BasketStats: An Educational Program Leveraging a Passion for Basketball Towards an Entry Point to Statistical & Computational Literacy

Christopher T. Boughter, Ph.D. – Postdoctoral Scholar, National Institutes of Health - 12/07/21

Brief Program Summary: The *BasketStats* program aims to introduce fundamental concepts in mathematics, science, and computer programming via a curriculum centered around basketball statistics. Students will be enticed to join through the promise of time in an indoor basketball court to play games, with a second session held in the same week focused on the subsequent analysis of the students' personal statistics and the statistics of their favorite NBA or WNBA players.

Target Audience: With its unique approach to teaching quantitative skills, *BasketStats* was specifically created to change the attitudes of students who think they "aren't good at math" or "don't like science" alongside the educational content presented each week. We hope to educate students aged 12 or above, old enough to both understand some of the most fundamental concepts in the class and have a sincere passion for basketball.

Resource Requirements: This course requires access to an hour or more of dedicated time on a full-sized basketball court with two hoops for the first session, and a computer for each student for the second session. If computers have IT restrictions, a suite of programming and analysis tools are required including Anaconda, Jupyter, Microsoft Excel, VScode, and Python.

Program Objectives: Lessons will focus on making connections between concepts the students will likely already be familiar with (i.e. points per game, field goal percentage, true shooting percentage, player efficiency rating, etc.) to concepts in mathematics (averages, fractions, weighted averages, normalization). The goal is for students to finish the program with a newfound appreciation for math, a proficiency in computer programming, and an understanding that a career in data analysis, especially in sports, is a viable and potentially lucrative career path.

Detailed Description: After-school robotics, computer programming, and science programs are incredibly enriching experiences for the students that participate, teaching key concepts that can improve critical thinking, develop financial literacy, and cultivate hireable skills. However, many existing STEM after-school programs self-select for students already interested in the sciences. Students who consider themselves bad at math or science may miss out on these opportunities due to disinterest or a lack of confidence in their abilities. *BasketStats* intends to expand the intended audience of STEM programs through an immersive foray into the use of mathematical tools and computer programming to analyze basketball statistics. Leveraging an existing interest in basketball, we aim to instill confidence in the quantitative abilities of students and generate new skills in mathematics and computer programming.

One of the fundamental strengths of *BasketStats* is the incorporation of physical activity and a direct link from this physical activity to the educational programming. One of the meetings for this after-school program will be dedicated solely to playing pickup basketball, an added incentive for students who otherwise might not want to learn statistics and computer programming to get involved. Their in-game stats will be tracked during these sessions, and these are the datasets (henceforth referred to as *personal statistics*) that will be used for the subsequent lessons in each class meeting. Personal statistics will then be analyzed alongside the statistics of NBA and WNBA players for fun comparisons between students and some of their favorite stars. Students will discover if they score as many points as LeBron James, grab as many rebounds as Joel Embiid, or dish out as many assists as Sue Bird, helping to connect basketball statistics the students are very familiar with to the mathematical concepts behind them.

In addition to the direct benefit this program will have to the students in Prince George's County, it also has significant potential to be adopted by schools or community centers across the country. Course materials will be posted online and shared on GitHub and social media. The course can effectively be run anywhere with available computers and basketball courts. By running the pilot version of this course, potential gaps or shortcomings in the course material will be addressed before other organizations attempt to adopt the curriculum. Furthermore, similar courses could be conceivably developed for baseball, football, soccer, and hockey. However, basketball is the most conceptually straightforward to create a curriculum for, primarily because it is a widely played sport with the simplest metrics for performance (i.e. rebounds, points, blocks, steals, etc.).

The curriculum for this program will progress along a set track of computer code and written text developed in Python using the open-sourced Jupyter Notebook. In each subsequent class, students will go from learning about basic statistics and computer programming to more advanced concepts in statistics. Lessons will involve a mix of analyzing the student's personal statistics and analyzing data from Basketball Reference, an online repository of statistics from current and past NBA and WNBA players. The course will run for four months, but the potential exists for a second half of the course, focused on data visualization and the reinforcement of concepts introduced in the first half of the course. **An abridged version of the syllabus is outlined below**:

The Basics: Computer Programming and Statistics (Weeks 1-4): To start the course, students will be introduced to the power of statistics through a conversation they are very familiar with; the debate over who is the greatest basketball player of all time. This will serve as the basic motivation for the course, specifically how simple averages cannot tell the whole story. Basic basketball statistics will be the central focus (i.e. rebounds, assists, points, etc. *per game*) for this section.

Student Objectives:

- *Build personal statistics through gym sessions for later use in the course*
- *Learn to navigate the pre-programmed Jupyter Notebook*
- *Become comfortable with fundamental statistical concepts (the concept of a "dataset", averages, standard deviation, using statistics to make data-based arguments)*
- *Create the "Course Leaders" page, a leaderboard page to track the statistics of students in the class, to be updated weekly. Participants will be required to update their own stats each week as homework, which will necessitate repeated review of the most fundamental concepts.*

Making Reasonable Comparisons: The Use of Extrapolation and Weighted Averages (Weeks 5-8): Once statistical foundations are sufficient, students will learn the more advanced concepts of weighted averages and extrapolation. By this time the students should have a working dataset of personal statistics. However, shorter games in the class make it difficult to compare personal statistics to NBA and WNBA players who play much longer games. For these lessons, students will extrapolate their own statistics and use Basketball Reference to find the professional players who play a similar style of basketball.

Student Objectives:

- *Understand the use of Per 36 minutes and Per 100 Possessions in basketball statistics and the general concept of extrapolation.*

- *Compare two all-time shooters (e.g. Steph Curry vs. Larry Bird) and use weighted averages (through Effective Field Goal % and True Shooting %) to determine who is better, crafting short arguments to discuss in class.*
- *Use extrapolated personal statistics to compare and contrast student playing style to that of NBA players.*

Can We Boil It Down to One Metric: Normalization (Weeks 9-12): While all of the statistics we have utilized in the course are pretty comprehensive, they are quite hard to compare (e.g. Jayson Tatum scores a lot, but Ben Simmons gets more rebounds and assists). For this portion of the course we will learn about normalization through the Player Efficiency Rating (PER), a metric that attempts to normalize the NBA to the most "average" player, and use this simple scoring system to determine the best and worst players in the NBA.

Student Objectives:

- *Calculate the PER of NBA players and use personal extrapolated statistics to determine student PER, using a provided formula*
- *Understand PER and normalization sufficiently to think critically about it and discuss whether it is an adequate metric for determining the "best" player.*
- *Propose alternative approaches to determining the best basketball players.*

Settling it Once and for All: Who is the GOAT? (Weeks 13-16): The course will culminate in the students presenting to the class and revisiting the argument "Who is the GOAT (Greatest Of All Time)" in both the class and the NBA. This will be a lighthearted, data-based exercise where the students can demonstrate their computer programming and statistical skills, their presentation skills, and their ability to (presumably) argue why they're the best basketball player in the class.

Student Objectives:

- *Use Jupyter Notebook and Basketball reference to craft an argument for crowning the greatest player in the course or the greatest player in the NBA. Presentations need not all focus on the same metrics but MUST be quantitative.*
- *Be able to think critically about basic and advanced statistical metrics and their use for comparing NBA players.*
- *Understand that a career focused on the use of numbers (in mathematics, engineering, or science) does not necessitate a life stuck in a lab. Quantitative thinkers who are able to use statistics to make arguments are hired everywhere, even by the NBA, NFL, etc.*

The Jupyter notebook being used in the course is currently under development by Christopher Boughter and can be found on GitHub: https://github.com/ctboughter/basketstats