SEI's Transition Partner

Continental Reaching Solutions

Statistical Thinking and Correct Use of Statistical Techniques in Achieving CMMI High Maturity

Dr. Bin Cong
SEI Certificated High Maturity Lead Appraiser
SEI authorized CMMI Instructor
CRS Tech
Professor and Director
Master of Software Engineering
Cal State Univ. at Fullerton
bcong@fullerton.edu or bcong@crs-tech.com

Challenges in implementing CMMI HM

- Limited successful implementations shared in the community
- Lack of knowledge in correctly applying statistical knowledge to product development and process improvement
- The descriptions of HM practices defined in the model do not provide very meaningful guidance
- … …

Outline: provide a flavor on:

- What is statistical thinking all about?
- Misuse of Statistics
- Understanding variation
- On building and using a process performance model
- Summary

A Couple of Definitions (SEI)

- **Statistical management**

Management involving statistical thinking and the correct use of a variety of statistical techniques, such as run charts, control charts, confidence intervals, and etc.

- **Quantitative management**

The process of using data from statistical management to help the project

  - **predict whether it will be able to achieve its quality and process-performance objectives, and**
  - **identify what corrective action (if any) should be taken.**

Definition (Triola)

- **Statistics**

  a collection of methods for planning studies and experiments, obtaining data, and then organizing, summarizing, presenting, analyzing, interpreting, and drawing conclusions for the population based on the sample data.

Population and Sample in Our Context

- Population: the set of projects (past, present, and future) defined by the operational definition of the relevant measure(s).
- Sample: a set of data randomly selected from projects in the population.
- Objective: draw conclusion for the population based on the sample data.

Inferential Statistics

- The two major applications of inferential statistics involve the use of sample data to (1) estimate the value of a population parameter, and (2) test some claim (or hypothesis) about a population.
- Estimating values of these important population parameters: proportions, means, and variances.
- Determining sample sizes necessary to estimate those parameters.

Rare Event Rule for Inferential Statistics

If, under a given assumption, the probability of a particular observed event is exceptionally small, we conclude that the assumption is probably not correct.

Example: Gender Choice – Statistical Thinking

- **Gender Choice, according to advertising claims, allowed couples to "increase your chances of having a girl up to 80." What should we conclude about the effectiveness of Gender Choice if the 100 babies include**

  (Triola)

a. **52 girls?**

b. **97 girls?**

Can you answer this question?

- **The passing rate of acceptance tests is 85% for the software group of ABC company. Suppose 30 out of 100 such tests failed to pass in the past three months, should we worry about it?**

How about answering this one?

29 peer reviews done at ABC group were selected to be analyzed. 13 of the 29 found at least 1 Major/Severe/Critical error. Can we conclude that less than 50% of the peer reviews conducted in the group can detect at least one major or worse error?

1.96* ((0.45*0.55)/29)**0.5 -> 0.18

(0.27, 0.63)

Is there really a difference?

Example: For the sample data listed in the Table below, use a 0.05 significance level to test the claim that the proportion of black drivers stopped by the police is greater than the proportion of white drivers who are stopped.

Racial Profiling Data

	Race and Ethnicity	
	Black and Non-Hispanic	White and Non-Hispanic
Drivers stopped by police	24	147
Total number of observed drivers	200	1400
Percent Stopped by Police	**12.0%**	**10.5%**

Value to Prepare for PR

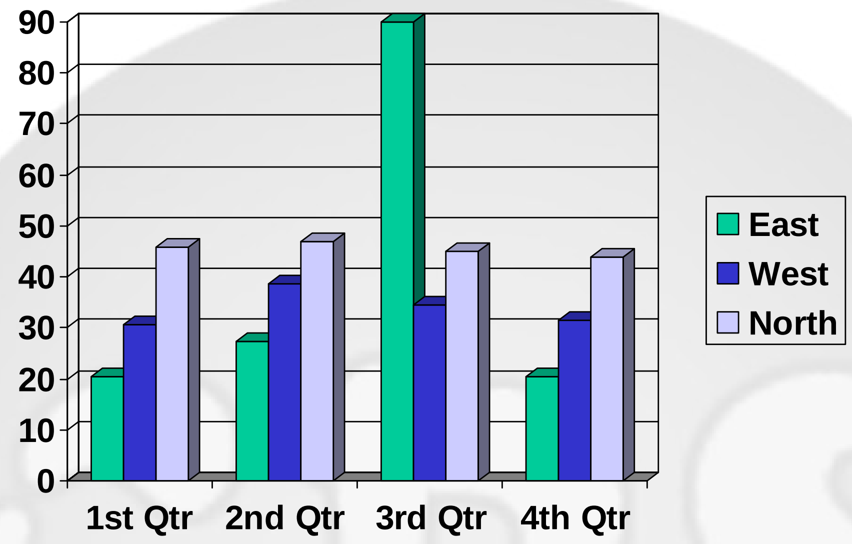
Sample statistics are shown for number of errors per page identified by the PR with preparation and without preparation in ABC. Use a 0.05 significance level to test the claim that the numbers come from populations with different means. Can you conclude that PR with prep can identify more errors?

	With Prep	Without Prep
n	40	43
$\bar{x}$	5.5	3.4
s	2.5	1.6

CVDOT: Required Dada Validation

- 1. Center: A representative or average value that indicates where the middle of the data set is located.
- 2. Variation: A measure of the amount that the values vary among themselves.
- 3. Distribution: The nature or shape of the distribution of data (such as bell-shaped, uniform, or skewed).
- 4. Outliers: Sample values that lie very far away from the vast majority of other sample values.
- 5. Time: Changing characteristics of the data over time.

Note: Data Validation represents over 95% of overall work.



Example: Productivity Variation

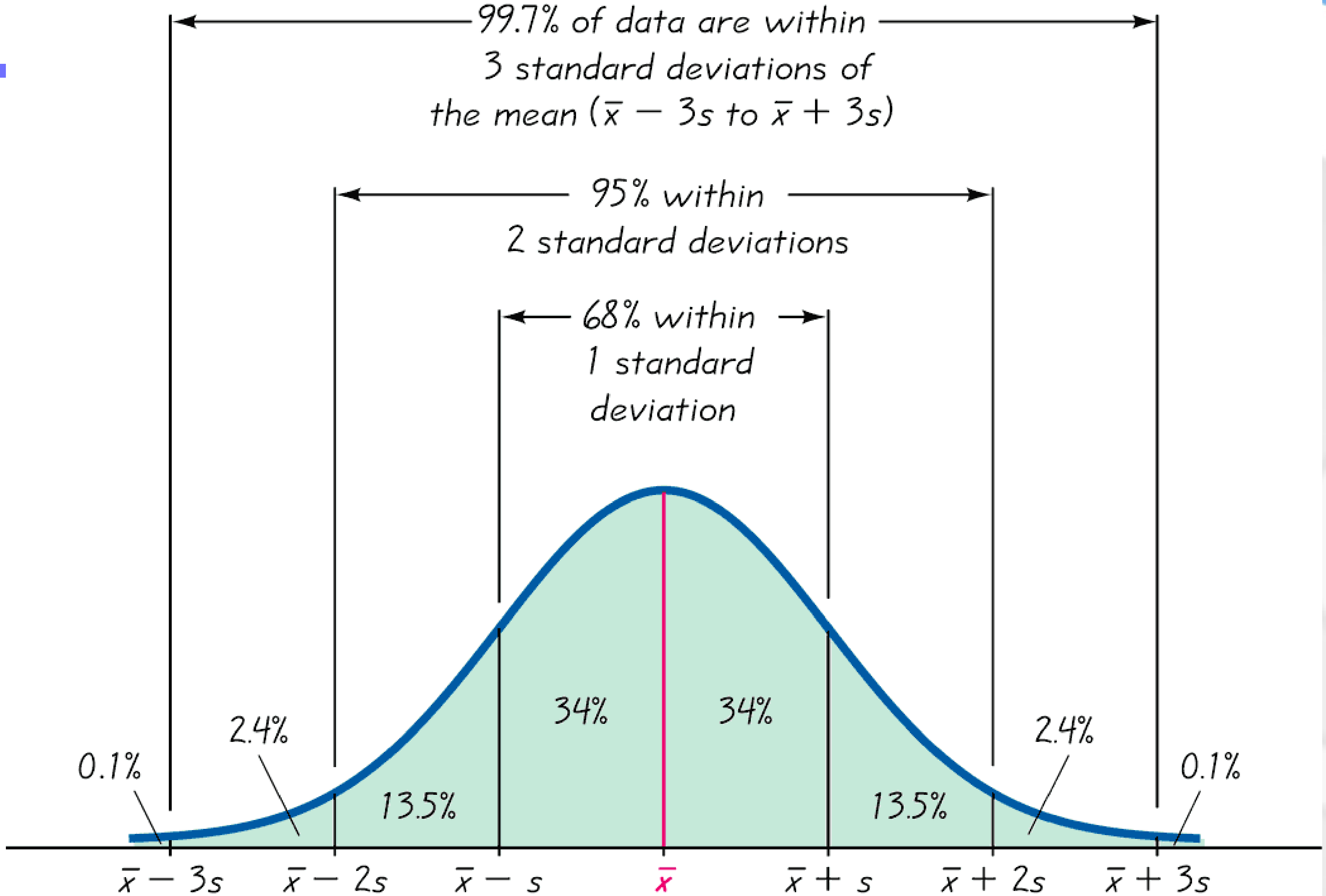
Project Team A:	600 LOC	600 LOC	600 LOC
Project Team B:	400 LOC	700 LOC	700 LOC
Project Team C:	100 LOC	300 LOC	1400 LOC

- Observation: There is no difference if only mean is considered.
- They are very different in the amounts that individual productivity vary.
- Question: How do we measure variation?

Definition

The standard deviation of a set of sample values is a measure of variation of values about the mean.

The Empirical Rule



Definition

- ***z*** **Score** **(or standardized value)**

  the number of standard deviations that a given value *x* is above or below the mean

Example: Comparing Heights of US Presidents and NBA Basketball Players.

- Facts:
- Tallest US president in 20th Century: Lyndon Johnson – 75 in.
- Tallest NBA Basketball Player in Houston Rocket: Yao Ming – 89 in.
- Question: who is *relatively taller?*

Mean (Heights of all presidents in last century) = 71.5 in.

Standard Deviation = 2.1 in.

Mean (Heights of players in Houston Rocket) = 83 in.

Standard Deviation = 3.7 in

Lyndon Johnson: $z = 1.67$ -> LBJ's height is 1.67 SDs above the mean.

Yao Ming: $z = 1.62$ -> Yao Ming's height is 1.62 SDs above the mean.

Examples of Statistical Methods:

- Distributions
- Statistical Process Control
- Hypothesis Testing
- Confidence Intervals
- Regression
- Multinomial Experiments
- ANOVA
- … …

Misuses of Statistics

- *Figures don't lie; liars figure.*
- *Statistics can be used to support anything – especially statisticians.*

Sources of deception:

1. *Evil intent on the part of dishonest persons.*
2. *Unintentional errors on the part of people who don't know any better.*

Misuse # 1- Bad Samples

- Voluntary response sample
- (or self-selected sample)

one in which the respondents themselves decide whether to be included

- In this case, valid conclusions can be made only about the specific group of people who agree to participate.

Which poll result you trust more?

Misuse # 2- Small Samples

Conclusions should not be based on samples that are far too small.

Example: Build a process performance baseline based on a sample of only five projects

What is the acceptable sampling size?

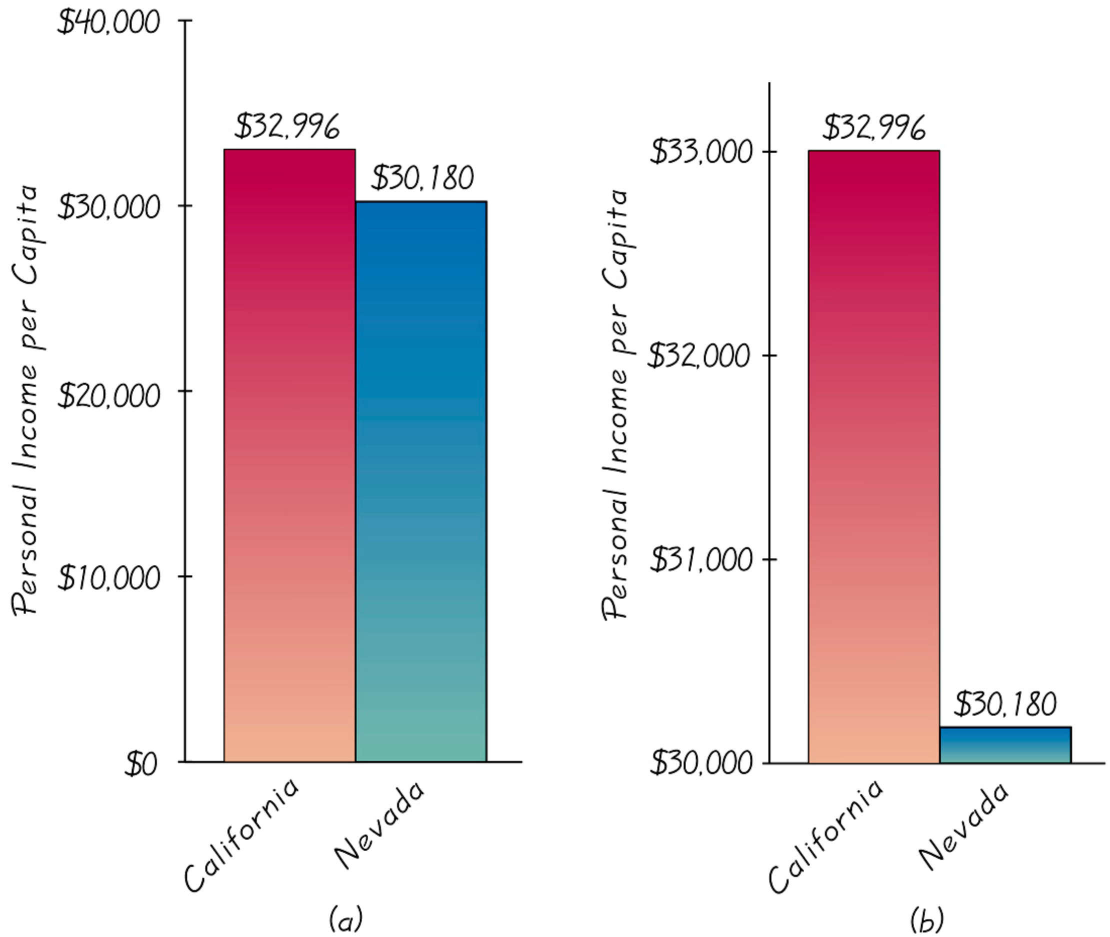
It all depends:

- population distribution

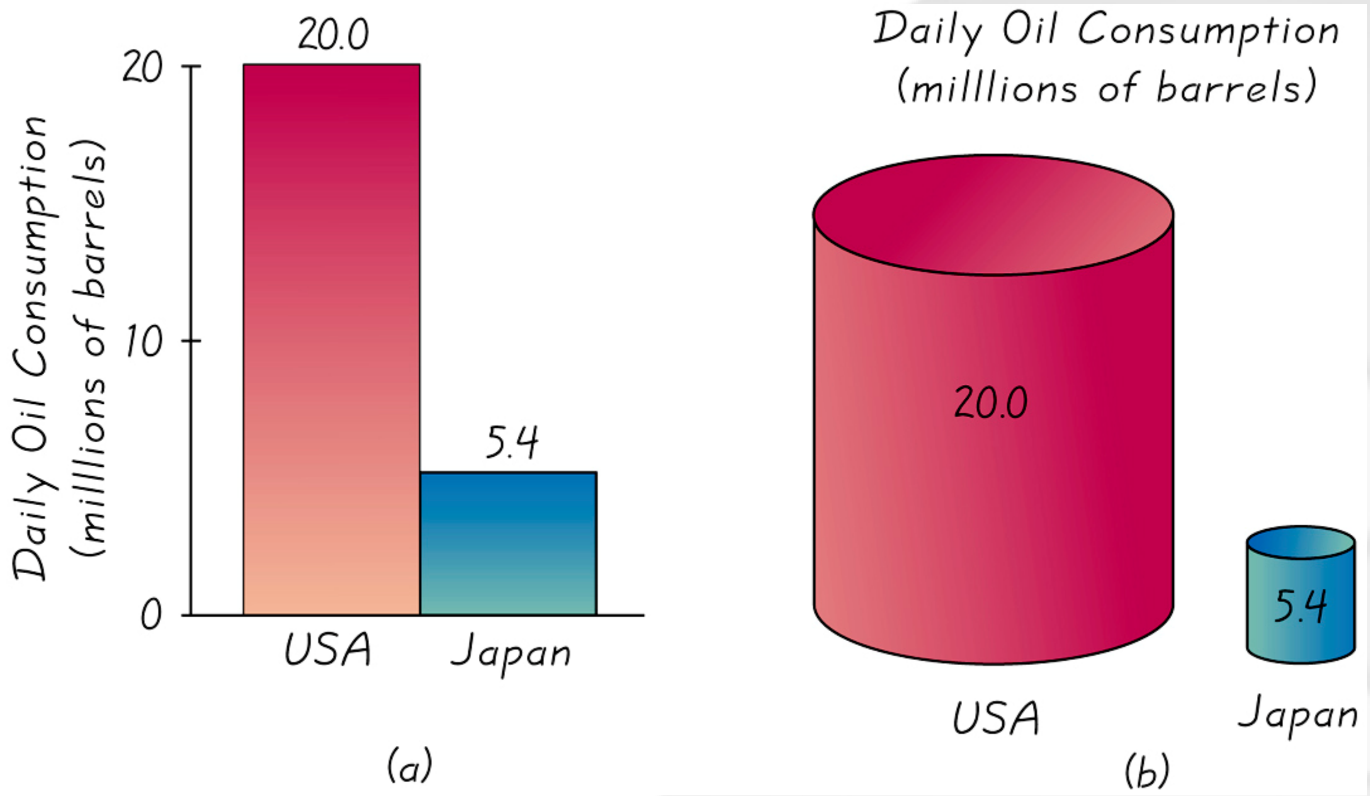
- statistical technique used

- a good dose of common sense

Misuse # 3a - Presentation Matters



Misuse # 3b - Pictographs



Statistical *Process* Control: a way to understand the variation.

- Statistical process control (SPC) **is a method for achieving quality control in "development/manufacturing processes." It is a set of methods using statistical tools such as control chart and others, to detect whether the process observed is under control. (Std definition)**

- **It is a statistical way to identify** signals **and** noises **of a process. That is, it is a way to understand the variation of the process.**

- **One of the core concept in CMMI HM**

Analyzing Process Behavior

- **Data are generally collected as basis for action.**
- **Process stability and capability.**
- **Variation exists in all data and consists of both noise (common cause variation) and signal (assignable cause variation).**

Total variation = common cause variation + assignable cause variation

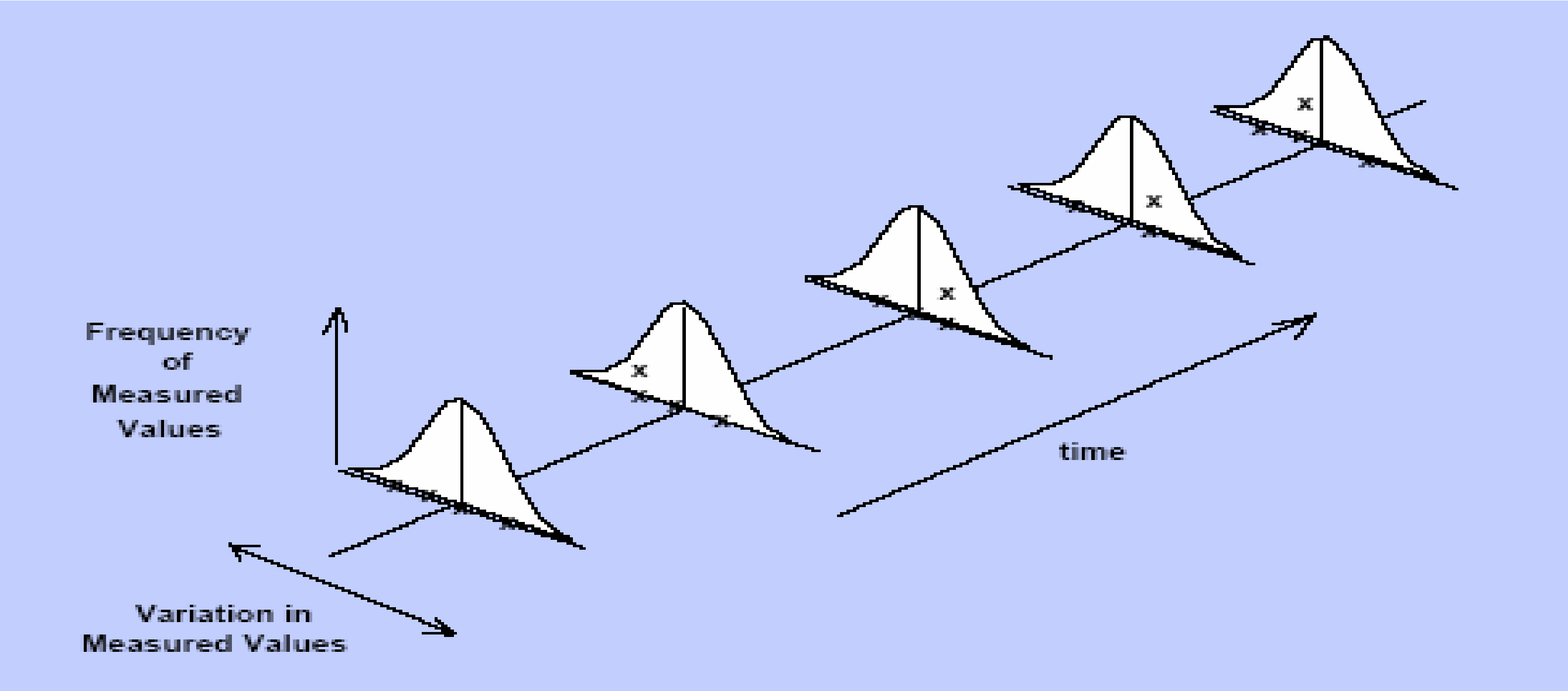
- **We wish to draw inferences that can be used to guide decisions and actions.**
- **One needs simple and effective ways to detect assignable causes surrounded by common causes.**

The only 2 mistakes a manager could make: (Deming)

- ***Mistake #1 is interfering when everything is normal – when the process exhibits statistical control. He called this*** *tampering*.
- ***Mistake #2 is a failure to intervene when a process in out of control. A process gets out of control when something external affects it. In common language, we call these*** *issues*.

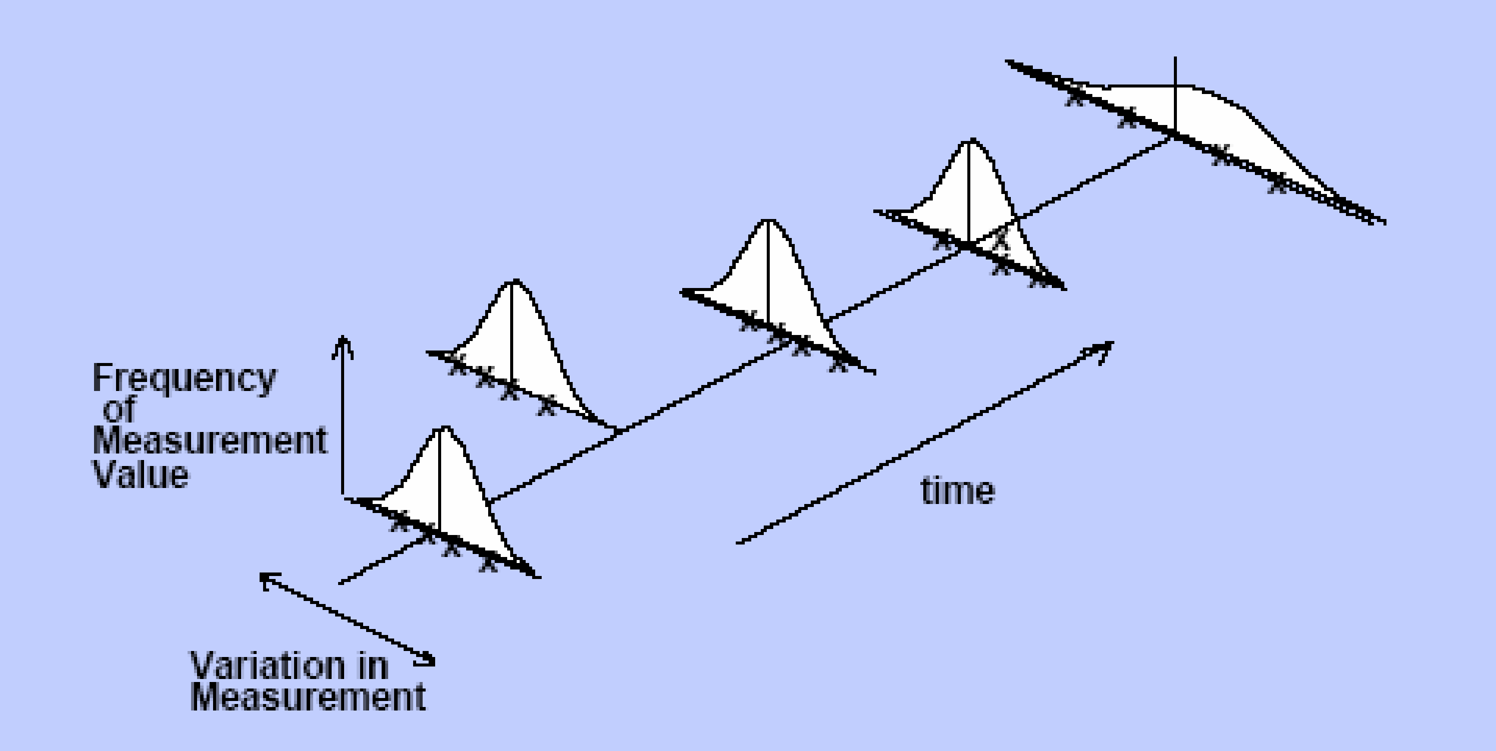
Question: how do you know if everything is normal or not?

For your review…



- Is there variation?
- Is this process under control?

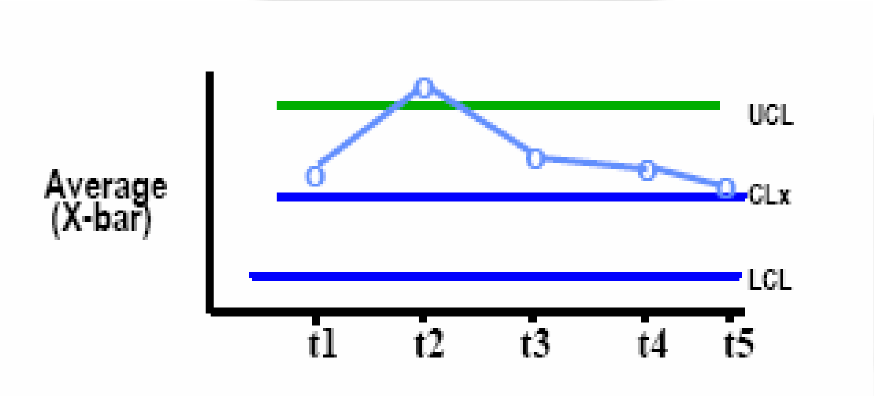
For your review…



- Is there variation?
- Is this process under control?

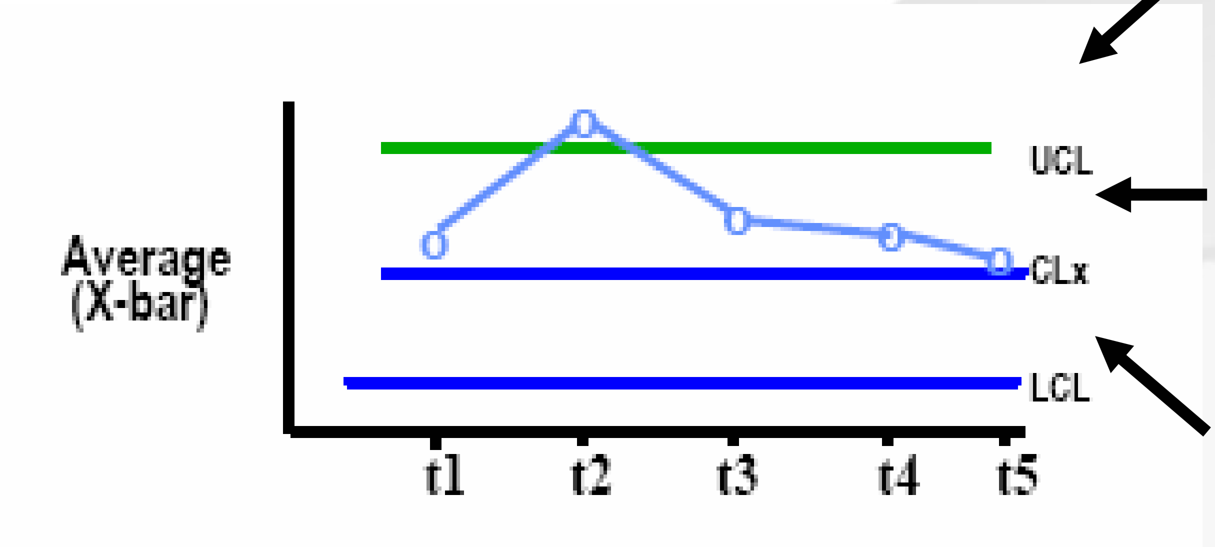
Control Charts: most commonly used technique

Control charts separate signal from noise, so that you can recognize a process change when it occurs.



Control charts identify unusual events. They point you to fixable problems (assignable causes) and to potential process improvements.

Anatomy of a Control Chart



Upper Control Limit
+ 3 sigma

Centerline

Upper Control Limit
- 3 sigma

A few things you need to know about control charts (I)

- **Check the distribution:**

  (for non normal distribution: 75% of all values lie within 2 standard deviations of the mean. at least 89% of all values lie within 3 standard deviations of the mean.)

- **A process is statistically stable (or within statistical control) if it has natural variation, with no patterns, cycles, or any unusual points.**
- **The control limits need to be maintained. Never stop control charting your process.**
- **Plan sampling**

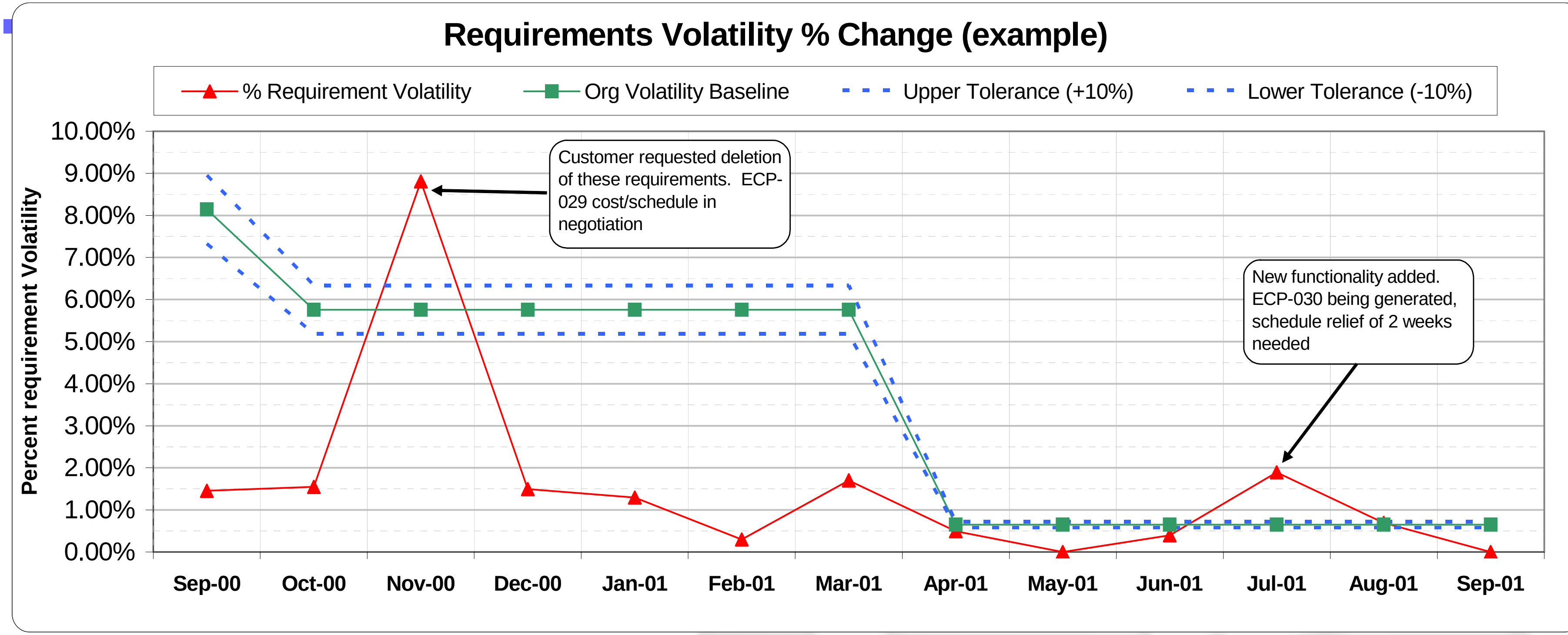
A few things you need to know about control charts (II)

- Aggregation and decomposition of process performance data: operational definition
- Control chart is only one of the many statistical techniques that support understanding of process variations. For example, confidence interval can often be used in certain situations.
- Focus on the action when signal is on: removal of the special causes.

High Maturity Is Management with a Navigation System - SEI

- **Measurement is used routinely by those who are proactive:**
  - **Are we confident we know where we are, where we are going, and our performance outcomes (quantitative understanding)?**
  - **Do we understand variation?**
- **Use measurement results to answer the questions "Will we be successful?," "Are our customer expectations and what we are capable of doing aligned?," and "What if we were to do something different?"**

Sample Requirements Volatility Report (Raytheon)



Totals	Sep-00	Oct-00	Nov-00	Dec-00	Jan-01	Feb-01	Mar-01	Apr-01	May-01	Jun-01	Jul-01	Aug-01	Sep-01
Total # of Requirements	1100	1101	1004	1004	997	999	1006	1006	1006	1002	1011	1007	1007
# of Reqts Added	7	5	0	0	3	2	7	0	0	0	14	0	0
# of Reqts Modified	1	8	0	15	0	1	10	5	0	0	0	3	0
# of Reqts Deleted	8	4	97	0	10	0	0	0	0	4	5	4	0
Total # Reqts changed	16	17	97	15	13	3	17	5	0	4	19	7	0
% Requirement Volatility	1.45%	1.55%	8.81%	1.49%	1.29%	0.30%	1.70%	0.50%	0.00%	0.40%	1.90%	0.69%	0.00%
Org Volatility Baseline	8.14%	5.76%	5.76%	5.76%	5.76%	5.76%	5.76%	0.65%	0.65%	0.65%	0.65%	0.65%	0.65%
Upper Tolerance (+10%)	8.95%	6.34%	6.34%	6.34%	6.34%	6.34%	6.34%	0.72%	0.72%	0.72%	0.72%	0.72%	0.72%
Lower Tolerance (-10%)	7.33%	5.18%	5.18%	5.18%	5.18%	5.18%	5.18%	0.59%	0.59%	0.59%	0.59%	0.59%	0.59%
Program Lifecycle Phase		CDR						System Integration					

How do you know that your std process is stable?

- Consistently enforced process
- Proper operational definition
- Large enough data sample used to compute the control limits (ranges)
- *Done a Hypothesis Testing*
- *a good understanding on the sources of variation*

Another Example of Using HT

A survey of n = 880 randomly selected peer review meetings showed that 56% (or p = 0.56) of those meetings lasted more than one hour. Find the value of the test statistic for the claim that the majority of all peer review meetings last more than 1 hour

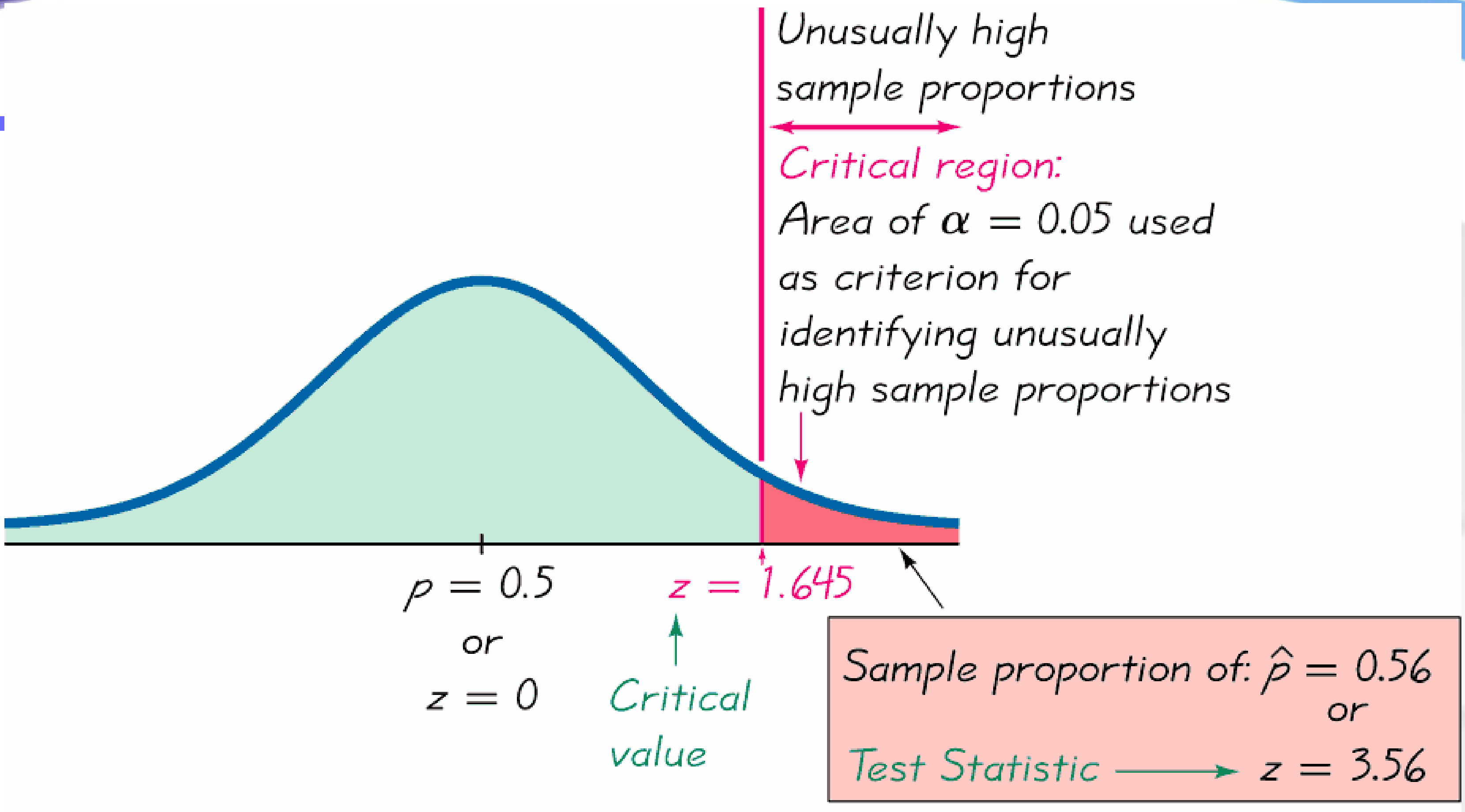
Solution: The preceding example showed that the given claim results in the following null and alternative hypotheses: H_0: $p = 0.5$ and H_1: $p > 0.5$. Because we work under the assumption that the null hypothesis is true with $p = 0.5$, we get the following test statistic:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.56 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{880}}} = 3.56$$

Interpretation: A *z* score of 3.56 is exceptionally large. It appears that in addition to being “more than half,” the sample result of 56% is **significantly** more than 50%.

See figure following.

Critical Region, Critical Value,Test Statistic



Proportion of Peer Review Meetings that Last More Than One Hour in Mykotronic.

Key Concept

Mathematical model: which is a function that “fits” or describes real-world data.

All models are wrong, some are useful.

Some Generic Models

- **Linear:** $y = a + bx$
- **Quadratic:** $y = ax^2 + bx + c$
- **Logarithmic:** $y = a + b \ln x$
- **Exponential:** $y = ab^x$
- **Power:** $y = ax^b$

Follow the recipe given in the here.

- Ingredients:
- high quality data
- one statistical analysis software package (Mini tab)
- a good dose of common sense
- - Validate your data (CVDOT)
- - Select the variables (possible sources of variations)
- - Perform preliminary analysis (using graphs, table, correlation and stepwise regression analyses)
- - Build the multi-variable model (using analysis of variance, ANOVA)
- - Check the model
- - Extract the equation

HOW AND WHERE TO USE STATISTICAL TECHNIQUES

Statistics for Describing, Exploring, and Comparing Data

- **CVDOT: How we use them?**

-**Preliminary data analysis and validation**

-**Identifying outliers**

-**DAR**

Probability Distribution

- Bionomial, Poisson, and Normal: How can we use them?

-Predicative Modeling

-Data validations before applying statistical methods

Control Chart

- XandR, XmR, P, and etc.: How to use them?

-Development and maintenance of Process Performance Baselines

-SPC: separating signal and noise

-Providing guidance for project mgt and process improvement

Estimating a Population Parameter from a Sample

- **The Central Limit Theorem, Confidence Interval, Chi-Square Distribution, and etc: How to use them?**

-**Establishing Process Performance Benchmarks**

-**Demonstrate quantitative process improvements**

Hypothesis Testing

- HT on proportion, mean, and variation of one data sample and two data samples: How to use them?
- -Verify Process Performance Baselines
- -Comparing the performance of different groups/processes/etc
- -Show the quantitative process improvement

Correlation and Regression

- Correlation, Regression, Variation and Predication Intervals, Multiple Regression, Stepwise Regression, Modeling: How to use them?
- -Key activities in developing a process performance model using your own projects data.
- -Data Validation and Analysis

Goodness-of-Fit

- Multinomial Experiments: How to use them?
- -Distribution frequency verification
- -Detection of false claims

ANOVA

ANOVA: How to use it?

- -Data validation and analysis
- -A key step in development of a process performance model
- -Comparison of multiple groups performances

Basic Steps Procedure to Develop a Process Performance Models

- **The one you should follow in developing a PPM. Pay attention to Data Types and Data Transformation.**

Exercise

Can you map the applications of these statistical techniques to SPs of OPP, QPM, CAR, OID?

A little promotion:

An entirely online Master Program in Software Engineering:

www.ecs.fullerton.edu/mse