Notes:

Introduction to Hypothesis Testing

Key Learning Points

1. List the steps on hypothesis testing.
2. Formulate Null and Alternative hypotheses.
3. Explain the significance of P Value.

What is a Hypothesis Test?

Hypothesis testing enables you to test the theories (or hypotheses) regarding probable Xs so that you can identify the vital few Xs in Y = f (X). The goal in a project is to determine whether there is enough evidence to conclude that the alternative hypothesis is true.

The Null Hypothesis

In hypothesis testing, the null hypothesis (Ho) is the default assumption in the absence of evidence to the contrary. You never prove the null hypothesis, you can only reject it (when the alternative hypothesis is proven), or fail to reject it.

The Alternative Hypothesis

In hypothesis testing, the alternative hypothesis (Ha) is the hypothesis that you must prove to be true. When you can't prove the alternative hypothesis, you fail to reject the null hypothesis.

Correctly establishing the hypothesis for testing is critical to achieving correct and meaningful results.

Example: The American Justice System

The American justice system can be used to illustrate the concept of hypothesis testing.

In the American justice system innocence is assumed until guilt is proven.

It requires strong evidence, "beyond a reasonable doubt," to convict a defendant. This corresponds to rejecting the null hypothesis and accepting the alternative hypothesis.

Rejecting the null hypothesis, Ho, is a strong conclusion because it requires enough evidence to find the defendant guilty. If there is insufficient evidence, you fail to reject Ho and the verdict is not guilty. Note that you never say innocent (you never accept Ho). Always say you do not reject, or fail to reject Ho.

You are the prosecuting attorney. You must provide evidence that disproves this presumption beyond reasonable doubt.

The certainty required for being beyond this reasonable doubt is 1-α.

Ho: The Defendant Not Guilty
Ha: The Defendant is Guilty

Possible Outcomes in The American Justice System

There are only four possible outcomes in the American Justice System:

Outcome	Result	Probability
An innocent defendant is set free	This is the correct decision	
An innocent defendant is jauled	This is a Type I error	The probability of this type of error occurring is α
A guilty defendant is set free	This is a Type II error	The probability of this type of error occurring is β
A guilty defendant is jailed	This is the correct decision	

Definitions

Null Hypothesis H_o

A statement of no change or no difference. This statement is assumed to be true until sufficient evidence is presented to reject it.

Alternative Hypothesis H_a

A statement of change or difference. This statement is considered to be true if Ho is rejected.

Type I Error

The error in rejecting Ho when it is in fact true, or in saying there is a difference

Notes:

when, in fact, there is no difference.

Type II Error

The error in failing to reject Ho when it is in fact false, or in saying there is no difference when there really is a difference.

Alpha Risk

The maximum risk or maximum probability of making a Type I Error. This probability is preset, based on how much risk the researcher is willing to take in committing a Type I error (rejecting Ho wrongly), and it is usually established at 5% (or 0.05). If the P-Value is less than or equal to alpha, then reject Ho.

It is also described as the probability level below which the outcome of the hypothesis test is significant.

Beta Risk

The risk or probability of making a Type II Error, or overlooking an effective treatment or solution to the problem.

Significance Level

This is the same as alpha risk.

Significant Difference

This describes the results of a statistical hypothesis test, where a difference is too large to be reasonably attributed to chance.

P-Value

The probability of committing a Type I Error. P-Value is the actual probability of incorrectly rejecting the Null Hypothesis (Ho), i.e., the chance of rejecting the Null when it is in fact true. When the P-Value is less than or equal to alpha, reject Ho. If the P-Value is greater than alpha, fail to reject Ho.

Remember: "If the P is Low, The Null Must Go!"

Power

This is the ability of a statistical test to detect a real difference when there really is one, or the probability of being correct in rejecting Ho. It is commonly used to determine if sample sizes are sufficient to detect a difference in treatments if one exists.

Test Statistic

A Test Statistic is calculated from sample data and compared against a known probability distribution (z, t, F, etc.). This comparison determines the probability of a significant difference and is necessary to create a p-value.

Notes:

Formulating the Null and Alternative Hypotheses

Notes:

The goal in formulating the Null and Alternative Hypotheses is that you must show that the values you observed were so unlikely to come from the same population that H_o must be wrong! You want to reject H_o, thus defaulting to H_a.

Null Hypothesis (H_o)

This is the default proposition you hope to reject.

Statistical Interpretation: There is no difference in population means for Processes A and B.

Practical Interpretation: There is no difference in the average yields of the two processes, i.e., your changes did not help.

H_o: $\mu a = \mu b$

Alternative Hypothesis (H_a)

This is the proposition you want to prove.

Statistical Interpretation: The population means for Process A and Process B are different.

Practical Interpretation: The average yield of Process B is different than that of Process A.

H_a: $\mu a \neq \mu b$

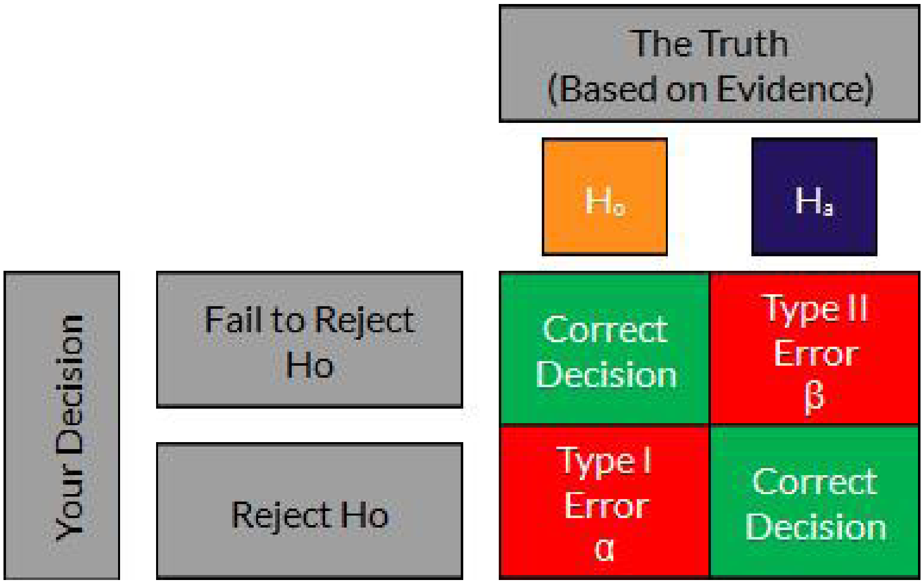
Risk

Experimenters use sample data to make conclusions about larger populations. Wrong decisions occasionally happen when testing hypotheses. Hypothesis testing methodology allows you to limit this risk.

There are two potential types of incorrect decisions:

- Type I error (α): This is called Producers' Risk. Rejecting Ho, but Ho is correct. This is a false positive, where the experimenter concludes there is a statistically significant difference, but there truly is no difference.
- Type II Error (β): This is called Consumers' Risk. Failing to reject Ho, but Ha is correct. This is a false negative, where the experimenter cannot prove a statistically significant difference, but there truly is a difference.

Alpha and Beta Risk



Significance Level (alpha)

- Significance refers to the probability level at which the outcome of the test would be considered non-random.

  You would like there to be less than 10% chance that these observations could have occurred randomly ($\alpha = 0.10$).
- Maybe you would like there to be less than 5% chance that the difference in observations occurred randomly ($\alpha = 0.05$).
- Or, conservatively, you want there to be less than 1% chance that the difference in observations occurred randomly ($\alpha = 0.01$).

P-Value

P-Value is the probability of being wrong if you say there is a difference.

If P-Value $\leq \alpha$: There is a statistical difference

1. P-Value is the actual probability of making a Type I error.
2. Ranges from 0.0 - 1.0.
3. The higher the P-Value, the more evidence you have to support the null hypothesis.
4. or most projects, you are looking for P-Values less than or equal to 0.05 to reject Ho.
5. Typically set Type I error probability of a at 0.05.
6. P-Value less than or equal to alpha means you reject the null hypothesis and accept the alternate hypothesis.

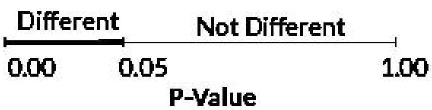
Notes:

Additional Ways of Stating Item 5

If P-Value is less than or equal to alpha, that means you reject the null hypothesis and accept the alternative hypothesis.

If the P is low, the Null Must Go!

P-Value $\leq \alpha$: Reject H_o
P-Value $\geq \alpha$: Fail to Reject H_o



Steps in Hypothesis Testing

1. State the practical problem.
2. Establish the hypotheses.
3. Decide on appropriate statistical test.
4. Set the alpha level.
5. Determine initial sample size. Calculate the power.
6. Develop the sampling plan and collect the data.
7. Use the appropriate graphical tool to explore the data.
8. Check data assumptions (change choice of tool if appropriate).
9. Run the statistical test and determine the statistical conclusion.
10. Translate statistical conclusion to a practical conclusion.

What Hypothesis Testing Proves

Hypothesis testing simply answers the question “Is there a real difference between two factors. The following example displays this practically and statistically.

Process Improvement Example

Suppose a project team working on improving a process made changes to that process. After some time, the team began to wonder if their changes made a difference. They decided to ask the following questions based on the yields of the process. Process A was the process before the change, and Process B was the process after the change.

The team asked: “After we made our changes, has yield been significantly improved? How can we determine if there is a real difference between the before and after yields?”

In basic terms, is there a difference between the average yield of Process A and

Notes:

Process B?

The Data

The team took a random sample of 10 process yields from before and after their change so they could compare them.

Process A Yield	Process B Yield
89.7	84.7
81.4	86.1
84.5	83.2
84.8	91.9
87.3	86.3
79.7	79.3
85.1	82.6
81.7	89.1
83.7	83.7
84.5	88.5

The Practical Question

The team next asked the following practical question:

Did the modifications on the process (resulting in Process B) improve the yield when compared to the original process (Process A)? Stated in another way using mean values, “Has the mean value of yield improved?”

To answer this question, the team calculated some descriptive statistics on their data.

Descriptive Statistics				
Variable	**Process**	**Total Count**	**Mean**	**St. Dev.**
Process A Yield	A	10	84.240	2.902
Process B Yield	B	10	85.540	3.650

Graphical Representation of the Data

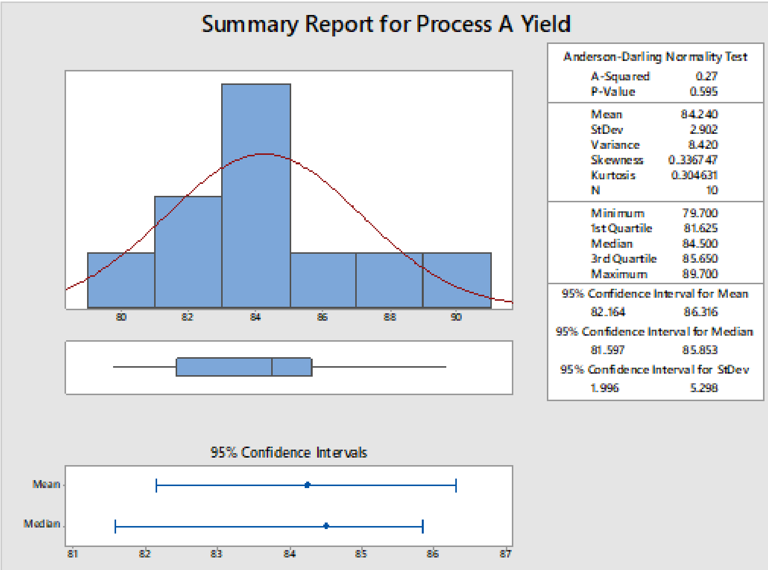
The team next asked the following question:

Does the observed difference in the mean between Process A and Process B occur by chance?

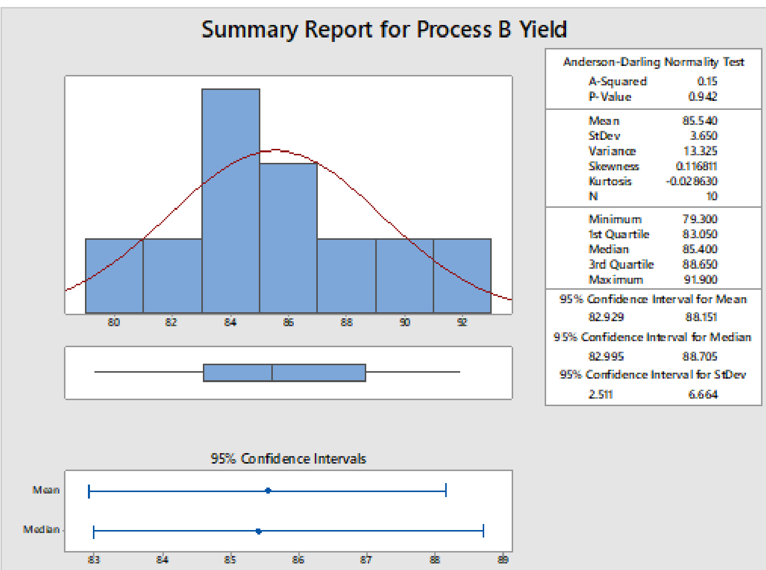
Notes:

Notes:

Graphical Summary: Process A Yield



Graphical Summary: Process B



Example Statistical Question

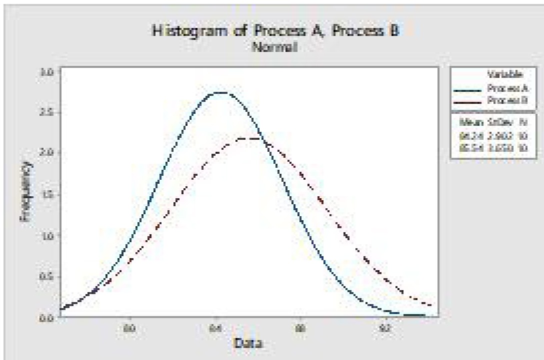
Is the mean for Process B (85.54) different enough from the mean for Process A (84.24) to be considered statistically significant?

Are the means close enough to have occurred just by chance?

In actuality, do the yield measurements from the processes represent different populations, or do the yield measurements from the processes come from one population?

$$H_o: \mu_a = \mu_b \quad H_a: \mu_a \neq \mu_b \quad \text{or} \quad H_o: \mu_a \leq \mu_b \quad H_a: \mu_a > \mu_b$$

Notes:



When Should Hypothesis Testing Be Used?

Use hypothesis testing whenever you need to prove that an observed condition is a true difference, not random chance.

- Enables testing of the hypotheses regarding probable Xs so that you can identify the vital few Xs in Y = f(x)
- Identify unplanned changes in performance for root cause analysis
- Proves improvement efforts produced superior results

Pitfalls to Avoid

- Always use randomly sampled, representative data to prevent biased results.
- Set hypotheses and significance levels before data collection.
- Be sure to use the correct hypothesis test.
- If concluding there is a statistically significant difference, ask if it is a practically significant difference.
- If concluding there is not a statistically significant difference, check for sufficient power – the ability to detect a true difference.

