Sports Analytics

Aaron Nielsen, Department of Statistics, Colorado State University

2022-05-23

Contents

About

This book serves as the course textbook for:

- STAT 351 (Sports Statistics and Analytics 1)
- STAT 451 (Sports Statistics and Analytics 2)

This project was first created during the summer of 2022 by:

- Aaron Nielsen
- Levi Kipp
- Ellie Martinez
- Isaac Moorman

Current Tasks

Updated: "2022-05-23"

Team Tasks and Tips

1. Find datasets from various sports to use as examples for EDA and later chapters
2. Show how to get basic sumamry statistics from these datasets using dplyr, tidy
3. Describe and calculate useful team and individual (descriptive statistics). Example: Baseball: calculate AVG, OBP, OPS, WOBA
4. (High quality) Visualizations using ggplot
5. Look for relevant "sports" R packages
6. Include examples from CSU and Colorado sports teams when possible
7. Sports to be included: Baseball/Softball, Football, Basketball, Soccer, Hockey, Volleyball
8. Sports to be potentially included: Lacrosse, Cricket, Handball,

---

Aaron:

Sports:

Chapters: Currently working to add content to chapters 1-4

Ellie:

Sports: Soccer, Volleyball

Chapters: EDA, Probability

Levi:

Sports: Basketball, Hockey

Chapters: EDA, Probability

Isaac:

Sports: Baseball, Football, Tennis

Chapters: EDA, Scraping

Chapter 1

Exploratory Data Analysis

Chapter 2

Probability

2.1 Definitions and Axioms

2.2 Theorems and Laws

2.3 Random Variables

Chapter 3

Simulation

Chapter 4

Statistical Inference

4.1 One Sample and Two Sample t-tests and confidence intervals

Chapter 5

Correlation

Chapter 6

Linear Regression

Chapter 7

Data Scraping

Chapter 8

Principal Component Analysis

Chapter 9

Clustering

Chapter 10

Classification

Chapter 11

Decision Trees

11.1 Random Forests

11.2 Gradient Boosting

Chapter 12

Non-parametric Statistics

Chapter 13

Baseball

Chapter 14

Football

Chapter 15

Basketball

Chapter 16

Soccer

Chapter 17

Hockey

Chapter 18

Volleyball

18.1 Resources

Women's Volleyball D1 Statistics

Chapter 19

Other Sports

Chapter 20

Ellie's stuff

Chapter 21

Levi's stuff

Chapter 22

Isaac's stuff

Chapter 23

Aaron's stuff

23.1 Notes for Chapter 2 (Probability)

Axioms of Probability:

1. $P(A) \geq 0$
2. $P(\Omega) = 1$
3. If $A_1, A_2, \ldots, A_n$ are disjoint events, then $P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$

Theorem 23.1 (Bayes theorem)**.** *Let A and B be events in Ω such that $P(B) > 0$. Then we have the following:*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

23.2 Notes for Chapter 4 (Simulation)

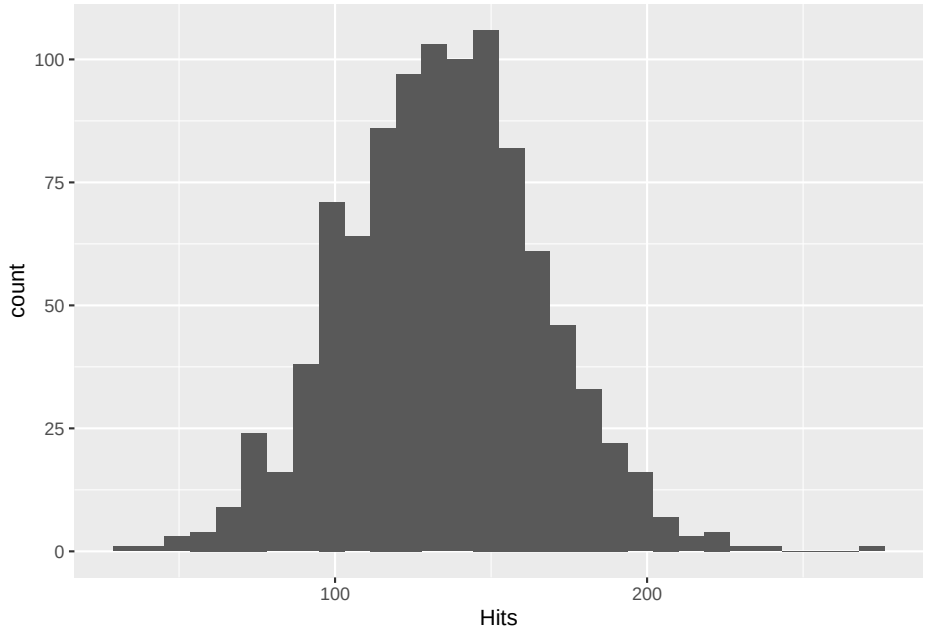
23.2.1 Baseball Simulation Example

```
library(tidyverse)
```

This is a baseball example for chapter 4.

```
set.seed(2022)
n.sims <- 1000
hits <- rep(0,n.sims)
avg <- 0.300
atbats.mean <- 450
atbats.sd <- 100
sim.atbats <- round(rnorm(n.sims,atbats.mean,atbats.sd))
```

```
for(i in 1:n.sims){
  sim.hits <- rbinom(1,sim.atbats[i],avg)
  hits[i] = sim.hits
}
hits.df <- data.frame(Hits=hits)
hits.df %>% ggplot(aes(x=Hits)) + geom_histogram()
```



Reference: Blocks

23.3 Equations

Here is an equation.

$$f\left(k\right) = \binom{n}{k} p^k \left(1-p\right)^{n-k} \tag{23.1}$$

You may refer to using `\@ref(eq:binom)`, like see Equation (23.1).

23.4 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref(thm:tri)`, for example, check out this smart theorem 23.2.

Theorem 23.2. *For a right triangle, if c denotes the* length *of the hypotenuse and a and b denote the lengths of the* ***other*** *two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html.

23.5 Callout blocks

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html

Reference: Footnotes and citations

23.6 Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one [1].

23.7 Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package (Xie 2016) (check out the last code chunk in index.Rmd to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** (Xie 2015) (this citation was added manually in an external file book.bib). Note that the `.bib` files need to be listed in the index.Rmd with the YAML `bibliography` key.

The RStudio Visual Markdown Editor can also make it easier to insert citations: https://rstudio.github.io/visual-markdown-editing/#/citations

Xie, Yihui. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. http://yihui.name/knitr/.

———. 2016. *Bookdown: Authoring Books and Technical Documents with r Markdown*. https://github.com/rstudio/bookdown.

[1] This is a footnote.