UNIVERSITY of CALIFORNIA Riverside

STATISTICAL CONSULTING COLLABORATORY @ UCR

SCCUCR – Southern California Conference on Undergraduate Research
Riverside, California - November 19, 2005

Streaky Baseball: A Statistical, Interactive, Online Multimedia Tool Designed for Self-Learning

Raymond Holguin

Department of Computer Science & Engineering, Bourns College of Engineering, University of California, Riverside, CA 92521

Mentors: Prof. Daniel R. Jeske (Dept. of Statistics), Ted Younglove (Dept. of Statistics), Pengyue James Lin (College of Humanities, Arts, and Social Sciences)

Abstract

Streaky Baseball is a game whose goal is to elicit interest in statistics among high school and undergraduate students as a possible field for further study. The objective of the game is to introduce simple, yet important, statistical concepts by illustrating them in a familiar and fun context.

The statistical methods illustrated in Streaky Baseball are the Binomial Distribution, Markov Chains and the Runs Test. The methodologies used to implement these statistical concepts were developed by a team of Statistics students. I worked with the team on the design of the game, and then took responsibility for the development and testing of the game. Throughout the game, users receive explanations for each new statistical concept, scores and feedback about their results to encourage them to play again to improve their scores.

I created this application using the computer technologies Java, PHP and MySQL. By using Java I was able to create an Applet that allows the user to run the application from their current web-browsers without having to download and install additional software. A MySQL database contains all the relevant information such as user high scores and MLB™ baseball teams. The PHP technology was used as an interface between the applet and the database to pass and return information. The combination of these computer technologies with statistical concepts allowed for this online multimedia self-learning tool.

Objectives

Often the concept of statistics can be intimidating to students and will discourage them from exploring the possibilities and nature of the subject. This tool will be designed to introduce to high-school students in an informative and fun way. By taking a well known sporting event and exploring it statistically, we hope that the student can have a greater appreciation for the ideas and methodology behind statistics and want to further pursue an education in the field.

Statistical Concepts

- The tool introduces the following statistical concepts to the user:
  - Binomial
  - Markov Chain
  - Longest Winning/Losing Streaks
  - Runs Test
- A binomial distribution created in R software was used to generate a winning percentage for each of the 30 baseball teams by using data from the past 30 years when applicable.
- A Binomial distribution created in Java was used for the scoring of the user predictions vs. actual results.
- A season simulation was created to model the Markov Chain for the graphs where 10,000 162-game seasons were simulated.

Architecture

The design of this tool is made up of 2 major components: the first being the client while includes a computer with a working internet connection, a web browser and the Java Run-time Environment 5.0 (JRE) to locally run the Java applet. The second component is the Apache HTTP server will communicate with the Client via HTTP requests. The Apache web server has installations of PHP and MySQL for a database interface that the applet will communicate with through file data streaming.

One reason Java was chosen to design this tool because of portability capabilities. Java is a platform independent language that will run on most OS installations including Windows, Mac OS, and the various permutations of Linux because it is a software-based platform that runs above the hardware level unlike some other programming languages.

The second reason Java was chosen was because of its extensive GUI libraries and easy to use API. By utilizing a combination of the AWT and Swing libraries I was able to develop a user interface using components such as button and combo-boxes that is both user-friendly and appealing.

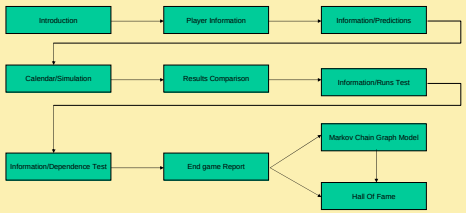
Client:

The applet runs in a non-resizable window that is launched from the web browser upon completion of the initial applet setup and image pre-loading. The structure of the applet consists of a pre-determined set of java frames that swapped into and out of the applet window. The client advances through the layout of frame by completing the required task for each frame. When database access is needed from the applet the Java will send an HTTP request to run a PHP script that is used for reading and writing to the database.

Web Server:

The applet-server communication is done through an Apache HTTP server. Apache allows for a communication line and data transfer between the server side components PHP and MySQL. Two PHP scripts were developed as an middle-man interface between the client side applet and the server side MySQL database. One of the PHP scripts was designed reading and the other for writing. When the applet needs database access it creates an HTTP request to the web server for one of those specific PHP scripts. If the request is a write, then the applet will also pass query string parameters to the script so the script will know what data needs to be stored. To pull information from the database, built-in database access calls are used and the PHP script will query the database and dynamically generate a plain text HTML file containing the necessary data. The applet will then create a streaming read buffer to this page and store the information locally in the applet's memory.

Methods



Streaky Baseball Architecture

- User can select their favorite MLB™ team and obtain a unique winning percentage.
- User can enter season predictions and compare them with the results of a season simulation model
- User can view their season progression in real time with a month-by-month calendar.
- User can explore Markov Chains using a dynamic graphing model with adjustable parameters.
- Effectively capture and simulate statistical models in Java.

Results

Markov Chain Dynamic Graph Model

Real time, Month-by-Month Calendar

Predictions vs. Simulated Model Results

Hall Of Fame

Conclusion

The success of this project has been recognized by many including staff, faculty and the public community. The positive feedback for this tool also includes a newspaper article that was written about it in a Kansas City newspaper.

From a computer engineering standpoint this project was a success and I was able to prove that teaching does not have to strictly be from a book. I proved that through use of modern technologies, interfaces and freeware, a student can be provided with a learning tool that is easy to use, helpful and enjoyable.

Acknowledgement

This Project was funded by the Reed Institute/Claremont McKenna College, UCR Computing & Communications, and Statistical Consulting Collaboratory

Statistics Team: Michelle Breaux, Yingtao Bi, Qi Zhang, Veronica Montes De Oca