PDHengineer.com

Course № F-6004

Finding the Root Cause

This document is the course text. You may review this material at your leisure before or after you purchase the course. If you have not already purchased the course, you may do so now by returning to the course overview page located at:

http://www.pdhengineer.com/pages/I-6001.htm
(Please be sure to capitalize and use dash as shown above.)

Once the course has been purchased, you can easily return to the course overview, course document and quiz from PDHengineer's My Account menu.

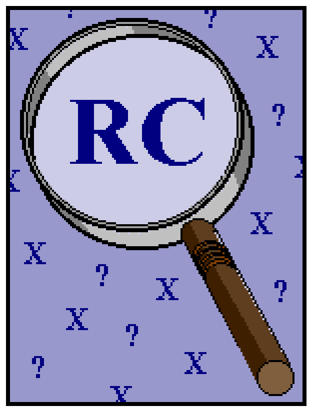
If you have any questions or concerns, remember you can contact us by using the Live Support Chat link located on any of our web pages, by email at administrator@PDHengineer.com or by telephone toll-free at 1-877-PDHengineer.

Thank you for choosing PDHengineer.com.

PDHengineer®.COM
Continuing Education for Engineers

17350 State Highway 249 | Suite 140 | Houston, TX 77064

FINDING THE ROOT CAUSE:
Statistical Solutions for Problems In Engineering Systems



Course offered via

PDHengineer.com

Authored by

Richard L. Duggan, PE (Ret.)

FINDING THE ROOT CAUSE:
Statistical Solutions for Problems In Engineering Systems

Richard L. Duggan, PE (Ret.)
10915 SW Berkshire St.
Portland OR 97225-4405
503-646-4252
duggandj@comcast.net

BS, General Engineering, University of Portland

MS, Electrical Engineering, Oregon State University

Professional Engineer (Ret.), Mechanical, State of Oregon

My career includes thirty-three years at Tektronix, Inc. working in Design Engineering, Manufacturing Engineering, and Management. Several years at Tektronix involved programs for solving quality problems in the development of color printers. Since retiring, I have taught courses for engineers in the high-tech industry to improve their efficiency in solving quality problems for new products or processes.

Prologue

Whether in electrical, industrial, or mechanical systems, problems always seem to pop up at the worst time . Engineers are often seen as resorting to "shot-gun" methods to find solutions. There are few, if any, classical tools available to help them. This material is designed specifically to address the needs of problem solvers in this situation.

TABLE OF CONTENTS

PART I - FUNDAMENTALS
Introduction

Scenario #1: A consulting firm has designed and supervised the construction of a water supply system for a large residential/commercial development. This hillside project spans over 500 feet difference in elevation. During fire hydrant flow testing, the upper and lower sections met all specifications, but a middle section hydrant failed to make minimum flow. The Chief Engineer has three days left to pass all tests before contractual penalties kick in.

Scenario #2: The Air Force had contracted with the county PUD to design and install a dedicated load switching network which would insure uninterruptible power for their new missile defense control center. Acceptance testing uncovered an under-voltage problem occurring sporadically for 100 to 800 ms periods, as the load was switched from the main supply node to any one of several back-up nodes. The PUD engineering team is under pressure because they are already two weeks behind schedule.

Scenario #3: Because of a bountiful local coal mine, the huge new pulp and paper mill elected to build a CHP steam plant. The extraction turbine would provide steam at various levels for all the internal process needs, while the 15 Mw generator would supply all its power needs, as well as co-generate power back into the local grid. After the plant was in operation a few weeks, the 450psig X 500^{o}F steam supply line developed large random pressure fluctuations. If the problem isn't corrected in a few days, damage may occur to the pulp process equipment.

*

While these scenarios are fictional, engineers face similar situations all too often during their professional lives. Very few have been trained to deal with these kinds of problems. In fact, their education and instincts typically work against them as they search for solutions. This course is intended to help engineers solve these problems reliably and efficiently.

Introduction

Course Objectives: Below are the four main areas of knowledge and skill which the student can achieve through this course:

- Learn the natural pitfalls which lead to divergent problem solving investigations
- Learn how to conduct a convergent investigation
- Learn how to design simple statistical experiments that work on small sample sizes
- Be able to solve problems in engineering systems more reliably and efficiently

Case Study Format: Many of the strategies and methods taught in this course are developed in a case study format from actual experiences of the author as a design and manufacturing engineer of computer printers. While the cases are not directly connected with the experiences of the vast majority of Professional Engineers, the technical nature of the stories should be relatable to this group. As you see how the principles of good problem solving develop in these stories, you gain insight on how they might apply in your own engineering system. No matter how different the technologies, there is a universality about engineering systems and the human minds that create and improve them. It is the author's hope that telling his story about one small piece of this universe can benefit a much larger part of the engineering community.

Natural Pitfalls

Conventional Wisdom - Many times during my career, I found myself in a large conference room in a session lasting many hours. Seated around the table would be the most creative technical minds in the organization. Lively conversations and friendly arguments would draw onward as a list slowly grew on the newsprint. It was an action item list to fix the latest problem-crisis in our area of the company. The list would eventually be prioritized and given to the engineering team assigned to solve the problem.

After a few days of trial modifications and testing, the group would reconvene to review progress. The engineering team would present their findings which invariably contained new information. This would trigger a little different track in the thought processes, and of course another round of lively conversations and friendly arguments would last late in the day. Finally, the engineering team would leave with new marching orders.

It was not uncommon for this cycle to be repeated many times until a solution was found. Often the "solution" just reduced the problem – it really didn't eliminate it. This process was not fun because it didn't work very well and it gobbled up precious time.

But that's the only way we knew to proceed. Design reviews were common during the development process and they gave us lots of creative ideas. We designed things by inventing and re-inventing. But there comes a point when the invention is good enough and the re-inventing must stop. At that point, "thinking like an inventor" is counterproductive.

Natural Pitfalls

Thinking Like an Inventor - What's wrong with "thinking like an inventor"? Two things: First, when you reinvent, the system gets changed and you lose the chance to learn what really went wrong. Second and most importantly, it can lead to a diverging problem solving investigation. That is, the investigation keeps growing bigger and bigger.

When any expert suggests changes, he/she is really formulating a theory based on expertise and creativity. When these changes are implemented and tested, the theory will be proven. This works great when it's right. When it's wrong, you are stuck at the next level with little progress and more data. Multiply that by more expert minds and more levels of "make a change and test", and you can see how divergence is created.

Insidiously Natural - Reducing the size of the team is good but it is not the answer. Even though I know better, when working alone I sometimes find myself "thinking like an inventor". It is simply a human trait. I know my system and I think I know what is causing the problem. If I just fix it, I can save a lot of time. Besides, the "scientific method", so successful for centuries, is to formulate a hypothesis and then test it for truth. Why shouldn't that work with my problem? But alas, this approach leads to disappointment too often. Not that my hypotheses are bad – they are always close. But this is not a horseshoe game.

Shot-gunning – The process described above could also be called "cut and try". If you get it right on the first "cut and try" it is "good engineering". If you don't get it right on the first "cut and try", it becomes "shot-gunning" Either way it is still "cut and try". And success depends to some degree on pure luck. If you don't want to be caught "shot-gunning", then don't do "cut and try". It's that simple.

Natural Pitfalls

Signs to Watch for - If you are working on a problem and have a sense of losing focus or getting nowhere, somebody is probably "thinking like an inventor". You are in a land of divergence. Maybe the culprit is someone on your team or your manager. But be careful before you count yourself as innocent. You, your teammates, and your manager need to be vigilant against falling into these natural pitfalls. Below is a checklist.

Natural Pitfalls
That Lead to Divergent Investigations

DO NOT:

- Theorize about causes without lots of clues
- Focus on critical elements of your system
- Make a list of possible problem causes
- Rely only on system expertise
- Try to fix the problem by modifying the system

…. In other words, **DON'T THINK LIKE AN INVENTOR!**

Basic Concepts

Thinking Like a Detective - If you are at the point where you accept the idea, "don't think like an inventor", then the question is, how should you think? The opposite of divergent is convergent, meaning the investigation continually gets smaller and smaller, converging on the solution. A model for that is how a good detective thinks. He or she avoids theories and works to collect clues. After sealing up the crime scene, the detective collects evidence to narrow down the suspect list and to eventually prove the last remaining suspect guilty.

This is how the problem solver must think. Avoid theories, preserve the existing system (and data), find evidence to narrow down the list of possible causes and to eventually identify the Root Cause and prove it. In other words, **the problem solver must commit to a Converging Process of Elimination.** The following is a fictional example of this process.

The Secret Number – The great detective Columbo was working with the FBI to break up a massive drug cartel operating across the entire U.S. According to a reliable informant, the cartel used a box at the Las Vegas post office to communicate with the big boss. The informant had been working for weeks to find the box number.

Columbo met with the informant in a park every Thursday morning. As they sat on a bench, each reading a newspaper, the informant said "I finally got the number!" Just then a shot rang out from a passing car, and the informant slumped to the ground. Columbo could see that the poor fellow had only a minute or two left. He was barely conscious, paralyzed. He couldn't speak, but Columbo noticed he could make faltering movements with his right hand.

In a stroke of genius, Columbo rasped into the informant's ear,

"Is it 5000?"

He looked at the quivering thumb moving upwards and nodded.

"Is it 7500?"

The thumb shook toward a down position.

"Is it 6250?"

Down again

"Is it 5620? …is it 5930? …is 5770? …is it 5690? …"

The thumb wobbled up or down appropriately at each question. On the fourteenth try, Columbo said,

"It's 5684!"

The hand clenched into a fist, rose up in a sign of victory, and collapsed limp and lifeless on the ground. In a little over one minute, Columbo had overcome nearly impossible odds. The rest is history. The PO box contained detailed information that enabled the FBI to nail the big boss and most of his organization. Detective Columbo had won a major battle in his long war against crime.

*

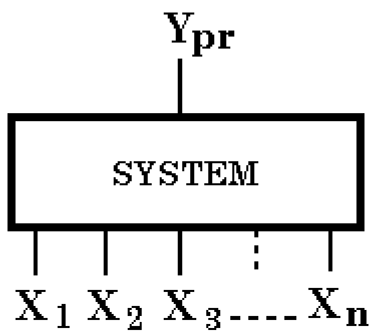
As corny as this story is, there are two important points to be made:

- There are 4 digits in a PO box number. Columbo made sure he contained it by splitting 9999 in half for his starting point.
- He never tried to find the number directly. He stuck with the process of eliminating half the remaining numbers with each question.

That is the essence of a convergent process.

Characteristics of the Root Cause

Model for Finding the Root Cause:



Y_{pr} is the variance (or range) of an output variable, the problem response of a system (or subsystem), caused by input variances X_1, X_2, X_3, …X_n. We only consider the single system output which exhibits the problem. This output must be a variable, measurable in numbers over a range with seven or more levels of discrimination. Some part of the range Y_{pr} must be acceptable for the problem to be solved without changing the system. The system has many output responses Y as shown below:

$Y_1 = f(X_1)$	$Y_{1,2} = f(X_1, X_2)$	$Y_{1,2,3} = f(X_1, X_2, X_3)$
$Y_2 = f(X_2)$	$Y_{1,3} = f(X_1, X_3)$	$Y_{n,p,q} = f(X_n, X_p, X_q)$
$Y_3 = f(X_3)$	$Y_{2,3} = f(X_2, X_3)$	
$Y_n = f(X_n)$	$Y_{n,p} = f(X_n, X_p)$	

$Y_{n,p}$ is called an interaction of two factors X_n and X_p.
Similarly, $Y_{n,p,q}$ is called an interaction of three factors X_n, X_p, and X_q.
If $Y_{n,p} = (Y_n^2 + Y_p^2)^{1/2}$, there is no interaction. (True also for three factors.) An interaction is defined as: $Y_{n,p} > (Y_n^2 + Y_p^2)^{1/2}$

Characteristics of the Root Cause

Since Y is a statistical variance,

$$Y_{pr} = [(Y_1)^2+(Y_2)^2+(Y_3)^2..+(Y_n)^2+(Y_{1,2})^2+(Y_{1,3})^2+(Y_{2,3})^2+..(Y_{n,p})^2 +(Y_{1,2,3})^2+..(Y_{n,p,q})^2]^{1/2}$$

This is a basic rule of statistical variance, the root-sum-square law.

It is important to note that Y_{pr} is the **full range** of the problem. If you measure the problem output at any one time, you will get a single value somewhere within that range.

*

The Pareto Distribution

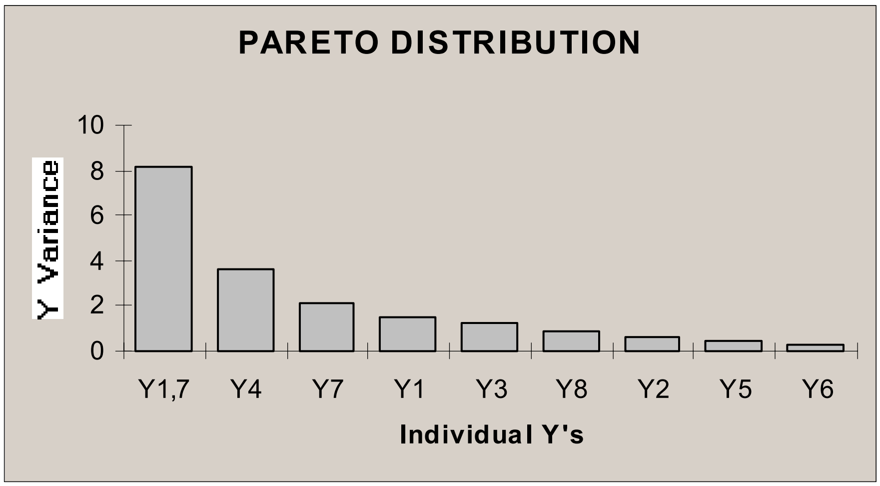
Vilfredo Pareto (1848 – 1923) was an Italian scholar who made major contributions to economic analysis, statistics, and the social sciences. He achieved international fame for his research in income distribution, which became known as Pareto's law. (20% of the people own 80% of the wealth?) The mathematical function derived from his study is used in many statistical applications and is popularly called the Pareto Distribution.

Few of us studied the Pareto Distribution in engineering school, but you hear the term often in the workplace. Our experience tells us that Pareto applies to many things. In the problem solving arena, it is an excellent descriptor for the distribution of values of :

$$Y_1, Y_2, Y_3, \ldots Y_n, Y_{1,2}, Y_{1,3}, Y_{2,3}, \ldots Y_{n,p}, Y_{1,2,3}, \ldots Y_{n,p,q}.$$

Characteristics of the Root Cause

There is no mathematical proof for this, but my experience tells me it's true. Reserve your own judgment for now, but please continue with the logical development. The following graph approximates such a distribution:



Note that the largest individual variance is $Y_{1,7}$ caused by an interaction of input variables X_1 and X_7. [$Y_{1,7} = f(X_1,X_7)$]

Find Y_{pr}:

	Y^1	Y^2
$Y_{1,7}$	8.1	65.6
Y_4	3.6	13.0
Y_7	2.1	4.4
Y_1	1.5	2.3
Y_3	1.2	1.4
Y_8	0.9	0.8
Y_2	0.6	0.4
Y_5	0.4	0.2
Y_6	0.3	0.1
		88.2

$$Y_{pr} = (\text{Sum } Y^2)^{1/2} = (88.2)^{1/2} = 9.4$$

Characteristics of the Root Cause

If the interaction $f(X_1,X_7)$ is eliminated, that is $Y_{1,7} = 0$, find Y_{pr}:

	Y^1	Y^2
$Y_{1,7}$	0	0
Y_4	3.6	13.0
Y_7	2.1	4.4
Y_1	1.5	2.3
Y_3	1.2	1.4
Y_8	0.9	0.8
Y_2	0.6	0.4
Y_5	0.4	0.2
Y_6	0.3	0.1
		22.6

$$Y_{pr} = (\text{Sum } Y^2)^{1/2} = (22.6)^{1/2} = 4.7$$

When $f(X_1,X_7)$ is functional, $Y_{pr} = 9.4$.

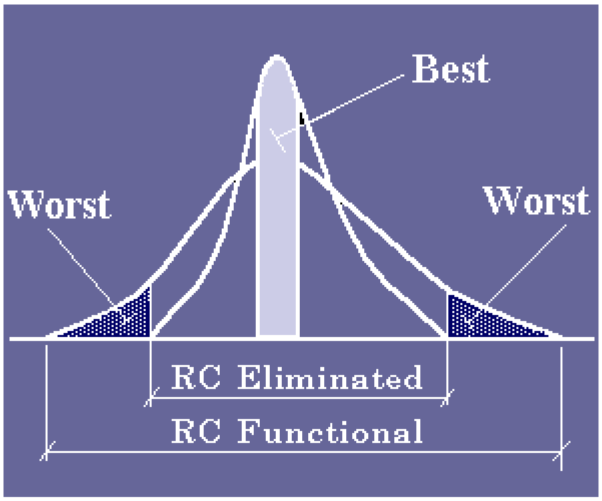
When $f(X_1,X_7)$ is eliminated, $Y_{pr} = 4.7$.

In other words, $f(X_1,X_7)$ **alone** drives the problem response to twice the value it was with all the other eight factors combined. Because it is dominant over any other single cause, $f(X_1,X_7)$, is called the **Root Cause.** (Abbreviated: **RC**)

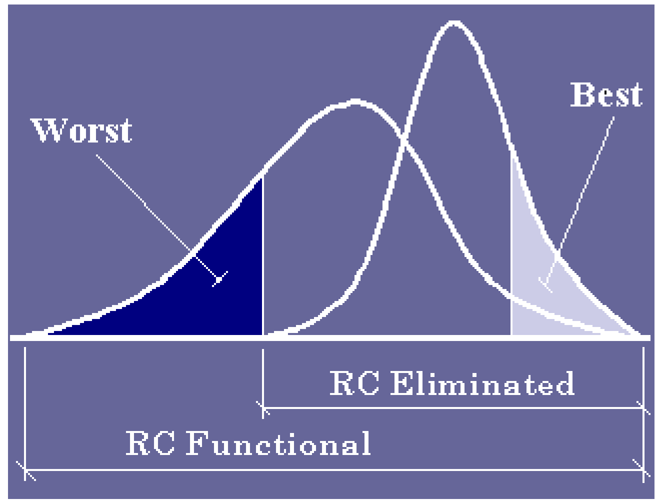
Obviously, to solve the problem you must correct the RC first. It just creates confusion to work on any of the other cause factors when the RC is driving large swings in Y_{pr}. That is why correctly identifying the RC is critical.

Characteristics of the Root Cause

These are graphical representations of Y_{pr} vs RC:



Double-Sided Distribution



Single-Sided Distribution

Best vs Worst – In the pre-production environment, we have only a few samples to represent a statistical population that would be grown by continuing to build units. Above are distributions of such populations that would exist if RC was active or inactive, respectively. The pre-production sample representing the best Y_{pr} should come from or near the areas marked BEST. Conversely, the worst Y_{pr} sample should come from or near the areas marked WORST. In the worst Y_{pr} sample, RC is functional, and in the best Y_{pr} sample, RC is eliminated.

Therefore, the "Best" and the "Worst" samples constitute a contrast in which a difference between them is RC itself. This fact is the key. Our investigations will focus on finding the full range of "Best" to "Worst", because in this contrast is all the information we need to discover (and prove) the identity of RC.

Best vs. Worst: Key to Convergent Investigations - Like Detective Columbo, we start our investigation by defining boundaries to include all possible outcomes. Then we set up partitions to begin the process of elimination. We seek a partition that shows full range "Best" to "Worst" on one side and less than full range on the other side. The side that has less than full range gets eliminated. The side that contains the full range is then set up for further partitioning. The full range "Best" to "Worst" operates just like the poor informant's quivering thumb. If the full range of variation is not exposed by a partition, it is useless. The search must continue until a partition is found that differentiates "Best" to "Worst" on one side or the other.

Natural Partitions – We naturally tend to generate these when problems pop up. Lot to lot, time to time, run to run, operator to operator, machine to machine: these are just a few of the possibilities. When something goes wrong, you'll usually hear someone ask "What changed?" The answer is likely a basis for a natural partition.

Start every investigation by gathering all the Y_{pr} data available. Spend some time studying it. Often, natural partitions are obvious. The "Hot Print Case", covered later, has a good example to demonstrate this. If not, organize it to search for suspect patterns. An example of this is a "measles chart" showing the location of defects on a circuit board. (Except in our case we'd want it to show variables (i.e. where the high end and low end Y_{pr}'s were located.)

The Extended Feeder Case – We had a beta site customer testing ten of our newest model printers hot off the pilot production line. The customer was running them 24/7 to print marketing brochures which were urgently needed. One morning, we got a frantic call informing us that our printers were having numerous paper jams, and the time it took to clear the jams was far too burdensome. Would we PLEASE send someone to fix the printers?

Convergent Investigations

Actually, getting a chance to analyze printers while failing on site is a rare opportunity. The project engineer and myself were on a plane that evening. The next morning, a harried fulfillment manager escorted us to the room where the printers were humming along, except for one that had a paper jam. The project engineer immediately removed the cover of the machine and probed inside with skilled fingers and experienced eyes.

Meanwhile, I asked the fulfillment manager if logs of performance problems were kept. He nodded and pointed at a binder over on the desk. Every paper jam was recorded by machine number, time of day, and name of the recording person. As I scanned down the list it became apparent that 5 out of the 10 machines had no jams at all!

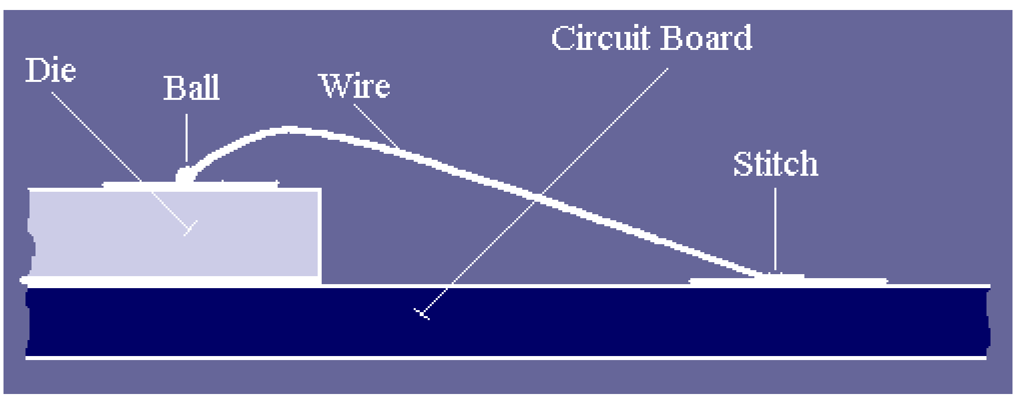
Five printers were on a table against the east wall of the room. I checked them out and found they were all the good ones. "Wow", I thought, "what are the odds of this happening?" As I turned toward the west wall where the project engineer was working, something caught my eye. These printers were different! The west wall printers had extended feeders (larger size for greater volume) while the east wall printers had standard feeders.

It didn't take the project engineer long to pinpoint the difference in the two feeder paths which was causing the jams. We called home to have five more standard feeders shipped to the fulfillment manager and departed, one hour after our arrival. In a few days, the extended feeder was modified, tested, and the case was officially closed.

Not all natural partitions are as easy as the one in this case. However, they are usually there. It just takes a little time and work to find them.

Experimental Partitions – In the unlikely event that no natural partitions are found or in the more likely event that we are well into our convergent investigation and exhausted all the natural partitions, experimental partitions are required. These are made on the basis of technical knowledge and the potential for eliminating the largest areas of the problem, thus providing the greatest convergence. The case studies to follow will gradually present experimental methods to create effective partitions.

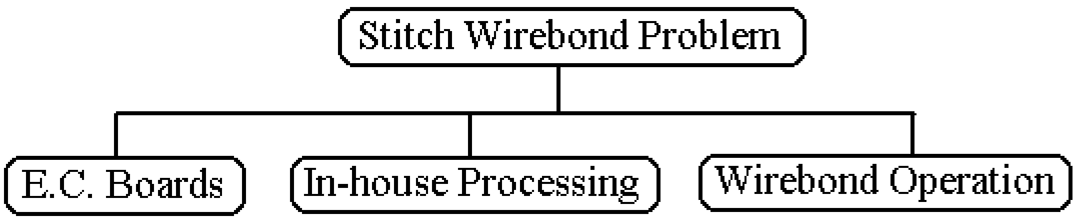
The Stitch Wirebond Case – A contract circuit manufacturer was having a problem with a new process. Too many wirebonds were failing during stress testing. They had invested in equipment to connect large processor and memory chips to circuit boards by thermosonic bonding of wires. The so-called "ball" end of the wire is bonded to a metalized pad on the die and the "stitch" end to a pad on the circuit board. See the sketch below:



Quality of wirebonding was measured with a puller which hooked under the wire and pulled upwards. The connection broke at the stitch 99% of the time. There was a force transducer in the puller that measured the force applied, ranging from 0g to 10g.

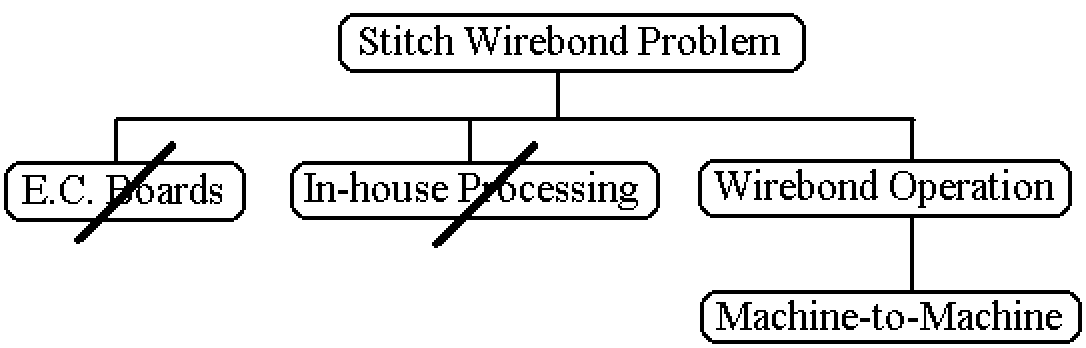
Because the ball end bonds were so much better than the stitch bonds, people were worried about surface contamination on the etched circuit boards. Another concern was contamination from in-house processing, in which components were attached to the circuit board with thermally cured epoxy. Finally, there was the wirebonding operation itself but not many people were concerned about that.

This led to the first set of natural partitions. It is shown below in a tool called an Investigation Tracking Chart (ITC). It looks like a Troubleshooting Chart common in service manuals. The difference is that you fill it out as you go. Its purpose is to keep track of your process of elimination.



Because it was handy, it was decided to study puller data from the Wirebond Operation first. They searched for patterns of "Best" (7 to 10g) and "Worst" (0 to 3g) across natural partitions of operator-to-operator, machine-to-machine, shift-to-shift, and week-to-week. Soon it was apparent that one machine always produced 7-10g puller results. The other machine, although it also produced good results, was clearly the one producing the 0-3g readings.

At that point, the lead problem solver made these changes to the ITC:



(This is a fictional case, but what follows is a very real occurrence.)

At that point, people who had been working on the problem for weeks (especially the manager) were very upset. “You didn’t look at the circuit boards! It’s obvious that contamination is part of the problem! It could be interacting with something in the machine. You can’t just ignore it!” They complained and pleaded. Many continued to work on contamination anyway.

But the problem solver was undaunted. She knew the natural pitfalls that problem solvers can fall into. The “quivering thumb” of full range “Best” to “Worst” had pointed toward Wirebond Operation, and that meant Circuit Boards and In-house Processing were eliminated. She designed experimental partitions exploring the differences between the good and the bad machine. With simple experiments which we will examine in the next case study, she soon found the Root Cause, and subsequently wirebond defects dropped dramatically!

Later on, contamination on the circuit boards from In-house Processing was shown to be a minor cause for wirebond defects. After fixing that, the stitch wirebond problem disappeared completely.

This good result only happened because the problem solvers focused on following the full range “Best” to “Worst”. Quite a different result would have ensued by following conventional wisdom. **Trying to deal with contamination would have been impossible while the machine-to-machine Root Cause was active.**

Remember the natural pitfalls. Do not focus on possible causes. Search only for Y_{pr} contrasts that show full range “Best” to “Worst”. Those are the keys to convergent investigations.

Designing Simple Experiments

General – There are two categories of experiments: partitioning and proof. Partitioning experiments continue the search for "Best" and "Worst" when natural partitions have been exhausted. They have a statistically sound basis but usually do not provide a statistical confidence number. On the other hand, proof experiments are designed specifically to provide statistical confidence that the RC has been correctly identified.

To design either kind of experiment, you don't need advanced statistical math. With a little basic algebra and common sense, you can design simple experiments for your convergent investigation. The following few pages contain all the math you will need.

Experiments are commonly used to prove a hypothesis true or false. In fact, our proof experiments do just that: proving that our claim to know the identity of the RC is either true or false.

But be careful when conducting your convergent investigation! Isn't postulating a hypothesis and proving it true or false just a fancy version of "cut and try"? Remember your detective hats when you design partitioning experiments. The purpose is to find out which sides of partitions contain the full range "Best" to "Worst" Y_{pr} variation, and which sides don't, so you can eliminate those that don't.

The Soul of an Experiment – There is a unique aspect of experiments, which I call casino rules. Casino gaming works this way: the rules are stated clearly beforehand, and you bet on an outcome before the game starts. You cannot bet after it starts (unless it's an additional bet). The gaming commission assures that pure chance is not compromised. At the end of the game, either you win or you lose. Experiments are similar. You design an experiment with clearly stated rules and procedures. You define the consequences of all outcomes before the experiment starts. You insure that pure chance is not compromised by randomizing (to be explained later). At the end of the experiment, you accept the consequences as you pre-defined them. If your experiments are as rigorous as casino gaming, your statistics will always be trustworthy.

Designing Simple Experiments

Math for Simple Experiments – Probability is the ratio of a particular outcome over the total number of outcomes possible, given any outcome has an equal possibility of occurrence (randomness). Thus, the probability of a single coin toss coming up "heads" is ½. What if the coin is tossed five times? The probability of all five coming up "heads" is:

$$p = (1/2)(1/2)(1/2)(1/2)(1/2) = 1/32 = 0.03125$$

Confidence is the complement of probability, or:

$$c = 1\text{-}p = 1\text{-}(0.03125) = 0.96875$$

Thus, your confidence that all five tosses will **not** be "heads" is 96.9%

Combinations - A combination is any unique way a set of "r" things can mix in a total set of "n" things ("n" includes "r"). You can differentiate the "r" things from the "(n-r)" things, but you can't differentiate among "r" things, nor among "(n-r)" things. The total number of combinations possible for a set of "r" things mixed with another set of "(n-r)" things is:

$${}_nC_r = n!/r!(n-r)!$$

What is the number of combinations possible for two black marbles to be arranged in a line with three white marbles?
In this case, $r = 2$ and $n = 5$.

$${}_5C_2 = (5)(4)(3)(2)(1)/[(2)(1)][(3)(2)(1)] = 10$$

Combinations in Proof Experiments - In proof experiments, there are two types of samples: Type E (RC eliminated) and Type F (RC functional). We use the formula for combinations to calculate how many ways are possible for E and F to be ordered. In the next case study, the details of designing proof experiments will be presented.

Designing Simple Experiments

Probability in Repeated Trials – Back to the coin example on the previous page: the probability of a coin coming up "heads" in five successive tosses is 0. 03125. But that is for the first try only. What if you try it more than once? The following is a calculation for the probability of getting five successive "heads" once out of four attempts (that is four trials of five tosses each).

The probability getting five "heads" in a row is p, and the probability for not getting five in a row is q. In this case, p will happen once, and q will happen three times. Moreover, these outcomes could occur in different combinations.

The general equation that takes this all in account is:

$$_nP_r = {}_nC_r p^r q^{n-r} = [n!/r!(n-r)!]p^r q^{n-r}$$

where $_nP_r$ is the probability an event will happen exactly r times in n trials. In our example, r = 1 and n = 4. Thus:

$$_4P_1 = [(4)(3)(2)(1)/(1)(3)(2)(1)](0.03125)^1 (0.96875)^3 = 0.1136$$

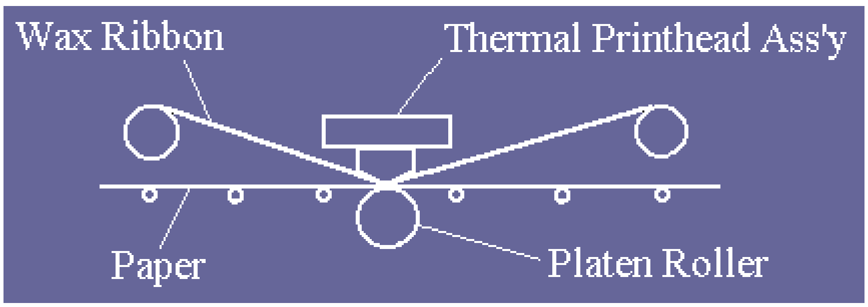
and conversely, the confidence of **not** getting 5 heads one time in four trials is 88.6%

Rules for Simple Experiments - In an experiment, you are setting decision criteria. If a certain outcome occurs, you will go one way; if not, you will go the other.

- Run the experiment like a casino game. You must make your "bet" before the game starts. When it's over, you either win or lose.
- All arbitrary events in an experimental procedure must be randomized (e.g. order of testing, etc.).
- Know the math of possible outcomes. In other words, get your simple statistics right.
- Decision criteria must be based on Y_{pr} outcomes only.

PART II - THE "HOT PRINT" CASE STUDY
Case Background

In the early days of desktop computer printers, wax transfer technology produced the best color. It had developed out of the mature thermal printing technology. After overcoming the challenge of precise registration for the three pass process, wax transfer took this advantage of beautiful bright color rendition into the printer wars of cost, speed and resolution. And it was successful for many years. Below is a sketch of the basic elements of a wax transfer printer:



I happened to be working on a new wax transfer printer that was cheaper, faster, and had higher resolution than the current competition. Our development project had passed the Design Completion milestone, and we built twenty prototypes from quasi-permanent tooling. Since the new printer used a new low-cost wax ribbon design, five of the prototype printers were sent to the ribbon vendor for testing. In this particular case of concurrent engineering, the vendor was well ahead of us, with $500K of early production ribbon in the warehouse.

Another five prototypes were sent to our reliability lab to print marketing prints 24/7 for the rapidly approaching product announcement. Within a day, the reliability engineer was at my desk with a big problem. The machines were set up to print ten copies of photo quality image and then stop for an attendant to reload another image. Blotched areas were sporadically appearing in the latter images of the ten copy run. Nothing like this happened in earlier engineering tests. I was dumbfounded!

Case Background

The reliability engineer told me not to worry because the ribbon was the problem. He had been using both the older Lot 2000 and the newer, low-cost Lot 3000 ribbon. The blotching problem only occurred in machines using Lot 3000. Soon, my boss and the VP of Procurement were on a conference call with the vendor to announce the bad news.

That conference call turned ugly fast. The vendor had noticed the same blotching problem in their lab, but their data indicated that only two specific printers did it. They believed that the problem occurred because of heat build-up during consecutive print runs. The printhead needed improved heat dissipation, they said. Irate, our VP of Procurement hung up.

Several stalemated conference calls went by. Of course, the vendor was worried about the $500K of lot 3000 ribbon in the warehouse. But he also had a good argument with valid data to back it up. On our side, redesigning the heat dissipation would set us back three months. Marketing had already pre-announced the product discreetly . We couldn't afford that kind of delay.

My boss and I decided the best way to resolve the situation was to invite the vendor to join us in a problem solving investigation. Within a week, we had a team from the vendor in our reliability lab along with their five printers. They were not familiar with our problem solving techniques, but this did not turn out to be an obstacle. The problem solvers compromised more easily than their higher level managers.

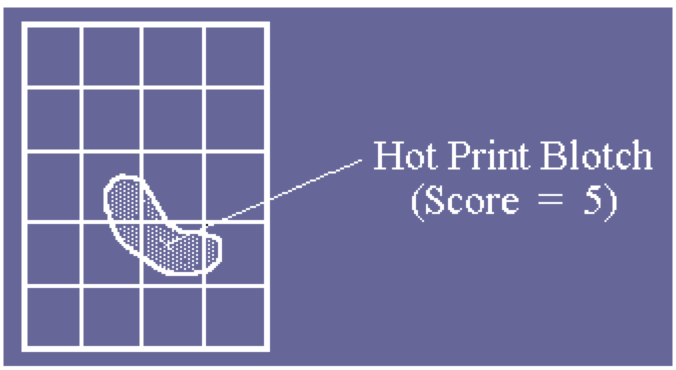
We had a total of ten printers in the lab, their five plus our five. We also had plenty of both types of ribbon, Lot 2000 and Lot 3000. (To have this many samples available is unusual in product development work. It felt luxurious!) Normally, we would spend more time looking for natural partitions in the existing data, but the vendor's team wanted to do a factorial type experiment. That was OK with us – we could find partitions that way.

Prerequisites – Before the investigation starts, three things need to be verified:

- The problem response (Y_{pr}) must be expressed as a numerical variable with seven or more levels of discrimination.
- The full range of Y_{pr} (or a reasonable estimate) must be defined.
- The system used for measuring the Y_{pr} must be repeatable, contributing less than 5% added variability to Y_{pr} measurements.

In the "Hot Print Case", many problem solving team members would have looked at a print and counted it as either good or bad, because even a small blotch was unacceptable. This is what statisticians call an attribute. But attribute statistics don't work well with small sample sizes. So the first step was to figure out a way to express blotching as a variable.

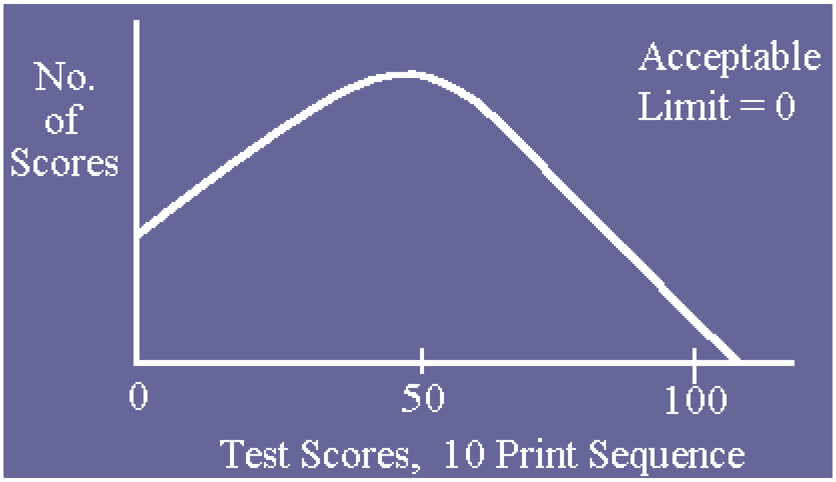
Attribute to Variable Conversion - There were several ways to do it, but we settled on an overlay grid to score a print by the size of the blotch, as shown in the sketch below:



Since there are twenty squares in the overlay, it provides twenty-one levels of discrimination (0 through 20). This is well above the minimum requirement of seven.

As you can see, all it takes is a little creativity to express Y_{pr} as a variable. Although you don't want to "think like an inventor" in conducting the main investigation, this is just one of many little opportunities you will have to vent those creative urges.

The Second Prerequisite is to define the full range of Y_{pr}. We still had several hundred reliability lab prints that were done in 10 copy sequences. So we decided to make 10 copy sequences (rather than a single copy) the basis for defining Y_{pr}. We used the overlay to score the old sequences and then plotted the distribution:



Notice that now we had over 100 levels of discrimination, which **far** exceeds the minimum of seven. In general, the more levels of discrimination, the easier it is to deal with measurement error.

The Third Prerequisite is verify that Y_{pr} measurements are repeatable so that no more than 5% variability is added to Y_{pr} from measurement repeatability error. Normally, people think about things like resolution, accuracy, and trace-ability to NBS when dealing with measurements. Since we're dealing only with patterns and comparisons, we don't need those things. We need repeatability.

Many team members thought that the overlay measurement system didn't need to be checked. But experience tells us otherwise. Always check it. The measurement system to be checked includes everything: production samples, measuring tools, operators, fixtures, procedures, etc.

Deriving the Quad-Five Repeatability Test - This is an easy repeatability test. It was derived as follows from these specs:

- 5% max. chance that
- 5% max of all measurements would add
- 5% variability in total Y_{pr} due to repeatability error, with
- 5 samples, each measured twice

The first step in designing the test was to apply the root-sum-square law to the third bullet above:

Let r = measurement repeatability limit as a % of Y_{pr}.
Then,

$$(Y_{pr}^2 + r^2 Y_{pr}^2)^{1/2} = 1.025\, Y_{pr}$$
$$Y_{pr}\,(1 + r^2)^{1/2} = 1.025\, Y_{pr}$$
$$r^2 = (1.025)^2 - 1$$
$$r = 0.225$$

So the repeatability limit is 22.5% of Y_{pr}.

Next was dealing with the second bullet above. In order to proceed, an assumption had to be made about the shape of the measurement distribution. We used the Normal Distribution, which is most likely true, but if not, is reasonably safe. Using the table in the back of the booklet:

sigma of (95%) = 3.92 **:** where 5% is left in tails of distribution

Finally, the first and fourth bullets. The worst allowable distribution has 3.92 sigma = 22.5% of Y_{pr}. How can that distribution be detected with a 5% chance of error (95% confidence) with only five samples? Answer - narrow the boundaries to the area which contains 5%-for-5-in-a-row probability:

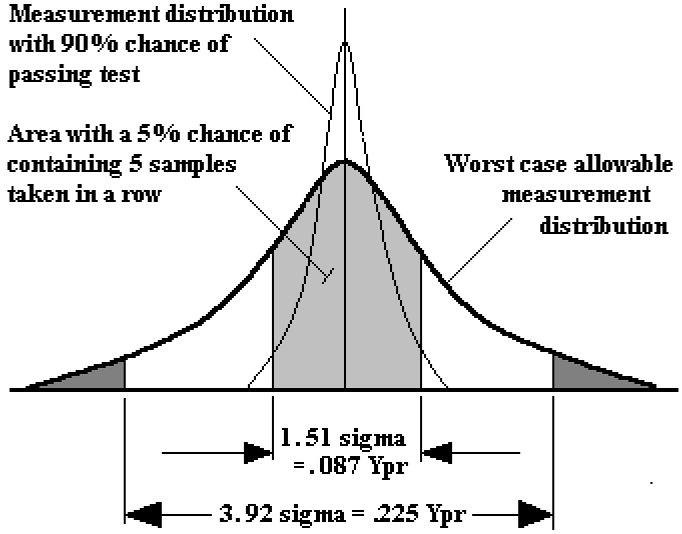
$$(p)^5 = 0.05$$
$$p = 0.5493$$

From the Normal Distribution table:

Sigma of (54.93%) = 1.51

Therefore, the test limit $= (1.51/3.92)(0.225)(Y_{pr})$
$= (0.087)(Y_{pr})$
$=$ 8.7% of Y_{pr}.

Stating it another way: when the test limit is set at 8.7% of Y_{pr}, the worst case measurement distribution has only a 5% chance of passing. The graph at the right shows the narrower measurement distribution required to pass the test 90% of the time.



$$(p)^5 = 0.90$$
$$p = .9791$$

From the Normal Distribution table:

Sigma of 97.91% = 4.62

Thus, a "typically good" measurement system will have 4.62/6 (slightly over ¾) of its six-sigma spread inside the 8.7% of Y_{pr} limit. If more test samples are taken, the difference between the "good" and the worst case will be less. However, rather than design a large sample test to pass a marginal measurement system, it is better to just improve the system.

Using Quad-Five in the Hot Print Case – From page 27, $Y_{pr} = 107$ for a 10 print sequence. Thus, the test limit was set at 8.7% of 107, or 9. The measurement system consisted of 2 fixtures to hold the print and the overlay and 2 designated operators (1 from the vendor's team and 1 from our team). We had several hundred prints on hand that were separated in to 5 stacks: severe blotching, moderately severe, moderate, moderately light, and light-to-none. Ten prints were randomly selected from each of the 5 stacks. Each group of ten would be a measurement sample. By spreading it out this way, we were verifying the measurement system over the entire range of Y_{pr}.

Since each of the 5 samples would be measured twice, there were a total of 10 measurements in the system, which were randomized relative to order, operators and fixtures, with the provision that each operator and fixture would do 5 measurements. The results are tabulated below:

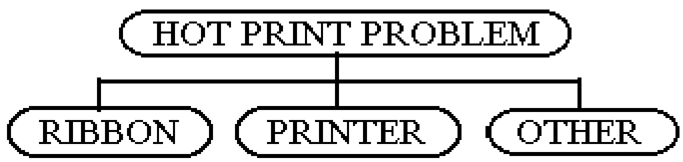
Severe:	meas.#1 = 92	meas.#2 = 98	repeat. error = 6
Mod. severe:	meas.#1 = 71	meas.#2 = 67	repeat. error = 4
Moderate:	meas.#1 = 46	meas.#2 = 51	repeat. error = 5
Mod. light:	meas.#1 = 23	meas.#2 = 27	repeat. error = 4
Light/none:	meas.#1 = 5	meas.#2 = 4	repeat. error = 1

The results were all within the limit of 9. The measurement system **passes** the Quad-Five repeatability test! If it had **not** passed, we would have a measurement system problem to solve. With repeatability error as the new Y_{pr}**,** we would begin looking for natural partitions (fixtures, operators, etc.)

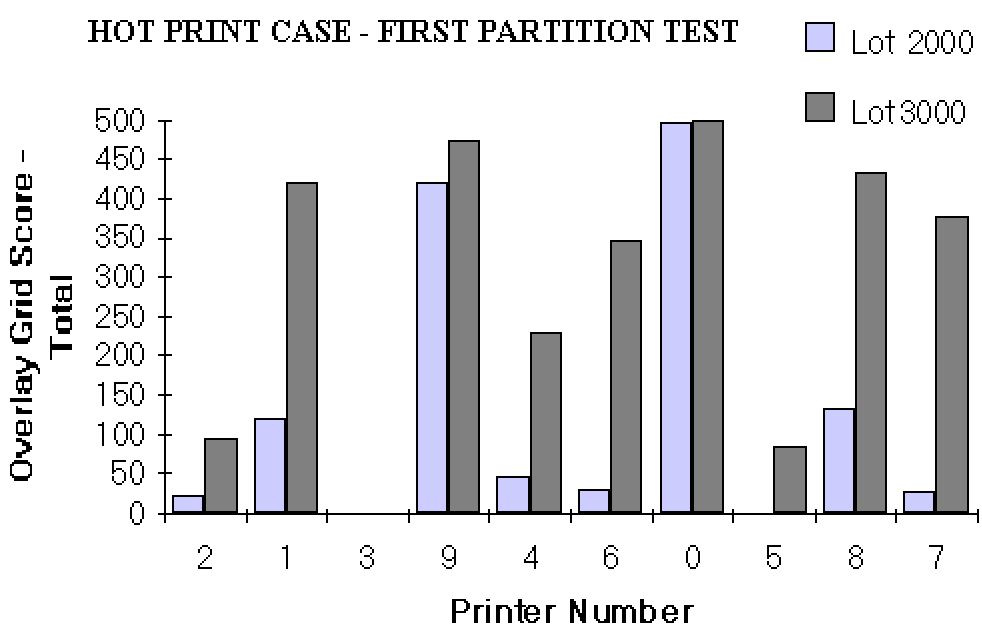
Randomization – On page 23, the second rule for simple experiments states: "all arbitrary events in an experimental procedure must be randomized". It is actually an easy rule to follow and it prevents human and non-human bias from contaminating the results. In the above test procedure, note that we randomized the selection of samples, test sequence order, fixture assignment, and operator assignment.

Randomizing is simple. The things to be randomized must have a serial number (1,2,3,etc.) or a natural sequence (top to bottom of a stack). Then toss a pen or pencil on top of a random number table (see page 55). The number nearest the pen tip (one or two digits as required) is the first random number. The serial number matching that random number will go first in order or assignment. Having predetermined the direction to proceed through the random number table (left-to-right, then down to the next row, etc. - your choice), stop at the next non-repeating random number. The serial number matching that random number will go second in order or assignment. And so on, until the order or assignment is completed.

The First Partition - As we were saying on page 25, the vendor's team wanted to do a factorial type of experiment. It was easy to agree to a ribbon vs. printer partition, obviously. We added one more category, "other", just in case it was something else such as paper, seasonal humidity, etc. This fulfills the requirement that the first partition contains the entire range of possible causes (ala Columbo). The ITC started like this:

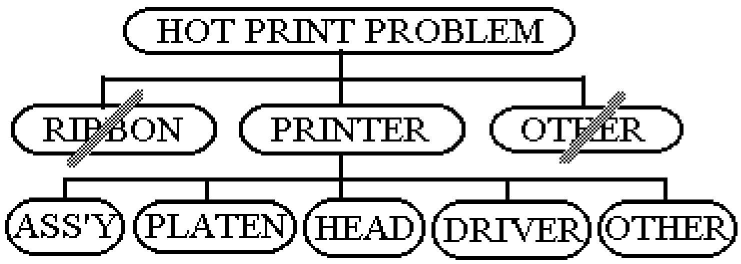


We had the five vendor printers and the five reliability lab printers for a total of ten printers. So five lot 3000 ribbon canisters and five lot 2000 ribbon canisters were selected at random from the warehouse (yes, we flipped the pen onto the random number table to get the canister numbers). There would be 10 runs for each of the 10 printers, so that each printer ran a set with each ribbon. As noted earlier, the run set was 10 successive prints. The test sequence was randomized (of course). These are the results:



The Investigation

Examining the results carefully, we see the ribbon has a **very** strong effect, but **the full range of variation (and the RC) is in the printer** (printer #3 vs. printer #0). At this point, we decided not to report the results to management, fearing another brew-haw. Instead, we'd continue the investigation with an experimental partition of the printer as follows: assembly issues, platen, print-head, driver circuit board, and other (the rest of the printer). The ITC was changed to look like this:



From now on, we would work only with printer #3, printer #0, and lot 3000 ribbon. That is the best and worst printer with the ribbon that gives the most best-to-worst separation.

Swap Test (The Second Partition)– The Swap Test is appropriate whenever two full range best-to-worst machines are available. The first step is to take Y_{pr} measurements, disassemble both machines into the experimental partition as defined, then reassemble them and re-measure Y_{pr}. The following results were recorded:

Hot Print Score, 10 successive prints, 3000 ribbon

	Printer 3	**Printer 0**
Original Assembly	0	101
After Disassembly/Reassembly	3	95

The RC is **not** an assembly issue as far as removing the platen, print-head, or diver circuit board are concerned. These units can be disassembled and reassembled with consistent Y_{pr}'s.

The Investigation

The second step is to disassemble and swap the platens between Best and Worst and re-measure Y_{pr}'s. Continuing the test record:

Hot Print Score, 10 successive prints, 3000 ribbon

	Printer 3	**Printer 0**
Original Assembly	0	101
After Disassembly/Reassembly	3	95
After Swapping Platens	2	103

Third step is swapping print-heads:

Hot Print Score, 10 successive prints, 3000 ribbon

	Printer 3	**Printer 0**
Original Assembly	0	101
After Disassembly/Reassembly	3	95
After Swapping Platens	2	103
After Swapping Print-heads	94	5

After swapping print-heads, the Best printer became the Worst, and vice-versa. Obviously, **the RC is contained in the print-head**. We did a final verification run by returning the print-heads back to their original printers:

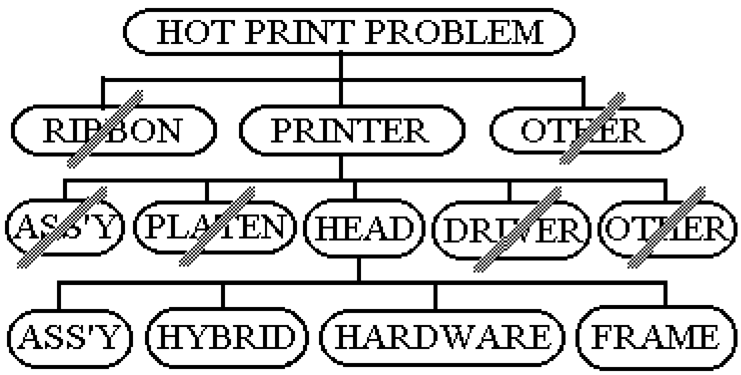
Hot Print Score, 10 successive prints, 3000 ribbon

	Printer 3	**Printer 0**
Original Assembly	0	101
After Disassembly/Reassembly	3	95
After Swapping Platens	2	103
After Swapping Print-heads	94	5
After Returning Print-heads	1	98

The Third Partition - The print-head was actually a sub-assembly which contained a print-head hybrid, mounting hardware, and head frame. The natural impulse was to blame the print-head hybrid, but we continued the investigation like good detectives would.

The Investigation

The ITC now looked like this:



The first step in the new partition was to take the print-heads out of both printers, disassemble the print-head subassemblies, reassemble and remeasure Y_{pr}:

Hot Print Score, 10 successive prints, 3000 ribbon

	Printer 3	**Printer 0**
Original Assembly	0	101
After Disassembly/Reassembly	3	95
After Swapping Platens	2	103
After Swapping Print-heads	94	5
After Returning Print-heads	1	98
After Print-heads Dis/Reassembly	72	48

THAT'S IT!   The RC is somewhere in the assembly of the print-heads.

At this point, we asked for help from Julie, considered to be the best image quality technician in the pre-production area.   When told the RC was in the assembly of the print-head, she said,

"Well, it must be the head pressure spring adjustment. Engineering set the spec at 150 grams, but I've been getting better image quality at 200. I'm writing a mod to get it changed."

The Investigation

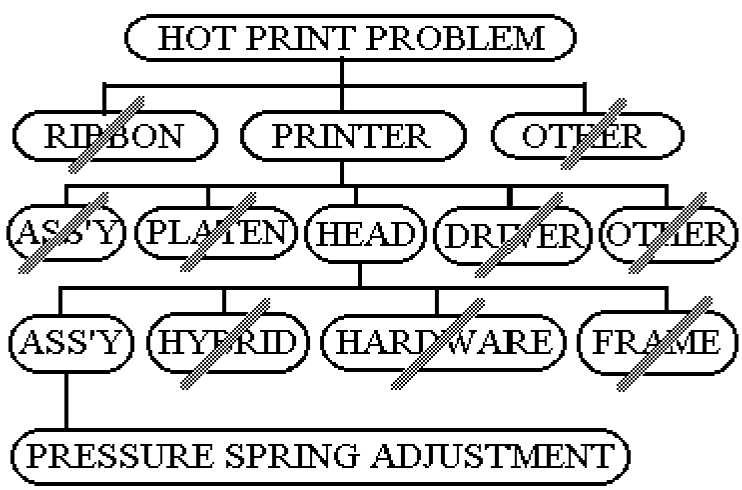
The Principal Engineer scratched his head and said,

"That would improve heat transfer from the media to the platen. I'll bet that's it! We can tolerate 200 grams from a wear standpoint."

Everything was falling into place. We now had a prime suspect for the RC with 100% agreement that changing the pressure spring adjustment would fix the hot print problem. At this point, most engineering teams in my experience would have expedited the mod to 200 grams, considered the problem solved, and gone on to other important matters.

But we still had on our detective hats. What we really had was some evidence and a plausible theory, but no proof. A prime suspect had been identified, so we didn't need any more partitioning for the time being. What we needed was an experiment that would give us proof that the RC either was or was not the print-head pressure spring adjustment.

At that point, the Investigation Tracking Chart looked like this:



Getting Proof

A rank order test is a simple but effective proof experiment. We decided to do this test using six samples for 95% statistical confidence, although as few as five samples can generate 90% confidence.

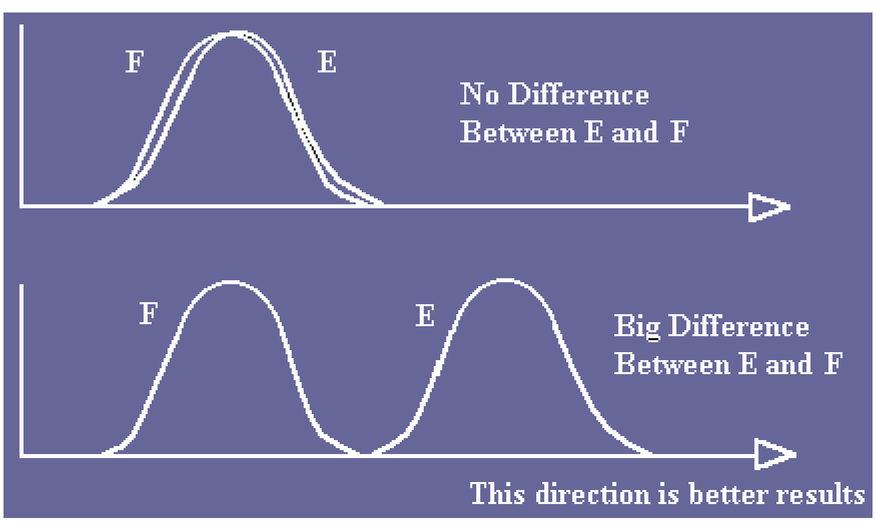
Rank Order Proof Test - Six of the ten printers were selected by tossing a pen onto a random number table. Another random order was used to pick three printers for Julie's new head pressure setting of 200 grams. They were labeled "E" (RC Eliminated), and the remaining three - set to 150 grams - were labeled "F" (RC Functional). All six were tested simultaneously and the results were rank ordered.

RANK ORDER PROOF TEST
Score, 10 successive prints, 3000 ribbon

Machine	Score
E_1	0
E_2	0
E_3	0
F_2	32
F_1	58
F_3	74

The RC is proven to be the head pressure adjustment with 95% confidence!

How It Works – Why is it 95%? Well, the full range of outcomes is this:



Getting Proof

If there were no difference between E and F, what is the probability of the outcome that all three “E”s would outrank all three “F”s ? As indicated in the section on math for simple experiments, it is the ratio of this single outcome to the total number of outcomes possible. The total number of combinations possible for a set of “r” (or E) things mixed with another set of “(n-r)”(or F) things is:

$$
\begin{aligned}
{}_nC_r &= n!/r!(n-r)! \\
{}_6C_3 &= 6!/3!(3)! \\
&= (6)(5)(4)(3)(2)(1)/[(3)(2)(1)][(3)(2)(1)] \\
&= 20
\end{aligned}
$$

Therefore, if there were to be no difference between E and F, there is one chance in twenty (5%) that all three “E”s would outrank all three “F”s. Conversely, we have 95% confidence there **IS** a difference between E and F. But we can’t say how much that difference is without applying more complicated statistical math. However, when the three top-ranked E results also fall in a range acceptable for a problem solution (as they did in the hot print case), this test gives you dependable proof.

The Rank Order Proof Test has this advantage: that the confidence goes up rapidly as the sample size is increased modestly. For example, the total number of combinations for 8 samples of 4 “E”s and 4 “F”s is 70, which gives a confidence of 98.6% if all the “E”s outrank all the “F”s.

Another advantage of this test is adaptability. In the pre-production environment, it is often difficult to get matching sample sizes for “E” and “F”. Typically, one type is hard to get and the other is easy. The test will work with just one or two samples in one (either “E” or “F”) type. The other type just requires more samples to get the desired confidence number, as calculated by the formula for combinations. As always, all “E”s must outrank all “F”s, and all “E”s must fall in a range acceptable for the problem solution.

Epilogue

Win-Win Situation – Except for the VP of Procurement (who was embarrassed), everyone came out a winner in the Hot Print Case. The vendor was happy because $500K of lot 3000 ribbon in the warehouse was saved. We were happy because the printer was fixed with a simple procedure modification which had no schedule impact at all. When you can identify interactions or, as in this case, two very strong main effects, you have flexibility to implement solutions. This increases the chance for win-win situations to occur.

Finding RC in the Ribbon – Before the vendor team left, we all agreed that the lot 3000 ribbon needed to be fixed in the near future. Since we were better at problem solving investigations and since our labs were more available for experiments than their labs were, we agreed to conduct the investigation (instead of the vendor) for a modest consideration: discounting the lot 3000 ribbon already in the warehouse (a deal cleverly negotiated by the VP of Procurement).

We talked about looking for natural partitions based on known differences between lot 2000 and lot 3000. Alas, there was little data to help us. We had samples of Best and Worst, but an experimental partition such as a Swap Test wouldn't work because the ribbon was monolithic, meaning it can't be taken apart and put back together again.

X-Y Correlation Test– This is a partitioning experiment for monolithic systems. It is also one of the few times we think about possible "X" or causal factors. We do so with strict discipline not to theorize (not to think like inventors, that is). The ribbon vendor's team identified six "X" factors that are suspect due to process redesign from lot 2000 to lot 3000. We used those six "X"s in a X-Y Correlation Test. The idea of this test is to take n sample pairs where each pair represents the full range Best-to Worst Y_{pr}. Otherwise, the sample pairs should be as much alike as possible. Measuring the "X"s for the n sample pairs, we look for directional matches of any X with Y_{pr} for all n sample pairs. Any X that doesn't have a directional match for all pairs is eliminated.

Epilogue

After identifying the six "X" factors, the next step is to calculate "n", the number of sample pairs required. This is just like the coin flip example in the section on probability in repeated trials, because there are six "X"s, which is six opportunities for any one "X" to align with Y_{pr} n times by pure chance. The equation for probability in repeated trials is:

$$_nP_r = {_nC_r}p^rq^{n-r} = [n!/r!(n-r)!]p^rq^{n-r}$$

Solving this equation wasn't easy, so we brought in our resident math and computer "geek" to do the computing. He furnished us with this table:

Number of "X"s	Number of Pairs Required for 90% min. Confidence
1-3	5
4-7	6
8-14	7
15-28	8

The table tells us that six sample pairs are required. On printer #7 (which gave us the best separation on ribbon), set at 150 grams, we checked out several lot 2000 (good) and lot 3000 (bad) ribbons, until we found six lot 3000 ribbons which had high hot print scores (>60) and six more lot 2000 with low scores (<10). Next, the lab did Quad-Five Repeatability Tests to verify the measurement systems for the six "X"s. Finally, we set up the random order and got the measurements. The results were tabulated as follows:

	Pair 1		Pair 2		Pair 3		Pair 4		Pair 5		Pair 6	
Ribbon Type	2k	3k	2k	3k	2k	3k	2k	3k	2k	3k	2k	3k
Y_{pr} (Hot Print Score)	08	91	00	74	07	89	04	63	09	69	02	85
Film thickness	3.1	3.2	3.2	3.0	3.3	3.2	3.2	3.1	3.0	3.1	3.0	3.0
Backcoat thickness	.2	.5	.1	.5	.2	.4	.1	.3	.2	.4	.1	.5
Yellow thickness	2.2	2.4	2.4	2.3	2.4	2.2	2.1	2.2	2.3	2.3	2.2	2.1
Magenta thickness	2.7	2.6	2.6	2.4	2.4	2.5	2.4	2.6	2.6	2.7	2.5	2.4
Cyan thickness	2.3	2.8	2.5	2.8	2.7	2.6	2.4	2.8	2.9	2.2	2.5	2.7
Film glass trans. temp.	88	83	84	87	87	83	89	85	87	86	82	88

Epilogue

Only the backcoat thickness went from low to high consistently, aligning with the hot print score for all six pairs. All the other "X"s went low to high in some pairs, but from high to low in others, so they are eliminated. From the X-Y Correlation Test we had 90% statistical confidence that the RC was backcoat thickness. Nevertheless, the vendor did a rank order proof test during the next pilot run to prove beyond any doubt that it was indeed the RC. The vendor adjusted the process to keep the backcoat thin and applied statistical process control, and we never saw hot print again.

One last comment about X-Y Correlation Tests: often it is useful just to make visual observations, noting the differences of successive Best/Worst pairs. As you go along, any difference not continually seen on succeeding pairs can be eliminated from the investigation. Any consistent difference remaining after "n" pairs (per the table) have been observed, is strong evidence to identify the RC.

PART III - WRAP UP
Serial Factorial

The Serial Factorial Test - Serial Factorial is an effective experiment for processes or problem systems that cannot be disassembled or when there is only one system available. It is similar to the Swap Test described in the Hot Print case study, except that instead of swapping subassemblies we swap "X" factors. It has the feature of factorial experiments in that two levels for each input "X" variable are tested. However, here it is done one at a time in a converging process of elimination. This creates the opportunity to significantly reduce the number of trials needed for the experiment.

Range Limits - In the following case study, the idea of establishing range limits is introduced. This enhances the capability of the experiment to detect interactions. Range limits can be used in the Swap Test experiment as well, but was left out of the Hot Print case study for instructional purposes.

Ajax Plating Case Study – In this fictitious case study, Ajax Plating Specialists, Inc. was having difficulty meeting a customer's thickness specification for nickel plating on a new part. As usual, the customer was getting upset, threatening to take the business elsewhere. The Ajax manager didn't want to create a special plating line for this part. He had to find out if the existing line could be set up to get the job done right.

Ajax's best plating engineer and another problem solving engineer were assigned the problem. Their first step was to take care of the three prerequisites. (Remember them?)

Months of prior work had narrowed the list to six suspect "X" variables. Using the technical expertise of the plating engineer, they arranged the list in expected order of decreasing importance. (For Serial Factorials, four to eight "X"s is a good number. If the list is too long, other partitions should be used to eliminate "X"s.) They made a prediction which level of each "X" variable would likely produce a "Best" or "Worst" Y_{pr} result. Their tabulation is shown on the next page.

Serial Factorial

	"X"- Factor	Best	Worst
Most likely	Conveyor Speed (V)	2.2 fpm (V_B)	2.8 fpm (V_W)
|	Anode Current (I)	35.0A (I_B)	37.5A (I_W)
|	Solution Temp. (T)	50C (T_B)	40C (T_W)
|	Electrolyte Sat. (E)	80% (E_B)	70% (E_W)
|	Solution Ph (P)	7.6 (P_B)	7.0 (P_W)
Least likely	Mounting Config (M)	Mirror (M_B)	Uniform (M_W)

The first experimental trial started by flipping a coin which came up tails for "Worst", then they set up the process for $\mathbf{V_W I_W T_W E_W P_W M_W}$, ran the line and took a sample. The plating thickness measured 45 uM. Earlier, while completing the prerequisite defining the full range of Y_{pr}, they had set estimated limits of <50uM for "Worst", and >65uM for "Best". Next they set up the process for $\mathbf{V_B I_B T_B E_B P_B M_B}$, ran the line, took a sample, and got 69uM.

"Hooray", they said, "we are in business!" That was because both readings fit the predetermined limits for "Best" and "Worst". Because they wanted to get statistical limits to replace the estimated limits, they ran two more trials of "Best" and two more "Worst". (Of course, they flipped a pen onto a random number table to determine the order of the four trials.) Below are the results of all six trials:

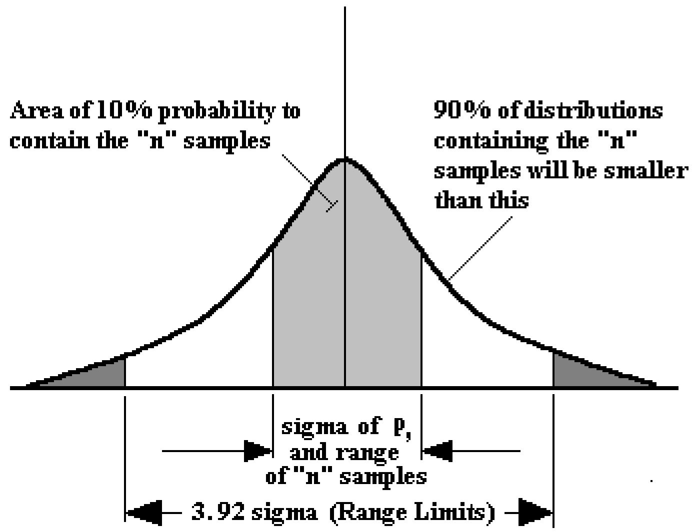
Y_{pr} (Plating thickness, uM)

Best	Worst
69	45
76	44
71	40

Deriving Range Limits – The evidence indicates that the difference between "Best" and "Worst" is the RC. Therefore we can reasonably assume that "Best" and "Worst" are two different distributions. We want a statistical estimate of those distributions to use as a guide to help us identify outlying results. The specifications are: Limits will be set at 3.92 sigma (95%) of the largest distribution computable at 90% confidence as defined by "n" Samples.

Serial Factorial

In the Ajax Plating Case, n =3 for either "Best" or "Worst". The distribution which will be the 90th percentile largest is the one with the center area defined by sigma of p = range of "n" samples, where:



$$p^n = 0.10$$
$$p^3 = 0.10$$
$$p = 0.4642$$

From the Normal Distribution table:

sigma of p = 1.24

For the "Best" sample set of three, the range is 76-69 = 7, and the average is (69+76+71)/3 = 72. For the "Worst" sample set, the range is 45-40 = 5, and the average is (45+44+40)/3 = 43. Range limits are centered on the average.

For the "Best" sample set:

Range Limits = 72 +/- (3.92/1.24)(7/2) = 72 +/- 11 = 61 and 83

For the "Worst" sample set:

Range Limits = 43 +/- (3.92/1.24)(5/2) = 43 +/- 8 = 35 and 51

The term (3.92/1.24) is called a Range Multiplier. Below is a table of Range multipliers for various "n" values:

"n"- Sample Size	Range Multiplier
2	4.81
3	3.16
4	2.53
5	2.18

This table gives the experimenter a trade-off between the cost of running samples and range limits that are appropriate for the situation. Obviously, n = 1 cannot produce statistical range limits, but it is nevertheless viable by continuing the experiment with the original estimated limits.

Serial Factorial

What if the range limits overlap? Minor overlap is allowable, but as a "rule of thumb", no range limit should intrude into the "sigma of p" range of the opposite distribution. Overlap may be reduced by taking more samples as indicated in the Range Multiplier table. Remember that any overlap increases the chance of confounding results and reduces the chance of identifying the RC. Too much overlap means the test cannot continue.

Back to the Ajax Plating Case – The problem solvers were indeed "in business". They had calculated good range limits and were ready to swap variables. (Of course, they had a coin ready for flipping to keep the pair order random.) The whole story can be told by the test results table:

Test #	**Process Settings**	**Y_{pr} (Plating, uM)**		**Notes**
		Best	**Worst**	
1	$V_WI_WT_WE_WP_WM_W$		45	Passes <50 Estimate
2	$V_BI_BT_BE_BP_BM_B$	69		Passes >65 Estimate
3	$V_BI_BT_BE_BP_BM_B$	71		
4	$V_WI_WT_WE_WP_WM_W$		44	Calculated Limits:
5	$V_BI_BT_BE_BP_BM_B$	76		Best: 61 to 83
6	$V_WI_WT_WE_WP_WM_W$		40	Worst: 35 to 51
7	$\underline{\mathbf{V_B}}I_WT_WE_WP_WM_W$		49	Inside Range Limits
8	$\underline{\mathbf{V_W}}I_BT_BE_BP_BM_B$	65		Inside –**V eliminated**
9	$V_B\underline{\mathbf{I_W}}T_BE_BP_BM_B$	55		Outside – **I significant**
10	$V_W\underline{\mathbf{I_B}}T_WE_WP_WM_W$		50	Inside
11	$V_BI_B\underline{\mathbf{T_W}}E_BP_BM_B$	74		Inside
12	$V_WI_W\underline{\mathbf{T_B}}E_WP_WM_W$		42	Inside – **T eliminated**
13	$V_BI_BT_B\underline{\mathbf{E_W}}P_BM_B$	64		Inside
14	$V_WI_WT_W\underline{\mathbf{E_B}}P_WM_W$		59	Outside – **E significant**
15	$V_W\underline{\mathbf{I_B}}T_W\underline{\mathbf{E_B}}P_WM_W$		71	**Worst to Best Swap**
16	$V_B\underline{\mathbf{I_W}}T_B\underline{\mathbf{E_W}}P_BM_B$	41		**Best to Worst Swap**
				IE interaction is the RC

The first "X" swap was "V" (the Conveyor Speed) which produced in-range values for "Best" and "Worst". This result eliminates "V". Note that this step has value because we don't have to worry about "V" any more. The next swap was "I" (the Anode Current) and the "Best" value was out of range, which makes "I" significant. The third swap resulted in eliminating "T" (the Solution Temperature). On the fourth swap, "E" (the Electrolyte Saturation) was shown to be significant by an out-of-range "Worst" value.

At this point, the team decided to take a chance that "I" and "E" were the only significant factors. The next swap was "I" and "E" together. Voila! The "Best" became "Worst" and "Worst" became "Best". Or, in other words, "Best" and "Worst" followed the "IE" combination and the other "X"s made no difference. The "IE" interaction was identified and proven to be the RC!

The plating engineer said, "Wow, I've been playing with those settings for a long time – but I never got the right combination!" He was right. It is very difficult to find interactions by "cut and try".

Notes on Serial Factorial Tests – Typically, the most difficulty is encountered in the beginning, getting the true "Best" and "Worst" results. The initial estimated limits should be set around the top third of the full range of Y_{pr} for "Best" and the bottom third for "Worst". If results don't fall in these ranges, or if the calculated ranges have too much overlap, the test cannot continue. There are two possible causes:

a) The "Best" and "Worst" levels were incorrectly assigned for one or more "X"s. Try swapping variables one-at-a-time in top down order. If any of the swaps give acceptable "Best"-"Worst" separation, continue with the experiment from that point.

b) The RC may have been left off the list of "X"s. Review prior partitions in the investigation which may have incorrectly eliminated a potential RC, or conduct further partitioning tests as required.

Interactions

An interaction is a change in the Y_{pr} caused by two or more "X"s acting together which exceeds the sum of their separate effects. Two factor interactions are by far the most common. Three factor interactions are somewhat rare and four or more factors are extremely rare.

Problems caused by interactions often go unsolved if a "cut and try" approach is used. But interactions can be reliably identified with a persistent, converging process of elimination. As an investigation proceeds, certain signs give warning that the RC may be an interaction.

Signs of an Interaction:

1. Best or Worst Y_{pr}'s seem to occur intermittently, randomly, or seldom. The full range of variation is difficult to find.

2. The results of tests are often confusing. Patterns of variation seem inconsistent. A reported problem is difficult to reproduce.

Finding an Interaction:

1. Look for the signs of an interaction. If in doubt, assume the interaction exists.

2. Stick with a converging process of elimination. Be relentless with partitions until you find full range "Best"-to-"Worst" results and then use those opportunities to eliminate as many "X"s as possible. The key is to be relentless, because the shorter the list of "X"s the better.

3. For "n" "X" factors, 2^n combinations of "B" and "W" levels exist, because each "X" can be at either of the two levels. (Remember that assigning "B" and "W" to predetermined levels is purely arbitrary.) Make a list of every combination. Next, rank every "B"/"W" combination using technical expertise according to the likelihood of being the RC.

4. Run tests in that order until you find a "B"/"W" combination which produces a "Best" or a "Worst" Y_{pr} result. When that happens, run the mirror image "B"/"W" combination for verification, needing to get the opposite full range Y_{pr} result to pass the test. See the example on the next page.

5. If the entire list of combinations does not produce a "Best" and "Worst" Y_{pr} result, the right "X"s weren't on the list. Go back to step 2.

Interactions

Example:

Three "X" factors are left after an exhaustive convergent investigation. (This does not necessarily mean it will be a three factor interaction.) The number of combinations is 2^n. In this case n = 3, so there are $2^3 = 8$ combinations. Using technical expertise, the problem solving team ranked the combinations from left to right as shown:

	<Most likely to be RC				Least Likely to be RC>			
"X" Factor	1	2	3	4	5	6	7	8
A	B	B	B	W	W	W	B	W
B	B	B	W	B	W	B	W	W
C	B	W	B	B	B	W	W	W

Results:

Tests 1 and 2 produced a mid-range Y_{pr} result.

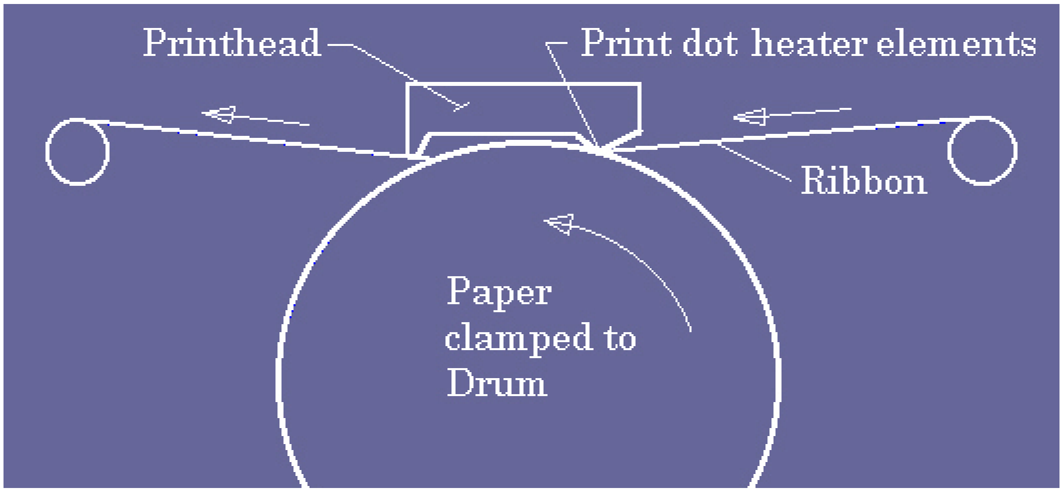
Test 3 produced a "Best" Y_{pr} result.

The mirror image of Test 3 is Test 6, which was the next (verification) run. It resulted in a "Worst" Y_{pr}.

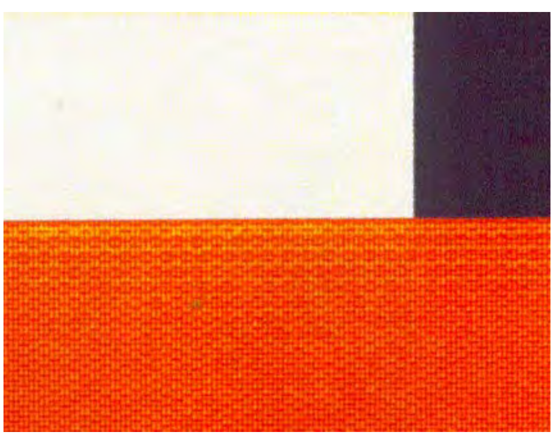
The RC is an interaction involving the factors ABC, combinations 3 and 6. It could be a three factor interaction ABC, or a two factor interaction AB (BC is eliminated because it was the same in tests 2 and 6, and AC is eliminated because it was the same in tests 1 and 3). There are two options at this point: run another test to solve the dichotomy or control all three variable levels to combination 3.

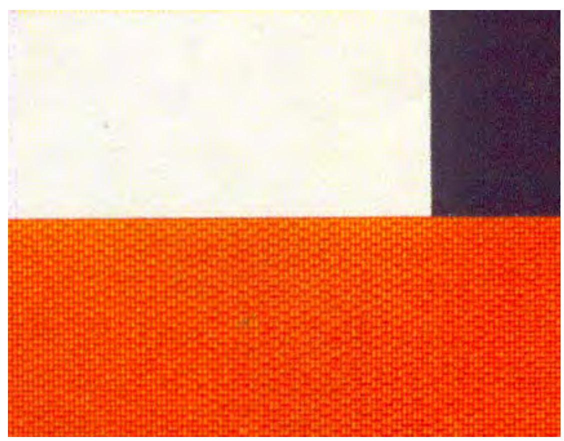
Test 4 produced a "Worst" Y_{pr} result, which positively identifies the RC as the two factor interaction AB.

The Case of The Ripples – The culmination of wax transfer technology was reached with the Model 222 Turbo printer. The platen roller of previous designs was replaced with a rotating drum which clamped the paper for rapid execution of the multiple pass printing process with the ribbon and printhead. This design increased print speed about 2X over previous paper path configurations. Of course, color richness and image resolution were also improved to produce breathtaking prints.



All prototype testing had gone well and the Ongoing Reliability Test program (ORT) was well underway as the Pilot Production run was nearing its end. Then we got a phone call. One of our beta site customers, a graphic arts firm, was complaining about a ripple effect showing up on their output. We got some of their prints and a copy of the image file which we sent to ORT to run. None of the printers in ORT showed any ripples printing that image.

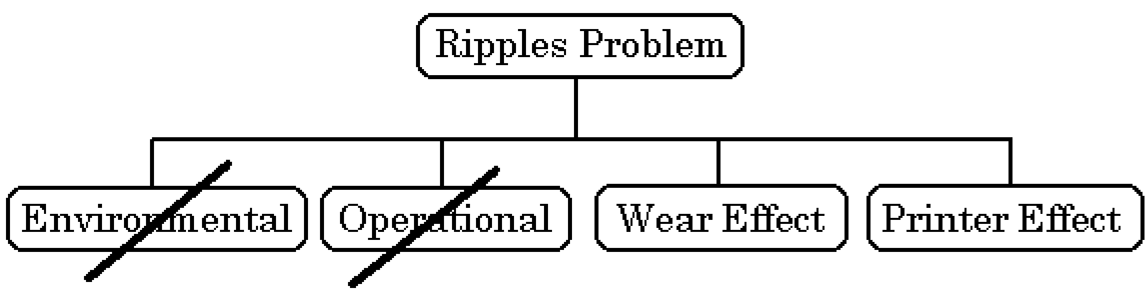

Beta Customer Image


ORT Image

After more discussions with the beta site customer, we discovered that the problem did not occur when the printers were new. It developed and grew worse over time. Since all our machines had copy counters embedded in the firmware, we did a comparison. To our dismay, several of the printers in ORT had about the same copy count as the beta site printer (about 2000 prints). If it was a wear out type of problem, why didn't it show up in ORT?

What to do next? We were stumped. Then the light dawned. This was a classic sign of an interaction – "a reported problem is difficult to reproduce". Since the only "Worst" printer was at the beta site, we decided to go there in case some environmental factor was interacting. And, of course, we took one of our "Best" printers from ORT.

At the beta site customers print lab, we found there was no significant operational or environmental differences between it and our own ORT lab. So we eliminated those areas as possibly containing interacting factors. Next, we spent a little time plotting strategy. The best clue at hand was that the problem developed and worsened over time, so we had to keep "Wear Effect" as an area for further partitioning. But we couldn't pursue it there at the beta site, because our "Best" and "Worst" samples had about the same copy count. That left only the "Printer Effect" area for immediate partitioning. At this point, our ITC looked like this:

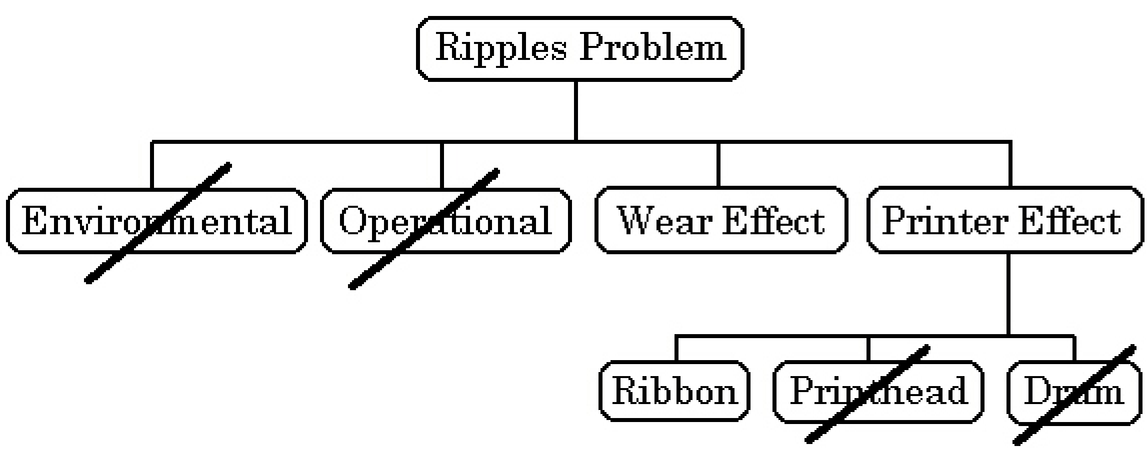


The next step was to take care of the prerequisites. Fortunately, our beta site customer had an archive file of prints which we could use. This file contained one print taken daily from the output of each printer. We could see the severity of "ripples" progressing with time from one faint "ripple" at the onset to six or seven "ripples" at 2000 copies. "Ripple" count was an obvious choice for the Y_{pr} variable – it met the minimum requirement of seven levels of discrimination with the full range of variation spanning zero to six. I was designated the official "ripple" counter who had to pass a Quad-Five repeatability test. My repeatability error limit was .087 X 7 = .61 "ripples". Since "ripples" are by nature integers, my actual repeatability error limit was (gulp!) zero.

I used a random number table to select five sample prints and used it again to determine my measurement sequence. Because there was a fair degree of subjectivity counting the last faint "ripple", I was not confident that I would pass. My jaw dropped when the tally sheet showed I had actually rated each print twice at exactly the same number!

With the prerequisites completed, the next step was to decide on the first level partitions. We asked the printer expert on our team this question: given the nature of the "ripples" artifact and obvious "wear" effect, how could we best partition these "Best" and "Worst" machines for a Swap Test? He told us to look at the Ribbon, the Printhead, and the Drum. (The first two because they have continual sliding frictional contact, and the latter because torsional vibration in the Drum assembly could conceivably produce a "ripple" artifact.)

As we began setting up the Swap Test, the test technician pointed out that the ribbon in the "Worst" machine was different. It was a four-pass ribbon (yellow-cyan-magenta-black) used by certain customers because it produced a superior black. The ribbon in the "Best" machine from ORT was a three-pass ribbon (yellow-cyan-magenta) used by a majority of our customers because it had lower print cost and higher print speed. Before starting the formal Swap Test, we simply swapped the ribbon canisters (done without any dis-assembly). We should have bought a lottery ticket that day because luck was with us—swapping ribbons caused a full reversal! The score on the "Best" printer went from zero to five, and the "Worst" printer from six to two (which we repeated three times). The ITC now looked like this:



So we discovered the four-pass ribbon contained one factor of the RC interaction. But there was more work to do. We went back home to begin partitioning the "Wear Effect" area. Starting with a brand new printer from the pilot production line we verified that it made zero-"ripple" prints with either three-pass or four-pass ribbon. A printer from ORT with over 2000 copy count scored a seven on four-pass ribbon and one on three-pass ribbon. These two printers would be our "Best" and "Worst" for the next Swap Test. Since the ribbon was already accounted for, the remaining partitions would be Printhead, Drum, and Other.

Obviously, we would use four-pass ribbon for this swap test. On the printhead swap, the "Best" printer scored a six and the "Worst" printer scored a one. Bingo! The Printhead contains another factor of the RC interaction. For proof, we did a capping run on a single printer (the one from ORT) with all four combinations. Results as follows:

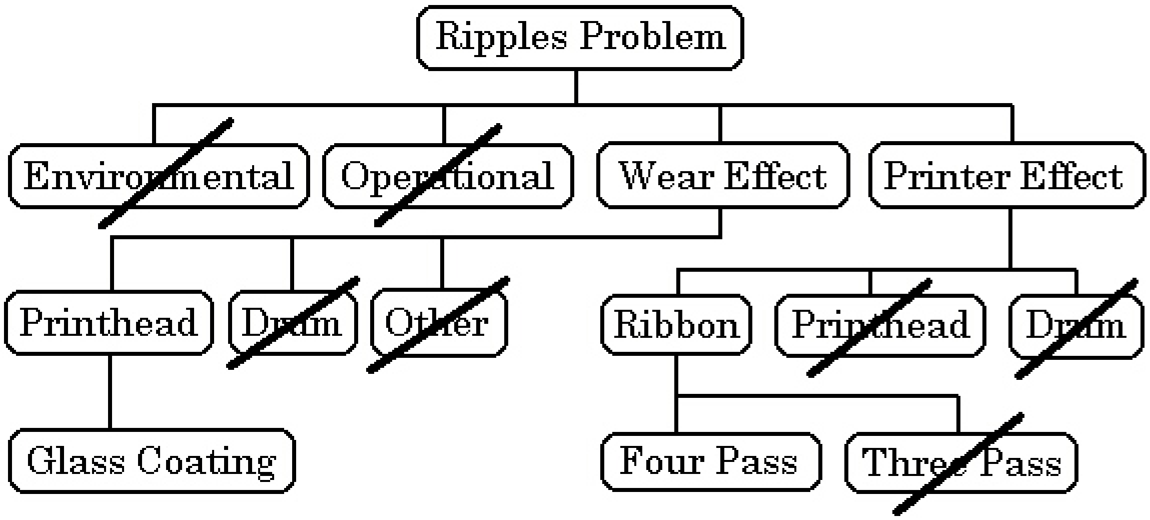
	Four-pass Ribbon	Three-pass Ribbon
0 Copy Printhead	One	Zero
2000 Copy Printhead	Six	Three

It takes both factors to produce the full range of variation. The RC is proven to be the Ribbon-Printhead interaction.

To make a long story shorter, we used X-Y Correlation tests to further partition both Ribbon and Printhead. Auger analysis and physical tests showed several differences between the three-pass and four-pass ribbons. It turned out that three-pass ribbon came from a large Asian manufacturer, and four-pass from a smaller US manufacturer.

Meanwhile, comparing 0 Copy Printheads with 2000 Copy Printheads showed the natural micro-cobbled surface of glass coating the thermal elements consistently wore smooth over 2000 copies. As the glass surface grew smoother, the four pass ribbon tended to stick to it and stretch before breaking away. Of course, we verified that with a Rank Order Proof Test. The final ITC is shown on the next page.

Interactions



During the several months lead time for the US manufacturer to change his ribbon making process, we worked out a simple process to rework smooth printheads with ultra fine sandpaper. So we began shipping printers and ribbons to our four-pass customers whom we could manage with routine service calls and discreet sandpapering.

Interactions may require hard work to solve. But correctly identified and understood, interactions always provide the opportunity for multiple solution options.

Managing Investigations - Managers should not micro-manage problem solving teams. Their job is to set the goal, set up the team, provide resources, review progress, and recognize success. During the reviews, a good manager would simply confirm that the team is executing the fundamentals: that Y_{pr} is expressed as a variable, that the full range of Y_{pr} is defined, that the measurement systems are proved repeatable, that all natural partitions have been explored, that experimental partitions follow a convergent process-of-elimination, and that statistical proof is presented for the RC.

The worst thing that can happen to a good problem solving team is to have a manager that thinks like an inventor. That type of manager does not "get it"! Typically the team will be directed to investigate ever increasing lists of potential causes, which conflicts with a convergent process of elimination. The conflict will only get worse as time goes on. In this case, the only recourse for the team is to find a way to educate their manager.

Growing the FRC Culture - Like engineering, convergent problem solving is as much art as science. For individuals, it's a skill to develop. Some will excel and others will not. For organizations, it's a culture to grow, complete with it's own terminology and social events. The leaders of the cultural development are those who excel. Those less skillful get help from the culture to solve their problems.

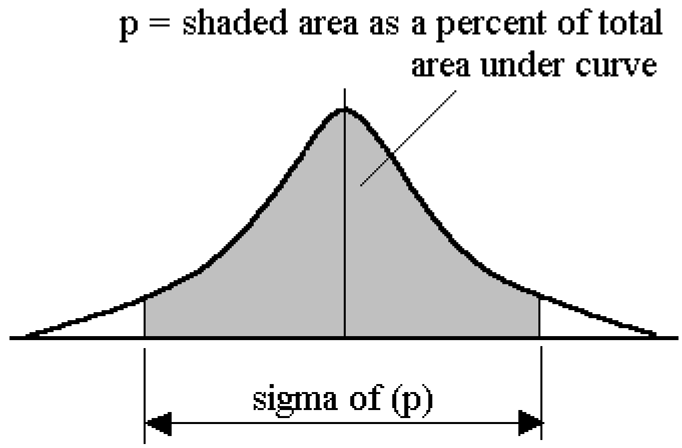
Problem solving is best done by small teams. The ideal team would be one system expert and one problem solving expert. If the team is more than three people, the process tends to slow down dramatically.

When a team solves a tough problem, that success is the seed for growth. Other problem solving teams want to know about it. For that matter, people who are not problem solvers also want to know about it. Managers should promote formal and informal communication opportunities for them. After all, detective stories have fascinated people for centuries. So stories about problem solvers who think like Detective Columbo or Sherlock Holmes will make easy listening, and as people listen the culture grows.

Normal Distribution

Values for p and sigma are double-sided and symmetrical.

Values for p are in lighter faced type.



sigma of (p)	**+0.00**	**+0.02**	**+0.04**	**+0.06**	**+0.08**	**+0.10**	**+0.12**	**+0.14**	**+0.16**	**+0.18**
0.0	0.00	0.80	1.60	2.40	3.20	3.98	4.78	5.58	6.38	7.18
0.2	7.96	8.76	9.56	10.34	11.14	11.92	12.72	13.50	14.28	15.06
0.4	15.86	16.64	17.42	18.20	18.96	19.74	20.52	21.28	22.06	22.82
0.6	23.58	24.34	25.10	25.86	26.62	27.36	28.12	28.86	29.60	30.34
0.8	31.08	31.82	32.56	33.28	34.00	34.72	35.44	36.16	36.88	37.58
1.0	38.30	39.00	39.70	40.38	41.08	41.76	42.46	43.14	43.80	44.48
1.2	45.14	45.82	46.48	47.14	47.78	48.44	49.08	49.72	50.34	50.98
1.4	51.60	52.22	52.84	53.46	54.08	54.68	55.28	55.88	56.46	57.04
1.6	57.62	58.20	58.78	59.34	59.90	60.46	61.02	61.56	62.12	62.66
1.8	63.18	63.72	64.24	64.76	65.28	65.78	66.30	66.80	67.30	67.78
2.0	68.26	68.76	69.22	69.70	70.16	70.62	71.08	71.54	71.98	72.42
2.2	72.86	73.30	73.72	74.16	74.58	74.98	75.40	75.80	76.20	76.60
2.4	76.98	77.38	77.76	78.14	78.50	78.88	79.24	79.60	79.94	80.30
2.6	80.64	80.98	81.32	81.64	81.98	82.30	82.62	82.94	83.24	83.54
2.8	83.84	84.14	84.44	84.72	85.02	85.30	85.58	85.84	86.12	86.38
3.0	86.64	86.90	87.14	87.40	87.64	87.88	88.12	88.36	88.58	88.82
3.2	89.04	89.26	89.48	89.68	89.90	90.10	90.30	90.50	90.70	90.90
3.4	91.08	91.28	91.46	91.64	91.82	91.98	92.16	92.32	92.50	92.66
3.6	92.82	92.98	93.12	93.28	93.42	93.56	93.72	93.86	93.98	94.12
3.8	94.26	94.38	94.52	94.64	94.76	94.88	95.00	95.12	95.22	95.34
4.0	95.44	95.56	95.66	95.76	95.86	95.96	96.06	96.16	96.24	96.34
4.2	96.42	96.52	96.60	96.68	96.76	96.84	96.92	97.00	97.08	97.14
4.4	97.22	97.28	97.36	97.42	97.50	97.56	97.62	97.68	97.74	97.80
4.6	97.86	97.92	97.96	98.02	98.07	98.12	98.17	98.22	98.27	98.32
4.8	98.36	98.40	98.45	98.49	98.53	98.57	98.61	98.65	98.69	98.72
5.0	98.76	98.79	98.83	98.86	98.89	98.92	98.95	98.98	99.01	99.04
5.2	99.07	99.09	99.12	99.15	99.17	99.20	99.22	99.24	99.26	99.29
5.4	99.31	99.33	99.35	99.37	99.39	99.40	99.42	99.44	99.46	99.47
5.6	99.49	99.50	99.52	99.53	99.55	99.56	99.58	99.59	99.60	99.61
5.8	99.63	99.64	99.65	99.66	99.67	99.68	99.69	99.70	99.71	99.72
6.0	99.73									

Random Number Table

39	52	87	24	84		81	61	61	87	11
07	58	61	61	20		90	76	70	42	35
40	18	82	81	93		18	61	91	36	74
74	62	77	37	07		32	39	21	97	63
78	46	42	25	01		62	09	53	67	87
12	30	28	07	83		76	37	84	16	05
05	04	14	98	07		46	97	83	54	82
47	66	56	43	82		34	41	48	21	57
63	43	97	53	63		67	04	90	90	70
79	49	50	41	46		91	70	43	05	52
83	76	16	08	73		14	38	70	63	45
51	32	19	22	46		72	47	20	00	08
05	46	65	53	06		25	16	30	18	89
65	25	10	76	29		36	81	54	63	25
64	39	71	16	92		04	51	52	56	24
71	85	71	59	57		92	78	42	63	40
04	92	17	37	01		45	19	72	53	32
15	19	11	87	82		01	29	14	13	49
38	38	47	47	61		66	16	44	94	31
54	15	58	34	36		72	84	81	18	34

Index