Errata for *Quantitative Social Science: An Introduction* (Princeton University Press)

Kosuke Imai

September 30, 2022

Corrections applicable to QSS: tidyverse first printing

Chapter 3

Section 3.6.2

- page 120-121. 0 and 1 should be flipped when describing the Gini coefficient, in both text and text box. Text should read "In a perfectly equal society, the Gini coefficient is 0. In contrast, a society where one person possesses all the wealth has a Gini coefficient of 1." Text box should read "It ranges from 1 (one person possesses all the wealth) to 0 (everyone has the same amount of wealth)."

Corrections applicable to the first, second, and third printings

Chapter 2

Section 2.8.1.

- page 70. Move the last sentence of Question 1, which begins with "Recall that", to the end of Question 2.
- pages 70-71. "aid" should be "aide". The mistake appears in Table 2.6, Questions 4 and 5.

Chapter 3

Section 3.3.2

- pages 81–82. The intervals should be right-closed, which is the default of R, rather than left-closed. For example, in the last paragraph of page 81, change "which results in the intervals $[15, 20), [20, 25), [25, 30)$, and so on" to "which results in the intervals $(15, 20], (20, 25], (25, 30]$, and so on"

Chapter 4

Section 4.1.3

- page 136, second paragraph. Change "while Illinois (`IN`) and North Carolina (`NC`)" to "while Indiana (`IN`) and North Carolina (`NC`)"
- page 136, third paragraph. Change "which was 364 votes." to "which was 364 votes (see footnote~3)."

Chapter 5

Section 5.1.2

- page 197, third paragraph. Change "for which the default value is `FALSE`. If this argument is set to `TRUE`, then term frequency `tf(w, d)` will be divided by the total number of terms in document `d`." to "for which the default value is `TRUE`. If this argument is set to `FALSE`, then term frequency `tf(w, d)` will not be divided by the total number of terms in document `d`."

Section 5.2.2

- page 210, second paragraph. Change "uniquely paired with 105 other nodes" to "uniquely paired with 15 other nodes"

Chapter 6

Section 6.1.5

- page 253, Figure 6.4. Change "gone with out the major reforms" to "gone without the major reforms".

Section 6.3.4

- page 288, first paragraph. Change "the second inequality holds since" to "the second equality holds since"

Chapter 7

Section 7.3.5

- page 384, textbox. Change the last equation to

$$\mathbb{V}(aX + bY + c) \;=\; a^2\mathbb{V}(X) + b^2\mathbb{V}(Y) + 2ab\ \mathrm{Cov}(X, Y)$$

Section 7.5.3.

- pages 394 - 396. The year of election should be 1933 rather than 1932

Corrections applicable to the first and second printings

Chapter 3

Section 3.7.3

- page 114, first paragraph. Change "produces 1 cluster containing only Democrats and the other consisting only of Republicans" to "produces one cluster containing all Democrats except one and the other consisting only of Republicans".

Chapter 4

Section 4.1.1

- page 126. In the second code chunk, the results should be "`1.5 NA NA`" rather than "`1 NA NA`"

Section 4.2.1

- page 140, first paragraph. Add the following sentence immediately after the first sentence, "Here, the competence measure for a Democratic candidate, for example, represents the proportion of experimental subjects who rated the Democrat more competent than the Republican."

Section 4.2.3

- page 148. Change "roughly 13 percentage points" to "12 – 13 percentage points"

Chapter 5

Section 5.1.1

- page 192. In the `tm` package version 0.7-1, which was released after this book went into the printer, it introduced `SimpleCorpus`, which became the default output object class for the `Corpus()` function. This change affects the results though the overall conclusions of the analysis remain similar. To obtain the results in the book, you will have to specify `VCorpus` as the output class using the following syntax:

```
corpus.raw <- VCorpus(DirSource(directory = "federalist", pattern = "fp"))
```

Section 5.1.4

- page 200, third paragraph. Change "we focus on the usage of articles, prepositions, and conjunctions" to "we focus on the usage of adjectives, adverbs, prepositions and conjunctions"

Section 5.2.2

- page 209, first paragraph. Add the following footnote to the end of the paragraph, "If a node is not connected to any other node, then the number of nodes in the graph, i.e., 16 in this case, is used instead of the length of the shortest path. Thus, Pucci family's closeness is equal to $1/(15 \times 16)$"

- page 211, first paragraph. In the latest version of the `igraph` package, the `closeness()` function returns `NaN` for this family. To prevent an error, we must manually code this as follows:

```
close <- closeness(florence)
close["PUCCI"] <- 1 / (15 * 16)
plot(florence, vertex.size = close * 1000,
     main = "Closeness")
```

Section 5.3.1

- page 221, first paragraph. Change "the neighborhoods along the River Thames (indicated by the blue region)" to "the area further south (indicated by the red region)". Change "the area further south (indicated by the red region)" to "the neighborhoods along the River Thames (indicated by the blue region)".
- page 222, Figure 5.5 caption. Change "blue (Lambeth) and red (Southwark and Vauxhall)" to "red (Lambeth) and blue (Southwark and Vauxhall)".

Section 5.3.5

- page 231, first paragraph. Change "Bentonville, Arkansas" to "Rogers, Arkansas".

Section 5.3.6

- page 235, first pragraph. Change "Midwestern" to "Southern".

Chapter 6

Section 6.1.5

- page 253, first paragraph. Change "$0.84 = 1 - 0.016 - 0.135$" to "$0.85 \approx 1 - 252/15504 - 2100/15504$".
- page 253, footnote 1. Add the following sentence to the footnote, "Although there are a total of 86 words from `For` to `time`, we follow the original article and use 85".

Section 6.2.1

- page 259, Table 6.2. We are missing the last row for `Other` whose entries should be 0.017 (Female), 0.017 (Male), 0.034 (Marginal prob.)
- page 273, the third set of equations should be changed to:

$$\begin{aligned} P(\text{surname} \mid \text{race and residence}) &= \frac{P(\text{surname and residence} \mid \text{race})}{P(\text{residence} \mid \text{race})} \\ &= \frac{P(\text{surname} \mid \text{race})P(\text{residence} \mid \text{race})}{P(\text{residence} \mid \text{race})} \\ &= P(\text{surname} \mid \text{race}). \end{aligned} \tag{6.23}$$

Section 6.2.4

- page 274, the following line

```
race.prop <-
    apply(FLCensus[, c("white", "black", "api", "hispanic", "others")],
          2, weighted.mean, weights = FLCensus$total.pop)
```

should be changed to

```
race.prop <-
    apply(FLCensus[, c("white", "black", "api", "hispanic", "others")],
          2, weighted.mean, w = FLCensus$total.pop)
```

In addition, the last sentence of this page, "The `weighted.mean()` function can be used to compute weighted averages, in which the `weights` argument takes a vector of weights.", should be changed to "The `weighted.mean()` function can be used to compute weighted averages, in which the `w` argument takes a vector of weights."

- pages 276 and 277. As a result of the above change in the code, the numerical results on these two pages change somewhat. Please see the `probability.pdf` file for details. The first sentence of the final paragraph on page 276 should be changed to "The true positive rate for blacks has jumped from 16% to 63%."

Corrections only applicable to the first printing

Table of Contents

- page ix. Change "Heterogenous" to "Heterogeneous" in the Section 4.3.3 title

Chapter 1

Section 1.3.8

- page 27, first paragraph. Change "The `lintr()` function in the **lintr** package" to "The `lint()` function in the **lintr** package"
- page 27, code output. Change "`## UNpop.R:8:7: style`" to "`## UNpop.R:7:7: style`"

Chapter 3

Section 3.3.2

- page 83, first paragraph. Change "skewed towards the left" to "right-skewed"

Chapter 4

Section 4.3.3

- page 170. Change "4.3.3 HETEROGENOUS TREATMENT EFFECTS" to "4.3.3 HETEROGENEOUS TREATMENT EFFECTS".
- page 170. Chage "helpful for exploring *heterogenous treatment effects*" to "helpful for exploring *heterogeneous treatment effects*".
- page 170. Change "To illustrate the analysis of heterogenous treatment effects" to "To illustrate the analysis of heterogeneous treatment effects"
- pages 170 – 176. Throughout this section, the `primary2008` variable should be labeled as `primary2006` so that it matches with the `social.csv` data file introduced in Chapter 2. For now, we include another version of `social.csv` in this chapter's folder so that users can apply the code.
- page 181. Change "We discussed how to estimate heterogenous treatment effects" to "We discussed how to estimate heterogeneous treatment effects"

Chapter 5

Section 5.2.3

- page 212. The code chunk that loads the data sets need to be changed to:

```
twitter <- read.csv("twitter-following.csv", stringsAsFactors = FALSE)
senator <- read.csv("twitter-senator.csv", stringsAsFactors = FALSE)
```

so that the names of senators will be treated as a character variable instead of a factor variable (default). Unfortunately, this changes the results of the rest of this subsection 5.2 although it does not change the code.

Section 5.3.6

- page 235, second paragraph. Change "by opening a web browser and clicking `File > Open file...` in the menu." to "opening the resulting `walmart.html` file in a web browser."

Chapter 6

Section 6.2.2

- page 265. The original code for the Monty Hall problem does not return the right answer when the order of doors is changed. This is due to the fact that the `sample()` function behaves differently when an integer is supplied as an input. The correct code that avoids this problem is below:

```
sims <- 1000
doors <- c("goat", "goat", "car")
result.switch <- result.noswitch <- rep(NA, sims)

for (i in 1:sims) {
    ## randomly choose the initial door
```

```
    first <- sample(1:3, size = 1)
    result.noswitch[i] <- doors[first]
    remain <- doors[-first] # remaining two doors
    ## Monty chooses one door with a goat
    if (doors[first] == "car") # two goats left
        monty <- sample(1:2, size = 1)
    else # one goat and one car left
        monty <- (1:2)[remain == "goat"]
    result.switch[i] <- remain[-monty]
}

mean(result.noswitch == "car")
mean(result.switch == "car")
```

Section 6.3.3.

- page 284, second paragraph. Change "$\{HTHTHT\}$" to "$\{HTHTH\}$".

Section 6.4.2.

- page 304, first paragraph. Change "we expect a binomial random variable to approximate the normal distribution as the sample size, or the number of balls in this case, increases." to "we expect the binomial random variable to approximate the normal random variable as the sample size, or the number of lines of pegs in this case, increases. Here, the sample size refers to the number of lines of pegs, not the number of balls. Increasing the latter reduces the Monte Carlo error." Also, change "The central limit theorem applies not only to the binomial distribution, but" to "The central limit theorem applies not only to the Bernoulli random variable, but".
- page 304, equation (6.42). The second term is missing X_i, which is highlighted in the correct equation below:

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \mathbb{E}(X)$$

- page 305. Add a sentence to the end of the last paragraph whose last sentence ends with "approximated by the standard normal distribution." The sentence to be added is "To illustrate the quincunx through Monte Carlo simulations, we sample from the Bernoulli distribution or equivalently the Binomial distribution with size $n = 1$."

Chapter 7

Section 7.1.3

- page 327, last paragraph. Change "such that $P(Z > \alpha/2) = 1 - P(Z \leq \alpha/2) = 1 - \alpha/2$" to "such that $P(Z > z_{\alpha/2}) = 1 - P(Z \leq z_{\alpha/2}) = 1 - \alpha/2$"
- page 329, last paragraph. Change "Consider the probability that $(1-\alpha/2)\times 100\%$ confidence interval" to "Consider the probability that $(1-\alpha)\times 100\%$ confidence interval"
- page 330, Step 4 in the box. Change "Compute the critical value $z_{\alpha/2}$ as the $(1-\alpha)\times 100$ percentile value" to "Compute the critical value $z_{\alpha/2}$ as the $(1-\alpha/2)\times 100$ percentile value"

Section 7.2.3

- page 354, first paragraph. The first sentence should read: "We can confirm this result using the current example by checking that 0.5 is contained in the 99% confidence interval (we fail to reject the null hypothesis when $\alpha = 0.01$) but not in the 95% confidence interval (we reject the null when $\alpha = 0.05$)."

General Index

- page 402. Change "heterogenous treatment effects, 170" to "heterogeneous treatment effects, 170"

Acknowledgements

Thanks to Jeff Arnold, Masahiko Asano, Matt Blackwell, Lori Bougher, Alison Durham, Calvin Garner, Bo Coleman, Michael Donnelly, Fallend Franz, Kentaro Fukumoto, Joel Gautschi, Mamiko Hamakado, Masataka Harada, Raymond Hicks, Danya Lagos, Michael Lewis, Yuko Kasuya, Soyoung Lee, Rocio Titiunik, and Sandy Weisberg for pointing out the errors.