Finding the True MVP*

An NBA Analysis Through Linear Regression

Yan B. Mezhiborsky

December 2, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

Table of contents

*Code and data are available at: https://github.com/RohanAlexander/starter_folder.

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2….

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023)…. Our data (Toronto Shelter & Support Services 2024)…. Following Alexander (2023), we consider…

Overview text

2.2 Measurement

Some paragraphs about how we go from a phenomena in the world to an entry in the dataset.

2.3 Outcome variables

Add graphs, tables and text. Use sub-sub-headings for each outcome variable or update the subheading to be singular.

Some of our data is of penguins (**?@fig-bills**), from Horst, Hill, and Gorman (2020).

Talk more about it.

And also planes (**?@fig-planes**). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

Talk way more about it.

2.4 Predictor variables

Add graphs, tables and text.

Use sub-sub-headings for each outcome variable and feel free to combine a few into one if they go together naturally.

3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix B.

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$
\begin{aligned}
y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) && (1)\\
\mu_i &= \pi_i + \alpha_i + \rho_i + \beta_i + \xi_i + \omega_i && (2)\\
\mu &\sim \text{Normal}(0, 2.5) && (3)\\
\pi &\sim \text{Normal}(0, 2.5) && (4)\\
\alpha &\sim \text{Normal}(0, 2.5) && (5)\\
\rho &\sim \text{Normal}(0, 2.5) && (6)\\
\beta &\sim \text{Normal}(0, 2.5) && (7)\\
\xi &\sim \text{Normal}(0, 2.5) && (8)\\
\omega &\sim \text{Normal}(0, 2.5) && (9)\\
\sigma &\sim \text{Exponential}(1) && (10)
\end{aligned}
$$

```
# Fit a linear regression model
lm_model <- lm(MVP_index ~ PTS + AST + TRB + BLK + STL + WS, data = nba_master)

# View the summary of the model
summary(lm_model)
```

```
Call:
lm(formula = MVP_index ~ PTS + AST + TRB + BLK + STL + WS, data = nba_master)

Residuals:
     Min       1Q   Median       3Q      Max
-0.27967 -0.06232 -0.00013  0.06198  0.33044

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.1545172  0.0136329   11.33   <2e-16 ***
PTS         0.0337247  0.0005182   65.08   <2e-16 ***
AST         0.0445034  0.0012761   34.88   <2e-16 ***
TRB         0.0295472  0.0010966   26.95   <2e-16 ***
BLK         0.1125198  0.0041166   27.33   <2e-16 ***
STL         0.1060966  0.0056427   18.80   <2e-16 ***
WS          0.0199776  0.0009258   21.58   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09672 on 1536 degrees of freedom
Multiple R-squared:  0.8992,    Adjusted R-squared:  0.8988
F-statistic:  2284 on 6 and 1536 DF,  p-value: < 2.2e-16
```

```r
next_year_mvp_stats <- mvp_table %>%
  filter(Year == 2024) %>%  # Replace with the target year
  select(PTS, AST, TRB, BLK, STL, WS)

# Predict MVP_index for the next year's MVP
predicted_index_mvp_table <- predict(lm_model, newdata = next_year_mvp_stats)
print(paste("Predicted MVP Index (mvp_table):", predicted_index_mvp_table))
```

```
[1] "Predicted MVP Index (mvp_table): 2.45027833674002"
```

```r
next_year_real_mvp_stats <- merged_real_mvp %>%
  filter(Year == 2024) %>%  # Replace with the target year
  select(PTS, AST, TRB, BLK, STL, WS)

# Predict MVP_index for the next year's MVP
predicted_index_real_mvp <- predict(lm_model, newdata = next_year_real_mvp_stats)
print(paste("Predicted MVP Index (merged_real_mvp):", predicted_index_real_mvp))
```

```
[1] "Predicted MVP Index (merged_real_mvp): 2.40118723560444"
```

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ.

3.1.2 Model Results

```
library(dplyr)

# Filter MVPs from 2014 to 2024 from `mvp_table`
mvp_table_filtered <- mvp_table %>%
  filter(Year >= 2014 & Year <= 2024) %>%
  select(Year, Player, MVP_index)

# Filter MVPs from 2014 to 2024 from `merged_real_mvp`
real_mvp_filtered <- merged_real_mvp %>%
  filter(Year >= 2014 & Year <= 2024) %>%
  select(Year, Player, MVP_index)

# Get stats for predicting the 2025 MVP_index from `mvp_table`
next_year_mvp_stats <- mvp_table %>%
  filter(Year == 2025) %>%
  select(PTS, AST, TRB, BLK, STL, WS)

predicted_index_mvp_table <- NA
if (nrow(next_year_mvp_stats) > 0) {
  predicted_index_mvp_table <- predict(lm_model, newdata = next_year_mvp_stats)
}

# Get stats for predicting the 2025 MVP_index from `merged_real_mvp`
next_year_real_mvp_stats <- merged_real_mvp %>%
  filter(Year == 2025) %>%
  select(PTS, AST, TRB, BLK, STL, WS)

predicted_index_real_mvp <- NA
if (nrow(next_year_real_mvp_stats) > 0) {
  predicted_index_real_mvp <- predict(lm_model, newdata = next_year_real_mvp_stats)
```

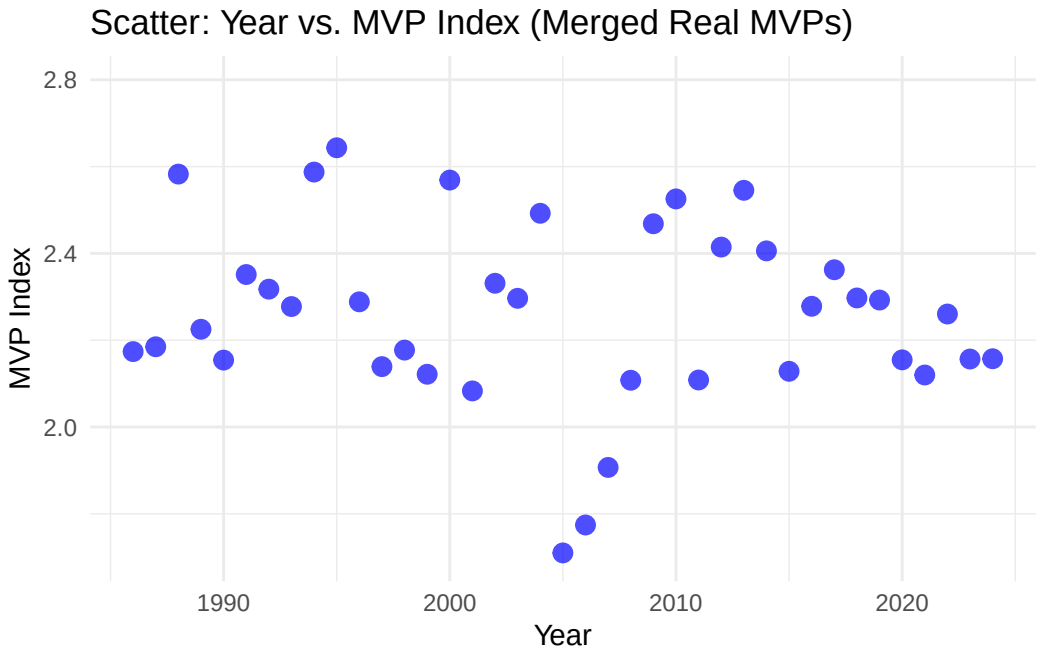
```r
}

# Combine results from 2014-2024 and add the 2025 predicted index
results_table <- bind_rows(
  mvp_table_filtered %>% mutate(Source = "mvp_table"),
  real_mvp_filtered %>% mutate(Source = "merged_real_mvp")
)

# Add 2025 predictions to the table
results_table <- results_table %>%
  mutate(
    Predicted_2025_Index = case_when(
      Source == "mvp_table" ~ predicted_index_mvp_table,
      Source == "merged_real_mvp" ~ predicted_index_real_mvp,
      TRUE ~ NA_real_
    )
  )

# View the table
print(results_table)
```

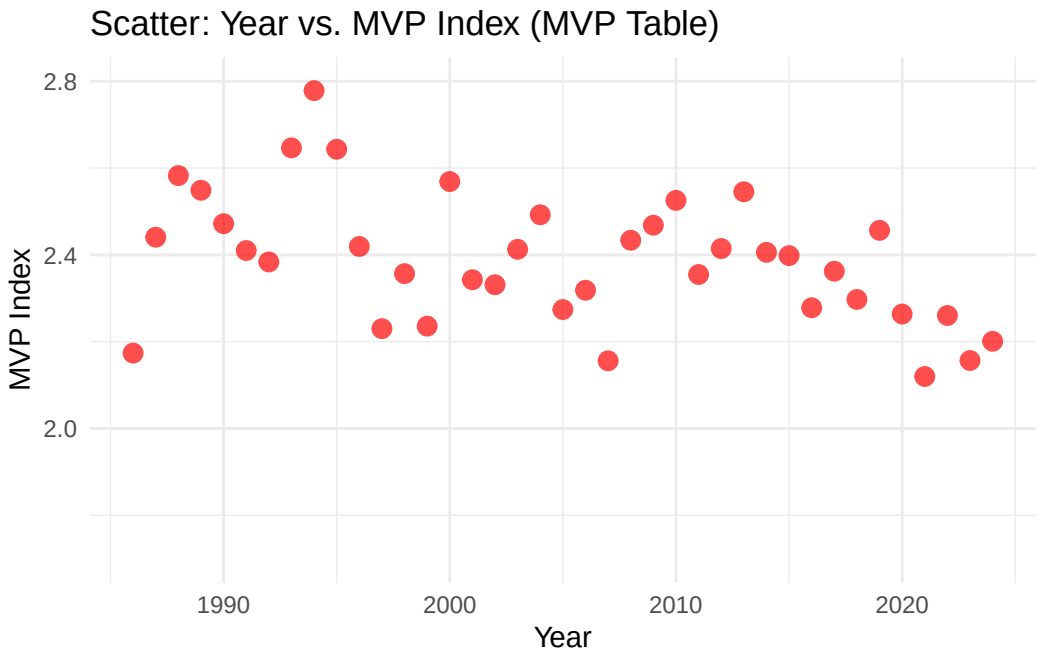
```
   Year                Player MVP_index          Source Predicted_2025_Index
1  2014          Kevin Durant  2.405868       mvp_table                    NA
2  2015     Russell Westbrook  2.398289       mvp_table                    NA
3  2016         Stephen Curry  2.278181       mvp_table                    NA
4  2017     Russell Westbrook  2.362328       mvp_table                    NA
5  2018          James Harden  2.297349       mvp_table                    NA
6  2019          James Harden  2.456584       mvp_table                    NA
7  2020          James Harden  2.263751       mvp_table                    NA
8  2021          Nikola Jokić  2.119867       mvp_table                    NA
9  2022          Nikola Jokić  2.260379       mvp_table                    NA
10 2023           Joel Embiid  2.156639       mvp_table                    NA
11 2024           Luka Dončić  2.200904       mvp_table                    NA
12 2019 Giannis Antetokounmpo  2.292537 merged_real_mvp                    NA
13 2020 Giannis Antetokounmpo  2.154718 merged_real_mvp                    NA
14 2018          James Harden  2.297349 merged_real_mvp                    NA
15 2023           Joel Embiid  2.156639 merged_real_mvp                    NA
16 2014          Kevin Durant  2.405868 merged_real_mvp                    NA
17 2021          Nikola Jokić  2.119867 merged_real_mvp                    NA
18 2022          Nikola Jokić  2.260379 merged_real_mvp                    NA
19 2024          Nikola Jokić  2.157337 merged_real_mvp                    NA
20 2017     Russell Westbrook  2.362328 merged_real_mvp                    NA
```

```
21 2015          Stephen Curry  2.128396 merged_real_mvp                NA
22 2016          Stephen Curry  2.278181 merged_real_mvp                NA
```

4 Results

Our results are summarized in **?@tbl-modelresults**.





5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

B.2 Diagnostics

Checking the convergence of the MCMC algorithm

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *palmerpenguins: Palmer Archipelago (Antarctica) penguin data*. https://doi.org/10.5281/zenodo.3960218.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Toronto Shelter & Support Services. 2024. *Deaths of Shelter Residents*. https://open.toronto.ca/dataset/deaths-of-shelter-residents/.