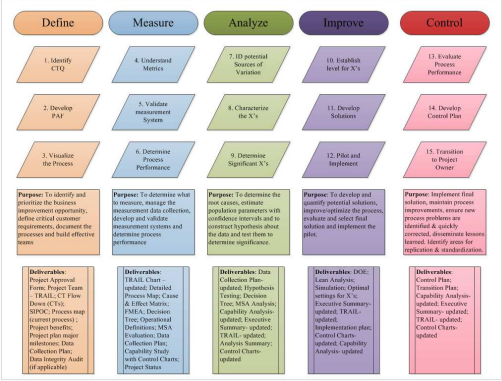
LEANOhio

Transforming the Public Sector

Green Belt

1 Variable Testing

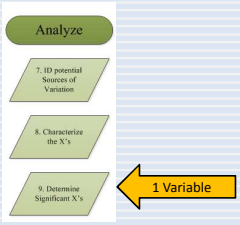
SIMPLER. FASTER. BETTER. LESS COSTLY.



DMAIC Flow

- Analyze:
  - ID Potential Sources of Variation
  - Characterize the X's
  - Determine Significant X's

Analyze Purpose: To determine the root causes, estimate population parameters with confidence intervals and to construct hypothesis about the data and test them to determine significance.



SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Agenda

- Review Types of Data: Discrete Continuous
- Review of 4 Characteristics of Data
- 1-Variable Testing Definition-Uses
- Parametric vs. Non-parametric

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Types of Data Review

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Types of Data

- It's important to know what type(s) of data you're collecting… Different types of data require different types of tests

Discrete		Continuous	
Discrete Data can be Qualitative AND Quantitative		Continuous Data is Quantitative ONLY	
↓		↓	
Nominal	Ordinal	Interval	Ratio
Finite Numbers No Fractions Here!		Usually Associated with Measurements & Fractions Work Here!	
← Weakest		Strongest →	

Discrete Data

Can only have the values
2, 3,4,5,6,7,8,9,10,11,and 12

Continuous Data

Time

Height

Discrete: Nominal Data

Counties

Nominal: Name Only
You can't perform arithmetic operations, like addition or subtraction.

Discrete: Ordinal Data

Very Interested	Somewhat Interested	Neutral	Not Very Interested	Not at All Interested
5	4	3	2	1

Likert Scale

BIG 10 Conference Football
Revenue Premium Based Brand Equity Rankings

University	Rank
The Ohio State University	1
Pennsylvania State University	2
University of Michigan	3
University of Nebraska	4
Michigan State University	5
University of Wisconsin	6
University of Iowa	7
University of Illinois	8
Purdue University	9
University of Minnesota	10
Indiana University	11
Northwestern University	12

Ordinal: Implied Order
Cannot state whether the intervals between each value are equal

Continuous: Interval Data

Interval: Like Ordinal except we can say the intervals between each value are equally split. No natural zero point.
Without a true zero, it is impossible to compute ratios

Continuous: Ratio Data

Ratio: Interval data with a zero starting point.

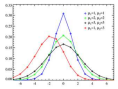
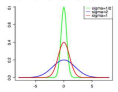
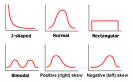
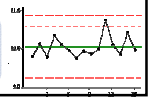
Data Summary

- Nominal data used to "name," or label a series of values
- Ordinal data provide good information about the order of choices, such as in a customer satisfaction survey
- Interval data give us the order of values + the ability to quantify the difference between each one
- Ratio data give us the ultimate–order, interval values, plus the ability to calculate ratios since a "true zero" can be defined

4 Characteristics of Data

The Key Characteristics of a Distribution

Where on the measure scale does the data appear to gather or "clump"?	• What is the **center** of the data?
How does the data distribute around the center?	• What is the **spread** of the data?
What values are more frequent and less frequent?	• What is the **shape** of the data?
How do the above characteristics behave over time?	• What is the **stability** of the data?

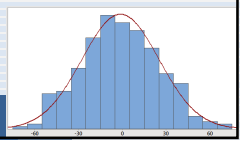
Normal vs. Non-normal

SIMPLER. FASTER. BETTER. LESS COSTLY.

LEANOhio

Is your data normal?

- Some tests assume normality and are sometimes called “parametric tests”
  - “Parametric” implies that a distribution (shape) is assumed for the population - commonly the Normal Distribution (Bell Curve)
  - Advantage of a parametric test is higher statistical power (more robust tests)



SIMPLER. FASTER. BETTER. LESS COSTLY.

Six Sigma Mantras

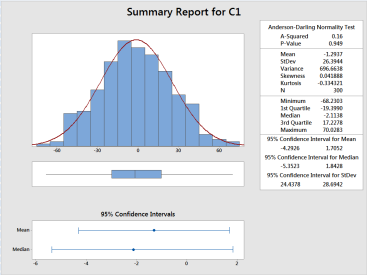
- Zero Mantra: Bring Home the Bacon
- First Mantra: Y is a Function of X
- Second Mantra: Variation is Evil
- Third Mantra: Shift Happens!
- **Fourth Mantra: Stat, Basic Stat, Graphical Summary**

SIMPLER. FASTER. BETTER. LESS COSTLY.

LEANOhio

Is your data normal?

Stat-Basic Stat-Graphical summary in minitab to test for normality or graph data and see if it follows the bell curve



What if my data is not normal!!!!

- Give up
- Change the data until it is normal
- Call 911???
- NO!

Nonparametric Testing

- If you find that your data is not normal we have a series of tests that can be used. These test are known as Nonparametric tests
- These tests give you more flexibility because they do not assume a shape. This allows you to use them with **any** data

Nonparametric Testing

- If we can use nonparametric tests for any data – why don't we use them every time….
- Because nonparametric tests are not as robust or statistically sound as parametric tests (often do not have as many bells and whistles)

SIMPLER. FASTER. BETTER. LESS COSTLY.

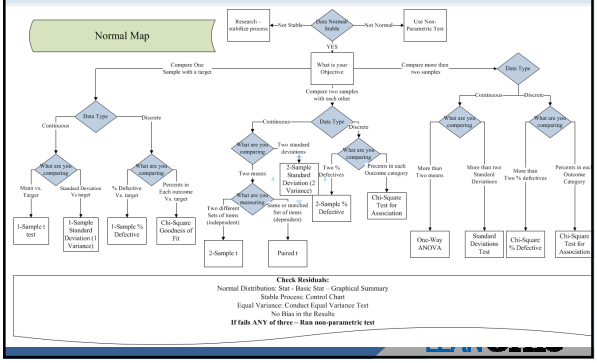
LEANOhio

Normal Road Map

SIMPLER. FASTER. BETTER. LESS COSTLY.

LEANOhio

Normal Road Map



Normal Map Assumptions

- Is Data Stable? What Tool?
  - Research
  - Lean Up Process
- Is Data Normal

Normality Tests

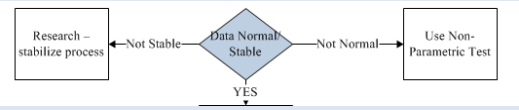
- Normality Tests is just another Hypothesis Test
- In essence we are conducting a hypothesis test in order to know if my data is normal or non-normal (to know which test to run)

Anderson Darling Test

- The most common Normality Test is the Anderson Darling Test.
- HO: Data = Normal Data
- HA: Data≠ Normal Data

Non-Normal Data

- If data not-normal: Use non-parametric test



SIMPLER. FASTER. BETTER. LESS COSTLY.

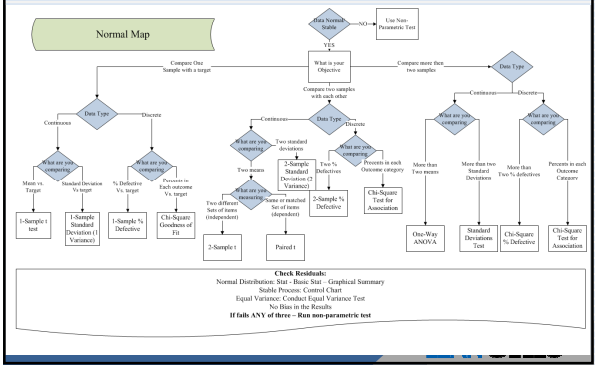
LEANOhio

1 Variable Testing: Continuous

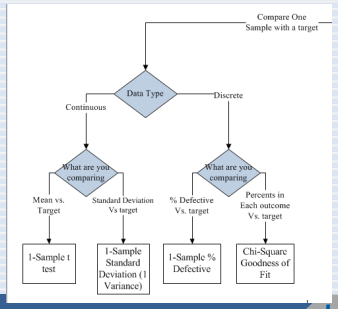
SIMPLER. FASTER. BETTER. LESS COSTLY.

LEANOhio

Normal Road Map



1-Variable
Campare One Sample with a Target



SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

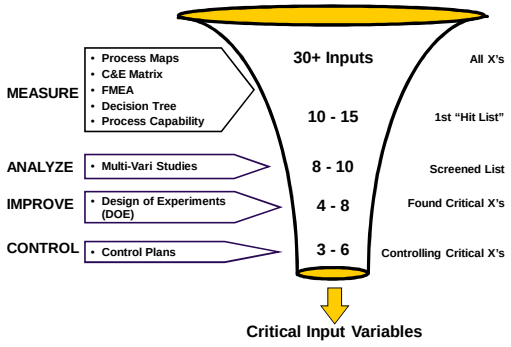
"1-Variable" Testing

1-Variable" testing involves evaluating: ***1 column of data at a time***

- Your project will require you to collect data on a number of different variables
- With this class of tools, we are studying any one variable, but only one at a time
- We are **NOT** trying to study the effect of an input variable on an output: that is 2-Variable testing

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

The Funneling Effect



1 Variable Testing: Example

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Industry Standards

Are there any examples of Industry Standards currently used in State Government?

Let's use a Call Center Example

- What type of metrics measure success or - are there benchmarks?
- Let's review data from a call center

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Possible Industry Standards Examples

- Call Length (Public Safety 2.5 minutes per call)
- Number of Abandoned Calls
- Number of Complaints
- Number of Calls Received
- Number of Busied Calls

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Questions we can ask in the 1 Variable World

- Is the DOP better then the Standard for:
  - Call Length Time: 2:30
  - Abandoned Calls: 0
  - Busied Calls?
  - Number of Calls Handled?
  - Incoming Calls?

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Hypothesis Form

What is our Y?

What Type of Data is our Y?

What is the Y?	**Call Length**	What Type of Data?	**Continuous**
What is the X?	**NA**	What Type of Data	**NA**
			NA
What type of tool would you use?			
Is my data Normal? (Outliers?)			
	Ho: Call Length = Standard (150 seconds)		
	Ha: Call Length (not equal) Standard (150 seconds)		
P value: (0.05)			
Interpret results:			

What is our X?

Trick Question – No X!

What is our Ho?

What is our Ha?

Is DOP different then the standard?

How do we test for Normality?

- In minitab: Stat-Basic Stat-Graphical Summary (if p is less than 0.05 then distribution is non normal)
- Graph (Histogram or Probability Test)

Hypothesis Test for Normality

- Ho: Data = Normal
- Ha: Data ≠ Normal (non-normal)

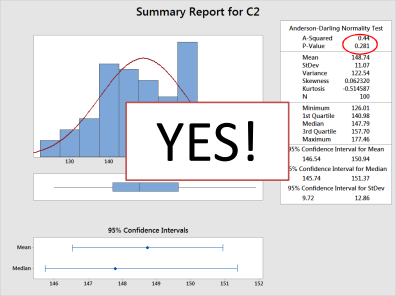
SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Normality Test

- Is our data normal?

40

Is the data Normal?



P- Value is .281

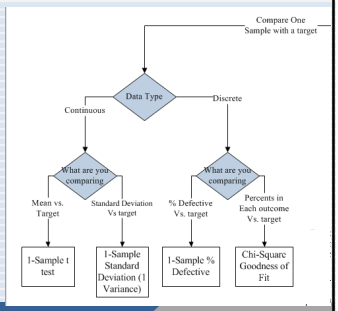
Let's review the hypothesis test

- Remember our mantra: **If the p is low the Ho must go!**
- The p value for this data is: 0.281 which indicates that the data is Normal!!!

Is the process better than the standard?

Check your Road Map:

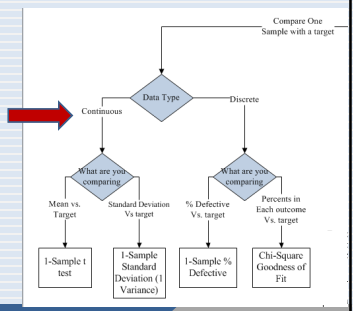
Is your data: Continuous or Discrete?



SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Is the process better than the standard?

Call Length is Continuous Data



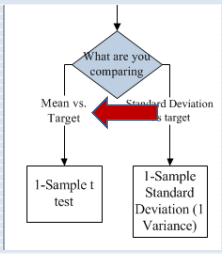
SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Is DOP different then the standard?

What are we comparing?

In this scenario we are comparing the center call time vs. an industry standard

Mean vs. Target



SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

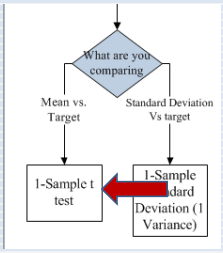
Is the process better than the standard?

- 1 sample t test is comparing the mean to the standard or customer request
- 1 sample sign test is comparing the median to the standard or customer request
  - This test is less powerful then the 1 sample t test
- 1 sample z test if the population standard deviation is known
  - This is pretty RARE
- 1 variance test is comparing standard deviation to the standard or customer request standard deviation
  - This is rare too because management often does not think in standard deviation terms.

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Is DOP different then the standard?

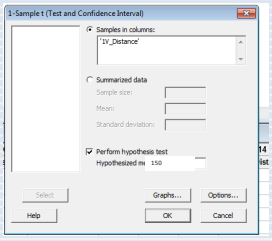
For this example we would run a 1 sample t test



SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Minitab Instructions

- Stat-Basic Stat – 1 sample t test
- Add in industry or customer standard in Hypothesized Mean
- In this example 150



SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Mini-tab says

One Sample T: Ave Call Length

- Test of mu = 150 vs not = 150

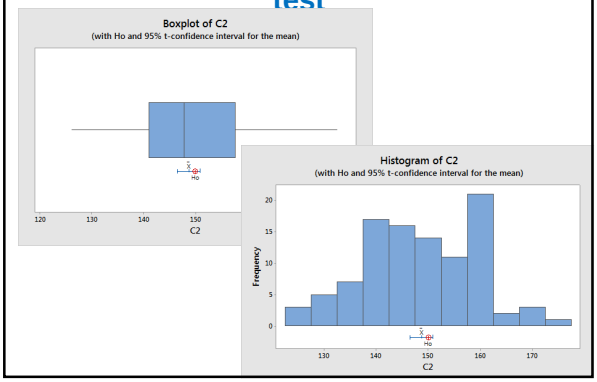
Sign test of mean = 150.0 versus not = 150.0

	N	Mean	StDev	SE Mean	95% CI	T	P
Ave Call Length	100	148.74	11.07	1.11	(146,150)	-1.14	0.258

P = 0.258

Remember if the p is low the Ho must go! (Below 0.05)
The p is not low…so the means are statistically equal and there is NO statistical difference.

Graphical Displays from 1 sample t test



1 Variable Testing Activity

SIMPLER. FASTER. BETTER. LESS COSTLY.

LEANOhio

1 Variable Testing Activity

- Cup Stacking Activity
- You vs. the Standard (Is there a statistical difference?)

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Cup Stacking Standard

The world record is 7.43 seconds. Are you better then the standard?

https://www.youtube.com/watch?v=usAWvTKpIIs

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Hypothesis Testing Steps

1: Write your Hypothesis: (Ho and Ha)

Hypothesis Form

2: Collect Data (a sample of reality)

Operational Definitions

3: DECIDE:

What does the evidence suggest?

Reject H_o? or Don't Reject H_o?

Step One: Write your Hypothesis

Are you better then the Standard: 7.43 Seconds

- HO: Time to cup stack = 7.43 seconds
- HA: Time to cup Stack ≠ 7.43 seconds

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Hypothesis Form

Hypothesis Testing Form			
What is the Y?		What Type of Data?	
What is the X?			
How many "levels" do			
Is my data Stable?			
What type of tool wou			
Is my data Normal? (Outliers?)			
Comparing Median or Means?			
Ho: (=)			
Ha:			
P value: (0.05)			
Interpret results:			

Good practice to begin completing form when conducting hypothesis tests!

Hypothesis Form

What is our Y?

What Type of Data is our Y?

What is our X?

Trick Question – No X!

What is our Ho?

What is our Ha?

Hyp			
What is the Y?	**Time to Stack Cups**	What Type of Data?	**Continuous**
What is the X?	**NA**	What Type of Data	**NA**
have?		**NA**	
Is my data Stable?			
What type of tool would you use?		**1 Variable**	
Is my data Normal? (Outliers?)			
Comparing Median or Means?			
Ho: (=)	**Time to cup stack = 7.43 seconds**		
Ha:	**Time to cup Stack ≠ 7.43 seconds**		
P value: (0.05)			
Interpret results:			

Step Two: Collect Data

Operational Definitions:

- Start
- End
- Time Accuracy
- Time Mechanism
- Timer Instructions

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Step Two: Collect Data

Go Collect Data:

- Ten Trials per Operator
- Collect name and time for each
- Collect number of defects

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Step Three: Analyze Data/Decide

Review data in minitab

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

1 Variable Testing: Discrete

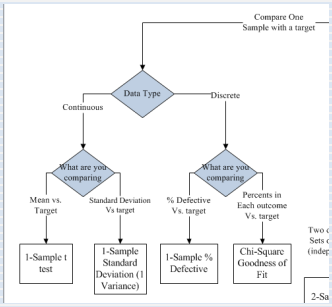
SIMPLER. FASTER. BETTER. LESS COSTLY.

LEANOhio

Discrete Data

Let's Look at the 1-Variable Road Map

Look at the data: is your variable Continuous or Discrete.



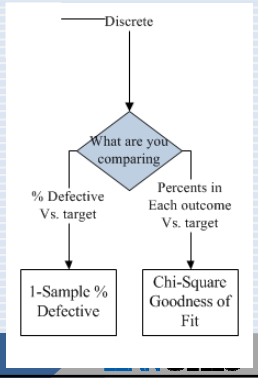
SIMPLER. FASTER. BETTER. LESS COSTLY.

LEANOhio

Discrete Data

Pick your test based on your data:

- For Defectives data, run 1 proportion test (ex. First aid kits—good/bad or pass/fail)
- For Defects data, run 1-sample Poisson rate test (defect rates)
- For percents in each outcome vs. target, run Chi-Squared or goodness of fit



SIMPLER. FASTER. BETTER. LESS COSTLY.

Analyzing Discrete Data

- Remember that discrete data is “information poor” relative to continuous data.
  - Using discrete data requires much larger sample sizes (often times 100 or more data points) to achieve statistical power comparable to continuous data tests

Questions?