Estimation

Last modified: 2026-11-12 22:02:32 (PST)

Scientific models

Definition 0.1 (Scientific models)**. Scientific models** are attempts to describe *physical conditions or changes* that occur in the world and universe around us.

Example 0.1 (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.

Models as approximations

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

[Box and Draper [1], p. 424; emphasis added]

See also [2, Sec. 1.8].

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 those data.

Physical processes do not produce data on their own. Data are only produced when scientists implement an *observation process* (that is, 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 interaction is called the "observer effect".

To learn about the physical processes we are ultimately interested in, we often need to account for the observation process that produced the data 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, we may need to investigate why those participants dropped out, as opposed to other participants who completed the study.

These kinds of *missing data* issues are outside the scope of these notes; see Van Buuren [3] for more details.

Estimands, estimates, and estimators

Estimands

Definition 0.1 (Estimand)**.** An **estimand** is an unknown quantity θ whose value we want to know [4], [5].
Source: Wikipedia contributors [6].

Example 0.1 (Mean height of students)**.** If we are trying to determine the population 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 parameters and other estimands are often symbolized using lower-case Greek letters. The notation choices page gives our convention and the alternatives.

Estimates

Definition 0.2 (Estimate/estimated value)**.** In statistics, an **estimate** or **estimated value** $\hat{\theta}$ is an informed guess of an estimand θ's value, computed from the observed data $x_1, \ldots, x_n$:

$$\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n).$$

Example 0.2 (Mean height of students)**.** Suppose we measure the heights of 50 randomly sampled students from our school, and their sample mean is 175 cm. We might use 175 cm as an *estimate* of the population mean.

Estimators

Definition 0.3 (Estimator)**.** An **estimator** is the function $\hat{\theta}(\cdot)$ that transforms data into an estimate; applied to a random sample $X_1, \ldots, X_n$, it is a random variable:

$$\hat{\theta}(X_1, \ldots, X_n).$$

Source: Wikipedia contributors [7].

i Estimators are random variables

When an estimator is applied to random variables $X_1, \ldots, X_n$ rather than to their observed values, the result $\hat{\theta}(X_1, \ldots, X_n)$ is also a random variable.

i Notation for estimators

Estimators are often symbolized by placing a hat over the symbol of the corresponding estimand; for example, $\hat{\theta}$. The notation choices page gives our convention and the alternatives.
Usually, their dependence on the data is implicit:

$$\hat{\theta} \stackrel{\text{def}}{=} \hat{\theta}(x_1, \ldots, x_n)$$

Example 0.3 (Mean height of students)**.** Suppose we want to estimate the population mean height of students at our school, which we will represent as μ, and we measure the heights of $n = 50$ randomly sampled students as random variables $X_1, \ldots, X_n$. Then we could use the function

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

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:

$$\hat{\mu}^{(2)}(X_1, \ldots, X_n) \stackrel{\text{def}}{=} X_1$$

A third possible estimator would be the sample mean of all sampled students' heights,

except for the two most extreme. Using the order statistics $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ (the observations sorted in increasing order), this estimator drops $X_{(1)}$ and $X_{(n)}$ and averages the remaining $n-2$ observations:

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

Which of these estimators is best? The answer depends on how we evaluate them (see Section).

Definition 0.4 (Sampling distribution)**.** The **sampling distribution** of an estimator $\hat{\theta}$ is the probability distribution of $\hat{\theta}(X_1, \ldots, X_n)$ over repeated random samples $X_1, \ldots, X_n$ from the same population: the distribution of the values the estimator would take across all the samples we might have drawn.
Source: Wikipedia contributors [8].

Example 0.4 (Sampling distribution of a sample mean)**.** Suppose student heights are $X_i \sim_{\text{iid}} \mathrm{N}(\mu, \sigma^2)$ with standard deviation $\sigma = 10$ cm, and we sample $n = 25$ students. The sample mean $\bar{X}$ is a linear combination of independent Gaussian variables, so it is also Gaussian, with mean $\mathbb{E}[\bar{X}] = \mu$ and variance $\mathrm{Var}(\bar{X}) = \sigma^2/n$. Its sampling distribution is therefore

$$\bar{X} \sim \mathrm{N}(\mu, \sigma^2/25),$$

which has standard deviation $\sigma/\sqrt{25} = 10/5 = 2$ cm. Across repeated samples of 25 students, about 95% of the sample means land within $2 \times 2 = 4$ cm of the true mean μ.

Contrasting estimands, estimates, and estimators

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

- estimands are numbers (or vectors of numbers);
- estimates are also numbers (or vectors);
- estimators are functions; an estimator applied to random data, $\hat{\theta}(X_1, \ldots, X_n)$, is a random variable.

i Read

- Hogg et al. [9, Sec. 6.4], pp. 264–276 covers maximum likelihood and method of moments estimation.

- Casella and Berger [10, Sec. 7.2], pp. 312–329 covers methods of finding estimators.
- Rice [11, Secs. 8.3–8.4], pp. 257–266 covers parameter estimation and the method of moments.

Accuracy of estimators

Accuracy

To determine which estimator is best, we need to define *best*. Accuracy is usually most important, and ease of computation is usually secondary.

Estimation error

Definition 0.1 (Estimation error)**.** The **estimation error** of an estimate $\hat{\theta}$ of a true value θ is the difference between the estimate and the estimand θ:

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

Example 0.1 (Estimation error of a mean height)**.** Continuing Example 0.2, suppose the population mean height of students at our school were actually $\mu = 172$ cm. Then the estimate $\hat{\mu} = 175$ cm would have estimation error:

$$\begin{aligned} \varepsilon(\hat{\mu}) &= \hat{\mu} - \mu && \text{(definition of estimation error)} \\ &= 175 - 172 && \text{(substitute the estimate and the estimand)} \\ &= 3 \text{ cm} && \text{(subtract)} \end{aligned}$$

In practice we never observe an estimation error, because we do not know the estimand's value; if we did, we would not need to estimate it.

The accuracy of an estimator has no single, agreed formal definition. The usual measures of accuracy, including the bias, population mean squared error, and population mean absolute error defined in this section, are all functions of the distribution of the estimator's estimation error.

Predictions and residuals

Predictions, fitted values, residuals, prediction errors and the sample mean squared error of predictions have their own page, Prediction. That page also contrasts them with the estima-

tion quantities defined here: estimation error and the population mean squared error of an estimator.

Bias

Definition 0.2 (Bias)**.** The **bias** of an estimator $\hat{\theta}$ for an estimand θ is the expected value of the estimation error:

$$\text{Bias}\left(\hat{\theta}\right) \stackrel{\text{def}}{=} \mathbb{E}\left[\varepsilon\left(\hat{\theta}\right)\right] \tag{1}$$

Source: Wikipedia contributors [12].

Exercise 0.1 (Bias in terms of the expectation)**.** Derive an expression for the bias $\text{Bias}\left(\hat{\theta}\right)$ of an estimator $\hat{\theta}$ in terms of its expected value $\mathbb{E}\left[\hat{\theta}\right]$ and the estimand θ.

Solution

Solution 0.1.

$$\begin{aligned}
\text{Bias}\left(\hat{\theta}\right) &\stackrel{\text{def}}{=} \mathbb{E}\left[\varepsilon\left(\hat{\theta}\right)\right] && \text{(definition of bias)} \\
&= \mathbb{E}\left[\hat{\theta} - \theta\right] && \text{(definition of estimation error)} \\
&= \mathbb{E}\left[\hat{\theta}\right] - \mathbb{E}[\theta] && \text{(linearity of expectation)} \\
&= \mathbb{E}\left[\hat{\theta}\right] - \theta && (\theta \text{ is a constant})
\end{aligned}$$

Theorem 0.1 (Bias equals expectation minus truth)**.**

$$\text{Bias}\left(\hat{\theta}\right) = \mathbb{E}\left[\hat{\theta}\right] - \theta$$

Proof

Proof. This is the solution to Exercise 0.1. □

Example 0.2 (Bias of two estimators of a population mean)**.** Let $X_1, \ldots, X_n$ each have expectation $\mathbb{E}[X_i] = \mu$, as in Example 0.3. By Theorem 0.1, the bias of the sample mean $\bar{X}$ is:

$$
\begin{aligned}
\operatorname{Bias}(\bar{X}) &= \mathbb{E}[\bar{X}] - \mu && \text{(bias equals expectation minus truth)} \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] - \mu && \text{(definition of } \bar{X}\text{)} \\
&= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] - \mu && \text{(factor the constant } \tfrac{1}{n} \text{ out of the expectation)} \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] - \mu && \text{(linearity of expectation)} \\
&= \frac{1}{n}\sum_{i=1}^{n} \mu - \mu && (\mathbb{E}[X_i] = \mu \text{ for every } i) \\
&= \frac{1}{n} \cdot n\mu - \mu && \text{(sum of } n \text{ copies of } \mu\text{)} \\
&= \mu - \mu && \text{(cancel } n\text{)} \\
&= 0 && \text{(subtract)}
\end{aligned}
$$

and the bias of the single-observation estimator $\hat{\mu}^{(2)} = X_1$ is:

$$
\begin{aligned}
\operatorname{Bias}(X_1) &= \mathbb{E}[X_1] - \mu && \text{(bias equals expectation minus truth)} \\
&= \mu - \mu && (\mathbb{E}[X_1] = \mu) \\
&= 0 && \text{(subtract)}
\end{aligned}
$$

So both estimators have zero bias, even though $\bar{X}$ uses all n observations and X_1 uses only one.

Population mean squared error of an estimator

Definition 0.3 (Mean squared error)**.** The **population mean squared error (pMSE)** of an estimator $\hat{\theta}$, also called its **mean squared error** or its **expected squared estimation error (ESEE)**, denoted $\operatorname{pMSE}\left(\hat{\theta}\right)$ or $\operatorname{ESEE}\left(\hat{\theta}\right)$, is the expectation of the square of the estimation error:

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

Source: Wikipedia contributors [13].

Exercise 0.2 (Expanding the squared bias)**.** Using Theorem 0.1, expand the squared bias $\left(\operatorname{Bias}\left(\hat{\theta}\right)\right)^2$ in terms of $\mathbb{E}\left[\hat{\theta}\right]$ and the estimand θ.

Solution

Solution 0.2.

$$\begin{aligned}\left(\text{Bias}\left(\hat{\theta}\right)\right)^2 &= \left(\mathbb{E}\left[\hat{\theta}\right] - \theta\right)^2 && \text{(bias equals expectation minus truth)}\\ &= \left(\mathbb{E}\left[\hat{\theta}\right]\right)^2 - 2\mathbb{E}\left[\hat{\theta}\right]\theta + \theta^2 && \text{(expand the binomial square)}\end{aligned}$$

Theorem 0.2 (Expanded squared bias)**.** *For any one-dimensional estimator* $\hat{\theta}$ *of an estimand* θ*, the squared bias is:*

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

Proof

Proof. This is the solution to Exercise 0.2, worked in Solution 0.2. □

Exercise 0.3 (Squared bias plus variance)**.** Using Exercise 0.2, write $\left(\text{Bias}\left(\hat{\theta}\right)\right)^2 + \text{Var}\left(\hat{\theta}\right)$ in terms of $\mathbb{E}\left[\hat{\theta}^2\right]$, $\mathbb{E}\left[\hat{\theta}\right]$, and the estimand θ.

Solution

Solution 0.3. The variance is (simplified expression for variance):

$$\text{Var}\left(\hat{\theta}\right) = \mathbb{E}\left[\hat{\theta}^2\right] - \left(\mathbb{E}\left[\hat{\theta}\right]\right)^2 \tag{2}$$

Let the population mean of the estimator $\hat{\theta}$ be $\mu_{\hat{\theta}} = \mathbb{E}\left[\hat{\theta}\right]$. Then the variance of $\hat{\theta}$ is $\mathbb{E}\left[\hat{\theta}^2\right] - \left(\mu_{\hat{\theta}}\right)^2$, and the squared bias of $\hat{\theta}$ for the estimand θ from Exercise 0.2 is $\left(\mu_{\hat{\theta}}\right)^2 - 2\mu_{\hat{\theta}}\theta + \theta^2$. Add them and simplify:

$$
\begin{aligned}
\left(\operatorname{Bias}\left(\hat{\theta}\right)\right)^2 + \operatorname{Var}\left(\hat{\theta}\right) &= \left(\mu_{\hat{\theta}}\right)^2 - 2\mu_{\hat{\theta}}\theta + \theta^2 + \operatorname{Var}\left(\hat{\theta}\right) && \text{(substitute the squared bias)} \\
&= \left(\mu_{\hat{\theta}}\right)^2 - 2\mu_{\hat{\theta}}\theta + \theta^2 + \mathbb{E}\left[\hat{\theta}^2\right] - \left(\mu_{\hat{\theta}}\right)^2 && \text{(substitute the variance)} \\
&= \mathbb{E}\left[\hat{\theta}^2\right] - 2\mu_{\hat{\theta}}\theta + \theta^2 + \left(\mu_{\hat{\theta}}\right)^2 - \left(\mu_{\hat{\theta}}\right)^2 && \text{(reorder the terms)} \\
&= \mathbb{E}\left[\hat{\theta}^2\right] - 2\mu_{\hat{\theta}}\theta + \theta^2 + \left(\left(\mu_{\hat{\theta}}\right)^2 - \left(\mu_{\hat{\theta}}\right)^2\right) && \text{(group the last two terms)} \\
&= \mathbb{E}\left[\hat{\theta}^2\right] - 2\mu_{\hat{\theta}}\theta + \theta^2 + 0 && (a - a = 0) \\
&= \mathbb{E}\left[\hat{\theta}^2\right] - 2\mu_{\hat{\theta}}\theta + \theta^2 && (a + 0 = a)
\end{aligned}
\tag{3}
$$

Theorem 0.3 (Expanded squared bias plus variance)**.** *For any one-dimensional estimator $\hat{\theta}$ of an estimand θ, the squared bias plus the variance is:*

$$
\left(\operatorname{Bias}\left(\hat{\theta}\right)\right)^2 + \operatorname{Var}\left(\hat{\theta}\right) = \mathbb{E}\left[\hat{\theta}^2\right] - 2\mathbb{E}\left[\hat{\theta}\right]\theta + \theta^2
$$

Proof

Proof. This is the solution to Exercise 0.3, worked in Solution 0.3. It adds the variance to the squared bias of Theorem 0.2. □

Exercise 0.4 (Expanding the population mean squared error)**.** Derive an expression for the pMSE pMSE$\left(\hat{\theta}\right)$ in terms of $\mathbb{E}\left[\hat{\theta}^2\right]$, $\mathbb{E}\left[\hat{\theta}\right]$, and the estimand θ.

Solution

Solution 0.4.

$$
\begin{aligned}
\operatorname{pMSE}\left(\hat{\theta}\right) &\overset{\text{def}}{=} \mathbb{E}\left[\left(\varepsilon\left(\hat{\theta}\right)\right)^2\right] && \text{(definition of MSE)} \\
&= \mathbb{E}\left[\left(\hat{\theta} - \theta\right)^2\right] && \text{(definition of estimation error)} \\
&= \mathbb{E}\left[\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2\right] && \text{(expand the binomial square)} \\
&= \mathbb{E}\left[\hat{\theta}^2\right] - \mathbb{E}\left[2\hat{\theta}\theta\right] + \mathbb{E}[\theta^2] && \text{(linearity of expectation)} \\
&= \mathbb{E}\left[\hat{\theta}^2\right] - 2\mathbb{E}\left[\hat{\theta}\right]\theta + \mathbb{E}[\theta^2] && \text{(factor the constant } 2\theta \text{ out of the middle expectation)} \\
&= \mathbb{E}\left[\hat{\theta}^2\right] - 2\mathbb{E}\left[\hat{\theta}\right]\theta + \theta^2 && \text{(the expectation of a constant is that constant)}
\end{aligned}
$$

Theorem 0.4 (Expanded population mean squared error)**.** *For any one-dimensional estimator $\hat{\theta}$ of an estimand θ, the pMSE is:*

$$\operatorname{pMSE}\left(\hat{\theta}\right) = \mathbb{E}\left[\hat{\theta}^2\right] - 2\mathbb{E}\left[\hat{\theta}\right]\theta + \theta^2$$

Proof

Proof. This is the solution to Exercise 0.4, worked in Solution 0.4. □

Theorem 0.5 (Population mean squared error equals bias squared plus variance)**.** *For any one-dimensional estimator $\hat{\theta}$:*

$$\operatorname{pMSE}\left(\hat{\theta}\right) = \left(\operatorname{Bias}\left(\hat{\theta}\right)\right)^2 + \operatorname{Var}\left(\hat{\theta}\right) \tag{4}$$

Proof (Shorter proof)

Shorter proof. Apply the second moment is variance plus squared mean lemma to the estimation error $Z = \varepsilon\left(\hat{\theta}\right)$:

$$\begin{aligned}
\operatorname{pMSE}\left(\hat{\theta}\right) &= \mathbb{E}\left[\left(\varepsilon\left(\hat{\theta}\right)\right)^2\right] && \text{(definition of pMSE)} \\
&= \operatorname{Var}\left(\varepsilon\left(\hat{\theta}\right)\right) + \left(\mathbb{E}\left[\varepsilon\left(\hat{\theta}\right)\right]\right)^2 && \text{(second moment is variance plus squared mean)} \\
&= \operatorname{Var}\left(\varepsilon\left(\hat{\theta}\right)\right) + \left(\operatorname{Bias}\left(\hat{\theta}\right)\right)^2 && \text{(definition of bias)} \\
&= \operatorname{Var}\left(\hat{\theta} - \theta\right) + \left(\operatorname{Bias}\left(\hat{\theta}\right)\right)^2 && \text{(definition of estimation error)} \\
&= \operatorname{Var}\left(\hat{\theta}\right) + \left(\operatorname{Bias}\left(\hat{\theta}\right)\right)^2 && \text{(subtracting a constant does not change a variance)}
\end{aligned}$$

Unlike the expansion route, this proof never expands the square.
Source: Wikipedia contributors [13]. □

Proof

Proof. By Theorem 0.4 and Theorem 0.3, $\operatorname{pMSE}\left(\hat{\theta}\right)$ and $\left(\operatorname{Bias}\left(\hat{\theta}\right)\right)^2 + \operatorname{Var}\left(\hat{\theta}\right)$ both equal $\mathbb{E}\left[\hat{\theta}^2\right] - 2\mathbb{E}\left[\hat{\theta}\right]\theta + \theta^2$, so they are equal to each other. □

Remark (Estimation versus prediction)

Remark 0.1 (Estimation versus prediction). The squared-error split has two forms, and they differ in what the estimate is compared with.

- **Estimation.** Theorem 0.5 compares an estimator $\hat{\theta}$ with a fixed parameter θ. The target is a constant, so the expected squared error is squared bias plus variance, with no other term.
- **Prediction.** The target is a new outcome, which is itself random. The model validation notes prove that expected squared prediction error equals squared bias plus variance plus the noise variance, the residual variance of that outcome. No choice of prediction using the same covariates can remove the extra term.

Example 0.3 (Population mean squared error of two estimators of a population mean)**.** Continuing Example 0.2, suppose also that $X_1, \dots, X_n$ are mutually independent, each with variance $\mathrm{Var}(X_i) = \sigma^2$. Both $\bar{X}$ and X_1 have zero bias, so by Theorem 0.5 each estimator's pMSE equals its variance.
For X_1,

$$\begin{aligned} \mathrm{pMSE}(X_1) &= 0^2 + \mathrm{Var}(X_1) \\ &= \sigma^2. \end{aligned}$$

For $\bar{X}$:

$$\begin{aligned} \mathrm{pMSE}(\bar{X}) &= \big(\mathrm{Bias}(\bar{X})\big)^2 + \mathrm{Var}(\bar{X}) && \text{(pMSE equals bias squared plus variance)} \\ &= 0^2 + \mathrm{Var}(\bar{X}) && \text{(zero bias)} \\ &= \mathrm{Var}(\bar{X}) && (0^2 = 0) \\ &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) && \text{(definition of } \bar{X}) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) && \text{(variance of a linear combination of independent variables)} \\ &= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 && (\mathrm{Var}(X_i) = \sigma^2 \text{ for every } i) \\ &= \frac{1}{n^2} \cdot n\sigma^2 && \text{(sum of } n \text{ copies of } \sigma^2) \\ &= \frac{\sigma^2}{n} && \text{(cancel one factor of } n) \end{aligned}$$

The step to $\frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i)$ uses the variance of a linear combination, whose covariance

terms are all zero for independent variables. The result σ^2/n is the variance of an IID sample mean.
So for any sample size $n > 1$,

$$\begin{aligned} \text{pMSE}\big(\bar{X}\big) &= \sigma^2/n \\ &< \sigma^2 \\ &= \text{pMSE}(X_1) : \end{aligned}$$

by mean squared error, the sample mean is the more accurate estimator. With $n = 50$ students, as in Example 0.3, the sample mean's mean squared error is 1/50 of X_1's.

Definition 0.4 (Root mean squared error)**.** The **population root mean squared error (pRMSE)** of an estimator $\hat{\theta}$, also called its **root mean squared error**, is the square root of its pMSE:

$$\text{pRMSE}\big(\hat{\theta}\big) \stackrel{\text{def}}{=} \sqrt{\text{pMSE}\big(\hat{\theta}\big)}$$

Source: Wikipedia contributors [14].

Example 0.4 (Population root mean squared error of the sample mean)**.** In Example 0.3, the pRMSE of $\bar{X}$ is $\sigma/\sqrt{n}$.

Remark (Units of the population root mean squared error)

Remark 0.2 (Units of the population root mean squared error). The pRMSE has the same units as the estimand.

Read

Casella and Berger [10, Sec. 7.3.1], pp. 330–333 covers mean squared error.

Unbiased estimators

Definition 0.5 (Unbiased estimator)**.** An estimator $\hat{\theta}$ is **unbiased** if its bias is zero:

$$\text{Bias}\big(\hat{\theta}\big) = 0$$

Definition 0.6 (Biased estimator)**.** An estimator $\hat{\theta}$ is **biased** if its bias is not zero:

$$\text{Bias}\big(\hat{\theta}\big) \neq 0$$

Example 0.5 (The sample mean is unbiased)**.** By Example 0.2, $\text{Bias}\big(\bar{X}\big) = 0$ and $\text{Bias}(X_1) = 0$, so both the sample mean $\bar{X}$ and the single-observation estimator X_1 are unbiased estimators of μ.

Example 0.6 (A biased estimator of the variance)**.** Let $X_1, \ldots, X_n$ be mutually independent, each with expectation μ and variance σ^2, and let $n \geq 2$. Consider two estimators of σ^2:

- the sample variance $S^2 \overset{\text{def}}{=} \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$;
- the divide-by-n estimator $\hat{\sigma}^2 \overset{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Both estimators are built from the sum of squared deviations $\sum_{i=1}^{n}(X_i - \bar{X})^2$, so we first find its expectation. Using

$$\begin{aligned}\mathbb{E}\big[X_i^2\big] &= \text{Var}(X_i) + \big(\mathbb{E}[X_i]\big)^2 \\ &= \sigma^2 + \mu^2\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\big[\bar{X}^2\big] &= \text{Var}\big(\bar{X}\big) + \big(\mathbb{E}\big[\bar{X}\big]\big)^2 \\ &= \sigma^2/n + \mu^2\end{aligned}$$

(by Example 0.3 and Example 0.2):

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right] &= \mathbb{E}\left[\sum_{i=1}^{n}\left(X_i^2-2X_i\bar{X}+\bar{X}^2\right)\right] && \text{(expand each square)}\\
&= \mathbb{E}\left[\sum_{i=1}^{n}X_i^2-\sum_{i=1}^{n}2X_i\bar{X}+\sum_{i=1}^{n}\bar{X}^2\right] && \text{(distribute the sum over the terms)}\\
&= \mathbb{E}\left[\sum_{i=1}^{n}X_i^2-2\bar{X}\sum_{i=1}^{n}X_i+\sum_{i=1}^{n}\bar{X}^2\right] && \text{(factor } 2\bar{X} \text{ out of the middle sum)}\\
&= \mathbb{E}\left[\sum_{i=1}^{n}X_i^2-2\bar{X}\sum_{i=1}^{n}X_i+n\bar{X}^2\right] && \text{(sum of } n \text{ copies of } \bar{X}^2)\\
&= \mathbb{E}\left[\sum_{i=1}^{n}X_i^2-2\bar{X}\cdot n\bar{X}+n\bar{X}^2\right] && \left(\textstyle\sum_{i=1}^{n}X_i=n\bar{X}\right)\\
&= \mathbb{E}\left[\sum_{i=1}^{n}X_i^2-2n\bar{X}^2+n\bar{X}^2\right] && \text{(multiply } 2\bar{X}\cdot n\bar{X})\\
&= \mathbb{E}\left[\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right] && \text{(collect the } \bar{X}^2 \text{ terms)}\\
&= \mathbb{E}\left[\sum_{i=1}^{n}X_i^2\right]-\mathbb{E}\left[n\bar{X}^2\right] && \text{(linearity of expectation)}\\
&= \sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right]-\mathbb{E}\left[n\bar{X}^2\right] && \text{(linearity of expectation)}\\
&= \sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right]-n\mathbb{E}\left[\bar{X}^2\right] && \text{(factor the constant } n \text{ out of the expectation)}\\
&= \sum_{i=1}^{n}(\sigma^2+\mu^2)-n\mathbb{E}\left[\bar{X}^2\right] && \text{(substitute } \mathbb{E}\left[X_i^2\right])\\
&= n(\sigma^2+\mu^2)-n\mathbb{E}\left[\bar{X}^2\right] && \text{(sum of } n \text{ copies of } \sigma^2+\mu^2)\\
&= n(\sigma^2+\mu^2)-n\left(\frac{\sigma^2}{n}+\mu^2\right) && \text{(substitute } \mathbb{E}\left[\bar{X}^2\right])\\
&= n\sigma^2+n\mu^2-n\left(\frac{\sigma^2}{n}+\mu^2\right) && \text{(distribute } n \text{ over } \sigma^2+\mu^2)\\
&= n\sigma^2+n\mu^2-\left(n\cdot\frac{\sigma^2}{n}+n\mu^2\right) && \text{(distribute } n \text{ over } \frac{\sigma^2}{n}+\mu^2)\\
&= n\sigma^2+n\mu^2-(\sigma^2+n\mu^2) && \text{(cancel } n \text{ in } n\cdot\frac{\sigma^2}{n})\\
&= n\sigma^2+n\mu^2-\sigma^2-n\mu^2 && \text{(distribute the minus sign)}\\
&= n\sigma^2-\sigma^2 && \text{(cancel } n\mu^2)\\
&= (n-1)\sigma^2 && \text{(factor out } \sigma^2)
\end{aligned}
$$

So

$$\begin{aligned}\mathbb{E}[S^2] &= \frac{1}{n-1}(n-1)\sigma^2 \\ &= \sigma^2,\end{aligned}$$

and by Theorem 0.1, $\text{Bias}(S^2) = 0$: the sample variance is unbiased.
For the divide-by-n estimator, $\mathbb{E}[\hat{\sigma}^2] = \frac{1}{n}(n-1)\sigma^2$, so:

$$\begin{aligned}\text{Bias}(\hat{\sigma}^2) &= \mathbb{E}[\hat{\sigma}^2] - \sigma^2 && \text{(bias equals expectation minus truth)} \\ &= \frac{n-1}{n}\sigma^2 - \sigma^2 && \text{(substitute } \mathbb{E}[\hat{\sigma}^2]\text{)} \\ &= \frac{n-1}{n}\sigma^2 - \frac{n}{n}\sigma^2 && \text{(write } \sigma^2 \text{ over the common denominator } n\text{)} \\ &= \frac{n-1-n}{n}\sigma^2 && \text{(combine the fractions)} \\ &= \frac{-1}{n}\sigma^2 && (n-1-n=-1) \\ &= -\frac{\sigma^2}{n} && \text{(multiply)}\end{aligned}$$

This estimator is biased: on average it underestimates σ^2, by an amount that shrinks to zero as n grows.

Exercise 0.5 (Expected value of an unbiased estimator)**.** Using Theorem 0.1, find $\mathbb{E}\left[\hat{\theta}\right]$ for an unbiased estimator $\hat{\theta}$ of θ.

Solution

Solution 0.5.

$$\begin{aligned}0 &= \text{Bias}\left(\hat{\theta}\right) && \text{(definition of unbiased)} \\ &= \mathbb{E}\left[\hat{\theta}\right] - \theta && \text{(bias equals expectation minus truth)}\end{aligned}$$

Adding θ to both sides gives $\mathbb{E}\left[\hat{\theta}\right] = \theta$.

Exercise 0.6 (Population mean squared error of an unbiased estimator)**.** Using Theorem 0.5, write the pMSE of an unbiased estimator $\hat{\theta}$ in terms of its variance.

Solution

Solution 0.6.

$$\begin{aligned} \mathrm{pMSE}\big(\hat{\theta}\big) &= \big(\mathrm{Bias}\big(\hat{\theta}\big)\big)^2 + \mathrm{Var}\big(\hat{\theta}\big) && \text{(MSE equals bias squared plus variance)} \\ &= 0^2 + \mathrm{Var}\big(\hat{\theta}\big) && \text{(definition of unbiased)} \\ &= \mathrm{Var}\big(\hat{\theta}\big) && (0^2 = 0) \end{aligned}$$

Theorem 0.6 (Properties of unbiased estimators)**.** *If $\hat{\theta}$ is an unbiased estimator of θ, then:*

$$\mathbb{E}\big[\hat{\theta}\big] = \theta \tag{5}$$

$$\mathrm{pMSE}\big(\hat{\theta}\big) = \mathrm{Var}\big(\hat{\theta}\big) \tag{6}$$

Proof

Proof. Equation 5 is the solution to Exercise 0.5, and Equation 6 is the solution to Exercise 0.6. □

Watch

5MinuteAI covers the properties of estimators, including bias, following the University of Washington's CSE 312 [15].

Population mean absolute error

Definition 0.7 (Mean absolute error)**.** The **population mean absolute error (pMAE)** of an estimator, also called its **mean absolute error**, is the expectation of the absolute value of the estimation error:

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

Source: Wikipedia contributors [16].

Example 0.7 (Population mean absolute error of a Gaussian estimator)**.** Suppose an estimator $\hat{\theta}$ has a Gaussian distribution with mean θ and standard deviation τ, so that its estimation error $\varepsilon\left(\hat{\theta}\right) = \hat{\theta} - \theta$ has the same distribution as τZ, where Z has a standard Gaussian distribution with density $\phi(z) = (2\pi)^{-1/2} e^{-z^2/2}$. Then:

$$
\begin{aligned}
\mathrm{MAE}\left(\hat{\theta}\right) &= \mathbb{E}[|\tau Z|] && \text{(definition of MAE)} \\
&= \mathbb{E}[\tau |Z|] && (|\tau Z| = \tau |Z| \text{ because } \tau > 0) \\
&= \tau \mathbb{E}[|Z|] && \text{(factor the constant } \tau \text{ out of the expectation)} \\
&= \tau \int_{-\infty}^{\infty} |z| \phi(z)\, dz && \text{(expectation of a function of } Z) \\
&= \tau \cdot 2 \int_{0}^{\infty} |z| \phi(z)\, dz && (|z|\phi(z) \text{ is symmetric about 0}) \\
&= \tau \cdot 2 \int_{0}^{\infty} z \phi(z)\, dz && (|z| = z \text{ for } z \geq 0) \\
&= \tau \cdot 2 [-\phi(z)]_{0}^{\infty} && (\phi'(z) = -z\phi(z)) \\
&= \tau \cdot 2 \left(\lim_{z \to \infty} [-\phi(z)] - (-\phi(0)) \right) && \text{(evaluate at the limits)} \\
&= \tau \cdot 2 \left(\lim_{z \to \infty} [-\phi(z)] + \phi(0) \right) && \text{(subtracting a negative is adding)} \\
&= \tau \cdot 2 (0 + \phi(0)) && (\phi(z) \to 0 \text{ as } z \to \infty) \\
&= \tau \cdot 2 \phi(0) && \text{(drop the zero term)} \\
&= \tau \cdot 2 (2\pi)^{-1/2} && (\phi(0) = (2\pi)^{-1/2}) \\
&= \tau \cdot 2 \cdot \frac{1}{\sqrt{2\pi}} && (\text{write } (2\pi)^{-1/2} \text{ as } 1/\sqrt{2\pi}) \\
&= \tau \cdot \frac{2}{\sqrt{2\pi}} && \text{(multiply)} \\
&= \tau \cdot \frac{\sqrt{4}}{\sqrt{2\pi}} && (\text{write } 2 \text{ as } \sqrt{4}) \\
&= \tau \sqrt{4/(2\pi)} && \text{(quotient of square roots)} \\
&= \tau \sqrt{2/\pi} && (\text{cancel the common factor 2 in } 4/(2\pi))
\end{aligned}
$$

So for an unbiased Gaussian estimator, the pMAE is about 0.80 times the standard deviation τ, while the pRMSE is τ itself (Theorem 0.6).

Standard error

Definition 0.8 (Standard error)**.** The **standard error** of an estimator $\hat{\theta}$ is the standard deviation of $\hat{\theta}$:

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

Source: Wikipedia contributors [17].

Example 0.8 (Standard error of the sample mean)**.** In Example 0.3, $\mathrm{Var}\big(\bar{X}\big) = \sigma^2/n$, so

$$\begin{aligned}\mathrm{SE}\big(\bar{X}\big) &= \sqrt{\sigma^2/n} \\ &= \sigma/\sqrt{n}.\end{aligned}$$

With $n = 50$ students, the standard error of the sample mean height is $\sigma/\sqrt{50} \approx 0.14\sigma$.

Exercise 0.7 (Spread of the estimation error)**.** Find the standard deviation of the estimation error $\varepsilon\big(\hat{\theta}\big)$ in terms of the standard error of $\hat{\theta}$.

Solution

Solution 0.7.

$$\begin{aligned}\mathrm{Var}\Big(\varepsilon\big(\hat{\theta}\big)\Big) &= \mathrm{Var}\big(\hat{\theta} - \theta\big) && \text{(definition of estimation error)} \\ &= \mathrm{Var}\big(\hat{\theta}\big) && \text{(subtracting a constant does not change a variance)}\end{aligned}$$

Taking square roots of both sides,

$$\begin{aligned}\mathrm{SD}\Big(\varepsilon\big(\hat{\theta}\big)\Big) &= \mathrm{SD}\big(\hat{\theta}\big) \\ &= \mathrm{SE}\big(\hat{\theta}\big).\end{aligned}$$

Theorem 0.7 (Standard error is the spread of the estimation error)**.**

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

Proof

Proof. This is the solution to Exercise 0.7. □

Remark (The name "standard error")

Remark 0.3 (The name "standard error"). "Standard error" is a confusing name in two ways. It is defined through the estimator's own spread, not through the estimation error (although Theorem 0.7 shows that the two spreads are equal). It is also a synonym for the standard deviation of an estimator, so it can look redundant. The name persists because standard errors are the building blocks of p-values and confidence intervals, so they come up often enough to deserve their own name.

Exercise 0.8 (Squared standard error from pMSE and bias)**.** Using Theorem 0.5, write the squared standard error $\left(\mathrm{SE}\left(\hat{\theta}\right)\right)^2$ in terms of the pMSE and the bias of $\hat{\theta}$.

Solution

Solution 0.8.

$$\begin{aligned}
\mathrm{pMSE}\left(\hat{\theta}\right) &= \left(\mathrm{Bias}\left(\hat{\theta}\right)\right)^2 + \mathrm{Var}\left(\hat{\theta}\right) && \text{(MSE equals bias squared plus variance)} \\
\mathrm{Var}\left(\hat{\theta}\right) &= \mathrm{pMSE}\left(\hat{\theta}\right) - \left(\mathrm{Bias}\left(\hat{\theta}\right)\right)^2 && \text{(subtract } \left(\mathrm{Bias}\left(\hat{\theta}\right)\right)^2 \text{ from both sides)} \\
\left(\mathrm{SD}\left(\hat{\theta}\right)\right)^2 &= \mathrm{pMSE}\left(\hat{\theta}\right) - \left(\mathrm{Bias}\left(\hat{\theta}\right)\right)^2 && \left(\mathrm{Var}\left(\hat{\theta}\right) = \left(\mathrm{SD}\left(\hat{\theta}\right)\right)^2\right) \\
\left(\mathrm{SE}\left(\hat{\theta}\right)\right)^2 &= \mathrm{pMSE}\left(\hat{\theta}\right) - \left(\mathrm{Bias}\left(\hat{\theta}\right)\right)^2 && \left(\text{definition of standard error: } \mathrm{SE}\left(\hat{\theta}\right) = \mathrm{SD}\left(\hat{\theta}\right)\right)
\end{aligned}$$

Corollary 0.1 (Standard error squared equals pMSE minus squared bias)**.** *The squared standard error is what remains of the pMSE after the squared bias is removed:*

$$\left(\mathrm{SE}\left(\hat{\theta}\right)\right)^2 = \mathrm{pMSE}\left(\hat{\theta}\right) - \left(\mathrm{Bias}\left(\hat{\theta}\right)\right)^2$$

Proof

Proof. This is the solution to Exercise 0.8. □

Exercise 0.9 (Standard error of an unbiased estimator)**.** Using Equation 6, write the standard error of an unbiased estimator $\hat{\theta}$ in terms of its pMSE.

Solution

Solution 0.9. By Equation 6, $\text{pMSE}\big(\hat{\theta}\big) = \text{Var}\big(\hat{\theta}\big)$. Taking square roots of both sides,

$$\begin{aligned}\sqrt{\text{pMSE}\big(\hat{\theta}\big)} &= \text{SD}\big(\hat{\theta}\big) \\ &= \text{SE}\big(\hat{\theta}\big).\end{aligned}$$

Corollary 0.2 (For unbiased estimators, SE equals pRMSE)**.** *If $\hat{\theta}$ is unbiased, then its standard error equals its pRMSE:*

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

Proof

Proof. This is the solution to Exercise 0.9. □

Further reading

Another explanation

Evans and Rosenthal's free textbook [18] covers statistical inference in Chapter 5 and likelihood inference in Chapter 6.

References

[1] G. E. P. Box and N. Richard. Draper, *Empirical model-building and response surfaces*. in Wiley series in probability and mathematical statistics. Applied probability and statistics. New York: Wiley, 1987.

[2] P. K. Dunn and G. K. Smyth, *Generalized linear models with examples in R*, vol. 53. Springer, 2018. doi: 10.1007/978-1-4419-0118-7.

[3] S. Van Buuren, *Flexible imputation of missing data*. CRC press, 2018. Available: https://stefvanbuuren.name/fimd/

[4] M. Pohl, L. Baumann, R. Behnisch, M. Kirchner, J. Krisam, and A. Sander, "Estimands—A Basic Element for Clinical Trials," *Deutsches Ärzteblatt International*, vol. 118, no. 51–52, pp. 883–888, 2021, doi: 10.3238/arztebl.m2021.0373.

[5] R. Lawrance *et al.*, "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*, vol. 4, no. 1, pp. 1–8, 2020, doi: 10.1186/s41687-020-00218-5.

[6] Wikipedia contributors, "Estimand — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Estimand

[7] Wikipedia contributors, "Estimator — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Estimator

[8] Wikipedia contributors, "Sampling distribution — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Sampling_distribution

[9] R. V. Hogg, E. A. Tanis, and D. L. Zimmerman, *Probability and statistical inference*, Tenth edition. Boston: Pearson, 2019.

[10] G. Casella and R. Berger, *Statistical inference*, 2nd ed. Cengage Learning, 2002. Available: https://www.cengage.com/c/statistical-inference-2e-casella-berger/9780534243128/

[11] J. A. Rice, *Mathematical statistics and data analysis*, 3rd ed. Belmont, CA: Cengage Learning, 2007.

[12] Wikipedia contributors, "Bias of an estimator — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Bias_of_an_estimator

[13] Wikipedia contributors, "Mean squared error — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Mean_squared_error

[14] Wikipedia contributors, "Root mean square deviation — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Root_mean_square_deviation

[15] 5MinuteAI, "Probability & statistics." Video playlist on YouTube. Accessed: Oct. 12, 2026. [Online]. Available: https://www.youtube.com/playlist?list=PLeB45KifGiuHesi4PALNZSYZFhViVGQJK

[16] Wikipedia contributors, "Mean absolute error — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Mean_absolute_error

[17] Wikipedia contributors, "Standard error — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Standard_error

[18] M. J. Evans and J. S. Rosenthal, *Probability and statistics: The science of uncertainty*, 2nd ed. Freeman, 2010. Available: https://utstat.utoronto.ca/mikevans/jeffrosenthal/