

Sports Analytics

Aaron Nielsen, Department of Statistics, Colorado State University

2022-05-24

Contents

About	5
Current Tasks	7
1 Exploratory Data Analysis	9
1.1 Using dplyr, tidyverse, ggplot	9
1.2 Baseball	9
1.3 Football	9
1.4 Basketball	9
1.5 Soccer	9
1.6 Volleyball	9
1.7 Hockey	9
2 Probability	11
2.1 Definitions and Axioms	11
2.2 Theorems and Laws	11
2.3 Random Variables	11
3 Simulation	13
4 Statistical Inference	15
4.1 One Sample and Two Sample t-tests and confidence intervals . .	15
5 Correlation	17
6 Linear Regression	19
7 Data Scraping	21
8 Principal Component Analysis	23
9 Clustering	25
10 Classification	27

11 Decision Trees	29
11.1 Random Forests	29
11.2 Gradient Boosting	29
12 Non-parametric Statistics	31
13 Baseball	33
14 Football	35
15 Basketball	37
16 Soccer	39
17 Hockey	41
18 Volleyball	43
18.1 Resources	43
19 Other Sports	45
20 Ellie's stuff	47
21 Levi's stuff	49
22 Isaac's stuff	51
23 Aaron's stuff	53
23.1 Notes for Chapter 2 (Probability)	53
23.2 Suggested Readings	53
23.3 Notes for Chapter 4 (Simulation)	54
Reference: Blocks	55
23.4 Equations	55
23.5 Theorems and proofs	55
23.6 Callout blocks	55
Reference: Footnotes and citations	57
23.7 Footnotes	57
23.8 Citations	57
24 References	59

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: “2022-05-24”

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 summary statistics from these datasets using dplyr, tidy
3. Describe and calculate useful team and individual (descriptive statistics).
Example: Baseball: calculate AVG, OBP, OPS, WOB
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

1.1 Using dplyr, tidyverse, ggplot

1.2 Baseball

1.3 Football

1.4 Basketball

1.5 Soccer

1.6 Volleyball

1.7 Hockey

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, \dots, 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 Suggested Readings

23.2.1 Moneyball

Moneyball, Chapter 2, How to Find a Ballplayer (Lewis, 2004)

Near the end of the chapter (page 40), Michael Lewis give a list of players the Oakland Athletics hoped to draft. How did these players turn out? Find the WAR for each of the players in their pre-free agency years and compare it against the Rockies draft picks in the same rounds from the same draft.

23.2.2 Future Value

Future Value, Chapter 7, How to Scout (Longenhagen and McDaniel, 2020)

If a player receives a running grade of 40, approximately what proportion of MLB players have a lower have a lower running grade?

For a given tool, about 95% of all player grades fall between what two bounds? (Consider the middle 95% of the distribution of grades.)

23.3 Notes for Chapter 4 (Simulation)

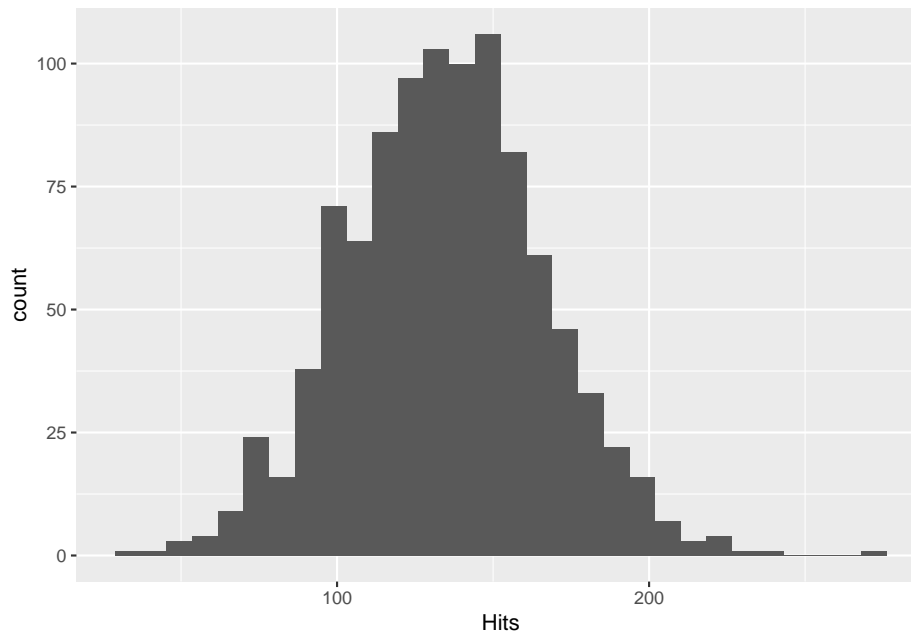
23.3.1 Baseball Simulation Example

```
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

23.4 Equations

Here is an equation.

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

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

23.5 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>.

23.6 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

23.7 Footnotes

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

23.8 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>

¹This is a footnote.

Chapter 24

References

Bibliography

- Lewis, M. (2004). *Moneyball: The art of winning an unfair game*. WW Norton & Company.
- Longenhagen, E. and McDaniel, K. (2020). *Future Value: The battle for baseball's soul and how teams will find the next superstar*. Triumph Books.
- Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.
- Xie, Y. (2016). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.3.9.