Definition 2.2. An *estimator* $\hat{\theta}$ is a statistic (that is, it is a random variable) which after the experiment has been conducted and the data has been collected will be used to estimate θ.

Example 2.3. Recall that if $Y \sim \text{Exp}(1/\theta)$, then Y has density function

$$f_Y(y) = \begin{cases} \frac{1}{\theta} e^{-y/\theta}, & y > 0, \\ 0, & y \le 0, \end{cases}$$

and so

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y)\,\mathrm{d}y = \int_0^{\infty} \frac{y}{\theta} e^{-y/\theta}\,\mathrm{d}y = \theta \int_0^{\infty} u e^{-u}\,\mathrm{d}u = \theta \Gamma(2) = \theta.$$

Suppose that Y_1, Y_2, Y_3 are a random sample from an $\text{Exp}(1/\theta)$ population and that estimation of θ is desired. In particular, θ is the population mean. The natural question, therefore, is what should be used as an estimator of θ? A similar question was asked in Example 2.1 and, as in that example, two possible estimators of θ are the sample median and the sample mean. Consider the sample mean

$$\hat{\theta} := \frac{Y_1 + Y_2 + Y_3}{3}.$$

Is this a good estimator of θ? Is there a better estimator of θ? Since the population distribution is skewed, the rule from Stat 160 suggests that the sample median is a better measure of centre than the sample mean. Is it? And how should we go about verifying this.

In fact, both the previous example and Example 2.1 suggest the following.

Question. What does it mean for an estimator to be good? What does it mean for one estimator to be better than another estimator? How should we chose estimators?

Classical Statistics Answer. We choose the "minimum variance unbiased estimator" (MVUE) as our preferred estimator. In general, this will be our criterion. We will search for unbiased estimators, and then we will select the unbiased estimator whose variance is the smallest.

Hence, we begin by defining what it means for an estimator to be unbiased. Recall that an estimator is a statistic (and hence a random variable) so that we can take its expectation, and that a parameter is not a random variable, but just a number.

Definition 2.4. If $\hat{\theta}$ is an estimator of θ, then the *bias* of $\hat{\theta}$ is

$$B(\hat{\theta}) := \mathbb{E}(\hat{\theta}) - \theta.$$

We say that $\hat{\theta}$ is an *unbiased estimator* of θ if $B(\hat{\theta}) = 0$.

Note that $\hat{\theta}$ is an unbiased estimator of θ if and only if $\mathbb{E}(\hat{\theta}) = \theta$.

Example 2.5. As a first example, we will show that if $Y_1, \ldots, Y_n$ are a random sample from a population whose population mean is μ, then the sample mean

$$\overline{Y} := \frac{1}{n} \sum_{i=1}^{n} Y_i$$

is *always* an unbiased estimator of μ. That is, since $\mathbb{E}(Y_1) = \cdots = \mathbb{E}(Y_n) = \mu$ by assumption, we deduce that

$$\mathbb{E}(\overline{Y}) = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} Y_i\right) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(Y_i) = \frac{1}{n}(\mu + \cdots + \mu) = \frac{n\mu}{n} = \mu.$$

That is, if $\theta := \mu$ and $\hat{\theta} := \overline{Y}$, then

$$B(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta = \mathbb{E}(\overline{Y}) - \mu = \mu - \mu = 0.$$

Actually, this example is so important that we record it as a theorem.

Theorem 2.6. *The sample mean is an unbiased estimator of the population mean.*

Example 2.7. We now revisit Example 2.3. If Y_1, Y_2, Y_3 are a random sample from an $\text{Exp}(1/\theta)$ population, then the sample mean

$$\hat{\theta} := \frac{Y_1 + Y_2 + Y_3}{3}$$

is an unbiased estimator of θ since

$$\mathbb{E}(\hat{\theta}) = \mathbb{E}\left(\frac{Y_1 + Y_2 + Y_3}{3}\right) = \frac{\mathbb{E}(Y_1) + \mathbb{E}(Y_2) + \mathbb{E}(Y_3)}{3} = \frac{\theta + \theta + \theta}{3} = \frac{3\theta}{3} = \theta.$$

As we know from Stat 251, we can also compute the variance of the sample mean.

Theorem 2.8. *Suppose that $Y_1, \ldots, Y_n$ are a random sample from a population having common mean μ and common variance σ^2. If*

$$\overline{Y} := \frac{1}{n} \sum_{i=1}^{n} Y_i$$

denotes the sample mean, then

$$\text{Var}(\overline{Y}) = \frac{\sigma^2}{n}.$$

Proof. Since $Y_1, \ldots, Y_n$ are independent, we deduce that

$$\text{Var}\left(\overline{Y}\right) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(Y_i),$$

and since $\mathrm{Var}(Y_1) = \cdots \mathrm{Var}(Y_n) = \sigma^2$, we conclude

$$\mathrm{Var}\left(\overline{Y}\right) = \frac{\sigma^2 + \cdots + \sigma^2}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

as required. □

Using this result, we can show that the sample variance is an unbiased estimator of the population variance.

Example 2.9. Show that if $Y_1, Y_2, \ldots, Y_n$ is a random sample from a population with mean μ and variance σ^2, then

$$S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

is an unbiased estimator of σ^2. This illustrates one reason for dividing by $n-1$ in the definition of the sample variance S^2, instead of dividing by (the seemingly more natural) n.

Solution. We begin by observing that $(Y_i - \overline{Y})^2 = Y_i^2 - 2\overline{Y}Y_i + \overline{Y}^2$ and so

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(Y_i^2 - 2\overline{Y}Y_i + \overline{Y}^2) &= \sum_{i=1}^{n} Y_i^2 - 2\sum_{i=1}^{n}\overline{Y}Y_i + \sum_{i=1}^{n}\overline{Y}^2 \\ &= \sum_{i=1}^{n} Y_i^2 - 2\overline{Y}\sum_{i=1}^{n} Y_i + n\overline{Y}^2 \\ &= \sum_{i=1}^{n} Y_i^2 - 2n\overline{Y}^2 + n\overline{Y}^2 \\ &= \sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2.\end{aligned}$$

We now observe

$$\mathbb{E}\left(\sum_{i=1}^{n}(Y_i - \overline{Y})^2\right) = \mathbb{E}\left(\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\right) = \sum_{i=1}^{n}\mathbb{E}(Y_i^2) - n\mathbb{E}(\overline{Y}^2). \qquad (2.1)$$

If X is any random variable, then $\mathbb{E}(X^2) = \mathrm{Var}(X) + [\mathbb{E}(X)]^2$. This implies that $\mathbb{E}(Y_i^2) = \mathrm{Var}(Y_1) + [\mathbb{E}(Y_i)]^2 = \sigma^2 + \mu^2$ and

$$\mathbb{E}(\overline{Y}^2) = \mathrm{Var}(\overline{Y}) + \mathbb{E}(\overline{Y})^2 = \frac{\sigma^2}{n} + \mu^2.$$

Therefore, substituting into (2.1), we find

$$\mathbb{E}\left(\sum_{i=1}^{n}(Y_i - \overline{Y})^2\right) = \sum_{i=1}^{n}(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) = n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2 = (n-1)\sigma^2$$

and so

$$\mathbb{E}(S^2) = \mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2\right) = \frac{1}{n-1}\cdot(n-1)\sigma^2 = \sigma^2.$$

Thus, we have shown that S^2 is an unbiased estimator of σ^2 as required.