MA388: Sabermetrics
AY24-2 Course Project: SCORE Module

April 23, 2024

1 Introduction

The SCORE Network describes itself as "a national network for developing and disseminating Sports Content for Outreach, Research, and Education in data science." The Handbook for Authors and Creators of SCORE Network Modules states the purpose of SCORE modules: "The main goal of the SCORE Network is the creation of a curated repository of education materials for teaching statistics and data science." Your course project is to construct such a module designed to teach statistics or data science using a sports application.

The expected organization of a SCORE module includes the following sections as outlined in the Handbook for Authors and Creators of SCORE Network Modules.

1. Learning Goals: A list of the learning outcomes for those who complete the module.
2. Introduction: Motivates the problem to be solved and provides high-level context on how it is to be solved (e.g., can we predict whether a batted ball is going to be a home run based on launch angle and launch speed using a classification tree model?). The accompanying introductory video would also be part of this section.
3. Data: Section in which data are provided and described.
4. Methods / Instructional Content: Section(s) that provides background material on the methods or main instructional content of the module (e.g., a high-level description of what a classification tree is and how it is learned).
5. Exercises/Activities: Section that provides exercises that students are to complete; if an exercise involves data analysis, that analysis is to be done using the provided dataset. Note: Exercises could be incorporated throughout the module rather than in a separate section.
6. Wrap-Up/Conclusions: A conclusion (perhaps with an accompanying video) that summarizes the information that students are to take away when completing the module.

The course project requirements will include a module proposal, an interim progress review, a peer review (providing feedback to peers about their modules), an oral presentation, and the submission of the final module products.

2 Module Proposal

Your SCORE module proposal will outline your plan for each of the required sections of a SCORE module. To get started, you'll need to decide on a statistics topic you want to teach as well as a relevant sports data set you can use to teach it. Explore the SCORE module repository for examples of topics and methods of delivery others have chosen. Take note of what you like and what you dislike. Ultimately, a great SCORE module is one that hooks the reader with an interesting application and convinces the reader they *need* to know how to do what you're offering to teach them how to do. This will likely be easiest if you're teaching a method you find particularly helpful/powerful and/or answering a question you find especially interesting.

Cadets can excel in this project in a number of ways: impressively tackling a hard topic, excellent delivery of presentation and written materials, creating a compelling context for the reader to learn the topic, etc. The intended audience for your module should be a college student enrolled in an introductory undergraduate statistics course.

Submit your module proposal as a PDF with the following sections (start in R Markdown):

1. Learning Goals: A list of the learning outcomes for those who complete the module.
2. Introduction: What is an interesting question that your statistics concept or method can help us answer? Describe the question or problem you will use to motivate the reader to learn about your topic.
3. Data: Describe the data and its origin. Provide a sample and summary of the data that demonstrates the variables contained, the range of values included in the data set, the sample size, etc.
4. Methods / Instructional Content: List two or more scholarly references (e.g., textbooks – print or online) that you can use to learn about the topic and see how it is normally presented.
5. Exercises/Activities: List any incremental or related exercises you think could be helpful in practicing the targeted skills for your module.
6. Wrap-Up/Conclusions: No requirement for the module proposal.

3 Interim Progress Review

You should have a complete draft of your SCORE module for this IPR. Submit a PDF version of your module with the following sections:

1. Learning Goals: Provide the reader a list of the learning outcomes for students who complete the module.
2. Introduction: Provide the reader a complete introduction to the module and convey both the topical coverage and the sports question(s) addressed by the module.
3. Data: Present a complete introduction of the data set to the reader. What does the data contain? How and where did you get it? Is there anything special the reader should know about this data set? Provide a sample and summary of the data that demonstrates the variables contained, the range of values included in the data set, the sample size, etc.
4. Methods / Instructional Content: Using the references identified in your project proposal (or others as needed), present your topic with appropriate attention given to mathematical notation, assumptions, interpretation, etc. As an example, teaching linear regression would require introducing the standard formulation ($y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \epsilon$), some discussion of how we model the error term ($\epsilon \sim N(0, \sigma)$), assumptions and how to check them (L-I-N-E), etc. You should then demonstrate how to do this in thoroughly commented code using straightforward variable names. Bottom line: this module seeks to teach the reader what your technique is and how to do it.
5. Exercises/Activities: Present the exercises you think could be helpful in acquiring/practicing the targeted skills for your module. Ask questions and provide the answers. Think of this module as being similar to a set of lesson notes or a problem set with the solutions provided.
6. Wrap-Up/Conclusions: Recap the lessons learned both in terms of statistical techniques and in terms of the sports research question. Provide the reader at least one other sports application (could be the same sport) for this particular skill and at least one idea for a future skill to learn that builds on what you've presented.

4 Peer Reviews

You will be provided a Word document template to fill out and (1) submit on Canvas and (2) email to the project group members whose work you are reviewing. Listed below are the sections / questions you'll be addressing.

4.1 Purpose

What is the statistical topic covered by this module? What is the sports context used by the authors to teach this statistical topic?

4.2 Learning Goals

Are the learning goals clear? Do you know what to expect from this module? How can they be clearer?

4.3 Introduction

Would the average reader be able to gain the necessary context about the statistical technique and sports application? How can it be clearer to better set up the reader for what's coming next?

4.4 Data

Do you understand what data was used, where it came from, what cleaning / preprocessing may have been done, etc.? Do the authors sufficiently introduce the different variables, typical values, etc.? How can it be improved?

4.5 Methods / Instructional Content

Is the explanation of the technique sufficiently rigorous? Do the authors present a standard formulation of the technique with appropriate mathematical notation? Can you imagine this topic being presented similarly in a textbook? Are important details about assumptions, interpretations, limitations, etc. presented? Does the module clearly demonstrate how to implement the technique? Is the code clear, appropriately commented, and as intuitive as possible by using sensible variable names, etc.?

4.6 Exercises / Activities

Are there interim questions or other activities provided to guide the reader toward a fuller understanding of the topic? Do the authors provide the solutions for each question / exercise?

4.7 Wrap-Up / Conclusions

Do the authors revisit the learning goals and recap the major takeaways for the technique presented and the sports question addressed? Do the authors present a future direction to apply this skill (some other context) and a related concept or skill that might be helpful to explore or learn next?