

Sports Analytics

Aaron Nielsen

2022-05-23

Contents

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

11.2 Gradient Boosting	27
12 Non-parametric Statistics	29
13 Baseball	31
14 Football	33
15 Basketball	35
16 Soccer	37
17 Hockey	39
18 Volleyball	41
19 Other Sports	43
20 Ellie's stuff	45
21 Levi's stuff	47
22 Isaac's stuff	49
23 Aaron's stuff	51
23.1 Notes for Chapter 2 (Probability)	51
23.2 Notes for Chapter 4 (Simulation)	51
24 Reference: Footnotes and citations	53
24.1 Footnotes	53
24.2 Citations	53
25 Reference: Blocks	55
25.1 Equations	55
25.2 Theorems and proofs	55
25.3 Callout blocks	55

About

This book serves as the course textbook for STAT 351 (Sports Analytics 1) and STAT 451 (Sports Analytics 2) at Colorado State University.

This project was first created during the summer of 2022 by Aaron Nielsen, Levi Kipp, Ellie Martinez, and Isaac Moorman.

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

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 Notes for Chapter 4 (Simulation)

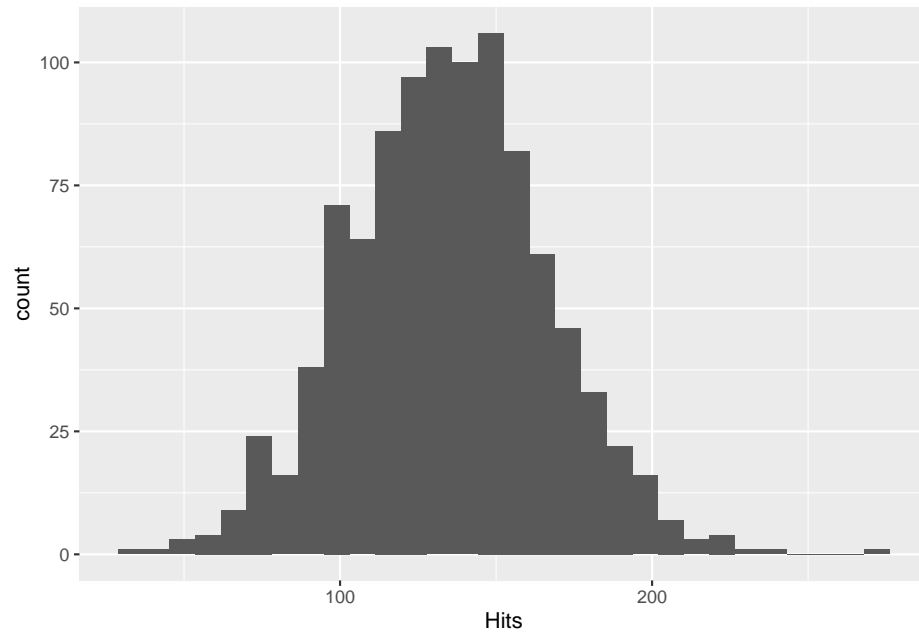
23.2.1 Baseball Simulation Example

```
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()
```



Chapter 24

Reference: Footnotes and citations

24.1 Footnotes

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

24.2 Citations

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

For example, we are using the **bookdown** package [Xie, 2022] (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 25

Reference: Blocks

25.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (25.1)$$

You may refer to using `\@ref{eq:binom}`, like see Equation (25.1).

25.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref{thm:tri}`, for example, check out this smart theorem 25.1.

Theorem 25.1. *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>.

25.3 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>

Bibliography

Yihui Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition, 2015. URL <http://yihui.org/knitr/>. ISBN 978-1498716963.

Yihui Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*, 2022. URL <https://CRAN.R-project.org/package=bookdown>. R package version 0.26.