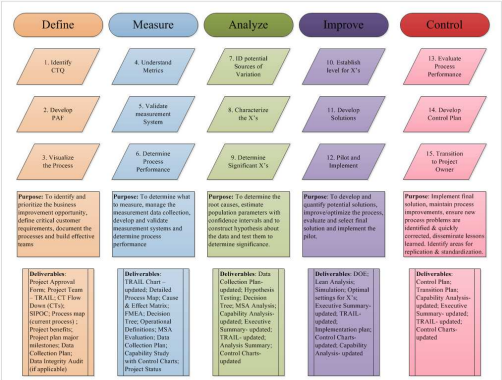
LEANOhio

Transforming the Public Sector

Green Belt

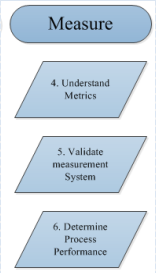
Intro to Measure

SIMPLER. FASTER. BETTER. LESS COSTLY.



DMAIC Flow

- Measure:
  - Understand Metrics
  - Validate Measurement System
  - Determine Process Performance



SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Measure Purpose

- To determine what to measure
- Manage the measurement data collection
- Develop and validate measurement systems
- Determine process performance.

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

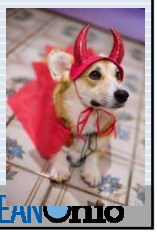
Measure Deliverables

- TRAIL Chart Updated (Week One)
- Detailed Process Map (Week One)
- **Cause and Effect Matrix (Week Two)**
- **FMEA (Week Two)**
- **CT Flowdown (Week Two)**
- Operational Definitions (Week One)
- **MSA Evaluation (Week Two)**
- Data Collection Plan (Week One)
- **Capability Study (Week Two)**
- **Control Charts (Week Two)**

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

The Fundamental Principle of Six Sigma

- Variation is Evil!
  - But…
- These causes of variation can be identified, measured, prioritized, and then controlled



SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Why Measure

- Our Processes:
  - Can be chronically off target
  - Can be occasionally off target
- Measure to see if it is a systematic problem or exception

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Remember our first Mantra…

$$Y = f(X_1, X_2, X_3 \ldots)$$

"Y is a function of X"

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

How do Statistics Relate to Six Sigma?

- Statistics deals with the study of variation
- Statistical tools will allow us to quantify the problem, investigate the root causes, develop a solution, and confirm the improvements over time.
- **Let the Data Take You There!**

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

First Lets Take a Moment and Think about Data

- Define your Primary Metric for your Project
  - What was the Primary Metric for the DOP example?
- Define your Secondary Metrics
  - What are the Secondary Metrics for the DOP example?

Be intentional when you collect data:
Think it through

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Secondary Metrics

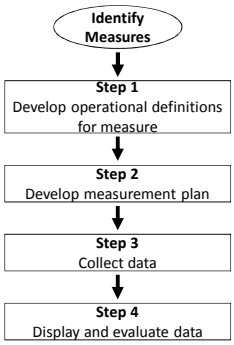
Keep Me Honest!

Helps avoid sub-optimization!

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Data Collection Method

Measurement management starts with a data collection methodology.



Develop a Measurement Plan - Types of Data

- Before data collection starts, classify the data into different types: continuous or discrete.
- This is important because it will:
  - Provide a choice of data display and analysis tools
  - Dictate sample size calculation
  - Provide performance or cause information
  - Determine the appropriate control chart to use
  - Determine the appropriate method for calculation of Sigma Level

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Types of Data

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Two Basic Types of Data

Discrete (Categories)

Ordinal
- Count defects
- # approved
- # of errors
- Rank

Nominal
- Yes/No
- Group A, B..
- Good/Bad
- Pass/Fail

Continuous (Measured on a continuum)

- Time
- Money
- Weight
- Length
- Speed
- Temperature

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
↓	↓
Finite Numbers No Fractions Here!	Usually Associated with Measurements & Fractions Work Here!
← Weakest	Strongest →

Discrete Data

JOB APPLICA

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

Discrete Data

Discrete Variable	Possible values for the variable
The number of defective needles in boxes of 100 diabetic syringes	0,1,2, …, 100
The number of individuals in groups of 30 with a Type A personality	0,1,2, …, 30
The number of surveys returned out of 300 mailed in a customer satisfaction study.	0,1,2, … 300
The number of employees in 100 having finished high school or obtained a GED	0,1,2, … 100
The number of times you need to flip a coin before a head appears for the first time	1,2,3, …

Continuous Data

Time

Height

Continuous Data

Continuous Variable	Possible Values for the Variable
The length of prison time served for individuals convicted of first degree murder	All the real numbers between ***a*** and ***b***, where ***a*** is the smallest amount of time served and ***b*** is the largest.
The household income for households with incomes less than or equal to $30,000	All the real numbers between ***a*** and $30,000, where ***a*** is the smallest household income in the population
The blood glucose reading for those individuals having glucose readings equal to or greater than 200	All real numbers between 200 and ***b***, where ***b*** is the largest glucose reading in all such individuals

Identify Discrete or Continuous Data

- Classify each set of data as discrete or continuous

1. The number of suitcases lost by an airline	
2. The height of corn plants	
3. The number of ears of corn produced	
4. The number of green M&M's in a bag	
5. The time it takes for a car battery to die	
6. The weight of luggage on an airplane	

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	

Characteristics of Discrete Data

Discrete Data	
Discrete Data can be Qualitative AND Quantitative *Finite Numbers – No Fractions Here!*	
Nominal	Ordinal
- The weakest level of data. - Made up of values that are distinguished by name only. - There is no standard ordering scheme.	Stronger than nominal data, but not as strong as Continuous data. - Like nominal data, also made up of values that are rank ordered values or ranked groups. - Different than nominal data because there is an ordering scheme.

Discrete: Nominal Data

Counties

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

Nominal Scale

Qualitative Variable	Possible nominal level data values for the variable
Blood Types	A, B, AB, O
State of Residence	Alabama, …, Wyoming
Country of Birth	United States, China, other

Time to weigh in!

Examples of Discrete Data

Discrete Data	
Discrete Data can be Qualitative AND Quantitative *Finite Numbers – No Fractions Here!*	
Nominal	Ordinal
The colors of M&M candies is an example of nominal level data. This data is distinguished by name only. There is no agreed upon order of this data to distinguish between colors and determine which color is "better".	Movies are classified as 2 thumbs up, 1 thumb up, or 0 thumbs up. There is an order here. A movie that receives 2 thumbs up is better than a movie that receives 1 thumb up. How much better is a movie that receives 2 thumbs than 1 thumb? Ordinal data tells you there is an order to distinguish between classes but does not tell you the distance between "good, better and best".

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

Ordinal Scale

Qualitative Variable	Possible Ordinal level data values
Automobile Sizes	Subcompact, compact, intermediate, full size, luxury
Product rating	Poor, good, excellent
Baseball team classification	Class A, Class AA, Class AAA, Major League

Characteristics of Continuous Data

Continuous Data	
Continuous Data is Quantitative ONLY	
Interval	Ratio
Interval data has an ordering scheme, but unlike ordinal data, the differences between data is meaningful and can be measured. Arithmetic operations are possible. However, interval data lacks a zero starting point, a characteristic unique to Ratio data.	Ratio data has an ordering scheme, has differences that are meaningful, can be measured, and arithmetic operations are possible. Ratio data has a natural starting point and can identify absolute zero.

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

Interval Scale

Interval Variable	Possible Scores
IQ scores of students in Green Belt Training	100… (the difference between scores is measurable and has meaning but a difference of 20 points between 100 and 120 does not indicate that one student is 1.2 times more intelligent)

Examples of Continuous Data

Continuous Data	
Continuous Data is Quantitative ONLY	
Interval	Ratio
Boiling temperatures of different liquids vary. We can tell whether a temperature is higher or lower than another, so we can put them in order. If water boils at 212 degrees and another liquid boils at 284 degrees, the differences between data are measurable and meaningful. However, because there is no natural starting point, ratios between 2 data values are meaningless. Is 75 degrees 3 times as hot as 25 degrees? No, because the ratio of 75 to 25 is meaningless here.	Four people are randomly selected and asked how much money they have with these results: $21, $50, $65 and $300. Is there an order to this data? Are the differences between the data values meaningful? Can we calculate ratios based on this data?

Continuous: Ratio Data

Ratio: Interval data with a zero starting point.

Data Summary

SIMPLER. FASTER. BETTER. LESS COSTLY.

LEANOhio

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

SIMPLER. FASTER. BETTER. LESS COSTLY.

LEANOhio

Data Summary ✓

Provides:	Nominal	Ordinal	Interval	Ratio
"Counts," aka "Frequency of Distribution"	✓	✓	✓	✓
Mode, Median		✓	✓	✓
The "order" of values is known		✓	✓	✓
Can quantify the difference between each value			✓	✓
Can add or subtract values			✓	✓
Can multiple and divide values				✓
Has "true zero"				✓

Samples vs. Populations

SIMPLER. FASTER. BETTER. LESS COSTLY.

LEANOhio

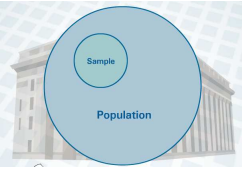
Samples vs. Populations

SIMPLER. FASTER. BETTER. LESS COSTLY.

LEANOhio

Sample vs. Population

- A sample is a portion of the population from which it is selected
- If selected properly, the sample will display characteristics similar to the parent population



Why Sample?

- Resource Constraints - Due to resource constraints (e.g., time and money) we often cannot measure everything in a given population
- Conclusion Validity - After a point, additional measurements will not have a meaningful return to the validity of the conclusion

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

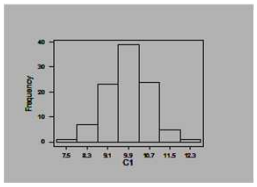
Sampling Strategies

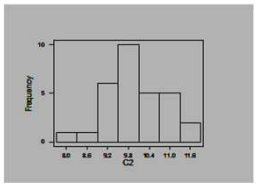
- Random
- Systematic-nth
- Stratified

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Sample Information

- Population (100 Items)
- Sample (30 Items)





The Sample distribution should mirror the population distribution.

One Sample Calculator - Continuous Data

- Note that this calculator uses the Standard Deviation from the population in place of proportion
- If the standard deviation is unknown, use Range/5

Calculation for Sample Size - Continuous Data

Population Size (N)	1,000	
Confidence Level (1-alpha)	0.95000	(Percent expressed as decimal)
Power Level (1-beta)	0.80000	(Percent expressed as decimal)
Standard Deviation (s)	0.50000	(Best prior estimate or Range/5)
Required Precision (+/- delta)	0.10000	(Margin of Error)
Initial Sample Size	197	(Assumes infinite population)
Corrected for Population	165	(Correction for finite population)
Recommended Sample Size	165	

Central Limit Theorem

- When studying averages, as the sample size gets larger, the distribution of averages more closely approximates normality
- The average of the sample averages is approximately equal to the population average

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

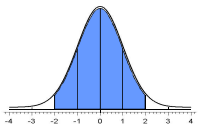
Confidence Intervals

- Confidence intervals tell us how confident we can be that a sample is representative of the population.
- If independent samples are taken repeatedly from the same population, and a confidence interval calculated for each sample, then a certain percentage (confidence level) of the intervals will include the unknown population parameter.

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Why Worry About Confidence Intervals

- Statistics such as the mean and standard deviation are only estimates of the population mean and standard deviation, and are based on one sample only.
- Because there is variability in these estimates from sample to sample, we can quantify our uncertainty using statistically based Confidence Intervals (C.I.'s)



Confidence Intervals

- In order to estimate confidence intervals, we need to know the population distribution.
  - While we never know the exact distribution, we know through experience and theory that certain mathematical distributions can be safely used to represent any conceivable situation.
- There are a dozen or so of these most common distributions, we almost always rely on just a small handful of them.

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Confidence Intervals

- Confidence intervals are usually calculated so that this percentage is 95%, but we can produce 90%, 99%, 99.9% (or whatever) confidence intervals for the unknown parameter.

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Another Mantra

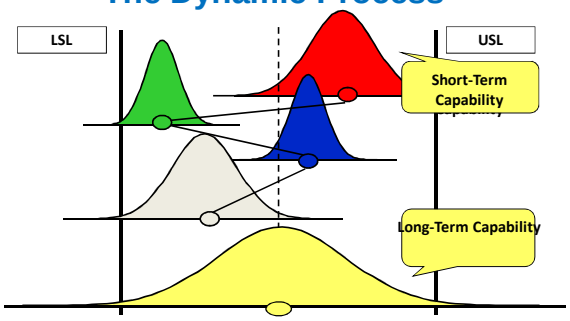
- Why is data so important?

SHIFT Happens

– http://www.youtube.com/watch?v=FdTOFkhaplo

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

The Dynamic Process



Questions

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio