# Calculating Sample Size

**Data Science for Quality Management: Two Sample Hypothesis Testing**

with **Wendy Martin**

**Learning objective:**

Calculate sample size for means and variance

# Sample Size Calculations

- For the industrial researcher, proper sample size is not an opinion

# Factors to Considered

- The minimum effect size ($\Delta$) to be detected
- That is, the smallest degree of shift in the parameter that the researcher wishes to identify

# Factors to Considered

- The number of treatment levels (or groups)
- The population variance ($\sigma^2$)

# Factors to Considered

- The probability of committing a Type I error (α)
- The probability of committing a Type II error (β)

# Calculating Sample Size for Two Sample Tests of Means

- Assumptions:
  - σ unknown
  - Continuous Data, Independent Samples
  - Two Normal distributions
  - Non-directional test

# Calculating Sample Size for Two Sample Tests of Means

- When σ is unknown, hypothesis tests for means will use the t distribution
- Unfortunately, the t distribution is based upon degrees of freedom, which is determined by sample size

# Calculating Sample Size for Two Sample Tests of Means

- As such, sample size must be solved iteratively, where the sample size is determined to be the smallest n that satisfies the following formula

# Calculating Sample Size for One Sample Tests of Means

- Formula for non-directional hypotheses

$$n \geq \left(t_{\alpha/2,(n-1)df} + t_{\beta,(n-1)df}\right)^2 \frac{\sigma^2}{\Delta^2}$$

# Example

- If the requirements of a pull test are to be $\alpha = 0.05$, $\beta = 0.02$, $\Delta = 1$ lbs, and $\sigma = 2$, what would the appropriate minimum sample size be for a non-directional test for means?

# Example

- In Rstudio

```
> sample.size.mean.t.onesample
```

# Calculating Sample Size for One Sample Tests of Variance

- Formula for non-directional hypotheses

$$\chi^2 = \frac{s^2(n-1)}{\sigma_0^2}$$

# Calculating Sample Size for One Sample Tests of Variance

- For a non-directional test, we must consider two cases
  - One in which the variance increases
  - One in which the variance decreases

# Example

- If the requirements of a pull test are to be $\alpha = 0.05$, $\beta = 0.02$, $\Delta\sigma = 1$ lbs, and $\sigma = 2$, what would the appropriate minimum sample size be for a non-directional test for variances?

# Example

- If the variance increases, $\sigma = 3$ (2 + 1)
- If the variance decreases, $\sigma = 1$ (2 - 1)

# Example

- In Rstudio

```
> sample.size.variance.onesample
```

# Sources


- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982