Estimation

Contents

1 Estimation

1.1 Probabilistic models

Definition 1.1 (Scientific models). **Scientific models** are attempts to describe *physical conditions or changes* that occur in the world and universe around us.

Example 1.1 (Scientific models in epidemiology). Epidemiologists typically study *biological conditions and changes*, such as the spread of infectious diseases through populations, or the effects of environmental factors on individuals.

1.1.1 All models are wrong, some are useful

Box and Draper (1987), p424 (emphasis added):

> ...Essentially, all models are wrong, but some are useful. **However, the approximate nature of the model must always be borne in mind.**

see also Dunn and Smyth (2018), §1.8

1.1.2 Statistical analysis of scientific models

When we perform statistical analyses, we use data to help us choose between models - specifically, to determine which models best explain that data.

However, physical processes do not produce data on their own. Data is only produced when scientists implement an *observation process* (i.e., a *scientific study*), which is distinct from the underlying

physical process. In some cases, the observation process and the physical process interact with each other. This phenomenon is called the "observer effect"[1].

In order to learn about the physical processes we are ultimately interested in, we often need to make special considerations for the observation process that produced the data which we are analyzing. In particular, if some of the planned observations in the study design were not completed, we will likely need to account for the incompleteness of the resulting data set in our analysis. If we are not sure why some observations are incomplete, we may need to model the observation process in addition to the physical process we were originally interested in. For example, if some participants in a study dropped out part-way through the study, we may need investigate why those participants dropped out, as opposed to other participants who completed the study.

These kinds of *missing data* issues are outside of the scope of this course; see Van Buuren (2018) for more details.

1.2 Estimands, estimates, and estimators

1.2.1 Estimands

Definition 1.2 (Estimand)**.** An **estimand** is an unknown quantity whose value we want to know (Pohl et al. 2021; Lawrance et al. 2020).

Example 1.2 (Mean height of students)**.** If we are trying to determine the mean height of students at our school, then the *population mean* is our **estimand**.

In statistical contexts, most estimands are parameters of probabilistic models, or functions of model parameters.

> **i** Notation for estimands
>
> Model paramaters and other estimands are often symbolized using lower-case Greek letters: $\alpha, \beta, \gamma, \delta$, etc.

1.2.2 Estimates

Definition 1.3 (Estimate/estimated value)**.** In statistics, an **estimate** or **estimated value** is an informed guess of an estimand's value, based on observed data.

Example 1.3 (Mean height of students)**.** Suppose we measure the heights of 50 random students from our school, and the sample mean was 175cm. We might use 175cm as an *estimate* of the population mean.

1.2.3 Estimators

Definition 1.4 (Estimator)**.** An **estimator** is a function $\hat{\theta}(x_1, ...x_n)$ that transforms data $x_1, ...x_n$ into an estimate.

> **i** Estimators are random variables
>
> When estimators are applied to random variables, the estimators are also random variables.

> **i** Notation for estimators
>
> Estimators are often symbolized by placing a ⌢ ("hat") symbol on top of the corresponding estimand; for example, $\hat{\theta}$.
> Usually, their dependence on the data is implicit:
>
> $$\hat{\theta} \stackrel{\text{def}}{=} \hat{\theta}(x_1, ...x_n)$$

---

[1] https://en.wikipedia.org/wiki/Observer_effect

Example 1.4 (Mean height of students). If we want to estimate the mean height of students at our university, which we will represent as μ, we might measure the heights of $n = 50$ randomly sampled students as random variables $X_1, ..., X_n$. Then we could use the function

$$\hat{\mu}(X_1, ..., X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i \stackrel{\text{def}}{=} \bar{X}$$

as an *estimator* to produce an *estimate* $\hat{\mu} = \bar{x}$ of μ.

Another estimator would be just the height of the first student sampled:

$$\hat{\mu}^{(2)}(X_1, ..., X_n) = X_1$$

A third possible estimator would be the mean of all sampled students' heights, except for the two most extreme; that is, if we re-order the observations $X_{(1)} = \min_{i \in 1:n} X_i$, $X_{(2)} = \min_{i \in \{1:n\} - \arg X_{(1)}} X_i$, ..., $X_{(n)} = \max_{i \in 1:n} X_i$, then we could define the estimator:

$$\hat{\mu}^{(3)}(X_1, ..., X_n) = \frac{1}{n} \sum_{i=2}^{n-1} X_{(i)}$$

Which of these estimators is best? It depends on how we evaluate them (see Section 1.3 below).

1.2.4 Contrasting estimands, estimates, and estimators

It's helpful to keep in mind the mathematical type of each estimation concept:

- estimands are numbers (or vector of numbers)
- estimates are also numbers (or vectors)
- estimators are functions of random variables, so they are also random variables

1.3 Accuracy of estimators

1.3.1 Accuracy

To determine which estimator is best, we need to define *best*. "Accuracy" is usually most important; easy computation is usually secondary.

Definition 1.5 (Accuracy). The **accuracy** of an estimator for a given estimand does not have a consensus formal definition, but all of the usual candidates are related to the distributions of the *errors* made by the resulting estimates.

1.3.2 Error

Definition 1.6 (Error). The **error** of an estimate $\hat{\theta}$ of a true value θ, often denoted $\varepsilon(\hat{\theta})$, or more completely $\varepsilon(\hat{\theta}, \theta)$, is the difference between the estimate and its estimand θ; that is:

$$\varepsilon(\hat{\theta}) \stackrel{\text{def}}{=} \hat{\theta} - \theta$$

Some frequently-used measures of accuracy include:

1.3.3 Mean squared error

Definition 1.7 (Mean squared error). The **mean squared error** of an estimator $\hat{\theta}$, denoted $\mathrm{MSE}(\hat{\theta})$, is the expectation of the square of the error[2]:

$$\mathrm{MSE}(\hat{\theta}) \stackrel{\text{def}}{=} \mathrm{E}\left[(\varepsilon(\hat{\theta}))^2\right]$$

---

[2]https://en.wikipedia.org/wiki/Does_exactly_what_it_says_on_the_tin

1.3.4 Mean absolute error

Definition 1.8 (Mean absolute error)**.** The **mean absolute error** of an estimator is the expectation of the absolute value of the error:

$$\mathrm{MAE}\big(\hat{\theta}\big) \stackrel{\text{def}}{=} \mathrm{E}\big[\big|\varepsilon(\hat{\theta})\big|\big]$$

1.3.5 Bias

Definition 1.9 (Bias)**.** The **bias** of an estimator $\hat{\theta}$ for an estimand θ is the expected value of the error:

$$\mathrm{Bias}\big(\hat{\theta}\big) \stackrel{\text{def}}{=} \mathrm{E}\big[\varepsilon(\hat{\theta})\big] \tag{1}$$

Theorem 1.1 (Bias equals Expectation minus Truth)**.**

$$Bias\big(\hat{\theta}\big) = E\big[\hat{\theta}\big] - \theta$$

Proof.

$$\begin{aligned}
\mathrm{Bias}\big(\hat{\theta}\big) &\stackrel{\text{def}}{=} \mathrm{E}\big[\varepsilon(\hat{\theta})\big] \\
&= \mathrm{E}\big[\hat{\theta} - \theta\big] \\
&= \mathrm{E}\big[\hat{\theta}\big] - \mathrm{E}[\theta] \\
&= \mathrm{E}\big[\hat{\theta}\big] - \theta
\end{aligned}$$

The third equality is by the linearity of expectation. $\square$

Theorem 1.2 (Mean Squared Error equals Bias Squared plus Variance)**.** *For any one-dimensional estimator $\hat{\theta}$:*

$$MSE\big(\hat{\theta}\big) = \big(Bias\big(\hat{\theta}\big)\big)^2 + Var\big(\hat{\theta}\big) \tag{2}$$

Proof. Let's start by expanding each term of the right-hand side:

$$\begin{aligned}
\big(\mathrm{Bias}\big(\hat{\theta}\big)\big)^2 &= \big(\mathrm{E}\big[\hat{\theta}\big] - \theta\big)^2 \\
&= \big(\mathrm{E}\big[\hat{\theta}\big]\big)^2 - 2\mathrm{E}\big[\hat{\theta}\big]\theta + \theta^2
\end{aligned}$$

$$\mathrm{Var}\big(\hat{\theta}\big) = \mathrm{E}\big[\hat{\theta}^2\big] - \big(\mathrm{E}\big[\hat{\theta}\big]\big)^2$$

Now, add them together and simplify:

$$\begin{aligned}
\big(\mathrm{Bias}\big(\hat{\theta}\big)\big)^2 + \mathrm{Var}\big(\hat{\theta}\big) &= \big(\mathrm{E}\big[\hat{\theta}\big]\big)^2 - 2\mathrm{E}\big[\hat{\theta}\big]\theta + \theta^2 + \mathrm{E}\big[\hat{\theta}^2\big] - \big(\mathrm{E}\big[\hat{\theta}\big]\big)^2 \\
&= \mathrm{E}\big[\hat{\theta}^2\big] - 2\mathrm{E}\big[\hat{\theta}\big]\theta + \theta^2
\end{aligned}$$

Now let's expand the left-hand side to reach the same expression:

$$\text{MSE}(\hat{\theta}) = \text{E}\left[(\text{e}^{(}\hat{\theta}))^2\right]$$
$$= \text{E}\left[(\hat{\theta} - \theta)^2\right]$$
$$= \text{E}\left[\hat{\theta}^2 - 2\hat{\theta}\theta - \theta^2\right]$$
$$= \text{E}\left[\hat{\theta}^2\right] - \text{E}\left[2\hat{\theta}\theta\right] + \text{E}[\theta^2]$$
$$= \text{E}\left[\hat{\theta}^2\right] - 2\text{E}\left[\hat{\theta}\right]\theta + \theta^2$$

$\text{MSE}(\hat{\theta})$ and $\left(\text{Bias}(\hat{\theta})\right)^2 + \text{Var}(\hat{\theta})$ both equal $\text{E}\left[\hat{\theta}^2\right] - 2\text{E}\left[\hat{\theta}\right]\theta + \theta^2$. Equality is transitive, so $\text{MSE}(\hat{\theta})$ and $\left(\text{Bias}(\hat{\theta})\right)^2 + \text{Var}(\hat{\theta})$ are equal to each other:

$$\text{MSE}(\hat{\theta}) = \left(\text{Bias}(\hat{\theta})\right)^2 + \text{Var}(\hat{\theta})$$

$\square$

Unbiased estimators

Definition 1.10 (unbiased estimator). An estimator $\hat{\theta}$ is **unbiased** if $\text{Bias}(\hat{\theta}) = 0$.

Theorem 1.3 (properties of unbiased estimators). *If $\hat{\theta}$ is unbiased, then:*

$$E[\hat{\theta}] = \theta \qquad (3)$$
$$MSE(\hat{\theta}) = Var(\hat{\theta}) \qquad (4)$$

Proof. If $\hat{\theta}$ is unbiased, then:

Equation 3:

$$\text{Bias}(\hat{\theta}) = 0$$
$$\text{E}[\hat{\theta}] - \theta = 0$$
$$\text{E}[\hat{\theta}] = \theta$$

Equation 4:

$$\text{MSE}(\hat{\theta}) \overset{\text{def}}{=} \text{E}\left[\left(\varepsilon(\hat{\theta})\right)^2\right]$$
$$= \text{E}\left[\left(\hat{\theta} - \theta\right)^2\right]$$
$$= \text{E}\left[\left(\hat{\theta} - \text{E}[\hat{\theta}]\right)^2\right]$$
$$\overset{\text{def}}{=} \text{Var}(\hat{\theta})$$

(*Alternative proof of Equation 4*) We could have started from Theorem 1.2 instead:

$$\text{MSE}(\hat{\theta}) = \left(\text{Bias}(\hat{\theta})\right)^2 + \text{Var}(\hat{\theta})$$
$$= (0)^2 + \text{Var}(\hat{\theta})$$
$$= 0 + \text{Var}(\hat{\theta})$$
$$= \text{Var}(\hat{\theta})$$

$\square$

1.3.6 Standard error

Definition 1.11 (Standard error)**.** The **standard error** of an estimator $\hat{\theta}$ is just the standard deviation[3] of $\hat{\theta}$; that is:

$$\text{SE}\big(\hat{\theta}\big) \overset{\text{def}}{=} \text{SD}\big(\hat{\theta}\big)$$

"Standard error" is a confusing concept in a few ways. First of all, it isn't even defined as a characteristic of the error, $\varepsilon(\hat{\theta})$! Moreover, it is just a synonym for standard deviation, so it seems like a redundant concept. However, standard errors help us construct p-values and confidence intervals, so they come up a lot - often enough to give them their own name.

We can relate standard error to actual error, by rearranging the result from Theorem 1.2:

$$\begin{aligned}\text{Var}\big(\hat{\theta}\big) &= \text{Var}\big(\hat{\theta} - \theta\big) \\ &= \text{Var}\big(\varepsilon(\hat{\theta})\big)\end{aligned}$$

So the variance of the estimator is equal to the variance of the error, and the standard error is equal to the standard deviation of the error:

$$\text{SE}\big(\hat{\theta}\big) = \text{SD}\big(\varepsilon(\hat{\theta})\big)$$

Corollary 1.1 (Standard error squared equals MSE minus squared bias)**.** *standard error is what is left over of MSE after bias is removed:*

$$\big(SE\big(\hat{\theta}\big)\big)^2 = MSE\big(\hat{\theta}\big) - \big(Bias\big(\hat{\theta}\big)\big)^2$$

Proof.

$$\begin{aligned}\text{MSE}\big(\hat{\theta}\big) &= \big(\text{Bias}\big(\hat{\theta}\big)\big)^2 + \text{Var}\big(\hat{\theta}\big) \\ \therefore \text{Var}\big(\hat{\theta}\big) &= \text{MSE}\big(\hat{\theta}\big) - \big(\text{Bias}\big(\hat{\theta}\big)\big)^2 \\ \therefore \big(\text{SE}\big(\hat{\theta}\big)\big)^2 &= \text{MSE}\big(\hat{\theta}\big) - \big(\text{Bias}\big(\hat{\theta}\big)\big)^2\end{aligned}$$

$\square$

Corollary 1.2 (For unbiased estimators, SE = RMSE)**.** *If $E\big[\varepsilon\big(\hat{\theta}\big)\big] = 0$, then:*

$$SE\big(\hat{\theta}\big) = \sqrt{MSE\big(\hat{\theta}\big)}$$

(this result is equivalent to Equation 4)

Box, George E. P., and Norman Richard. Draper. 1987. *Empirical Model-Building and Response Surfaces.* Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. New York: Wiley.

Dunn, Peter K, and Gordon K Smyth. 2018. *Generalized Linear Models with Examples in R.* Vol. 53. Springer. https://link.springer.com/book/10.1007/978-1-4419-0118-7.

Lawrance, Rachael, Evgeny Degtyarev, Philip Griffiths, Peter Trask, Helen Lau, Denise D'Alessio, Ingolf Griebsch, Gudrun Wallenstein, Kim Cocks, and Kaspar Rufibach. 2020. "What Is an Estimand, and How Does It Relate to Quantifying the Effect of Treatment on Patient-Reported Quality of Life Outcomes in Clinical Trials?" *Journal of Patient-Reported Outcomes* 4 (1): 1–8. https://link.springer.com/article/10.1186/s41687-020-00218-5.

---

[3]probability.qmd#def-sd

Pohl, Moritz, Lukas Baumann, Rouven Behnisch, Marietta Kirchner, Johannes Krisam, and Anja Sander. 2021. "Estimands—A Basic Element for Clinical Trials." *Deutsches Ärzteblatt International* 118 (51-52): 883–88. https://doi.org/10.3238/arztebl.m2021.0373.

Van Buuren, Stef. 2018. *Flexible Imputation of Missing Data.* CRC press. https://stefvanbuuren.name/fimd/.