Final Project: Predicting Kobe Bryant's Shots

Belino Xhafa, Christopher Chen, Josh Felizardo, & Wenzhe (Harry) Xue

May 4, 2016

1 Research Questions

2 Motivation

3 Hypotheses

4 Methods

4.1 Overview

Our code is divided into the following files as follows:

- **analysis.R** contains preliminary exploratory analysis of the datasets
- **cleaning.R** contains the commands we ran to clean/preprocess the data
- **hypothesis_testing.R** contains various hypothesis tests for statistical significance
- **models.R** contains the models we created based on the data

Datasets:

- **raw.csv**: Raw data downloaded from Kaggle
- **cleaned.csv**: Processed data with our modifications and dummy variables added
- **clutch_shots.csv**: Data on Kobe's shooting performance in clutch situations
- **win_loss.csv**: Win-loss data scraped from `landofbasketball.com`

4.2 Data Preprocessing

4.3 Addition of Win-loss Data

4.4 New Variables Created

The following is a list of new columns that we added to the original dataset from Kaggle:

- `win`: Dummy variable for whether the Lakers won the game.
- `home`: Dummy variable for whether the game was at home or away.

- `three_pointer`: Dummy variable for whether the shot was a three-pointer.
- `jump_shot`: Dummy variable for whether the shot was a jump-shot. 0 indicates a default of layup.
- `dunk`: Dummy variable for whether the shot was a dunk. 0 indicates a default of layup.
- `tip_shot`: Dummy variable for whether the shot was a tip-shot. 0 indicates a default of layup.
- `hook_shot`: Dummy variable for whether the shot was a hook-shot. 0 indicates a default of layup.
- `bank_shot`: Dummy variable for whether the shot was a bank-shot. 0 indicates a default of layup.
- `game_date_formatted`: Reformatted date into R's native format for boolean comparisons during data processing. Pretty useless otherwise.
- `game_number`: Normalized game date. First game is 1, for game i, $game\_number[i] = i$.
- `avg`: Average shot percentage for each game.
- `shots_made`: Shots made for each game (may seem redundant, but useful for calculating averages over multiple games since we can't just average the averages)
- `shots_taken`: Shots taken per game (may seem redundant, but useful for calculating averages over multiple games since we can't just average the averages)
- `clutch_threshold`: Number of minutes remaining at which we begin counting shots as clutch shots.
- `clutch_perc`: Average shot percentage for clutch shots (shots attempted with below `clutch_threshold` minutes remaining) for each game.
- `clutch_shots_made`: number of clutch shots made for each game.
- `clutch_shots_taken`: number of clutch shots taken for each game.
- `ot`: dummy variable for whether or not the game went overtime.
- `ot_taken`: number of shots taken in OT.
- `ot_made`: number of shots made in OT.
- `ot_avg`: OT shooting percentage for each game.
- `season_norm`: represents the number of seasons Kobe has been in the NBA.

5 Assumptions

6 Results

7 Limitations

8 Conclusion

9 Challenges Faced

10 References

11 Acknowledgements

Special thanks to Phillip Huang for his kind assistance in helping us scrape win-loss data using the Web Scraper Chrome extension.

12 Theoretical Analysis