Introduction to Hypothesis Testing
(Green Belt)

Module Objectives

- ❑ To *introduce* statistical hypothesis testing
- ❑ To define *null* and *alternate* hypotheses
- ❑ To demonstrate a method for *constructing* a hypothesis test

Using Data to Improve Our Understanding

- Questions asked repeatedly during a Six Sigma Project:
  - Does this X cause a change in Y? I.e. What is Y=f(X)?
  - Is my understanding of process location, spread, shape, and consistency correct?

- These questions are answered using:
  - Assumptions made based on knowledge of the process:
    - Design intent
    - Historical performance
  - Real-time data (collected during the Measure Phase)

- Current assumptions ***might be*** modified because of new data

What Is a Hypothesis Test?

- A hypothesis test is simply comparing reality to an assumption and asking, “Did things change?”

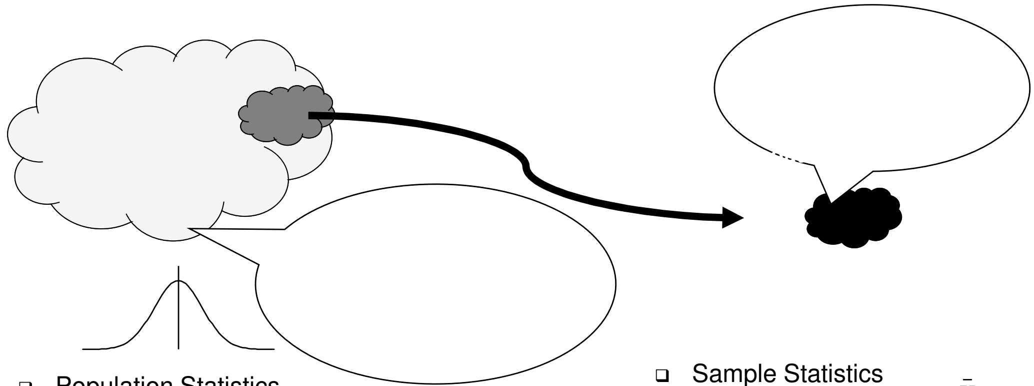
Or

- A hypothesis test is testing whether real data fits a model

Or

- A hypothesis test is comparing a statistic to a hypothesis

Review: Populations vs. Samples



- Population Statistics
  - Mean – μ
  - Standard Deviation – σ
- Sample Statistics
  - Sample Mean – $\bar{x}$
  - Estimate for Standard Deviation – **s**

Population Values Vs. Sample Statistics

	Population Parameters	Sample Statistics
Mean	μ	$\bar{X}$
Standard Deviation	σ	s
Proportion (Percentage)	P	p

Populations and Samples. Know the difference!

The Purpose of Hypothesis test

- The purpose:
  - To draw conclusions (*inferences*) about *populations* from measurements of samples
- The truth about populations
  - Populations have parameters which are fixed, and known only to God
  - Statistics is the science of guessing, from sample data, what God knows about population parameters
- Our initial guesses, before we collect data, are called hypotheses

Hypotheses

- Hypotheses:
  - Hypotheses are statements about *population parameters*, not statements about samples.
- The Null Hypothesis:
  - Abbreviated as H_0
  - Usually H_0 is a statement of ***no effect*** or ***no difference***
  - We reject or fail to reject H_0 based on statistical evidence
- The Alternative Hypothesis:
  - Abbreviated as H_a
  - A statement about a population parameter that is suspected of being true, if H_0 rejected

Null hypothesis = no difference

Hypotheses for Order Processing

- Order processing needed to improve, throughput time, due to the increase in demand and an organizational goal to hold head count steady. A change in process flow and form design was implemented to improve order cycle time. The sample mean processing time prior to the change was 125 minutes, and the sample mean processing time after the change was 118 minutes.

- What is **H_0** in common language?
  - There is no difference in average processing time between the old and the new processes
- What is **H_0** in statistical terms?
  - $\mu_{New} = \mu_{Old}$
- What is **H_a** in common language?
  - The new process completes orders faster, on average, than the old process.
- What is **H_a** in statistical tems?
  - $\mu_{New} < \mu_{Old}$

Another Hypothesis Example

- Regarding the testing of the anti-psychotic properties of an experimental drug: is the average shock avoidance response for rats treated with the new drug the same as that which has been observed from rats treated with a standard drug?

- What is **H_0** in common language?
  - There is no difference between the effects of the experimental and standard drugs
- What is **H_0** in statistical terms?
  - $\mu_{Expt} = \mu_{Std}$
- What is **H_a** in common language?
  - Rats react differently when treated with the experimental drug
- What is **H_a** in statistical terms?
  - $\mu_{Expt} \neq \mu_{Std}$

Order in the Court

- Hypothesis testing is like the American legal system where a man is assumed innocent until proved guilty
  - H_0: Defendant is innocent (Assumed)
  - H_a: Defendant is guilty (Must be proved)

Verdict	True State of Nature: Did Not Commit Crime	True State of Nature: Commited A Crime
Did Not Commit Crime	**Correct Verdict**	**Guilty Goes Free**
Commited A Crime	**Innocent Person Convicted**	**Correct Verdict**

What kind of mistakes are made?

Hypothesis Testing Errors

- Type I error
  - Rejecting the null hypothesis when it is, in fact, true
  - Also known as producer risk
  - The probability of a Type I error is denoted by α $(0< \alpha <1)$
- Type II error
  - Accepting the null hypothesis, when it is, in fact, false
  - Also known as consumer risk
  - The probability of a Type II error is denoted by β $(0< \beta <1)$

Conclusion Drawn	True State of Nature: Ho	True State of Nature: Ha
Ho	**Correct Decision**	**Type II Error**
Ha	**Type I Error**	**Correct Decision**

Evidence is required to reject null hypotheses.

Hypothesis Example

- To meet an established performance standard the average daily balance of non-interest generating funds must be less than Rs,.10 Lacs.

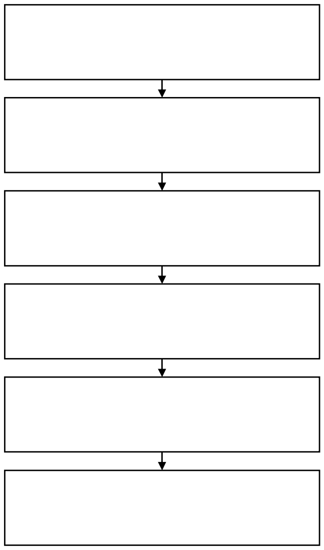
  - What is **H_0** in common language?
    - The average daily balance of non-interest generating funds is equal to Rs.10Lacs.
  - What is **H_0** in statistical terms?
    - $\mu_{actual} = \mu_{Std}$ = Rs.10Lacs
  - What is **H_a** in common language?
    - The average daily balance of non-interest generating funds is less than Rs.10Lacs
  - What is **H_a** in statistical terms?
    - $\mu_{actual} < \mu_{Std}$ = Rs.10Lacs

Hypotheses for Order Processing

- The Order processing department needs to improve throughput time due to the increase in demand and an organizational goal to hold head count steady. A change in process flow and form design was implemented to improve order cycle time. The sample mean processing time prior to the change was 125 minutes, and the sample mean processing time after the change was 118 minutes.

  - What is **H_0** in common language?
    - There is no difference in average processing time between the old and the new processes.
  - What is **H_0** in statistical terms?
    - $\mu_{New} = \mu_{Old}$
  - What is **H_a** in common language?
    - The new process completes orders faster, on average, than the old process.
  - What is **H_a** in statistical terms?
    - $\mu_{New} < \mu_{Old}$

A Hypothesis Testing Flow Chart



Null Hypothesis – H_0: No difference
Alternate – H_a: Difference exists (<, ≠, >)

Usually 95% (α = 0.05)

What type of data? Variable or Attribute? Means, sigmas, counts? How many sets?

Or if the sample size is fixed, calculate the power (1-β) from the fixed sample size.

Run the test. Collect the data. Analyze in Minitab. Interpret p-value against α.

What does it mean to accept H_0? To accept H_a? Verify the results before final implementation

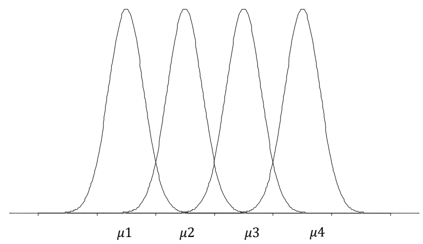
Definitions

- **Null Hypothesis (H_0)** – Statement of no change or difference; assumed to be true until sufficient evidence is presented to reject it.
- **Alternate Hypothesis (H_a)** – Statement of a change or difference; assumed to be true if the null hypothesis is rejected
- **Alpha Risk** – The maximum probability of making a Type I error. This probability is established by the experimenter at 5%.
- **Beta Risk** – The risk or probability of making a Type II error.
- **Test Statistic** – A standardized value (z, t, etc.) which represents the likelihood of H_0, and is distributed in a known manner such that the probability for this value can be determined.

Tests of Means

ANOVA One-way –

Hypothesis testing for means



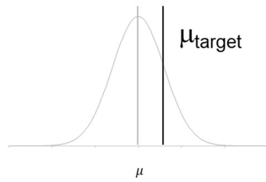
Module Objectives

- ❑ Understanding Means Testing
- ❑ Demonstrating method to carry out means testing
  - ❍ One Sample
  - ❍ Two samples
  - ❍ Multiple samples

Testing Means

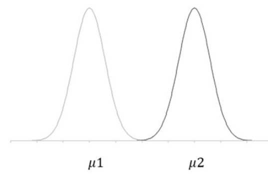
H_o: $\mu = \mu_{target}$

H_a: $\mu < \mu_{target}$
$\mu > \mu_{target}$
$\mu \neq \mu_{target}$



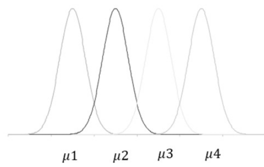
H_o: $\mu_1 = \mu_2$

H_a: $\mu_1 < \mu_2$
$\mu_1 > \mu_2$
$\mu_1 \neq \mu_2$



H_o: $\mu_1 = \mu_2 = \ldots = \mu_n$

H_a: @ least one is different from another



- **Testing for one mean**
  - **Z-test for large samples or when s is known**
  - **t-test for smaller samples or when σ is unknown**
- **Testing for two means**
  - **2-sample t-test**
- **Testing for three or more means**
  - **1-way ANOVA**

Exercise – Mean Vs. Target (t-test)

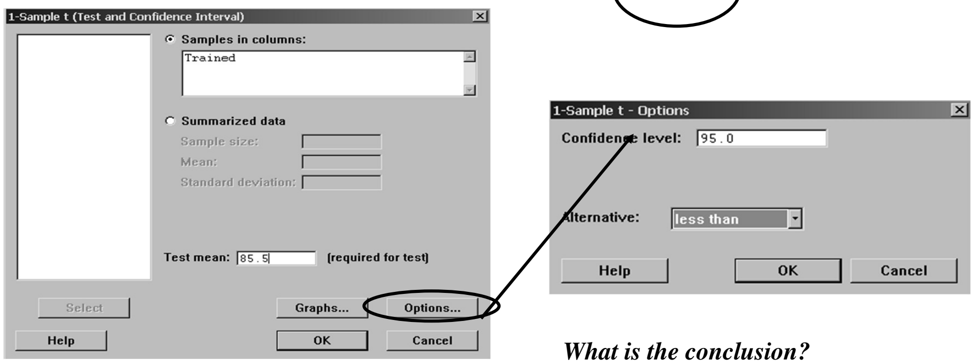
- A tuition center manager evaluates the maths scores of 15 students who have recently attended maths coaching. The class-wide average score is 75.5. Their typical scores were 80.

1. Did the coaching have an impact on scores vs. typical performance? Use $\alpha = .05$.

2. Conduct a 1-sample t-test in Minitab to check your results. The data are in worksheet MeanExamples.mtw. (Column: Trained) (Hint: Stat > Basic Statistics > 1-sample t…)

1-sample t-test – Minitab Results

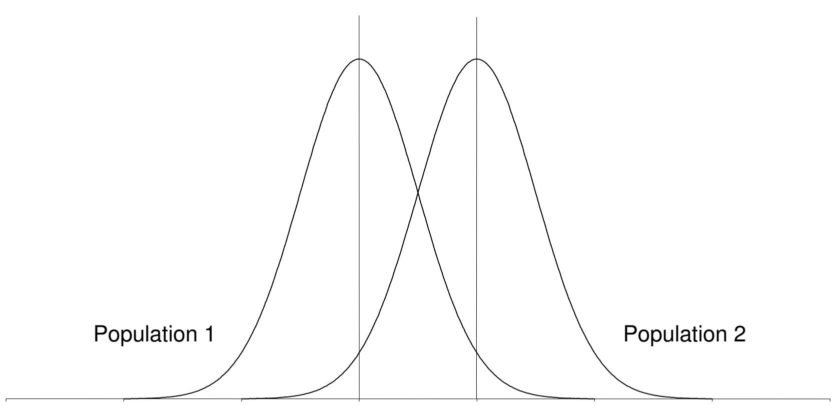
```
Test of mu = 85.5 vs mu < 85.5

Variable          N      Mean     StDev   SE Mean
Trained          15     80.79      8.93      2.31

Variable      95.0% Upper Bound        T        P
Trained                   84.86    -2.04    0.030
```



What is the conclusion?

Comparing Two Means

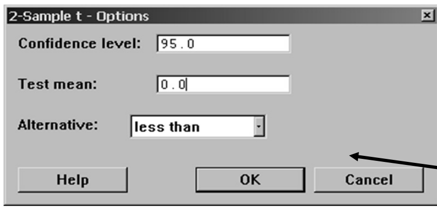
- Fundamental analysis question: "Is there a real difference between these two groups?"
- The 2-sample t-test is the hypothesis test for a difference in means.

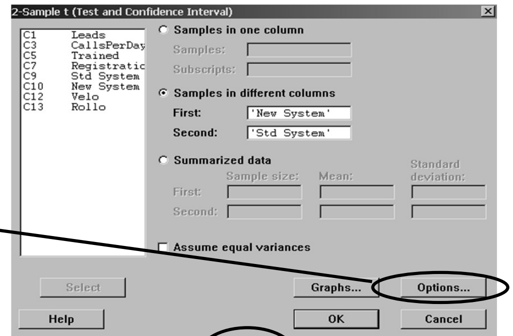


2-sample t-test – Minitab Results

Two Sample T-Test and Confidence Interval

```
Two-sample T for New System vs Std System
             N   Mean  StDev
New System  10  44.30   4.64
Std System  10  46.50   5.40
```





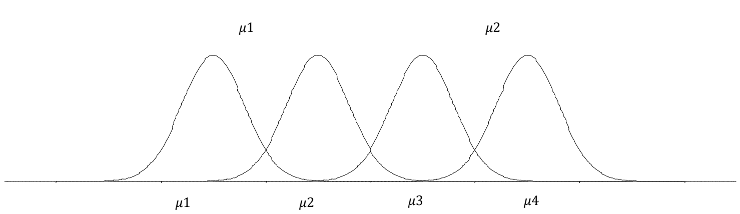

```
Estimate for difference:  -2.20
95% upper bound for difference: 1.71
T-Test of difference = 0 (vs <): T-Value = -0.98  P-Value = 0.171  DF = 18
Both use Pooled StDev = 5.04
```


Testing Means

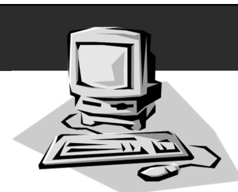
- Testing means
  - 1-way ANOVA

H_o: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

H_a: at least one μ is different from another



ANOVA Introduction



- Example:
  - The Finance Department of a company wishes to compare effectiveness of different software to increase speed of their data processing. Speed was evaluated on software packages procured from four different suppliers and tested with similar data. One set of data was processed with existing package.
  - Ten different operators were deputed to evaluate each package and the effectiveness was measured on a scale of 1:10 and recorded.

- Problem:
  - Were any of the packages different from each other and the existing package?

? ***How can we analyze this data?***

Solution – Analysis of Variance

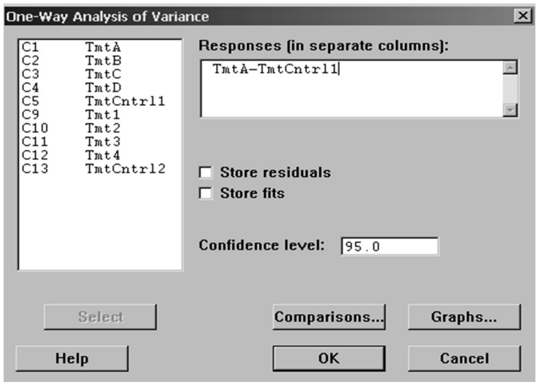
- ANOVA is a method of detecting differences between multiple means of samples

- Why is it called Analysis of Variance?
  - ANOVA compares/analyzes variances
    - Variance within a group
    - Variance between groups

ANOVA is the mathematics behind the intuitive evaluation

ANOVA in Minitab – A Quick Demo

- In Minitab select Stat>ANOVA>One-way (Unstacked)…



$H_0: \mu_1 = \mu_2 = \ldots = \mu_j$

H_a: @ least one μ_j is different