MATH 23C FINAL PROJECT OUTLINE

Harvard University, Extension Studies

Spring 2022

Submitted by:

Maria Cristina An

Daniel Lebedinsky

Julio Solis Arce

1. Introduction

* Lakers statistics dataset of Game Log of the 2019-2020 Season for both the Regular Games and the Playoffs
* Filename is Lakers 2019-20 Game log
* Reference was <https://www.basketball-reference.com/teams/LAL/2020/gamelog/>
* Lakers were the champion that year against the Miami Heat

1. Explanation of the dataset

* Dataset is a record of the Lakers’ entire season game log.
* It lists both the figures of the team as well the the opponents.
* It has in the table form the criteria(ex. points , rebounds, etc) for both the team and the opponents as columns.
* It has the games played in chronological order as rows.
* There is only one character variable - the symbol name of the opposing team, one date variable - the date of the game event, two logical variables - the won or lost W\_L variable and the Western or Eastern W\_E variable.
* There are 40 numerical columns.
* This dataset is part of the final project which include 2 other files: Term Project - NBA Lakers' Stat Analysis.R and this file Math 23C Final Project.

1. Analysis includes
   1. Bar charts of wins against losses for our team.
   2. Bar charts of 3-point shots from our team.
   3. Bar charts of opponent blocks.
   4. Histogram of 3-point shots from our team.
   5. Histogram of opponent blocks
   6. Histogram of free throws from our team
   7. Histogram with probability density graph overlay of our team’s total rebounds
   8. Histogram with probability density graph overlay of our team’s scores
   9. Contingency table of the opponents scoring greater than 100 in a game
   10. Contingency table of the Lakers scoring less than 100 in a game
   11. Permutation test with p-value of the team’s scores in games played against opponents from the East and opponents from the West.
   12. Analysis of the charts and contingency tables.
   13. Comparison of analysis by classical methods and simulation methods.
2. Additional attempts for creativity or complexity
   1. A graphical display that is different from those in the textbook or in the class scripts.
   2. Appropriate use of R functions for a probability distribution other than binomial, normal or chi-square.
   3. Appropriate use of integration to calculate a significant result.
   4. A convincing demonstration of a relationship that might not have been statistically significant but that turns out to be so.
   5. A convincing demonstration of a relationship that might have been statistically significant but that turns out not to be so.
   6. Professional-looking software engineering (e.g defining and using your own functions).
   7. Nicely labeled graphics using ggplot, with good use of color, line styles, etc., that tell a convincing story.
   8. An example where permutation tests or other computational techniques clearly work better than classical methods.
   9. Appropriate use of novel statistics (e.g. trimmed mean, maximum or minimum, skewness, ratios).
   10. Use of linear regression.
   11. (2 points) Integration of well-written LaTeX
   12. (2 points) Integration of an R Shiny application using techniques taught in Paul’s weekly classes
   13. Calculation and display of a logistic regression curve.
   14. Appropriate use of covariance or correlation.
   15. Use of theoretical knowledge of chi-square, gamma, or beta distributions.
   16. Use of theoretical knowledge of sampling distributions.
   17. A graphical display that is different from those in the class scripts.
   18. Calculation of a confidence interval.
   19. Appropriate use of quantiles to compare distributions.
   20. A video of the short script is posted on YouTube and a link to it is left in your long script.