Estimation

Last modified: 2026-09-29 00:05:17 (PDT)

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 (1987), p. 424; emphasis added]

See also (Dunn and Smyth 2018, 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 (2018) for more details.

Estimands, estimates, and estimators

Estimands

Definition 0.2 (Estimand)**.** An **estimand** is an unknown quantity whose value we want to know (Pohl et al. 2021; Lawrance et al. 2020).

Example 0.2 (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 parameters and other estimands are often symbolized using lower-case Greek letters: $\alpha, \beta, \gamma, \delta$, etc.

Estimates

Definition 0.3 (Estimate/estimated value)**.** In statistics, an **estimate** or **estimated value** is an informed guess of an estimand's value, based on observed data.

Example 0.3 (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 0.4 (Estimator)**.** An **estimator** is a function $\hat{\theta}(x_1, \dots, x_n)$ that transforms data $x_1, \dots, x_n$ into an estimate.

i Estimators are random variables

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

i Notation for estimators

Estimators are often symbolized by placing a $\widehat{\ }$ ("hat") symbol on top of the corresponding estimand; for example, $\hat{\theta}$.
Usually, their dependence on the data is implicit:

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

Example 0.4 (Mean height of students)**.** Suppose we want to estimate the 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 $X_1, \dots, X_n$. Then we could use the function

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

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:

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

A third possible estimator would be the 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:

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

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

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.

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 0.5 (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 0.5 (Estimation error of a mean height)**.** Continuing Example 0.3, 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:

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

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, mean squared error, and mean absolute error defined in this section, are all functions of the distribution of the estimator's estimation error.

Residuals

See Linear-model residual definitions and terminology for residual definitions and for the relationship between residuals, model deviations, and estimation error.

Bias

Definition 0.6 (Bias)**.** The **bias** of an estimator $\hat{\theta}$ for an estimand θ is the expected value of the estimation error:

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

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

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

Proof

Proof.

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

□

Example 0.6 (Bias of two estimators of a mean)**.** Let $X_1, \ldots, X_n$ each have expectation $\mathrm{E}[X_i] = \mu$, as in Example 0.4. By Theorem 0.1, the bias of the sample mean $\bar{X}$ is:

$$\begin{aligned}
\operatorname{Bias}(\bar{X}) &= \mathrm{E}\left[\bar{X}\right] - \mu && \text{(bias equals expectation minus truth)} \\
&= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] - \mu && \text{(definition of } \bar{X}\text{)} \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[X_i] - \mu && \text{(linearity of expectation)} \\
&= \frac{1}{n} \cdot n\mu - \mu && (\mathrm{E}[X_i] = \mu \text{ for every } i) \\
&= 0
\end{aligned}$$

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

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

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

Mean squared error

Definition 0.7 (Mean squared error)**.** The **mean squared error** of an estimator $\hat{\theta}$, denoted $\text{MSE}\big(\hat{\theta}\big)$, is the expectation of the square of the estimation error:

$$\text{MSE}\big(\hat{\theta}\big) \stackrel{\text{def}}{=} \text{E}\Big[\big(\varepsilon\big(\hat{\theta}\big)\big)^2\Big]$$

Theorem 0.2 (Mean squared error equals bias squared plus variance)**.** *For any one-dimensional estimator $\hat{\theta}$:*

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

Proof

Proof. Let's start by expanding each term of the right-hand side. By Theorem 0.1, the squared bias is:

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

The variance is (simplified expression for variance):

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

Now, add them together and simplify:

$$\begin{aligned} \big(\text{Bias}\big(\hat{\theta}\big)\big)^2 + \text{Var}\big(\hat{\theta}\big) &= \big(\text{E}\big[\hat{\theta}\big]\big)^2 - 2\,\text{E}\big[\hat{\theta}\big]\theta + \theta^2 + \text{E}\big[\hat{\theta}^2\big] - \big(\text{E}\big[\hat{\theta}\big]\big)^2 && \text{(substitute both expansions)} \\ &= \text{E}\big[\hat{\theta}^2\big] - 2\,\text{E}\big[\hat{\theta}\big]\theta + \theta^2 && \text{(cancel } \big(\text{E}\big[\hat{\theta}\big]\big)^2\text{)} \end{aligned}$$

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

$$\begin{aligned}
\mathrm{MSE}\big(\hat{\theta}\big) &\stackrel{\text{def}}{=} \mathrm{E}\Big[\big(\varepsilon\big(\hat{\theta}\big)\big)^2\Big] && \text{(definition of MSE)}\\
&= \mathrm{E}\big[(\hat{\theta}-\theta)^2\big] && \text{(definition of estimation error)}\\
&= \mathrm{E}\big[\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2\big] && \text{(expand the binomial square)}\\
&= \mathrm{E}\big[\hat{\theta}^2\big] - \mathrm{E}\big[2\hat{\theta}\theta\big] + \mathrm{E}[\theta^2] && \text{(linearity of expectation)}\\
&= \mathrm{E}\big[\hat{\theta}^2\big] - 2\,\mathrm{E}\big[\hat{\theta}\big]\theta + \theta^2 && (\theta \text{ is a constant})
\end{aligned}$$

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

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

□

Example 0.7 (Mean squared error of two estimators of a mean)**.** Continuing Example 0.6, suppose also that $X_1, \ldots, X_n$ are mutually independent, each with variance $\mathrm{Var}(X_i) = \sigma^2$. Both $\bar{X}$ and X_1 have zero bias, so by Theorem 0.2 each estimator's mean squared error equals its variance.
For X_1, $\mathrm{MSE}(X_1) = 0^2 + \mathrm{Var}(X_1) = \sigma^2$.
For $\bar{X}$:

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

The third line uses the variance of a linear combination, whose covariance terms are all zero for independent variables.
So for any sample size $n > 1$, $\mathrm{MSE}\big(\bar{X}\big) = \sigma^2/n < \sigma^2 = \mathrm{MSE}(X_1)$: by mean squared error, the sample mean is the more accurate estimator. With $n = 50$ students, as in Example 0.4, the sample mean's mean squared error is 1/50 of X_1's.

Definition 0.8 (Root mean squared error)**.** The **root mean squared error** of an estimator $\hat{\theta}$ is the square root of its mean squared error:

$$\mathrm{RMSE}\big(\hat{\theta}\big) \stackrel{\text{def}}{=} \sqrt{\mathrm{MSE}\big(\hat{\theta}\big)}$$

It has the same units as the estimand. In Example 0.7, the root mean squared error of $\bar{X}$ is $\sigma/\sqrt{n}$.

Unbiased estimators

Definition 0.9 (Unbiased estimator)**.** An estimator $\hat{\theta}$ is **unbiased** if its bias is zero: $\mathrm{Bias}\big(\hat{\theta}\big) = 0$. An estimator whose bias is not zero is **biased**.

Example 0.8 (The sample mean is unbiased)**.** By Example 0.6, $\mathrm{Bias}\big(\bar{X}\big) = 0$ and $\mathrm{Bias}(X_1) = 0$, so both the sample mean $\bar{X}$ and the single-observation estimator X_1 are unbiased estimators of μ.

Example 0.9 (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 $S^2 \stackrel{\text{def}}{=} \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$;
- the divide-by-n estimator $\hat{\sigma}^2 \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

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 $\mathrm{E}\big[X_i^2\big] = \mathrm{Var}(X_i) + \big(\mathrm{E}[X_i]\big)^2 = \sigma^2 + \mu^2$ and $\mathrm{E}\big[\bar{X}^2\big] = \mathrm{Var}\big(\bar{X}\big) + \big(\mathrm{E}\big[\bar{X}\big]\big)^2 = \sigma^2/n + \mu^2$ (by Example 0.7 and Example 0.6):

$$
\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right] &= \mathrm{E}\left[\sum_{i=1}^{n}X_i^2 - 2\bar{X}\sum_{i=1}^{n}X_i + n\bar{X}^2\right] && \text{(expand each square and sum)}\\
&= \mathrm{E}\left[\sum_{i=1}^{n}X_i^2 - 2\bar{X}\cdot n\bar{X} + n\bar{X}^2\right] && \left(\textstyle\sum_{i=1}^{n}X_i = n\bar{X}\right)\\
&= \mathrm{E}\left[\sum_{i=1}^{n}X_i^2 - n\bar{X}^2\right] && \text{(collect the } \bar{X}^2 \text{ terms)}\\
&= \sum_{i=1}^{n}\mathrm{E}[X_i^2] - n\,\mathrm{E}[\bar{X}^2] && \text{(linearity of expectation)}\\
&= n(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right) && \text{(substitute both second moments)}\\
&= (n-1)\sigma^2 && \text{(cancel } n\mu^2 \text{ and simplify)}
\end{aligned}
$$

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

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

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

Theorem 0.3 (Properties of unbiased estimators)**.** *If $\hat{\theta}$ is an unbiased estimator of θ, then:*

$$\mathrm{E}\left[\hat{\theta}\right] = \theta \tag{3}$$

$$\mathrm{MSE}\left(\hat{\theta}\right) = \mathrm{Var}\left(\hat{\theta}\right) \tag{4}$$

Proof

Proof. For Equation 3, apply Theorem 0.1:

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

Adding θ to both sides gives $\operatorname{E}\left[\hat{\theta}\right] = \theta$.
For Equation 4, apply Theorem 0.2:

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

□

Mean absolute error

Definition 0.10 (Mean absolute error)**.** The **mean absolute error** of an estimator is the expectation of the absolute value of the estimation error:

$$
\operatorname{MAE}\left(\hat{\theta}\right) \stackrel{\text{def}}{=} \operatorname{E}\left[\left|\varepsilon\left(\hat{\theta}\right)\right|\right]
$$

Example 0.10 (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:

$$
\begin{aligned}
\operatorname{MAE}\left(\hat{\theta}\right) &= \operatorname{E}[|\tau Z|] && \text{(definition of MAE)} \\
&= \tau \operatorname{E}[|Z|] && (\tau > 0 \text{ and linearity of expectation}) \\
&= \tau \int_{-\infty}^{\infty} |z| \phi(z)\, dz && \text{(expectation of a function of } Z) \\
&= 2\tau \int_{0}^{\infty} z \phi(z)\, dz && (|z|\phi(z) \text{ is symmetric about 0}) \\
&= 2\tau [-\phi(z)]_0^{\infty} && (\phi'(z) = -z\phi(z)) \\
&= 2\tau \phi(0) && (\phi(z) \to 0 \text{ as } z \to \infty) \\
&= \tau \sqrt{2/\pi} && (\phi(0) = (2\pi)^{-1/2})
\end{aligned}
$$

So for an unbiased Gaussian estimator, the mean absolute error is about 0.80 times the standard deviation τ, while the root mean squared error is τ itself (Theorem 0.3).

Standard error

Definition 0.11 (Standard error)**.** The **standard error** of an estimator $\hat{\theta}$ is the standard deviation of $\hat{\theta}$:

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

Example 0.11 (Standard error of the sample mean)**.** In Example 0.7, $\mathrm{Var}\big(\bar{X}\big) = \sigma^2/n$, so $\mathrm{SE}\big(\bar{X}\big) = \sqrt{\sigma^2/n} = \sigma/\sqrt{n}$. With $n = 50$ students, the standard error of the sample mean height is $\sigma/\sqrt{50} \approx 0.14\sigma$.

Theorem 0.4 (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.

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

Taking square roots of both sides, $\mathrm{SD}\big(\varepsilon\big(\hat{\theta}\big)\big) = \mathrm{SD}\big(\hat{\theta}\big) = \mathrm{SE}\big(\hat{\theta}\big)$. □

"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 0.4 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.

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

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

Proof

Proof. Start from Theorem 0.2:

$$\begin{aligned}
\mathrm{MSE}\big(\hat{\theta}\big) &= \big(\mathrm{Bias}\big(\hat{\theta}\big)\big)^2 + \mathrm{Var}\big(\hat{\theta}\big) && \text{(MSE equals bias squared plus variance)} \\
\mathrm{Var}\big(\hat{\theta}\big) &= \mathrm{MSE}\big(\hat{\theta}\big) - \big(\mathrm{Bias}\big(\hat{\theta}\big)\big)^2 && \text{(subtract } \big(\mathrm{Bias}\big(\hat{\theta}\big)\big)^2 \text{ from both sides)} \\
\big(\mathrm{SE}\big(\hat{\theta}\big)\big)^2 &= \mathrm{MSE}\big(\hat{\theta}\big) - \big(\mathrm{Bias}\big(\hat{\theta}\big)\big)^2 && \big(\mathrm{Var}\big(\hat{\theta}\big) = \big(\mathrm{SD}\big(\hat{\theta}\big)\big)^2 = \big(\mathrm{SE}\big(\hat{\theta}\big)\big)^2\big)
\end{aligned}$$

□

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

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

Proof

Proof. By Equation 4, $\mathrm{MSE}\big(\hat{\theta}\big) = \mathrm{Var}\big(\hat{\theta}\big)$. Taking square roots of both sides, $\sqrt{\mathrm{MSE}\big(\hat{\theta}\big)} = \mathrm{SD}\big(\hat{\theta}\big) = \mathrm{SE}\big(\hat{\theta}\big)$. □

References

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. Wiley.

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

Lawrance, Rachael, Evgeny Degtyarev, Philip Griffiths, et al. 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://doi.org/10.1186/s41687-020-00218-5.

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/.