Raytheon

Customer Success Is Our Mission

Using the Scientific Method to Achieve Level 4 and 5

Inferential Statistical Models and their Relationship to CMMI Levels 4 and 5

Jeff N. Ricketts, Ph.D.

od and inferential
ed

Raytheon

The scientific method is a body of techniques for investigating phenomena, acquiring new knowledge, or correcting and integrating previous knowledge. It is based on gathering observable, empirical and measurable evidence subject to specific principles of reasoning. The scientific method consists of the collection of data through observation and experimentation, and the formulation and testing of hypotheses. The scientific method is used to explain and predict the causes of variability in natural phenomena.

Inferential statistics or statistical induction comprises the use of statistics to make inferences concerning relationships within a population. These relationships are expressed in causal terms

- Current state
- General Measurement Issues
- Burning Platform
- Measurement in the Model
- Steps in the Scientific Method
- More issues
- Example Statistical Model
- Summary

Raytheon

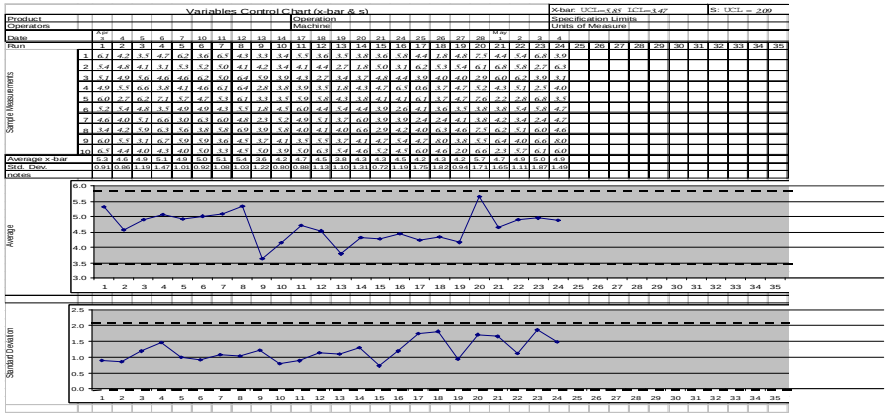
f the Practice

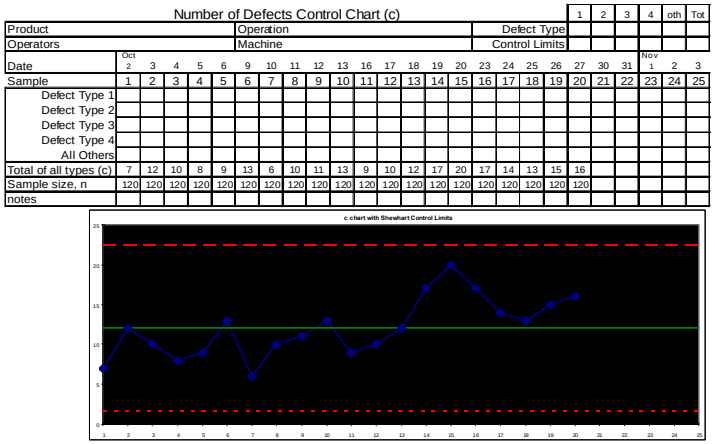
Engineering Measures:

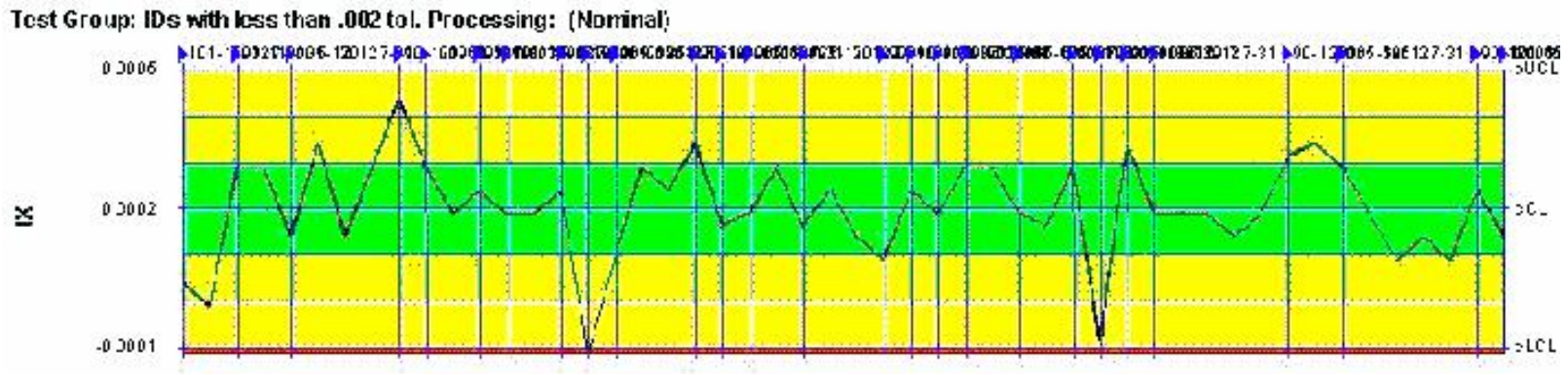
- Staffing
- CPI/SPI
- Defect Density
- Defect Containment
- Problem Report Open and Closure status
- Requirements Volatility
- Stoplight Charts

eering Review Charts

Raytheon







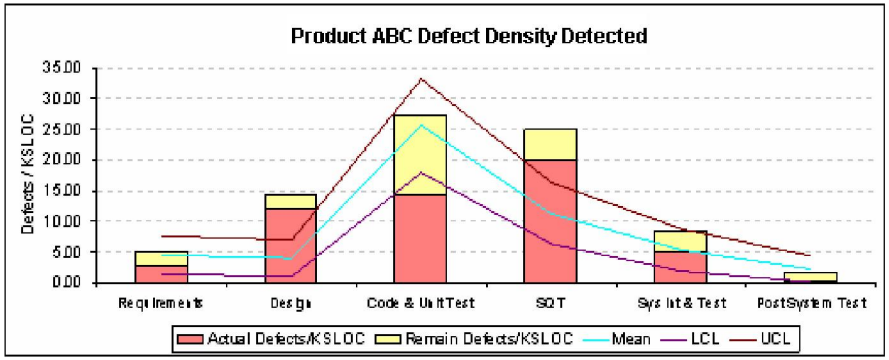
Run charts for CPI/SPI/RVOL etc

Raytheon

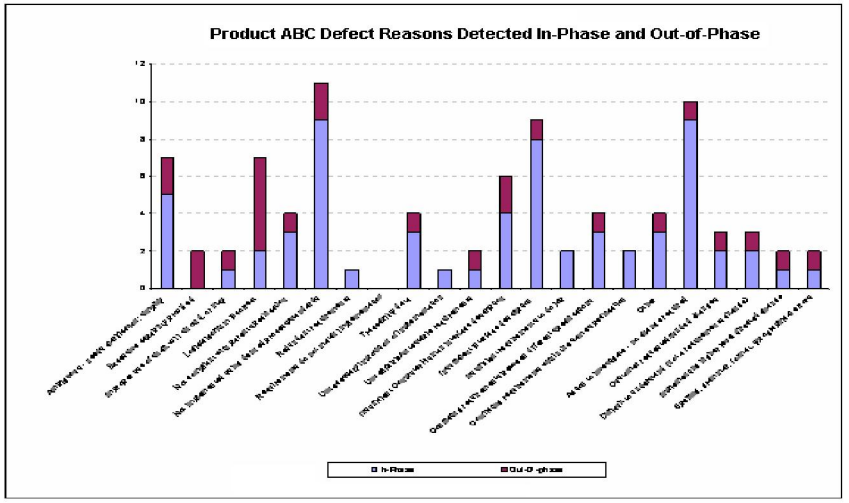
eering Review Charts

	Baseline	System Requirements and Architecture	Product Requirements and Architecture	Requirements Analysis	Preliminary Design	Detailed Design	Implementation	Integration	Product Verification and Validation	System Integration	System Acceptance Test	System Field Test	Production and Deployment	Operations and Support	Total
Baseline	0														0
System Requirements and Architecture	0	1													1
Product Requirements and Architecture	3	3	4												10
Requirements Analysis	2	2	23	3											30
Preliminary Design	3	2	2	2	2										11
Detailed Design	1	11	1	1	1	11									26
Implementation	0	0	0	0	0	0	0								0
Integration	0	0	0	0	0	0	0	3							3
Product Verification and Validation	0	0	0	0	0	0	0	0	0						0
System Integration	0	0	0	0	0	0	4	0	3	0					7
System Acceptance Test	0	0	0	0	0	0	0	0	2	0	2				4
System Field Test	0	0	0	0	0	0	0	3	0	0	0	1			4
Production and Deployment	0	0	0	0	0	0	0	0	0	0	0	0	3		3
Operations and Support	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	9	19	30	6	3	11	4	6	5	0	2	1	3	0	99

Defect Containment



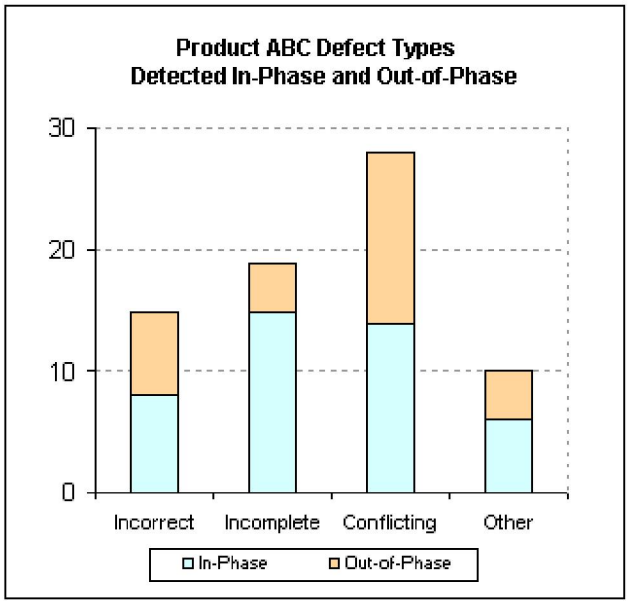
Defect Density



Defect Types

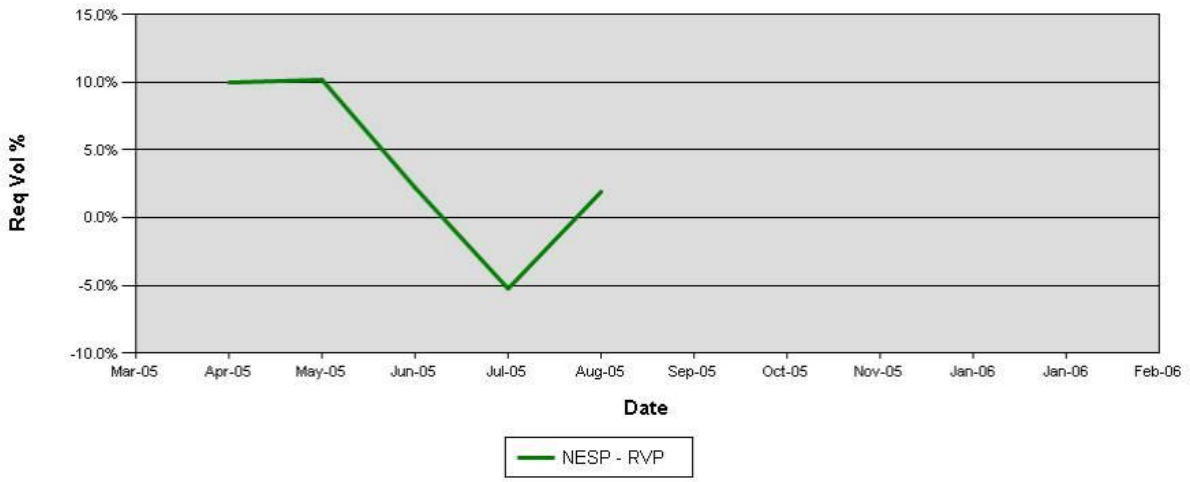
eering Review Charts

Raytheon



Defect type histogram

Requirements Volatility run chart



Raytheon

ırement Issues

The standard measures commonly in use today all have one thing in common: they are historical vs. predictive

They are all reactive vs. proactive

Some metrics have little relationship with the real questions that need to be answered

Corrective actions are usually haphazard and unverifiable as to their effectiveness

There are no standard measurement definitions

Platform

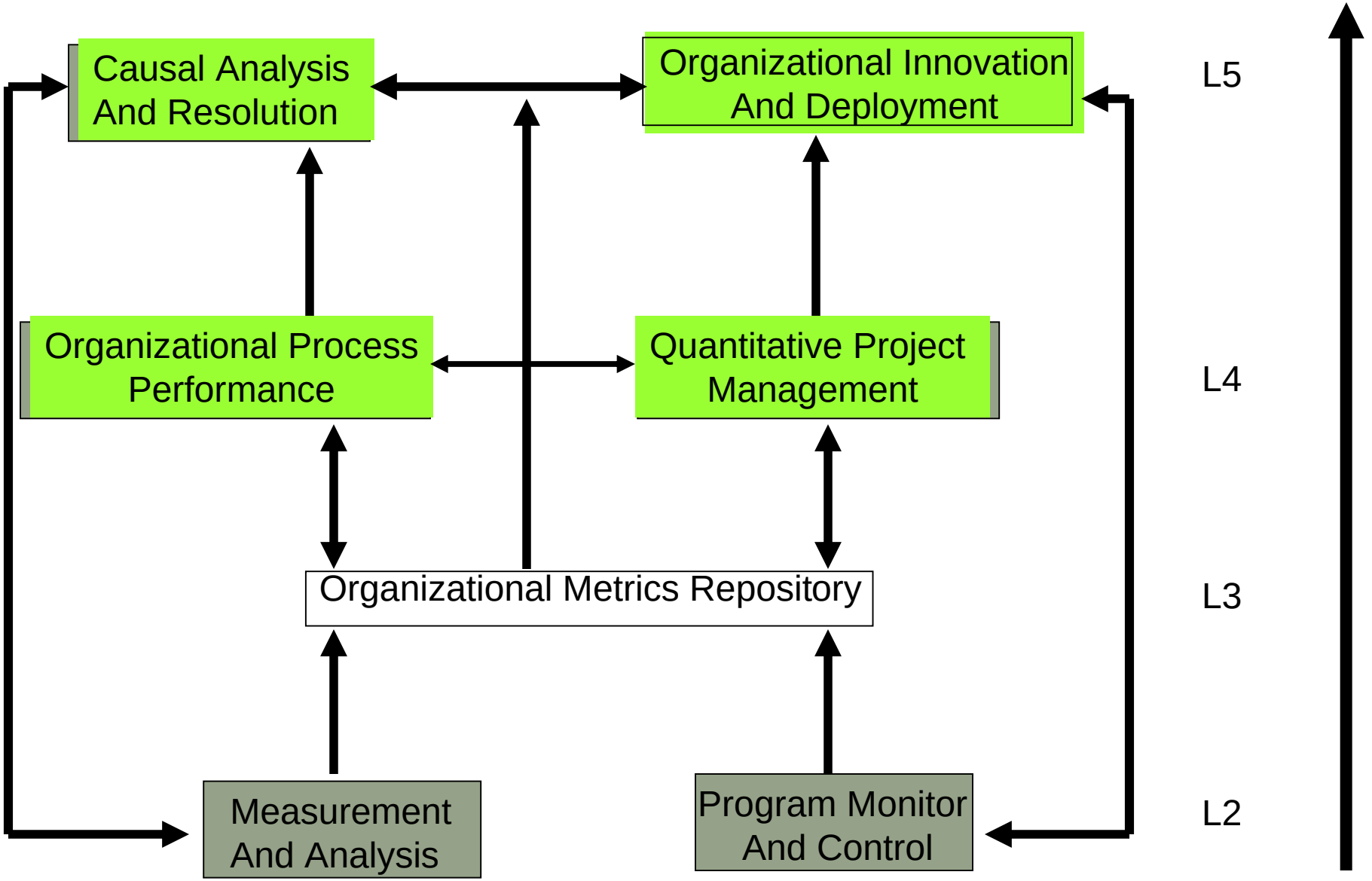
Raytheon

We need to do a better job applying scientific methods and inferential statistical models to our business to determine what causal relationships exist between the variables that we can control in order to optimize our processes and tools and reduce development costs

Level 4-5 processes can be optimized through the use of causal analysis and predictive measurement

Raytheon

ysis Spans the Model



Raytheon

nkage

QPM SG2- Statistically Manage Sub process Performance
- SP2.1 - Select Measures and Analytic Techniques
- SP2.2 - Apply Statistical Methods to Understand Variation
- SP2.3 - Monitor Performance of the Selected Sub processes
- SP2.4 - Record Statistical Management Data

OPP SP1.5 - Establish and maintain process performance models for the organizationôs set of standard processes

OID SG1 - Select Improvements
- SP1.3 - Pilot Improvements

CAR SG1 - Determine Causes of Defects
- SP1.1 - Select Defect data for Analysis
- SP1.2 - Analyze Causes

CAR SG2 - Address Causes of Defects
- SP2.1 - Implement Action Proposals
- SP2.2 - Evaluate the Effect of Changes

e Process

Raytheon

The product development process consists of many variables (tools, people, processes, inputs, outputs)

There is a lot of variation in these factors and consequences of the variation:

- stability of requirements
- makeup of peer review teams
- stability of design
- types of tools and technology used
- number of defects identified in peer reviews
- amount of hrs of training per engineer
- maturity of technology
- types of development environments used
- skill sets/mix
- programming language or design methods used

ots
nditional associations)

Raytheon

X seems to happen more often when Y is around

We always seem to do better when we use this product/method/tool/process

Do we really save time by conducting formal peer reviews for reused and ported code?

Are peer reviews even necessary on a product line?

Use cases take a long time to develop. Are they really necessary?

The key is to identify factors that appear to be associated with each other or are not reducing cost and schedule

Raytheon

Null Hypotheses

If you suspect that there is a causal relationship between two variables, the relationship is stated in the form of ñno differenceò

e.g. Systems engineers find the same number of defects during peer reviews as software engineers.

e.g. The amount of preparation time one takes for a peer review has no relationship to the number of defects identified

Raytheon

e Process

Measurements must be consistent, precise and repeatable

Measures are targeted for the type of statistics that will be generated

Nominal - categorical/dichotomous- systems engineers vs. software engineers

Ordinal - categorical -low medium high- complexity factors, lift/mod/reuse

Interval - frequency distributions- 1é n - years of experience

Ratio - frequency distributions with an absolute zero

Raytheon

ategory of data

Nominal	Difference in proportions, Chi square, Lambda, student's t test
Ordinal	Analysis of Variance, Exactness tests, Rank Order correlation, Gamma
Interval	Correlation and regression, Multiple and stepwise regression, path analysis
Ratio	Correlation and regression, multiple and stepwise regression, path analysis

mple (test) Statistic

Raytheon

Samples must be representative of the population under study

Samples must be randomly selected (can be simple, stratified, cluster, etc)

Samples cannot be the whole population

Statistics computed must be appropriate for the level of measurement

Raytheon

neses

What is the observed difference between Group A and Group B?

What is the measure of association between the independent variable (X) and the dependent variable (Y)?

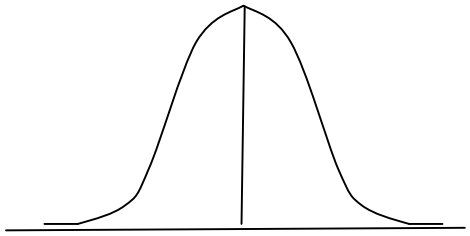
Significance levels tell you if the observed difference is statistically significant

Given no relationship between what you measured, this is the probability (.05, .01, .001) that you would observe this result in a randomly drawn sample from the population of this size?

Raytheon

Important

Study 1

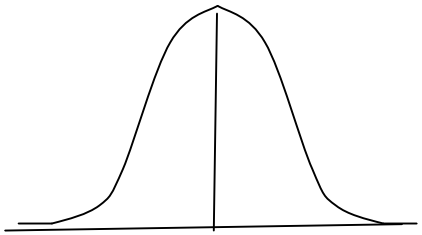


$$\bar{x}_1 - \bar{x}_2 = 0$$

$\bar{x}_1 = 14.25$ $\bar{x}_2 = 21.75$

$N_2 = 25$

Study 2



$$\bar{x}_1 - \bar{x}_2 = 0$$

$\bar{x}_1 = 30.0$ $\bar{x}_2 = 34.75$

$N_2 = 450$

Group 1 was composed only of requirements developers
Group 2 was composed of testers and requirements developers

Which observed difference between these groups is statistically significant given their sample sizes?

Raytheon

ment

What is a line of code?

What is a defect?

What is productivity?

What is rework?

What is a requirement?

g

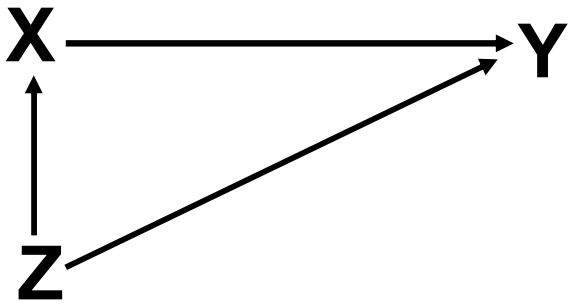
Typically not done

Typically not random

Samples need to be representative of the population that they are drawn from

Raytheon

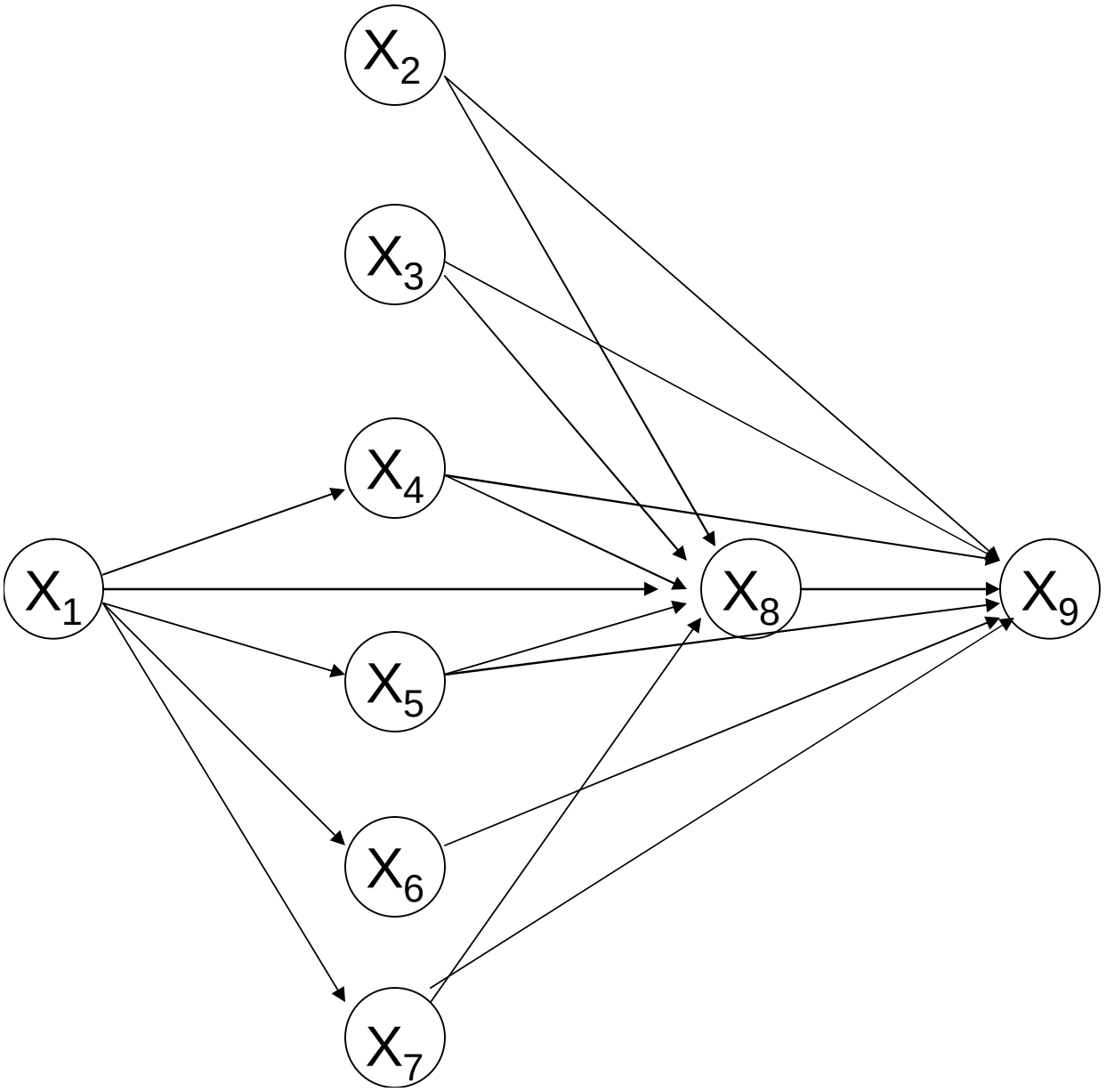
rious Relationships



Changes in X appear to be causing changes in Y when in fact Z is associated with both X and Y so when Z varies both X and Y vary

riation in Integration SPI/CPI?

Raytheon



X_1 = Training
X_2 = Technology Maturity
X_3 = Team Composition
X_4 = Hrs Spent In Peer Review
X_5 = Type of Review
X_6 = Domain
X_7 = Development Env
X_8 = Peer Review Efficiency
X_9 = IV&V CPI/SPI

- We could be doing a much better job and adding more value to our level 4-5 processes by incorporating the use of the scientific method and inferential statistical models into our measurement and analysis processes
- The data is there, but being collected inconsistently
- Random samples allow us to create probability distributions, generate sample statistics and to test null hypotheses that will aid us in being able to predict the effect of fine tuning our methods used to build our products and Dispel myths and non truths regarding the value of non-value added tasks.
- Statistically significant results typically warrant further investigation
- Correlation is not necessarily causation