Chapter 8 Sampling Distributions: Measuring the accuracy of sample results

8.1 Sampling distribution of a statistic defined

Definition: The sampling distribution of a statistic is the distribution (probability) of the values of the statistic in all possible samples of the same size n taken from the same population.

Empirical sampling distribution:
- the sample proportion
- the sample mean
8.2 Sampling distribution of a Sample Proportion

Notation:

1. \( p \) denotes the population proportion of some event \( A \).
   \[ p = P(A) \] — parameter

2. \( \hat{p} \) denotes the sample proportion of times \( A \) is observed.
   \( \hat{p} \) relative frequency of \( A \) in a sample.
   Let \( X \) denotes the number of times \( A \) occurs in a sample of size \( n \).
   \[ \hat{p} = \frac{X}{n} \] — statistic

Sample 1: 9 women, 11 men.
\[ \hat{p} = \frac{9}{20} \]
Sample 2: 10 women, 10 men
\[ \hat{p} = \frac{10}{20}. \]

Example: Suppose a student guesses on a multiple choice test \( n = 2 \)
questions each with 4 choices.
What is the probability distribution of \( \hat{p} \), the observed proportion of questions correct?

- \( X \), number of questions correct

\( X \sim \text{Bin}(2, 0.25) \)

\[ P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2 \]

<table>
<thead>
<tr>
<th>( x )</th>
<th>( P(X = x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.56</td>
</tr>
<tr>
<td>1</td>
<td>0.38</td>
</tr>
<tr>
<td>2</td>
<td>0.06</td>
</tr>
</tbody>
</table>

\[ \hat{p} = \frac{X}{n} \]

<table>
<thead>
<tr>
<th>( \hat{p} )</th>
<th>0</th>
<th>0.5</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What do we expect of sample proportions?

1. The values of \( \hat{p} \) vary from one random sample to the next in a predictable way.

2. The shape of the distribution of \( \hat{p} \) is approximately symmetric and bell-shaped (large sample size).

3. The center of the distribution of \( \hat{p} \) is at the true proportion \( p \) (large sample).

4. With larger sample sizes, the values of \( \hat{p} \) vary less around the true proportion.
Example: $n = 10$ question.

$X \sim \text{Bin}(10, 0.25)$

<table>
<thead>
<tr>
<th>$X$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = x)$</td>
<td>0.056</td>
<td>0.19</td>
<td>0.28</td>
<td>0.25</td>
<td>0.15</td>
</tr>
<tr>
<td>$P(\geq)$</td>
<td>0.1</td>
<td>0.3</td>
<td>0.5</td>
<td>0.7</td>
<td>1.0</td>
</tr>
</tbody>
</table>

\[
\binom{10}{x} (0.25)^x (0.75)^{10-x}
\]
83. Bias and Variability

- Center
- How the values varied
- Overall shape

Definition:
A statistic is **unbiased** if the center of its sampling distribution is equal to the corresponding population parameter value.

The **variability** of a statistic corresponds to the spread of its sampling distribution. A statistic whose distribution shows values that are very spread out and dispersed is said to lack precision.
Goal: low bias, low variability. For reducing variability, we can have larger sample size.