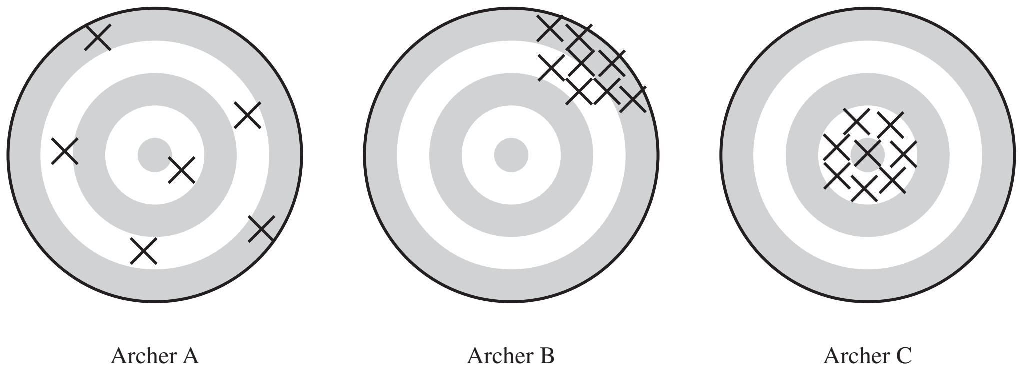
FIGURE 2.3
Unbiased, precise, and accurate archers. Archer A is unbiased—the average position of all arrows is at the bull's-eye. Archer B is precise but not unbiased—all arrows are close together but systematically away from the bull's-eye. Archer C is accurate—all arrows are close together and near the center of the target.



Thus, an estimator $\hat{t}$ of t is **unbiased** if $E(\hat{t}) = t$, **precise** if $V(\hat{t}) = E[(\hat{t} - E[\hat{t}])^2]$ is small, and **accurate** if $\text{MSE}[\hat{t}] = E[(\hat{t} - t)^2]$ is small. A badly biased estimator may be precise but it will not be accurate; accuracy (MSE) is how close the estimate is to the true value, while precision (variance) measures how close estimates from different samples are to each other. Figure 2.3 illustrates these concepts.

In summary, the finite population $\mathcal{U}$ consists of units $\{1, 2, \ldots, N\}$ whose measured values are $\{y_1, y_2, \ldots, y_N\}$. We select a sample $\mathcal{S}$ of n units from $\mathcal{U}$ using the probabilities of selection that define the sampling design. The y_i's are fixed but unknown quantities—unknown unless that unit happens to appear in our sample $\mathcal{S}$. Unless we make additional assumptions, the only information we have about the set of y_i's in the population is in the set $\{y_i : i \in \mathcal{S}\}$.

You may be interested in many different population quantities from your population. Historically, however, the main impetus for developing theory for sample surveys has been estimating population means and totals. Suppose we want to estimate the total number of persons in Canada who have diabetes, or the average number of oranges produced per orange tree. The population total is

$$t = \sum_{i=1}^{N} y_i,$$

and the mean of the population is

$$\bar{y}_U = \frac{1}{N} \sum_{i=1}^{N} y_i.$$

Almost all populations exhibit some variability; for example, households have different incomes and trees have different diameters. Define the **variance** of the population values about the mean as

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y}_U)^2. \tag{2.6}$$