

Errata to *Thinking Clearly with Data: A Guide to Quantitative Reasoning and Analysis*

Ethan Bueno de Mesquita and Anthony Fowler

March 1, 2024

This errata corrects a few small errors that have been brought to our attention. We appreciate all of the readers who have pointed out mistakes. We want to especially acknowledge Jeff Ludwig, Kevin McConway, and Hyungjune Im who identified the majority of these issues in their careful readings of the book. If you find any further errors, we would very much appreciate your letting us know so that we can update this document.

Chapter 2, page 34, definition of “Correlation coefficient”: “product of the **variances**” Should be “product of the **standard deviations**”.

Chapter 3, page 45, third to last paragraph: “Individuals 1, 3, 4, **5**, and 8 are always-sick types” Should be “1, 3, 4, and 8 are always-sick types”.

Chapter 5, page 90, second paragraph: “That term was coined by the **late eighteenth-century** scholar, Francis Galton.” Should be “**nineteenth-century**”.

Chapter 6, page 97, first paragraph of the Noise section: “Suppose your **estimand** is unbiased.” Should be: “**estimator**”.

Chapter 6, page 103, second paragraph: “We have to admit a larger range of possible Republicans if we want to be sure **our estimate** is within that range 99 percent of the time rather than just 95 percent of the time.” Should be: “**the estimand**”.

Chapter 7, pages 116: We write “David **Spiegelharter**”, but the correct name is “David **Spiegelhalter**”. This mistake is repeated in the index on page 377. In addition, we refer

to the relevant analysis as having been carried out by “two mathematicians, Chris Budd and David Spiegelhalter”; Professor Budd is a mathematician, but Professor Spiegelhalter is a statistician (who, incidentally, has written a very nice book on learning from data).

Chapter 8, page 139: We say Galton discovered mean reversion in the 1860s. While Galton’s first studies of heredity were indeed in the 1860s, his discovery of mean reversion and regression were not until the study of sweet peas that was published in 1877.

Chapter 8, page 140, Figure 8.1: The text at the bottom of the figure says, “When Mid-Parents are **taller** than mediocrity, their Children tend to be taller than they.” Should be: “**shorter**”.

Chapter 9, page 170: The would be golfer’s name is Dan **McLaughlin**.

Chapter 10, page 204, second paragraph “That number will be a weighted average of the slopes of the two gray lines in **figure 10.11**.” Should be: “**figure 10.6**”.

Chapter 15, page 319, fourth paragraph of Abe’s Celiac Revisited section: “That test has a false positive rate of 50 percent—that is, $\Pr(\text{Positive on Test 2} \mid \text{Celiac}) = .5$.” Should be: “ $\Pr(\text{Positive on Test 2} \mid \text{No Celiac}) = .5$ ”

Chapter 15, 321, chapter 15, first full paragraph: “This happens with probability $.99 \times .5 = .495$.” Should be: “ $.95 \times .5 = .475$ ”. This error carries through to the two displayed equations below, which should be:

$$\begin{aligned} \Pr(\text{Neg on Test 1 \& Pos on Test 2}) &= \Pr(\text{Celiac}) \Pr(\text{Neg on Test 1 \& Pos on Test 2} \mid \text{Celiac}) \\ &\quad + \Pr(\text{No Celiac}) \Pr(\text{Neg on Test 1 \& Pos on Test 2} \mid \text{No Celiac}) \\ &= 0.01 \times 0.04 + 0.99 \times \mathbf{0.475} \\ &= \mathbf{0.47065}. \end{aligned}$$

and

$$\begin{aligned}
\Pr(\text{Celiac} \mid \text{Neg on Test 1 \& Pos on Test 2}) &= \frac{\Pr(\text{Neg on Test 1 \& Pos on Test 2} \mid \text{Celiac}) \Pr(\text{Celiac})}{\Pr(\text{Neg on Test 1 \& Pos on Test 2})} \\
&= \frac{\mathbf{0.04} \times 0.01}{\mathbf{0.47065}} \\
&\approx 0.001
\end{aligned}$$