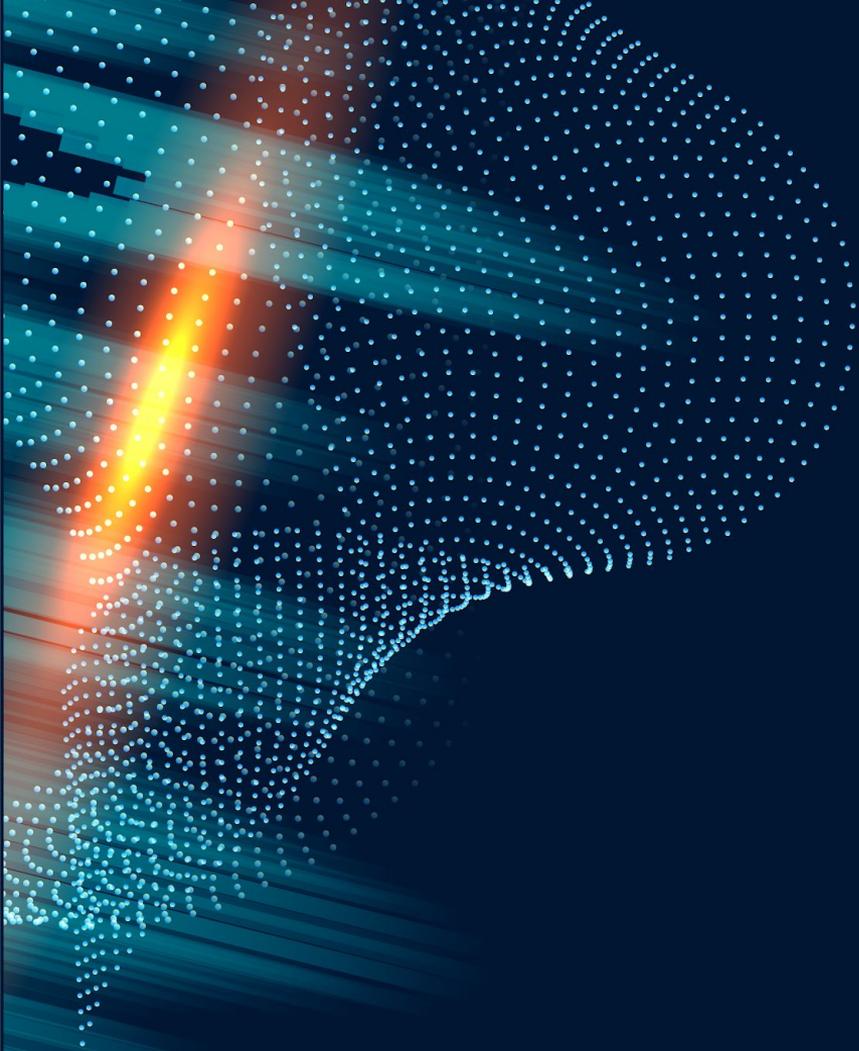
Statistical **M**entoring in **A**pplication,
Research, and **T**echnology

Department of Statistics

Winter 2024 Info Session

Emily Ouyang
eouya001@ucr.edu

Dr. Analisa Flores
analisa.flores@ucr.edu



The mission of the SMART program is to provide undergraduate students with the opportunity to conduct statistical projects under the supervision of graduate students and to provide graduate students with an opportunity to develop projects and mentor students.

01 Who can participate?

MENTORS

- STAT graduate students (MS/PhD)

FELLOWS

- Undergraduate students majoring or minoring in STAT or DS; active HiSS members.
- Interested in attending graduate school.
- Must be able to commit to an average of 5 hours per week for one quarter.
- Can only participate in the program once per year.

02 What do participants do?

MENTORS

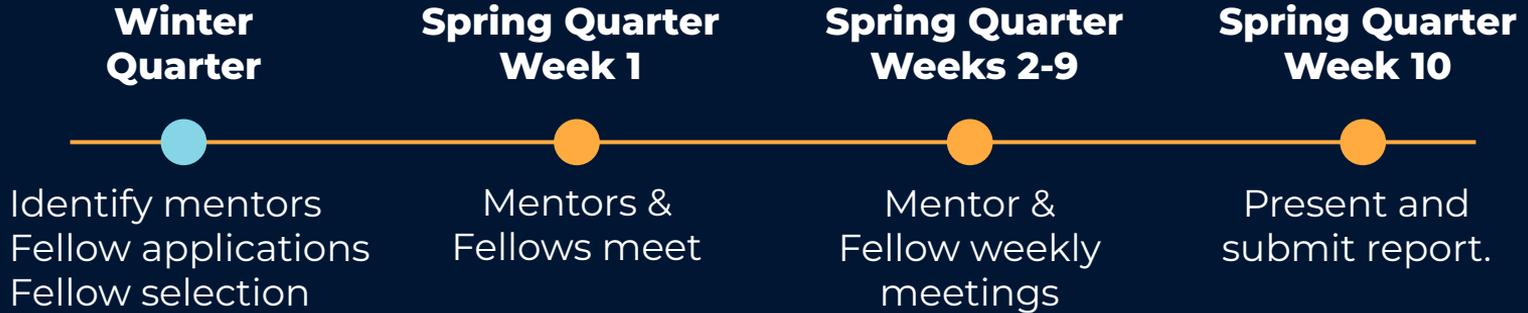
- Create a project that an undergraduate student can complete within 6-8 weeks.
- Set prerequisites.
- Gather teaching materials.
- Mentor an undergraduate student for the duration of the quarter.
- Meet with an undergraduate student for 1-2 hours a week.
- Monitor student progress and adjust the project as needed.

02 What do participants do?

FELLOWS

- Work on a project under the guidance of a graduate student.
- Spend an average of 5 hours per week on the project.
- Meet with a graduate student mentor for 1-2 hours per week.
- Develop a 1-page report and 10 minute presentation for the end of the quarter.
- Present their work at a department seminar.

TIMELINE OVERVIEW



03 Why is this program valuable?

MENTORS

- Leadership Opportunities
- Mentoring Experience
- Collaboration
- Communication
- **Community Building**

FELLOWS

- Project experience
- One-on-one guidance
- Exposure to topics outside of curriculum
- Communication: one-on-one and presenting
- Informally learn about grad school experiences and process
- **Community Building**

7 Mentors



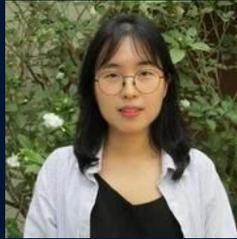
Brian



Emily



Jericho



Nahee



Nayla



Namhwa



Noe

Project Overview: Brian Neal

Topic: Statistical Data Mining w/ R



Description: Statistical learning and data mining involves the use of statistical models/methods to learn from data, with potential objectives being prediction (predict quantities/characteristics), inference (understand/quantify relationships), or clustering (identify meaningful subgroups within the data).

For this project, you will choose a data set of interest to you, determine a relevant question/objective, and identify one or more suitable statistical learning method(s). You will learn about the method(s) and apply it/them to your data in R. Some examples of methods include linear/polynomial/local/logistic regression, SVM, PCA, splines, LDA/QDA, KNN, decision trees, and complementary methods such as variable/model selection or regularization. The choice of method(s) will depend on the data, objective and student's background/interests.

In addition to presenting the method(s) and your results, you will also have the opportunity, if desired (and time permitting), to develop your R code in a usable, presentable and redistributable format that can be hosted on GitHub and linked directly on your resume/CV to showcase your statistics/programming experience. This could be an R package, an RMarkdown document, or an R script.

Prerequisites:

- STAT 160
- STAT 107 (or experience in R)

Project Overview: Brian Neal



Goals / Topics:

- 1) Learn conceptually about one or more statistical learning methods
- 2) Learn to apply the method(s) to relevant data in R
- 3) Develop a presentation to explain the methods to others and present/interpret your results
- 4) If desired, create a deliverable to showcase project/skills in future applications/resume/CV
 - E.g. R script implementation, R package, RMarkdown

Project Overview: Emily Ouyang

Topic: Data Visualization

Description: Data visualization is an important skill due to its ability to communicate results and highlighting key points in a comprehensive manner. Students will learn about common tools for data visualization tools (including different R packages and Tableau) and different methods of data visualization. They will then create a project to highlight the skills gained.

Prerequisites:

- None!



Project Overview: Emily Ouyang



Goals / Topics:

- 1) Become familiar with common data visualization tools
- 2) Learn common data visualization methods
- 3) Apply the techniques to create a research project that focuses on clear and concise communication and visuals

Project Overview: Jericho Lawson



Topic: Forecasting, Regression, & Sampling Techniques for Sports Analytics

Description: The student will learn about how data is utilized in sports to develop better strategies and analyze player/team performance. Specifically, they will utilize bootstrapping, advanced regression techniques (such as logistic regression and generalized linear models), and some form of forecasting based on data collected from various sports. A background into sports analytics and a development of coding and soft skills will also be emphasized during the project.

Prerequisites:

- STAT 107
- STAT 170

Project Overview: Jericho Lawson



Goals / Topics:

- 1) Develop an understanding of various statistical techniques in sports analytics
- 2) Carry out a project that utilizes these techniques w/ data related to sports
- 3) Practice coding and presentation skills to clearly communicate work being done

Project Overview: Nahee Kim



Topic: Regularization: Ridge and Lasso

Description: Ridge and Lasso estimator are the ones commonly used in many fields. This project will have undergraduate students learn how Ridge and Lasso work and their application with simulation (or real data if time allows). A related chapter is Chapter 6 from ISLR textbook.

Prerequisites:

- STAT 170

Project Overview: Nahee Kim



Goals / Topics:

- 1) Develop an understanding of ridge and LASSO estimators
- 2) Learn how to use glmnet, an R package
- 3) Develop an understanding of tuning parameter selections (optional)

Project Overview: Namhwa Lee



Topic: Statistical Data Mining w/ R

Description: In statistical data analysis, prediction and inference are both based on data, but they have different purposes. The primary focus in inference is on interpreting the model we fit. However, the primary goal of prediction model is to accurately predict new or future data based on a given or past dataset.

Throughout the program, we will learn how to employ statistical data mining methods (e.g. regression, and tree-based techniques) using R. Additionally, we will learn how to evaluate or assess the model performance based on its prediction error using data splitting method. Finally, if time allows, we will do a small data analysis project to apply what we have learned during the program.

Prerequisites:

- STAT 107
- STAT 170

Project Overview: Namhwa Lee



Goals / Topics:

- 1) Distinguish between inference and prediction
- 2) Application of regression and tree-based methods
- 3) Model assessment and selection for prediction models

Project Overview: **Nayla Liu**



Topic: Regression Basics

Description: Regression is a foundational topic in statistics; however, it can be heavily built upon. Starting from linear regression, the student will work their way up to more complex models such as including interactions and discuss more quadratic, polynomial, and logistic regression (as well as important assumptions related to these models). The student will then utilize these tools on a data set of their choosing.

Prerequisites:

- STAT 107 (or programming experience)

Project Overview: **Nayla Liu**



Goals / Topics:

- 1) Learn when to use different types of regression
- 2) Utilize different model selection techniques
- 3) Learn how to interpret estimated coefficients

Project Overview: Noe Vidales

Topic: Diversity within the Statistics Major



Description: Learn the factors that have contributed to underrepresentation of minorities within the Statistics discipline. Specifically, why it has been so difficult for Latinos and African Americans to gain representation within this discipline and the barriers of entry that may be the main culprits. If you are interested in developing a better understanding as to what has caused this phenomenon, join me in reviewing the statistical literature for answers.

Prerequisites:

- STAT 169
- STAT 170
- STAT 171

Project Overview: Noe Vidales



Goals / Topics:

- 1) Develop skills in applying statistics to real world problems
- 2) Gain an understanding of problems that plague minorities
- 3) Learn to read statistic literature as applied to the social science and prepare a concise 10 minute presentation to relay findings

APPLICATION PROCESS

Winter 2024

Thursday (Week 6)
2/15/2024

Information
Session
&
Application
Opens

Friday (Week 9)
3/8/2024

Applications
Close

Applications
Reviewed

Friday (Finals)
3/22/2024

Fellows
Notified

APPLICATION Materials

- 1) Courses taken and R experience
- 2) What are your career goals? (Word Limit: 300)
- 3) How would you benefit in participating in the SMART Program? (Word Limit: 300)
- 4) Do you plan to attend graduate school?
- 5) Upload a 1-2 page Resume (Submit a PDF)
- 6) Project Interest Ranking

APPLICATION

<https://forms.gle/sMv4yQ4gjppu7TUB6>



THANK YOU!

Emily Ouyang (eouya001@ucr.edu)

Dr. Analisa Flores (analisa.flores@ucr.edu)