Previously Defined Terms

Parameter—a numerical descriptive measure of a population

Sample statistic—a numerical descriptive measure of a sample; calculated from the observations in a sample

Sample values are measurements or observations of RVs; the value for a sample statistic will vary in a random manner from sample to sample

Sample statistics are RVs because different samples can lead to different values for the sample statistics

Sampling Distributions—Statistics as RVs

New Definition

Sampling distribution (of a sample statistic)—the distribution of possible values of a statistic for repeated samples of the same size from a population

Statistical Inference

Want to make conclusions about population parameters on the basis of sample statistics.

Confidence Intervals—interval of values that the researcher is fairly certain will cover the true, unknown value of the population parameter

Hypothesis Tests (significance testing)—uses sample data to attempt to reject a hypothesis about the population

Mean midterm score = 81.01; SD midterm score = 18.34

<table>
<thead>
<tr>
<th>Sample</th>
<th>mean10</th>
<th>mean20</th>
<th>mean30</th>
<th>mean50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>86.89</td>
<td>83.05</td>
<td>84.45</td>
<td>82.22</td>
</tr>
<tr>
<td>Sample 2</td>
<td>77.40</td>
<td>73.16</td>
<td>82.76</td>
<td>82.90</td>
</tr>
<tr>
<td>Sample 3</td>
<td>85.90</td>
<td>78.68</td>
<td>84.30</td>
<td>82.06</td>
</tr>
<tr>
<td>Sample 4</td>
<td>79.90</td>
<td>73.15</td>
<td>82.14</td>
<td>81.06</td>
</tr>
<tr>
<td>Sample 5</td>
<td>86.40</td>
<td>81.10</td>
<td>80.31</td>
<td>78.51</td>
</tr>
<tr>
<td>Sample 6</td>
<td>80.40</td>
<td>82.61</td>
<td>81.47</td>
<td>82.52</td>
</tr>
</tbody>
</table>
Many common, practical problems involve estimating mean values from samples; interest is really in making an inference about the mean, $\mu$, of some population.

**What do we know:** the sample mean, $\bar{X}$, is often a good estimator of $\mu$.

**Example:** Midterm exam scores—we considered the class data to represent the population.

Look again at histograms for $Y$ based on $n = 10, 20, 30$ and $50$ samples.
The normal probability distribution approximates the computer-generated sampling distributions very well.

### Sampling Distributions

**Properties of the sampling distribution of \( \bar{x} \)**

1. **Mean of sampling distribution equals mean of sampled population:**
   \[ \mu_{\bar{x}} = E(\bar{x}) = \mu \]
2. **Standard deviation of sampling distribution equals standard deviation of sampled population:**
   \[ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \]

### Central Limit Theorem (CLT)

**Consider a random sample of \( n \) observations selected from a population (any population) with mean \( \mu \) and standard deviation \( \sigma \).**

Then, when \( n \) is sufficiently large, the sampling distribution of \( \bar{x} \) will be **approximately a normal distribution** with mean \( \mu_{\bar{x}} = \mu \) and standard deviation \( \sigma_{\bar{x}} = \sigma / \sqrt{n} \).

The larger the sample size, the better will be the normal approximation to the sampling distribution of \( \bar{x} \).
Accessing Normality

• Look at a histogram—look for unimodal, approximately symmetric, no profound outliers
• Identify potential outliers; reject assumption of normality if there is more than one outlier present
• Normal quantile (probability) plot—use software
• Statistical tests for normality
  • Chi-square goodness-of-fit
  • Shapiro Wilks
  • Others

If data deemed not normal, what can you do?
• Consider data transformations to achieve approximate normality
• Use Nonparametric methods