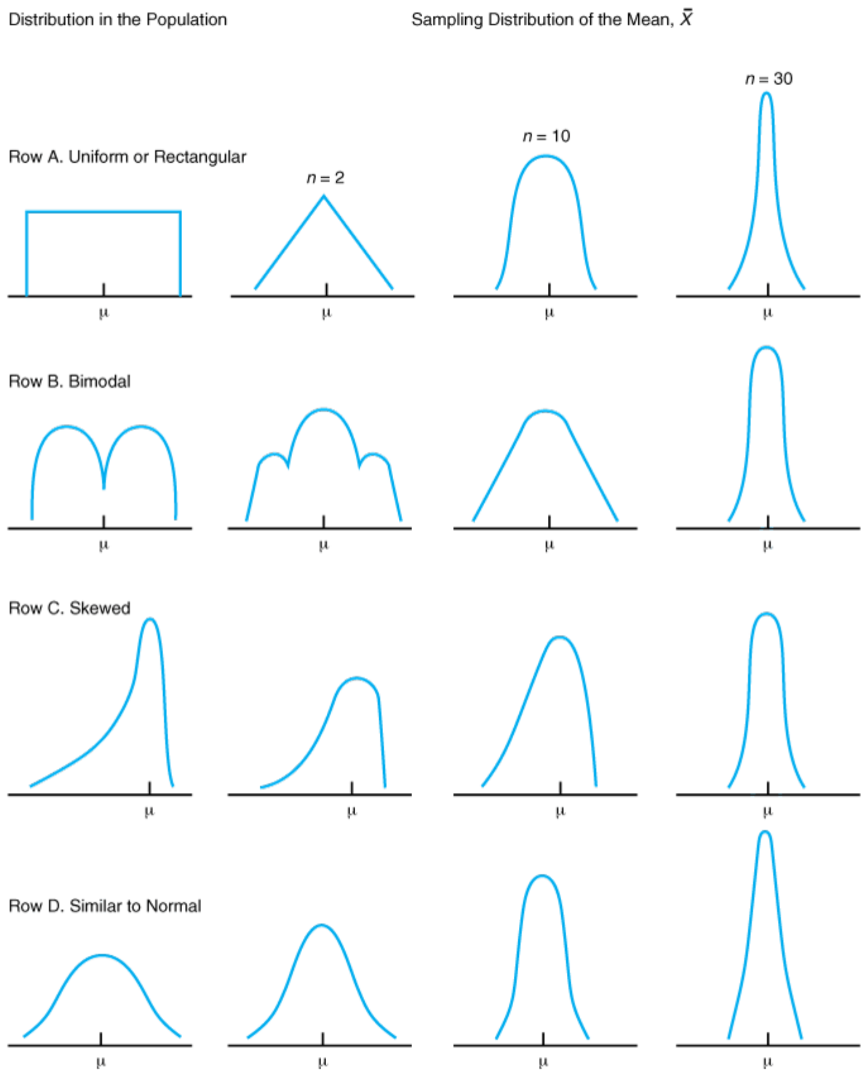
Form of Sampling Distribution of $\bar{x}$

1. **Population has a normal distribution**
   In many situations it is reasonable to assume that the population from which we are selecting a sample has a normal or nearly normal distribution.
2. **Population does not have a normal distribution**
   When the population does not have a normal distribution, the **central limit theorem** is helpful in identifying the shape the sample distribution of $\bar{x}$.

Central Limit Theorem

In selecting simple random samples of size n from a population, the sampling distribution of the sample mean $\bar{x}$ can be approximated by a *normal distribution* as the sample size becomes large.

General statistical practice is to assume that, for most applications, the sampling distribution of $\bar{x}$ can be approximated by a normal distribution whenever the sample is size 30 or more. In case where the population is highly skewed or outliers are present, samples of size 50 may be needed.



Other Sampling Methods

Stratified Random Sampling

In stratified random sampling, the elements in the population are first divided into groups called *strata*, such that each element in the population belongs to one and only one stratum. The basis for forming the strata, such as department, location, age, industry type, and so on, is at the discretion of the designer of the sample. However, the best results are obtained when the elements within each stratum are as much alike as possible. Stratified random sampling works best when the variance among elements in each stratum is relatively small.
(Notes)

Cluster Sampling

In cluster sampling, the elements in the population are divided into separate groups called *clusters*. Each element of the population belongs to one and only one cluster. All elements within each sampled cluster form the sample. Cluster sampling tends to provide the best results when the elements within the clusters are not alike. It works best when each cluster provides a small-scale representation of the population. A good example of cluster sampling is area sampling: clusters can be city blocks or other well-defined areas.
(Notes)

Systematic Sampling

In some sampling situations, especially those with large populations, it is time-consuming to select a simple random sample by first finding a random number and then counting or searching through the list of population until the corresponding element is found. An alternative to simple random sampling is *systematic sampling*.
(Notes)

Convenience Sampling

Convenience sampling is a *nonprobability sampling* [1] technique. The sample is identified by convenience. Elements are included in the sample without pre-specified or known probabilities of being selected. Convenience samples have the advantage of relatively easy sample selection and data collection. However, it is impossible to evaluate the "goodness" of the sample in terms of its representativeness of the population. No statistically justified procedure allows a probability analysis and inference about the quality of the sample results. Therefore, we should be cautious in interpreting the results of convenience samples that are used to make inferences about populations.
(Notes)

Judgment Sampling

In this approach, the person most knowledgeable on the subject of the study selects elements of the population that he or she feels are most representative of the population. However, the quality of the sample results depends on the judgment of the person selecting the sample. We should also be cautious in interpreting the results based on judgment sampling.
(Notes)

[1] On the flipside, for probability sampling, elements selected from the population have a known probability of being included in the sample.

Chapter 4: Interval Estimation

Because a point estimator cannot be expected to provide the exact value of the population parameter, an *interval estimate* is often computed by adding and subtracting a value, called the *margin error*, to the point estimate. The general form of an interval estimate is as follows:

$$\text{Point estimate} \pm \text{Margin of error}$$

The general form of an interval estimate of a population mean is:

$$\bar{x} \pm \text{Margin of error}$$

The general form of an interval estimate of a population proportion is:

$$\bar{p} \pm \text{Margin of error}$$

Population Mean: σ Known

Interval estimate of a population mean: σ Known

$$\textbf{Confidence Interval:} \quad \bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Table 1: Values of $z_{\alpha/2}$ for the most commonly used confidence levels

Confidence Level	**α**	**α/2**	$z_{\alpha/2}$
90%	0.10	0.05	1.645
95%	0.05	0.025	1.960
99%	0.01	0.005	2.576

(Notes)

Exercises

1. Given $\bar{x}$ = 82, n = 100, and σ = 20, find the interval estimate using 90%, 95%, and 99% confidence intervals.

2. A simple random sample of 50 items from a population with σ = 6 resulted in a sample mean of 32.
    a. Provide a 90% confidence interval for the population mean.
    b. Provide a 95% confidence interval for the population mean.
    c. Provide a 99% confidence interval for the population mean.

3. A 95% confidence interval for a population mean was reported to be 152 to 160. If σ = 15, what sample size was used in this study?

4. In an effort to estimate the mean amount spent per customer for dinner at a major Atlanta restaurant, data were collected for a sample of 49 customers. Assume a population standard deviation of $5.
    a. At 95% confidence, what is the margin of error?
    b. If the sample mean is $24.80, what is the 95% confidence interval for the population mean?

5. Nielsen Media Research reported that the house hold mean television viewing time during the 8 pm to 11 pm time period is 8.5 hours per week (*The World Almanac 2003*). Given a sample size of 300 households and a population standard deviation of σ = 3.5 hours, what is the 95% confidence interval estimate of the mean television viewing time per week during the 8 pm to 11 pm time period?

6. Playbill magazine reported that the mean annual household income of its readers is $119,155 (Playbill, December 2003). Assume this estimate of the mean annual household income is based on a sample of 80 households, and based on past studies, the population standard deviation is known to be σ = $30,000.
    a. Develop a 90% confidence interval estimate of the population mean.
    b. Develop a 95% confidence interval estimate of the population mean.
    c. Develop a 99% confidence interval estimate of the population mean.
    d. Discuss what happens to the width of the confidence interval as the confidence level is increased. Does this result seem reasonable? Explain.