The Big Dance

Our goal on this assignment is to evaluate three-point shooters in NCAA women's basketball. This is an example of an important job function fo analysts working in basketball front offices because evaluating talent is the starting point for player personnel decisions (draft, trades, free agency, etc.).

Instructions

Use the wehoop R package to download player box data and schedule data for the current NCAA women's basketball season. Estimate a hierarchical model for player-game three-point field goal percentage with random effects for player, conference, opposing conference, and opposing team. Extract estimates of player true talent three-point shooting percentage from this model. Think carefully about which effects to include when estimating player talent.

Write a report (max 2 pages) summarizing your findings, assuming the reader is a sports data scientist who does not know your assignment. Tell a STORY about your findings. Your report is a narrative about what the analysis reveals and why it matters. For this assignment, you must follow a specific prescribed format for your report:

- **Introduction.** Explain why your analysis matters (one paragraph).

- **Methods.** Write down and explain the mathematical specification of the model (one paragraph).

- **Results.** Display two results:

  - a table of top ten shooters by estimated true talent shooting percentage
  - a figure plotting estimated true talent (y-axis) vs. observed (x-axis) shooting percentage

  Do the results look correct? Explain whether they pass the "sniff" test and why (one paragraph).

- **Discussion.** Write two paragraphs:

  - Escape from Model Land. Choose a player for whom you think the estimated true talent is too high or too low. Provide your best guess of the player's actual true talent (guess an actual number), and explain what assumption(s) caused the model to get it wrong.
  - Write a concluding paragraph for your report. Recall that the purpose of the concluding paragraph is to (1) signal the ending; (2) restate the purpose of the report; (3) restate the key results; and (4) state why the findings matter (end with a bang).

Extra Credit

Fit another hierarchical model with random effects only for player and opposing team (not conference or opposing conference). Is the population variance for opposing team larger or smaller than in the full model? Explain why. Submit your response as a separate PDF file named `extra_credit.pdf`.

Submission Requirements

Anonymize your submission by removing any personally identifiable information (including file paths in your R script that contain things like a username!).

- A PDF report named `the_big_dance.pdf`, summarizing your findings (max 2 pages)

- An R sript named `the_big_dance.R`, containing all of the code you usd to perform the analysis

- A CSV file named `the_big_dance.csv` with your player picks (not graded, see page 2)

	Needs Improvement (60%)	Meets Expectations (80%)	Exceeds Expectations (100%)
Modeling	*Your mathematical specification is missing, incomplete, or contains significant errors.*	*You correctly specified the model mathematically, but some notation may be slightly inconsistent.*	*Your mathematical specification is flawless and precise, fully defining the model parameters and structure.*
Coding	*You did not implement the model correctly.*	*You implemented the model correctly, but your code is disorganized, lacks comments, or ignores Tidyverse style.*	*You implemented the model correctly with clean code: well commented, easy to read, and adhering to Tidyverse style.*
Critical Thinking	*Your analysis reveals a misunderstanding, or your results don't pass the "sniff" test, and you do not notice.*	*You discuss the results and identify model limitations in a general sense.*	*You connect specific results to model assumptions, identifying exactly where and why those assumptions limit the findings.*
Communication	*Your writing is disorganized; your figures/tables are messy, unlabelled, or difficult to interpret.*	*Your writing is clear and figures are functional, but the report lacks a polished professional feel or strong paragraph structure.*	*Your figures and tables look polished, and your paragraphs have clear topic sentences with supporting evidence.*

Player Picks

For fun, we will run a contest. Pick 10 players and rank them 1 through 10. For every three-pointer made by a player on your list, you score points based on your rank for the player. Specifically, your total score is

$$\sum_{r=1}^{10}(11 - r) \cdot Y_r,$$

where Y_r is the total number of three-pointers hit by your r-ranked player during the tournament. Submit your picks as a CSV with columns `rank` (integer-valued) and `player_id` (integer-valued). The results of this contest will determine draft order for project presentation time slots at the end of the semester.