Hypothesis Testing

Dr. Bob Gee
Dean Scott Bonney
Professor William G. Journigan
American Meridian University

Learning Objectives

Upon successful completion of this module, the student should be able to:

- Understand statistical and practical significance
- Understand Hypothesis Tests
- Demonstrate the ability to conduct Hypothesis Testing

Hypothesis Testing

- Definition
  - A Hypothesis Test is a method of using sample data to making decisions about a population, whether from a controlled experiment or an observational study
- Purpose
  - Determine whether a change in a process input (X) significantly changes the Output (Y) of the process
  - Statistically determine if there are differences between two or mores process outputs (Y)
- Application to the Lean Six Sigma process
  - Test the viability of the project team's process changes based on the experimental or observed data gathered during pilots

Hypothesis Testing

- Null Hypothesis (H_0): A statement about the value of a population (sample) parameter, that we hope to prove or disprove
- Alternative Hypothesis(H_1 or H_a): The statement that is accepted to be true Null Hypothesis is rejected

Disproving the null hypothesis may NOT allow us to say much about truth

e.g., H_0: The cup is full; if disproven, one can only state that the cup is not full - you cannot say the cup is empty

Non-statistical Hypothesis Testing

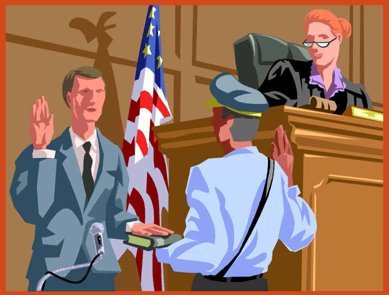
- A criminal trial is an example of hypothesis testing without the statistics.
- In a trial a jury must decide between two hypotheses. The null hypothesis is

  H_0: The defendant is innocent
- The alternative hypothesis or research hypothesis is

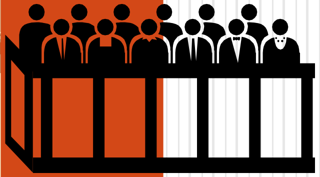
  H_1: The defendant is guilty
- The jury does not know which hypothesis is true. They must make a decision on the basis of evidence presented.

Hypothesis Testing

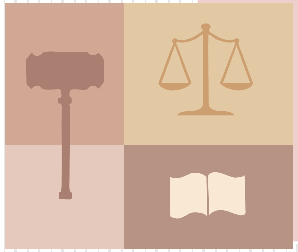
Hypothesis



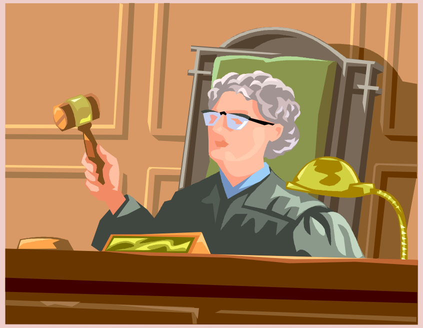
Significance



Collect Data



Decision



Non-statistical Hypothesis Testing

- Convicting the defendant is called rejecting the null hypothesis in favor of the alternative hypothesis.
  - The jury is saying that there is enough evidence to conclude that the defendant is guilty
    - There is enough evidence to support the alternative hypothesis
- If the jury acquits it is stating that there is not enough evidence to support the alternative hypothesis.
  - Notice that the jury is not saying that the defendant is innocent, only that there is not enough evidence to support the alternative hypothesis.

Hypothesis Testing Errors

- A Type I ($\boldsymbol{\alpha}$, alpha) Error occurs when the Null Hypothesis is rejected when it is actually true
  - Also called the Producer's Risk/ False Negative
  - $\boldsymbol{\alpha}$ is the probability that a Type I Error has occurred
    - Type I error occurs when the jury convicts an innocent person
- A Type II ($\boldsymbol{\beta}$, beta) Error occurs when the null hypothesis is not rejected when it should be rejected
  - Also called Consumer's Risk/ False Positive
  - $\boldsymbol{\beta}$ is the probability that a Type II Error has occurred
    - Type II Error occurs when a guilty defendant is acquitted

Hypothesis Testing Errors

- As $\boldsymbol{\alpha}$ increases, $\boldsymbol{\beta}$ decreases
- As $\boldsymbol{\beta}$ increases, $\boldsymbol{\alpha}$ decreases
- Increasing sample size simultaneously reduces $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$
  - Run your tests with enough samples!

Hypothesis Testing

Decision Table and Types of Risk

		Conclusion (Based on Data)	
		***Actual* Decision** is for Null Hypothesis (Fail to Reject H_0)	***Actual* Decision** is for Alternate Hypothesis (Reject H_0 in favor of H_1)
Reality (True State of Nature)	**Correct Decision** should be Null Hypothesis (Fail to Reject H_0)	Right Decision (No Error) ***Correctly*** Fail to Reject the Null $1-\alpha$ Producer's Confidence	Wrong Decision (Type I Error) ***Incorrectly*** Reject the Null α Risk Producer's Risk (Action may be taken when it shouldn't)
	Correct Decision should be Alternate Hypothesis (Reject H_0 in favor of H_1)	Wrong Decision (Type II Error) ***Incorrectly*** Fail to Reject the Null β Risk Consumer's Risk (Action may not be taken when it should be)	Right Decision (No Error) ***Correctly*** Reject the Null $1-\beta$ Consumer's Confidence

P Value and Significance Levels

- The p-value is:
  - The probability of obtaining a test statistic at least as extreme in either direction as the one observed assuming the null hypothesis (H_0) is true.
- Significance Level: The degree of risk deemed acceptable
  - The α level: probability of rejecting the null hypothesis when true
  - Usually set based on the criticality of an error
    - 0.05 if not critical (i.e., normal processes)
    - 0.01 if reasonably critical (i.e., Safety)
    - 0.001 if critical (i.e., Life vs. Death)

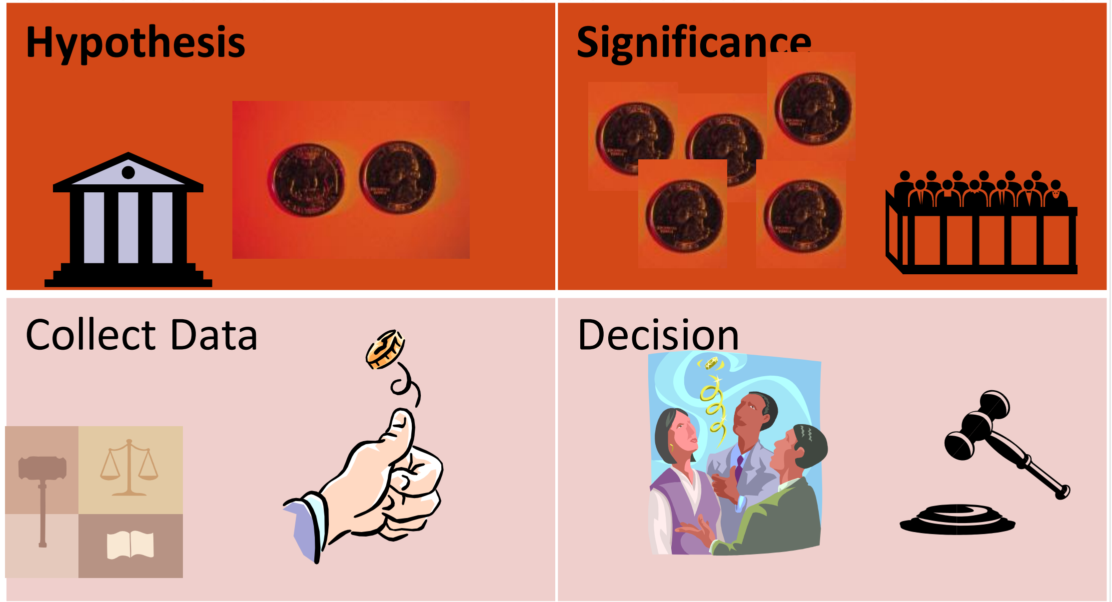
P Value and Significance Levels

- Compare the p-value to a:
  - If the p-value < a then you reject the null hypothesis.
    - If the p is low; the Null must go!
  - If p-value > a then you fail to reject the null hypothesis.
    - If the p is high; the null must fly!

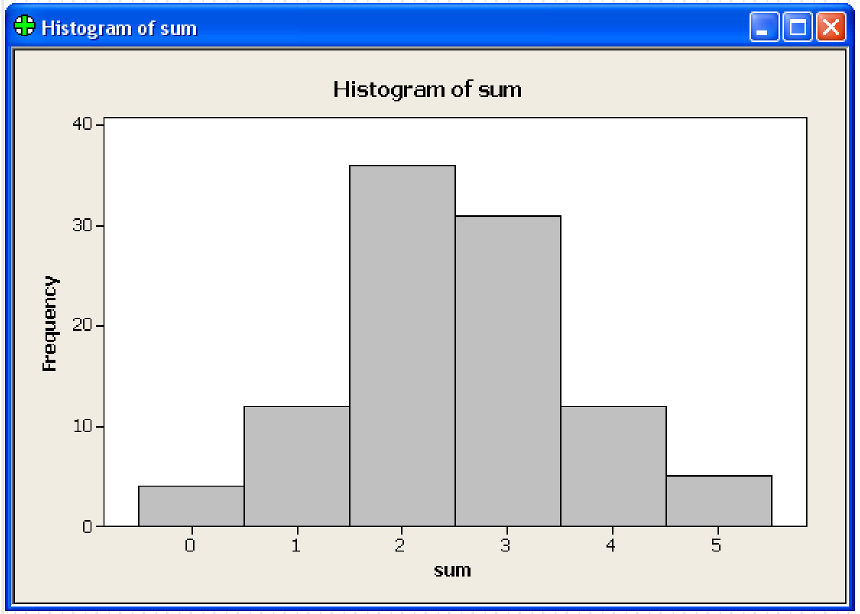
Hypothesis **H_0: Equality** **H_a: Difference**	**Significance** **Set α**
Collect Data **Experimental Design**	**Decision** **Calculate a p – value** **Compare p to α**

A Practical Illustration

- You and a friend routinely flip a coin to decide who is going to buy drinks before dinner.
- You're not sure the friend is using a fair coin, so you decide to conduct a covert experiment.
- Using the friends coin, you flip the coin five times and write down the number of times heads comes up and the number of times tails comes up.



Fair Coin Toss



Decision Table and Types of Risk

		Conclusion (Based on Data)	
		Actual **Decision** is for Null Hypothesis (Fail to Reject H_0) **Coin is not fair**	***Actual*** **Decision** is for Alternate Hypothesis (Reject H_0 in favor of H_1) **Coin is Fair**
Reality (True State of Nature)	**Correct Decision** should be Null Hypothesis (Fail to Reject H_0) **Coin is not fair**	Right Decision (No Error) ***Correctly*** Fail to Reject the Null 1-α Producer's Confidence	Wrong Decision (Type I Error) ***Incorrectly*** Reject the Null α Risk Producer's Risk (Action may be taken when it shouldn't)
	Correct Decision should be Alternate Hypothesis (Reject H_0 in favor of H_1) **Coin is fair**	Wrong Decision (Type II Error) ***Incorrectly*** Fail to Reject the Null β Risk Consumer's Risk (Action may not be taken when it should be)	Right Decision (No Error) ***Correctly*** Reject the Null 1-β Consumer's Confidence

- The experiment is simulated 100 times.
- A 0 is recorded if a coin toss comes up tails.
- A 1 is recorded if a coin toss comes up heads.
- The total number of heads is counted for each experiment.
- Even with a fair coin, all heads or all tails may come up during the experiment.

Situations for Hypothesis Testing

The following situations for conducting hypothesis tests are covered in this module:

1. Testing equality of population mean to a specific value.
2. Testing equality of means from two populations.
3. Testing equality of means from more than two populations
4. Testing equality of variances
5. Testing equality of population proportions (Binomial data)
6. Testing equality of population defect rates (Poisson data)
7. Testing for association

Important!

When conducting a hypothesis test, first determine which of these seven situations fits your application. Then follow the corresponding decision tree to determine the appropriate test. (See "References" in your manual.)

Hypothesis Tests

Mann-Whitney

1 Proportion

2 Sample t

Levene

Anderson-Darling

1 Sample Wilcoxon

1 Sample t

You must be JOKING!

Kruskal-Wallis

Chi Square

2 Proportion Test

Paired t

F-Test

Mood's Median

1 Sample Z

Example

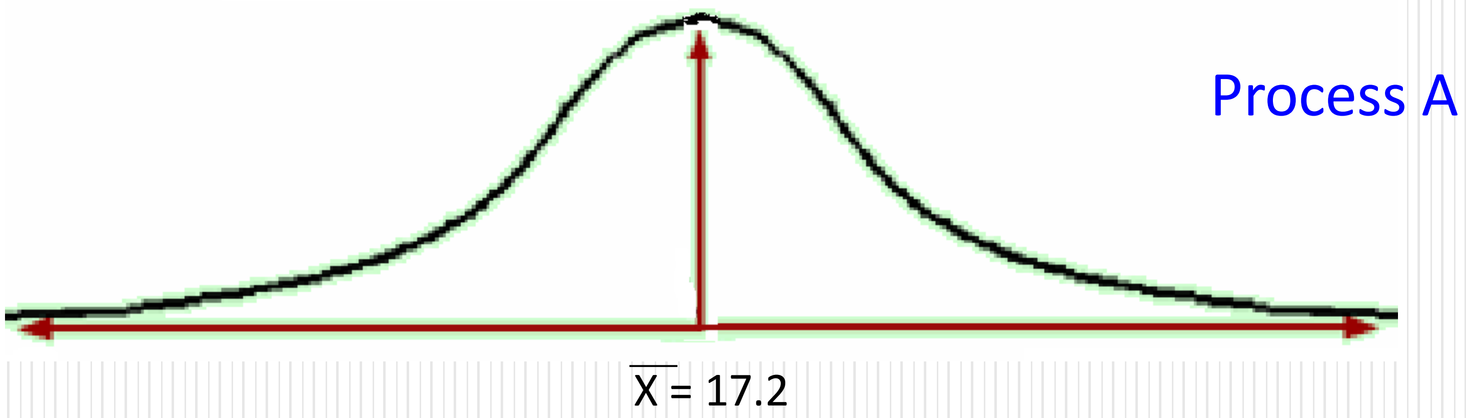
- Times to cool down a shaft from 400˚F to ambient temperature, using two different processes, have been recorded.

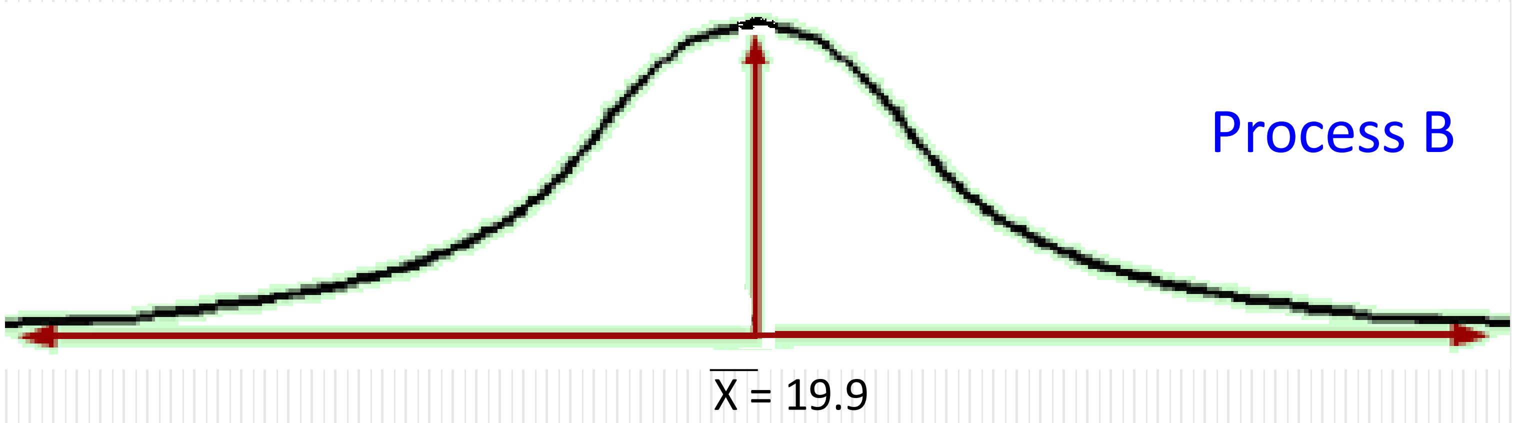
- The following table summarizes the results:

	Process A	Process B
n	11	10
$\bar{X}^{[hrs]}$	17.2	19.9
$S^{[hrs]}$	2.1	2.0

Question?

- Is the average of Process A significantly different from the average of Process B?





Answers:

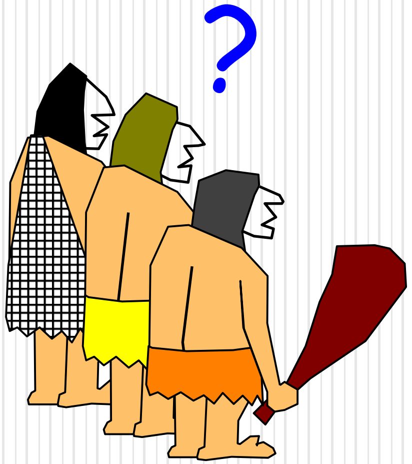
- Option #1
- We can guess!

Answers:

- Option #2: We can number crunch!

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{(S_1)^2}{n_1} + \frac{(S_2)^2}{n_2}}}$$

2 Sample t Test

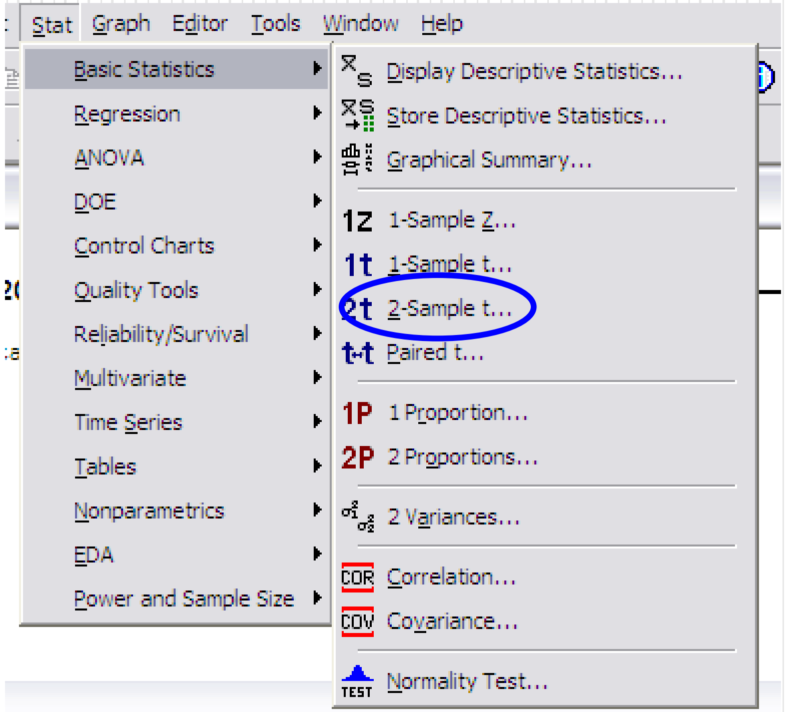


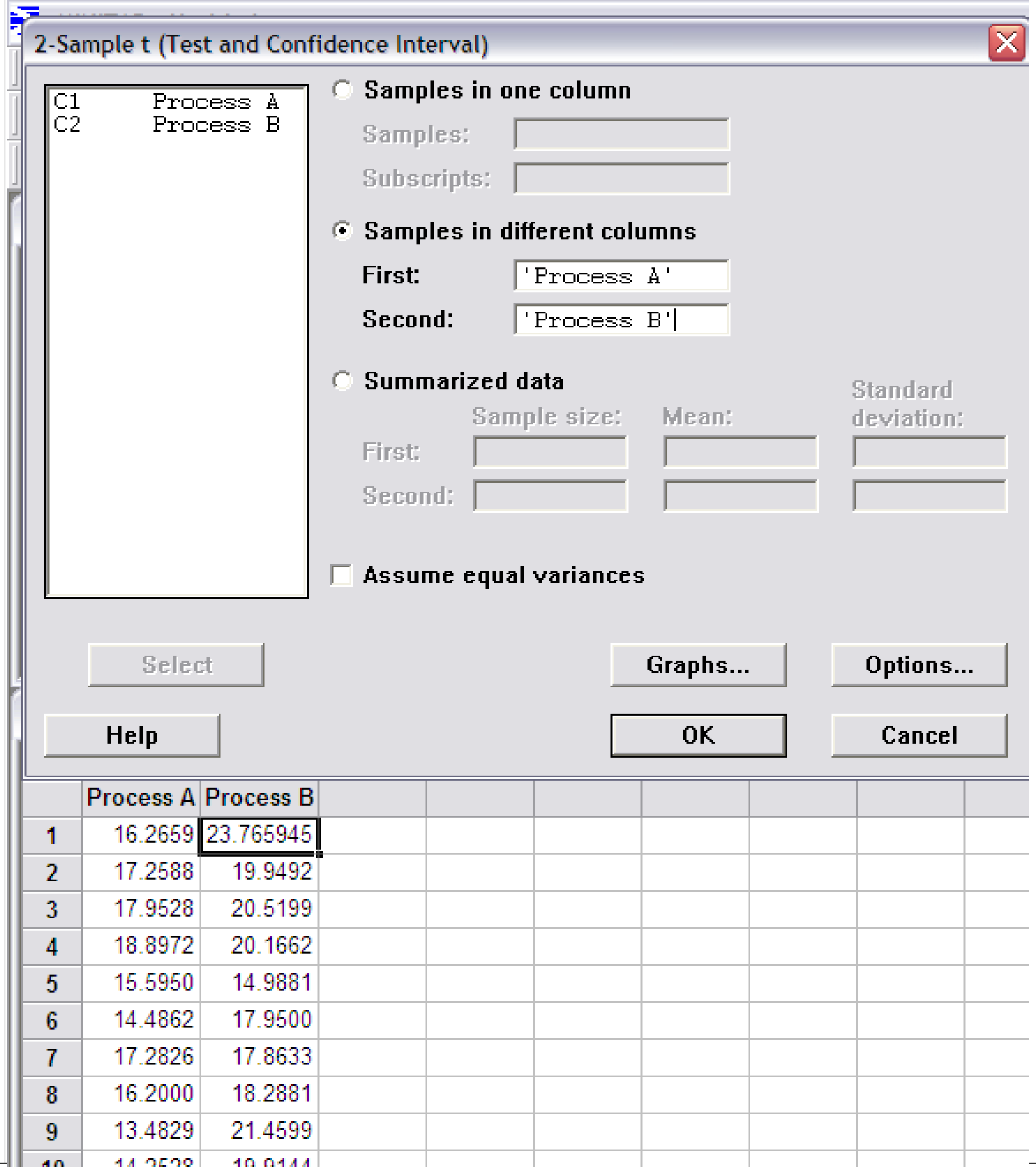
$$t = \frac{(17.2 - 19.9) - 0}{\sqrt{\frac{4.216}{11} + \frac{4.216}{10}}}$$

$\bar{X}_1$ and $\bar{X}_2$ = The Sample Means of the two populations
D_0 = The target value or Population Mean
S_1 and S_2 = The Standard Deviations of the two Samples
n_1 and n_2 = The number of test samples

Answers:

- Option #3 – Minitab!





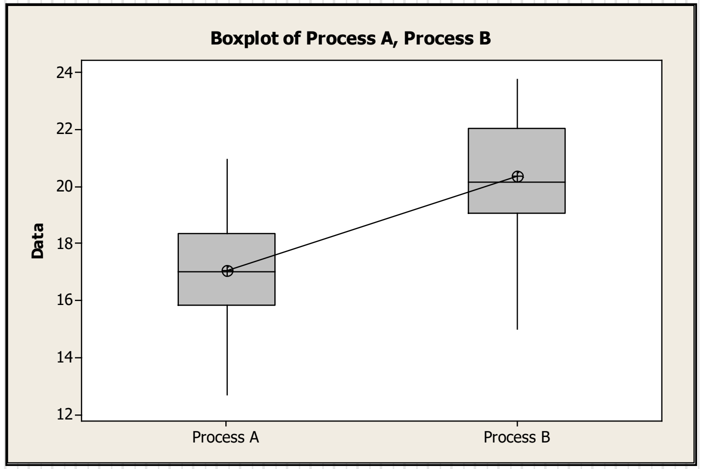
Minitab: My Hero!

Two-Sample T-Test and CI: Process A, Process B

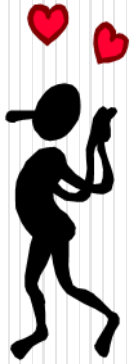
Two-sample T for Process A vs Process B

	N	Mean	StDev	SE Mean
Process A	50	17.08	1.92	0.27
Process B	50	20.38	2.03	0.29

Difference = mu (Process A) - mu (Process B)
Estimate for difference: -3.30073
95% CI for difference: (-4.08331, -2.51815)
T-Test of difference = 0 (vs not =): T-Value = -8.37 P-Value = 0.000 DF = 97



So, does the average of process A differ from the average of process B?

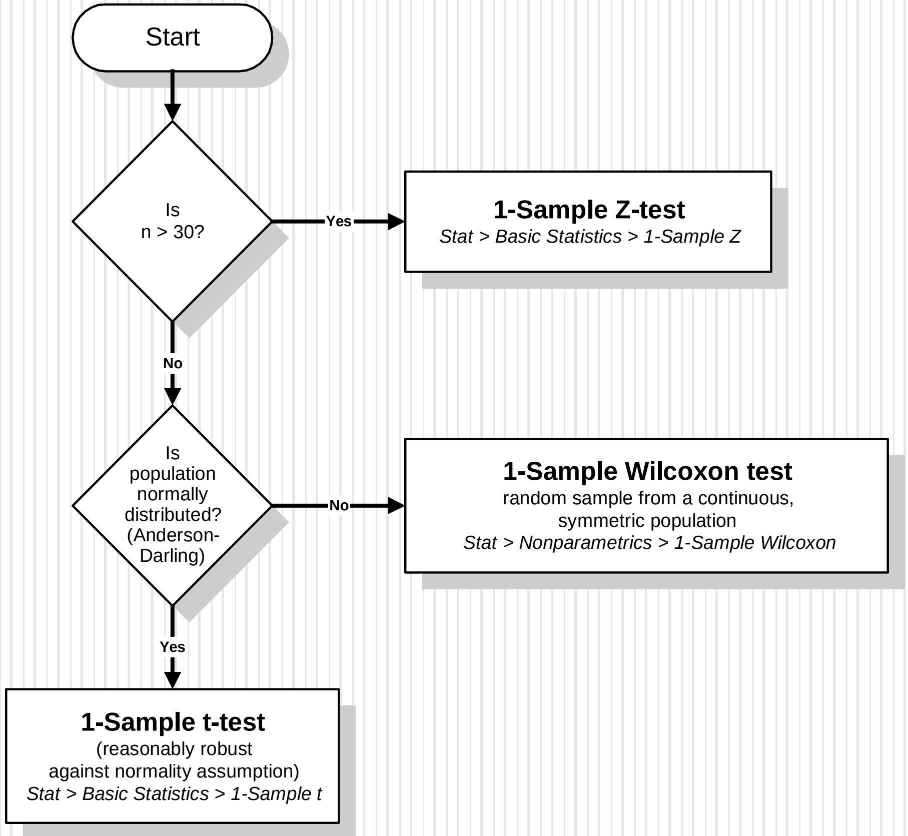


Decision Trees

Data Type	Hypothesis to be Tested	Decision Tree
Variable	Testing equality of population **MEAN** to a **specific value**	1
Variable	Testing equality of population **MEANS** from **two populations**	2
Variable	Testing equality of population **MEANS** from **more than two populations**	3
Variable	Testing equality of population **VARIANCES** from **two or more populations**	4
Attribute - Binomial (Go/No-Go)	Testing equality of population **PROPORTION OF DEFECTIVES** from **one or more populations**	5
Attribute - Poisson (Count)	Testing equality of population **PROPORTION OF DEFECTS** from **two or more populations**	6

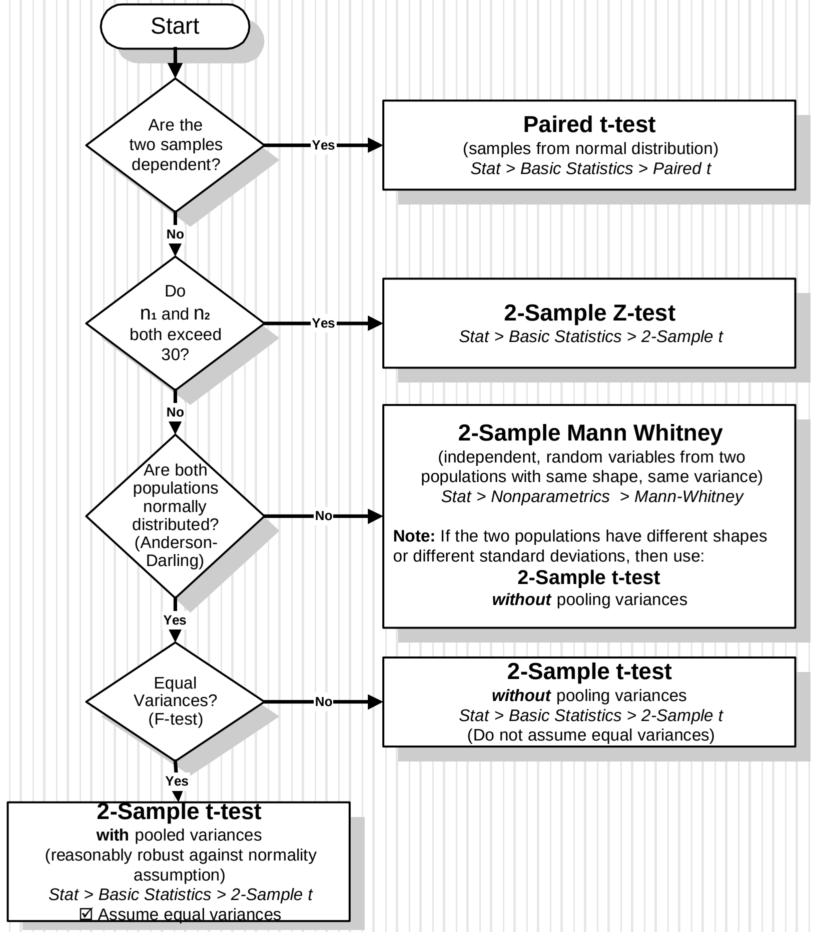
Decision Tree #1

Application:	Testing Equality of Population Mean to a Specific Value
Type of Data:	Variable (Continuous)
Example:	Has the average button diameter from the welder changed from its historical value?



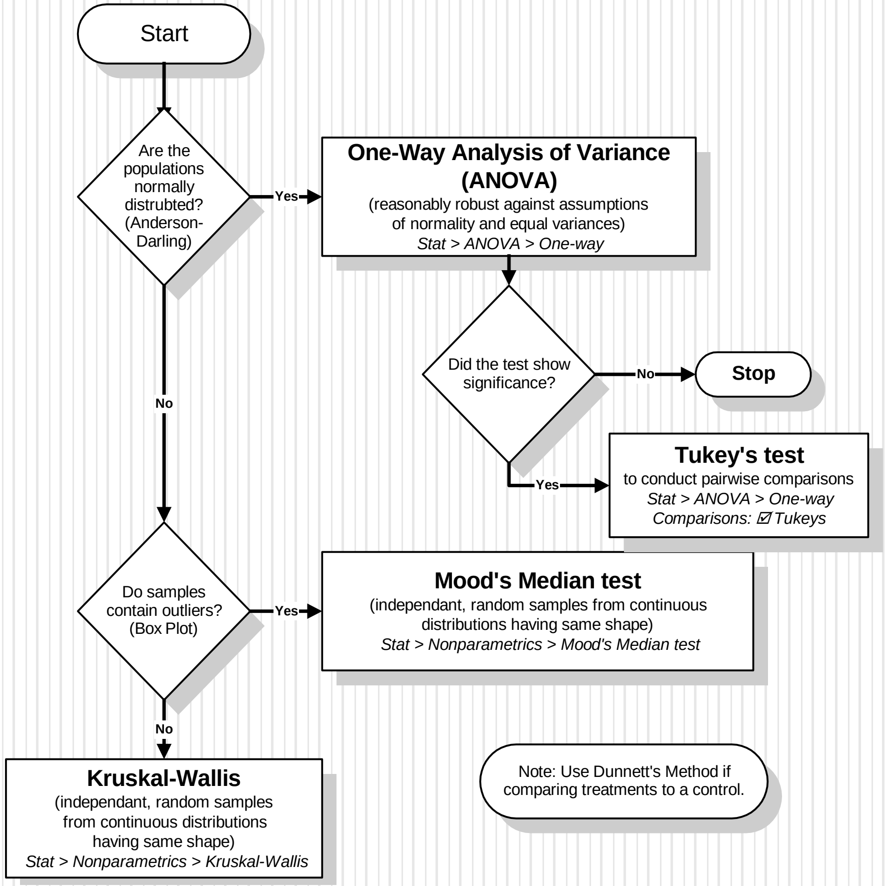
Decision Tree #2

Application:	Testing Equality of Means from Two Populations
Type of Data:	Variable (Continuous)
Example:	Is the average button diameter from Welder A different from that of Welder B?



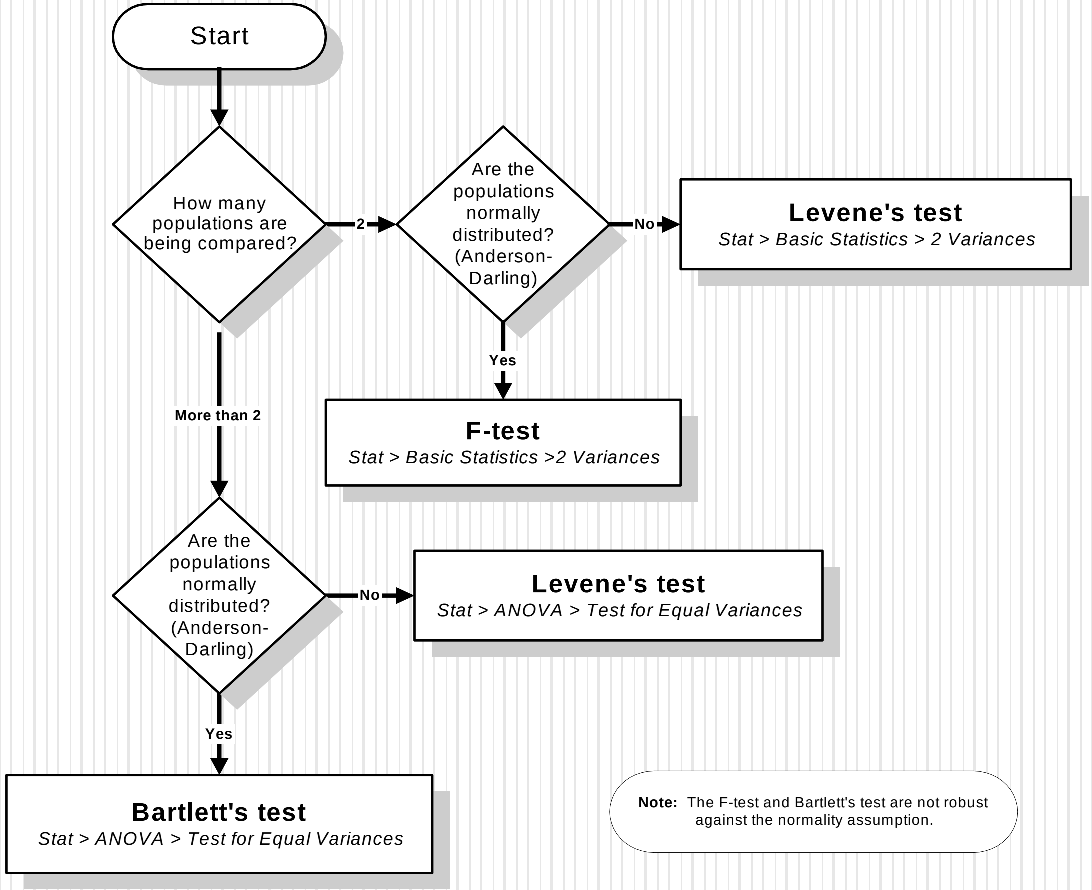
Decision Tree #3

Application:	**Testing Equality of Means from More than Two Populations**
Type of Data:	Variable (Continuous)
Example:	Do the average button diameters from Welders A, B and C differ from one another?



Decision Tree #4

Application:	Testing Equality of Variances
Type of Data:	Variable (Continuous)
Example:	Do the variances in button diameter from the three welders differ from one another?



Decision Tree #5

Application:	Testing Equality of Population Proportions
Type of Data:	Attribute (Discrete) - Binomial Distribution

Case 1:
Testing Population Proportion Against a Specific Value

Example: Has the % defective rate on Line 1 changed from its historical value?

→ ***Stat > Basic Statistics > 1-Proportion***

Case 2:
Testing Equality of Proportions from Two Populations

Example: Are Lines 1 and 2 running at the same % defective rate?

→ ***Stat > Basic Statistics > 2-Proportions***
Ho:P1=P2 no difference in popluation proportions
MiniTab - Options select pooled p

Case 3:
Testing Equality of Proportions from More than Two Populations

Example: Are Lines 1, 2 and 3 running at the same % defective rate?

→ **Use Chi-Square test**
MiniTab
Stat>Tables>Chi-square test

Decision Tree #6

Application:	Testing Equality of Population Defect Rates
Type of Data:	Attribute (Discrete) - Poisson Distribution
Examples:	1) Is the number of errors on invoices different between Dept. A and Dept. B? 2) Does the number of seat defects differ among shifts 1, 2 and 3?

Comparing two Poission Distributions

Use 2 Sample t-test

Stat > Basic Stat > 2 Sample t

Comparing more than two Poisson Distributions

Use One-Way Analysis of Variance

Stat > ANOVA > One-way

Caution
No Extreme Outliers

Decision Tree #7

Application:	**Testing for Association**
Type of Data:	Attribute (Contingency Table Data)
Example:	Does the type of defect that occurs depend on which product is being produced?

Chi-square test

Minitab:

Stat > Tables > Chi-square test

Hypothesis Testing - Steps

Step 1: Define the problem to be studied

Step 2: Define the objective

Step 3: Specify the null (H0) hypothesis

Step 4: Specify the alternative (H1) hypothesis

Step 5: Determine the practical difference

Step 6: Establish the a and b risks for the test (degree of risk acceptable), usually .05 or .01

The BB Memory Jogger, p. 152 - 153

Hypothesis Testing - Steps

Step 7: Determine the number of samples needed to obtain the desired b risk

Step 8: Calculate the probability value (p value).

- The probability of obtaining a statistic as different or more different from the value specified in the null hypothesis as the statistic computed from the data

Step 9: Compare the p value to the significance level. If the probability value is less than or equal to the significance level, then the null hypothesis is rejected. (If p-value < a reject H0)

The BB Memory Jogger, p. 152 - 153

Situation 1

Testing the Equality of …

A Population Mean to a Specific Value

NOTE

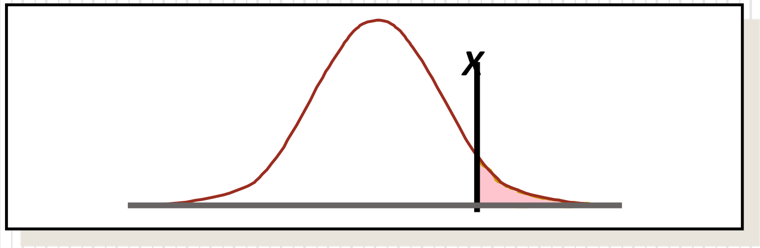
Review the Hypothesis Testing Decision Tree #1 in References section.

Definitions

- One-Tail / One-Sided Tests: To test whether a sample value is smaller or larger than a population. Hypothesis statement in a specific direction:

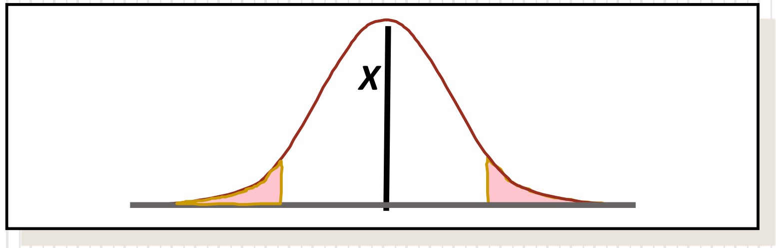
  H_0 : m ≥ X, H1: m < X

  or H_0 : m ≤ X, H1 : m > X



- Two-Tail / Two-Sided Tests: To determine if a shift has occurred in either direction. Hypothesis sets null hypothesis equal to a value:

  H_0: m = X, H1: m ≠ X



Large vs. Small Samples

- When performing hypothesis test with variable data…

- The sample size is considered large when $n > 30$.
- The sample size is considered small when $n \leq 30$.

Parametric vs. Non-parametric Tests

- Parametric tests require assumptions about the nature or shape of the population. (For example, the population must be normally distributed).

- Non-parametric tests do not require such assumptions. (For example, normality is not required).

NOTE

Whenever possible, conduct parametric tests. Non-parametric tests are less efficient, and therefore less powerful

In this module, we cover the most commonly used Parametric tests.

Testing a Claim about a Mean: Large Samples

The 1-sample Z test is used when….

- Testing the equality of a population mean to a specific value, and
- Sample size is large (n > 30)

Note: The t-test (appropriate for small sample sizes) may also be used for large sample sizes.

Exercise 1.1
Delivery Speed Example

You are attempting to assess the speed of delivery when ordering commodities utilizing two different methods. The use of a corporate credit card has traditionally been assumed to generate the best response, but that assumption is now going to be tested against a standard procurement requisition form process.

Historically, when utilizing the corporate credit card:

Average delivery time = 6 days
Standard deviation = 2 days

A random sample of size 36 was collected from the requisition process, yielding:

$\bar{x} = 4.7$ days
s = 2.0 days

Is there a difference in average delivery speed when the Corporate Credit Card process is used?

Working Together
Follow the Steps

a.) Establish both the Alternative and Null Hypotheses.

$H_0 : \mu = 6.0 \text{ days}$

(Average delivery time using blue requisitions equals the historical value of 6 days)

$H_1 : \mu \neq 6.0 \text{ days}$

(The average delivery time using blue requisitions does not equal 6 days).

b.) Determine the level of significance, α.

$\alpha = 0.05$

Refer to Exercise 1.1 in your Workbook

Collect Data and Compute P-value

c.) Randomly select a representative sample of data.

36 data were collected

$\bar{x} = 4.7$

$s = 2.0$

d.) Compute the P-value: the probability of obtaining the observed sample if the null hypothesis is true.

Using Minitab…

Select: Stat > Basic Statistics > 1-sample Z

NOTE

We know to perform a 1-sample Z test because we are testing the equality of a population mean to a specific value (6 days), and we have a large sample ($n \geq 30$).

Minitab Output

One-Sample Z: Delivery

```
Test of mu = 6 vs mu not = 6
The assumed sigma = 2

Variable          N      Mean     StDev    SE Mean
Delivery         36     4.722     1.966      0.333

Variable              95.0% CI              Z       P
Delivery       (   4.069,    5.376)     -3.83   0.000
```

e.) Compare the P-value to the level of significance, α.

P-value = 0.000 , $< \alpha = 0.05$

If p is low, the null must go!

Therefore, we reject the null hypothesis. The data provides sufficient evidence that the average delivery time when using blue requisitions does not equal the historical value of 6 days.

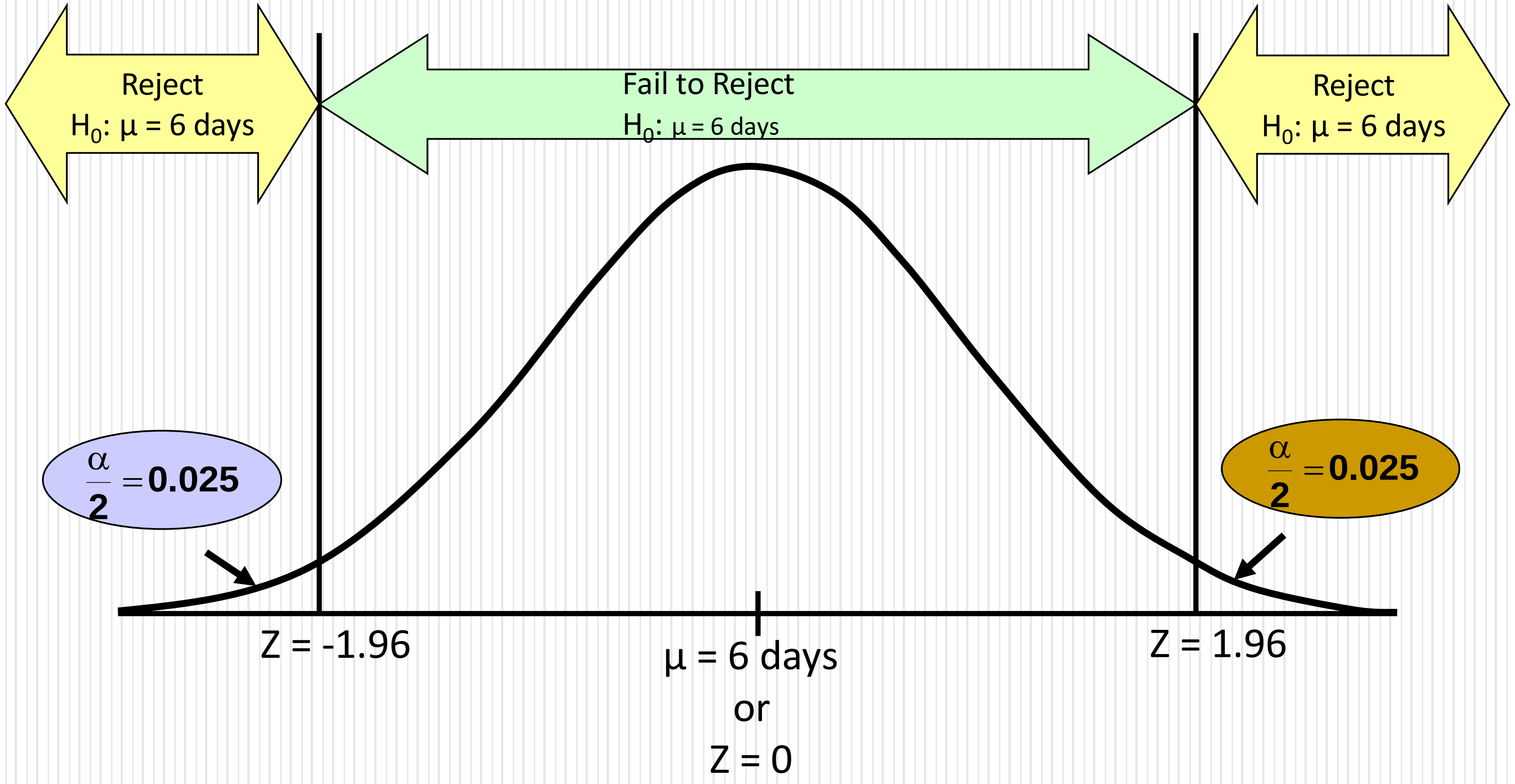
Gaining Insight

For the current example, we now show the details of how the hypothesis test is conducted.

(For future examples, we simply use the P-value provided by Minitab as was already demonstrated).

- We are testing a claim about a <u>population mean</u>.
- Because $n > 30$, the central limit theorem indicates that the distribution of sample means can be approximated by the <u>normal distribution</u>.

Distribution of Means of Delivery Time



Since this is a two-tailed test, we divide $\alpha = 0.05$ equally between the two tails.

We will reject the null hypothesis if the computed Z statistic is less than –1.96 or greater than 1.96.

Calculating the Test Statistic, Z

$\bar{x} = 4.7$
S = 2.0
n = 36

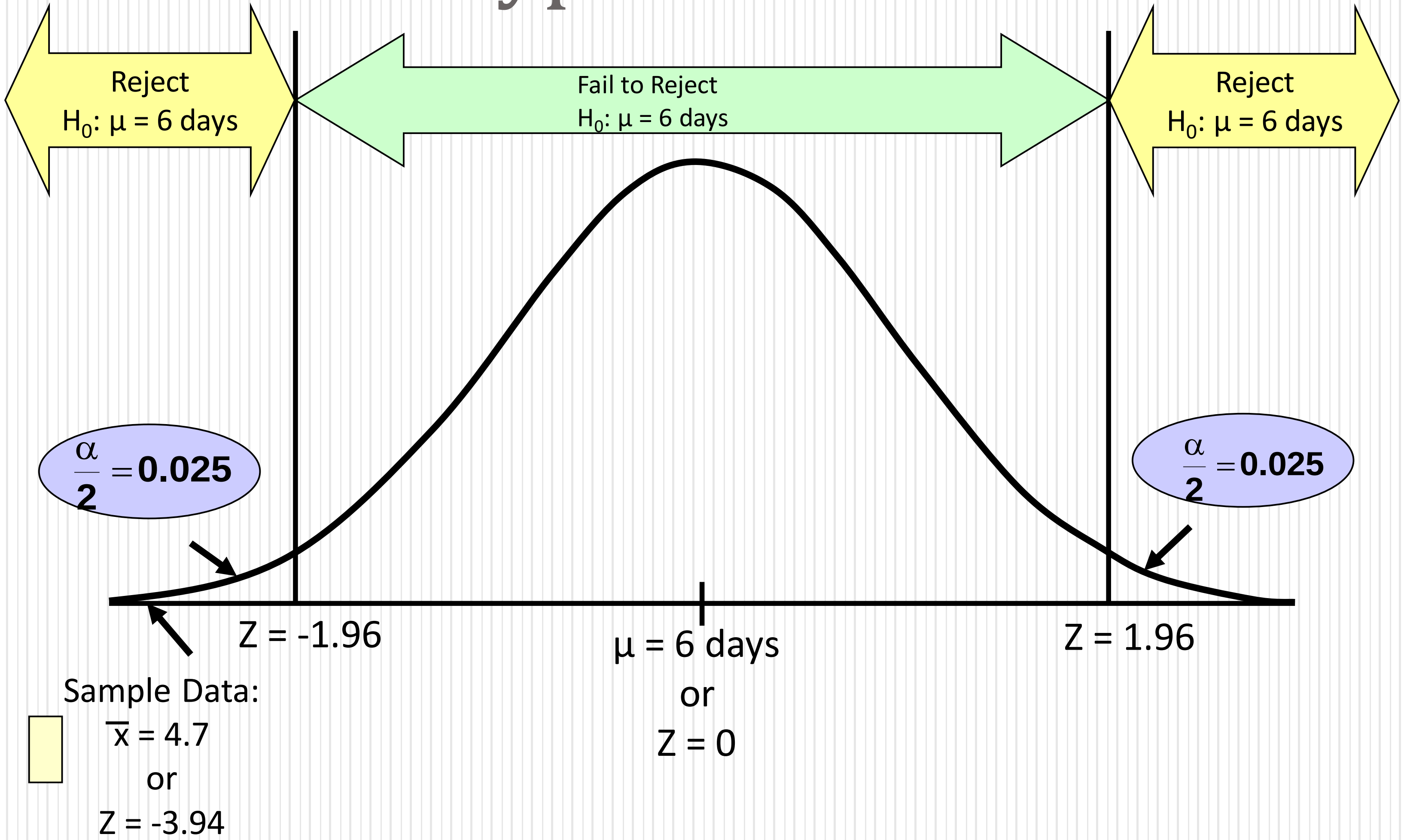
From the Central Limit Theorem:

$$\sigma_{Avg.} = \frac{\sigma_{population}}{\sqrt{n}}$$

We use the sample standard deviation, s, as our estimate of σ population. Then….

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{4.7 - 6.0}{\frac{2.0}{\sqrt{36}}} = \frac{-1.3}{0.33} = -3.94$$

Reject the Null Hypothesis



Since the computed Z value = -3.94 < -1.96, we reject the null hypothesis.

The P-Value

- The P-Value (probability value) is the probability of getting a value of the sample test statistic that is at least as extreme as the one found in the sample data, assuming the null hypothesis is true.

$$P = P(z < -3.94 \text{ or } z > 3.94) = 0.0000$$

- Since the P-value is less than $\alpha = 0.05$, we reject the null hypothesis that the average delivery time is 6 days.

Summary

In this module you have learned about:

- Understand statistical and practical significance
- Understand Hypothesis Tests
- Demonstrate the ability to conduct Hypothesis Testing