

# Developing Interactive Statistics Apps

*A look at the project and student experience*

Maria Tackett, Sean Li, Emmanuel Mokel, Shari Tian  
Duke University

USCOTS 2021

---

# Introduction

---

- **Overall project goal:** Use R Shiny to develop interactive learning apps on statistical concepts in 1st year courses
- Motivated by the Book of Apps for Statistics Teaching (BOAST) led by Dennis Pearl at Penn State
- Aims to *expand opportunities*:
  - ✓ Open to students at all levels (intro statistics course or equivalent required)
  - ✓ Positions posted on Muser, a university-wide platform
  - ✓ Opportunity to earn pay or course credit

# Program structure

---

## Summer 2020 (6 weeks)

- Team:
  - 4 undergrad students
  - 1 faculty member
  - 1 project manager
- ~ 15 hours per week
- 3 meetings per week
- Focused on app development

## Spring 2021 (12 weeks)

- Team:
  - 3 undergrad students
  - 1 faculty member
- ~ 6 - 8 hours per week
- 1 meeting per week
- Focused on preparing apps for classroom use

# Weekly workflow

---

## Team meeting

- Project updates
- Feedback and idea sharing
- Plan for upcoming week

## Independent work

- Implement updates
- Use R documentation and online resources for ideas and to learn new skills

## Communication

- Google docs for meeting notes and resources
- Slack for communication and idea sharing between meetings

# Learning outcomes

---

- ✓ **Acquired new skills** seldom covered through regular statistics coursework
  - Interactive app and web development
  - Effectively explain and teach complex statistical concepts
- ✓ **Improved understanding of statistical concepts and R** programming
- ✓ Learned how to **give effective and constructive feedback** and reflect on own work
- ✓ Encouraged to take on **future data-driven projects and research opportunities**

# Next steps & conclusion

- Conduct a formal research study in upcoming academic year to assess effectiveness of apps in a classroom setting
- Find the apps at <https://duke-shiny-ed.netlify.app/>

Fit the Model | Model Diagnostics | Approaches to deal with outliers

Click on the scatterplot to add new points. As you add points, the model will be refit using the a data set that includes the original data and the newly added points.

Notice how the model changes as new points are added to the data set.

- How does intercept change?
- How does the estimated coefficient of  $M\text{edi}$  change?
- How does the confidence interval for the coefficient of  $M\text{edi}$  change?

REMOVE ALL NEW POINTS

**Business Value Added vs. Median Income**

**Model fit using original data**

Term	Estimate	Std. Error	Statistic	P-value	Conf. Low	Conf. High
(Intercept)	-1823.59	1061.106	-1.719	0.098	-4008.978	361.799
medi	1.01	0.169	5.988	0.000	0.663	1.357

**Model fit using original AND new data**

Term	Estimate	Std. Error	Statistic	P-value	Conf. Low	Conf. High
(Intercept)	-1890.133	1234.048	-1.532	0.133	-4378.829	598.563
medi	1.148	0.148	7.755	0.000	0.850	1.447

Logistic Regression Curve

False Positive Rate (1 - Specificity)

How many observations were correctly/incorrectly predicted by the model. Click the buttons to see how each value in the table below are calculated.

REMOVE ALL NEW POINTS

**What is the F-statistic?**

If the sample means vary around the overall mean more than the individual observations vary around their sample means, we have evidence that the corresponding population means are different. We formally compare these variances with the F-statistic. Relating back to Step 1, we see that increasing the between-group variance and decreasing the within-group variance increases the F-stat, while doing the opposite decreases the F-stat.

If there is **no treatment effect** the F-stat will be very close to 1

$$F = \frac{s_b^2 / df}{s_w^2 / df}$$

- $s_b^2$  is the between groups variance \*
- $s_w^2$  is a pooled estimate of the within groups variance \*
- $df_b$  is the numerator degrees of freedom
- $df_w$  is the denominator degrees of freedom

The F-stat for our data is:

(1) 189.1498

**Sample Distributions**

Term	Degrees Freedom	Sum of Squares	Mean Square	F-Stat	P value
dataset	2	4.206	2.103	189.15	0
Residuals	297	3.302	0.011	NA	NA

	Actual Prediction	Negative Prediction
default		
0	2,799	18
1	82	45

	Negative Model Prediction	Positive Model Prediction
Actual Negative	True Negative: Model correctly predicts the negative	False Positive (Type 1 Error): Model incorrectly predicts the negative as a positive
Actual Positive	False Negative (Type 2 Error): Model incorrectly predicts the positive as a negative	True Positive: Model correctly predicts the positive

Thank you!

# Resources

---

- Project website: <https://duke-shiny-ed.netlify.app/>
- [Reflection on Summer 2020 program](#)
- Related Projects:
  - [Book Of Apps for Statistics Teaching \(BOAST\)](#)
  - [Shiny Ed](#)





What motivated  
you to get  
involved with the  
project?

*“A primary motivator to get involved in a statistics project outside of class was the curiosity of applying learned technical knowledge to help improve other students’ education in statistics. Furthermore, with online classes, I was looking for ways to connect more with my peers and professors.” - Sean*

*“I took an introductory data science course which was my first time using R. I wanted to expand my programming experience, and I found the concepts in the course extremely interesting. “ - Emmanuel*

*“Spring 2020, I took an Introduction to Data Science course and learned a significant amount of R programming. Toward the end of the course, we were introduced to the R Shiny package. While we did not get the chance to use it extensively, the introduction motivated me to take advantage of its app-building capabilities to create something meaningful.” - Shari*



How do you plan to get involved in future statistics projects?

*“This summer, I am conducting data-driven urobiome research at Loyola University Chicago while simultaneously working on a machine learning project for a microbiology lab at Duke. In the future, I plan on continuing to find computational research opportunities that pertain to health and medicine, whether through individual projects or working in a lab.” - Shari*

*“At the moment I am working on a project through Duke which focuses on using wearables data to predict COVID-19. Its primarily data engineering work which is new to me, but the premise of the object is interesting. In the fall, I will be working in a statistical genetics lab which focuses on predicting ancestry of admixed individuals. “ - Emmanuel*

*“This summer, I am working on a machine learning research project targeting to identify and remove bias and discrimination. In the future, I am pursuing data science research to improve educational outcomes.” - Sean*