7-3 Estimating a Population Mean

Key Concept This section presents methods for using a sample mean $\bar{x}$ to make an inference about the value of the corresponding population mean μ. Here are the three main concepts included in this section:

- **Point Estimate:** The sample mean $\bar{x}$ is the best *point estimate* (or single value estimate) of the population mean μ.
- **Confidence Interval:** We can use a sample mean to construct a *confidence interval* estimate of the true value of a population mean, and we should know how to construct and interpret such confidence intervals.
- **Sample Size:** We should know how to find the sample size necessary to estimate a population mean.

Part 1 of this section deals with the very realistic and commonly used case in which the population standard deviation σ is not known. Part 2 includes a brief discussion of the procedure used when σ is known, which is very rare.

Part 1: Estimating a Population Mean When σ Is Not Known

It's rare that we want to estimate the unknown value of a population mean but we somehow know the value of the population standard deviation σ. The realistic situation is that σ is not known. When σ is not known, we construct the confidence interval by using the Student t distribution instead of the standard normal distribution.

Point Estimate The sample mean $\bar{x}$ is an *unbiased estimator* of the population mean μ , and for many populations, sample means tend to vary less than other measures of center, so the sample mean $\bar{x}$ is usually the best point estimate of the population mean μ.

The sample mean $\bar{x}$ is the best *point estimate* of the population mean μ.

Although the sample mean $\bar{x}$ is usually the *best* point estimate of the population mean μ, it does not give us any indication of just how *good* our best estimate is, so we construct a *confidence interval* (or *interval estimate*), which consists of a range (or an interval) of values instead of just a single value.

Requirement of Normality or $n > 30$ The procedure we use has a requirement that the population is normally distributed or the sample size is greater than 30. This procedure is *robust* against a departure from normality, meaning that it works reasonably well if the departure from normality is not too extreme. Verify that there are no outliers and that the histogram or dotplot has a shape that is not very far from a normal distribution.

If the original population is not itself normally distributed, we use the condition $n > 30$ for justifying use of the normal distribution, but there is no exact specific minimum sample size that works for all cases. Sample sizes of 15 to 30 are sufficient if the population has a distribution that is not far from normal, but some other populations have distributions that are extremely far from normal and sample sizes greater than 30 might be necessary. We use the simplified criterion of $n > 30$ as justification for treating the distribution of sample means as a normal distribution, regardless of how far the distribution departs from a normal distribution.

Confidence Level The confidence interval is associated with a confidence level, such as 0.95 (or 95%). The confidence level gives us the *success rate of the procedure*