Estimated Margin of Error for \( \hat{p}_1 - \hat{p}_2 \). Consult page 567 and the supporting readings pp. 557-563.

**Random page sampling.** In the posted outline for lecture 1-12-09 we randomly sampled 36 page numbers from 001 to 767.

<table>
<thead>
<tr>
<th>Sampled Page</th>
<th>Pages Skipped</th>
<th>Total Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>716</td>
<td>32</td>
<td>463</td>
</tr>
<tr>
<td>473</td>
<td>200</td>
<td>731</td>
</tr>
<tr>
<td>727</td>
<td>39</td>
<td>759</td>
</tr>
<tr>
<td>764</td>
<td>(skip 944)</td>
<td></td>
</tr>
<tr>
<td>512</td>
<td>364</td>
<td>132</td>
</tr>
<tr>
<td>181</td>
<td>27</td>
<td>133</td>
</tr>
<tr>
<td>622</td>
<td>(skip 922)</td>
<td>666</td>
</tr>
<tr>
<td>310</td>
<td>(skip 844)</td>
<td>720</td>
</tr>
<tr>
<td>(skip 945)</td>
<td>639</td>
<td>112</td>
</tr>
<tr>
<td>285</td>
<td>(skip duplicate 112)</td>
<td>429</td>
</tr>
<tr>
<td>471</td>
<td>(skip duplicate 112)</td>
<td>647</td>
</tr>
<tr>
<td>(skip 770)</td>
<td>183</td>
<td>71</td>
</tr>
<tr>
<td>359</td>
<td>412</td>
<td>585</td>
</tr>
<tr>
<td>428</td>
<td>42</td>
<td></td>
</tr>
</tbody>
</table>

Every page of the book had the same chance of making it into this sample of \( n = 36 \) pages. I've highlighted in **bold** those sample pages having a picture or graphic.
Here are the 36 sample pages sorted:

**A Question.** Some pages of our book have a picture or graphic. How do the frequencies of such pages compare in the first half of the book versus the last half?

**Our solution.** We'll estimate the difference \( p_1 - p_2 \) between the probabilities:
\[ p_1 = \text{probability that a page} < 384 \text{ has a picture or graphic} \]
\[ p_2 = \text{probability that a page} > 383 \text{ has a picture or graphic} \]
using the data above.

Examining the data we find
a. Sample pages \(< 384\) vs \(> 383\) total 17 early, 19 late.
   b. Sample pages total 20 with picture/graphic, 16 without.

Try to be careful in the presence of such a *coincidence*. It will happen from time to time and can be a source of major error if we're not observant.

It helps to organize these counts in a **contingency table**:

<table>
<thead>
<tr>
<th></th>
<th>(&lt; 384)</th>
<th>(&gt; 383)</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contingency counts for 36 random pages:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with p/g</td>
<td>12</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>without</td>
<td>5</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>total</td>
<td>17</td>
<td>19</td>
<td>36</td>
</tr>
</tbody>
</table>
* We may regard the 17 sample pages \(< 384\) as a random sample of 17 pages from the first half of the book (383 pages).
* Likewise, we may regard the 19 sample pages \(> 383\) as a random sample of 19 pages from the second half of the book (384 pages).
* The sample of 17 pages from the first half is \textit{statistically independent} of the sample of 19 pages from the second half in the sense that \textit{neither can tell us anything about the other that we would not know anyway}.

\textit{Statistical independence is a backbone concept in statistics, and is largely responsible for the fabled "Law of Averages" which we study later.}

\textbf{Recall the Question:} Some pages of our book have a picture or graphic. How do the frequencies of such pages compare between the first half of the book and the last half?

\textbf{Recall our solution.} We'll estimate the difference \(p_1 - p_2\) between the probabilities:

\begin{align*}
p_1 &= \text{probability that a page } < 384 \text{ has a picture or graphic (of 383 pages)} \\
p_2 &= \text{probability that a page } > 383 \text{ has a picture or graphic (of 384 pages)}
\end{align*}

using the data from the contingency table above.

\textbf{1a. Needed estimates.}

\textbf{Point estimate of } \(p_1\) \text{ is } \frac{12}{17} \sim 0.70588

(this is \(\hat{p}_1\), we selected 17 pages \(< 384\) of which 12 had a picture or graphic).

\textbf{Point estimate of } \(p_2\) \text{ is } \frac{8}{19} \sim 0.42105

(this is \(\hat{p}_2\), we selected 19 pages \(> 383\) of which 8 had a picture or graphic).

\textbf{Our point estimate of } \(p_1 - p_2\) \text{ is } \hat{p}_1 - \hat{p}_2 \sim 0.70588 - 0.42105 \sim 0.28483

(this seems to be a fairly large difference, but what about the estimated margin of error of \(\hat{p}_1 - \hat{p}_2\)?)
1b. Estimated margin of error of the estimator \( \hat{p}_1 - \hat{p}_2 = \)

\[
1.96 \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} \frac{N_1-n_1}{N_1-1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} \frac{N_2-n_2}{N_2-1}}
\]

\[
= 1.96 \sqrt{\frac{.70588 (1-.70588)}{17} \frac{383-17}{383-1} + \frac{.42105 (1-.42105)}{19} \frac{384-19}{384-1}}
\]

\[
\approx 0.303185
\]

1c. Claim made for estimated margin of error: Provided \( n_1, N_1-n_1, n_2, N_2-n_2 \) are all "large enough"

Around 95% of random samples of 17 from 383 and 19 from 384 have

\( \hat{p}_1 - \hat{p}_2 \pm \) estimated margin of error

that will cover the actual value of \( p_1 - p_2 \).

Our \( n_1 = 17 \) and \( n_2 = 19 \) are on the small side so we'll have to say the confidence level will not so closely approximate 95%. But let's press on since the data are still informative.

Our random sample of 17 from 383 and 19 from 384 has produced the interval

\( \hat{p} \pm \) estimated margin of error

\[
= 0.28483 \pm 0.303185
\]

\[
= [-0.01835, 0.58801]
\]

This 95% confidence interval for \( p_1 - p_2 \) just contains 0, suggesting that we cannot confidently exclude the possibility that \( p_1 = p_2 \). It is within the estimated margin of error of our point estimate \( \hat{p}_1 - \hat{p}_2 = 0.28483 \) of \( p_1 - p_2 \).