SAMPLE SIZE
CONSIDERATIONS
Learning Objectives

• Understand the critical role having the right sample size has on an analysis or study.

• Know how to determine the correct sample size for a specific study.

• Understand the limitations of different data types on sample size.
How does it help?

**The correct sample size is necessary to:**

• ensure any tests you design have a high probability of success.
• properly utilize the type of data you have chosen or are limited to working with.
Common Uses

- Sample Size considerations are used in any situation where a sample is being used to infer a population characteristic.
KEYS TO SUCCESS

Use variable data wherever possible

Generally, more samples are better in any study

When there is any doubt, calculate the needed sample size

Use the provided excel spreadsheet to ease sample size calculations
CONFIDENCE INTERVALS

The possibility of error exists in almost every system. This goes for point values as well. While we report a specific value, that value only represents our best estimate from the data at hand. The best way to think about this is to use the form:

$$\text{true value} = \text{point estimate} \pm \text{error}$$

The error around the point value follows one of several common probability distributions. As you have seen so far, we can increase our confidence is to go further and further out on the tails of this distribution.

This “error band” which exists around the point estimate is called the **confidence interval**.
BUT WHAT IF I MAKE THE WRONG DECISION?

<table>
<thead>
<tr>
<th>Test Decision</th>
<th>Reality</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Not different (H₀)</td>
<td>Not different (H₀)</td>
<td>Correct Conclusion</td>
</tr>
<tr>
<td>Different (H₁)</td>
<td>Different (H₁)</td>
<td><strong>α</strong> Risk</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Type I Error</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Producer Risk</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>β</strong> Risk</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Type II Error</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Consumer Risk</strong></td>
</tr>
</tbody>
</table>

Test Reality = Different

Reality = Different

Decision Point

Test

β Risk

α Risk
WHY DO WE CARE IF WE HAVE THE TRUE VALUE?
How confident do you want to be that you have made the right decision?

A person does not feel well and checks into a hospital for tests.

<table>
<thead>
<tr>
<th>Test Decision</th>
<th>Reality</th>
<th>Error Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not different (H)&lt;sub&gt;0&lt;/sub&gt;</td>
<td>Not different (H)&lt;sub&gt;0&lt;/sub&gt;</td>
<td>Correct Conclusion</td>
</tr>
<tr>
<td>Different (H)&lt;sub&gt;1&lt;/sub&gt;</td>
<td>Different (H)&lt;sub&gt;1&lt;/sub&gt;</td>
<td>Correct Conclusion</td>
</tr>
</tbody>
</table>

- Type I Error = Treating a patient who is not sick
- Type II Error = Not treating a sick patient

Ho: Patient is not sick
H1: Patient is sick
HOW ABOUT ANOTHER EXAMPLE?

A change is made to the sales force to save costs. Did it adversely impact the order receipt rate?

<table>
<thead>
<tr>
<th>Test Decision</th>
<th>Reality</th>
<th>Error Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not different (H₀)</td>
<td>Not different (H₀)</td>
<td><strong>Correct Conclusion</strong></td>
</tr>
</tbody>
</table>
|               | Different (H₁) | • Type II Error  
|               |               | • β risk  
|               |               | • Consumer Risk  
| Different (H₁) |               | **Correct Conclusion** |
|               |               | • Type I Error  
|               |               | • α Risk  
|               |               | • Producer Risk  

Ho: Order rate unchanged
H₁: Order rate is different

Error Impact
Type I Error = Unnecessary costs
Type II Error = Long term loss of sales
These individual formulas are not critical at this point, but notice that the only opportunity for decreasing the error band (confidence interval) without decreasing the confidence factor, is to increase the sample size.
SAMPLE SIZE EQUATIONS

Mean

\[ n = \left( \frac{Z_{a/2} \sigma}{\mu - \overline{X}} \right)^2 \]

Allowable error = \( \mu - \overline{X} \)
(Also known as \( \delta \))

Standard Deviation

\[ n = \left( \frac{\sigma}{s} \right)^2 \chi^2_{a/2} + 1 \]

Allowable error = \( \sigma/s \)

Percent Defective

\[ n = \hat{p}(1 - \hat{p}) \left( \frac{Z_{a/2}}{E} \right)^2 \]

Allowable error = \( E \)
We want to estimate the true average weight for a part within 2 pounds. Historically, the part weight has had a standard deviation of 10 pounds. We would like to be 95% confident in the results.

**Calculation Values:**
- Average tells you to use the mean formula
- Significance: $a = 5\%$ (95\% confident)
- $Z_{a/2} = Z_{0.025} = 1.96$
- $s = 10$ pounds
- $m - x = \text{error allowed} = 2$ pounds

**Calculation:**

$$n = \left( \frac{Z_{a/2} \sigma}{\mu - X} \right)^2 = \left( \frac{1.96 \times 10}{2} \right)^2 = 97$$

**Answer:** $n=97$ Samples
SAMPLE SIZE EXAMPLE

We want to estimate the true percent defective for a part within 1%. Historically, the part percent defective has been 10%. We would like to be 95% confident in the results.

- Calculation Values:
  - Percent defective tells you to use the percent defect formula
  - Significance: $\alpha = 5\%$ (95% confident)
  - $Z_{\alpha/2} = Z_{.025} = 1.96$
  - $p = 10\% = .1$
  - $E = 1\% = .01$

- Calculation:

$$ n = \hat{p}(1 - \hat{p})\left(\frac{Z_{\alpha/2}}{E}\right)^2 = .1(1-.1)\left(\frac{1.96}{.01}\right)^2 = 3458 $$

Answer: $n=3458$ Samples
Learning Objectives

• Understand the critical role having the right sample size has on an analysis or study.

• Know how to determine the correct sample size for a specific study.

• Understand the limitations of different data types on sample size.