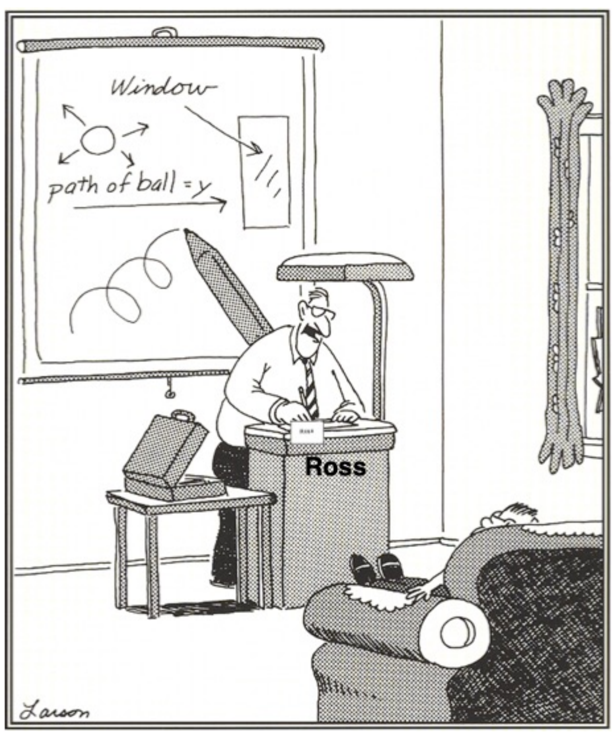
Ross' Rules

Practitioner Edition



Eventually, Billy came to dread his father's
lectures over all other forms of punishment

A Note to the Reader

This book represents several decades of experience investigating engineering and business problems. My hope is that the ideas presented here help engineers and scientists improve how they think about data, experimentation, and discovery. You are welcome to share this PDF with colleagues who may find it useful. In fact, I encourage it. The goal of Sigma Science is to improve investigative thinking and practice. If you find the material valuable, please consider:

- sharing the book with other engineers or scientists
- using the ideas in your own investigations
- citing the work when appropriate

Table of Contents

Preface

I am writing this book near the end of my career. Over four decades, I have accumulated experience, insight, and perspective, often as a result of my own failures. I would like to share my knowledge with the next generation of critical thinkers. Traditional mechanisms for dissemination have limitations. Papers reach limited audiences, and internal training and short courses are transient. Corporate motivation often disappears when people move on and, as my wife has reminded me for forty-five years, this kind of work is certainly not for everyone. I want to spark your interest. My objective is not to convince everyone, but to inspire curiosity and reflection to those who are inclined toward thoughtful inquiry. This book is one medium, along with others still in development, to help me disseminate these ideas. Over the years I have done project work, written papers, developed course materials, and taught engineers how to apply statistical methods in real settings. At the time, those efforts and the materials delivered represented my best understanding of the subject. More importantly, they were attempts to prevent others from repeating mistakes I had already made. What I did not realize, at the time, was that my own thinking was changing as I taught.

Early in my career, I placed too much confidence in tools. If an experiment was designed carefully and analyzed correctly, I expected solutions to follow. When it didn't, my first instinct was to look for execution errors, missing factors, or insufficient measures. Sometimes that was true. Often it wasn't. What I came to understand, slowly, is that many of those experiments were not "wrong," but **premature or misapplied**. I was optimizing before I had oriented myself. I was applying powerful methods before I understood where the leverage actually was. In particular, I learned *late* the importance of understanding variation before designing experiments (DOE). Components of Variation (CoV) was not something I used early enough in my career. Had I done so, several experiments would have been more effective, better targeted, and easier to interpret. That realization helped explain their limitations. It also changed how I approached every investigation afterward. This pattern repeated itself in other areas. Prediction became more important than analysis. Iteration became more important than efficiency or one-shot solutions. Engineering interpretation became more important than statistical.

This is not a statistics textbook. Many excellent books already exist for that purpose. I am not attempting to re-teach statistical methods, nor to turn engineers and scientists into statisticians. Instead, this book is about **augmenting thinking**, specifically how engineers, scientists, and technical professionals frame problems, collect data, analyze data, interpret results, and learn about causality. Engineering or any learning is not linear. It loops, doubles back, stalls, and occasionally jumps forward. Any diagram that pretends otherwise is a lie of convenience.

> *"Two equally competent investigators presented with the same problem would typically begin from different starting points, proceed by different routes, and yet could reach the same answer.* ***What is sought is not uniformity, but convergence****."*[1]

[1] Box, Hunter and Hunter, Statistics for Experimenters

This book is also not a workflow or recipe. Application of the methodology does not proceed in a straight line. The sequence that looks optimal in hindsight is rarely obvious at the outset. Any book that presents a single, canonical path or set of steps through problem solving is offering comfort, not truth. I do hope this book contains some of the best practices I have learned.

The intended audience is practicing engineers and scientists who are responsible for designing and improving products and processes, conducting investigations, or making decisions. I assume some familiarity with statistics, strong technical domain knowledge, and a working environment where real constraints, time, cost, organizational pressure, shape how studies are conducted. A recurring theme in this book is the distinction between **discovery** and **execution**. Execution assumes knowledge already exists. Discovery exists because it does not. Most managers want a solution, any solution will do. What they fail to recognize is the value of understanding the situation, discovery work, before a solution is proposed. The expectations, timelines, and metrics applied assume certainty that simply is not there.

Over decades of teaching, consulting, and collaboration, I have observed a recurring pattern. Substantial effort is expended. Data are collected. Sophisticated analyses are performed. Yet learning is often limited, fragile, or difficult to translate into sustained improvement. When this happens, the problem is rarely the statistical tool itself. More often, it is the **way the tool was applied**, the **way the data were acquired**, or the **lack of disciplined thinking before the study began**. Efficiency is a *retrospective judgment*. After learning occurs, we can see which steps were unnecessary, which experiments could have been skipped, which data added little value. Before learning, those judgments are guesses. The obsession with appearing efficient, especially early in an investigation, leads to misapplication of tools. Everything you will read here reflects **my** way of thinking about problems, shaped by successes, failures, blind alleys, and hindsight accumulated over decades. It is not *the* right way. In fact, your processes, constraints, incentives, and experiences are different from mine. They should shape how you think, just as mine shaped how I think. I want to be explicit about several biases that shape this book.

First, I have a strong bias toward **determinism**. I believe there is no effect without cause. When variation appears random, it is usually an indication that causes are interacting, unmeasured, or not yet understood. The objective is not to eliminate uncertainty, but to understand its origins.

> *"We do not know a truth without knowing its cause."*
>
> *Aristotle*

Second, I value **practical significance over statistical significance**. Early in my career, I argued loudly and publicly for statistically significant effects that ultimately did not matter. They did not move the response enough to be useful. That experience permanently shaped how I think about data, models, statistical methods and decision-making. Statistical software has become extraordinarily powerful. That power has not reduced the need for reasoning, it has increased it. Software will happily analyze inadequate data, fit inappropriate models, and produce precise answers to meaningless questions. Judgment decides whether data are worth analyzing, an experiment should be run, a result is actionable, and whether uncertainty has been reduced enough to move forward. No tool can do that for you.

Third, my career objective is to understand **causality**, how and why things work. Models are valuable only to the extent they help with understanding and prediction. A complex model that cannot be executed, interpreted, or used to guide decisions is of limited value. I am biased toward learning from data generated within a defined stochastic structure rather than relying on algorithmic models that treat the data-generating mechanism as unknown. I would rather understand an imperfect model than admire a powerful one that I cannot explain.

I do not believe there is a single "right" way to think critically. The most efficient and effective approach is often obvious only in hindsight. What I present here is **my way**, developed iteratively over several decades. I encourage readers to adapt it, critique it, and develop their own approach. Many methods fail not because they are flawed, but because they are misapplied.

> *"Intelligence consists not only in the knowledge but also in the skill to apply the knowledge into practice."*
>
> *Aristotle*

I have divided the book into three parts: Part I introduces and explains the Sigma Science Methodology, why and how to critically think and collect data to provide insight. Part II discusses data collection strategies and the subsequent analysis and interpretation of the acquired data. Part III debates deployment models, organizational constraints and management considerations.

This book is not meant to be followed mechanically. It is meant to provoke reflection, stimulate questions, and encourage deliberate practice. You will get the most value from it if you consider my thoughts and methods, reflect on them and develop your own. One concept that appears repeatedly in this book because it has proven indispensable in practice is the **User Guide**. A User Guide is a self-created, living document that captures how *you* think about applying methodology. Early versions often look like regurgitation, copied flowcharts, borrowed checklists, phrases lifted from lectures, papers or books. That is not a failure. It is how fluency begins. Over time, the User Guide evolves. It accumulates warnings, tips, tricks, heuristics, decision triggers, and lessons learned, not just about products or processes, but about the methodology itself. It becomes a personal reference for due diligence, sequencing, and reflection. The User Guide makes the process of engineering or scientific investigation visible, preserves learning, and accelerates judgment the next time you face uncertainty.

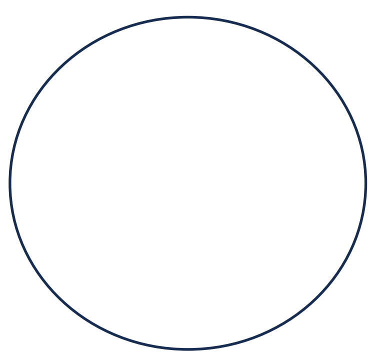
This book can be read as part of *my* User Guide. I hope it helps you begin, or continue, building your own. You will not get the most value from it by reading straight through once, underlining a few passages, and putting it back on the shelf. The ideas here are meant to be **used, tested, and revisited** as your understanding evolves. What follows is guidance, not instruction.

A note on tone and writing style. During my lectures, I use humor where it helps. While I cannot deliver witticisms in this book, I will include a few jokes, tell stories because they make the book more interesting and so people will remember them. I include quotes because they provide emotional identifiers for ideas and tools. Human learning requires sensory input and emotional tags. I will also repeat explanations as they change with context. As you apply ideas from this book, resist the urge to ask, “Am I doing this right?”, A better question is: “What did I learn, and how did that change what I’ll do next?” Failures, in and of themselves, are not bad. Unexamined failures are.

A further note regarding Artificial Intelligence (AI). AI is evolving rapidly, and what role it may ultimately play in scientific and engineering investigations remains uncertain. In my own work, I have found AI useful for expanding domain knowledge, mitigating some of my own biases, and helping me consider a broader range of possibilities. In many ways it functions as a remarkably powerful search engine. However, this strength also introduces limitations. How relevant is the information returned to the specific situation under investigation? What if the product being designed does not yet exist? What if the inference space is fundamentally different from previously documented cases? I have also found that AI systems are currently incapable of designing effective and efficient data collection plans. Such plans require careful reasoning about what must be learned, how uncertainty can be reduced, and how evidence will support or challenge competing explanations. Tools can assist investigation, but they cannot replace the disciplined thinking required to design it. For an example of the type of reasoning required, see the Thought map, Figure 4.2 (page I-49).

Appendices are provided as reference and support material. They are not required to read sequentially and are not substitutes for the thinking process described in the main text.

Lastly, I want to thank the many mentors, teachers, peers, participants, students and organizations that afforded me the opportunity to apply the methodology and learn.



Pie Chart of Procrastination

Part I
The Sigma Science Methodology

The Sigma Science methodology is intended to strengthen an organization by adding reliable, effective, and efficient data reasoning skills to its technical staff, engineers and scientists. Its purpose is not to replace their technical expertise, but to amplify it. By integrating structured deductive reasoning with process and product knowledge, the methodology enables technical professionals to make better decisions, design better studies, and facilitate meaningful improvement in processes and products.
Sigma Science provides guidance on how to:

- perform situational analyses,
- think critically,
- develop rational hypotheses,
- link appropriate tools to the situation,
- acquire data intentionally,
- analyze and interpret results responsibly, and
- learn iteratively within a scientific framework.

The goal is not to turn engineers or managers into statisticians. It is to equip them with disciplined reasoning skills so they can acquire data properly, draw sound conclusions, and influence the changes that must be implemented. Sigma Science is not about statistics. It is about disciplined, structured thinking.

In this book, Sigma Science refers to the deliberate integration of scientific method and statistical thinking as applied to engineering and scientific practice. Effective work depends on hypothesis generation, prediction, data collection, analysis, and interpretation, carried out iteratively.
In practice, Sigma Science relies on four primary tools:

1. Critical thinking about questions, hypotheses, and predictions.
2. Mapping of phenomena, processes, and products to expose potential causes.
3. COV studies to understand existing variation.
4. DOE to deliberately manipulate factors and learn about causality.

Part I focuses on Tools 1 and 2 and on building your own User Guide. Part II develops Tools 3 and 4. Part III examines how Sigma Science operates inside organizations.

"In the face of overwhelming odds, I'm left with only one option,
I'm gonna have to science the shit out of this."

The Martian

Chapter 1
Sigma Science Overview

I have been fortunate to have the opportunity to work with engineers, scientists and statisticians, some very notable, across a wide range of industries, problems, and organizational cultures. What has struck me most is not a lack of intelligence, motivation, or technical skill, but rather how often statistical tools are applied inappropriately and powerful analytical tools are applied in ways that limit learning. The power of Sigma Science is not due to the statistical toolbox, nor in the throwing of engineering resources at problems. Sigma Science achieves sustainable results by merging statistical thinking with engineering and process knowledge through the scientific method. In their most potent form, Sigma Science focuses on increasing the competency and critical thinking of the technical (engineering/scientific/business) organization via the creative merging of process & product knowledge, laws & principles of the physical sciences and individual contributor experience with statistical methods. This book is about how engineers and scientists can be **more effective and more efficient** in the way they problem solve, design studies, collect data, interpret results and ultimately gain knowledge.

At the heart of nearly all engineering and scientific work lies a fundamental desire to understand *how and why* things work. Throughout this book, $Y = f(x)$ is used as shorthand for causal structure, not as a promise that a complete or closed-form model exists. Y represents outcome or dependent variable(s) and x is the input or independent variable(s). Outcomes are a function of underlying causes. Whether we are improving a manufacturing process, designing a new product, diagnosing a failure, reducing the cost of production, or understanding product performance in the hands of the customer, the objective is the same. The expression, $Y = f(x)$ is deceptively simple. While few would argue with it, far fewer understand it or consistently act on its implications. The challenge is not accepting that causes exist, but determining which causes matter, under what conditions, and to what extent. That challenge is compounded by variation; variation in materials, processes, designs, environments, measurements, human behavior, and corporate culture.

Understanding the amount of variability and its sources is central to analytical statistics. Variability is both expected and challenging to explicate. To find associations between x's and Y's, x's need to vary while the Y's are measured. For example, to determine if mold temperature is associated with polymer chain length and so dimensional shrinkage of an injection molded part, mold temperature will need to vary in the study (either naturally over time or through manipulation in a designed experiment); Analytical statistics aim to evaluate the data and to assist in partitioning and assigning such variation. At times variability may also limit discovery such as when measurement precision (gage variation when measuring the same characteristic multiple times) is so poor product variation cannot be assessed.

The difference between inexperienced and skilled statistical practitioners becomes evident in how they plan data acquisition. COV studies and DOE effectively organize or generate data to provide insight to hypotheses. Inexperienced users of statistics tend to ignore **the link between data and context** and, as a result, jump directly to analysis rather than appropriately planning to collect the *right* data in the first place. For example, most organizations gather tremendous amounts of data and then "torture the data into submission". The data analysts haphazardly run programs in the statistical software hoping to expose something. What statistical tool should be used to look at the data? How can the data be organized to *prove* my hypothesis is correct? While this historical/observational data may be useful to develop hypotheses, it is not useful for drawing conclusions about causal structure.

Data Source Matters

Figure 1.1 shows a continuum of data sources from anecdotal to designed experiments. The order reflects their **effectiveness** in understanding causality, not their convenience or efficiency. A practical rule is: use red-zone data (anecdotal, research, simulation, historical, observational) to generate hypotheses, and green-zone data (Sampling and DOE) to test those hypotheses and learn about causality.

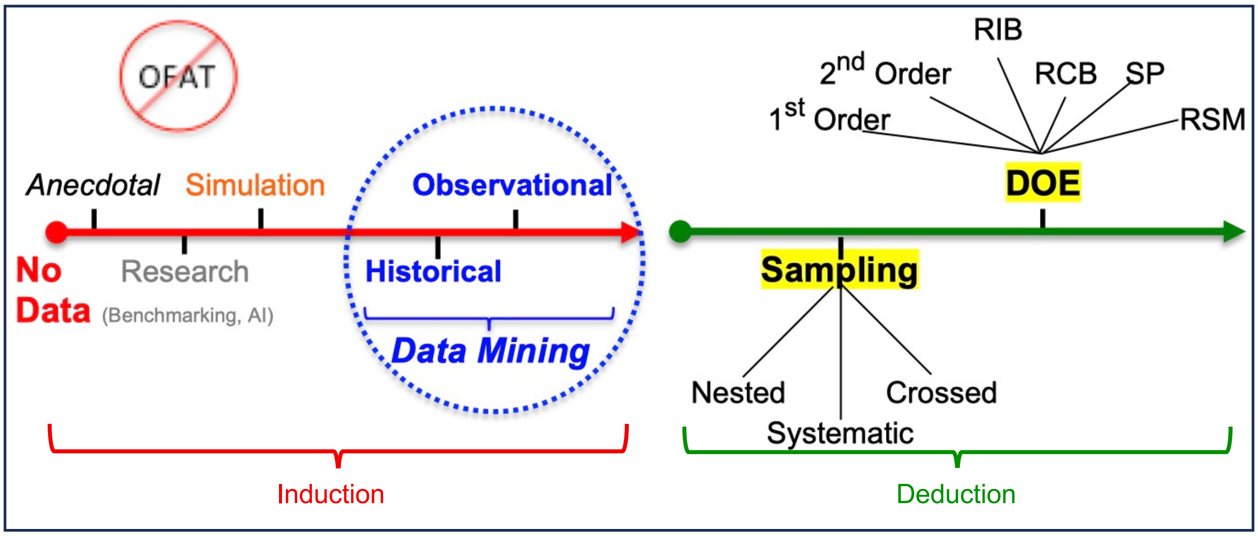


Figure 1.1: Data Source Continuum

The elements of the continuum are as follows. The red-zone of the continuum are data sources that are useful for hypothesis development.

- **Anecdotal**: another person, engineer, manager, peer, tells you a story of what they have experienced. This is often said with a tone of this is a fact, but more often, this is their hypothesis.
- **Research**: someone, somewhere else in a different inference space learned something and wrote it down. This also includes benchmarking, which is usually best served as a motivational activity and artificial intelligence search engines.
- **Simulation**: this one is tricky. If you are using an "off the shelf" simulation generator, the algorithm already exists. What data was used to create the algorithm? Likely not yours. Is it purely based on theory? How does it simulate real noise? Does it apply to your situation? If you are experimenting with a factor and that factor is not in the algorithm, it will conclude the factor is not significant. You risk the Type 2 error, that is abandon the factor for future considerations when it is actually useful.

- **Historical**: the good news, it is data from your processes. Unfortunately, it lacks context. We don't know what all of the variables were doing when the data was collected. We don't even know the measurement error.
- **Observational**: data being collected currently without plan. Data storage and adding sensors is relatively cheap. Is the data collected at the right location, on the right thing, with the right frequency, over the right time period?

In reality, there is a **huge gap between the red and the green-zones.** The green-zone of the continuum are the data sources that are used to provide insight to the hypotheses.

- **Sampling**: collecting samples as a function of hypotheses to specifically partition and assign sources of variation. Note, COV is a sampling strategy, not all sampling is COV.
- **DOE**: accelerate learning via manipulation of factors identified via hypotheses and sampling.

A practical rule: use red-zone data (anecdotal, historical, observational) to generate hypotheses, and green-zone data (sampling, DOE) to test them and learn about causality.

Execution and Discovery

Most organizations behave as if all work is execution. That belief causes more damage than any statistical mistake I have seen. Execution assumes the problem is well defined, the appropriate measures are understood, the causal structure is largely known, and uncertainty is limited around an expected outcome. Under those conditions, speed and efficiency are appropriate goals. Tools are selected, plans are followed, and results are measured against expectations. There is nothing wrong with execution, *when execution is appropriate*. Discovery exists when those assumptions fail. Discovery work begins with an incomplete understanding, ambiguous problem definitions, competing hypotheses, inadequate measurements and uncertainty that cannot be reduced by analysis alone. In discovery, the goal is not speed, the goal is **learning**. Discovery work should precede execution.

Learning requires iteration, prediction, data acquisition designed to discriminate between explanations, and a willingness to revise both questions, hypotheses and plans. Treating discovery as execution guarantees frustration. One of the most common, and costly, errors in engineering organizations is labeling discovery work as execution. This happens when management delineates timelines before understanding exists, early data is treated as confirmation, iteration is seen as unnecessary and costly rework, or experiments are expected to "prove" decisions already made. Under these conditions, engineers learn quickly what is rewarded; confidence over curiosity, speed over insight, compliance over improvement, short-term cost savings over long-term profitability. Learning slows, even as activity increases.

Iteration is unavoidable in discovery. Yet it is often interpreted as costly and time consuming because expectations were set incorrectly. If a project requires iteration, that is not evidence of poor planning. It is evidence that the system is not yet understood. A lack of iteration should be a warning sign. When a complex problem proceeds smoothly from definition to solution without revisiting assumptions, it usually means important variation has been ignored, key questions were never asked, or inconvenient data was filtered out.

Traditional problem-solving models assume an assignable cause[2]. They suggest the appropriate steps to take. Quality professionals enthusiastically decree their importance. It often seems like the application of the tools is the objective. In discovery, tools **probe**, **challenge**, and **expose ignorance**. Using quality metrics to judge discovery work is a category error. It punishes learning and rewards superficial certainty. Discovery does not end when a model fits, a p-value is small, or an experiment is "successful." Everything that follows in this book depends on recognizing whether you are operating in discovery or execution. It affects:

- how you define the problem,
- how you acquire data,
- when you sample,
- when you experiment,
- how you interpret results,
- and how you communicate with others.

Confusing the two is not a minor mistake. It is foundational. Discovery work places a cognitive burden on the engineer that execution work does not. You are asked to hold multiple hypotheses simultaneously, reason with incomplete and often contradictory information, make provisional decisions while resisting premature commitment, and explain uncertainty to others who may not tolerate it well.

Most organizations provide no infrastructure for this kind of thinking. Schedules, metrics, and reporting systems are designed for execution, not learning. Discovery may lead to execution when uncertainty has been reduced enough to justify confidence, the cost of being wrong is understood, and the next steps are robust to what is not yet known.

A more experienced user would, instead, start with an underlying set of questions, such as, what is the motivation for investigation? What are the phenomena of interest? How can the phenomena be quantified? Is the measurement system adequate? How much do I think I know about the phenomena? What is the basis for this knowledge? What are the factors (x's) affecting variation? What is the noise? From the set of questions, they would develop hypotheses and consider multiple ways the collection of data might provide insight to those hypotheses or in addition, try to find ways to provide evidence their hypotheses are wrong. Perhaps a statistical test would be useful, but other approaches might be more applicable, such as the use of graphical methods.

This shift in perspective from statistical technique to scientific investigation will likely change the way one approaches data collection and analysis. After creating a list of questions, skilled statistical thinkers discuss with their scientific/engineering collaborators the ways data might be collected to provide insight

[2] See Visualize Variation, page I-20

to their questions and, thus, what kinds of studies might be most useful. Together, they try to identify potential sources of variability and predict all possible outcomes of a study. This is a major reason why collaborating with statistical thinkers can be helpful, and also why the collaborative process works best when initiated early in an investigation. Of course, having engineers capable of integrating the scientific process and the statistical thinking together can be a huge advantage.

I suggest to start investigations by first determining where to work and the nature of the phenomena. Which set of x's provides greater opportunity for understanding the causal structure? Are the phenomena special or common[3]? Partitioning the x's, comparing and assessing leverage of each set of x's improves the efficiency of the study. Provided the study will be iterative,

> "*It is better to confound, than restrict*."
>
> *Ross*

Restricting factors in a study provides no opportunity to learn about their effects and constrains the inference space impacting confidence in extrapolating results. Typically, investigations start far from optimum and the initial work is to move in the direction of optimum. I suggest assume you know nothing. The study often proceeds by developing a first-order, linear model. Linear models work well as a first approximation or as a depiction of a general trend, especially when the amount of noise in the data makes it difficult to distinguish between linear and nonlinear relationships. The appropriateness of the model should be evaluated over the duration of the study. Keep in mind a good sampling plan, implemented well, can often allow simple methods of analysis to produce excellent results. Simple models help us to create order out of complexity, are more useful for prediction and are well suited for communication to others. Thoughtful data collection can greatly simplify analysis and make it more precise.

> "*Results of a well-planned experiment are often evident using simple graphical analysis. However, the world's best statistical analysis cannot rescue a poorly planned experimental program.*"[4]

When substantial effort (i.e., time and money) will be involved in collecting data, statistical issues may not be addressed in a question such as; What is the correct *sample size*? Sample size is seldom the right question for an analytical problem. More appropriately, is the sample **representative** of future considerations; over what changing conditions does the engineer want to observe and draw conclusions? Rather than focusing on a specific detail in the design of the experiment, someone with statistical experience is likely to step back and consider many aspects of data collection in the context of overall goals and may start by asking; What would be the possible outcomes of the sampling plan, and how would the data be interpreted? What could be done with this information? What if the hypotheses are wrong? How likely is this experiment representative of future conditions? How will this information

---

3 Deming, W. E., (1982) *Out of The Crisis,* Cambridge, MA: MIT Center for Advanced Engineering Study, Ch. 11
4 Hahn, Gerry, Doganaksoy, N. (2008) *The Role of Statistics in Business and Industry*, Wiley

increase understanding of the phenomena and the causal structure? What will the next iteration of study look like?

In trying to determine the relationship between x's and Y's, key issues involve:

- Hypotheses as to why the x's might affect the Y's,
- Understanding of the amount of change in Y's that is of engineering or scientific interest (i.e., practical significance),
- The way x and Y are measured (measurement uncertainty and discrimination),
- The extent to which the measurements represent the underlying causal relationships of x and Y (Have both varied enough during the study?),
- The ability to identify and account for the multitude of factors (perhaps confounded) that could affect the measurements, and
- Whether some of those factors might introduce systematic errors (bias) or act specially.

This process of data analysis often involves a multitude of outputs of statistical procedures, including many plots and graphs and a host of quantitative tables. These need to be interpreted. This is one reason I encourage engineers to develop a **personal User Guide**. In discovery, the User Guide serves as:

- a reminder of questions that should be asked before acting,
- a place to capture elements of iteration that assure due diligence,
- a record of what has challenged you in the past,
- advice on what to do given the "it depends" nature of investigation,
- and a guardrail against rushing into tools simply because they are available,
- Links between situation diagnostics and applicable tools.

The User Guide does not define *the* way to proceed. It helps you be thorough in deciding *your* next steps, given what you know at the time. As discovery progresses, and as understanding improves, the User Guide should evolve along with it. Discovery is iterative by nature. Your thinking infrastructure should be as well.

A note on optimization. We are continually trying to optimize our products performance, but optimization is not a steady state. I believe there are two reasons why we will always be *in the act of* optimizing (aka. continuous improvement). First, we never start with all possible variables, we always start with some subset. Second, as technologies, measurement systems and materials change, so changes the optimum.

Initial Questions That Motivate This Work

Engineering work often begins with deadlines. Someone proposes a solution, suggests an experiment or analyzes an existing data set. Only later, sometimes much later, does anyone stop to ask whether the **right questions** were ever articulated. This methodology is about treating questions as **first-class engineering products**, deserving the same care and scrutiny as designs, models, or experiments.

> *"The mere formulation of a problem is often far more essential than its solution, which may be merely a matter of mathematical or experimental skills. To raise new questions, new possibilities, to regard old problems from a new angle requires creative imagination and marks real advances in science."*
>
> *Albert Einstein*

In practice, engineers are handed complaints, symptoms, deadlines, or directives. "Reduce variation." "Improve yield." "Fix this process." "Make it cheaper." "Run a DOE." These are not questions. They provide motivation, almost always urgent, but they do not define what needs to be learned. The real work of discovery begins when those signals are translated into questions that can be investigated. Tools answer the questions they are asked, nothing more. If the question is vague, the answer will be ambiguous. If the question is mis-framed, the answer may be precise and useless. If the question is premature, the answer may mislead. No amount of analytical sophistication can compensate for a poorly formed question. There are stupid questions and I have asked some of those. Experienced engineers learn, sometimes the hard way, that progress accelerates when effort is invested **upstream**, in question formulation, rather than downstream, in analysis. Ironically, you need to be well versed in the tools and analysis to ask better questions.

Useful questions in discovery tend to share several characteristics, they:

- expose uncertainty rather than hide it
- lead to multiple plausible hypotheses
- suggest what could be measured to discriminate between those hypotheses
- indicate what data is necessary
- evolve as learning occurs

Early questions are often broad and imprecise. That is acceptable. What is not acceptable is treating them as fixed. A red flag in any discovery effort is a question that never changes. Every question implicitly defines what is in scope, what data will be considered relevant, whom should be asked, and what conclusions will be considered reasonable. When questions are inherited without critical thought, often from management, customers, or prior work, the scope is inherited as well, along with its biases. Questions that matter usually lead to **multiple hypotheses**, not one. For example: Is the issue driven by materials, process, environment, distribution, customer use, or measurement? Is the variation inherent, assignable, or induced by our own actions? Is this a stability, mean, variation, or a leverage problem? Good discovery work resists collapsing these possibilities too early. Thought maps and similar tools are useful here because they make competing hypotheses explicit and inspectable. They allow the investigator and teams to see not just *what* is being considered, but *what is being ignored*.

The methodology presented in this book emerged from a recurring set of questions I encountered repeatedly in practice and instruction. These questions are not academic; they shape decisions with real consequences. *What does it mean to apply statistics? What statistics are useful under what circumstances? What statistical technique should be used for different situations? How and where should* *data be collected? How should data be analyzed? How should results be interpreted? What does statistical significance actually mean? How does statistical significance relate to practical significance? How can data be used to help me understand how and why things work? How can I predict future performance?*

A common misconception is that statistical thinking is something applied *after* data exist. In reality, the most consequential statistical decisions are made **before any data are collected**. These decisions determine, what questions can be answered, how the data should be collected, what tools should be used for analysis, what conclusions are justified, and how confident one can be in extrapolating results.

Sigma Science: Merging Scientific Method and Statistical Thinking
In this book, Sigma Science refers to the deliberate integration of scientific method and statistical thinking as applied to engineering and scientific practice. Effective engineering and scientific progress depend on scientific method; hypothesis generation, prediction, data collection, analysis and interpretation. Learning occurs through iterative cycles of induction and deduction. Each cycle resulting in learnings, modified hypotheses, abandoned hypotheses and a new set of questions.

In its most potent form, Sigma Science focuses on increasing the **competency and critical thinking** of organizations; engineers, scientists, and business professionals alike. This is achieved by creatively merging:

- process and product knowledge,
- laws and principles of the physical sciences,
- individual contributor experience and intuition,
- collaboration with others,
- and appropriate application of statistical methods.

The power of Sigma Science is not due to the statistical toolbox, nor in the indiscriminate application of engineering resources. It comes from merging **statistical thinking with scientific method**.

> *"…Use of data requires knowledge about different sources of variation."* [5]

Statistical thinking emphasizes understanding **variation**; its sources, structure, and implications. Without this understanding, analysis becomes mechanical and conclusions become precarious. Statistical thinking does not replace this process. It strengthens it by providing a disciplined way to analyze, understand and interpret the data.

[5] Deming, 1995 The New Economics

Enumerative and Analytical Perspectives

It was Dr. W. Edwards Deming who first distinguished between enumerative and analytical studies[6]. Enumerative studies focus on the judgment, explanation, evaluation and description of what **already exists**. I sometimes refer to them as autopsy statistics, useful for describing, but limited for prediction. Distributional analysis, confidence intervals, regression models, statistical tests and summary statistics are all examples of enumerative statistics. In contrast, analytical studies aim to understand causality in order to predict what may happen under future conditions. I will be focusing on the analytical nature of investigation in this book.

I want to highlight two key differences between the types of studies: For the question of: How do we arrive at a more representative study? The enumerative thinker suggests a random sample (to mitigate bias) and to increase the sample size. An analytical thinker will suggest using hypotheses to guide sampling to make it more representative. For prediction, the enumerative thinker uses probability. For example, if you want to predict the force an object will exert upon being dropped, drop the object many times and record the force. Create a distribution of forces and state the next force will fall within this distribution. An analytical thinker develops a mathematical model, F = ma, measure the acceleration and mass and use the formula to predict the next force. Sampling plays a role in both contexts, but the intent differs. In all cases, sampling implies that not everything will be measured, only a subset. The selection of that subset is not a trivial detail; it provides the context for everything that follows. What questions can be answered, what analyses are appropriate, what actions are justified, and how confidently results can be extrapolated are entirely dependent on **how the data are acquired**.

The Dictionary

To understand Y = *f*(x), one must first identify the x's. I often illustrate this with a simple analogy.
I ask someone to think of a word in a common English dictionary, roughly 60,000 words, an exaggerated analogy of the number of variables in the system you are trying to understand. My task is to guess the word. This is a zero-knowledge sampling idea. That is, I have no hypotheses as to what the individual might be thinking about. I guess some singular words, elephant, lubricity, etc. I stop and ask; What do you think of my sampling plan? The response is, not good, we'll be here all day. Random guessing is inefficient and ineffective. A better strategy is to ask questions that systematically reduce the search space. For example, you might ask a question with a binomial response: *Is the word in the first half or the second half of the dictionary?* With a single question, approximately 30,000 words are eliminated. By repeating this strategy, each time dividing the remaining possibilities in half, you can identify the word in fewer than sixteen questions. Remarkably, this can be done **without any prior knowledge** of the person, the word, or its meaning.

Now consider what happens when prior knowledge is introduced. If you have hypotheses about the type of word the person might choose, you can partition the dictionary into smaller, more informative subsets.

[6] Deming, W. Edwards (June 1942). "On a Classification of the Problems of Statistical Inference". *Journal of the American Statistical Association*. 37 (218): 173–185

Each question becomes more efficient, and fewer questions are required. The more specific the hypothesis, the more obvious the sampling to provide insight to that hypothesis. This is the conceptual basis for **directed sampling**. Now consider a different situation: the word is *not* in the dictionary. No amount of questioning, data collection, partitioning, or analysis will identify it.

This is the dictionary illustration that I will refer to often in the book. In engineering and science, the "dictionary" is the set of factors, the x's, that could affect the outcome of interest. If a variable is not identified, it is impossible to be studied. If it is not studied, it cannot be analyzed. If it cannot be analyzed, it becomes a hidden source of confounding. The longer the problem has been around the more likely it is due to factors never studied. The objective of early-stage thinking and mapping is not to create the *best* dictionary, but the **most inclusive dictionary possible at the time**, knowing that it will evolve as learning occurs. Do not leave words out due to your preconceived ideas.

Why is it Important to Understand Data Acquisition

Inexperienced users of statistics often jump directly to analysis without understanding context. Large volumes of historical or observational data are examined with the hope that insight will emerge. Statistical software is run repeatedly until something appears interesting. This can be quite useful if the intent is to **develop** hypotheses. This is the basis for "data mining". Observations from these studies implore the investigator to explain what they see in the data. These explanations are hypotheses. While such data may be useful for generating hypotheses, they are rarely sufficient for drawing conclusions about causal structure.

More experienced practitioners begin elsewhere. They ask what do they want to know, why they want to know it, how the phenomena can be quantified, what should be measured and is the measurement system is adequate, what sources of variation might be affecting variation, and how will I get data to provide insight into my multiple hypotheses. Only after these questions are considered does it make sense to decide how data should be collected.

> *"Why we take data:*
> a) *To obtain quantitative information*
> b) *To obtain a causal explanation of observed phenomena"*
>
> *Shewhart Economic Control of Quality of Manufactured Product (p. 55)*

Prediction, Iteration, and Learning

All meaningful engineering and scientific work is iterative. Before collecting data, thoughtful practitioners make predictions. They predict what data they expect to observe, they enumerate **all** plausible outcomes, and they consider what actions they would take under each possible outcome. In doing this, they develop multiple options for data collection. Each differing in what will be estimable, confounded or restricted in the study and each having different resource requirements. They compare and contrast the multiple plans and then pick one knowing there will be iteration. I use a simple acronym, KISS, Keep It Simple and

Sequential. Prediction forces assumptions into the open. It provides a reference point for reflection. Learning occurs not only when predictions are confirmed, but when they are violated.

> *"We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances."* [7]*,*

> *"Pluralitas non est ponenda sine necessitate."* [8]

All things being equal, the simplest explanation is the best. This guideline has been included in operating procedures across many fields. This principle of economy can be a useful guide. Do not try to answer all questions, provide insight to all hypotheses and get a prediction model in one plan. The likelihood your first plan captures everything you will need to know is close to zero.

> "*The first sampling plan is intended to help design a better sampling plan*."
>
> *Ross*

It is recommended to start with simple approaches and only add complexity as needed. When building mathematical models, start with first-order terms and add higher order as necessary. This follows the Effect Sparsity[9] and Effect Hierarchy[10] Principles and also implied by Taylor series order. Interactions among explanatory x's, nonlinear models, missing data, confounding, sampling biases, measurement error and so on, can all complicate the ability to create a simple useful model.

Variation and Causality

All measurements exhibit variation, everything varies. Repeating a measurement under nominally identical conditions will not produce identical results. Understanding what creates variation, and how it is structured, is central to analytical statistics. To learn about causal relationships, x's must vary while Y's are measured. We look for the relationships between the changing x's and Y's. In order for a factor effect to be quantified, it must vary in the study. The more the factor varies, the easier it is to detect its effect. Conversely, if the factor does not vary in the study, there is no way to determine that factor's effect. Correlation may be evidence of, but does not by itself imply causation. This is one of the most serious and common errors of human reasoning[11]. To reasonably infer causality, correlation is necessary but not sufficient. The case for causation is strengthened when variables are deliberately manipulated in a designed experiment or when structured variation is introduced through a well-designed analytical study, and when rational, logical hypotheses support these observed relationships. Correlation without

---

[7] Sir Isaac Newton (1687) *Principia*

[8] Occam's razor, William of Ockham, 14th century

[9] Box, G.E.P.; Hunter, J.S.; Hunter, W.G. (2005*). Statistics for Experimenters: Design, Innovation, and Discovery*. Wiley. p. 208.

[10] Wu, Jeff and Michael Hamada (2000) *Experiments: Planning, Analysis, and Parameter Design Optimization*. New York: Wiley. p. 112.

[11] A couple of others: a temporal bias, when an event takes place, we look to what immediately preceded it and only getting data on factors we suspect matter, never getting data to support the factors that we are not suspicious of.

explanation is pattern recognition, not understanding. Without hypotheses grounded in physical, chemical, or biological mechanisms, correlation remains an observation, not an explanation.

Visualize Variation

Figure 1.2 includes a "picture" of variation. It displays to fundamental belief that variation exists in everything. Although a true normal distribution is theoretical and does not exist in reality, using it for data interpretation is useful. There are two models associated with the distribution. Shewhart calls data "under the curve" **Unassignable**, chance or random variation. Deming calls this variation **Common Cause**, a function of the many x's varying in the system. The X, Shewhart designates **Assignable** and Deming calls this **Special Cause** variation. The models are similar, but not identical.

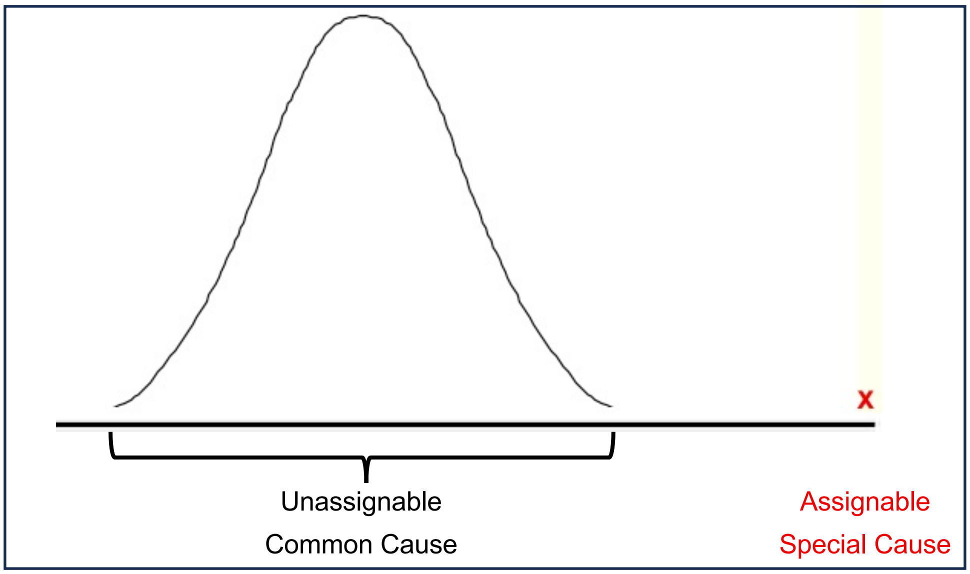


Figure 1.2: The Normal Distribution with an *Outlier*

Special causes are rare by definition. Since assigning them as special is time dependent, the appropriate reaction can be challenging. Taking common cause like action (e.g., change the process or design) assuming a special cause is usually ineffective and often increases variation[12]. Reacting to special as if it were common is also inefficient. Changing the process or design every time a defect is found can be expensive, disruptive and may not reduce variation anyway.

> *"There are common causes associated with the system (and also potentially special causes). It is unfair, wrong, and counter-productive to blame the individual for the failings of the system. Sure we don't like mistakes, complaints from customers, accidents —but if we weigh in at them without understanding, then we make things worse."*
>
> *W. Edwards Deming*

[12] See Deming funnel experiments, Out of the Crisis, p. 327

Guiding Perspective

Several principles guide the methodology presented in this book. Start with questions rather than tools. Make assumptions explicit. Identify variables before collecting data. Prefer representativeness over sample size. Use prediction to expose bias. Accept iteration as fundamental. Favor simple explanations when sufficient. Do not try to learn everything in the first study. The first iteration is designed to inform the next.

The Four Tools

The methodology described above is operationalized through four integrated tools. I intentionally limit the framework to four to avoid overwhelming learners and to support effective knowledge transfer. Each will be discussed further in separate chapters.

> First is **critical thinking**. The disciplined, independent, and deliberate process of framing questions, articulating assumptions, generating hypotheses, making predictions, and reflecting on outcomes to understand causality. Critical thinking is internal, iterative, and inherently personal. Because it is difficult to observe, it must be externalized to be examined, shared and improved. I use Thought maps to visualize critical thinking.
>
> Second is **mapping**. Mapping engages the visual sense, which plays a dominant role in human cognition. I distinguish between Thought maps and Process/product maps. Thought maps externalize the critical thinking process, while Process/product maps are visual aids to help the investigator be thorough in identifying the Y's, y's, x's and purpose, by location. Thought mapping helps manage complexity and reduce cognitive load. Process/product mapping supports the investigator in developing a comprehensive "dictionary". These maps provide contextual framework necessary for interpreting results from the study by facilitating identification of what the x's are doing in relation to the data collection strategy.
>
> Third is **directed sampling**, which I will refer to as COV studies. COV is used to understand what already exists, to assess stability, quantify sources of variation, determine leverage, and decide how an investigation should proceed. Sampling plans are designed as a function of hypotheses and questions, not convenience.
>
> Fourth is **designed experimentation**, which accelerates learning by intentionally varying factors. When used properly, experimentation provides insight into causal relationships that cannot be obtained through observation alone. Effective experiments are designed with iteration, inference space, and future learning in mind.

These tools are not linear steps. They are used iteratively, and each informs the other.

> *"While techniques are important…knowing when to use them and why to use them is more important"*
>
> *Tukey (1954)*

How Part I of this book proceeds

This chapter establishes the methodological foundation of Sigma Science. The remaining chapters in Part I develop the foundations of Sigma Science:

Chapter 2 treats **critical thinking**, not as an abstract idea, but as a disciplined practice for engineering and scientific work. Before variables are mapped, before data are collected, and before statistical tools are applied, someone must decide what questions are worth asking and why. This is where learning begins.

Chapter 3 discusses **prediction** as a strategy, a guide to action and self-assessment of knowledge.

Chapter 4 develops the different types of **mapping** to support learning and engage the visual sense.

Chapter 5 **links** critical thinking to data collection strategies.

Chapter 6 and 7 elaborate on **COV** and **DOE.**

Chapter 8 exposes the impact of **noise** on experimentation.

Chapter 9 describes **analysis** and interpretation to gain insight into hypotheses.

Chapter 10 introduces the **User Guide** as a self-reflective compendium of knowledge.

Chapter 11 contains some of my **reflections** and bias.

Chapter 2
Critical Thinking as a Discipline

What Is Critical Thinking?

Critical thinking, as used throughout this book, is not a personality trait, not intelligence, and not simply being skeptical. It is a *disciplined practice*. I define critical thinking as:

> *The intellectually disciplined, independent, and deliberate process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, evaluating, critiquing, and predicting information gathered from observation, experience, reflection, reasoning, collaboration, and data collection, as a guide to understanding causality.*

Several aspects of this definition are intentional. First, critical thinking is **independent**. Collaboration is valuable, and essential for mitigating your own bias While collaboration informs critical thinking, critical thinking itself requires individual ownership of questions, assumptions, hypotheses, and reasoning. Critical thinking itself is not a team sport.

Second, critical thinking is **deliberate and disciplined**. Humans are naturally curious and naturally ask questions, but undisciplined questioning often leads to scattered effort, fixation on the wrong issues, or endless iteration without learning. Discipline is what turns curiosity into progress.

> *"Critical thinking calls for a persistent effort to examine any belief or supposed form of knowledge in the light of the evidence that supports it and the further conclusions to which it tends."*[13]

Third, critical thinking is **meta-cognitive**. It includes reflection on *how* one is thinking, not just *what* one is thinking. This reflective component is what allows improvement over time.

> *"Critical Thinking is a deliberate meta-cognitive* (*thinking about thinking*) *and cognitive (*thinking*) act whereby a person reflects on the quality of the reasoning process simultaneously while reasoning to a conclusion. The thinker has two equally important goals: coming to a solution and improving the way she or he reasons."*
>
> *Moore, 2006*

Finally, while not required of critical thinking in general, the guiding principle of critical thinking in engineering and science is **understanding causality**, how and why things work, not merely describing patterns or producing answers.

[13] Edward M. Glaser, *An Experiment in the Development of Critical Thinking*, Teacher's College, Columbia University, 1941

Why Critical Thinking Matters in Engineering and Science

Most engineering and scientific studies fail to be as informative as they could be, not because of poor tools, insufficient data, or lack of intelligence, but because the thinking that precedes data collection is underdeveloped or implicit. We, collectively, fall back on explanations for the phenomena *after* the phenomena occurs and often implement solutions to those explanations without testing. We fail to realize those are hypotheses that need to be evaluated before solutions are implemented. Common failure modes include:

- collecting data without knowing what will be done with it,
- jumping directly to data analysis with existing data,
- insufficient data sets,
- unknown measurement errors,
- mistaking statistically significant results for meaningful ones,
- or discovering too late the study cannot answer the questions that motivated it.

Here is an example. I recall working a project on a chemical process to improve yields that included a batch reactor. There was an enormous amount of historical data. There was pressure from management to use the historical data, and to get a quick solution rather than designing an experiment or conducting a component of variation study. The historical data reflected one data point per batch for each of the variables (e.g., there was one thermocouple in the reactor, one measure of chemistry from one sample of the batch). As it turns out, the problem was within batch. It was impossible to estimate the within batch variation. Experiments were run on factors changing batch-to-batch. Variation was created in the experiment and factor significance was determined, but it did not have any effect on the problem at hand.

Critical thinking addresses these failures *before* data are collected or analyzed. A thoughtfully designed study often allows simple analysis to yield powerful insight. Conversely, no amount of sophisticated analysis can rescue a poorly conceived study. This is why critical thinking is not an abstract philosophical exercise in this book. It is a **practical engineering skill** that directly affects the effectiveness and efficiency of studies.

Critical Thinking and the Scientific Method

Engineering and scientific progress has always relied on the scientific method[14]: observation, hypothesis generation, testing, and refinement. Learning occurs through **iterative cycles of induction and deduction**.

> *"Once you have learned to ask questions - relevant and appropriate and substantial questions - you have learned how to learn and no one can keep you from learning whatever you want or need to know."*
>
> *Neil Postman, Teaching as a Subversive Activity, 1961*

[14] Invented by Sir Francis Bacon, Novum Organum, 1620

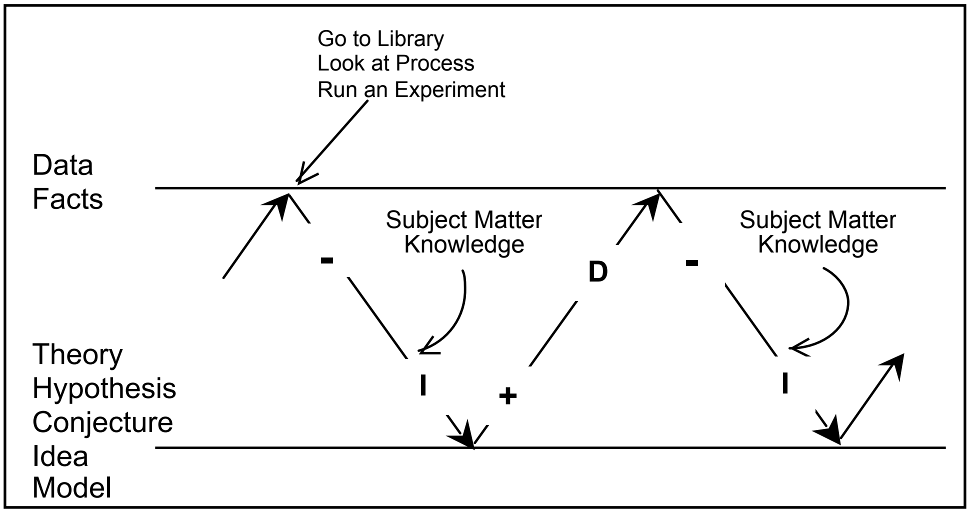


Figure 2.1: Diagrammatic Representation of The Iterative Learning Process[15]

Critical thinking is what makes those cycles productive. Observation without reflection leads to data analysis without understanding. Hypotheses without discipline lead to confirmation bias. Testing without prediction leads to post-hoc rationalization.

The purpose of critical thinking is not to be "right," but to **create a structure for learning**. Being wrong, when done thoughtfully, is often the fastest path to understanding.

> *"You have to have a willingness to repeatedly fail."*
>
> *Jeff Bezos*

This perspective aligns with Box's model (Figure 2.1) of the scientific method, which emphasizes iteration and learning rather than final answers. Models are provisional. Understanding evolves.

What Critical Thinking Looks Like (and What It Does Not)

When someone is thinking critically in an engineering or scientific context, several behaviors are consistently present:

- Questions are documented, not just virtual.
- Assumptions are made explicit.
- Hypotheses are articulated before data are collected.
- Predictions are made prior to data collection strategy selection.
- Alternative explanations are actively considered.
- Outcomes are reflected upon, regardless of whether they confirm expectations.

In contrast, a lack of critical thinking often appears as fixation on a single explanation early, reliance on tools to "find something" in the data, retroactive justification of results, implementing solutions before

[15] Box, G.E.P, (1999) Statistics as a Catalyst to Learning by Sigma Science Part II — A Discussion Journal of Quality Technology 31 (1)

understanding the problem, or confusion about what was actually learned. Critical thinking does not eliminate bias, but it **makes bias visible**.

> *Suppose you are holding an apple in your hand. Now you let the apple go. You observe the result and say, "The apple falls." That is a description. A prediction might have been the statement "The apple will fall if I open my hand." Both are valuable, and both can be correct. But an explanation is something more: It includes not only descriptions and predictions but also counterfactual conjectures like "Any such object would fall," plus the additional clause "because of the force of gravity" or "because of the curvature of space-time" or whatever. That is a causal explanation: "The apple would not have fallen but for the force of gravity. That is critical thinking."*[16]

Questions

Not all questions advance understanding. The discipline lies not in asking more questions, but in asking more informative ones. I have accumulated some advice for myself regarding the questioning process. First, understand what the purpose is for your question. Some Reasons for Asking Questions:

- Improve understanding, gain insight, get clarification
- Probe assumptions
- Evaluate, assess and get feedback (instructor)
- Explore implications/consequences
- Drive behavior
- Challenge knowledge
- Stimulate creativity (Elicit alternative perspectives)
- Establish rapport and inclusiveness
- Establish scope/boundaries
- Show interest
- Make a point (rhetorical)

Focus on the Questioning, not just any questions, but intelligent ones. How to improve the quality of questions:

1. Consider whom you are asking the question to. Is that person the correct one to answer the question? Do they have insight or perspective that will add to my current knowledge? Are they at the appropriate level in the organization to provide such insight?
2. Ask the question at a reasonable time. Not when there is an emergency or when personal issues may distract from the question
3. Consider how to phrase the question. How did you create this scrap? Vs. What do think contributes to the problem?

My advice is to predict, just before you ask, all possible answers to the question you are about to ask. What would you say if you asked yourself the question? Will the answer be useful? If not, modify the question accordingly or don't ask the question.

[16] Noam Chomsky, Ian Roberts and Jeffrey Watumull, "The False Promise of ChatGPT", New York Times, March 8, 2023

Externalizing Thinking

Critical thinking happens internally, but improvement requires externalization. When thinking remains internal: it cannot be reviewed, it cannot be challenged, and it cannot be improved efficiently. Externalizing thinking, through sketches, diagrams, notes, flowcharts or check sheets, creates an artifact that can be examined and refined. This does not replace thinking; it enhances it. Externalized thinking slows premature convergence, exposes hidden assumptions, supports parallel lines of inquiry, and allows others to engage with the reasoning rather than the conclusion. The goal is not to produce a perfect diagram. The goal is to make your current understanding explicit enough to be examined. This is the intent of the Thought map.

Thought maps (see chapter 4) do not start as documentation for others. They are working representations of how you think, independently. Eventually they can be used to collaborate and share ideas with others. They evolve. They are incomplete. And they are owned by the individual.

Prediction and Reflection

A central element of disciplined critical thinking is **prediction**. As a word of caution, a prediction is not a restatement of the hypothesis. Predictions are the data (specific values of the metrics being considered), all possible outcomes and actions for each possible outcome. Prediction prepares the mind to learn. As Pasteur observed:

> "*Dans les champs de l'observation le hasard ne favorise que les esprits prepares"*
>
> *Louis Pasteur, 1854*

> Or roughly, Chance favors the prepared mind.

Reflection completes the cycle. After data are collected. How did the actual data compare to the predicted data? The delta between predicted and actual data is what I call the Ross Metric. How and why did reality differ from expectation? What does that difference say about my understanding? What new or modified hypotheses are relevant?? Learning occurs in the **gap between prediction and reality**. I discuss prediction further in the next chapter (3).

I summarize the project investigation process with a model in Figure 2.2. The process begins with some motivation. The investigator performs situation diagnostics which culminate in a set of hypotheses. Data collection plans are developed to provide insight to those hypotheses. Predictions are used to help choose the appropriate plan. Data is collected, analyzed and interpreted through the lens of the engineer. Learnings are documented and implemented, new questions and hypotheses are uncovered, some hypotheses are archived and some are modified.

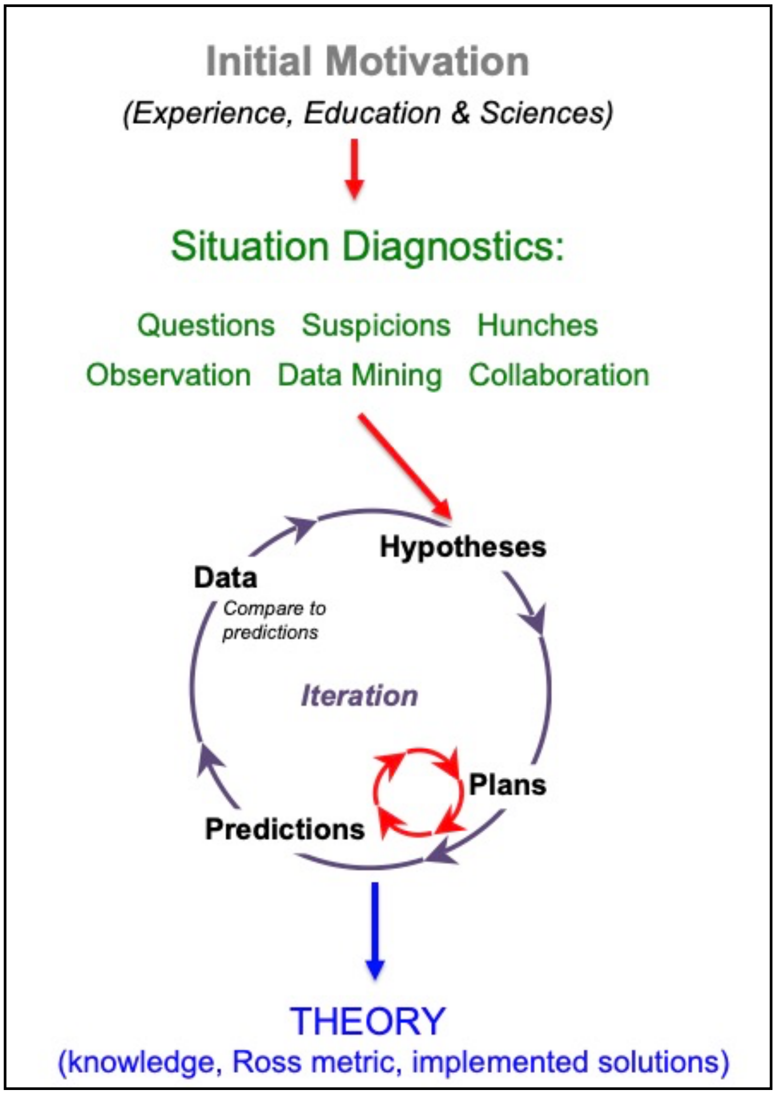


Figure 2.2: My Model for Project Investigation

The cycle, similar to Deming's Plan, Do, Study, Act (PDSA), or the Shewhart Cycle, iterates as questions are answered and new questions arise.

Chapter 3
Prediction as a Strategy

One of the most powerful, and underutilized, mental tools in engineering and scientific work is **prediction**. Prediction is often treated as a byproduct of analysis, something that happens after a model is built or a study is completed. In this methodology, prediction plays a very different role. Prediction is a precondition for learning. In analytic work, the goal is not to describe what has already happened. The goal is to learn enough about a system to **anticipate what will happen next**. Before data is collected, before COV, before an experiment is run, prediction requires engineers to answer a simple but revealing questions: *What do we expect to see in the data? What are all possible outcomes? What will I do for each possible outcome?* Prediction forces clarity of thinking. It exposes assumptions. It reveals what is known, what is believed, and what is merely hoped. Prediction is not speculation. A prediction is a **conditional statement**, grounded in explicit hypotheses and subject-matter knowledge. Predictions are an honest, although biased, assessment of knowledge and uncertainty.

Why Prediction Matters

Many investigations encounter difficulty not because the data are confusing, but because expectations were never met. When results differ from what was hoped or assumed, engineers often enter an excuse-and-blame cycle, questioning the data, the measurement system, the operator, or the method, rather than learning from the discrepancy. Prediction changes this dynamic. When all reasonable outcomes are predicted *before* data are collected, one of those outcomes will be correct. This prevents the post-hoc explanation game. The objective is not to be right, but to be **prepared**.

What Engineers Are Asked to Predict

When I ask engineers to predict, I am not asking for a single number or a narrow forecast. I expect prediction at multiple levels. First, engineers should predict the **actual data** they expect to get from their proposed data collection strategy, whether that strategy is a sampling plan or a designed experiment. For example, with a designed experiment there are treatment combinations that specify factor levels for a run of the experiment. If these factors are set at these levels what should be the value for the measured response variables? This prediction forces reflection on plausibility, measurement resolution, current understanding of the system and reflects the inherent bias of the engineer. In addition, this prediction is used to estimate the Ross metric, a quantified measure of your knowledge.

Second, engineers should predict **all possible outcomes**. Not just the ones the engineer is hoping for. The engineer should anticipate, if this is the result what will I do with that insight? Why might I get this outcome? How would I modify my hypotheses for each instance?

Third, identify what course of action would be taken for each possible outcome. This ensures the study is purposeful and that results will lead to decisions and actions rather than debate. For **each predicted outcome**, I expect engineers to answer this question: *If this occurs, what will I do next?* This may involve revising hypotheses, redesigning the sampling plan, running a follow-on experiment, or stopping entirely. If your course of action is the same regardless of the data, there is no need for the data. Prediction without a corresponding course of action is incomplete. Analytic work is not finished when results are obtained. It is finished when decisions are informed.

The following example is intentionally detailed. Its purpose is not to teach analysis techniques, but to illustrate how prediction structures thinking before data are collected. I will discuss the detail of analysis in Part II of the book. The project motivation is to understand customer issues with a refrigerator. It starts with a Thought map including motivation, questions, observations, and hypotheses leading to data collection opportunities.

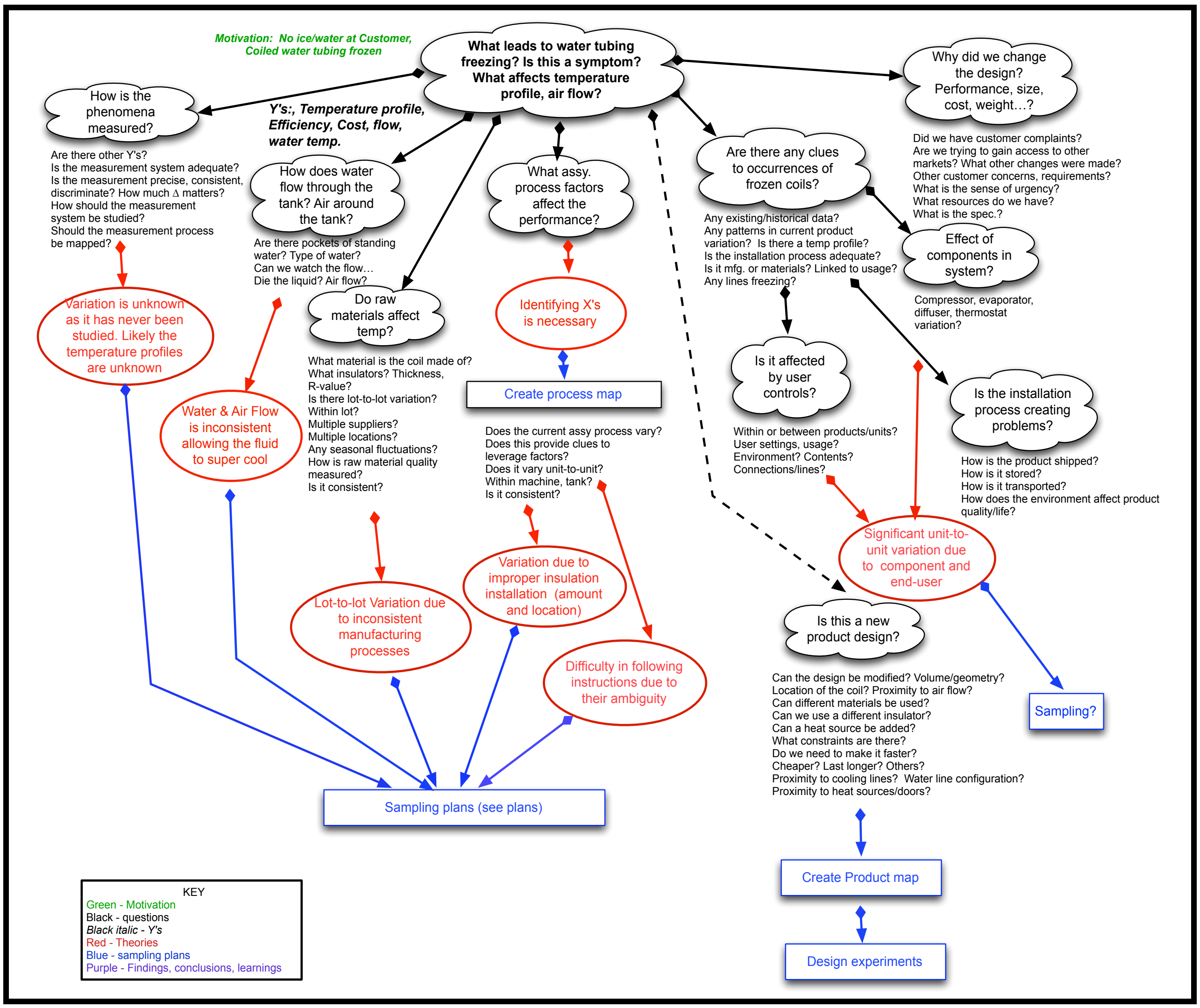


Figure 3.1: Thought map

Figures 3.2 & 3.3 are just two of multiple potential sampling plans. It is always recommended to design multiple plans. Each plan will differ in the potential for knowledge to be gained. What the plan can estimate, what is confounded, what is not in the study. That potential knowledge gain is contrasted with the resources required to execute the plan. Ultimately, one of the proposed plans is carried out.

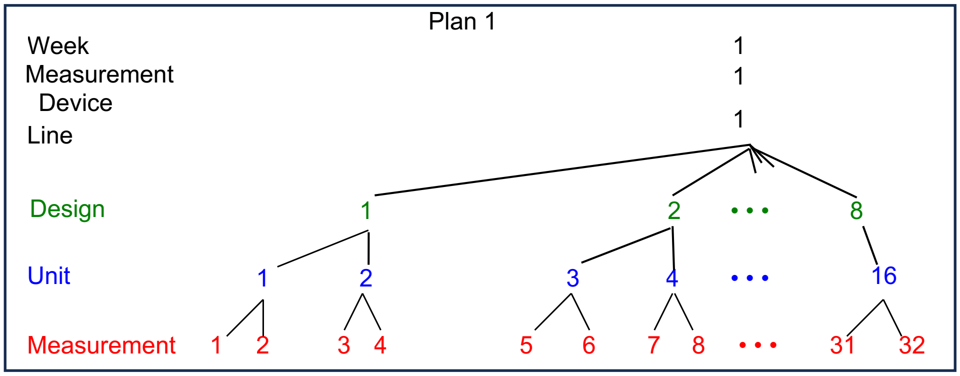


Figure 3.2: Sampling Plan 1

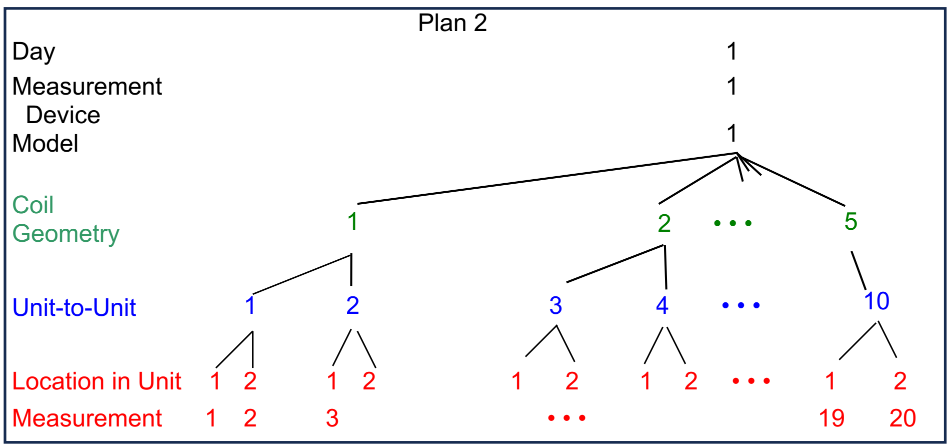


Figure 3.3: Sampling Plan 2

I will continue using plan 1. Three hypotheses, from the Thought map, associated with each "layer" in the sampling tree. This layer is correlated to x's on the product map which are color coded to match the plan:

Hypothesis 1: There is variation in the measurement system due to equipment precision, and connections to the thermocouple(s). Measurement system variation is unknown as it has never been studied. Designated **Measurement** for this illustration.

Hypothesis 2: There is variation in assembly due to ambiguous assembly instructions and supplied components due to supplier manufacturing variation (These confounded in this study). The air flow and water circulation are dependent on consistent performance of the components and the correct assembly of those components. Designated **Unit** for this illustration.

Hypothesis 3: There is variation of the proximity of the coiled tubing with respect to super cooled air due to a design flaw. The closer to super cooled air, the more likely freezing will occur. Designated **Design** for this illustration.

A change in 0.5 units of the measurement scale is of engineering interest. This is practical significance.

Predicting the data compels you to think through the plan. Is it reasonable? Are all of the combinations possible? It provides an opportunity to *practice* the data analysis and go through the possible *what-if* scenarios: outcomes and subsequent actions. Also, it provides a method for assessing your current state of knowledge as mentioned earlier, the Ross Metric (∆Actual – predicted). Predictions are a function of current understanding and biases as to factor effects. Here is an excerpt of predicted data:

Design	Unit	Measurement	Predicted Y1	Actual Y1
1	1	1	32.5	
1	1	2	33	
1	2	3	31.5	
1	2	4	32	
2	3	5	35.5	
2	3	6	36	
2	4	7	34	

Table 3.1: Subset of Predicted Data Table

Analysis can be performed with the predicted data. This enables me to plan how the data will be analyzed before I acquire it. I will show one sequence of each possible scenario. The objective is not to be *right*, but be thoughtful. While I won't redo predicted data, I will anticipate all possible outcomes and subsequent actions.

I will first take a **practical** look at the data without any summary statistics. Is there enough variation in the response to justify further analysis? In this case, the data must vary greater than the 0.5 practical significance. If there is not enough variation, it is either due to insufficient measurement discrimination or the inference space was too small. That is, the factors that are causing the variation did not vary enough in the study. Since the problem has already been identified with the current measurement system, it seems reasonable it is not a measurement discrimination issue. Going forward, I will need to expand the inference space allowing more time for variables to vary and for additional variables to be included in the study. If there is practically significant variation, I will look for any obvious deviations from my predictions. If there are, consider why? How confident am I with my predictions? Can I explain why the actual data is different from my predictions?

I will then take a **graphical** look at the data. If there is an obvious pattern in the data, I will be most suspicious of the x's associated with that pattern. For systematic plans, color/symbol coding the data can be helpful for identifying patterns.

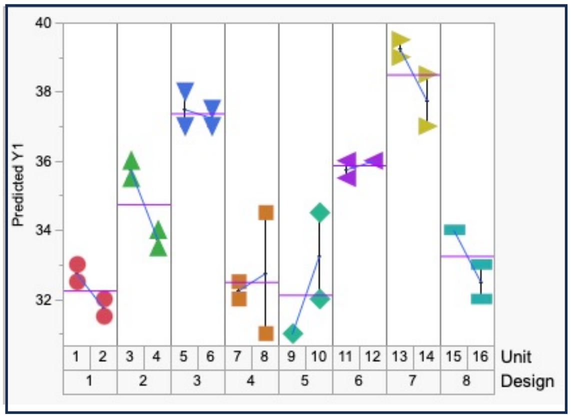


Figure 3.4: Variability Chart of Predicted Data (Marked by Design)

Given this plan, I will analyze each layer in order, starting with the bottom layer.

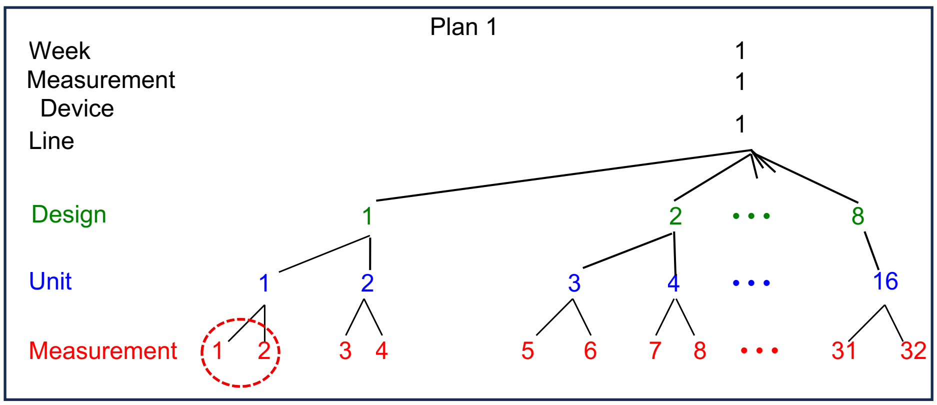


Figure 3.5: Sampling Plan with Subgroup Circled

With this first subgrouping strategy, I will be comparing the sources of variation associated with Hypothesis 1 to the other sources in the study (both hypothesis 2 & 3). Prior to comparison, I want to establish the basis for comparison is stable. A range chart will provide this insight.

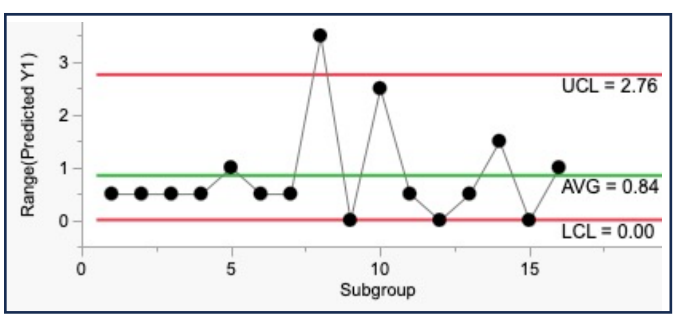


Figure 3.6: A Range Chart

If the range chart indicates inconsistency (as shown), I will investigate “special cause” like in the measurement system. I will review my notes to determine if anything unusual happened at that point in time:

1. Was the wrong number entered into the data table?

2. Was the time between measurements consistent?
3. Did the thermocouple disconnect?
4. Were there other thermocouple connection issues (e.g., contamination)?

I need to understand the possible causes, compare to my hypotheses and possibly change the measurement procedure accordingly. If the range chart indicates consistency, I can evaluate the measurement system for discrimination. If the discrimination is unacceptable, I will review the measurement system to ensure I am not inadvertently rounding the outputs. If still inadequate I will seek out alternative measurement systems or develop alternative Y's. If the discrimination is acceptable, I'll quantify the measurement component:

$$\hat{\sigma}_M^2 = \left(\frac{\overline{R_M}}{d_2}\right)^2$$

I will compare the estimate to the amount of variation that is important to me (of practical importance, 0.5 in this case).

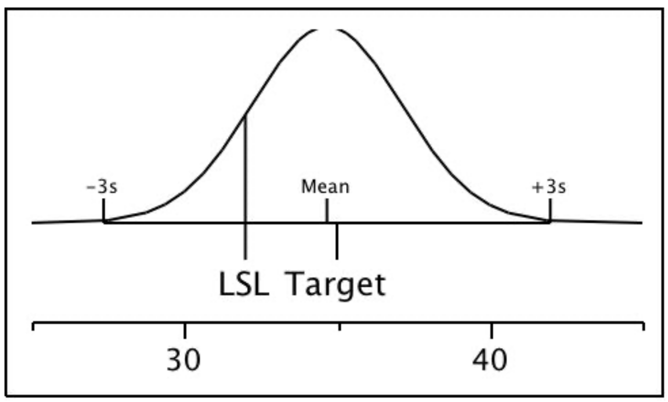


Figure 3.7: Visual Comparison of Measurement Error to Specifications

If this variation is large, I will need to understand how to reduce the variation in the measurement system *commonly*. There are no specific causes. Possibilities include:

1. Change calibration procedures and frequency
2. Investigate alternate measurement systems
3. Find other y's to measure
4. Change the location of measurement
5. Investigate thermocouple wire lengths

If the variation is small, I will continue my analysis by comparing the measurement variation to the other components of variation in the study (represented by Unit and Design layers of the tree).

If the y-bar chart is in-control, I will focus on improving the measurement system as it is the largest source of variation in the study. I might map the measurement process, look for factors that might affect variation and even run experiments to identify significant factors in the measurement system. If the y-bar chart is out-of-control, by a large margin since the basis for comparison is the measurement system, I will conclude the measurement system is adequate to understand sources of variation captured in the study.

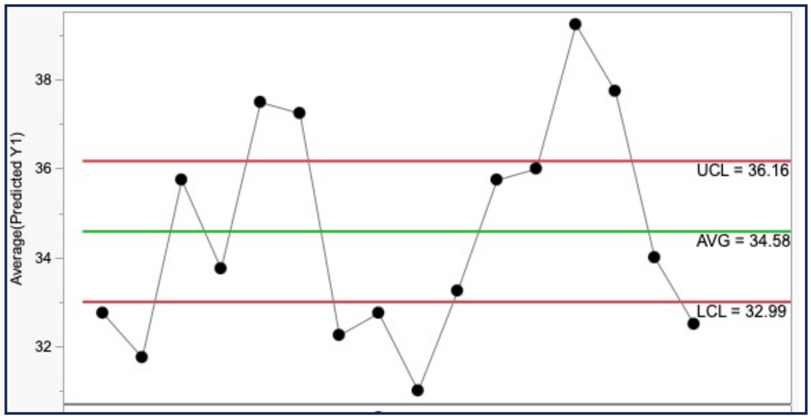


Figure 3.8: Average Chart Depicting Between Subgroup Sources of Variation are > Within

I will first average the two measurements, thereby reducing the variation due to measurement layer. Then re-subgroup those averages to determine if the Unit-to-Unit variation is consistent and compare it to the Design layer. This is often referred to as "rolling up the tree".

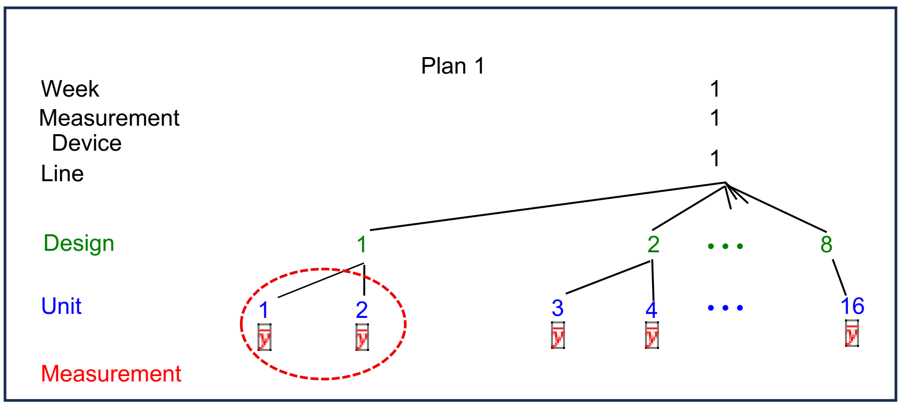


Figure 12: Subgrouping Unit-to-Unit

I will create a range chart for the *new* subgroup. If the range chart displays inconsistency, I will work on the hypotheses pertaining to unit-to-unit and examine the respective x's, *specially*. One of the units is different than the others. It could be:

1. A specific component issue
2. Tube assembly issue
3. Coil assembly issue

I will verify my notes on whether any of these things acted differently during the study or were there other items observed. I may need to work with suppliers, change geometry of the coil or location of the air flow deflectors to simplify assembly. If the range chart indicates consistency, I will calculate an estimate of variation and compare it to "practical" amount of variation I am interested in. If this is significant, I will continue to study the unit variation commonly. And so on, rolling up the tree until the Y-bar chart indicates the within subgroup sources dominate (i.e., in-control). Note there should be predicted data and predicted actions for each scenario and that in every case refer back to the Thought map and update it, modify hypotheses, add hypotheses, drop hypotheses, make hypotheses more specific, etc. and use my predictions to select the next iteration. Additional examples of data analysis and interpretation, including DOE and multi-iteration studies, are provided in appendices A, Thought Maps and F, DOE Analysis and Interpretation.

When experiments are involved, engineers should predict the **rank order of model effects**, at least through second-order terms in addition to the predictions already discussed. This is not done to be correct, but to inform design selection. If first-order effects are expected to dominate, lower-resolution designs may be appropriate. As interactions rise in the predicted rank order, higher-resolution designs are required to avoid confounding critical effects. Prediction, in this sense, is a planning tool.

Prediction as a Measure of Knowledge

Over time, I began using prediction as a way to quantify an engineer's understanding of a system. This led to what I formally call the **Ross Metric**. The Ross Metric is conceptually simple. Engineers predict the data they expect to obtain from a proposed study. After data are collected, the difference between predicted and actual results is examined. The smaller the difference, the greater the engineer's understanding. This idea is closely related to residuals in statistical modeling, but with an important distinction: the comparison is between **engineering predictions and reality**, not between a fitted mathematical model and data. There is also a Bayesian flavor (á priori – á posteriori) to this approach. The purpose is not to score performance, but to calibrate intuition and provide a quantitative self-assessment of knowledge.

Prediction Mitigates Bias

Prediction made *after* seeing the data is rationalization. Prediction made *before* seeing the data is discipline. By explicitly predicting all plausible outcomes, engineers reduce confirmation bias and protect themselves from interpreting data in ways that merely reinforce prior beliefs. This practice also reduces emotional attachment to specific results. One of your predictions will be correct, though many may not be.

Prediction and Iteration

Meaningful engineering and scientific work is iterative. Prediction provides continuity across iterations. As shown in chapter 2, figure 3, each cycle refines understanding:

- multiple data collection plans are proposed,
- predictions are made,
- a plan is selected,
- data are collected,
- results and analyzed and interpreted,
- hypotheses are updated, some are dropped, some are modified and some new ones are formulated,
- and the next *predicted* study is executed (these have already been designed).

This cycle mirrors the scientific method and reinforces the connection between thinking, data collection, and learning.

Critical thinking provides the *structure* for inquiry, but it must be supported by tools that manage complexity. One of the hardest disciplines in discovery is resisting the urge to collapse uncertainty too early. Critical thinking requires parallel processing. Thought maps support **parallel paths** by allowing multiple explanations to coexist without forcing a premature choice. This is particularly important when knowledge is lacking, situational diagnostics have yet to be performed, what to measure has yet to be determined, measurement systems are uncertain, or organizational pressure favors quick answers. Parallel paths do not mean unfocused work. They mean **structured exploration**, guided by explicit reasoning rather than intuition alone.

As learning occurs, Thought maps should evolve. New questions and hypotheses appear. Hypotheses not supported by data are discounted. Learnings are documented and implemented. New data acquisition plans are created. I recommend keeping all versions of a Thought map. Evaluating your Thought map can be quite valuable for improving the efficiency of your critical thinking. The evolution of thinking is itself a record of learning. That record should be documented in your User Guide. When individuals skip this step, they lose the opportunity to reflect on *how* understanding developed, and why certain paths were abandoned and others emphasized.

As studies grow in scope, memory becomes unreliable and intuition becomes overloaded. This is where visual mapping becomes essential. When thinking is visible, disagreement becomes easier to manage. Instead of arguing over solutions, teams can argue over assumptions, missing variables, inference boundaries, or alternative explanations. This shifts the conversation from **who is right** to **what might be true**. That shift is essential in discovery.

Chapter 4
Mapping to Support Learning

Why Mapping is Indispensable

Human cognition has limits. Working memory is finite, and as the number of variables, hypotheses, and interactions grows, those limits are quickly exceeded. When that happens, thinking becomes fragmented. Important variables are forgotten. Questions and assumptions go unstated or undocumented. Decisions are made implicitly or emotionally rather than rationally.

As discussed in chapter 2, the visual sense plays a dominant role in human cognition. When information is externalized visually, it reduces cognitive load and frees mental capacity for higher-level thinking. Thought mapping leverages this by providing a mechanism to document, creating space for analysis, reflection, and iteration. How often have you written something down, only to change it after looking at it? Thought mapping also provides continuity. Engineering investigations often span weeks, months, or years. Without a visual record of how thinking evolved, it becomes difficult to understand why certain decisions were made or why particular paths were pursued or abandoned. This is still a huge challenge in business. How to document and communicate the depth of studies so that knowledge can be transferred into the next iteration. PowerPoint slides are not effective as they typically are results laden, and summarized, rather than extensive documentation of the work done.

Different Kinds of Maps

Although all maps are visual, they serve different purposes. A key distinction in this methodology is between **Thought maps** and **Process & Product maps**. Confusing these two leads to misuse of both. Thought maps belong to **critical thinking**. Process & Product maps belong to **variable identification and measurement**. They are complementary, but they are not interchangeable.

Thought maps: Visualizing Critical Thinking

Engineering projects rarely disappoint because people are not thinking. The project work is less than effective because **thinking is invisible**. Assumptions go unstated. Questions go undocumented. Hypotheses remain implicit. Decisions are made without anyone being able to explain why one path was chosen over another. When results disappoint, the discussion turns quickly to excuses or blame, rather than understanding. Making thinking visible changes that dynamic. When reasoning exists only in someone's head, it cannot be challenged constructively, improved collaboratively, be revisited after the fact, or be used to improve your own critical thinking effectiveness. This vulnerability is especially damaging in discovery work, where uncertainty is high and iteration is unavoidable. When assumptions are implicit, iteration looks like inconsistency rather than learning. Visible thinking provides continuity across iterations. Engineering and scientific problems are rarely simple. Even when the underlying phenomenon is governed by a small number of physical principles, the surrounding system, designs, materials, processes, environments, measurements, customer use, and human intervention, quickly becomes complex. As complexity grows, relying on memory, intuition, or informal discussion becomes increasingly ineffective. Thought mapping exists to address this problem. It is not about drawing for the sake of drawing. It is about **making thinking visible**, managing complexity, and creating structure so learning can occur deliberately and effectively rather than accidentally and haphazardly. Thought maps are used to externalize the critical thinking process. They document how you are thinking, not what exists physically in a system. They don't have to look pretty, they evolve.

The following example is included to illustrate the iterative nature of Thought maps, not to introduce AI as a topic. I am considering how artificial intelligence will affect my methodology. I started by writing some questions and notes on a sheet of graph paper, see Figure 4.1.

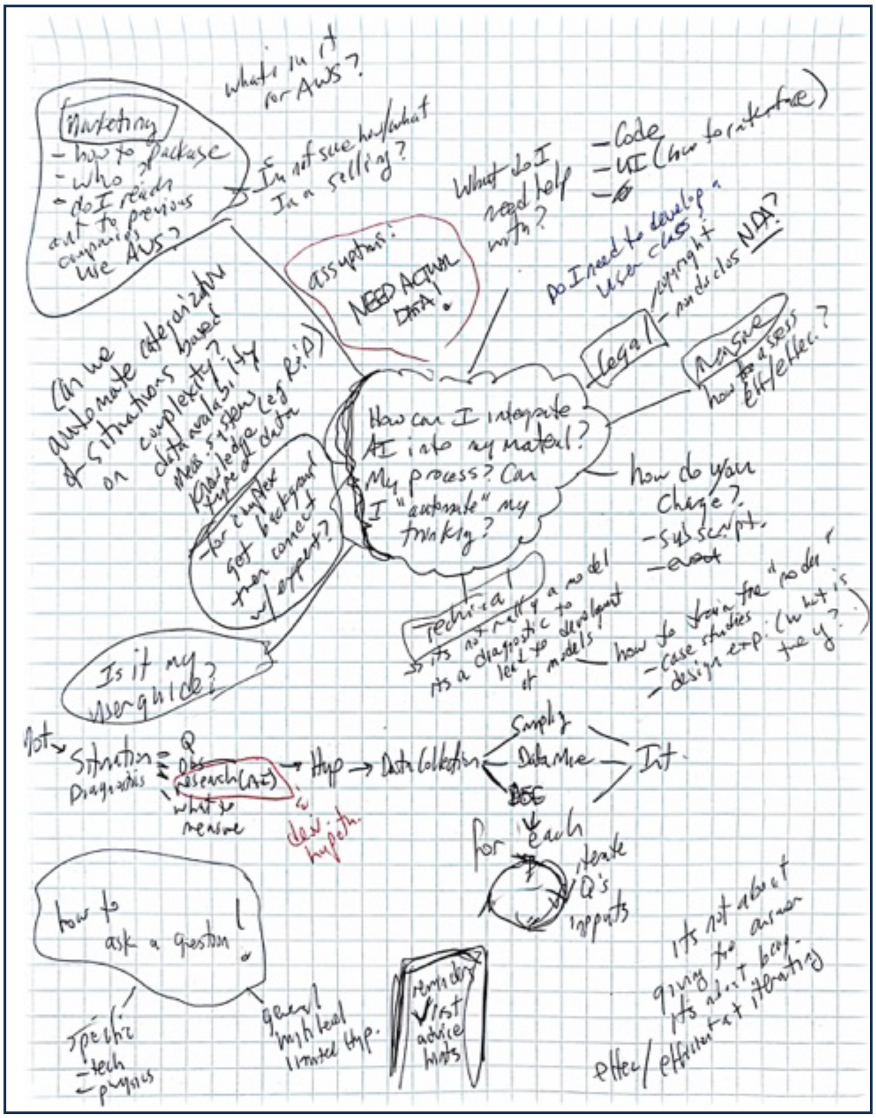


Figure 4.1: Initial Thought map notes

After multiple iterations, I created my latest Thought map, Figure 4.2.

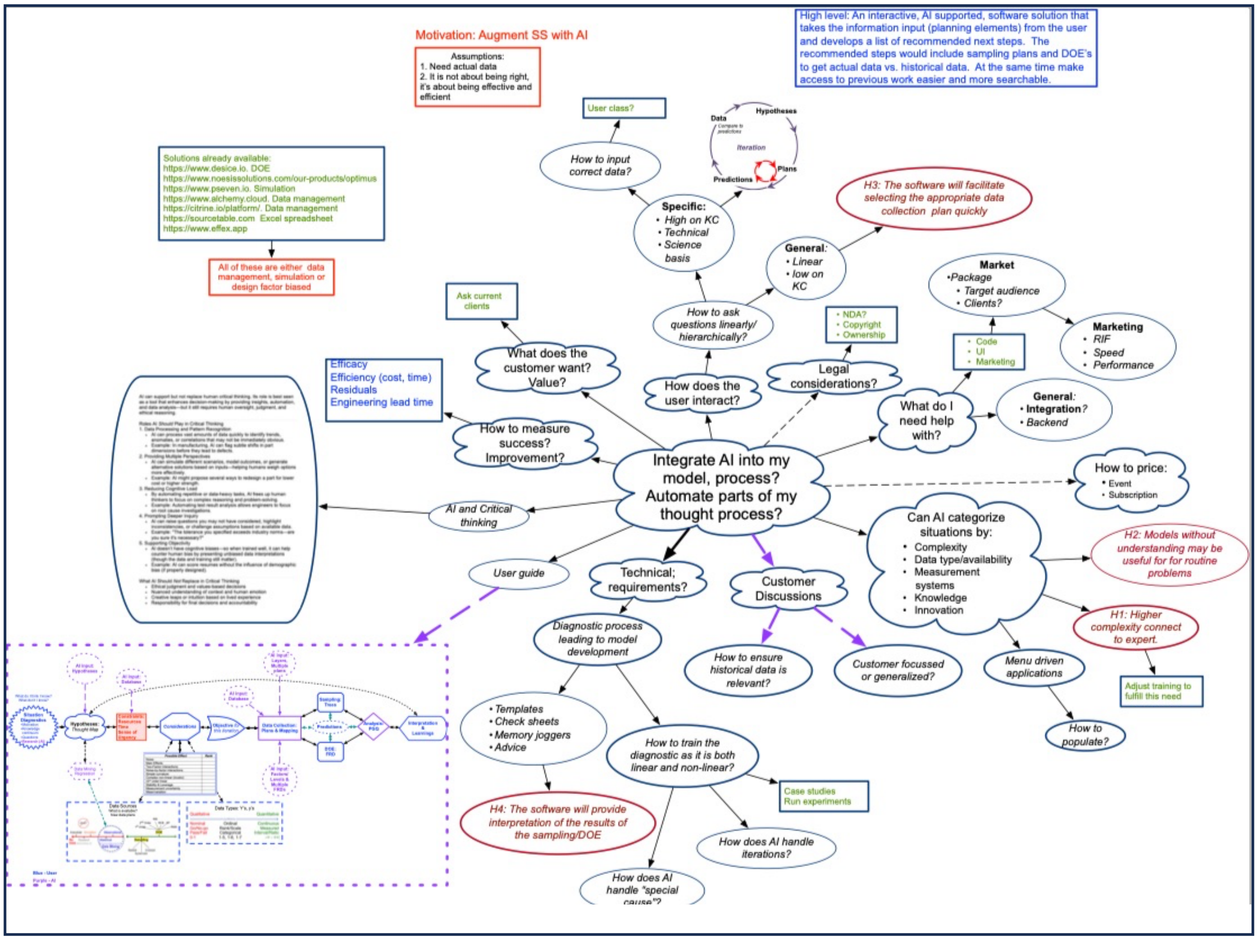


Figure 4.2: Updated Thought Map on Integrating AI into Sigma Science

Additional examples are provided in Appendix A, Thought Maps.

How and Why Thought Maps Emerged

The importance of externalizing critical thinking was not something I initially planned to teach, it emerged from necessity. Early in my career as a consultant, I was teaching a class of primarily PhD aerospace engineers at a Fortune 500 company. There were approximately 25 participants, the proverbial rocket scientists. I was assisted by my coworker, Dr. Doug Sanders. Prior to the class, we gave project assignments and recommended readings. During the delivery of the course material, participants were asked to present what they were working on in brief "project reviews." This served two purposes, first to engage the participants and second to gather information about the nature of project work being done to modify material to be delivered. The first engineer began presenting a traditional PowerPoint deck, following a familiar corporate template. After several slides, I leaned over to Doug and quietly asked if he understood what the engineer was actually doing. He shook his head. I stopped the presentation and apologized. I explained that what was being shown were summary results, but without any visibility into how the study had been conceived, designed, or executed. I reminded the class of something from grade school mathematics: *"Show your work."* Not because the answer is wrong, but because improving the **process** requires seeing the steps. I asked the engineer to do something different overnight: Draw a flowchart of the questions you asked, the assumptions you made, the paths you explored, the decisions you took, and how you arrived at the slides you presented. That exercise resulted in the first Thought map. A Thought map may include:

- motivation,
- observations,
- background knowledge,
- prior studies,
- a list of Y's
- assumptions,
- unanswered questions,
- hypotheses,
- candidate solutions,
- possible data collection strategies,
- predictions (data, all possible outcomes and anticipated actions for each outcome),
- and ideas for next iterations.

Thought maps require the documentation of information most often retained in the minds of those who own the process or the improvement work. The iterative nature of the maps requires those working on a process or product design to evaluate the logic of their thinking and actions with respect to the goals and objectives of the work. Thought maps, like any other tool or methodology, are only as valuable as the information captured on them. So, if a key path of thought is missed, then sub-optimization in process or product performance can result.

As a secondary benefit, Thought maps provide valuable communication. A typical question asked by those working on a project team is "What should we do next?" A typical question asked by others working within a process area or product design group is "What changes are being made to the process/product and why?" A typical question asked by managers is "How can I keep track of the progress and work being done by the engineering staff?" Often, notebooks full of copied documents and formal presentations are kept to provide information on the work performed on projects. However, these notebooks rarely aid in communicating to others what questions have been asked, what solutions have been obtained, and the breadth of work required realizing solutions. In fact, the work necessary to prepare for a presentation is often a distraction from the work performed to gain process knowledge. Thought maps allow for transfer of process/product knowledge and for handing-off components of work. They communicate the hierarchical nature of project work and the many factors, questions, issues associated with the overall objective. Therefore, they are useful for communicating and directing work across a team of individuals working on different paths within the same project. As the Thought map delineates the parallel paths of work required, the team members can each work on a path while maintaining linkages between the parallel paths. The Thought map provides a means of providing feedback on the work of others within the team and for developing an understanding of how individuals arrive at specific conclusions. The evolving nature of the Thought map brings all of the knowledge gained back to a central location, from which new ideas and questions can be generated. Thought maps are also effective at communicating where efforts can be combined to more efficiently arrive at solutions thus reducing redundancy of work efforts.

I have found use of Thought maps, in the technical or engineering community, result in two responses: either they find them extremely powerful or they fail to understand the relevance of the Thought map to their work. Thought maps can be extremely powerful tools. When their power is not realized, it is almost always due to three primary failure modes:

The first major failure mode is they are treated as a static document, done once and left unchanged throughout the project work (like FMEA's in many cases). The intent of the tool is to be used as an iterative document, illustrating the evolution of work through documenting questions that need answering and the new knowledge as it is gained. Updating the Thought map requires discipline.

A second related failure mode is questions and hypotheses are not explicitly defined and stated. Therefore, the work is not guided by rational and logical questions for defining data collection strategies and choice of analysis. Recently, an engineer gave a presentation on a technical analysis performed to understand the variability in paint thickness on various product types. Several pages of data were shown and the results from an experimental design. However, the results were primarily inconclusive. What followed was a series of questions from those observing the presentation. How did you set up the experiment? Why did you choose those factors? What questions were you interested in answering? How did you handle lot-to-lot variation of the raw materials? What sources of variation are you attempting to understand? How did you collect the data? Is the measurement system adequate? etc. In essence, all of these questions led to the conclusion the analysis were not providing relevant information because the data collection and analysis was not designed to answer specific questions. Obviously, the presence of an *iterative* Thought map was absent. In fact, the Thought map had been developed *after* the data had already been collected.

A third failure mode is only using Thought maps to guide the technical components of improvement work. It is important to keep in mind the Thought map provides the documentation, the linkages, and the organization of thought. It does not contain all of the vital backup information and data. Thus, the Thought map should include references, or links to actual data and supporting documentation.

Thought maps are inherently **iterative**. They change as understanding evolves. They are rarely neat or complete, and they are not intended to be. Their value lies in making the thinking process visible so it can be examined, challenged, and improved. Just as critical thinking is independent, Thought maps are also **owned by the individual**. While collaboration is encouraged, particularly to mitigate bias, the act of thinking critically is not a team sport. Doing Thought maps as a team becomes a negotiation. Each person must construct and refine their own understanding. Because Thought maps document thinking of the process to get to some result, they are often uncomfortable for organizations accustomed to polished presentations. That discomfort is a signal that something important is being exposed. Having established why Thought maps exist and how they function, it is useful to examine what they typically contain.

Anatomy of a Thought Map

A Thought map is a structured representation of what you believe might be true, why you believe it, and what would cause you to revise that belief. It is not a brainstorming exercise, a project plan, or a presentation slide. It is a working record for discovery. In early stages, Thought maps are often messy, incomplete, and uncomfortable. That is a feature, not a flaw. Neat diagrams often hide uncertainty rather than reveal it. While no two Thought maps look the same, effective ones tend to share common elements. They begin with a set of central questions often derived from a motivational statement, typically from management. From there, branches or paths emerge representing questions, hypotheses, observations, and relevant background knowledge. Additional branches connect these ideas to potential data collection strategies, sampling plans, experiments, or diagnostic studies. Over time, annotations capture what was learned, what needs to be done next, and what questions remain. The structure is less important than the function. A Thought map might delineate questions such as:

- Why this motivation?
- What clues are there from historical or observational data?
- What are the symptoms?
- What is our current state of knowledge and where is that grounded?
- What are my initial suspicions, hunches, ideas and intuition?
- What should be measured and are those measurement systems adequate?
- What data should I collect?
- How do I ensure the data is representative?
- What scientific theories are applicable (e.g., laws and principles of physics or chemistry)?
- What hypotheses am I evaluating?
- What assumptions am I making?
- What will I do with the results?
- What should the next iterations look like?

The goal is not to create a “correct” Thought map, but a **useful one**. The following sequence illustrates components of a Thought map. This is an example, not meant to be a template.

Example anatomy of the first iteration of a Thought map, Figures 4.3-4.6.

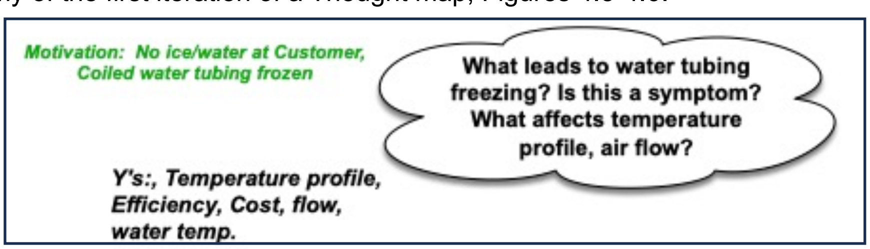


Figure 4.3: Initial Elements: Project Defining Questions, Motivation & Y’s

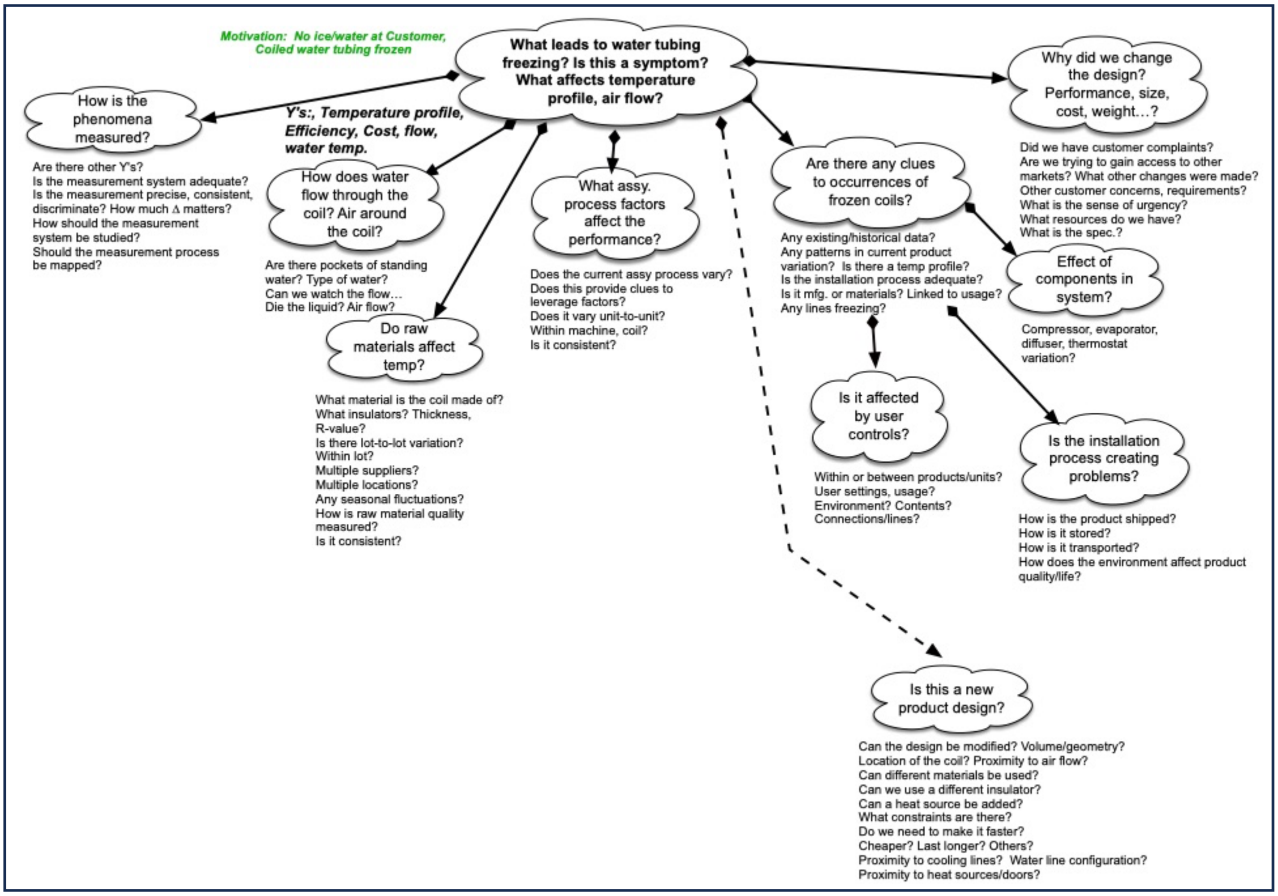


Figure 4.4: Initial Questions Organized

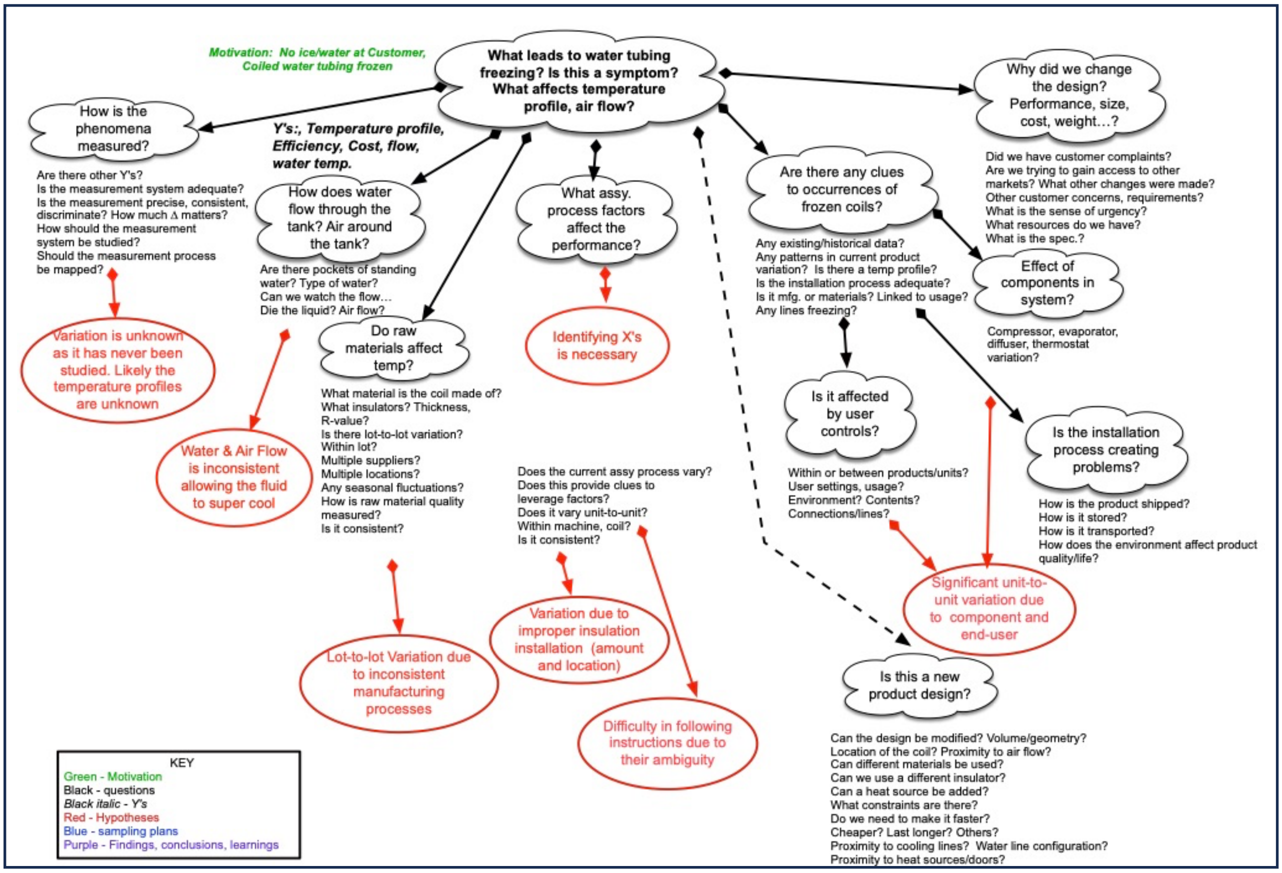


Figure 4.5: Hypotheses Stated

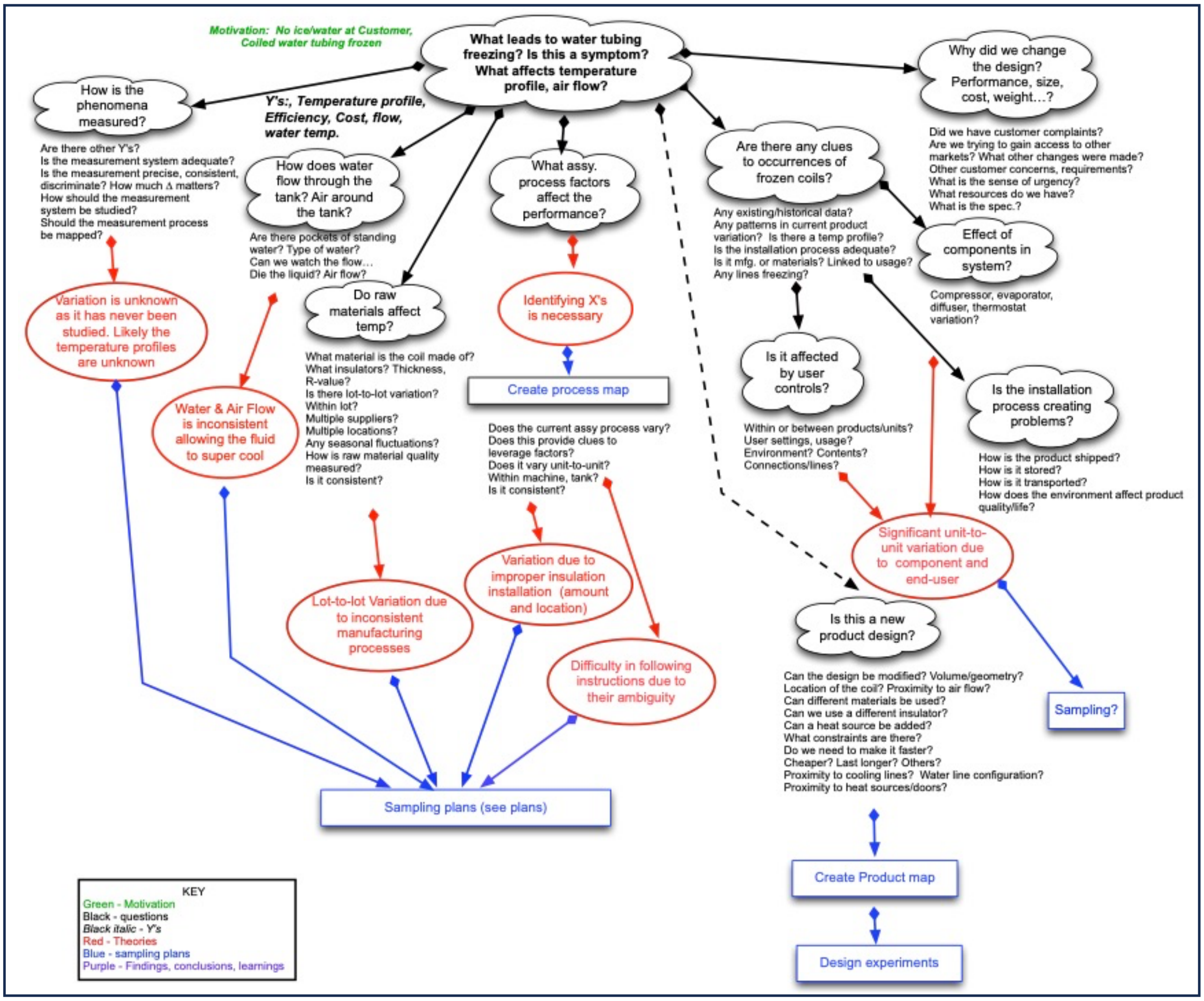


Figure 4.6: Data Collection Plans Linked to Hypotheses or Questions

A Thought map is not intended to be complete, correct or final. It is intended to be thorough and useful.

Process & Product Maps: Identifying Variables

Process & Product maps serve a different purpose. They are used to facilitate identification of variables by location. I invented this type of map as a result of getting blindsided by factors I had not identified before experimentation. A Process map describes the steps to transform inputs into outputs. A Product map describes components of a product assembly, subassembly or components. Together, these maps help identify:

- purpose (for the process, process step, component, subassembly, etc.)
- measurable outcomes of interest (Y's),
- intermediate response variables (y's),
- input variables (x's),

Unlike Thought maps, Process & Product maps are not primarily about thinking style. They are about **variable identification**. They help the investigator create the variable "dictionary" introduced in

Chapter 1. Importantly, these maps are not limited to what currently exists. They can also represent what the organization is **willing to manage in the future**. A factor that is currently uncontrolled may be treated as noise today, but if there is willingness to control it, it becomes a candidate factor for experimentation. The objective is not to produce a perfect or static map. Processes change continuously. The moment a map is declared "complete," it is already out of date. Instead, the goal is to create a reasonable reflection of the process as it exists now, and how it might exist, so variables can be identified.

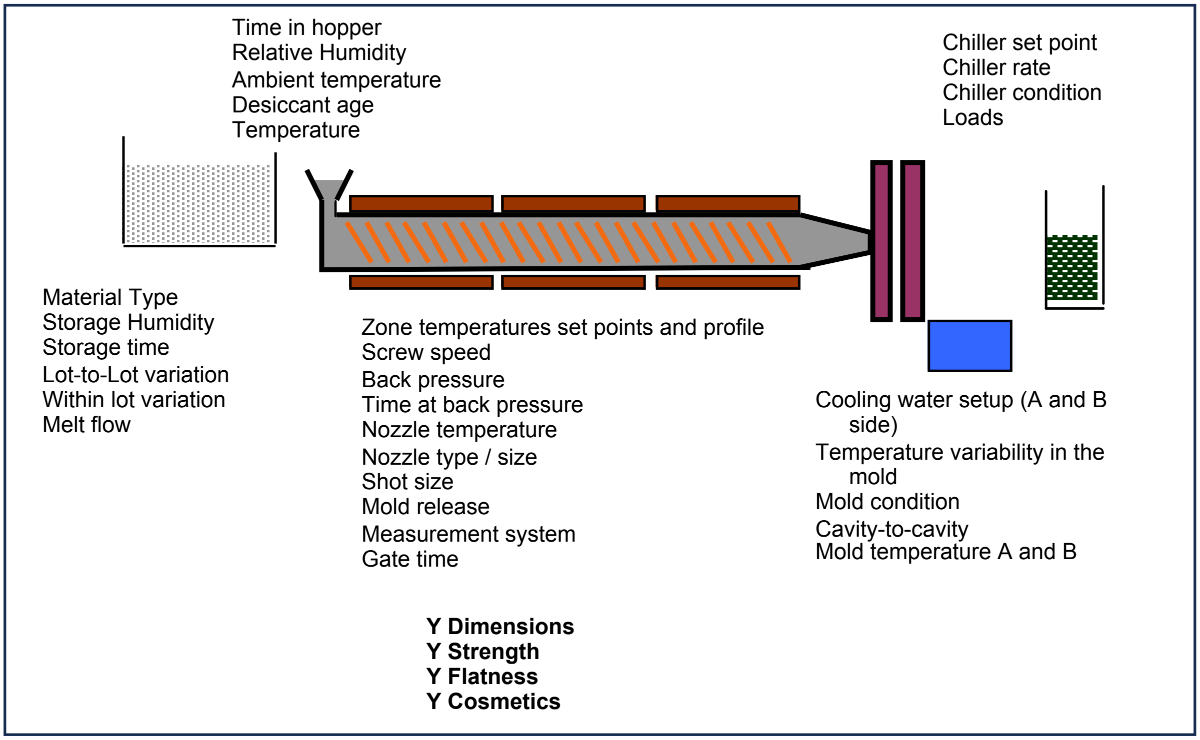


Figure 4.7: Process Map of Injection Molding

The Product map is similarly constructed and contains the comparable elements, see Figure 4.8. What is the purpose of this feature or component on the product? Are there other ways to have that function be performed? How do determine if it is performing as intended? How is that performance measured? What can you vary (e.g., dimension, material, orientation, location)?

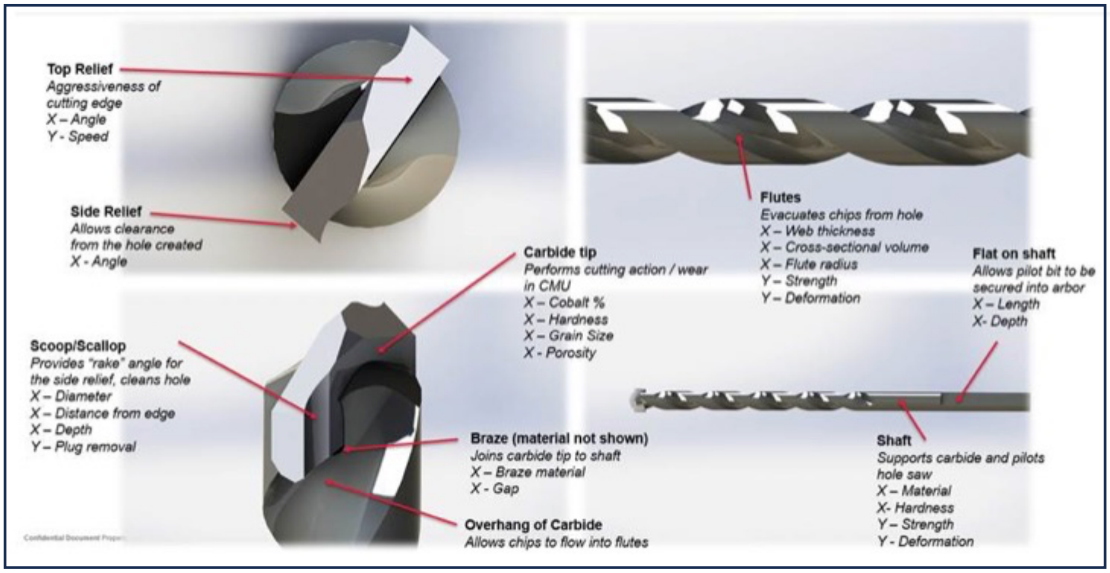


Figure 4.8: Product Map of a Drill Bit

Please see Appendix B for additional examples of Process& Product maps.

Process & Product Mapping and the Variable Dictionary

Earlier, I introduced the “dictionary”. Process & Product Mapping is a tool that assists in developing a comprehensive set of variables, the *dictionary*. It accomplishes this by engaging the visual sense during observation of the process in action or the product in use, over time. If a variable is not identified on a Process or Product map, it is unlikely to be included in a study. If it is not included in a study, it cannot be assigned and its effect cannot be determined. When this happens, the variable becomes a hidden source of confounding. This type of mapping does not guarantee completeness, but it dramatically reduces the likelihood of missing variables, some of which may be important. The objective is not to find the “right” set of variables, but the most **inclusive** set possible at the time. The detail of the Process or Product map should be commensurate with the detail of the hypotheses. As knowledge increases, variables identified earlier, that were actually chunks of variables, are disaggregated. For example, there may be a note of environmental variation in an early version of the map. Later that is broken down into humidity, ambient temperature, dew point, etc.

Linking Maps to Data Collection Strategies

Process & Product Mapping is not an end in itself. Its value lies in how it connects thinking to action. Process& Product maps help determine what variables are associated with the area of focus as a function of the hypotheses on the Thought map. Together, they provide the rationale for selecting appropriate data collection strategies. Sampling plans and designed experiments should not be created in isolation. They should be direct consequences of the hypotheses articulated on the Thought map, and the variables identified on the Process & Product maps. When mapping is skipped or rushed, data collection strategies are often chosen based on convenience, habit, or organizational pressure rather than learning potential.

Transition: From Mapping to Data Collection

Though there is no “right” order, it seems rational to proceed once thinking has been externalized and variables have been identified, to the next question: How to acquire data that meaningfully supports learning and provides insight to hypotheses. In the next chapter, we turn to **data collection strategies**, how Thought maps and Process & Product maps guide decisions about sampling, COV studies, and designed experiments (DOE). Mapping creates the dictionary. Data collection determines how the dictionary can be filtered to understand which of those variables are the most interesting and useful.

Chapter 5
Linking Critical Thinking to Data Collection Strategies

Data collection does not begin with a sampling plan or an experimental design. It begins with thinking. One of the most common errors I see in engineering and scientific investigations is the haphazard selection of a data collection method. If they were taught DOE, they apply DOE to everything.

> *"When you have a hammer, everything looks like a nail"*
>
> *Maslow's Hammer*

Engineers ask; *What tool should we use? Should we run an experiment? How many samples do we need?* These questions are often asked before the underlying problem has been framed, before hypotheses have been articulated, and before variables have been identified. In practice, engineers are often told to "use the available data.", "see what the data says.", "we already have plenty of data." This language is misleading. Data does not exist independently of how it was generated. Every dataset is the result of choices, some deliberate, many accidental, about where measurements were taken, when they were taken, under what conditions, and for what purpose. Ignoring those choices does not make them disappear. It simply hides their consequences.

In this methodology, **data collection strategies are a consequence of thinking**, not a starting point. What data will I need to answer my questions? How will I collect data to provide insight to my hypotheses? How do I ensure the data I collect will be representative of future conditions? Even if there is pressure to implement a solution, What data will I need to ensure the solution is adequate?

> *"Data have no meaning in themselves; they are meaningful only in relation to a conceptual model of the phenomenon studied"* [17]

From Thinking to Action

Thought maps document the critical thinking process including observations, assumptions, questions, and hypotheses. Process & Product maps identify variables and measurement locations. Together, they provide the context required to make intelligent decisions about how data should be collected. Without this context, data collection strategies are selected based on habit, convenience, time pressure, or organizational norms. When that happens, the data may be abundant, but learning is limited. The purpose of linking Thought maps to data collection strategies is to ensure that **every study is intentional**, that data are acquired because they are expected to provide insight into specific hypotheses about causality.

[17] Box, G.E.P., Hunter, Bill, Hunter, Stu "Statistics for Experimenters".

Representativeness Is Not a Statistical Property

A critical clarification is representativeness is not a statistical concept. It is a conceptual one. A study is representative only to the extent that it meaningfully reflects the real variation now and in the future. An enumerative thinker uses large sample size and randomization to increase the chances samples are representative. This may lead to inefficient studies. Statistical rigor applied to non-representative data only produces precise answers to the wrong questions. Which sample size is better, 10,000 or 5? Now the situation, the sources of variation acting day-to-day are significant. The 10,000 samples are gotten in one day. The 5 samples are one per day for five days. Which is more likely to uncover insight into the actual problem? It is not a question of sample size, but of how to represent the true variation in your sample. Representativeness must be evaluated *before* data are collected by asking:

- What phenomena am I trying to understand?
- Under what conditions do I want conclusions to be valid?
- Which sources of variation must be present for the study to be useful?
- Did those sources vary enough to draw operative conclusions?

Recall, in order to quantify the effect of a factor, it must vary in the study. Thought maps make these questions explicit. Process& Product maps help determine whether the variables needed to answer them are included.

Data Type

Some metrics are often Yes/No (or qualitative, discrete) responses such as "Does the product meet specification?". Inspection is also often qualitative, not for the same reason, but rather for expediency. Sometimes it is straightforward to change the response to quantitative or continuous (e.g., "How well does the product perform?", "Why is it failing?", "How does it work?"). In all cases, a relevant **quantitative** metric will be much more efficient in understanding causal structure.

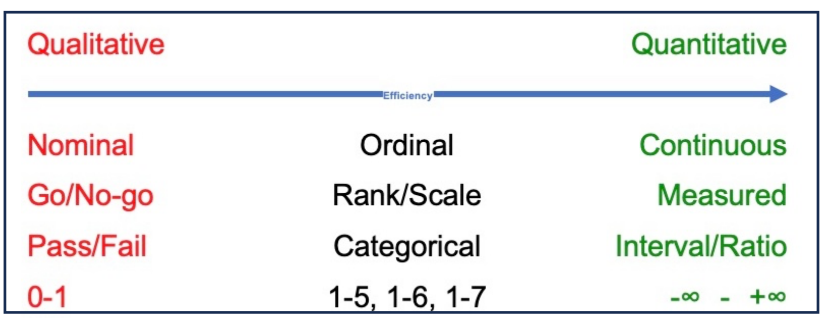


Figure 5.1: Summary of Data Types

COV and DOE Are Situation Dependent

Throughout this book, sampling, directed sampling, and COV studies refer to the same underlying intent: structuring data collection to expose sources of variation relevant to specific hypotheses. I do not advocate for COV *instead of* DOE, nor for DOE *instead of* COV. In many investigations, both are necessary, sometimes sequentially, sometimes in parallel. COV helps assess stability, quantify sources of variation, and identify where leverage lies (i.e., which set of factors should be studied in greater detail). If one is trying to understand an already existing phenomena or a rare event, my bias is to use COV. DOE intentionally varies independent factors to accelerate learning about causal relationships. Unfortunately, often engineers are forced into using methods that reduce time. Again, effectiveness is sacrificed for efficiency. Can you create the phenomena with experimentation? Perhaps. Is that why the phenomena occur naturally? Good question. Which approach is appropriate depends on the situation:

- What is already known and how was this knowledge acquired?
- What is unknown?
- How stable is the system?
- What resources are available?
- What is the sense of urgency?
- What risks are acceptable?

These questions cannot be answered by statistical rules. They require judgment informed by thinking and mapping.

Hypotheses Drive Data Collection

In analytical work, data collection should be **hypothesis-driven**. Hypotheses are explanations for a why a given phenomenon occurs. They also provide clues as to what might be measured and how data needs to be collected. COV and DOE are designed as functions of hypotheses, not convenience. This means:

- deciding which components or factors must vary,
- by how much or over what period of time?
- which may be constant or are held constant (inference space),
- and, which may be confounded to increase efficiency

When hypotheses are weak explanations or merely descriptive, data collection becomes ill-defined. When hypotheses are explicit, specific and include causal explanations, even simple studies can be highly informative. Here is an example: Suppose there is a problem with dimensional characteristics of an injection molded part. There might be several hypotheses: One hypothesis might be the measurements system is inadequate. Another might be it is due to material variation. And yet another might be it is due to variation in molding parameters. However, none of these are actually explaining why these possible hypotheses would result in dimensional issues with the part. Contrast with these hypotheses; the dimensional issues with the injection molded part is due to incorrect mold temperatures affecting the length of polymer chains or the measurement system is inadequate due to orientation of the part in the measurement device. Notice this also suggests measuring polymer chains and part orientation as intermediate responses (y's).

Multiple Candidate Data Collection Strategies

A key practice in this methodology is the deliberate generation of **multiple candidate data collection strategies** before selecting one. Each candidate strategy should be evaluated in terms of:

- what questions it can answer,

- what hypotheses it can compare,
- which variables are included or excluded,
- what assumptions are being made,
- what resources are required,
- and what the likely next iteration would be.

This comparison transforms data collection from a default activity into a **decision-making process**. It allows engineers to trade off potential learning against time, cost, and risk in a rational way. It is highly unlikely anyone can select the "best" COV or DOE. To do this you would need profound knowledge á priori, which means you wouldn't need the data.

Iteration Is Expected

No data collection strategy is perfect. The first plan rarely uncovers the defining equation. This is not a failure; it is the nature of learning.

> *"The purpose of the first COV/DOE is to design a better COV/DOE"*
>
> *Ross*

The objective of an initial study is often to design a better study. Thought maps, Process & Product maps, and data collection strategies should all evolve together as knowledge increases. Starting with simple approaches and adding complexity only as needed is generally more effective than attempting to do everything in one study. Recall the acronym that states this principle, KISS.

Transition: From Strategy to Execution

Once a data collection strategy has been selected, the focus shifts to execution. How data are collected, timing, grouping, sequencing, matters deeply. In the next chapter, we focus on COV studies, beginning with one of the most misunderstood tools in practice: control charts. See Part II for discussion of data analysis and interpretation.

Chapter 6
Components of Variation (COV)

In this book, sampling and COV studies are the same thing. The distinction is largely one of language. The term “sampling” is often more understandable to engineers, while “COV” emphasizes the analytical intent. Both describe a directed approach to data collection designed to understand how variation arises in a system.

COV is one of the most misunderstood data collection strategies in engineering and scientific work. It is often treated as a monitoring technique rather than as a deliberate strategy for learning. I recall when I was first taught statistical process control (SPC), the instructor said sample the process every hour and collect five samples. This was how the data collection sheet was set-up. No discussion of a rational subgroup. It has been observed most training in the use of these techniques is solely on data analysis rather than on planning for the acquisition of data. Here are the equations to calculate control limits, not how to interpret control limits.

Sampling is a technique used to assist in understanding a larger population. Sampling may be used to draw conclusions about that which already exists (enumerative) or to assist in prediction of what may happen (analytical). In both cases, sampling implies not all items or characteristics will be measured, only a select few (a sample). Selection of the items to be measured is an **important decision** and will provide the context for data analysis. The intent of COV is to correlate variation in outputs (Y’s) to variation in inputs (x’s) to gain understanding of potential causal relationships. Once the relationships are exposed and understood, the sources can be managed, improved, designed out, or their effects mitigated. COV is an effective and efficient means of investigating a large number of potential sources simultaneously.

The Purpose of COV

Most people who use control charts at all use it for what I call monitoring purposes, as a sort of early-warning mechanism. If all the data lie within two horizontal lines, called the control limits (computed as a function of how the data was collected), and continue to stay there, all is regarded as being well, and people may relax and think of other things. But if the process, say, starts to wander in some way, or there is an unusual event, the control chart signals the onset of trouble, so that corrective action may be taken before the trouble becomes too serious. Now, I am not saying that it is wrong to monitor outputs and react, but this is not how control charts were intended to be used. They may work well in that early-warning role. I’m simply saying if that is all you are using the control chart for, you are missing out on the main purpose for which Shewhart intended. Shewhart’s control chart method was to provide guidance for the type of actions to take which will lead to improvement, to making things better, not to just maintaining status quo. To merely maintain things as they are, or to improve: that’s the difference. COV is used to understand currently existing processes. Its purpose is not optimization, but rather to learn how a system behaves under natural conditions and expose the natural variation of the factors in the study. Specifically, COV is used to assess stability of the within subgroup sources of variation, quantify sources or components of variation, determine where leverage exists, and guide decisions about how an investigation should proceed (e.g., special or common).

Sampling plans are designed as a function of hypotheses and questions articulated on the Thought map and factors identified on Process& Product maps. When sampling is treated haphazardly, its ability to support learning is severely compromised. It is impossible for any sampling plan to capture all potential sources of variation that act on the process now and in the future or to completely answer all initial questions. The best industrial studies are sequential in nature, a sampling plan leading to another expanding and/or refining the information. Again, the KISS principle. It is generally possible to decrease the duration of the sequential study by initially considering as many potential sources of variation as possible. This does not imply the initial sampling plan should be where all potential sources of variation initially considered are estimable (separable). A more efficient approach would be to explicitly confound "subsets" of x's (and associated hypotheses) within each layer and compare these large subsets of x's to each other (layer-to-layer) to determine which layer has greater leverage. Then investigate and separate the confounded x's in the layer that had greater effect in a subsequent study.

Elements and Use of Sampling Trees

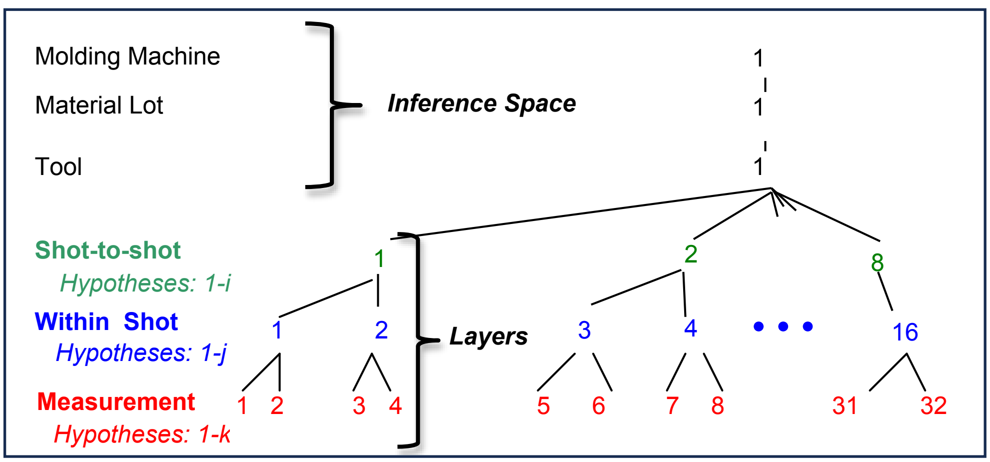


Figure 6.1: Sampling Tree for Injection Molding Process

Sampling trees consist of the following elements:

1. The inference space at the top of the diagram. The variables that are constant during the sampling. I suggest completing the inference part of the tree last when creating the tree.
2. Multiple layers labeled according to the hypotheses found on the Thought map and explicit instructions as to how the samples will be acquired. The layers are color coded to link the x's corresponding to the Process & Product map.
3. A numbering scheme that is always coded using whole numbers to indicate variation. Refrain from using actual values as this complicates analysis.

Sampling trees visually describe:

- how the samples are collected (mechanism),

- why those samples were chosen (hypotheses).
- how many samples/data points will be needed (impacting resources),
- potential patterns within layer (e.g., random or systematic), and
- potential relationships between layer (e.g., nested or crossed)

The sampling tree guides analysis by providing the grouping for variability plots, the subgroup size for control charts, how summary statistics will be calculated for "rolling up the tree" and what comparisons are being made (i.e., within and between subgroup). The tree links to the hypotheses found on a Thought map and the x's found on a Process or Product map for context. There is no limit to the number of measurements (Y's) that can be taken on the samples acquired. Figure 6.1 depicts a sampling tree for an injection molding process. Figure 6.2-6.4 depict a sampling plan with associated hypotheses and modified Process map.

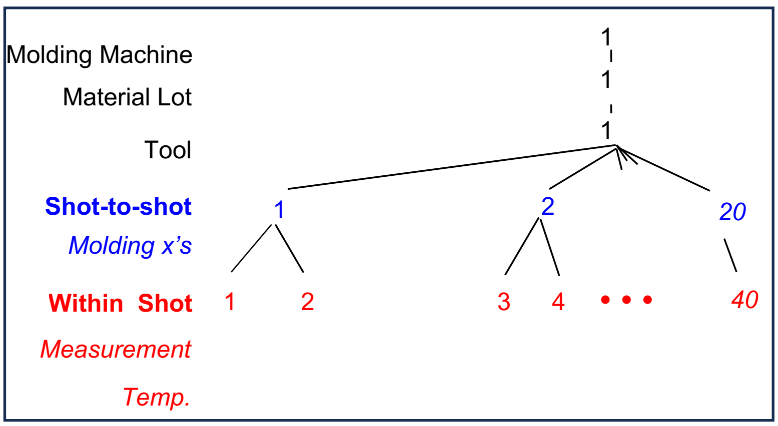


Figure 6.2: Sampling Plan

Hypotheses transferred from Thought Map:

1. Molding parameter variation (x's) contribute to variation in OD due to their effect on the length of polymer chains.
2. Measurement system varies due to part orientation and op. technique.
3. Temperature fluctuation inside the tool contributes to varying shrink rates effecting polymer chains.

Figure 6.3: Hypotheses Associated with Sampling Plan in Figure 6.2

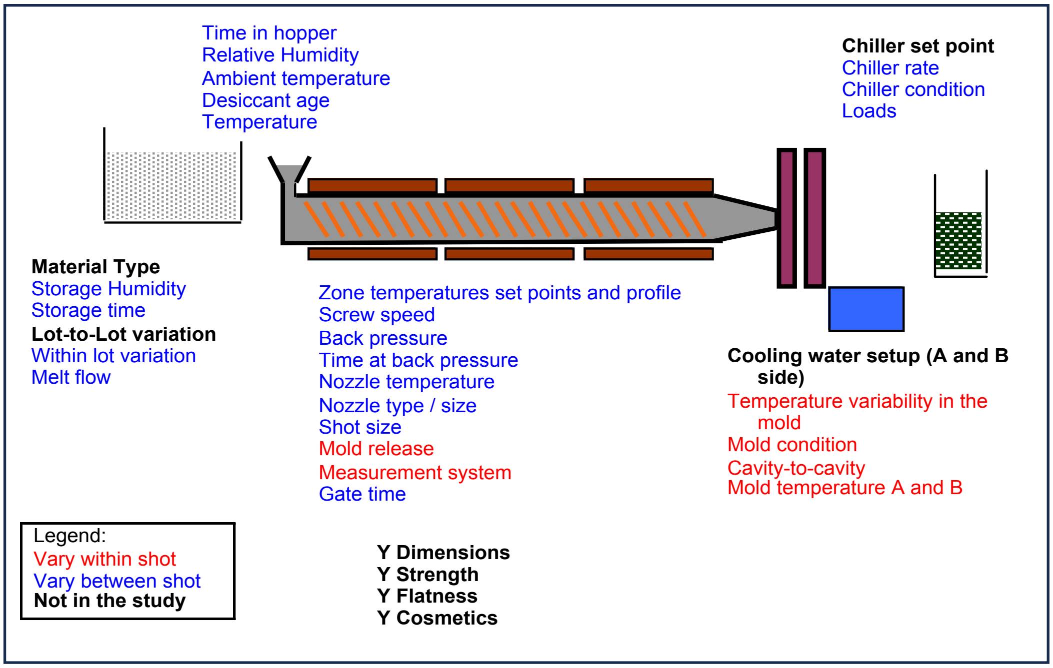


Figure 6.4: Process Map Color Coded Corresponding to the Sampling Plan in Figure 6.2

Stability should be evaluated first

Before attempting to quantify and compare sources of variation or interpret patterns in data, a more basic question must be answered: *Is the basis for comparison stable?* Stability refers specifically to the consistency of the **within-subgroup sources of variation**. If those sources are not stable, then any comparison of those sources (e.g., measurement error, within batch, within part) to the between sources of variation (e.g., machines, operators, lots, environmental conditions), is conditional and potentially misleading. This is why stability must be assessed *before* answering questions about leverage.

Appropriate Use of Control Charts

Control charts are a visual **diagnostic** and are **discovery tools**, not reactionary tools. They are designed to answer specific questions regarding potential sources of variation. Once they have answered those questions, change the question. That means change the sampling plan. They can be used as a discovery tool by proactively changing sampling and subgrouping strategies. By doing so, one can change how the x's are partitioned and what comparisons are being made. They do not tell you whether you need to investigate, they tell you **where** and **how** to investigate. The control chart method consists of two charts, the range chart and the average chart, each serving a different purpose.

The **range chart** (or moving range chart) assesses the stability of the within-subgroup sources of variation. It answers a single question: *Are the within subgroup sources of variation acting consistently, stably, predictably over time?* Because the range measures variation among observations collected under the same conditions, it isolates the variation that is intrinsic to the within subgroup and the x's that change at that frequency. The way consistency is determined is by comparing the same sources of variation to themselves, over time. Only range charts assess stability. No other chart performs this function. If the range chart indicates instability, the evidence suggests that something different happened at a specific point in time. These are candidates for **special cause like** investigation. The analogy I often use is imagine you are a homicide detective. When a special cause is captured in the data, it is like you walk into a room and see a dead body with a person standing over the body with a smoking gun. Case solved. In such cases, it is appropriate to investigate those specific instances. These events are rare, and when identified, Deming suggested acting locally rather than changing the entire system. If the range chart indicates stability, then the variation is unassignable at the individual-instance level. In this case, learning must focus on understanding the relationship between Y and the x's: Y = *f*(x).

> *"The first step in the examination of data is accordingly to question the state of statistical control that produced the data."*
>
> *Deming, Out of the Crisis, p.312*

Once stability of the within subgroup has been established, the second chart, an **average chart**, is useful. The average chart compares the variation between subgroup averages (plotted points), due to the sources of variation changing between subgroup, to the variation within subgroups (represented by the control limits) due to the sources of variation changing within subgroup. The control limits on the average chart are derived from the within-subgroup variation, making this a direct comparison of sources of variation. This comparison answers a critical question: *Where is the leverage, within or between?*
This comparison is what allows leverage to be identified. An *in-control* average chart indicates within sources dominate. An *out-of-control* average chart (or some other non-random pattern) indicates the between subgroup sources of variation dominate. It is not a question of good or bad, just an estimate of which source has a larger effect on the plotted metric. Leverage refers to which **set** of factors contributes most to variation in the response. For example, variation:

- within machine to variation between machines
- within lot to variation between lots,
- within operator to variation between operators,
- in a short period of time to variation over a longer period of time,
- within environmental conditions to variation across environmental conditions,

Rational subgrouping is the key to effective use of this tool.

> *"The engineer who is successful in dividing his data initially into rational subgroups based on rational theories is therefore inherently better off in the long run. . ."*
>
> *Shewhart*

Identifying leverage helps prioritize where effort should be applied next. Where you get the greatest opportunity to learn about causal structure.

Special Causes, Common Causes, and Interpretation

Special causes are rare. They are instances where the data suggest something different, unusual or according to Shewhart, something assignable has happened. Of course, this is completely dependent on the data set, how the data was acquired and how the data was partitioned. My own hypothesis is special causes are often the result of higher-order interactions, interactions among four or more variables, that occur infrequently. Deming initially indicated they about 20% of the time. I believe this is consistent with the 80-20 rule. However, through the years, he updated that probability to less than 2% of the time.

An analogy I often use is from the game *Clue*. A special cause is like Professor Plum, in the library, with a wrench, on the second Tuesday of a month that has two full moons. Such combinations are possible, but rare.

Designing a system to be robust to such rare events is usually not practical. Deming suggested reacting to such signals specifically and locally. Redesigning a process to prevent extremely rare events is usually inefficient, unless the consequences are catastrophic. Acting locally may be the appropriate response. Common causes, on the other hand, reflect the inherent variability of the system. When variation is common-cause dominated, there is no single instance to investigate. Learning must focus on understanding causal structure. This is where the causal framework, Y = *f*(x) becomes central. Improvement requires identifying which x's contribute meaningfully to variation in Y and how they interact. Control charts do not provide this understanding directly. They tell you *where* to work, and *how* to investigate.

Acting on Rare Events: A Cautionary Example

It is tempting to overreact, or take common cause action to rare events. One example I often use occurred when an explosive device was discovered in a shoe during airport security screening[18]. This was likely a special-cause event. The response was to take **common-cause action**, changing airport security procedures for everyone, everywhere. The result was a dramatic increase in the time and inconvenience experienced by millions of travelers. Whether this was effective or appropriate is debatable, but it illustrates a broader point: acting systemically on rare events can have unintended and widespread consequences. This distinction between local and systemic action is critical in engineering and organizational decision-making. What do you want engineers to be working on? Special or common cause events.

COV as a Learning Accelerator

COV can dramatically increase learning efficiency, particularly early in an investigation. It enables the investigator to partition the sources of variation and compare their contributions. Recall the dictionary illustration as discussed in chapter 1. Adding prior knowledge, hypotheses, allows the dictionary to be sliced even more efficiently, as well as eliminate some candidates right off the bat. The dictionary must be

[18] Richard Reid, December 2001

inclusive. If the word is not in the dictionary, no amount of clever questioning will find it. COV serves the same purpose: narrowing uncertainty efficiently by structuring how variation is exposed.

Confounding

Confounding is a **deliberate** strategy to improve the efficiency of a particular sampling strategy (e.g., chunking or grouping of x's). Its effectiveness relies on identifying what is confounded. The purpose of COV is not to gather information on all possible inputs and determine their contribution to the output **in one plan**. Rather, the COV is directed and focused by specific hypotheses. The purpose of the COV is to efficiently confirm, modify or invalidate those hypotheses. Further sampling (e.g., COV or DOE) is almost always required. Effective sampling does not answer all questions. It answers some of the questions and tells you **which questions are worth answering next**. COV can reveal dominant sources of variation, negligible sources of variation, measurement issues (stability and uncertainty), and unexpected variation in the system. That information determines which components should be pursued more deeply, or is DOE warranted, or does the measurement system need attention, or the problem itself needs reframing. COV is not a detour. It is orientation.

Limitations Must Be Recognized á Priori

Once data are collected, the investigator is constrained by the study design. Sampling plans define: what questions can be answered, what conclusions are justified, confidence in those conclusions, and how far into the future results can be extrapolated. There is no such thing as "the right" COV, only COV that is more or less **effective or efficient** for a given situation. The responsibility of the engineer or scientist is to recognize these limitations *before* data are collected. Or, at least, understand how these limitations impact conclusions.

Nested, Systematic and Crossed

Random vs. systematic sampling refers to the way **a particular layer** (within layer) is collected. Random sampling increases the chances the sample is representative **when there are no hypotheses** to explain the variation. This is also considered nested or hierarchical. Random sampling is shown by sequential numbers, as in Figure 6.5 and is depicted in an illustration of sampling within and between batches in figure 6.6.

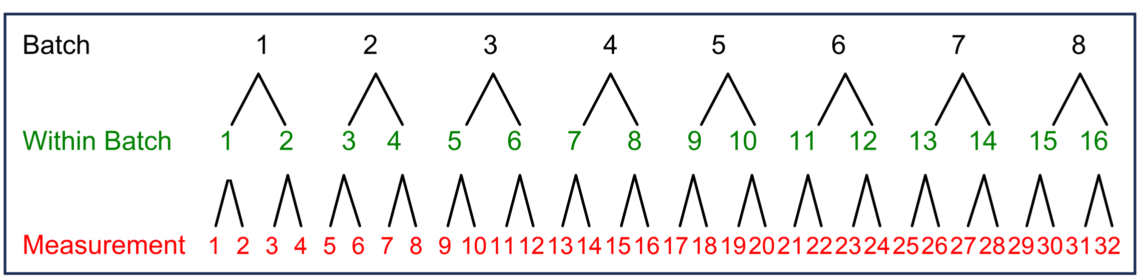


Figure 6.5: Nested, Random Sampling Plan

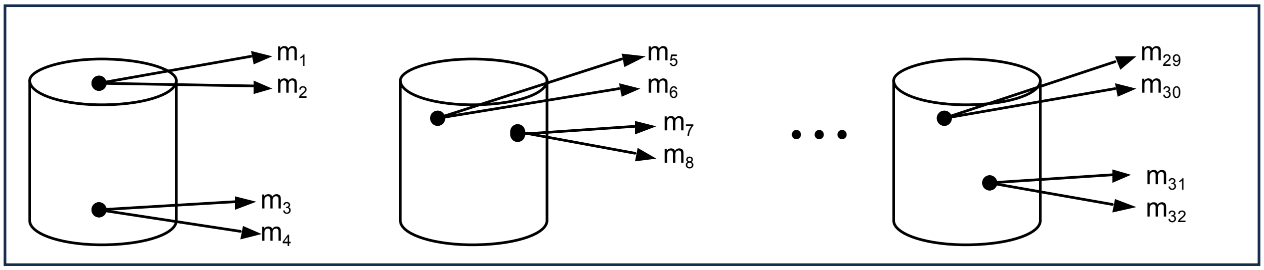


Figure 6.6: Depiction of Random Sampling Within Batch

Systematic sampling is shown by **repeating** numbers **and** a **random** sequence at the same layer. It is tied to a **more specific hypothesis** and may expedite the learning. Figure 6.7 depicts a systematic sampling plan and Figure 6.8 shows how the samples were acquired from just the top and bottom locations of the vat.

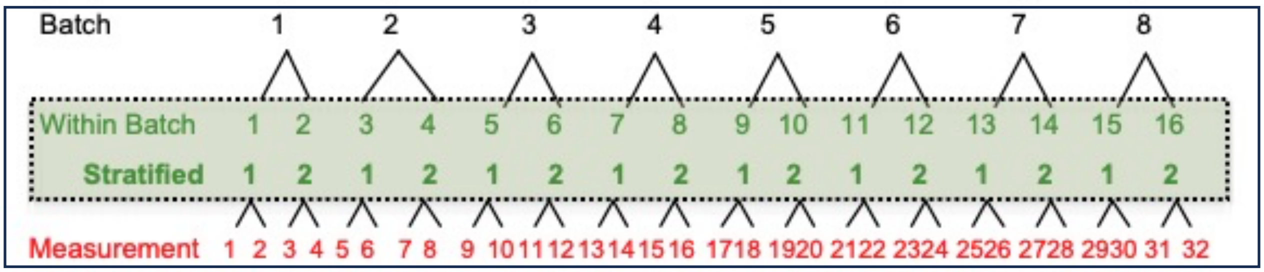


Figure 6.7: Systematic Sampling Plan

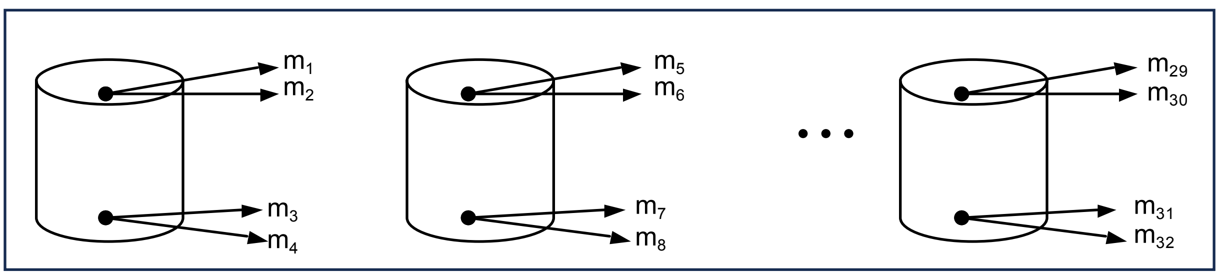


Figure 6.8: Depiction of Systematic Samples Within Batch

Nested (or hierarchical) and crossed designations refer to the relationship **between layers**. A nested layer is unique in relation to the layer above it. The layers cannot be in any other order. Crossed is the condition where a layer sees every item in another layer (a factorial). In crossed studies, there is no hierarchy, any layer could be in any hierarchical order. When layers are crossed, interactions can be estimated. Figure 6.5 depicts a nested sampling plan, Figure 6.7 a systematic layer and Figure 6.9 a study that includes both crossed layers and a nested layer.

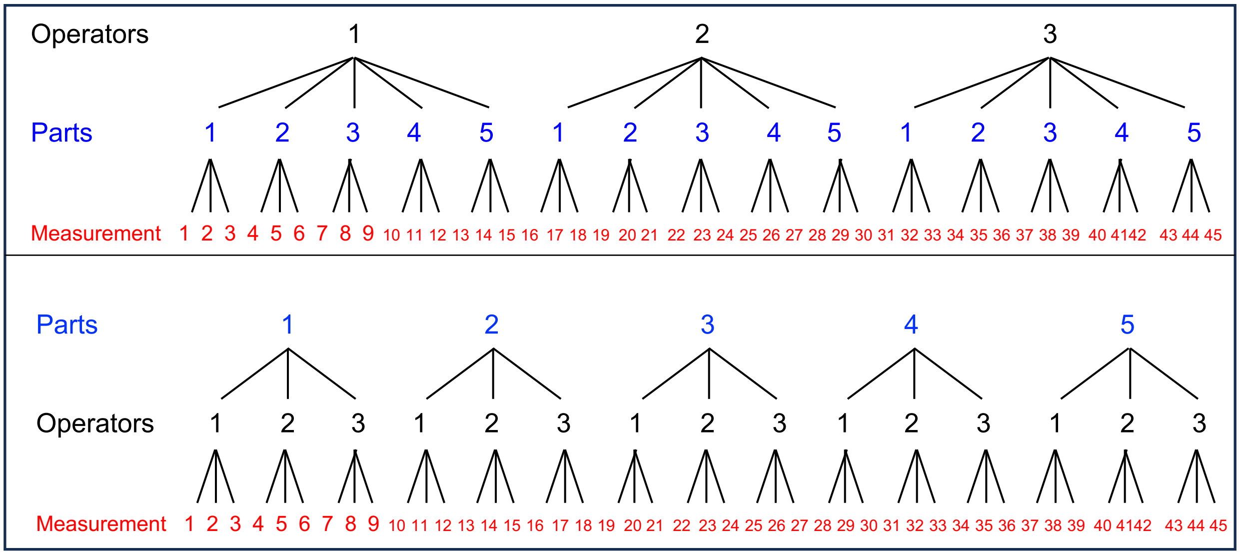


Figure 6.9: Two Identical Sampling Plans Depicting a Crossed Layer (Operator and Parts) and a Nested Layer (Measurement)

Figure 6.10 summarizes the differences between the types of sampling plans.

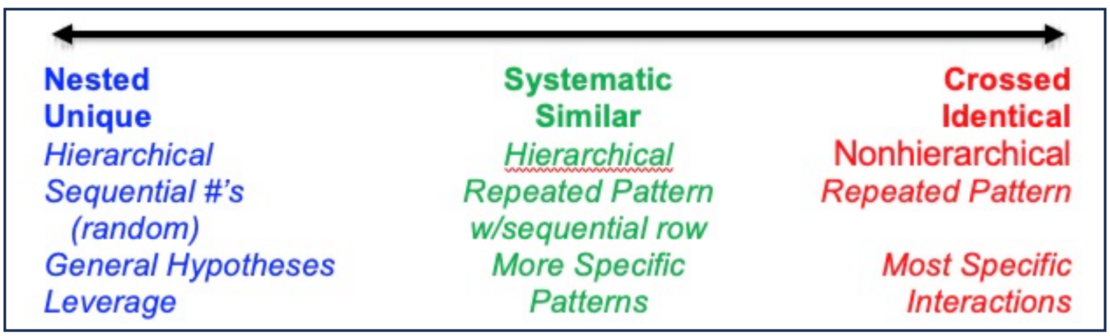


Figure 6.10: Comparison Summary

A Note About Measurement Systems (See Part II, Chapter 14, Measurement Systems)
The total variability of a process or product can be assigned to various sources. For instance, a common application of COV in industry is investigation of measurement processes. These investigations are variously called Measurement System Evaluations (MSE), Measurement System Analysis (MSA), Gauge Repeatability & Reproducibility (Gauge R&R) or Evaluating the Measurement Process (EMP). The specific intent of these studies varies, but a common intent is to understand whether a measurement process is capable of providing insight to hypotheses proposed by the engineer. If the measurement process is not capable of this task it may be possible to improve and reduce the variability of the measurement process, identify alternative measurement systems, or define alternative measures of the process or product performance (Y's).

A typical measurement system study (aka. Gauge R&R) might have multiple operators/technicians, measure multiple parts, multiple times, each in a manner consistent with their usual mode of inspection and in *blind* fashion (an operator is not aware of which sample he or she is measuring). A simple *tree diagram* depicting a sampling plan with three operators, measuring five parts, three times each is shown in figure 6.11. Each of these are layers of the COV.

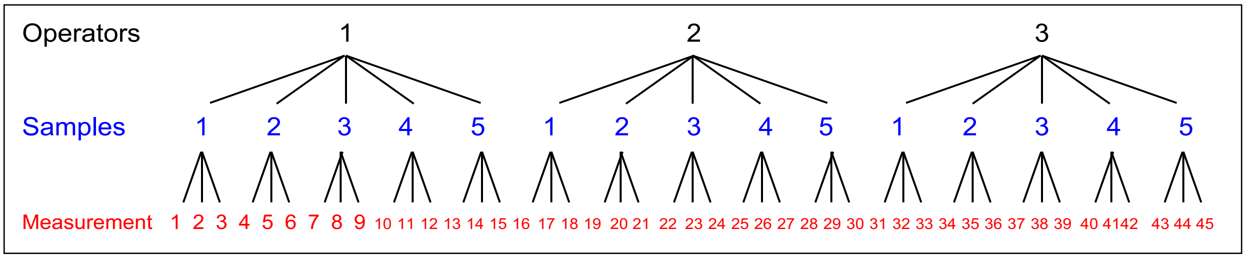


Figure 6.11: Typical Gage R&R Sampling plan

Problems with application of traditional Gage R&R studies[19]:

1. Biased to the measurement system components. Multiple layers are components of measurement system sources compared to one source for product variation.,
2. How are the samples in the study chosen? What do they represent? What do you want to compare the measurement system variation to? The samples need to represent variation as a function of the hypotheses.,
3. "One shot" studies where results are applied beyond the inference space. Typically, once the gage R&R is "passed", it is never re-evaluated. Unfortunately, as hypotheses change, so may measurement capability.,
4. The plan is a mixed model (crossed and nested), making it difficult to analyze for stability. Only nested layers can be analyzed for stability.

Experienced engineers often record COV lessons in their User Guides: where early samples misled them, warning signs of inadequate coverage or narrow inference space, measurement issues, and reminders to test stability before comparison. These notes are rarely found in textbooks. They are learned through experience. Capturing them accelerates judgment the next time uncertainty appears.
COV does not replace DOE. It prepares you for it and assesses stability after it. Before deliberately changing a system, it is wise to know: where leverage resides, whether stability exists, which components deserve focused attention, and which can safely be ignored. That preparation is the role of a COV study. Once sampling has oriented you, once you know where variation lives and where it does not, you are ready to ask a different class of questions: *How should variation be partitioned, compared, and understood?* For example, if there is significant material lot variation and only one lot of material is used when running the DOE, the results of the DOE will be unusable in the future. The next chapter discusses DOE. I have put together an overly simplified comparison in Figure 6.12.

[19] See Wheeler, Don, "*Problems With Gauge R&R Studies*", 46th Annual Quality Congress

COV	DOE
Understand an existing phenomena or rare events	Identify a better design or solution*
Explain	Predict
Low on knowledge cont.	Higher on knowledge cont.
Broader inference space	Narrower Inference
General hypotheses	More specific hypotheses
Lots of factors (x's)	≤15 Factors (x's)
Nature & leverage (components)	Effect magnitude, not nature (stability)
Ideas to handle noise	Noise strategies
Understand level setting	Bold, but reasonable
More time consuming	Faster
Less resource intensive	More resource intensive
Nested/systematic	Main and Crossed effects
Clues about model effects	1st & 2nd order+ polynomials
	*Note: DOE can assist in understanding an existing phenomena, but this imposes restrictions on factor selection and level setting

Figure 6.12: Sampling and DOE Comparison

Chapter 7
Design of Experiments (DOE)

Dr. George Box introduced me to a simple model for learning: In order to learn, two things must occur simultaneously: something must happen (a significant informative event) and someone must see it happen (a perceptive observer). Significant events occur naturally, all of the time. It is happenstance when the event and the perceptive observer meet. Consider gravity. Events associated with gravity have been around as long as humans, but it wasn't until Newton, hit on the head with a falling apple, that gravity was discovered. Rather than waiting for a significant event to occur and increasing the likelihood we will see it (through sampling and critical thinking), create the significant events (DOE) and measure them, thus obtaining knowledge faster. That is basically what we attempt to accomplish with a DOE. The bigger the event, the faster the learning. An experiment is an invitation for an informative event to occur. DOE, by definition, is the manipulation of controllable factors (x's, independent variables) at different levels to see their effect on responses (Y's or y's, dependent variables) **in the face of noise**. Where, **controllable** factors are the factors we are willing to manage or control and **Noise** are the factors we are NOT willing/able to manage or control for various reasons:

1. there is no current methodology/technology available to control,
2. management or control is too costly, or
3. management or control is inconvenient

I will not re-state the arguments against one-factor-at-a-time experiments (OFAT). Their issues including inefficiency, narrow inference space, suboptimization, unrealistic and inability to identify interaction effects is well documented. DOE is the most powerful tool we have for understanding causality. We are literally manipulating factors at multiple levels while other factors are also being manipulated and while noise is changing. When used well, they allow us to deliberately vary inputs and observe the resulting changes in performance in a way that cannot be achieved through observation alone. When used poorly, they consume time and resources while providing little insight, and sometimes misleading conclusions.

> *"Experience may be acquired in two ways; either, first by noticing facts without any attempt to influence the frequency of their occurrence or to vary the circumstances under which they occur; this is observation; or, secondly, by putting in action causes or agents over which we have control, and purposely varying their combinations, and noticing what effects take place; this is experiment."*
>
> *Sir John Herschel*

Learning vs. "Pick a Winner" Selection

In many organizations, experiments are run under intense schedule pressure. The objective is often to "pick a winner" rather than to understand how and why a system behaves as it does. Factors are selected based on prior beliefs, or "rounding up the usual suspects". Levels are chosen based on what is the current level and what is thought to be optimal, and the experiment becomes an exercise in justification rather than discovery. While this may satisfy organizational timelines, this approach may sabotage learning. When only favored factors are included, and when factor levels are set narrowly around current practice, the experiment severely restricts the design space. If effects are underestimated due to insufficient level separation, important factors are dropped prematurely, creating a **Type II error (Beta)**, which is often the most damaging error in an iterative learning process.

There is a fundamental distinction between experiments run to **understand causality** and experiments run to **select among options**. Learning experiments assume incomplete knowledge, intentionally explore broad design spaces, sacrifice efficiency early to gain insight, and prioritize inference space over immediacy. This distinction matters. An experiment designed to learn embraces uncertainty, breadth, and iteration. It is expected to inform the *next* experiment, not to finalize a design. This book is concerned primarily with using experiments to gain knowledge.

Level Setting

Two of the most influential elements of effective DOE planning are the handling of noise and level setting. In sampling, the time series is used to ensure the natural variation of factors is exposed. In designed experiments, the time series is purposefully removed. To expose the influence of factors, the factors have to be manipulated at multiple levels. The general guidance for initial experiments is: **bold, but reasonable**. For example, when experimenting on materials to use in a product where the hardness of the material may have an effect, the experimenter may choose two different grades of stainless steel, but would not consider making it out of platinum. In early experiments, factor levels should be set to create sufficient contrast to reveal potential factor effect. The objective is not to identify the best setting, but to determine whether the factor matters and to estimate the direction and relative magnitude of its influence. If levels are set too close together, important factors may appear insignificant simply because the study did not reveal enough variation. A common mistake is to select one level as the current setting and the other as the engineer's best estimate of optimum. Under typical time pressures, this approach often fails to create enough separation to reveal meaningful effects or to overcome the influence of noise. Because exaggerated levels move the study away from the current operating region, **all factors in the experiment should be set boldly**, not just a select few. This ensures that the experimental region remains balanced and interpretable. Noise should also be varied broadly when the goal is to create an experiment space indicative of the future. If noise is artificially constrained, factor effects may appear larger than they will be under real operating conditions, reducing the usefulness of the results. Initial experiments are intended to **expose effects**, not to select optimal settings. Subsequent experiments can then refine level selection and move toward practical operating conditions. Sampling studies can often provide valuable guidance for appropriate level ranges, particularly for understanding the natural variation of noise factors.

Experiments Designed to Provide Insight

An experiment that cannot provide insight to hypotheses and what you will do next is not useful. Before running an experiment, engineers should be able to answer:

- What decisions might change based on the results?
- What will we do if the results contradict expectations?

- What will we do if no effects are detected?
- What will we do if there is a huge amount of variation, but no factors are significant

These questions should be predicted. If these questions cannot be answered, the experiment is premature.

Visual Tool: Factor Relationship Diagrams (FRD)

FRD is to DOE as sampling trees are to COV. FRD's help visualize relationships among factors and noise (Figure 7.1). They are used during planning to decide when noise should be held constant, allowed to vary, or be deliberately partitioned. They are also used during analysis to interpret results in the context of those decisions (what are the appropriate comparisons to make). An FRD should have **design structure (DS)**, **unit structure (US)**, at least one **line of restriction (LOR)**[20] and the **model** to assign the degrees of freedom (DF's). All FRD's will have a row for treatments[21] and actual run order.

I. The **DS** is the factors being deliberately manipulating (i.e., controllable x's). The design factors are shown in **black** and are coded equidistant, centered on zero. (e.g., -1, +1 for 2-level designs) to indicate the number of levels.

II. The **US** is the noise related to the experiment. It is always **red** and coded using whole numbers (unless it is manipulated, then it follows the coding for **DS**) to indicate one or multiple levels. It distinguishes:

1. Noise factors held constant during the experiment. Conditions under which the experiment is run (i.e., inference space)., **AND**
2. Noise factors that vary but are not deliberately manipulated during the running of the experiment. This noise forms the basis for statistical tests (F-test).

III. LOR's are shown in **green** and are used to show:

1. Partitioning of the **US**,
2. Partitioning of DF's for analysis

Solid **LOR's** partition **both US** and DF's whereas dashed **LOR's** partition only the **US**.

IV. The **Model** is shown in **blue** and should assign the DF's with the appropriate **LOR's**.

---

[20] The phrase comes from restrictions on randomization. There is always something constant during the study.
[21] Treatments = Experimental Units

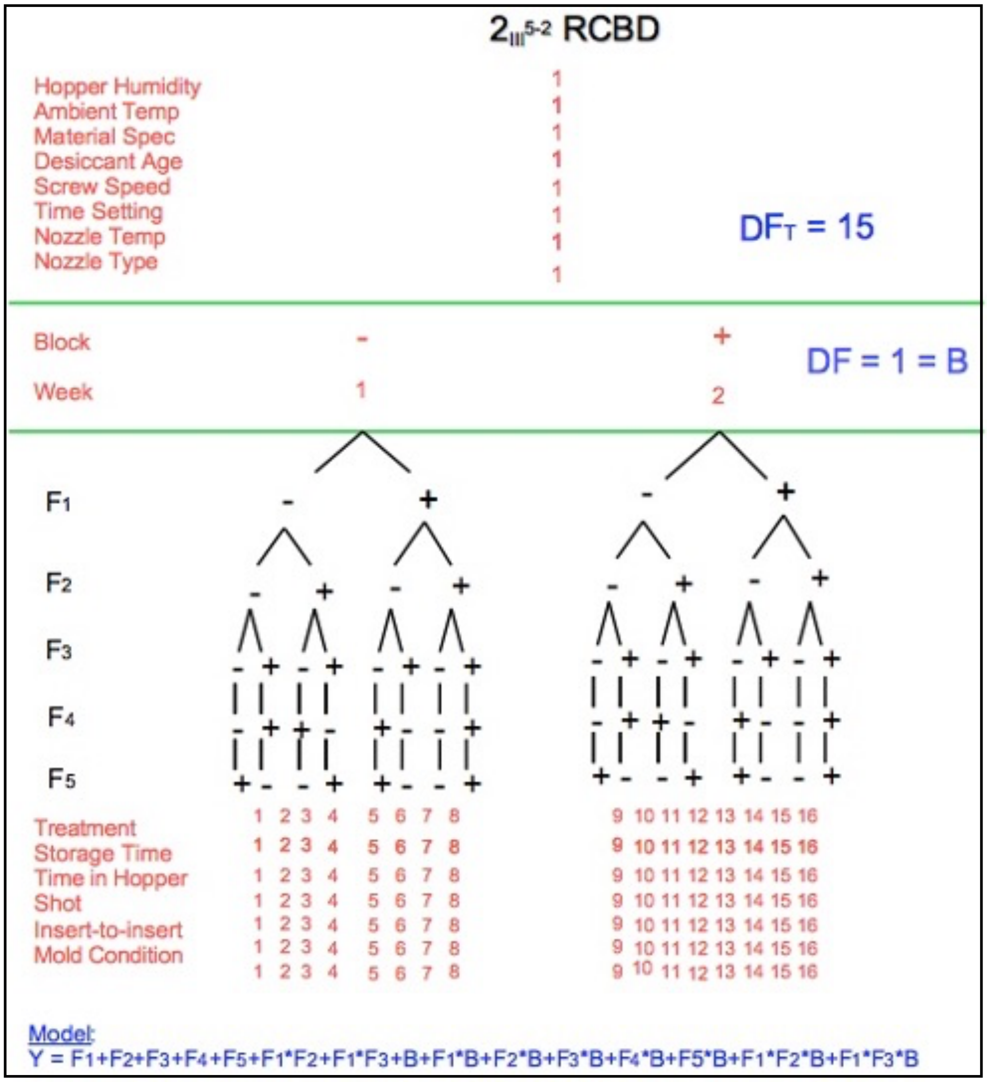


Figure 7.1: Example Factor Relationship Diagram for Injection Molding DOE

Rank Order of Model Effects

A central planning concept in this methodology is the **predicted rank order of model effects**. Before selecting an experimental design, engineers should predict which effects, main effects and two-factor interactions at a minimum, are most likely to matter. This prediction is not about correctness; it is about **design selection**. This is not a statistical exercise; it is an expression of engineering judgment. If first-order effects are expected to dominate, lower-resolution designs may be appropriate and efficient. As interactions rise in the predicted rank order, higher-resolution designs are required to avoid confounding effects that are believed to be important. Rank order prediction directly informs design resolution, alias structure acceptability, and the balance between resource use and learning. Skipping this step often leads to designs that are either overbuilt or insufficiently informative.

Sequential Experimentation: Start Wide, Then Focus

A recurring principle in this methodology is to **start wide and narrow later**. Most meaningful experimentation is sequential. Rarely do we begin near the optimum. Early experiments are exploratory, intended to reveal direction, magnitude, and relevance. Later experiments refine understanding, increase resolution, and address nonlinear behavior. I often describe this sequential process using the analogy of traveling from Philadelphia to Boulder. I was born in Philly. I had one objective since birth, get out of Philly. When I turned 17, I graduated high school and got a *get out of Philly* opportunity. I was able to select a location for my continuing studies. I set my sights on Boulder, CO. If my goal is to leave Philadelphia quickly, I do not need a detailed street map of center city Philadelphia. I need the fastest way out of the city. As I got closer to Boulder, I realized I needed the detailed map. Only near the end does it make sense to focus on specific streets and addresses.

Early experiments should include many factors, use bold level settings, tolerate confounding strategically for efficiency, and aim to identify direction rather than precision. As learning accumulates, subsequent experiments increase resolution, explore nonlinearity, and refine models. This mirrors how humans learn and aligns with effect sparsity and hierarchy principles. The mistake is not starting with low resolution, the mistake is thinking you can get all the information you need in one experiment and therefore running an optimization experiment in Philly. Assume you do not know much. This assumption is appropriate because, if the assumption is right, you have run experiments to provide the proper foundation for iterative work. If the assumption is wrong, you haven't wasted resources, you simply augment the current design space.

Design Selection Criteria

The following is a list of criteria to aid in experimental design selection.

1. **Constraints**: Time, money, material availability, measurement/equipment capability, etc. How many treatments can be made (this will likely need to be negotiated)?
2. How many **factors** are to be manipulated (the number of hypotheses to be compared)?
3. **How will noise be included?** (e.g., repeats, replicates)
4. Are there purposeful **restrictions on randomization**? Are some factors harder to change than others?
5. What **global effects**[22] do you want to estimate?
6. Are higher order **model effects** suspected/predicted (i.e., interactions, curvature)?
    - What is the desired **resolution**? (What interaction effects do you want to estimate/separate?)
    - What order **polynomial** is required? (How many levels do the factors need to be tested at?)

[22] See Chapter 9, Figure 9.1, page I-83

Chapter 8
Noise, Robustness, and Restrictions on Randomization

Perhaps the most important distinction between my approach and that of traditional enumerative statisticians lies in how noise is treated. The prevailing statistical instinct is to randomize noise *away*. While this may be useful for certain enumerative objectives, it fundamentally limits learning about causal structure. Randomization hides noise; it does not explain it. In engineering and scientific work aimed at understanding *how and why systems behave*, noise must be **studied**, not ignored.

> *"The exact standardization of experimental conditions, which is often thoughtlessly advocated as a panacea, always carries with it the real disadvantage that a highly standardized experiment supplies direct information only in respect to the narrow range of conditions achieved by the standardization. <u>Standardization, therefore, weakens rather than strengthens our ground for inferring a like result</u>, when, as is invariably the case in practice, these conditions are somewhat varied."*
>
> *R. A. Fisher (1935), Design of Experiments (p.99-100)*

What Noise Really Is

Noise consists of all the factors the investigator is unwilling to manage. Note the future tense, most noise may be able to be *controlled* over the course of an experiment. Recall, these factors may be uncontrollable, impractical or too costly to control, not of immediate interest or inconvenient to control. In any realistic system, the number of noise factors vastly exceeds the number of factors that can be explicitly studied. This imbalance creates a dilemma: how do we learn about important design factors while the rest of the system continues to vary? Ignoring noise is not an option. But neither is holding it constant. Noise affects experiments in two ways; it limits the precision for detecting factor effects and it determines the inference space over which conclusions remain valid. Examples of noise include environmental conditions, wear, raw materials, operator behavior, customer use, source of power, etc.

The Lake-Level Analogy

I often explain noise using a simple analogy. Imagine the effects of design factors as objects protruding above the surface of a lake (Figure 8.1). The water level represents the variation due to noise. Only the effects that rise above the water level can be detected. If the amount of noise that varies in the experiment increases, the water level rises, and smaller effects disappear beneath the surface. If noise decreases, more effects become visible, but the inference space shrinks. Precision is the term we use to describe the ability of the experiment to detect factors. The objective is not to drain the lake. The objective is to **understand how deep it is and why**. The optimum water level is the level that represents the true variation due to noise in the future.

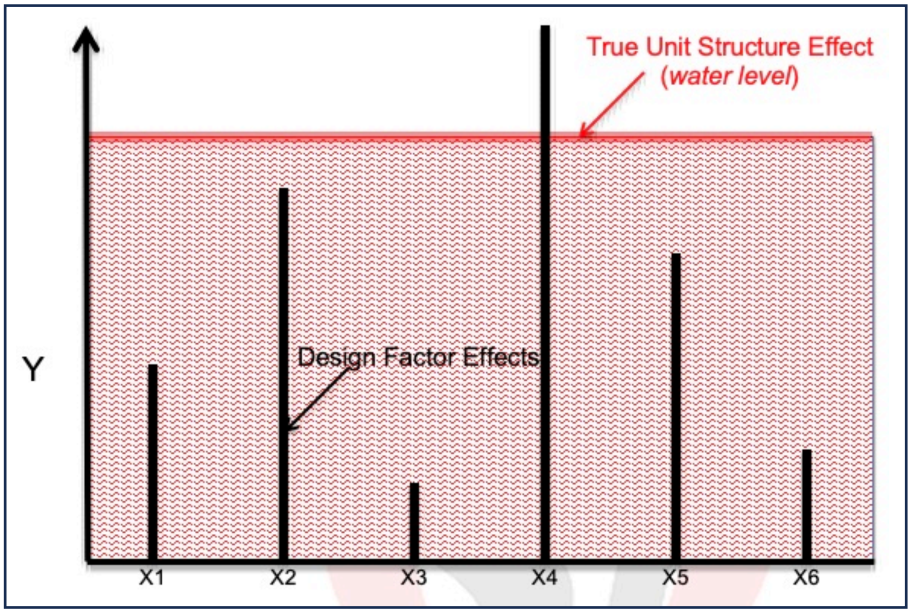


Figure 8.1: Graphic Representation of Water Level Analogy

The Cost of Holding Noise Constant

A common strategy, often unexamined, is to reduce noise by holding it constant. While this increases precision, it also dramatically restricts inference. An experiment conducted under tightly controlled conditions may yield precise estimates that are useless tomorrow, see figure 8.2. The system outside the laboratory does not behave that way. Holding noise constant trades short-term clarity for long-term irrelevance.

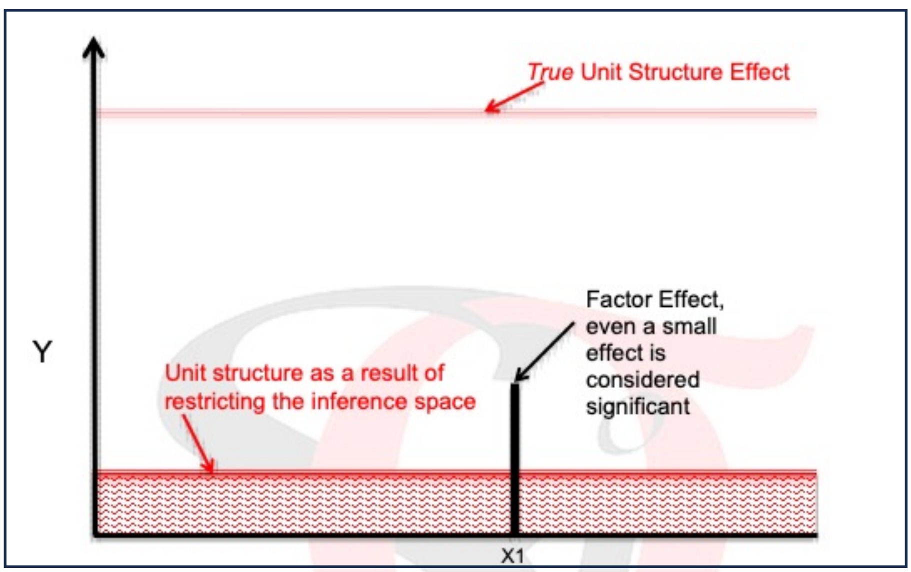


Figure 8.2: Depiction of Restricted Inference Space

Randomization has a role. It protects against bias and supports certain enumerative objectives. But randomization alone cannot reveal causal structure. If noise is always randomized, it is never understood.

Recall Shewhart's model; random variation is unassignable. Engineering progress requires more than statistical protection, it requires insight.

Robustness Defined

Robustness is often presented as a design technique. It is better understood as an **analytic obligation**. If a conclusion or decision is highly sensitive to small, uncontrolled changes, it is fragile, regardless of how compelling the analysis appears. Robust conclusions are those that continue to hold when the system is exposed to reasonable variation. I define robustness as **the absence of noise-by-factor interactions**. If the effect of a design factor depends on noise, the design is not robust. Noise-by-factor interactions are not inconveniences; they are opportunities. They tell us that performance depends on conditions. Identifying these interactions early, during the design phase, is critical, because this is when options are plentiful. Once a design is released, options disappear.

Partitioning Noise to Learn about It

The challenge, then, is to increase inference space while preserving precision. This is not only possible, it is essential. Noise can be deliberately partitioned through:

- **repeats**, to understand short-term variation,
- **blocking**, to account for long-term or systematic variation,
- **split-plot designs**, to manage specific noise efficiently.

These are not statistical tricks. They are learning strategies. Noise is not handled by statistical software. It is handled during **experimental planning**. The Factor Relationship Diagrams (FRDs) introduce in chapter 7, make noise visible. Noise placed at the bottom of an FRD defines the water level in the lake. As that list grows, it signals a need to partition variation, before data are collected. The software never asks for this information. Yet that is what determines the validity of every p-value and confidence interval produced.

Repeats

When ***repetition*** is used, the treatment combinations **are not** changed between the data points. Results obtained by repetition should **not** be considered independent events (there is only 1 experimental unit per treatment, that experimental unit is measured multiple times). The data points are useful for estimating averages and short-term variation components and may therefore provide estimates of mean and dispersion for use in analysis of the treatment effects. When the repeated data points are averaged, the effect of the within treatment variation (due to unit structure) is reduced increasing the precision of the design (ability to assign design factor effects). The measure of dispersion quantifies the unit structure within treatment and can be used to model the design structure effect on variation. The way we keep track of this partitioning of the unit structure is with a dashed line of restriction on an FRD (Figure 8.3).

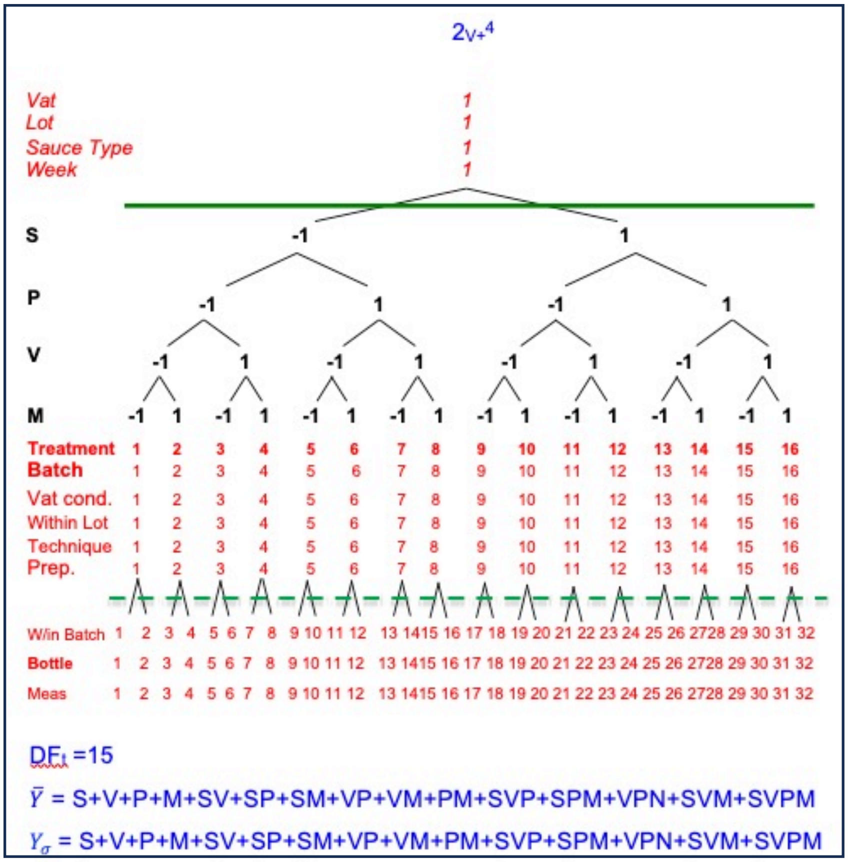


Figure 8.3: FRD depicting repeats

Replication: Completely Randomized Replicates

When ***replication*** is used, the treatments **are** changed between experimental units (Figure 8.4). Recognize this doubles the size of the experiment. Traditionally, statisticians recommend randomized replicates. When replicating the experiment, each treatment results in multiple experimental units. Each experimental unit can be considered an independent event. This is accomplished by changing the experimental conditions between each treatment. The advantages are it increases the likelihood the inference space is more representative (broader), the unit structure effects, estimated by the error term, are theoretically less biased, and the error term can be quantified and used to test statistical significance of the design factors. Essentially, the error term quantifies the water level in the experiment. Unfortunately, the error cannot be assigned to specific x's and the precision for detecting design factor effects may be compromised (the water level is higher).

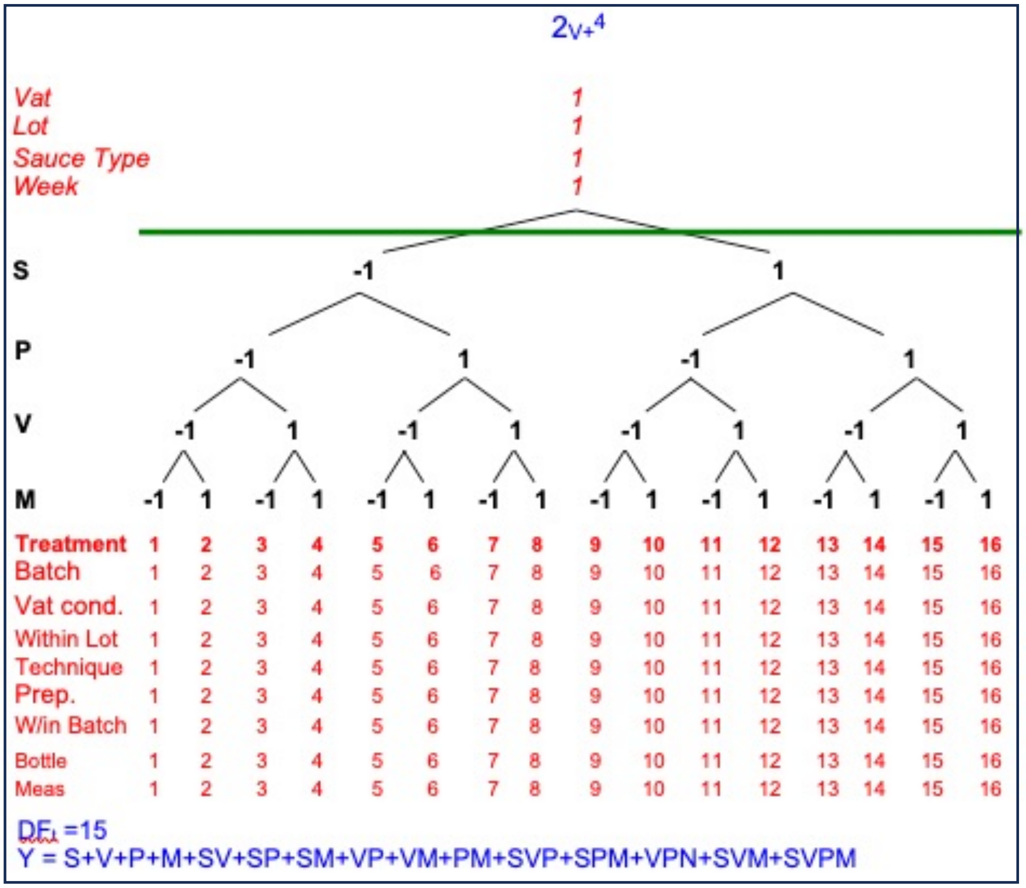


Figure 8.4: FRD depicting randomized replicates

Blocking: Randomized Complete Block Designs

Replication can also be accomplished in blocks allowing for the assignment and partitioning of the unit structure (Figure 8.5). In industrial experimentation, blocks are frequently a frame where **noise** variables can reasonably be expected to remain constant or are held constant while that part of the experiment takes place. This has the benefit of reducing the effect of noise and increasing the precision for detecting factor effects within the block (e.g., dropping the water level). Subsequent blocks are selected so the noise held constant *within* block changes *between* blocks. This allows for estimation of the effect of noise and often **noise-by-factor** interactions necessary for robust design. In this manner information regarding **design structure** (factors manipulated) is acquired across changing unit structure (environmental conditions, variation in raw materials, and other known and unknown noise variables) therefore **increasing inference space** with **increased precision**.

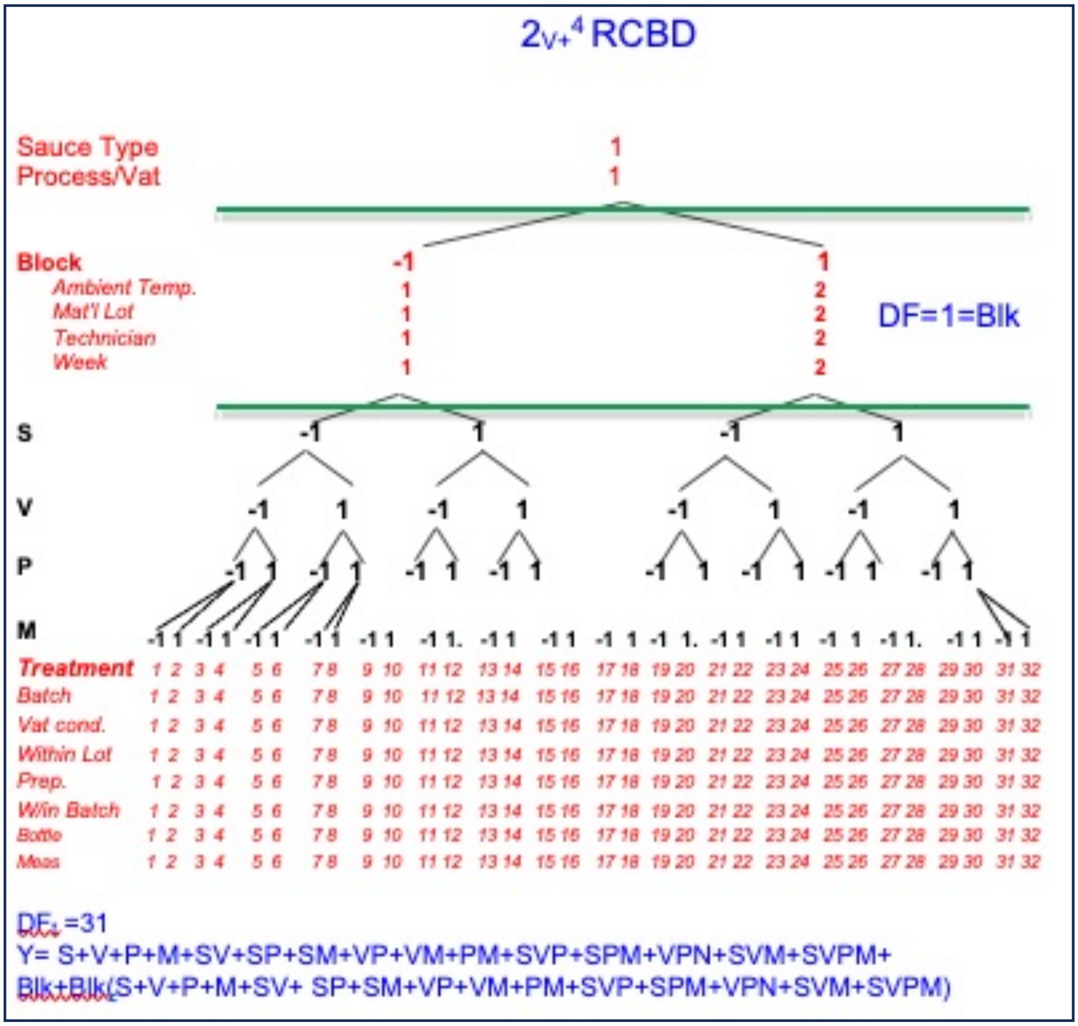


Figure 8.5: FRD depicting blocks

It is often the case you might make a decision to reduce the resolution of the design structure in order to increase the resolution of the unit structure. This is accomplished by fractionating the design structure. You might also fractionate the block effects (Figure 8.6), creating a randomized incomplete block design (aka balanced incomplete block design).

"Block what you can, randomize what you cannot"

G.E.P. Box

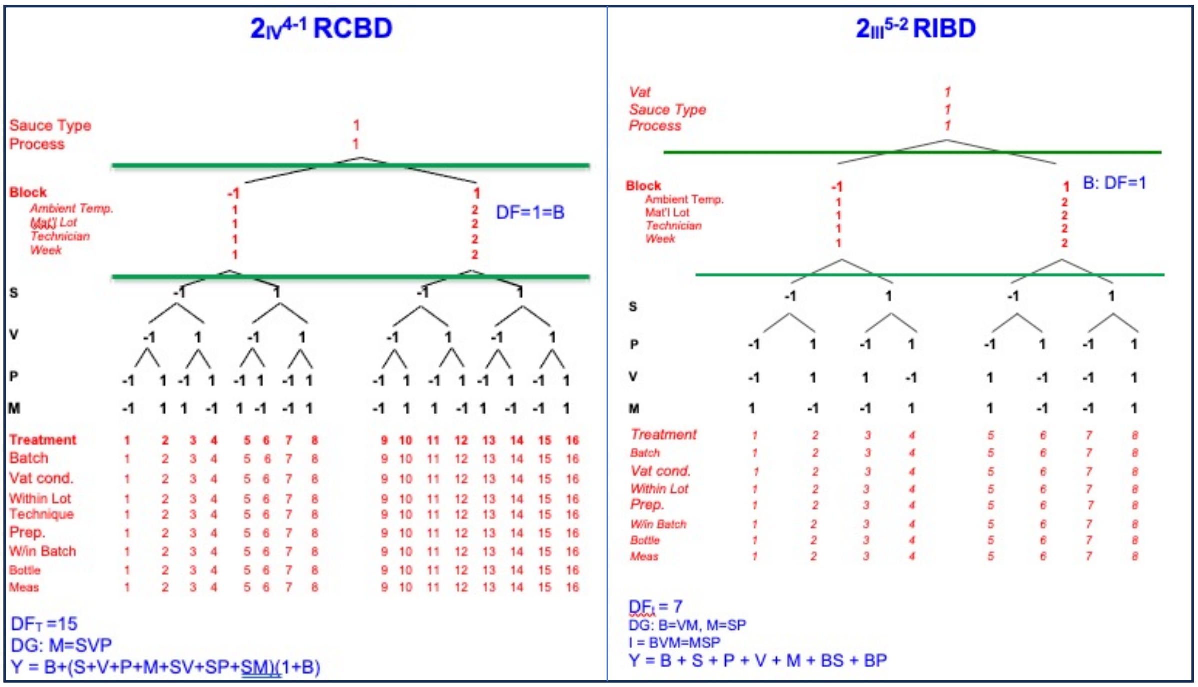


Figure 8.6: RCBD Fractional Factorial and an Incomplete Block

Figure 8.7 depicts the effect of blocking on design factor precision.

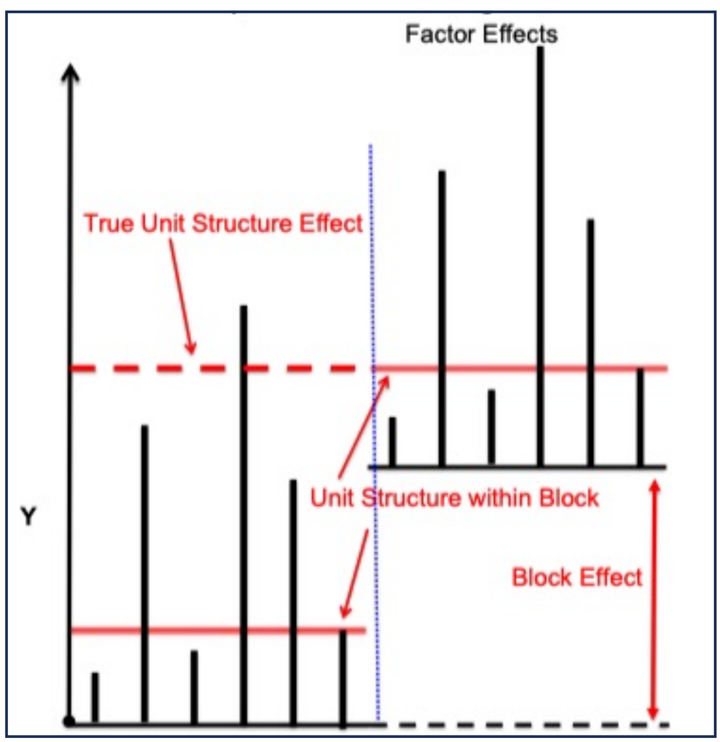


Figure 8.7: Depiction of a Significant Block Effect

Split-plots

Split-plot experiments can be seen as a practical way to deal with certain issues which preclude randomization. Completely randomized designs are the archetype, often advocated, design. This is the typical strategy to handle noise in an experiment **when the noise has not been identified**. When a design is not randomized, those restrictions must be taken into account. One such design used to handle restrictions is a split-plot design. There are two distinctly different reasons to restrict randomization and I make a deliberate distinction between **convenience split-plots** and **efficiency split-plots**.

Efficiency Split-plot

The second reason to use split-plot designs is due to the desire to partition the unit structure (lower the water level) and thereby **improve** the precision of the design structure using less resources. This is what I call an efficiency split-plot (Figure 8.8). Efficiency split-plots are designed intentionally to partition noise in a way that increases precision of both the whole plot and the subplot, preserves inference space, and allows meaningful interpretation of effects. This distinction is rarely made explicit, yet it fundamentally changes how experiments are planned and analyzed. To my knowledge these can only be run in situations where the experiment can be run in stages or in some time ordered sequence. The factors that make up the subplot are subordinate or subsequent to the factors in the whole plot. This can be extremely useful for product design as different designs can be tested over factorials of noise. Doing this increases the precision of understanding noise-by-factor interactions and therefore the robustness of the design to noise. This is essentially what Taguchi's inner-outer array or Cox's cross-over design is.

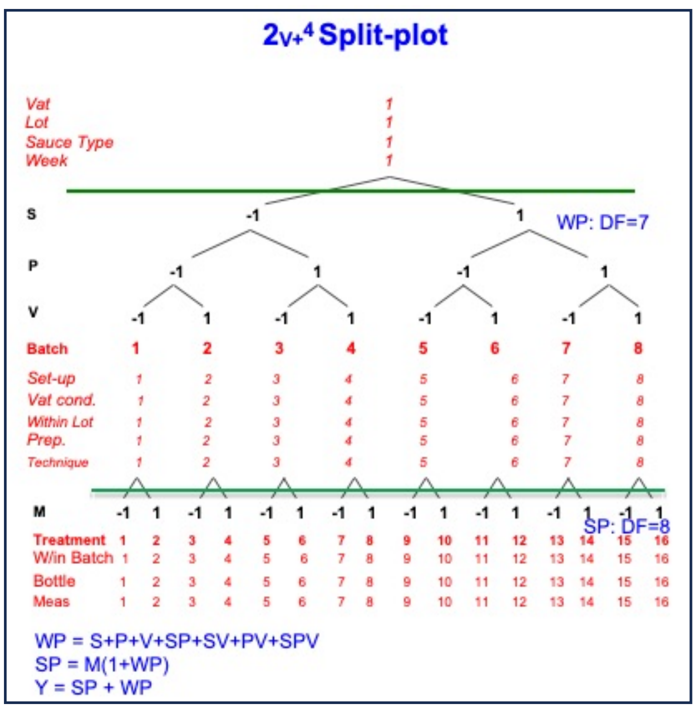


Figure 8.8: Efficiency Split-plot Design

Convenience Split-plots

An experiment that restricts randomization due to physical or economic reasons is what I call a convenience split-plot (Figure 8.9). Most split-plot designs are created for convenience, factors that are hard or costly to change. The **run order** is executed by changing the subplot (SP) factor(s) while the whole plot (WP) factor remains constant. Then the WP factor is changed and the runs replicated. This minimizes the number of times the whole plot factor is changed. the **information about the whole plot factor** is subject to *different* unit structure due to the minimized number of changes. For quantitative analysis, the whole plot error (rather than the mean square error of the subplot) should be used to determine whether the whole plot factor had a statistically significant effect. **Replication** of the treatments made varying the whole plot factor is necessary to have any estimate of the whole plot error and therefore a **quantitative** test of significance which means you have to change the whole plot factor more than once. If the unit structure is significantly partitioned and it is confounded with the whole plot factor, there will be a negative effect on the precision of the whole plot. In this case, it would NOT be desired to have a significant partitioning of the unit structure thereby not diminishing the precision of the whole plot.

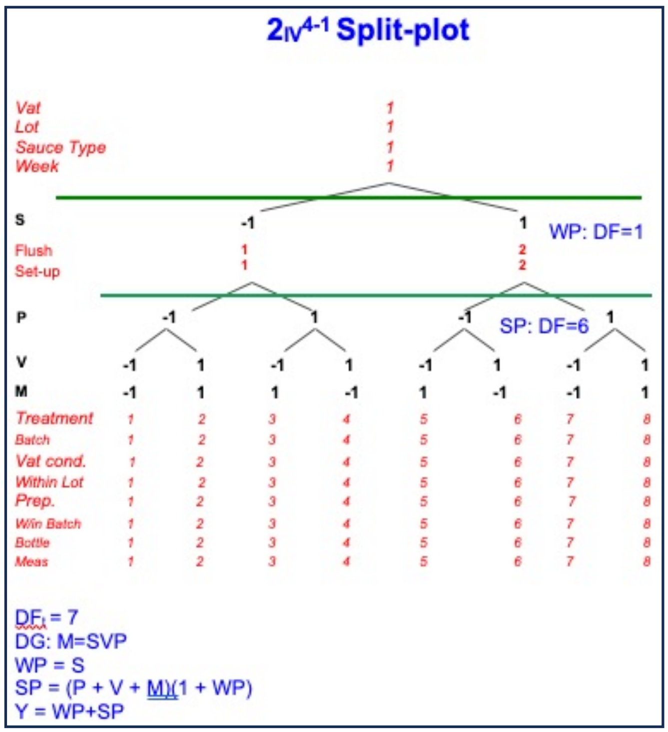


Figure 8.9: Convenience Split-plot

Options for Significant Noise Effect

Once noise (or noise-by-factor interactions) is found to have a significant effect, what can be done to mitigate the effect?

- Optimize levels of the controllable factors that reduce the noise effect,
- Discover other factors (e.g., not previously considered) that will mitigate the noise,
- Manage the noise (essentially consider the noise as a controllable factor and find best levels),

- Continuously adjust controllable factors to compensate for the noise effect (the noise would need to be measurable),
- Work to reduce the noise effect directly (e.g., work on the supplied/raw material processes),
- Create/develop a new process (e.g., innovative materials and technology),
- Accept the process will be producing product that will perform inconsistently and prepare to handle customer complaints.

Transition: From Noise to Analysis

Once noise has been deliberately accounted for in the design, analysis becomes clearer, simpler, and more meaningful. In the next chapter, we turn to **analysis**, emphasizing practical interpretation, graphical methods, and quantitative tools, applied in that order, to extract learning rather than merely produce statistics.

Chapter 9
Analysis for Learning

Analysis is often where engineering and scientific studies go wrong, not because the tools are incorrect, but because the **purpose of analysis is misunderstood**. Analysis is not an end point. It is not a ritual. It is not proof of rigor. Analysis exists for one reason only: **to support learning about causal structure**. Everything that analysis can reveal is constrained by how the data were acquired. No amount of analytical sophistication can recover information that was never collected.

> "*There are three kinds of lies: lies, damned lies, and statistics*."[23]

The Governing Constraint: Data Acquisition

The most important principle in analysis is this: **The information revealed by data is entirely dependent on how the data were acquired.** Before looking at plots, tables, or p-values, the analyst must understand:

- what varied,
- what was held constant,
- what is assignable,
- what was confounded,
- what was randomized,
- and what sources of noise varied during the study.

Without this context, analysis is disconnected from reality. This is why skilled practitioners spend far more time planning data collection than performing analysis. Analysis is downstream of design, and its value is bounded by it.

> "*To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.*"[24]

Ross' Rules of Analysis

Over time, I formalized a simple sequence for analysis that reflects how engineers actually learn. I refer to this as **Ross' Rules of Analysis**:

1. **Practical**
2. **Graphical**
3. **Quantitative**

This order is deliberate and non-negotiable. I learned the required order of analysis the hard way.

[23] Benjamin Disraeli, also attributed to Mark Twain.
[24] Fisher RA (1938) Presidential address. Sankhyā 4: 14–17.

As an advisory engineer working at StorageTek, I had the opportunity to design, review and analyze many experiments. In this role, I would assist in analysis by running procedures in SAS (e.g., proc GLM, Proc ANOVA, proc REG). I would provide a detailed quantitative analysis of the data. I would also support other engineers in their explanations of what was learned to others. At one such meeting, an engineer had finished their presentation on a project they were working on and asked if there were any questions. Sure enough, an individual in the meeting raised their hand. This individual was a manager and so I prepared for a less than challenging question. The manager said, I have no idea what ANOVA is or what a significant p-value is, but looking at the data, the amount of variation in the experiment is irrelevant. The response is nowhere near where we need to be. I considered his comment and I had an emotional event. It was then and there I realized I had been analyzing data backwards. Upon this realization I wrote a note to myself in my User Guide regarding the order of analysis.

Practical Analysis: Does This Make Sense?

Practical analysis asks the most basic, and most important questions:

- Does the data make sense?
- Is the magnitude of change meaningful?
- Is there enough variation to continue analyzing?
- Are results consistent with engineering intuition?
- Are the results supportive of hypotheses (or can you rationally explain the results?)
- How do results compare with predictions?
- Are there any obvious outliers or interesting patterns?

Prediction plays a crucial role here. When predictions were made before data collection, discrepancies become learning opportunities rather than sources of confusion.
Practical analysis is where engineers apply judgment. Statistical significance without practical significance is irrelevant.

Graphical Analysis: Seeing the Patterns

Humans are visual learners. Graphical analysis exploits this strength. Common graphical tools include:

- variability plots,
- time series plots,
- control charts,
- normal probability plots,
- Pareto chart of effects,
- scatter plots.

Variability plots for sampling studies or nested layers in an experiment are particularly powerful because they make structure visible without requiring distributional assumptions. They allow engineers to see:

- separation between groups,
- a comparison of each of the components,

- patterns due to systematic sampling,
- patterns suggesting interactions or instability.

I also use **ANOG (Analysis of Good)** as a practical counterbalance to ANOVA. Rather than asking whether differences are statistically detectable, ANOG asks whether the "good" (and "bad") outcomes are associated with specific levels of factors. This aligns analysis with engineering objectives.

> *"Results of a well-planned experiment are often evident using simple graphical analysis. However, the world's best statistical analysis cannot rescue a poorly planned experimental program."*
>
> *G. Hahn*

Quantitative Analysis: Augment, Don't Replace

Quantitative methods such as ANOVA and regression should **augment**, not replace, practical and graphical understanding. Statistical procedures are useful for detecting subtle patterns in complex data sets, preventing false pattern recognition, and increasing confidence in conclusions.

They are not substitutes for thinking. A common failure mode is to begin with quantitative output and work backward to interpretation. This reverses the learning process and encourages overconfidence in numerical results divorced from context.

Analysis of DOE vs Analysis of Existing Data

It is important to distinguish between analysis of designed experiments and analysis of existing (observational or historical) data. In DOE, model development is fundamentally **subtractive**. The experiment is designed to support a specific model. Analysis begins with a saturated model, all degrees of freedom are assigned, and terms are removed using engineering judgment, graphical evidence, and statistical analysis. In observational data analysis, modeling is most efficiently accomplished using an **additive** approach. One begins with simple models and adds complexity as warranted, often guided by Taylor series logic. Confusing these approaches leads to inappropriate modeling and false conclusions. Ignoring effects simply because they are not in the current model creates blind spots that often reappear later as "unexpected" behavior.

Model Effects vs Global Effects

Another frequent source of confusion is the distinction between **model effects** and **global effects**. Model effects are those included in a specific experimental design. I suggest rank ordering model effects up to second order during planning (this is, of course, a prediction). If the top of the list is all main effects, then lower resolution designs are appropriate. As the second orders percolate up the list, higher resolution designs become better options. This rank ordering is also useful in analysis when assigning aliased effects. **Global effects** represent the broader universe of considerations that could influence tool selection. The priority for estimating which effects are necessary, given your current knowledge, will change over the course of the investigation. Figures 9.1-9.3 are not prescriptions; they are interpretive scaffolding intended to help engineers link what they might want from a study to the most effective strategy.

Possible Effect
Noise
Main Effects (Factors)
Two-Factor Interactions (2nd order linear)
Noise-by-factor Interactions
Simple Curvature (2nd order non-linear)
Complex Non-linear (≥3rd order non-linear)
≥ Three-Factor Interactions (3rd order linear)
Stability
Leverage
Measurement Uncertainty
Mean
Variation

Figure 9.1: List of Global Effects

Ranking these effects, and updating that ranking as learning progresses, helps guide selection of sampling strategies and experiment design It also helps in iteration planning. Figure 9.2 offers some guidance on possible data collection strategies associated with estimating each global effect. Figure 9.3 is *one* possible ranking of global effects through phases of iteration.

Possible Effect	***Minimal Strategy***
Noise	RIBD, Sampling
Main Effects (Factors)	Res. III Fractional Factorial
Two-Factor Interactions (2nd order linear)	≥ Res. IV Fractional Factorial
Noise-by-factor Interactions	RCBD, Split-plots
Simple Curvature (2nd order non-linear)	Center Points
Complex Non-linear (≥ 3rd order non-linear)	RSM, ≥ 3 Level Factors
≥ Three-Factor Interactions (3rd order linear)	Full Factorial
Stability	Sampling, R, MR Charts
Leverage	Nested/Systematic Sampling
Measurement Uncertainty	Nested Layer, MSE
Mean	Any Data
Variation	Repeats

Figure 9.2: Possible Data Collection Strategies Linked to Global Effects

	Knowledge Continuum			
Possible Effect	**Low**			**High**
Noise	1	1	2	3
Main Effects (Factors)	2	1	3	4
Two-Factor Interactions (2^{nd} order linear)	3	2	1	1
Noise-by-factor Interactions	3	2	1	2
Simple Curvature (2^{nd} order non-linear)	4	3	2	1
Complex Non-linear (≥ 3^{rd} order non-linear)	5	5	5	4
≥ Three-Factor Interactions (3^{rd} order linear)	5	5	5	4
Stability	2	3	3	1
Leverage	1	2	4	4
Measurement Uncertainty	1	2	4	4
Mean	1	1	1	1
Variation	1	1	1	1

Figure 9.3: Hypothetical Ranking of Global Effects Through Iterations

Measurement Matters

Analysis cannot overcome poor measurement systems. If measurement discrimination is inadequate, real variation cannot be detected. Measurement error raises the water level of the lake, obscuring effects and misleading conclusions. Variability and control charts are invaluable for diagnosing measurement issues early. If the data cannot support learning, the correct response is not more analysis, it is better measurement. Measurement systems should be studied prior to experimentation or if not, nested within the experiment as repeats.

Simplicity Is a Feature, Not a Flaw

Simple models are not inferior models. They are often more useful. Simple models are easier to interpret, utilize, communicate, manage and guide action. Complex models that cannot be executed, explained, or trusted offer little value, regardless of statistical sophistication.

> *"A good model is an approximation, preferably easy to use, that captures the **essential features** of the studied phenomenon and produces procedures that are **robust** to likely deviations from assumptions"*
>
> *G.E.P. Box*

The interpretation of analytical outputs is inseparable from the structure of the data that produced them. In Part II, I examine how different data acquisition strategies shape the outputs and how engineers should interpret them.

Chapter 10
The User Guide

Engineering and scientific work aimed at understanding causality is inherently complex, iterative, and uncertain. There is no single "correct" sequence of steps, no universal decision tree, and no cookbook that can guarantee success. This reality creates a problem: how does an individual remain effective and efficient while navigating complexity repeatedly over the course of a career? My answer is the **User Guide**.

What a User Guide Is, and Is Not

A User Guide is not a textbook. It is not a set of rigid rules. It is not a standardized corporate procedure. A User Guide is a **personal, evolving representation of how *you* think about methodology**, captured in your own words, diagrams, checklists, flowcharts, and examples. It is both a reflection of what you currently understand and a mechanism for improving that understanding over time. It is your operating system for discovery.

Why Books Are Not Enough

Most books on statistics and experimentation do an excellent job explaining *how* to execute specific techniques. They describe different types of designed experiments, sampling methods, and analyses. What they do not explain is **when to use which approach**, or why. I believe this is due to the fact that techniques are always situation dependent. That decision cannot be fully prescribed. It depends on:

- the system under study,
- the questions being asked,
- the hypotheses in play,
- the resources available,
- and the engineer's current state of knowledge.

These judgments require ownership.

The First Purpose: Capturing Understanding

The first purpose of a User Guide is to document what you know, or think you know, about the methodology. Imagine not only your understanding, but your ability to explain to others. This is the true test of understanding. Early versions often contain little more than regurgitated content from classes, books, or mentors. This is normal and necessary. Over time, those notes transform into something more meaningful: a personalized guide grounded in experience. This process reveals misunderstandings, gaps, and contradictions and are signals that learning is occurring.

The Second Purpose: Guiding Action

The second purpose of a User Guide is navigation.
A User Guide may include:

- flowcharts to sequence thinking,

- checklists to ensure due diligence,
- tips, hints and tricks,
- reminders about common failure modes,
- templates for planning,
- criteria for design selection,
- links between situations and appropriate tools,
- steps of analysis,
- case studies illustrating effective and ineffective methods.

These artifacts reduce cognitive load, free mental capacity, and improve consistency without sacrificing judgment.

Iteration and Learning

A User Guide is never finished. It evolves with every project, every mistake, every surprise, and every insight. Lessons learned are captured, refined, and reused. Old assumptions are revised. New heuristics emerge. I am still working on my own User Guide after forty years. It will end when I end. An illustration of how I organize my User Guide is shown in Figure 10.1.

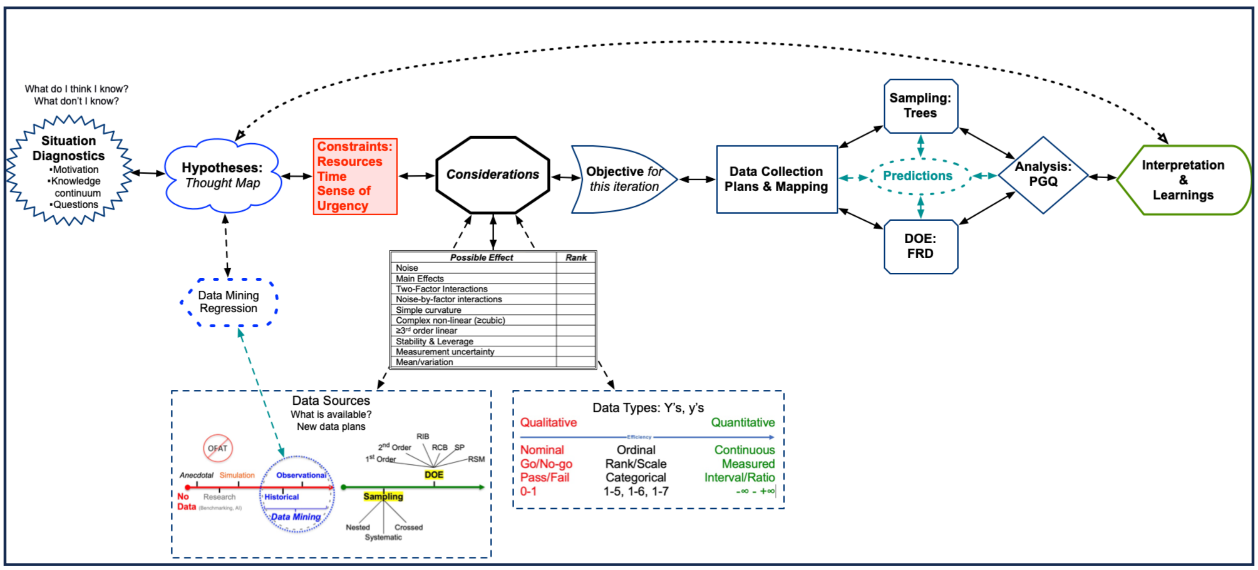


Figure 10.1: Diagram of My User Guide

Ownership Matters

Methodology that belongs to someone else may be useful, but when you internalize the methods, you are more likely to adhere to them. When guidance is self-authored, it carries power. When decisions reflect your own thinking, you are more likely to act on them, revise them, and defend them thoughtfully. This does not imply isolation. Collaboration is essential, particularly for mitigating bias. But collaboration does not replace ownership. Critical thinking is individualized and is not a team sport.

Examples Without Templates

I have collected many User Guides from past participants. They share similarities, but no two are the same. This diversity is intentional. Providing examples can inspire, but copying undermines the purpose. The goal is not to replicate my or anyone else's thinking, it is to develop your own. What works, and what doesn't, needs to be learned. The User Guide brings the methodology full circle. It connects critical thinking, mapping, sampling, experimentation, and analysis into a coherent, repeatable practice. In the final chapter, we step back and reflect on the journey, what this approach offers, what it does not promise, and how it can support a lifetime of thoughtful engineering and scientific work.

Chapter 11
Reflections, Biases, and an Invitation to Think

This book does not offer a formula. It offers a way of thinking. If there is one theme that runs through everything presented here, it is that **learning about causality requires deliberate thought, not haphazard application of tools**. Methods matter, but how and when they are applied matters more.

On Bias, Explicit and Otherwise

Everyone is biased. The danger is not bias itself, but un-recognized and **unexamined bias**. Throughout this book, I have been explicit about mine:

- a strong bias toward determinism, a belief there is no effect without cause,
- a preference for practical significance over statistical significance,
- a focus on understanding how and why systems behave,
- a skepticism of complex models that cannot be executed or used.

These biases were not adopted philosophically; they were earned through experience, including failure, embarrassment, and hard lessons. You should examine yours with equal rigor.

There Is No One Right Path

There is no single correct sequence for applying critical thinking, mapping, sampling, experimentation, and analysis. The most effective approach is often obvious only in hindsight. What I present here is **my path**, iterated over decades. It works for me. It may work for you. It will almost certainly need adaptation. What matters is not adherence to a method, but **ownership of a method**.

On Efficiency and Effectiveness

Speed is often treated as the primary metric in engineering and product development. Timelines are established before understanding what it will take to deliver. Designs are released prematurely. Problems appear downstream at great cost. Thoughtfulness early does not slow progress. It ultimately accelerates it. Taking the wrong action quickly is almost always slower than taking the right action deliberately.

Failure as a Teaching Aid

One of the most effective ways to learn is through failure. Prediction prepares you for failure by making it informative rather than emotional. When expectations are explicit, failure becomes data. Reflection turns failure into progress. This is not a comfortable process, but it is an effective one.

> *"Most large organizations embrace the idea of invention, but are not willing to suffer the string of failed experiments necessary to get there."*
>
> *Jeff Bezos*

The Invitation

If this book succeeds, it will not be because you adopt my terminology, my tools, or my preferences. It will succeed if:

- you ask better questions,
- you think more deliberately before collecting data,
- you treat noise as something to understand rather than eliminate,
- you design studies with learning in mind,
- you reflect honestly on what you thought would happen and what actually did,
- and you develop a methodology that you own.

I encourage you to create your own User Guide. Let it evolve. Let it contradict itself. Let it improve. The journey does not end here.

Closing Thought

I will close with a reminder that has guided much of my thinking:

> *"All models are wrong, some are useful."*
>
> *G.E.P. Box*

Use what is useful. Question everything else. And above all, remain thoughtful.

Part II transition

Part II shifts attention from critical thinking, mapping and data collection strategies to analysis and interpretation from data acquired through COV and DOE with a brief discussion on data mining, while retaining the same discipline of thought.

Part II
Data Collection, Analysis and Interpretation

The Sigma Science Methodology introduced in Part I is about how to think about situations. Part II is more about what to do in those situations. **Part II is organized by how data are acquired.**

The interpretation of data cannot be separated from the way the data were acquired. Sampling data, experimental data, and historical or observational data differ fundamentally in their intent, structure, and limitations. As a result, the questions they can answer, the conclusions that can be drawn, and the confidence with which those conclusions can be extended into the future are fundamentally different. Much confusion in applied statistics arises from treating these data sources as interchangeable and attempting to analyze them with the same tools and expectations. In this part of the book, analysis is approached not as a collection of statistical techniques, but as a disciplined interrogation of variation that respects the origin of the data and the hypotheses that motivated their collection.

"If your result needs a statistician, then you should design a better experiment."

Baron E. Rutherford

Chapter 12
Analysis Begins Before Software

Purpose of This Chapter

This chapter establishes a foundational premise for Part II: data analysis is dependent on how the data were acquired. This chapter is not about how to generate outputs of analysis, but how to interpret them once they exist. The same analytical technique applied to different types of data may yield similar numerical outputs, yet those outputs can differ profoundly in meaning, credibility, and usefulness. Much of the confusion surrounding statistical analysis in engineering and science arises not from misuse of mathematics, but from failure to respect the origin and structure of the data. Analysis performed without regard to how variation was generated often produces results that appear rigorous but are misleading, fragile, or irrelevant. The objective of this chapter is to demonstrate analysis techniques, to clarify what analysis is for, what it can and cannot do, and how to approach it thoughtfully.

Analysis Is Not a Phase

In many organizations, analysis is treated as a downstream activity, something that happens after data have been collected. This framing is misleading. The ability of analysis to reveal useful information is largely determined before the first data point is collected. Decisions regarding what to measure, where to measure, when to measure, and under what conditions measurements should be taken define the structure of variation available for interpretation. Analysis does not discover structure; it interprets structure that already exists. When data are acquired without clear hypotheses, without attention to representativeness, or without understanding how variation is generated, analysis becomes an exercise in rationalization rather than learning.

> *"Hunches and intuitive impressions are essential for getting the work started but it is only through the quality of numbers at the end that the truth can be told."*
>
> *Lewis Thomas*

Why Data Origin Matters

All data are not created equal. Sampling data, experimental data, and historical or observational data differ fundamentally in intent, structure, and limitation. Sampling data are collected to understand and assign sources of variation that already exist. The sampling data exposes the natural variation and are diagnostic in nature. Experimental data are created through deliberate manipulation to provide insight to hypotheses and understand causal relationships. Historical or observational data are artifacts of past operations, collected for the purpose of developing hypotheses. Treating these data sources as interchangeable and subjecting them to the same analytical expectations is a common and costly mistake. Each requires a different analytical mindset, different questions, and different standards of evidence. For this reason, Part II is organized by how data are acquired.

Analysis as Interrogation, Not Computation

Analysis is often misunderstood as computation, selecting a statistical test, pressing a button in software, and interpreting the resulting p-value or coefficient. This view confuses means with purpose. Proper analysis is an interrogation of variation. It asks:

- What variation is present?
- How is that variation structured?
- Which sources dominate?
- Which comparisons are meaningful?
- What remains unexplained?

Statistical procedures can assist in this interrogation, but they cannot replace it. Software does not know the hypotheses, the context, or the consequences of being wrong.

Practical, Graphical, Quantitative: IN THAT ORDER

Throughout this book, I advocate a disciplined sequence for analysis, which I refer to as Ross' Rules of Analysis:

1. **Practical** — View the data through the lens of engineering judgment. Does it make sense? Is the magnitude of variation meaningful? Does it align with predictions?
2. **Graphical** — Use plots and charts to reveal patterns, structure, instability, and anomalies.
3. **Quantitative** — Apply statistical procedures to augment, not replace, insight.

This sequence is not optional. Practical significance is always more important than statistical significance. Skipping directly to quantitative analysis increases the risk of misinterpretation and false confidence.

Statistical Significance Is Conditional

Statistical significance is not a property of the system being studied. It is a conditional statement based on:

- how the data were acquired,
- what variation was present,
- what was held constant,
- amount of measurement error,
- what comparisons are made,
- and what assumptions were made.

A statistically significant result may be practically irrelevant. A practically important effect may fail to reach statistical significance if noise dominates or if the measurement system lacks resolution. Understanding this conditionality is essential for responsible analysis.

Simplicity as a Restraint

Simple models are not simplistic. They are often the most useful. Simple models:

- are easier to interpret,
- are more robust to noise,
- support prediction,
- communicate well, and
- are useable.

Complexity should be added only when required by evidence and need, not curiosity. This discipline mirrors the principle that learning should proceed from simple to complex as understanding increases.

Prediction as a Measure of Understanding

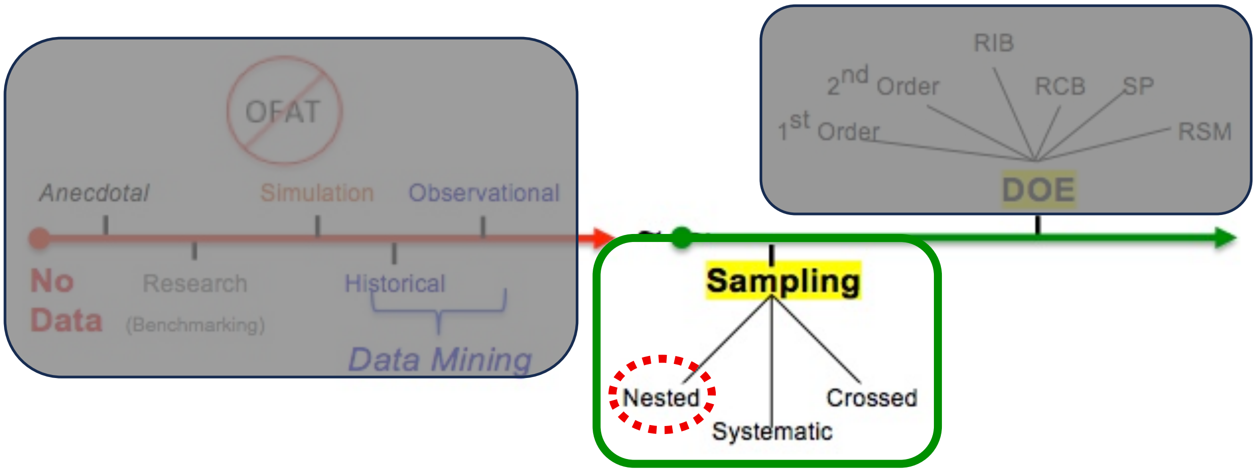
Analysis should ultimately support prediction, not explanation alone. Prediction forces clarity. It requires hypotheses to be explicit and exposes gaps in understanding. Comparing predicted outcomes to observed data is one of the most powerful ways to assess knowledge. Recall, I call this the Ross Metric. Discrepancies between prediction and observation are not failures; they are opportunities. They reveal where models are incomplete, assumptions are wrong, hypotheses are deficient, or important variables have been overlooked. Analysis that does not inform prediction rarely advances understanding.

Preparing for What Follows

The chapters that follow examine analysis through the lens of data origin: sampling data, experimental data, and historical or observational data. Each will require different questions, different analytical caution, and different humility. What remains constant is the guiding principle: analysis must respect how data were acquired. Statistical analysis is a powerful aid to learning when used thoughtfully. When used carelessly, it can create the illusion of knowledge without its substance. The distinction lies not in the tools chosen, but in the discipline with which they are applied.

Appendix F provides a detailed evaluation of potential outputs from different experimental design strategies.

Chapter 13
Analyzing Sampling Data: Stability, Leverage, and Diagnosis



Focus of This Chapter

Sampling data are collected to understand what currently exists and get clues about causal structure. Unlike experimental data, the investigator does not create events but observes variation that arises naturally over time, space, or operating conditions. The purpose of analyzing sampling data is not optimization or prediction of optima. Its purpose is orientation: to assess stability, identify dominant sources of variation, and determine where further investigation is most likely to be productive. Essentially where and how to work. This chapter focuses on how to analyze sampling data in a way that preserves context, avoids overinterpretation, and supports intelligent next steps.

Prioritized Global Effects: Below are the highlighted effects, introduced in Chapter 9, that are the focus of study.

Possible Effect	***Rank***
Noise	2
Main Effects (Factors)	2
Two-Factor Interactions (2nd order linear)	5
Noise-by-factor interactions	5
Simple curvature (2nd order non-linear)	5
Complex non-linear (≥3rd order non-linear)	6
≥ Three-Factor Interactions (3rd order linear)	6
Stability	1
Leverage	1
Measurement uncertainty	1
Mean	2
Variation	2

Sampling is Designed

Sampling data is specifically taken, plotted, summarized, and interpreted. The conclusions drawn from sampling data are entirely dependent on how the data were acquired and how variation is structured within the study. Sampling plans implicitly define which sources of variation are exposed for each layer of the tree, why are they being exposed, and which are not being considered. Analysis that ignores this structure risks answering questions that were never meaningfully posed.

Stability Comes Before Comparison

The first requirement when analyzing sampling data is to determine whether the sources of variation captured within subgroup are stable. Stability refers to the consistency of variation generated by the same underlying causes over time. If the nature of variation within subgroups changes, then comparisons across time, location, or condition are invalid. Without stability, there is no basis for comparison. Only range-based charts assess stability, because they track the same sources of variation repeatedly. Average charts and summary statistics do not perform this function. If instability is detected, the appropriate response is investigation, not adjustment. The goal is to understand what causes variation, not to smooth it away.

> *"What reliance, asks the engineer, can be placed on sampling results? The answer is that prediction based upon sampling from a non-controlled universe in which causes of the lack of control are unknown is likely always to be in error…"*
>
> *Walter Shewhart*

Sampling Data Provide Diagnostics

Sampling analysis does not tell you what to do. It tells you where and how to investigate. The role of sampling data is to:

- reveal whether variation is dominated by short-term or long-term sources (or other rational subsets),
- indicate whether differences between conditions are meaningful relative to inherent variation,
- provide confidence in extrapolating results,
- help prioritize which set of factors deserve focused investigation, and
- expand inference space.

Attempts to use sampling data to optimize performance typically exceed what the data is intended to support.

Considerations Prior to Sampling

- What questions are you trying to answer? What hypotheses are you trying to gain insight into, or compare? (reference Thought Map)
- Where & what might be hierarchical sources of variation (x's) in your process/product? (reference Process/product maps)
- What response variables (y's, Y's) are of interest?
- What and how adequate is the measurement process? What is your degree of belief in the measurement process and its outputs?
- How "representative" is the data of future conditions? Is the data sufficient and comprehensive?
- How much data will you need to capture the phenomena? Over what conditions do you want the study to be valid?
- Are there any potential data or system distortion issues?
- What resources are available? What is the sense of urgency of the effort?
- What are your predictions? How will you make them?

Understanding Leverage Through Comparison

Once stability has been established, sampling data can be used to compare sources of variation. This comparison is about assessing leverage. Leverage refers to the extent to which a source of variation contributes meaningfully to the observed variation in performance metric. Sources that dominate variation offer greater opportunity for learning and improvement than those whose effects are small relative to background variation. This comparison is inherently relative. It depends on how the study was structured, which sources of variation varied and by how much, and which were restricted (did not vary in the study). Leverage is not an intrinsic property of a factor; it is a property of a set of variables, components, in the study. Sampling analysis should stop when it has accomplished its purpose. When it has answered the questions posed and provided quantitative guidance as to which of your predicted next steps you should follow.

The following is a conceptual example of the analysis of a **nested** sampling plan. This example is intentionally detailed. It is not meant to be a template, it illustrates thinking. All possible outcomes from this study are discussed. Crazy Cajun Company has been making sauces, especially hot, for many years. They are interested in improving their product consistency and introducing a new, hotter sauce. Their objective is to identify factors that affect the consistency and hotness of the sauce. Figure 13.1 shows an initial Thought map.

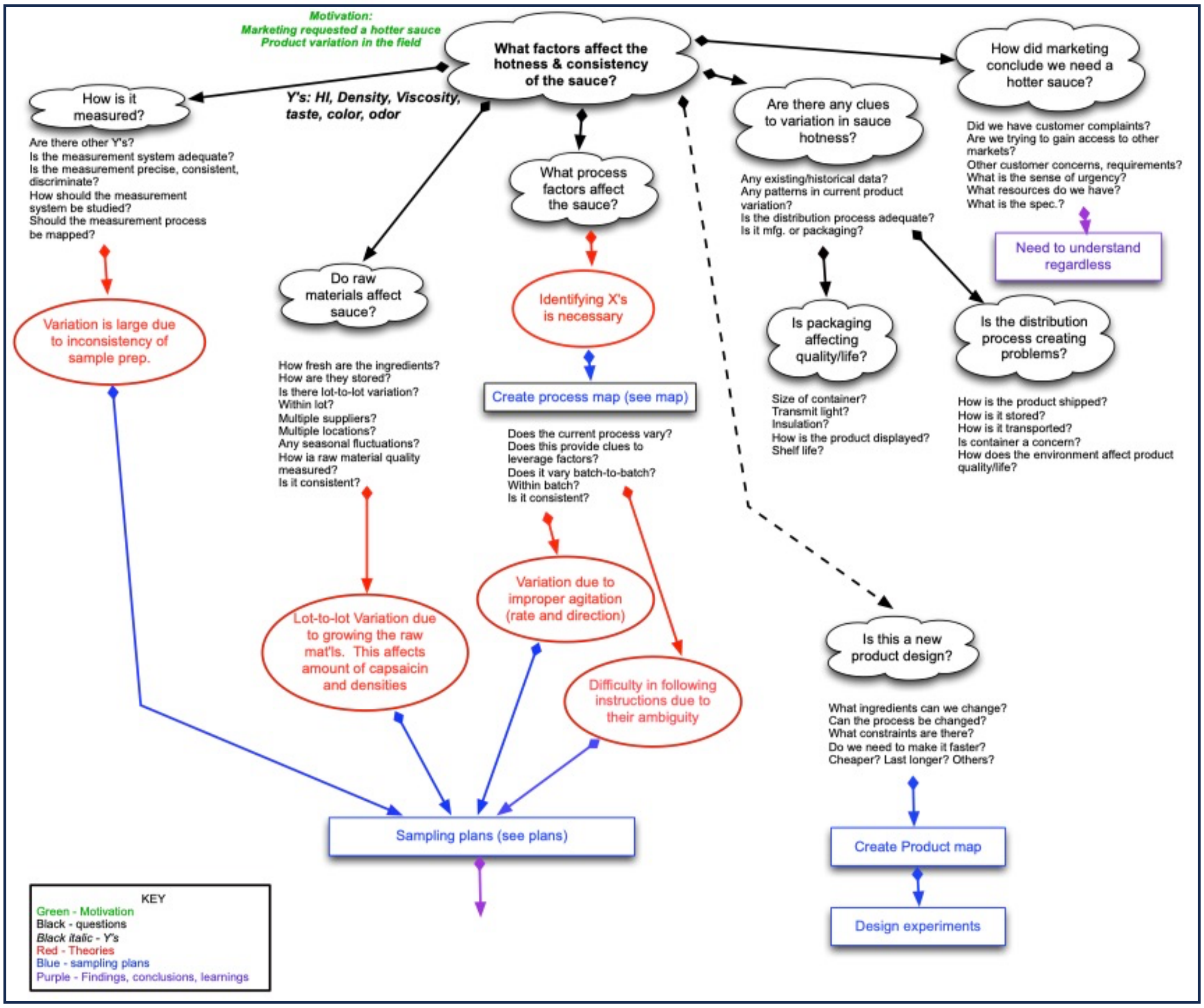


Figure 13.1: Thought Map

Crazy Cajun's manufacturing process is a batch type process, Figure 13.2, making batches one-at-a-time. Ingredients are added to a vat, following a recipe. The mixture is blended and heat is applied. Traditionally, one sample is taken per batch to determine if the sauce meets specification. Each sample is then measured for multiple y's: viscosity, color, smell, particle size, and most importantly, heat index. The heat index is measured on a high-performance liquid chromatograph (HPLC) using the Scoville Scale[1].

[1] The scale is named after its creator, American pharmacist Wilbur Scoville.

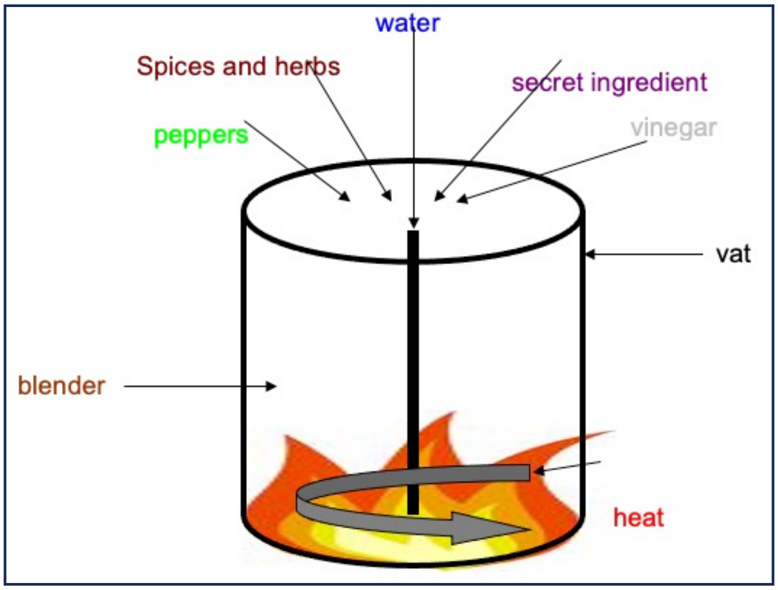


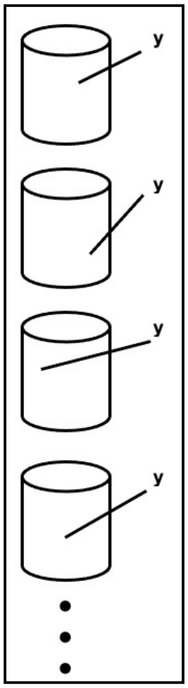
Figure 13.2: Illustration of Crazy Cajun Batch Process

These following is a list of sources of variation (x's), documented on the Process map. These can be roughly *grouped* into three categories or components. Grouping is a function of the frequency in which these x's change. Components, grouped and color coded:

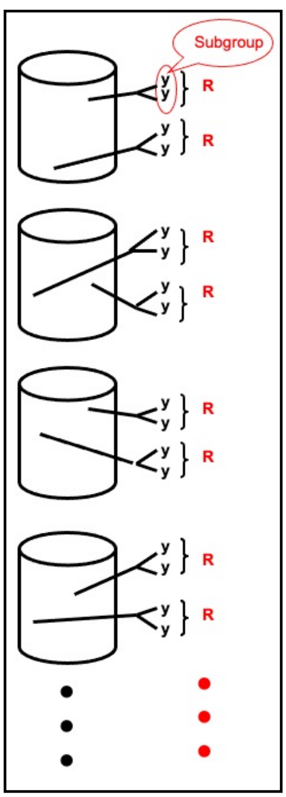
1. Measurement system (M),
2. Within batch (W), and
3. Between batch (B),

Ability to adhere to recipe
Vat cleanliness
Amount of raw materials
Lot-to-lot variation of raw materials
Water quality
Ambient conditions
Measurement equipment precision
Agitation rate
Agitation location
Agitator time
Agitator geometry
Order of adding ingredients
Raw material densities
Temperature rate of rise
Temperature stratification
Cook time
Sample preparation
Sample separation
Dwell time of measurement
Measurement device contamination
Consistency of agitation
Material preparation
Within lot of raw materials
Temperature setting

The *current*, perhaps convenient, data collection is to take and measure one sample per batch depicted below.



A modified sampling strategy is developed to separate and assign sources of variation where two random samples are gathered from each batch and those samples are measured twice for multiple batches depicted below.



The range for the two y's within subgroup is a quantification of the variation due to the measurement system. In this sampling strategy, the measurement component (M) can be separated from the other components (W & B). The measurement system can be assessed for stability and then averaged to reduce the measurement error. When the ranges have been calculated, they may be analyzed using a range chart. The range chart answers the question: Is the measurement system consistent, stable, predictable?

Two scenarios depicted in Figure 13.3:

1. Range chart inconsistent: out-of-control(*)
   This indicates the measurement process is not stable. The x's associated with the measurement system (red) should be further studied to understand what is causing the instability. Investigating that particular "event", when the variation is "out-of-control", is warranted.
2. Range chart consistent: in-control
   This indicates the within subgroup variation is stable. It is now possible to calculate an estimate of the measurement system variance and compare the measurement system to the other components (using the $\overline{y}$ chart).

   Measurement variance[2]:

$$\hat{\sigma}_M^2 = \left(\frac{\overline{R_M}}{d_2}\right)^2$$

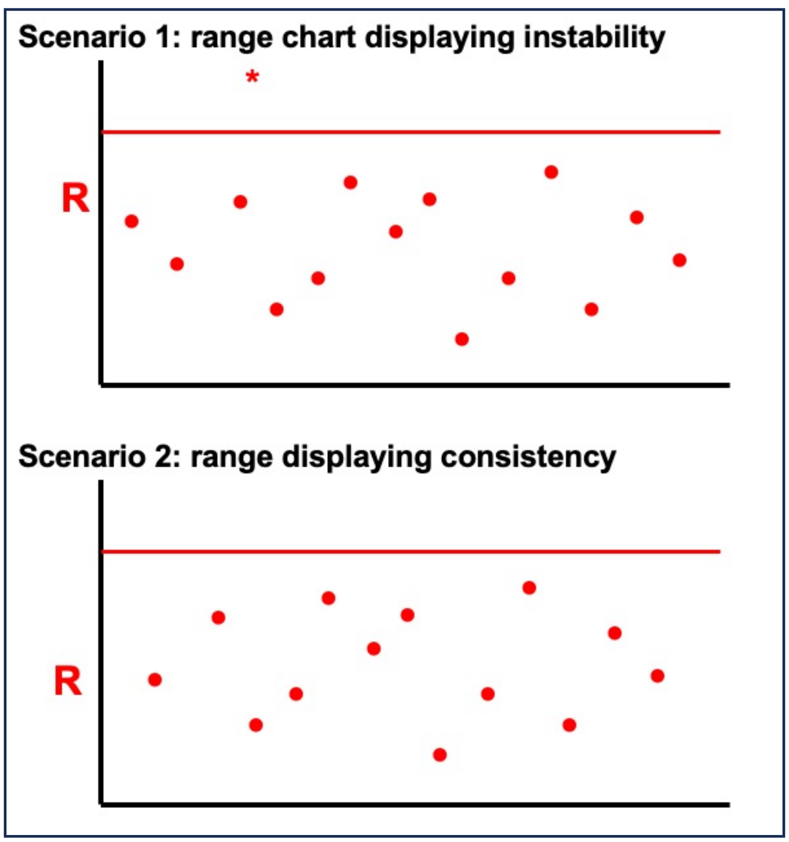
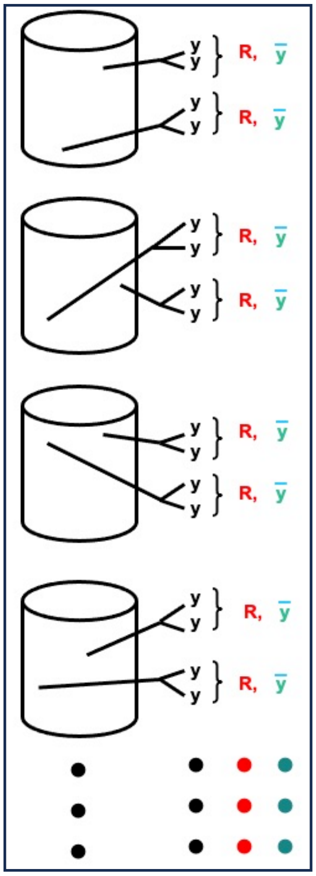


Figure 13.3: Two Possible Outcomes of Measurement System Variation

[2] Variances are used due to their additivity (Pythagorean's Theorem)

Given scenario 2, the range chart indicates consistency, $\bar{y}$'s can be calculated (this is an appropriate statistic). The measurement variation is reduced when the $\bar{y}$ is calculated (s^2/n). The $\bar{y}$'s vary due to factors associated with within and between batch. Note, both within batch and between batch components are present. A comparison of measurement variation and within/between batch variation can be done using the $\bar{y}$ chart.



The $\bar{y}$ chart answers the question: Which source of variation is greater within subgroup sources or between subgroup sources?

Two possible scenarios depicted in Figure 13.4:

1. $\bar{y}$ chart in-control

   Measurement system variation is greater than sum of the within and between batch components of variation. Investigate the measurement system, *commonly*. There is no specific instance that should be considered over others.

2. $\bar{y}$ chart out-of-control

   This indicates the measurement system variation is less than sum of the within and between batch variation. The within & between batch variation is greater than the measurement variation. It is desired for the measurement variation to be <<< (much, much, much less than) the within & between batch) variation.

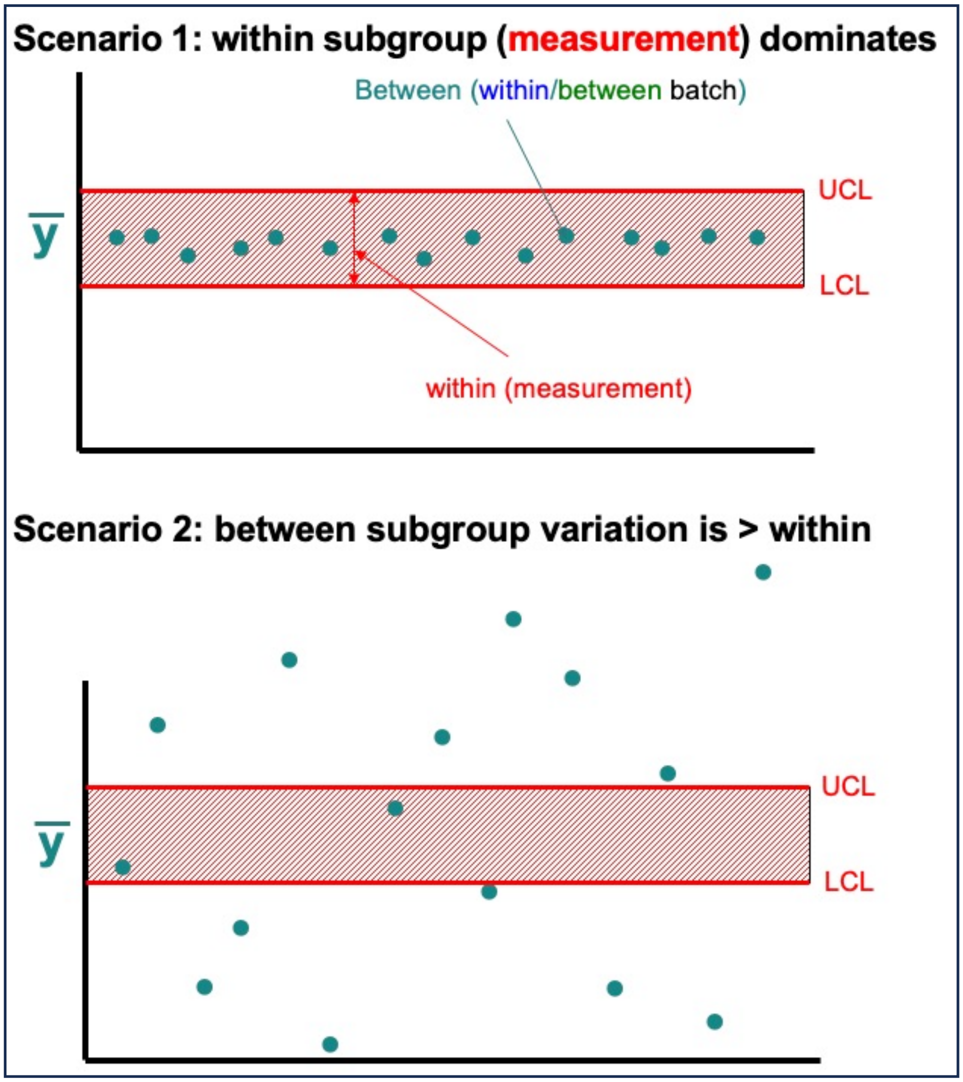
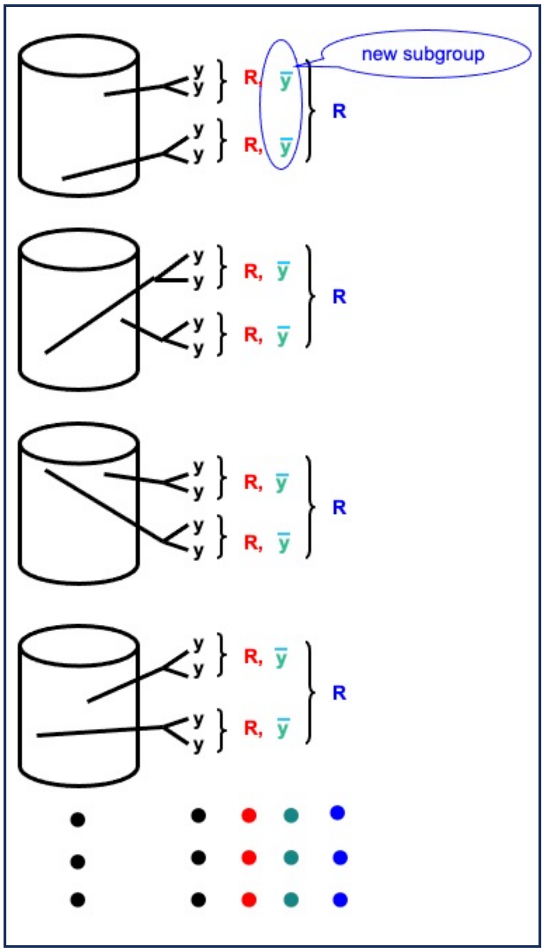


Figure 13.4: Two Possible Outcomes of the Comparison, Measurement to Within & Between Batch Variation

Given Scenario 2, it is now reasonable to analyze the within batch variation for consistency as the measurements system appears capable of discriminating variation in the sample. This is done with the subgrouping strategy depicted as follows. Once ranges have been calculated, a range chart can be used for analysis. *Note, there is still a measurement component present in the ranges calculated, but it is reduced when the y's were averaged.*



The range chart in this case answers: Is the within batch component of variation consistent?

Two possible scenarios depicted in Figure 13.5:

1. Range chart inconsistent: out-of-control(*)
   This indicates the within batch sources of variation are not stable. In this case work within batch, *specially*. Investigating that specific instance is warranted.
2. Range chart consistent: in-control
   This indicates the within batch component is stable. It is now reasonable to calculate an estimate of the within batch variance and compare the within batch to the between batch (using the $\bar{\bar{y}}$ chart). Within batch variance:

$$\hat{\sigma}_W^2 = \left(\frac{\overline{R}_w}{d_2}\right)^2 - \frac{\hat{\sigma}_M^2}{2}$$

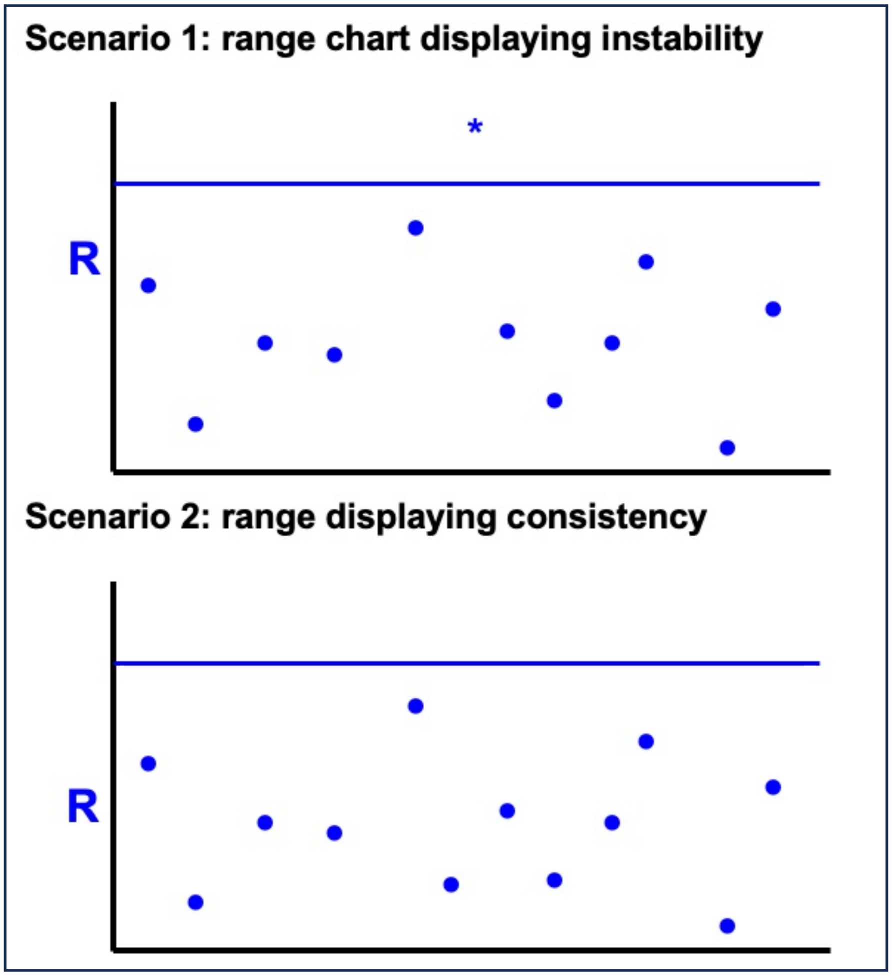
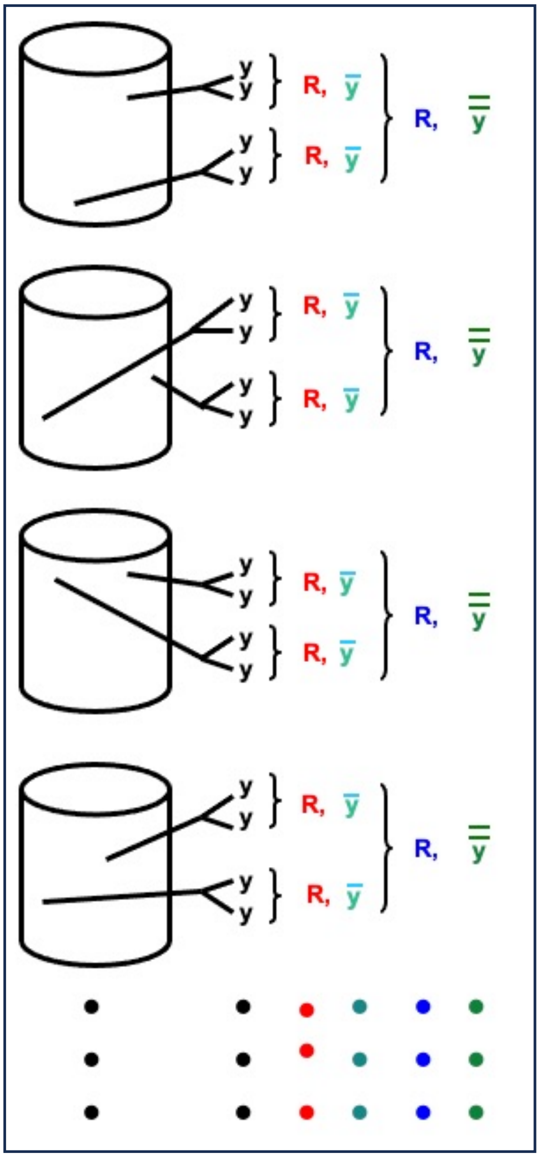


Figure 13.5: Two Possible Outcomes Assessing the Consistency of the Within Batch Variation

Given scenario 2 (range chart in control), $\bar{\bar{y}}$'s can be calculated. Both the measurement and the within batch variation is reduced when the $\bar{\bar{y}}$ is calculated.

The $\bar{\bar{y}}$'s vary due to between batch components of variation. A comparison of the within batch variation to the between batch variation can be done using the $\bar{\bar{y}}$ chart.



Two possible scenarios for the comparison of the within batch variation to the between batch variation depicted in Figure 13.6:

1. $\bar{\bar{y}}$ chart in-control

   This indicates the within batch variation is greater than the between batch component of variation. Work within subgroup (within batch), *commonly*.

2. $\bar{\bar{y}}$ chart out-of-control

   This indicates the between batch variation is greater than the within batch. Work between batch.

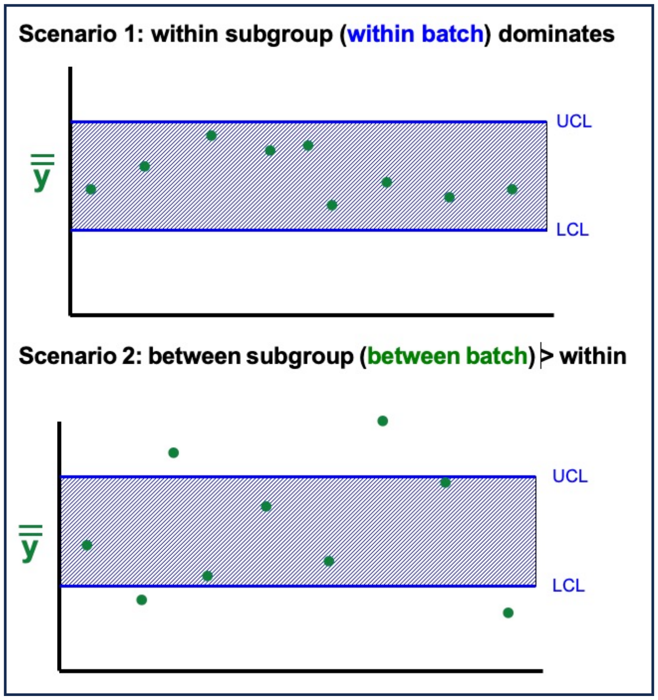
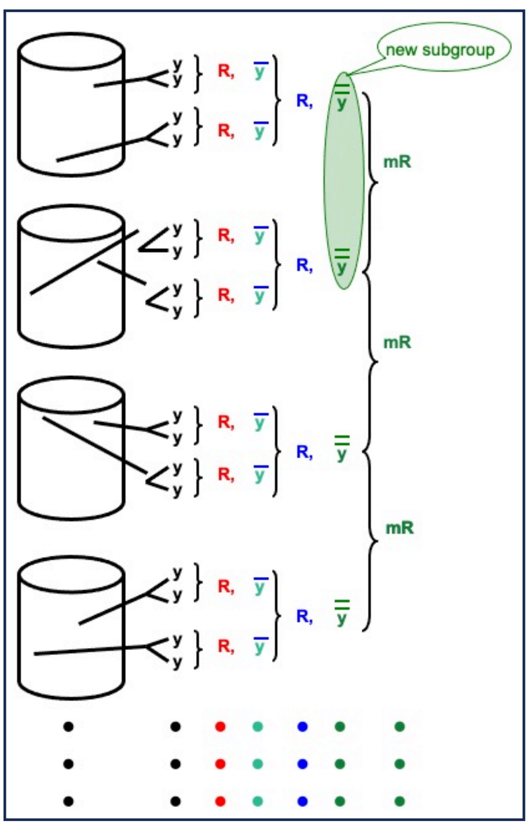


Figure 13.6: Two Possible Outcomes for the Comparison of the Within Batch Variation to the Between Batch Variation

Given scenario 2, the only question that remains: what is the nature of the investigation? Between batch variation can be analyzed for consistency. This is done by calculating and plotting the moving ranges. In this case the range of consecutive batch averages.



The moving range chart answers: Is the between batch variation consistent?

Two possible scenarios depicted in Figure 13.7:

1. Moving range chart out-of-control (*)
   This indicates the variation in consecutive batches is not stable. In this case work between batch, *specially*. A reasonable question, what happened there?
2. Moving range chart in-control
   This indicates the variation is stable. It is now possible to calculate an estimate of the between batch variance:

$$\hat{\sigma}_B^2 = \left(\frac{\overline{mR}}{d_2}\right)^2 - \frac{\hat{\sigma}_W^2}{2} - \frac{\hat{\sigma}_M^2}{4}$$

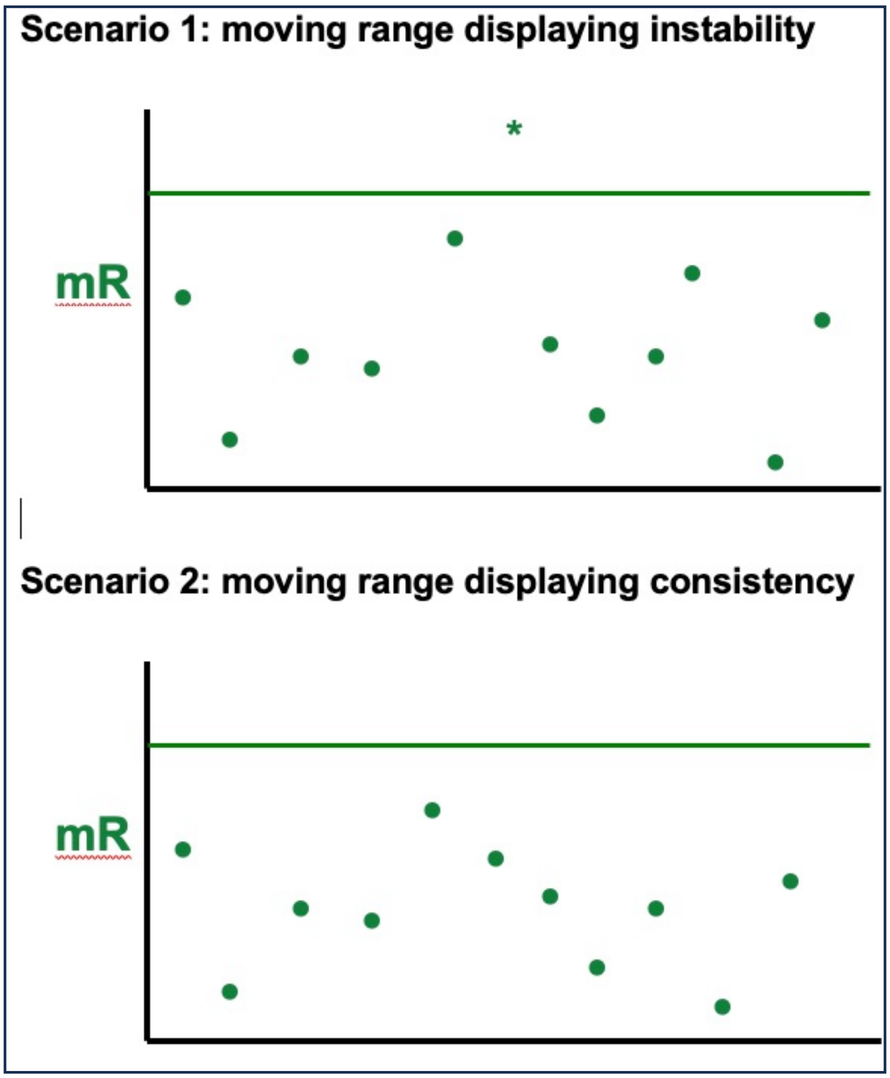


Figure 13.7: Two Possible Outcomes Evaluating Between Batch Variation for Consistency

The sum of each component of variance is the total variance *in the sampling plan*.

$$\hat{\sigma}^2_{total} = \hat{\sigma}^2_M + \hat{\sigma}^2_W + \hat{\sigma}^2_B$$

Of course, the percentage each component contributes to the total variation may also be estimated. The numbers calculated to estimate the variance components are conditional and only applicable to this study. If this study is representative of future conditions, you can be confident these values will be reasonable estimates in the future, if not they are not useful.

As discussed in Part I, sampling trees are used to visually describe the sampling plan. Each layer of the Sampling Tree corresponds to:

- a set of x's likely to be changing at that layer (color coded to correspond to the process/product map),
- a hypothesis or set of hypotheses (labels),
- a set of x's not changing during the study (inference space)

Each sample can be measured for multiple y's. The tree is necessary for planning and analysis. It describes how and why the samples will be acquired.

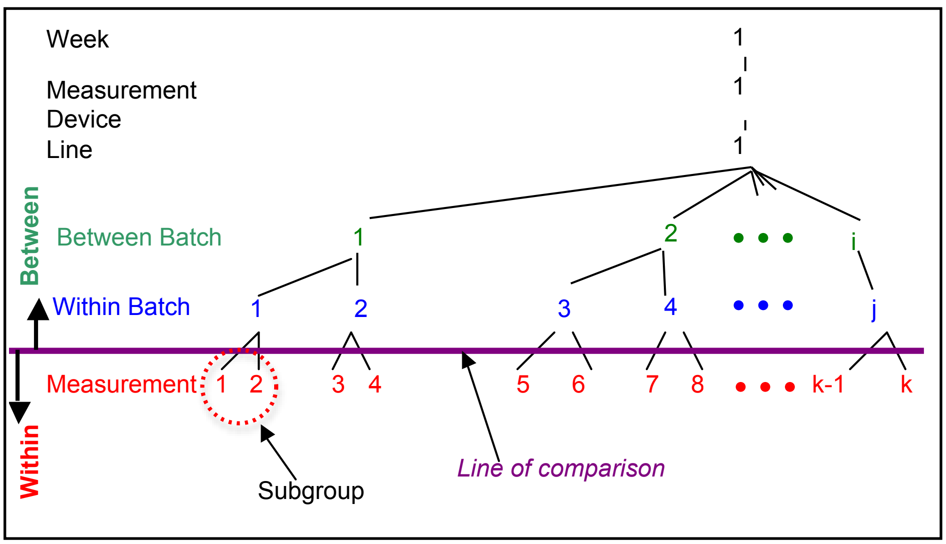


Figure 13.8: Samplig Tree Depicting Measurement Subgroup

In this tree, the measurement component is used as the initial subgroup. It will be assessed for stability using a range chart. If the range chart indicates the measurement system is stable, the measurement variation can be compared with the other components using a Y-bar chart. The *line of comparison* is an aid to see what comparisons (within and between) are being made. If the between variation is larger, then it may be possible to *roll up the tree* to separate more components of variation (e.g., the within batch from the between batch). This essentially means a change in the location of the *line of comparison* as depicted in Figure 13.9.

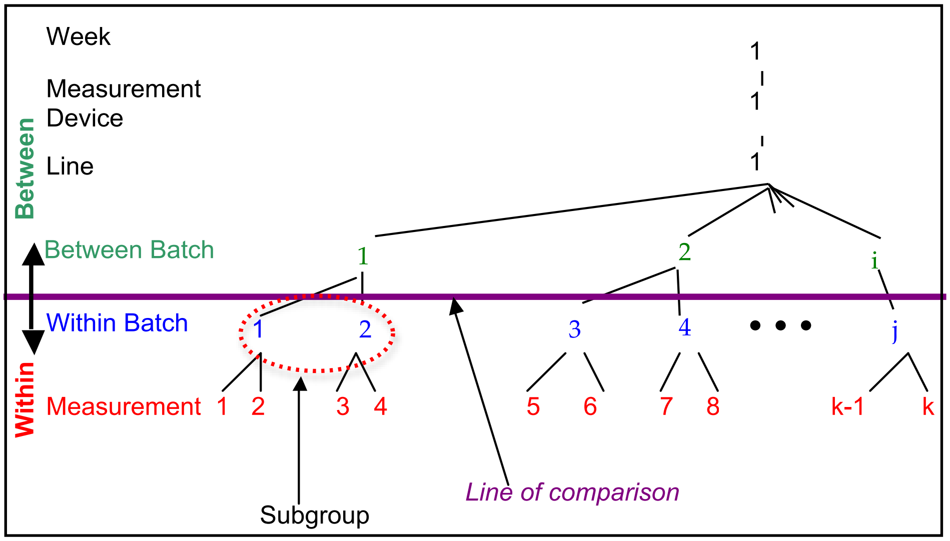


Figure 13.9: Sampling Tree Depicting New Subgrouping Strategy

Here, the *line of comparison* is positioned such that within batch (and some of measurement, recall the measurement was averaged) is within subgroup and between batch is between subgroup. The within batch (and some of the measurement) variation is the basis for comparison for the between batch variation. Possible outcomes that should be predicted prior to the sampling plan being executed:

- The response variable did not exhibit any practically significant variation,
- Measurement system inconsistent,
- Measurement system was the biggest source of variation,
- Within batch inconsistent,
- Within batch the biggest source of variation,
- Between batch inconsistent,
- Between batch the biggest source of variation.

Also predicted is what work would be done for each possible outcome.

Summary

Analyzing sampling data is an exercise in restraint. It requires resisting the urge to explain, optimize, or conclude prematurely. When used properly, sampling analysis provides orientation, highlights leverage, and informs the design of subsequent studies. Its value lies both in the answers it provides, and in the quality of the questions it helps you ask next and in how effectively it guides the next iteration. The next chapters discuss different types of sampling plans and what information each may provide.

Chapter 14
Measurement Systems: Context for Interpretation

Why Measurement Matters

Every investigation begins with some simple questions: How do I quantify the phenomena of interest? **Is the measurement system capable of providing insight to my hypotheses?** Measurement is one of the multiple components introduced in Chapter 13.

It is necessary to compare measurement variation to the sources you wish to detect. As those sources change, with changing hypotheses, so does measurement capability. All learning about causality depends on detecting changes in a response (Y) as inputs (x's) vary. If the measurement system cannot detect those changes, then:

- important effects may appear insignificant
- processes may appear stable when they are not
- experiments may fail to reveal meaningful relationships
- conclusions may be incorrect or misleading

Before analyzing data, the investigator should ask:

- Does the measurement system have enough **discrimination** (aka effective resolution)?
- Is the measurement system **stable**?
- Is the measurement variation small relative to the **sample or treatment variation**? Does it have adequate **precision**?
- Will the measurement system allow detection of changes that matter **practically**?

A lack of observed variation is not always good news. It may simply mean the measurement system cannot quantify the signal. This and other measurements system issues is one of the most common reasons studies fail to be informative.

I recall a discussion I had with a manager discussing the lack of improvement on a project. The measurement system was intended to provide a quality check for the rotational balance of helicopter blades. Anecdotal evidence also suggested it was ineffective as every blade had to be re-balanced after installation onto the helicopter. The measurement system was studied by an engineer who discovered the measurement system was incapable of providing any reasonable estimate of the balance of the blades. The manager said; "At least it's giving us a number".

Measurement as a Source of Variation

All measurements contain variation. When a characteristic is measured repeatedly on the same item, in the same location, the results will differ. This variation may come from:

- inherent measurement device precision
- operator technique
- environmental conditions
- sample orientation or positioning

- calibration drift
- inadequate effective resolution

Measurement systems can be evaluated using a nested or a crossed structure. Repeated measurements are taken within the same part/sample at the same location so that measurement variation can be separated from part-to-part or sample-to-sample variation. From a Sigma Science perspective, a measurement evaluation is best accomplished using a nested COV study. This allows for assessing measurement capability and stability over other changing sources of variation.

Practical Evaluation First

Consistent with Ross' Rules of Analysis, evaluation should begin with a **practical view**.
Before formal studies:

- Repeat measurements on the same sample, same location
- Examine the spread visually
- Compare measurement variation to practical significance or expected factor effects
- Look for obvious patterns, instability or drift

If repeated measurements vary more than expected process variation, further analysis is not useful. The priority should shift to improving the measurement system. This may include evaluating the measurement process, using alternative measurement equipment or re-defining the Y.

> *"Data of poor quality are a pollutant to clear thinking and rational decision making"*
>
> *Hunter*

Measurement Discrimination and Resolution

A measurement system must be capable of resolving meaningful differences. If the resolution is too coarse distinct values collapse into the same reading, variation is artificially reduced, and patterns are hidden. As a practical guideline: The measurement system should be capable of detecting variation significantly smaller than the engineering changes of interest. A practical rule is that measurement variation should be small, less than 10%, relative to the smallest change that matters. If the objective is to detect a 5% change, a system that varies by ±5% cannot support the investigation.

Stability of the Measurement System

Even a precise measurement system is not useful if it is not stable. Stability means the measurement process produces consistent results over time when measuring the same standard or reference. Control charts, repeated measurements of samples or standards, or periodic checks can reveal drift, step changes, environmental effects, and calibration issues. Unstable measurement systems introduce time-dependent variation that may exaggerate or hide the variation you are interested in learning about.

Measurement and Study Design

Measurement capability should be considered **before** data collection. The following questions should be part of planning:

- Will measurement error limit the ability to detect factor effects?

- How effective are the measurement units?
- Would averaging multiple measurements reduce measurement error?
- Should the study include nested repeats?
- Would a transformation or alternative metric be more sensitive?
- Should an alternative response be defined?

Measurement uncertainty affects:

- sample size requirements
- experimental precision
- inference space
- interpretation confidence

In many cases, improving the measurement system is more valuable than increasing sample size.

Destructive or Altering Measurement Systems

Destructive measurement systems are common in engineering and present a special case where precision repeatability cannot be assessed directly. In these situations, measurement variation cannot be separated from sample variation analytically and must be managed through sampling design. Some responses cannot be measured repeatedly on the same unit. Examples include:

- strength, fatigue tests
- life testing
- material and chemical composition analysis

In destructive measurement systems, the observed variation includes both product and measurement variation. These cannot be separated directly. This creates a fundamental challenge:

How can measurement variation be estimated when the same sample cannot be measured twice?

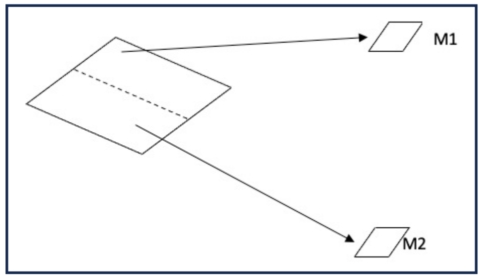
Strategies for Non-repetitive or Destructive Measurement (Figure 14. 4)

When repeated measurement is impossible, the investigator must change the data collection strategy. Common approaches include:

A. **Reduce product variation locally (Figure 14.1)**
   Sample adjacent units or items from a small, homogeneous region so that differences are primarily measurement related.
B. **Use matched or nearly identical samples (Figure 14.2)**
   Select units expected to have minimal inherent variation.
C. **Measure within restricted locations (Figure 14.3)**
   For spatial processes, sample from a confined area or time window.
D. **Use reference standards**
   Where it is impossible to ascertain measurement error on actual samples, perhaps use materials or calibration samples with known values. This more likely speaks to accuracy rather than precision.
E. **Develop surrogate responses**
   Identify a non-destructive measurement that is strongly associated with the destructive characteristic. Of course, some measure of the destructive mechanism must be estimated to perform the correlation.

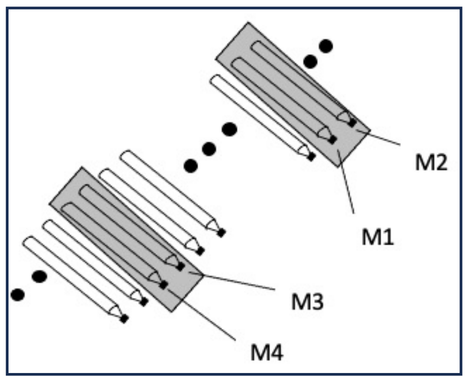
F. **Improve the measurement process experimentally**
Treat measurement factors (operator technique, preparation method, environment) as experimental variables, determine their effect on measurement error and reduce their impact.

Each of these approaches reflects a core Sigma Science principle: When measurement variation cannot be separated analytically, it must be managed through **sampling strategy**.



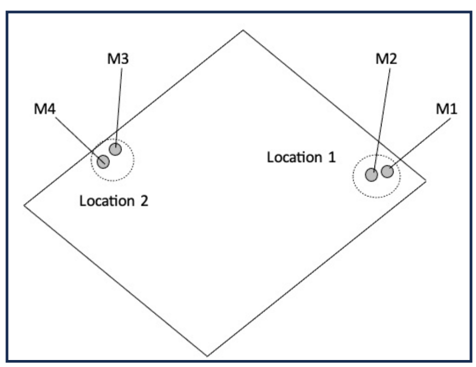
Assumption: within sample variation is smaller than measurement system variation

Figure 14.1: Samples in Close Proximity



Assumption: short term process variation is smaller than measurement system variation

Figure 14.2: Nearly Identical Samples



Assumption: Within location variation is smaller than the measurement system variation

Figure 14.3: Restricted Locations

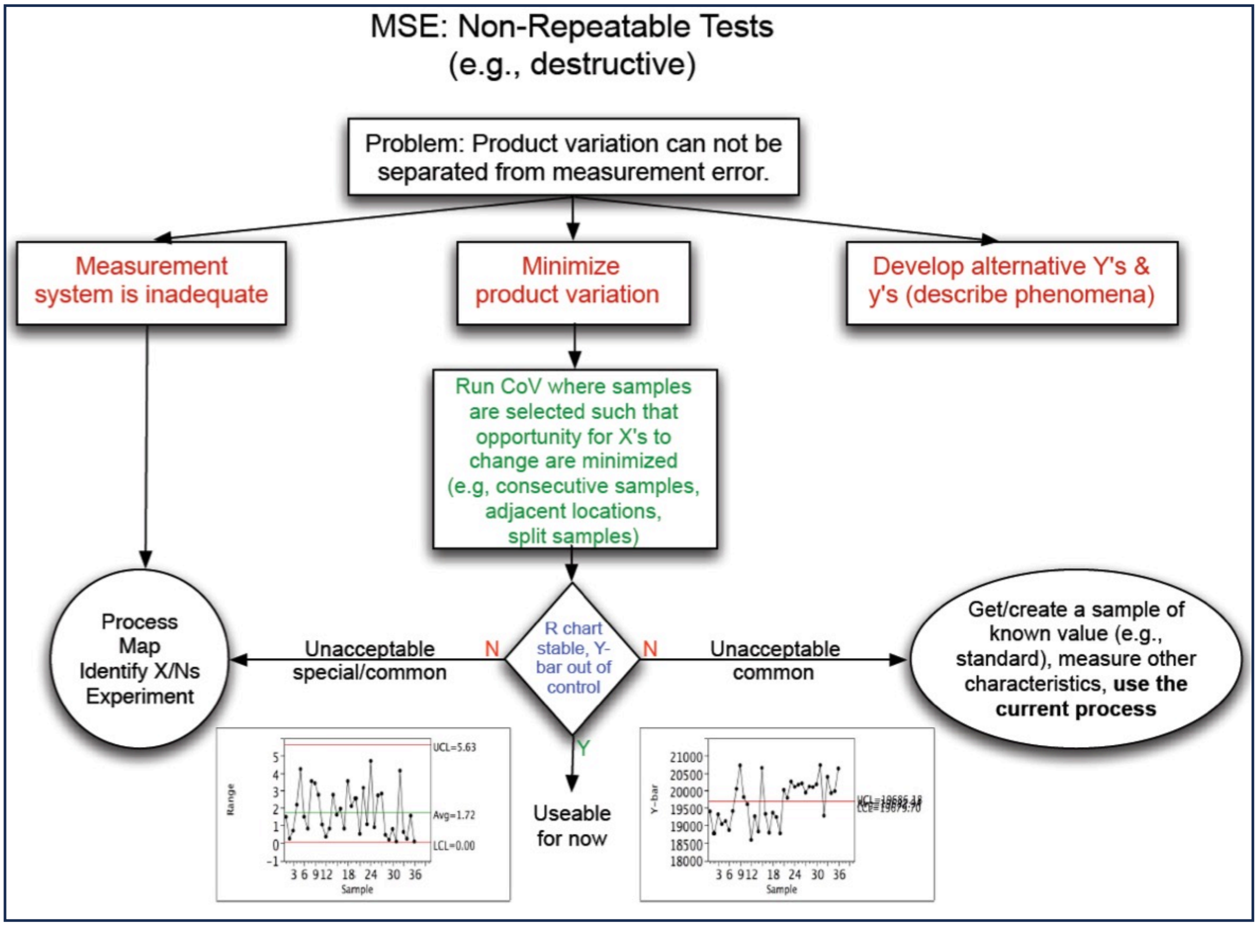


Figure 14.4: MSE for Non-repeatable Tests

Measurement and Learning Efficiency

Measurement systems directly influence learning efficiency.

If measurement noise is high:

- larger experiments are required
- stronger factor effects are needed for detection
- subtle relationships remain hidden

This is analogous to the **water level in the lake**:

Measurement variation raises the water level. Only large effects rise above it. Improving the measurement system lowers the water level and allows smaller effects to be seen.

Measurement in the Iterative Process

Measurement capability should not be treated as a one-time assessment.

As knowledge evolves:

- responses may be redefined
- more sensitive metrics may be developed
- surrogate measures may replace original ones
- precision requirements may change

Early studies often reveal that the original response was not appropriate. This is part of the iterative nature of Sigma Science.

Measurement as Context for Interpretation

Measurement systems provide context for interpreting all results.

When reviewing analysis outputs, the investigator should ask:

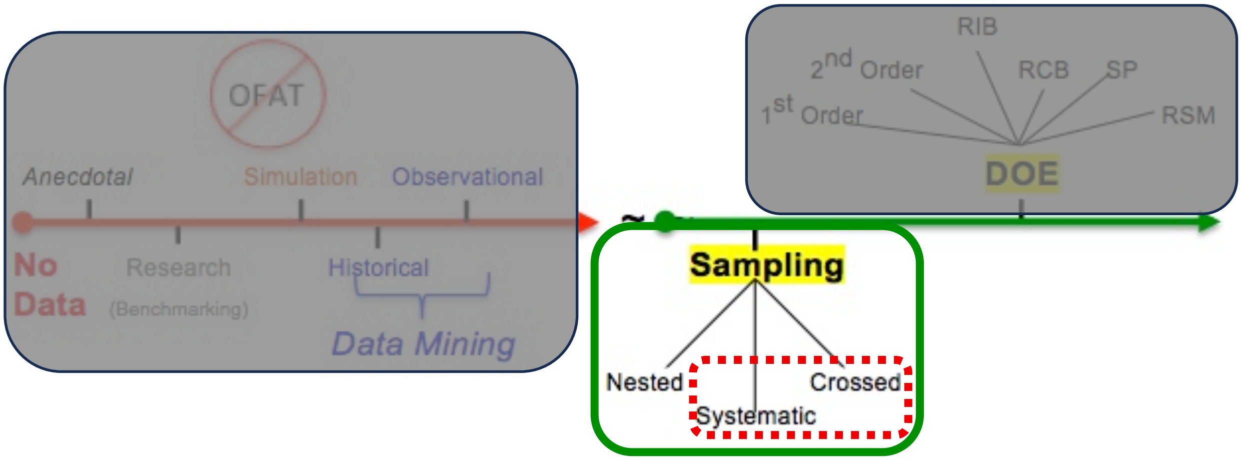
- Could measurement noise be masking effects?
- Could apparent stability be due to low resolution?
- Could non-normality or outliers be measurement artifacts?
- Is the lack of variation real or a measurement limitation?

Many analysis problems are actually measurement problems.

Key Principles

- All measurements contain variation.
- Measurement studies are most effective when they are nested either in a sampling plan or in an experiment.
- Unknown measurement variation can be a noise source.
- Learning requires measurement systems capable of detecting meaningful changes.
- Poor measurement cannot be fixed with statistical analysis.
- In destructive systems, measurement issues must be addressed through sampling design.
- Improving measurement capability often increases learning efficiency more than increasing sample size.

Chapter 15
Sampling Structure: Output and Interpretation
Systematic and Crossed



Purpose of the Chapter

To demonstrate how the structure of a sampling plan governs the meaning of analytical outputs, and how misinterpretation arises when structure is ignored. This chapter generalizes the logic from Chapter 12 and prepares the reader to interpret *different* sampling situations. These various sampling strategies will be demonstrated via case studies. If the structure is hierarchical, use control chart logic. If the structure is crossed, quantitative methods may be necessary. These examples are intended to provide insight and interpretation to the outputs of analysis. These are not intended to be models for all situations.

Prioritized Global Effects

In general, systematic and crossed sampling plans are focused on the highlighted **effects:**

Possible Effect	***Rank***
Noise	2
Main Effects (Factors)	2
Two-Factor Interactions (2nd order linear)	2
Noise-by-factor interactions	5
Simple curvature (2nd order non-linear)	5
Complex non-linear (≥3rd order non-linear)	6
≥ Three-Factor Interactions (3rd order linear)	6
Stability	1
Leverage	1
Measurement uncertainty	1
Mean	2
Variation	2

Nested-Systematic Sampling Plan: Lithography

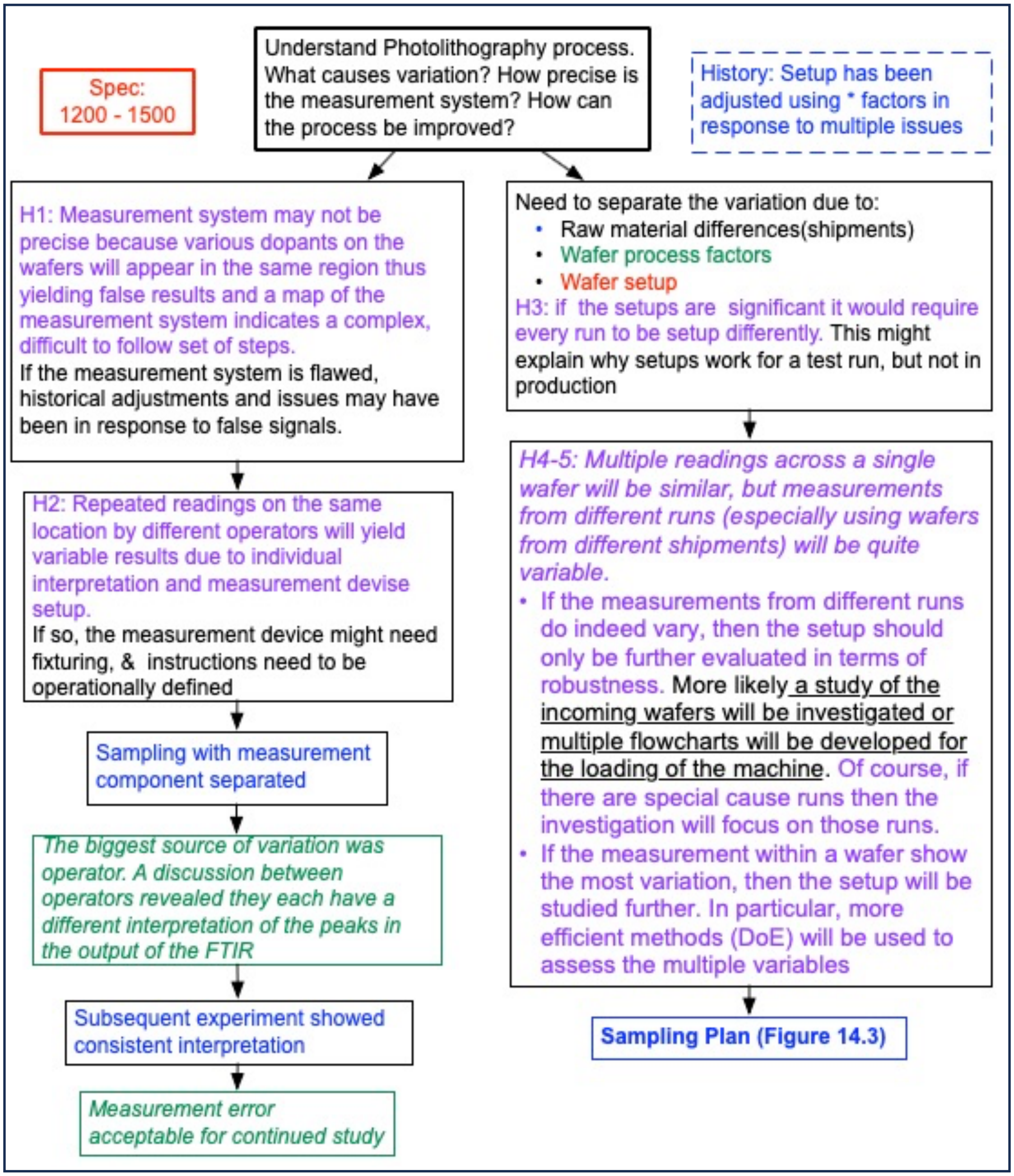


Figure 15.1: Thought Map

In the first sampling iteration, the measurement system was evaluated and improved. The portion of the project presented here will focus on the amount and variability of a particular impurity (dopant). It is measured using FTIR[3]. The value should be >1200 and preferably <1500. A change in 25 is of practical value.

[3] Fourier Transform Infrared spectroscopy measures reflected or transmitted IR beams at a particular wavelength for the impurity.

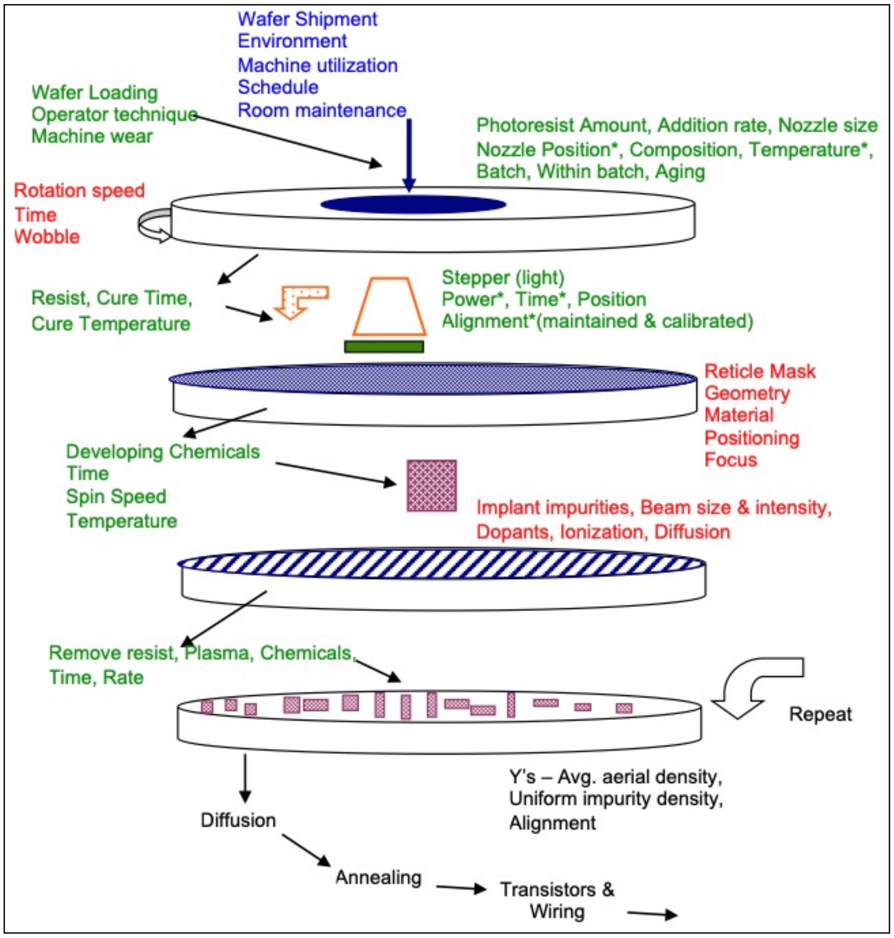


Figure 15.2: Process Map, Colored to Correlate to the Sampling Plan

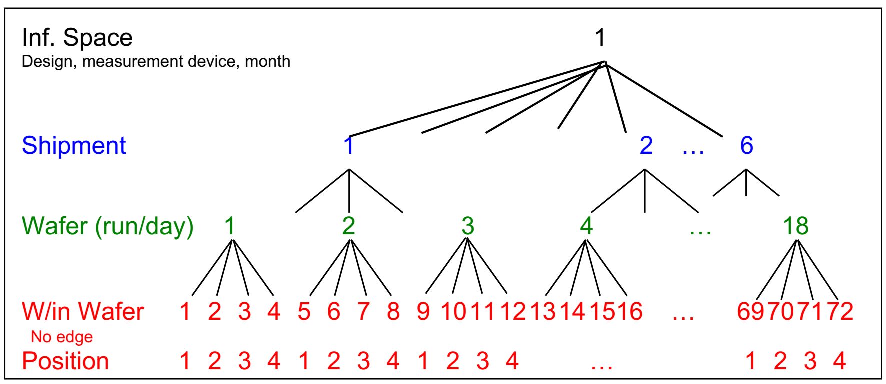


Figure 15.3: Sampling Plan (Tree)

Four systematic positions are measured per wafer/run/day (Figure 15.4). The instruction provided for the positions is to "avoid measuring near the edge" and measure in similar locations. Any specific instructions should be noted on the sampling plan. The reason for the restriction on the edge measurement is the edge has always been assumed to exhibit the most variation. Every third day in the sampling study will be using wafers from a different shipment. It has frequently been suggested these wafers (about 8%) are routinely scrapped. Note of all sources of variation that did NOT vary during the study. This defines the inference space of the study which impacts your ability to extrapolate the results into the future.

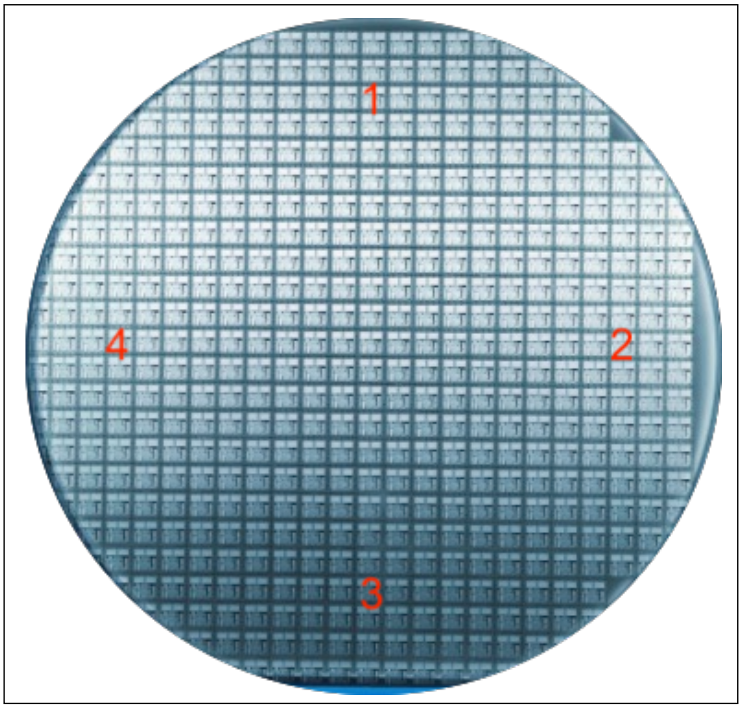


Figure 15.4: Systematic Positions Located 1-4

Predicted Next Steps

If certain positions show unusual values (special) then the investigation will focus on what happened on those certain positions. If position shows evidence of a systematic pattern, then we look to see what x's are correlated to the different positions. If position shows evidence of being large in comparison to wafer, without a systematic pattern, then further work should focus on the green x's in the process map and associated hypotheses. That work includes further directed sampling studies or experimentation to understand these sources of variability in the manufacturing process. If shipment is large, then work at the supplier would be prioritized or a possible re-design of the product robust to supplier variation.

Analysis of a sampling plan will typically include, but not be limited to:

- A description of the sampling plan, including the tree, explaining how and why the samples will be taken (linking to the Thought map and Process map for context),
- A practical look at the data for practical significance, comparison with predicted data, obvious unusual data points or patterns in the data,
- A graph of all of the data without any summary. This can be accomplished with a variability chart[4],
- A set of R and Y-bar control charts.
- Quantified variance components if appropriate (the components are consistent).

Analysis Sequence:

0 Does the data set match the tree? The data set should have the same columns as the layers of the sampling plan. **The data MUST be in the same order as the sampling plan.** If it is not,

[4] Variously known as a multi-vari, dot-frequency, individual value and/or Box plots.

reorganize it, to match how the data was acquired. Realize, statistical software programs sort time series vertically in the data table.

1 Is there enough practical variation in the data set (>25)? Has the phenomena of interest been exposed? If not, next steps are to work on factors in the inference space for the next iteration. This means expand the inference space by allowing variables in the inference space to change.

2 Are there any patterns in the data? Are there any interesting data points? How does the data compare with predictions?

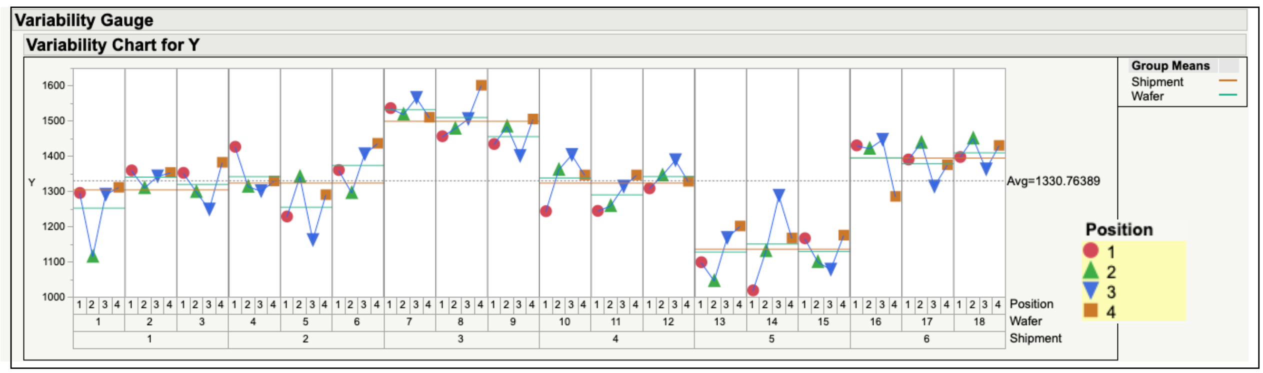


Figure 15.5: Variability Chart for Y[5]

There do not appear to be any obvious systematic patterns where one position or another is consistently at the top or bottom. Update the Thought map.

3 Are there any special causes within subgroup? How does the within subgroup compare to the between subgroup sources of variation? Circle the measurements that make up the subgroups.[6] (Figure 15.6) This helps recognize the subgroup size.

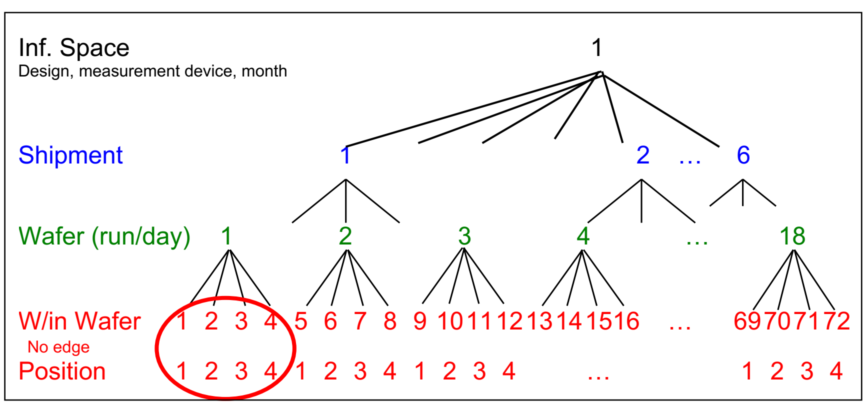


Figure15.6: Tree Showing Initial Subgroup Circled

[5] The purpose of the variability plot is to look at the data without summarizing it.
[6] Start with the smallest possible subgroup at the lowest layer.

4 Are the sources of variation are captured within subgroup consistent and stable? Create a range chart (Figure 15.7). Yes, they appear stable, so it is appropriate to summarize and compare with the other sources in the study.

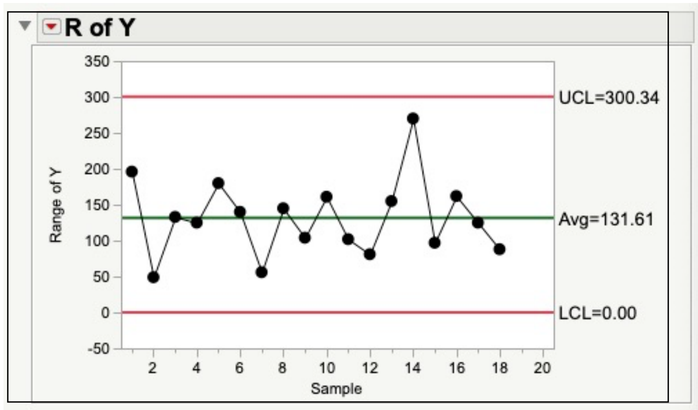


Figure 15.7: Range Chart for Within Wafer Variation

5 Is the within wafer variation more or less than the other sources in the study? Compare sources of variation, Figure 15.8.

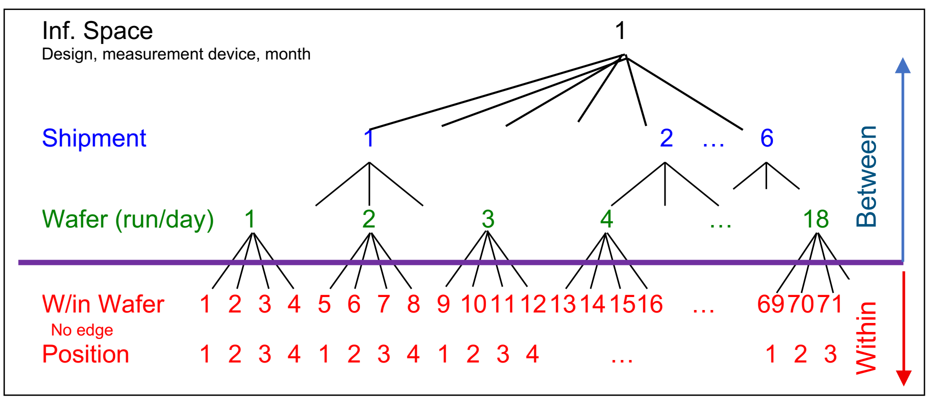


Figure 15.8: Visual of the Component Comparison

What is within and between is dependent on the subgrouping strategy. Sources of variation below the line of comparison are *within* and those above the line of comparison are *between.* The Y-bar chart (Figure 15.9) suggests the variation between is larger.

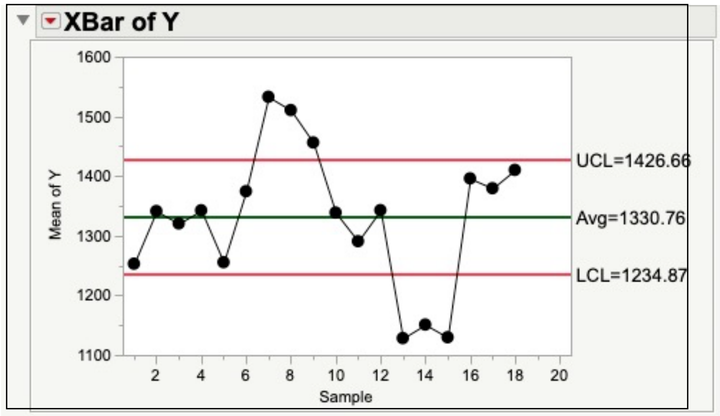


Figure 15.9: Y-bar Chart Suggesting Between Subgroup Sources are Dominant

6 Roll up the tree. Since the larger source of variation is between subgroup: Is the variation wafer consistent and stable? Is the largest source of variation wafer or shipment? Summarize the within wafer variation by averaging the four data points. Averaging reduces the variation due to within wafer and biases the averages to the wafer layer.

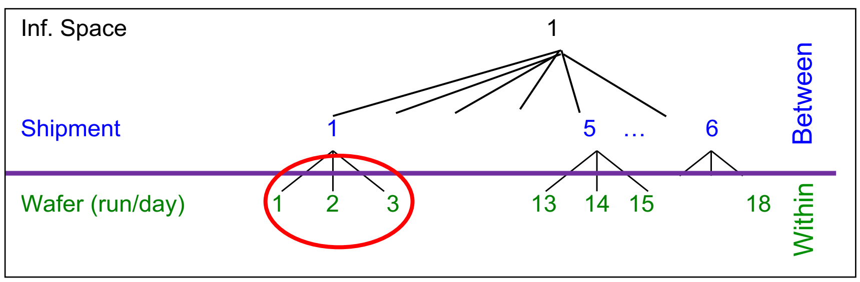


Figure 15.10: New Subgrouping Strategy

7 Create new control charts based on the new comparisons (subgrouping strategy). Is the variation wafer-to-wafer consistent and stable?

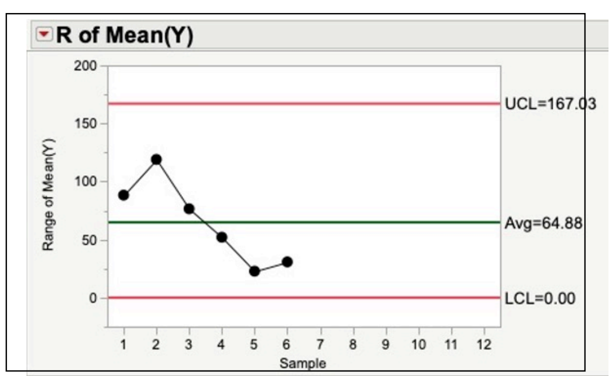


Figure 15.11: Range Chart for the Wafer Layer

It appears the wafer variation is consistent (Figure 15.11). Which source is greater wafer-to-wafer or shipment-to-shipment?

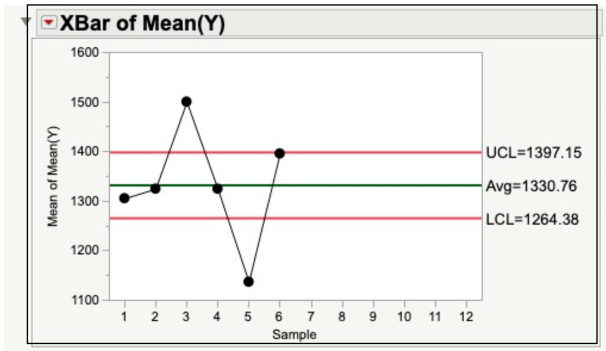


Figure 15.12: Average Chart Comparing Wafer (control limits) to Shipments

It appears the shipments have more variation than the wafers.

8 Is the shipment variation consistent? Roll up the tree again, averaging the 3 wafers per shipment. This again reduces the wafer variation and biases the averages to the shipments. There are no more subgrouping strategies, so use a MR chart to assess stability at the top of the tree. There is not a lot of data to assess stability.

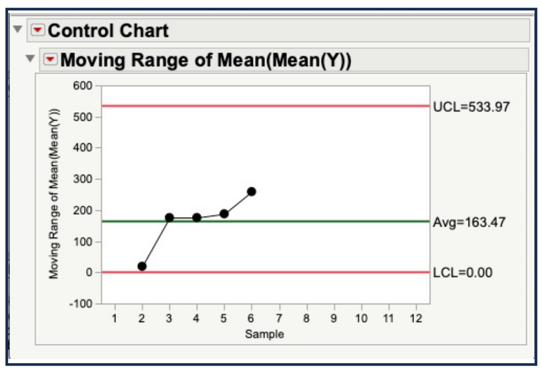


Figure 15.13: MR Chart Showing Consistent variation in Shipments

What sources of variation are not captured in the study? These are the remaining x's with no color or symbol.

9 Quantitative estimates of components of variation: $\hat{\sigma}_T^2 = \hat{\sigma}_s^2 + \hat{\sigma}_w^2 + \hat{\sigma}_p^2$

$$\hat{\sigma}_p^2 = \left(\frac{\bar{R}_p}{d_2}\right)^2 = \left(\frac{131.61}{2.059}\right)^2 = (63.92)^2 = 4085.7$$

$$\hat{\sigma}_w^2 = \left(\frac{\bar{R}_w}{d_2}\right)^2 - \frac{\hat{\sigma}_p^2}{2} = \left(\frac{64.88}{1.693}\right)^2 - \frac{4085.7}{4} = (38.3)^2 - 1021.4 = 445.5$$

$$\hat{\sigma}_s^2 = \left(\frac{\overline{mR_s}}{d_2}\right)^2 - \frac{\hat{\sigma}_w^2}{3} - \frac{\hat{\sigma}_p^2}{12} = \left(\frac{163.5}{1.128}\right)^2 - \frac{445.5}{3} - \frac{4085.7}{12} = 20520.6$$

$\hat{\sigma}_T^2$ = 20520.6 + 445.5 + 4085.7 = 25051.8

- ✓ $\hat{\sigma}_p^2$ = 16.3%
- ✓ $\hat{\sigma}_w^2$ = 1.7%
- ✓ $\hat{\sigma}_s^2$ = 81.9%

10 Summarize findings. What did you learn? How did the results compare to predictions?

11 Update thought map (Figure 15.14)

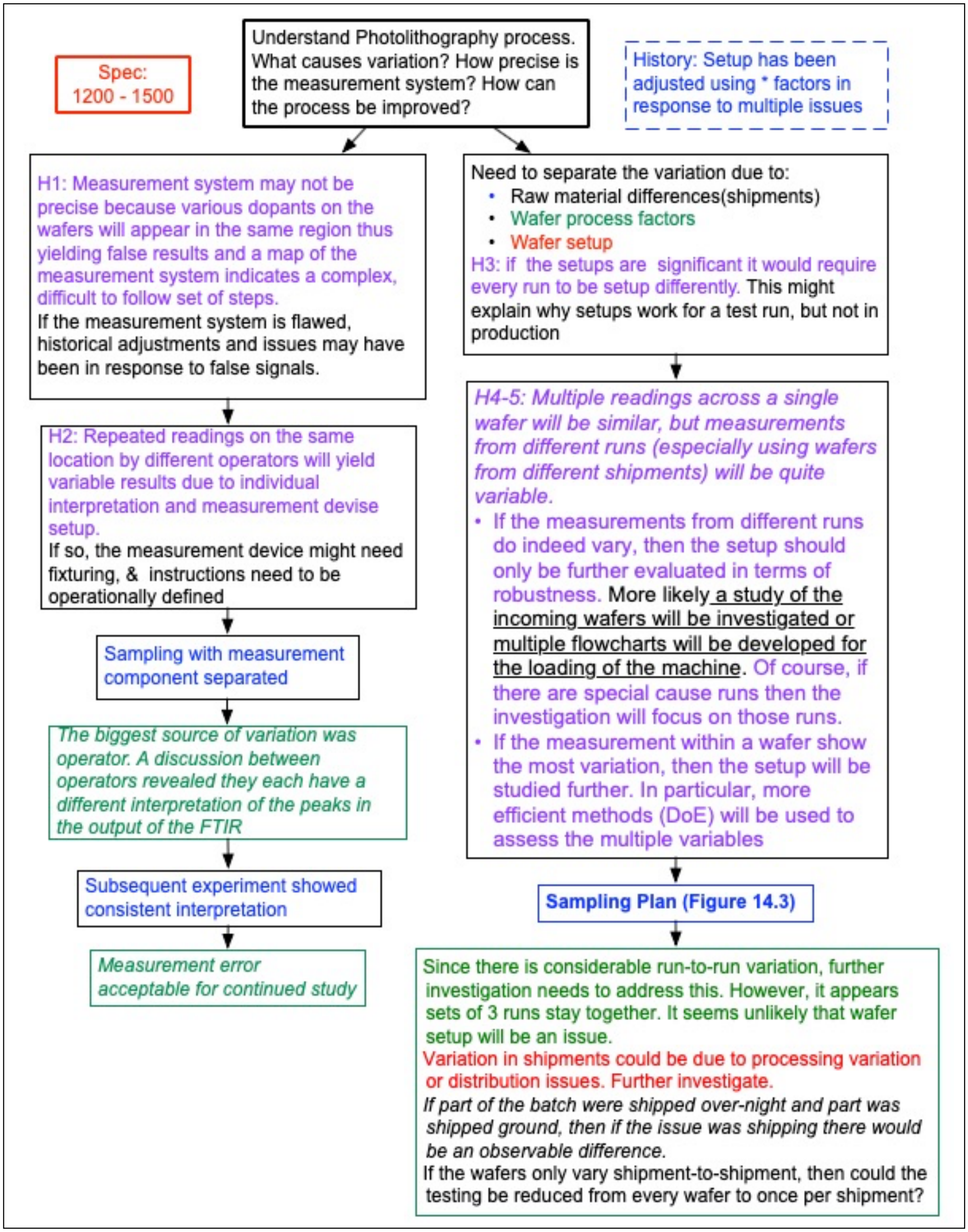


Figure 15.14: Updated Though Map

12 Next iteration? Follow predicted next steps.

Crossed Study: Injection Molding

Plastic parts are made in a multi-cavity mold. There are several dimensions that are critical. The measurement system has been questioned. The Process map is shown in Figure 15.15.

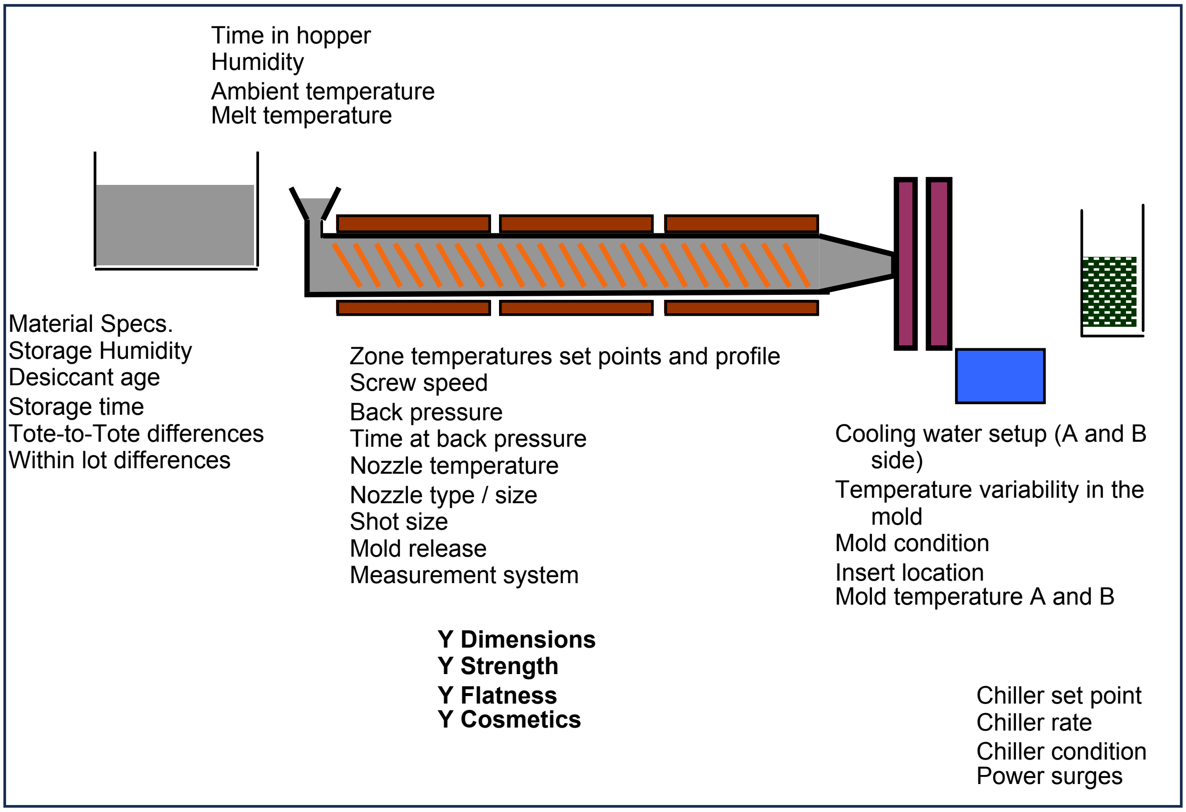


Figure 15.15: Injection Molding Process Map

Figure 15.16 depicts a crossed-nested sampling plan.

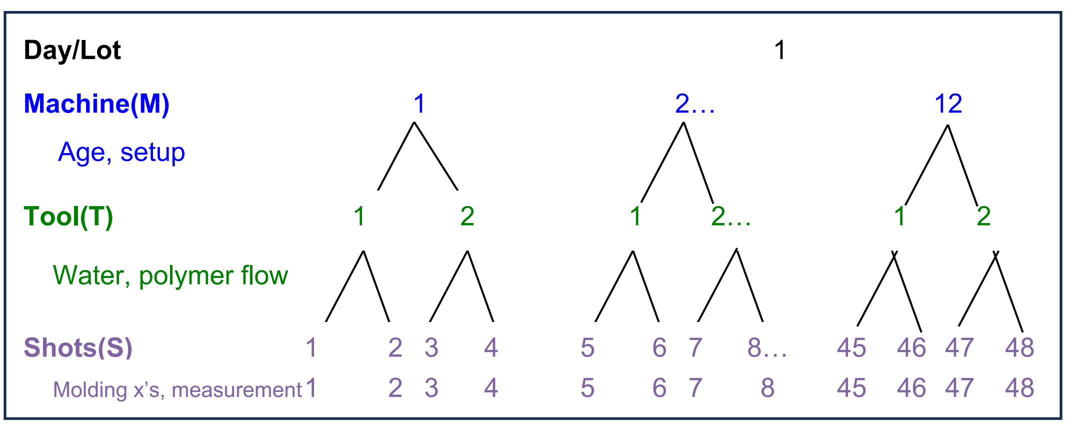


Figure 15.16 Sampling Plan with the Top Two Layers Crossed and the Bottom Layer Nested

For analysis of a crossed sampling plan, interactions need to be estimated. This precludes the use of control charts as demonstrated in the previous example. A range chart may be used to assess consistency at the bottom layer, but control charts otherwise do not handle interactions appropriately. In this case, we are left with quantitative analysis (e.g., variance components or ANOVA).

Prioritized Global Effects:

Possible Effect	***Rank***
Noise	3
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	1
Noise-by-factor interactions	5
Simple curvature (2nd order non-linear)	5
Complex non-linear (≥3rd order non-linear)	6
≥ Three-Factor Interactions (3rd order linear)	6
Stability	4
Leverage	3
Measurement uncertainty	3
Mean	3
Variation	3

Analysis Sequence:

0 Does the data set match the tree? The data set should have the same columns as the layers of the sampling plan. **The data MUST be in the same order as the sampling plan.** If it is not, reorganize it, to match how the data was acquired. Realize, statistical software programs sort time series vertically in the data table.

1 Is there enough variation in the data set? Has the phenomena of interest been captured? If not, next steps are to work on factors in the inference space for the next iteration. This means expand the inference space by allowing variables in the inference space to change.

2 Are there any patterns in the data? Are there any interesting data points? How does the data compare with predictions? See Figure 15.17.

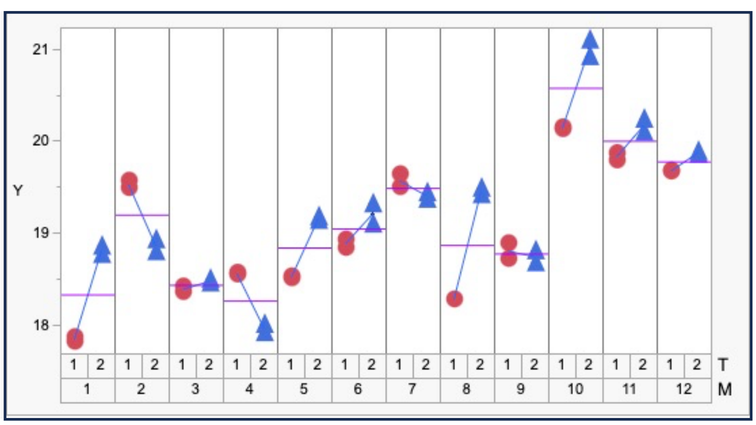


Figure 15.17: Variability Chart, Colored by Tool

3 Are there any special causes within subgroup? How does the within subgroup compare to the between subgroup sources of variation? Figure 15.18 is a range chart for the bottom layer of the tree.

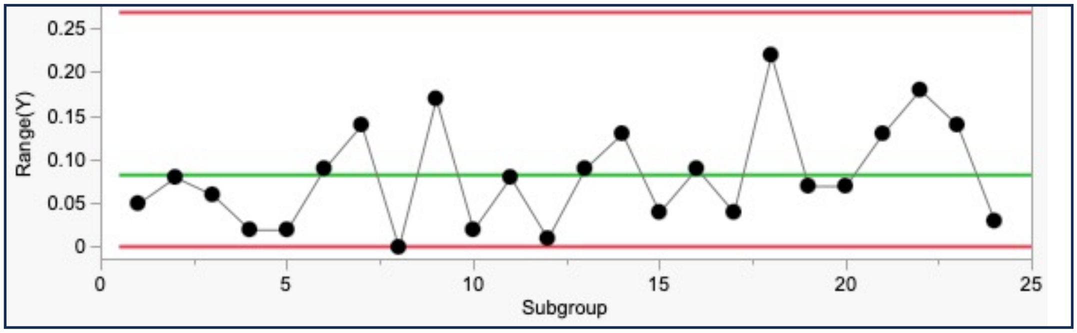


Figure 15.18: Range Chart

4 What are the quantitative components of variation? Perform ANOVA and calculate variance components:

Analysis of Variance

Source	DF	SS	Mean Square	F Ratio	Prob > F
M	11	22.21857	2.01987	5.8925	0.0033*
T	1	0.813802	0.8138	2.3741	0.1516
M*T	11	3.770623	0.34278	68.2440	<.0001*
S[M,T]	24	0.12055	0.00502	.	.
Within	0	0	0		
Total	47	26.92355	0.57284		

Variance Components

Component	Var Component	% of Total	20 40 60 80	Sqrt(Var Comp)
M	0.41927159	68.4		0.64751
T	0.01962576	3.2		0.14009
M*T	0.16888049	27.6		0.41095
S[M,T]	0.00502292	0.8197		0.07087
Within	0.00000000	0.0		0.00000
Total	0.61280076	100.0		0.78282

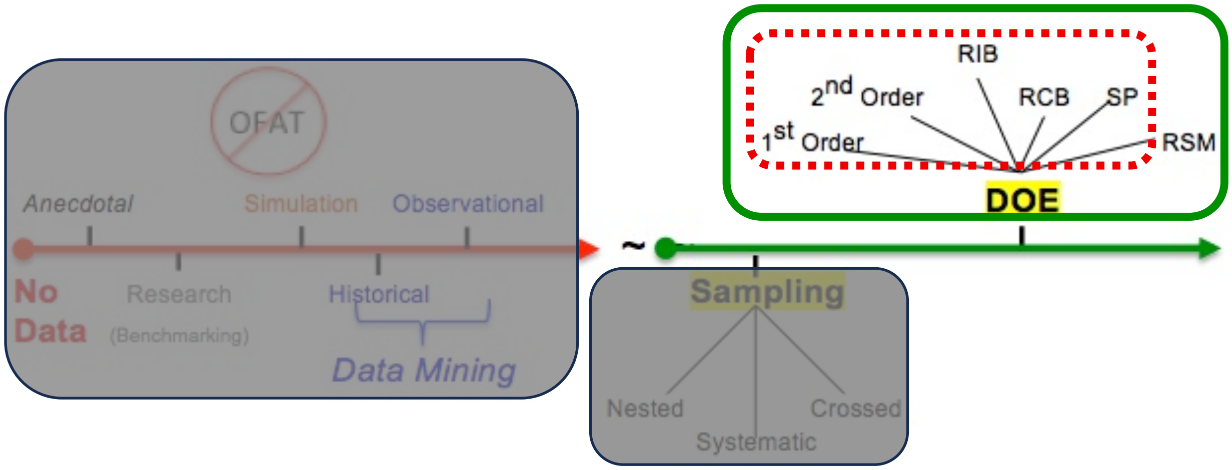
Figure 15.19: ANOVA and Variance Components

The Machine is the biggest source of variation in the study. There also appears to be a significant interaction between Machine and Tool. This means the effect of Tool variation depends on which Machine the Tool is in.

Transition to DOE

When the sense of urgency is high and sampling data cannot discriminate among competing hypotheses, the limitation is structural, not analytical. At that point, learning requires deliberately creating variation and changing the structure of the data, this is the role of designed experiments.

Chapter 16
Experimentation and Noise



Purpose of This Chapter

Designed experiments are powerful because they allow engineers to deliberately create variation in order to learn about causality. However, that power is easily misused when experimentation is treated as a method to pick the best result or as a software exercise rather than a thinking discipline. The purpose of this section is not to teach how to run DOE software, but to describe how designed experiments should be **planned, structured, analyzed, and interpreted** when the objective is learning. Experimental structure affects interpretation. The same statistical outputs can lead to very different conclusions depending on how noise was handled. Repeats, blocking, covariates, and split-plot structures do not change the model effects that can be estimated, but they change precision, stability, and the confidence with which conclusions can be drawn.

Experimental Planning Precedes Software

Before any statistical software is used, the investigator must be able to describe the experimental strategy clearly and completely. This description is best captured using a **Factor Relationship Diagram (FRD)**. The FRD externalizes the investigator's thinking about what is being manipulated, what is held constant, what varies unintentionally, and how degrees of freedom will be partitioned in analysis. At a minimum, the investigator should be able to answer the following questions *before* creating a design matrix in software:

- How many factors are being deliberately manipulated, and at how many levels?
- Which variables represent noise, and how will that noise be handled (e.g., repeats, blocks, covariate, split-plots)?
- What are the response variables (Y's), and how will they be measured?
- What model effects are prioritized for estimation?
- What resolution is required to separate those effects?
- What restrictions on randomization are being imposed, intentionally or otherwise?
- How many total runs are required, and are adequate resources available?

If these questions cannot be answered prior to software use, the experiment is not yet ready to be designed. Statistical software cannot supply this reasoning; it can only implement it. This planning discipline is emphasized throughout Sigma Science and is foundational to responsible experimentation. Experimentation is the manipulation of controllable factors, at different levels, to expose their effects on multiple response variables**, in the face of noise.** Deliberately manipulating factors, experiments create variation with intent. The analysis of experimental data must therefore respect the structure imposed by the design and the hypotheses that motivated it. Noise is often treated as an inconvenience or nuisance in experimentation, something to be randomized away, averaged out, or minimized. This view is incomplete and frequently counterproductive. Noise is not merely an obstacle to analysis; it is a fundamental feature of real systems. This chapter examines how noise affects the analysis of experimental data, how it limits precision, and how it constrains inference. More importantly, it explains how experiments can be structured and analyzed, to learn about noise rather than hide it. This distinction is central to robust design and to understanding causality under real-world conditions.

Precision Versus Inference Space

Precision and inference space are often in tension. Increasing precision by restricting noise improves the ability to detect small effects but limits applicability. Increasing inference space by allowing noise to vary improves relevance but reduces precision. There is no universally correct balance. The appropriate tradeoff depends on the purpose of the experiment and the stage of learning. Failing to acknowledge this tradeoff leads to experiments that are either elegant but useless, or realistic but indecipherable.

Restrictions on Randomization Are Analytical Facts

Restrictions on randomization are often introduced for practical reasons, equipment limitations, material batching, convenience of executing, or safety concerns. In some cases, they are introduced deliberately to manage noise. Blocking, split-plots, and repeats are not inconveniences to be "handled" by software. They are intentional design choices that define how variation is partitioned. Once imposed, these restrictions must be honored in analysis. Ignoring them produces misleading error terms, incorrect effect estimates, and false confidence.

> *"All industrial experiments are split-plots"*
>
> *Daniel*

Robustness is the Absence of Noise-by-Factor Interaction

Robustness is often described informally as "insensitivity to noise." More precisely, robustness refers to the absence of meaningful interactions between design factors and noise factors. If the effect of a design factor depends strongly on noise conditions, the design is fragile. If the effect is consistent across changing noise conditions, the design is robust. Identifying noise-by-factor interactions is therefore essential. These interactions are not nuisances; they are signals that design choices matter differently under different conditions.

Why This Must Be Done Early

Noise-by-factor interactions are best addressed during design, not after deployment. At the design stage, options exist: material changes, dimensional characteristics, positioning and orientations, and architecture modifications. Once a product or process is launched where materials are already selected and tooling is hardened, options narrow dramatically or options become financial not viable. What could have been designed out must now be managed operationally, often at great cost. This is why robust design is fundamentally a design-phase activity, even though it relies on the identification of future noise.

Prioritized Global Effects

Table 16.1 highlights the effects typically prioritized in experimental planning, execution and analysis. While experimentation can be used to estimate very complex models, the usefulness of such models is questionable.

Possible Effect
Noise
Main Effects (Factors)
Two-Factor Interactions (2^{nd} order linear)
Noise-by-factor Interactions
Simple Curvature (2^{nd} order non-linear)
Complex Non-linear ($\geq 3^{rd}$ order non-linear)
$\geq$ Three-Factor Interactions (3^{rd} order linear)
Stability
Leverage
Measurement Uncertainty
Mean
Variation

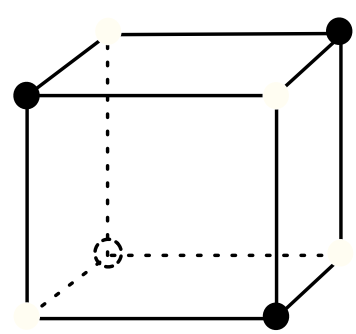
Table 16.1: Prioritized Effects for Experimentation

Graphical Depictions of Iterative Design

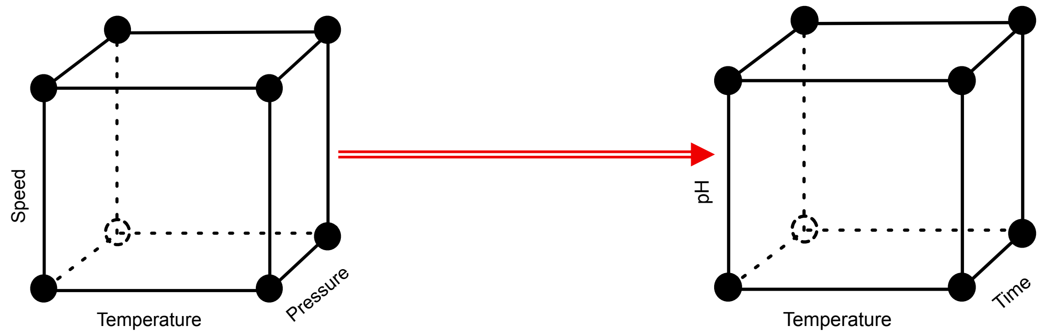
Experimental strategy evolves as knowledge increases. I always recommend starting with the assumption you do not know much. You don't know what you don't know. Depending on the assessment of where you are in the knowledge continuum, different strategies may be used to start or continue iteration.

The following graphical depictions illustrate common ways experiments iterate.

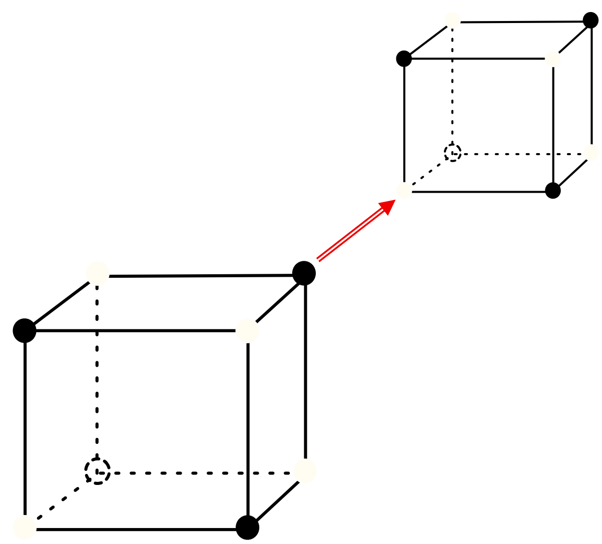
1. **Screening**. Identify and prioritize vital few factors from the many possibilities.



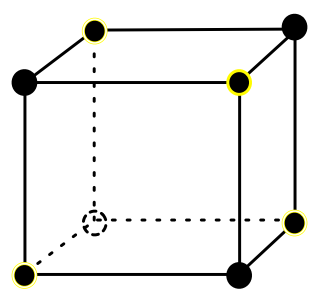
2. **Drop/Add factors.** Using process knowledge combined with data, insignificant factors may be removed from and others added to the investigation.



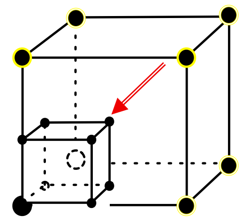
3. **Move space** (change factors and/or levels). Based on the initial screening design, a direction to follow for improvement should be available. If the data suggests a significant change in the experiment space is needed, you may want to use different factors.



4. **Add fractions.** This will make it possible to learn more about the space the experiment is currently being in. This is particularly important when the optimum is somewhere inside of the experiment space. Additional information can be obtained without running an entirely new design (interactions may be separated, increasing resolution). Fold over designs may be used to increase resolution and separate confounded terms.

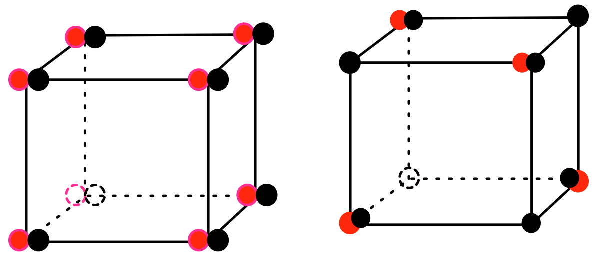


5. **Rescale**. Change factor level setting (conceptually).

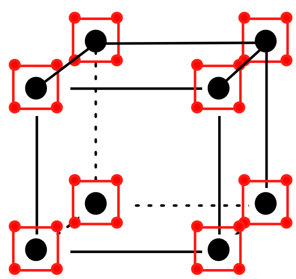


6. **Replication**. Complete or fractional replication of an experiment usually done over a different set of noise (Blocks). This is in contrast to repetition. Repeats will allow for estimation of short-term noise. Replication will increase the inference space and is be used to understand long-term noise and the noise-by-factor interactions. Incomplete blocks are useful for estimating the "main effect" of the block to look at large chunks of x's

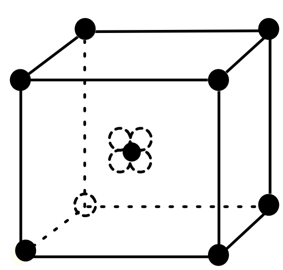
.



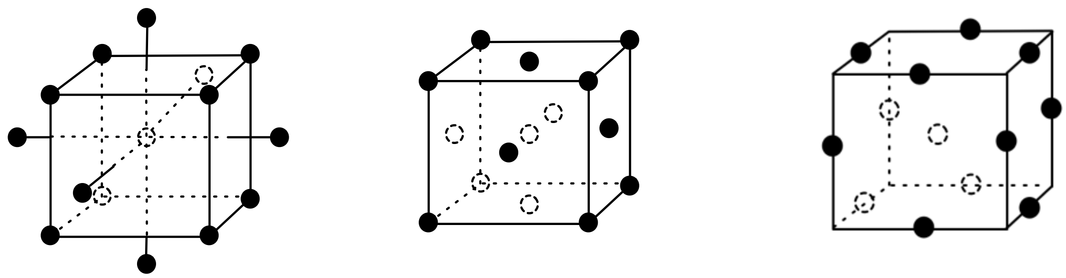
7. **Robust**. Simultaneously manipulate noise and design factors to identify opportunities to create robustness. Cross-over (inner-outer) arrays provide additional techniques to create robust designs.



8. **Curvature.** Test for non-linear relationships in the design space. Center points provide a reasonable estimate that is more efficient than three-level designs. To understand complex relationships more than three levels is necessary.



8. **Augmentation**. Performed to improve models, understand more complex relationships and build confidence in models. These augmented designs map the response surface more thoroughly. Central composite, Face Centered and Box-Behnken are examples:



Strategy for Handling Short-term Noise: Repeats[7]

In this strategy, repeats are introduced to partition short-term noise including components such as measurement system variation, within batch, within part, consecutive batches, consecutive parts and quantify within treatment variation. Repeats do not increase design resolution or separate confounded effects. They are used to reduce and quantify the within treatment noise, hence increasing precision. Repeats are shown on an FRD by dashed lines LOR's (see Appendix E for examples). There can be multiple layers of repeats in an experiment. These layers are nested within treatment. Each layer can be evaluated and summarized as appropriate.

Analysis starts with a practical look at the data and specifically, the within treatment data. If there are no unusual data points, it is acceptable to summarize the within treatment data. If there were unusual data points, further investigation should determine what to do with those unusual data points prior to further analysis (How much will the unusual data point effect the summary statistics? What happened there?) An average of the within treatment data reduces the variation within, hence increasing the precision of the experiment (lowers the water level without negatively impacting the inference space). In addition, a second Y is created, a measure of the variation of the within treatment data. The range (or any measure of the within treatment dispersion) is an estimate of the magnitude of the variation due to noise changing within treatment. It cannot be due to the treatments as the treatments are constant when the multiple data points are acquired. This Y can be analyzed separately from the mean. There are now two response variables describing the within treatment effects. It is good practice to look for correlations when you have a multivariate situation, not only to determine if there are relationships between the Y's, but also to test for any outliers. Correlation[8] between the mean and variation responses can indicate whether the same mechanisms influence both. It is wise to constantly evaluate the integrity of the data from an experiment. This is because DOE data sets are small, so singular, unusual data points can have a dramatic effect on quantitative analysis. Mahalanobis[9] outliers is a multivariate method for assessing if there are unusual data points. A second method is part of the Analysis of Good (ANOG). ANOG is simply an ordering of the treatments based on the response variable, best to worst. ANOG is used to reveal practical differences among treatments, patterns and anomalies that guide interpretation. The order you sort does not actually

[7] Crazy Cajun Case study analysis for a number of the noise strategies is in Appendix F
[8] A correlation coefficient |r| > 0.7 suggests a strong association
[9] Indian statistician, P.C. Mahalanobis

matter as you are simply looking for patterns of the Y and associated patterns of the factors and their levels. In addition, obtaining a moving range chart is a diagnostic to check for unusual data points. If there is a huge main effect, when the data is sorted in ANOG order, there will be a point out-of-control in the middle of the MR chart. If there are "jumps" in forth positions, that is indicative of two main effects or a 2nd order effect. If there is an out-of-control point at the beginning or end of a MR chart, that is evidence of noise effecting one treatment. The designs have an equal number of + and – level setting for each factor, one treatment being different could not be because of design structure. It must be due to noise changing during the experiment.

If there is evidence an unusual run based on the rest of the data from the experiment you should seek to understand why. What happened there? This is likely due to noise. This has the impact of raising the water level in the experiment, decreasing precision. A test of the impact of this run can be done with Rank[10] analysis which normalizes the jumps in the response variable between treatments.

The patterns observed from ANOG can be substantiated with statistical methods. The Normal/Half normal plots[11] assist in determining if those patterns are significant. These should only be used for the saturated model (all DF's assigned). Some software will add a line on the plot. This is a function of Lenth's pseudo-standard error. In some cases, this line provides guidance interpreting what are the random errors that are the basis for determining statistical significance (random errors would form a straight line on this plot). Unfortunately, there are instances where Lenth's PSE is not applicable to the effects being plotted. In those cases, interpreting which points make up the random errors must be judged. I use the "fat pencil" test to help interpret. Fit a pencil to the data, centered around an estimate of 0. Points covered by the pencil are random errors. The Pareto plot shows the practical significance of the factors. Remember to look for "jumps" between the bars of the Pareto chart, not just ordering.

Once the significant effects (both practical and statistical), are determined, simplify or reduce the model by removing the insignificant terms. This is done to create a more useful model and to get residuals, not to re-assess statistical significance. This reduced model is used to estimate residuals, a further check on model assumptions. There are a number of considerations that guide model reduction:

- engineering judgement augmented with predictions,
- magnitude of the R-square Adj (amount of variation in the data explained by the model),
- Δ between R-square Adj and R-square (test for overspecified model),
- root mean square error (standard deviation of the model),
- residuals analysis helps assess quantitative assumptions; NID(0, s^2), normally and independently distributed residuals with a mean of 0 and constant variance,
- leverage and residual plots (look for outliers, multicollinearity and confidence limits crossing horizontal line),
- AIC/BIC[12] (AIC favoring better prediction and BIC favoring simpler models), and
- p-values (less useful as statistical significance has already been assessed)

[10] In ANOG order, create a column with sequential whole numbers and analyze this additional Y.
[11] Application of half normal plots for analysis invented by Cuthbert Daniel
[12] Akaike information criterion and Bayesian information criterion used for comparing models

Residual plots help diagnose special cause and model inadequacy. This does not mean the raw data is unusual, it is an indication the model did not predict those actual values well. In some cases, this is due to model insufficiency or noise. Plotting those same residuals in run order may offer clues. The plots do not provide the answers, but stimulate questions about the model adequacy and possible further investigations.

Additional Application for Repeats

There can be multiple nested layers within treatment. When the response variable is ordinal, particularly involving human sensory perception, repeats or nested layers are useful[13].

Strategy for Handling Long-term Noise: Completely Randomized Replicates (CRR):

There are several options for replication. This is NOT the most effective strategy. The statisticians' default recommendation is randomized replicates. This is an appropriate strategy if noise has not been identified before the experiment is conducted. There are some advantages and disadvantages to this strategy. The advantages are:

- the inference space is more representative (broader),
- unit structure (noise) effects can be estimated by the error term (mean square error) and are theoretically less biased,
- the error term, quantification of the water level, can be used to test statistical significance of the design factors (with ANOVA).

Unfortunately, the error cannot be assigned to specific x's, and the precision for detecting design factor effects is compromised. Whether the error term is representative of the true random errors is still in question. If it is exaggerated, then you risk the type 2 error, if it is smaller than the true errors, you risk the type 1 error. It would be a good idea to compare the estimate of random errors with longer-term process data.

Since it is not possible to create a saturated model as there are unassignable DF's, Analysis of Variance (ANOVA) is used to assess the model effect significance. Residuals plots may provide insight to model integrity. Once the model is reduced, the model summary statistics and residuals and are re-evaluated.

Strategy for Handling Long-term Noise: Randomized Complete Block Design (RCBD):

In industrial experimentation, blocks are frequently a frame where noise, can reasonably be expected to remain constant (or are held constant) while that part of the experiment takes place. Subsequent blocks (replicates) are selected so the noise that was constant in the first block, **changes between** blocks. In this manner, there is increased precision for the **design structure and increased inference space**. Information regarding design structure is acquired across changing unit structure such as environmental conditions, variation in raw materials, and other known and identified noise factors. The purpose of blocking is to determine whether the effects of design factors are **repeatable** across blocks (changing

[13] See Printed Circuit Board Screening case study Appendix G

noise conditions) and to see whether the overall average of the experiment changes[14]. So, either one can gain confidence the results you have initially obtained will be useful in the future, there are opportunities to improve the robustness of the design, or the direction of your study needs to be expanded.

The advantages of the RCBD are increased inference space, greater precision and the ability to quantify noise-by-factor interaction effects for robust design. Analysis of RCBD follows the same sequence, PGQ. To analyze RCBD, start with the saturated model including block and block-by-factor interactions (full resolution of the block effects).

Strategy for Handling Long-term Noise: Randomized Incomplete Block Design (RIBD or BIB):
Blocking doubles the size of the experiment. If the investigation is still in the early stages, it may be appropriate to use incomplete, fractional blocks to increase inference space improve precision and investigate many confounded x's for further iteration.

Strategy for Efficiently Partitioning Noise with Design Factors: Split-plots
Split-plot experiments can be seen as a practical way to deal with certain issues which preclude randomization. When a design is not randomized, the restriction to randomization must be taken into account. One such design used to handle restrictions is a split-plot design.

There are two distinctly different reasons to restrict randomization. The first, an efficiency split-plot, is due to the desire to partition the unit structure and thereby *manage* the precision of the design structure using less resources. In this case, the whole plot experiment precedes the subplot experiment. The second, a convenience split-plot, is due to physical or economic reasons. The run-order for this type of split-plot design is executed by changing the subplot (SP) factor(s) while the whole plot (WP) factor(s) remains constant. Then the WP factor(s) is changed and the runs replicated. This minimizes the number of times the whole plot factor is changed. The runs of an experiment might be made in a split-plot fashion for the following reasons:

1. **Efficiency** of running the design:
   - Desired precision of design structure varies (e.g., design factors may interact with the noise and there is interest in estimating noise-by-factor interactions.)
   - Unit structure needs to be partitioned. The partitioning of the unit structure allows for greater precision when evaluating the design structure.

   To my knowledge, this type of split-plot design can only be run in situations where the experiment can be run in stages or in some time ordered sequence. The factors that make up the subplot are subordinate or subsequent to the factors in the whole plot. This can be extremely useful for product design as different designs can be tested over factorials of noise. While to overall noise in the experiment is the same as without restriction, the restriction partitions the noise into two smaller subsets of noise, thus lowering the water level for each plot. Doing this increases the precision of

[14] This approach differs from that in classical statistical textbooks where the concept of a block is presented as a tool to remove extraneous sources of variation (noise) so the effects of the design factors will be more clearly apparent.

understanding noise-by-factor interactions and therefore the robustness of the design to noise. This is the basis of Taguchi's inner/outer arrays and Cox's cross-over designs.

While practical analysis is similar as the other examples in this appendix, graphical and quantitative analyses are quite different. Because each plot of the split-plot design includes different unit structure, Normal and Pareto plots must be created for each split, a separate normal and Pareto plot for the whole plot and subplot. These designs often find more statistically significant effects due to the partitioning of the noise and lowering the water level in both the whole plot and the subplot. Again, it is important to assess practical significance on the Pareto Plot. The quantitative analysis is as if there were two separate experiments run.

2. **Convenience** of running the design:
   - One (or more) of the factors to be investigated is **hard or expensive** to change but is not noise. This 'hard to change" factor(s) will be designated the **whole plot (WP)**. The other factors will make up the **subplot (SP)**.
   - The experimenter wishes to make the experiment easier to execute.
   - If the unit structure is significantly partitioned, there will be a negative effect on the precision of the whole plot. In this case, it would NOT be desired to have a significant partitioning of the unit structure thereby not diminishing the precision of the whole plot.

   For convenience split-plots, the information about the whole plot factor is subject to *different* unit structure due to the minimized number of changes. For quantitative analysis, the whole plot error, rather than the mean square error of the subplot, should be used to determine whether the whole plot factor had a statistically significant effect. **Replication** of the treatments made varying the whole plot factor is necessary to have any estimate of the whole plot error and therefore a **quantitative** test of significance. Of course, replication requires the whole plot factor to be changed which is exactly what the experimenter wants to prevent. Note there will not be as much information available for these factors as there is for the other factors because these factors will not have been changed as many times. In the usual factorial situation, all variables are perturbed an *equal* number of times. It is also frequently the case the factors creating the WP, on which the amount of information is limited, are critical factors in the experiment. They are not typically used to adjust the product or process performance. Adjusting them multiple times may actually inflate the effect of noise in the experiment since the multiple manipulations are á typical.

   While convenience split-plots sacrifice information about the whole plot factors, they increase the precision of the subplot effects. Without replication, the analysis for the whole plot factor is solely practical.

Strategy for Handling Long-term Noise: Covariates

Understanding of the unit structure is most effectively accomplished by manipulation for at least the duration of an experiment using split-plot designs. If the noise cannot be specifically manipulated, it may be able to be "grouped/chunked" and evaluated using RCBD or RIBD (e.g., raw material may be sorted into two lots).

Another strategy to understand certain, **measurable** noise variables is to measure them and incorporate the measurements into the analysis. Covariates[15] can be recorded during any designed experiment (full or fractional factorial). ANCOVA (Analysis of Covariance) requires a degree of freedom in the experiment be used to estimate the (linear) effect of the covariate. Thus, you will lose a degree of freedom otherwise used to estimate some effects outlined in your aliasing (usually some higher order term) or used to estimate mean square error. Model effects are fixed effects. The effect of the covariate is a random effect. There are potential issues with using this strategy:

- Since the covariate is not specifically manipulated nor is it orthogonal to the other factors in the experiment, it is evidence of correlation vs. causation,
- It may be collinear with model effects,
- It adds another source of measurement error,
- Lagged effects may be misinterpreted,
- It can be a challenge to determine the one number for the covariate during the execution of the treatment.

Analysis requires differentiation of the type I (sequential) and type III (partial) sums of squares. Use Type I Sums of Squares to judge the effect of the covariate. This allows one to observe the effect of the covariate without the influence of other factors. Use the parameter estimates to help in the interpretation of the covariate. Use Type III Sums of Squares to judge the effect of the other model effects. Note, because of the effect of the covariate, the use of the parameter estimates for interpretation of the other model effects may be very difficult and misleading.

Tables 16.2 provides a summary of noise strategies for experimentation.

[15] There is a practical limit to the number of covariates in any model.

	Noise Strategy	Appropriate Usage	Implementation	Effect	Pros	Cons
1	Hold Constant	Only to be applied when it is known (data supports) that specific noise factors **do not** influence outcomes.	These noise factors are in the inference space.	Little effect as it is already known they have no impact.	Can reduce costs of testing.	Danger that while they have no main effect, they may interact with design factors. Smaller inference space may impact prediction.
2	Randomization	Noise has yet to be identified. Trying to prevent alignment of some noise with a factor in the experiment (hidden confounding).	Noise should be allowed to vary naturally. Randomizing should account for any items desired to be removed from the inference space.	Increases inference space and hopefully prevents hidden confounding of noise with model effects.	Increases the inference space while reducing bias and hidden confounding.	Decreases precision. Noise effects are unassignable. Practically difficult to execute.
3	CRR	Quantify variation between multiple treatments of the same factors. Used when the noise has not been identified.	Multiple experimentyal uints are created for each treatment combination. Run order is randomized. Execution requires complete setup for each replicate.	Theoretically unbiased estimate of error term. Used to estimate mean square error for statistical tests.	Increased inference space with unassignable error term. And an un-biased estimate of variation can be quantified for the basis of a statistical test (F-test)	Decrease in precision. Increase in cost to execute. Practically difficult to execute. Noise unassignable for future disaggregation.
4	Convenience Split-plot	Make the experiment easier and more economical to execute Paritions noise.	Vary levels of the subplot factors while the difficult to change factor (whole plot) is constant, then change the WP factor and replicate.	Allows estimation of the whole plot and sub-plot separately impacting the precision of both.	More practical and permits the running of the experiment in situations where randomization would precllude running the experiment. Increased precision of subplot.	WP precision is compromised.
5	Covariates	When noise is measurable.	Record values of the noise factor for each treatment.	Provides for assignment of the noise factor effect in the model. Creates a mixed model of both fixed and random effects.	Increases power of statistical tests. Model accounts for covariate effect.	No noise by factor interactions estimable. Additional measurement error. Only one value per treatment. Potential lage effect missed.
6	Center Points	As a means to get a reasonable estimate of the random errors and to test the linear assumtion during the execution of an experiment.	Run multiple (~≥8) center points randomly throughtout experiment.	Since the centerpoints may be least impacted by model effects, the variation in the center point runs may be a good estimate of the MSE for statistical tests.	Provides for estimating stability over the course of the experiment.	Only useful when all factors are quantitative.
7	Repeats/Nested	Effective when desire to understand variation within treatement & quantify short-term variuation (e.g., reduce measurement system, within part/batch).	Multiple measurements are taken for each experimental unit. Can also be structured for multiple layers (nested designs)	Increased precision, decreases variation due to within-treatment variation via averaging. Quantifies short-term noise for modeling.	Assessment of within treatment stability. Mean and variation can be assessed.	Increases cost relative to cost per measurement. No noise-by-factor interactions.
8	RIBD	Most useful in manufacturing processes where x's are currently unmanaged.	Noise is confounded with block, set at 2 levels, fractional blocks are run for each noise level. Block is created by aliasing with a higher-order effect.	Main effect of Block can be estimated to identify potential future factors for consideration.	Increased inference space, ability to analyze block effects.	Block effects are aliased with higher order effects and block-by-factor effects are aliased with other 2nd order effects.
9	RCBD	Effectively used with product design exposed to high levels of noise. Ability to alias multiple noise factors with block. Exposes noise-by-factor interactions leading to robust design.	Noise factors are confounded with the block, set at 2 bold levels, complete block is replicated for each noise level. Many noise factors can be aliased within each block.	Lowers effective noise present in each independent block, increasing precision. The effect of Block can be estimated. Broadens inference space.	Ability to analyze block and block-by-factor effects. Create increased inference space while increasing precision. Useful for robust design	Doubles the size of an unreplicated experiment.
10	Efficiency Split-plot	Partition noise to improve design precision while economizing resources	Restrictions are associated with sequential experiment or testing steps.	Allows for estimation of the whole plot and sub-plot separately impacting the precision of both.	Increases the precision of the WP and SP with significatly less resources. Specific WP-by-SP factor interactions identified.	More treatments than RCBD
11	Cross-over Array	Evaluate design factor effects over changing noise (robust design)	Essentially an efficiency split-plot where noise can be in either the whole plot or sub-plot.	Allows for precise estimation of noise-by-factor interactions.	Increases precision of both WP and SP and noise-by factor interactions.	More treatments necessary

Table 16.2: Summary of Experimental Noise Strategies

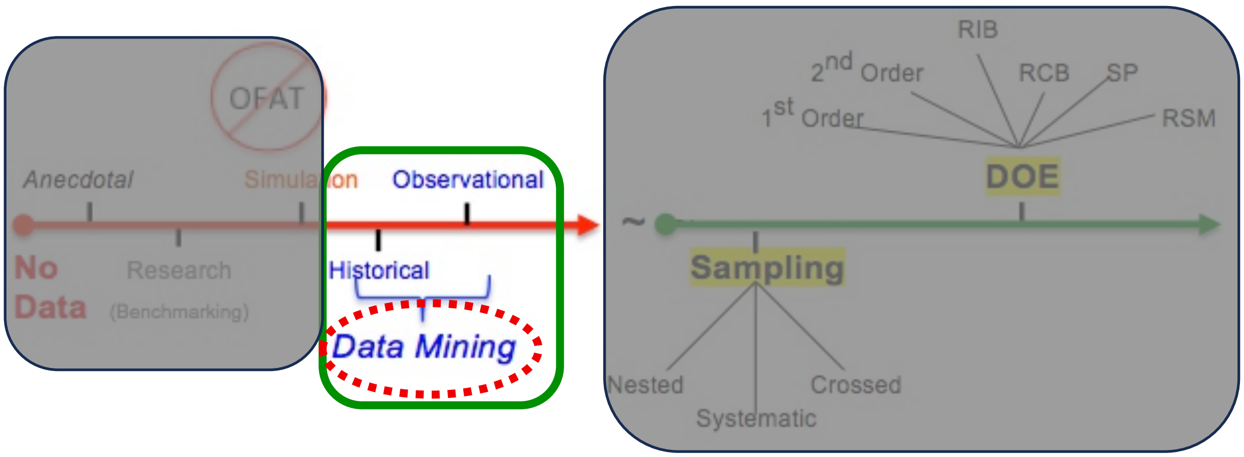
When Experimental Analysis Is Complete

Analysis of experimental data should stop when the dominant effects are understood, and ambiguity is reduced to an acceptable level. Proceed to the next iteration, which has already been predicted. Continuing to analyze beyond this point often creates false precision and distracts from learning. Experiments are part of an iterative process. Their value lies not in final answers, but in how effectively they guide the next question.

Summary

Analyzing experimental data requires respect for structure, resolution, and intent. When done thoughtfully, experiments provide powerful insight into causality. When analyzed carelessly, they produce results that appear rigorous but mislead. The discipline of experimental analysis lies not in statistical sophistication, but in intellectual restraint. An experiment does not reveal truth. It reveals what can be learned under the conditions you chose.

Chapter 17
Historical and Observational Data: What It Can and Cannot Tell You



Purpose of This Chapter

Historical and observational data are abundant. Modern organizations collect vast amounts of data as a by-product of operating processes, monitoring systems, quality checks, financial reporting, and compliance requirements. Because these data already exist, they are often viewed as "free" and therefore attractive for analysis. Historical data occupies the induction end of the data continuum. As we move toward strategic sampling and designed experiments, the control over how variation is manifested increases, and the strength of causal inference improves.

This chapter explains why historical data must be approached differently from sampling or experimental data, what kinds of insight it can legitimately support, and where its limitations are fundamental rather than technical. The objective is not to discourage its use, but to prevent misuse, particularly when the goal is to understand causality. Historical data is rarely sufficient for causal conclusions.

What Is Historical or Observational Data?

Historical (or observational) data are data that were collected for reasons other than the current investigation. The data is, however, from the organization. The analyst did not choose what variables were measured, how they were measured, when they varied, or why they varied.
Examples include:

- production records,
- warranty databases,
- sensor logs,
- transactional data,
- quality dashboards,
- and archived test results.

The defining characteristic of historical data is not age, but lack of design intent.

The Fundamental Limitation: No Control Over Variation

Causal learning depends on understanding how variation in inputs (x's) relates to variation in outcomes (Y's). With historical data, the analyst has no control over which x's varied, how they varied, or whether they varied independently. As a result:

- important factors may not have varied at all,
- multiple factors may have varied together,
- changes may be correlated with time, policy, or external events,
- and critical variables may not have been measured.

No amount of analytical sophistication can recover information that was never generated.

Correlation and Confounding is the Problem

It is common to caution that "correlation does not imply causation." While true, this statement is incomplete and often misunderstood. The deeper issue is confounding: when multiple factors change together, their individual effects cannot be separated. In historical data, confounding is the rule, not the exception. Statistical models can describe associations, but they cannot reliably determine which factor caused the change, or whether an unmeasured factor was responsible. Correlation may be evidence of causation, but it is not sufficient (see Figure 17.1) Without controlled variation and rational hypotheses grounded in physical or process understanding, correlation remains an observation rather than an explanation. This is not a software limitation. It is a structural limitation of the data.

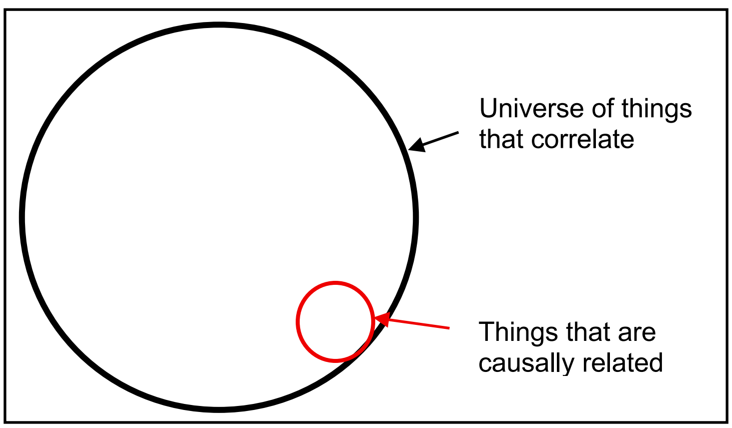


Figure 17.1: Relationship Between Correlation and Causation

A woman's husband had been slipping in and out of a coma for several months, yet she had stayed by his bedside every single day. One day, he motioned for her to come nearer. As she sat by him, he whispered, eyes full of tears, "You know what? You have been with me all through the bad times. When I got fired, you were there to support me. When my business failed, you were there. When I got shot, you were by my side. When we lost the house, you stayed right here. When my health started failing, you were still by my side. You know what?" "What dear?" she gently asked, smiling as her heart began to fill with warmth. "Maybe you should get the heck away from me!"

What Historical Data Is Good For

Despite these limitations, historical data can be valuable when used appropriately. Its strengths include hypothesis generation; Patterns in historical data can suggest where to look and stimulate what questions to ask. Contextual understanding; Historical data can reveal ranges of operation, rare events, and long-term trends. Screening and identification; Historical data can help identify candidate variables worthy of further investigation. Used in these ways, historical data complements, not replaces, designed learning.

Common Failure Modes in Data Mining

Several recurring failure modes appear when historical or observational data are treated as if they were experimental:

- Data sets are incomplete, labeled incorrectly, the wrong data type and need to be cleansed.
- Overfitting models that describe the past but fail to predict the future.
- Mistaking statistical significance for causal importance.
- Building complex models that cannot be executed or validated.
- Retrofitting narratives to explain coincidental patterns.
- Using the same data to generate and confirm hypotheses.

These failures are often reinforced by software that rewards model fit rather than understanding. Table 17.1 illustrates how common data integrity issues vary across the data continuum. As design intent increases, uncertainty decreases and interpretability improves. The progression reflects increasing control over variation and increasing confidence in causal interpretation.

Issue	**Typical Historical Data Set**	**Improved Historical Data Set**	**Observational**	**Strategic Sampling**	**Designed Experiment**
Hypothesis Prior to Acquisition[1]	Ø	Ø	?	√	√
Operational Definitions	Ø	√	√	√	√
Motivation for data acquisition	Ø	Ø	?	√	√
Measurement Error	Ø	?	?	√	√
Columns (Variables) Missing	Ø	Ø	?	√	√
Rows Missing[2]	Ø	√	√	√	√
Incorrect Data[3]	Ø	?	?	√	√
Correlation	Ø	Ø	√	?	√
Leverage	Ø	Ø	?	√	√
Cause - Effect	Ø	Ø	Ø	?	√
Balanced	Ø	Ø	Ø	√	√
Aggressive Level Setting	Ø	Ø	Ø	?	√
Changing Business conditions	Ø	Ø	√	√	√
Special Cause / Common Cause	Ø	Ø	Ø	√	√

Legend	
Rarely an Issue	√
Sometimes an Issue	?
Often an Issue	Ø

Table 17.1: Data Integrity Issues[16]

[16] Notes: 1. In historical data the data acquisition precedes the hypotheses. Analysis should NEVER precede hypotheses. 2.Too "busy" to enter data for a time period! Too much flux to enter data! Or chose not to sample over (non-representative data). 3. Putting "zero" to represent unknown. Typos (putting wrong number, putting a letter instead of a number). Filled in later. Lack of operational definitions (consistently or inconsistently misinterpreted). Deliberate distortion (games). Measurement issues.

Regression Is Not the Enemy, Misinterpretation Is

Regression is extremely flexible which makes it both a powerful and dangerous analytical tool. It is not inherently flawed. The danger lies in interpreting regression coefficients from historical data as causal effects. When multiple factors move together, regression may assign importance to variables that are simply correlated with the true cause. This is called multicollinearity. When regression is applied to observational data coefficients describe association, not causation, hidden variable bias is likely, and interpretation requires substantial domain knowledge. Regression can support hypothesis development, but it rarely settles causal questions on its own.

The Illusion of Predictive Power

It is tempting to evaluate historical models based on predictive accuracy alone. While prediction is valuable, it does not guarantee understanding. A model may predict well because underlying conditions have not changed, confounding relationships persist, or noise overwhelms structure. When conditions change, and they always do, models built on historical coincidence often fail catastrophically. Prediction without causal understanding is fragile. When operating conditions change, only causal relationships remain reliable.

When to Stop Mining and Start Designing

Historical analysis should prompt a decision point. Signals that it is time to move on include:

- ambiguity that cannot be resolved analytically,
- sensitivity of conclusions to modeling choices,
- reliance on untestable assumptions,
- or recognition that critical variables are missing.

At this point, the appropriate response is not more analysis, but better data, through directed sampling or designed experiments. Historical data might suggest where to dig. It does not dig for you.

Summary

Historical and observational data are indispensable in modern engineering and science, but their value lies primarily in exploration, orientation, and hypothesis generation. They are poorly suited for establishing causality or guiding irreversible decisions. Respecting these limits is not a weakness. it is a mark of professional judgment.

Closing Thought

Data collected without intent can inform curiosity, but data collected with intent is required to learn cause.

Part III
Deployment, Leadership, Organizational Integration and Sustainability

Part III explains how Sigma Science lives inside an organization. It addresses deployment, culture, leadership behavior, project selection, communication, metrics, and sustainability. This is where methodology meets reality. In Part I, the emphasis was on thinking, how to frame problems, develop rational hypotheses, and link tools to situations. In Part II, the focus shifted to structure, how the way data are acquired governs what analysis can legitimately say. In this final section, the objective is integration: applying the methodology iteratively in real organizational settings where uncertainty, constraints, urgency, and competing priorities are always present. Real improvement does not occur because a tool was applied correctly. It occurs when disciplined reasoning is sustained over time. Technical insight must be coupled with judgment. Data must be linked to action. Hypotheses must evolve as knowledge increases. Decisions must be made with incomplete information. Part III examines how Sigma Science operates in this environment, how studies are sequenced, how iteration is managed, how ambiguity is reduced, and how learning is translated into meaningful process and product improvement. This part is not about adding new statistical techniques. It is about developing intellectual discipline: knowing when to stop analyzing, when to redesign a study, when to change the response, when to broaden the inference space, and when the evidence is sufficient to act. Sigma Science is not a linear procedure. It is an iterative framework for learning, grounded in causality and guided by scientific inquiry. The ultimate objective is not elegant models, impressive analysis outputs, or technical sophistication. It is understanding, understanding strong enough to support prediction, confident enough to guide decisions, and robust enough to endure changing conditions. The first two parts of this book established the intellectual foundation of Sigma Science. This part examines how Sigma Science can be deployed within an organization, how projects should be selected to encourage discovery and learning, and how leadership behavior, communication, incentives, and metrics influence the success of investigative work. It also considers the role of the individual practitioner and the challenge of sustaining structured inquiry over time. Ultimately, Sigma Science succeeds when disciplined reasoning becomes part of how an organization thinks about problems. Part III explores how that transition occurs.

A man in a hot air balloon realized he was lost. He reduced altitude and spotted a guy below. He descended a bit more and shouted, "Excuse me, can you help me? I promised a friend I would meet him an hour ago, but I don't know where I am." The guy below replied, "You're in a hot air balloon hovering approximately 30 feet above the ground. You're between 40º and 41º north latitude and between 59º and 60º west longitude." "You must be an engineer," said the balloonist. "I am," replied the guy "How did you know?" "Well," answered the balloonist, "everything you told me is, technically correct, but I've no idea what to make of your information, and the fact is I'm still lost. Frankly, you've not been much help at all. If anything, you've delayed my trip." The guy below responded, "You must be in Management." "I am," replied the balloonist, "but how did you know?" "Well," said the guy, "you don't know where you are or where you're going. You have risen to where you are due to a large quantity of hot air. You made a promise which you've no idea how to keep, and you expect people beneath you to solve your problems. The fact is you are in exactly the same position you were in before we met, but now, somehow, it's my fault."

Chapter 18
From Method to Movement: What Deployment Really Means

The first two parts of this book addressed how to think and how to learn. Part I introduced the Sigma Science methodology as a disciplined framework for understanding causality. Part II addressed how data are acquired, structured, analyzed, and interpreted. Those sections focused on competence, on improving the technical capability of individuals to reason from data in a scientific and iterative way.

Part III shifts the focus from competence to integration. Deployment is not solely the act of delivering training. It is not the creation of slides, templates, or certification levels. It is not even the successful completion of a few well-executed projects. Deployment is the process of embedding critical thinking into the daily operation of a business or at least into the thinking of the individuals within it. It is the transition from method to movement. Most organizations recognize the importance of continuous improvement. They understand that product performance, customer satisfaction, cost, reliability, and throughput must improve to remain competitive. Continuous improvement implies and requires change. And change, particularly change that alters how people think, requires more than tools. I start a conversation with potential clients with a Thought map (Figure 18.1).

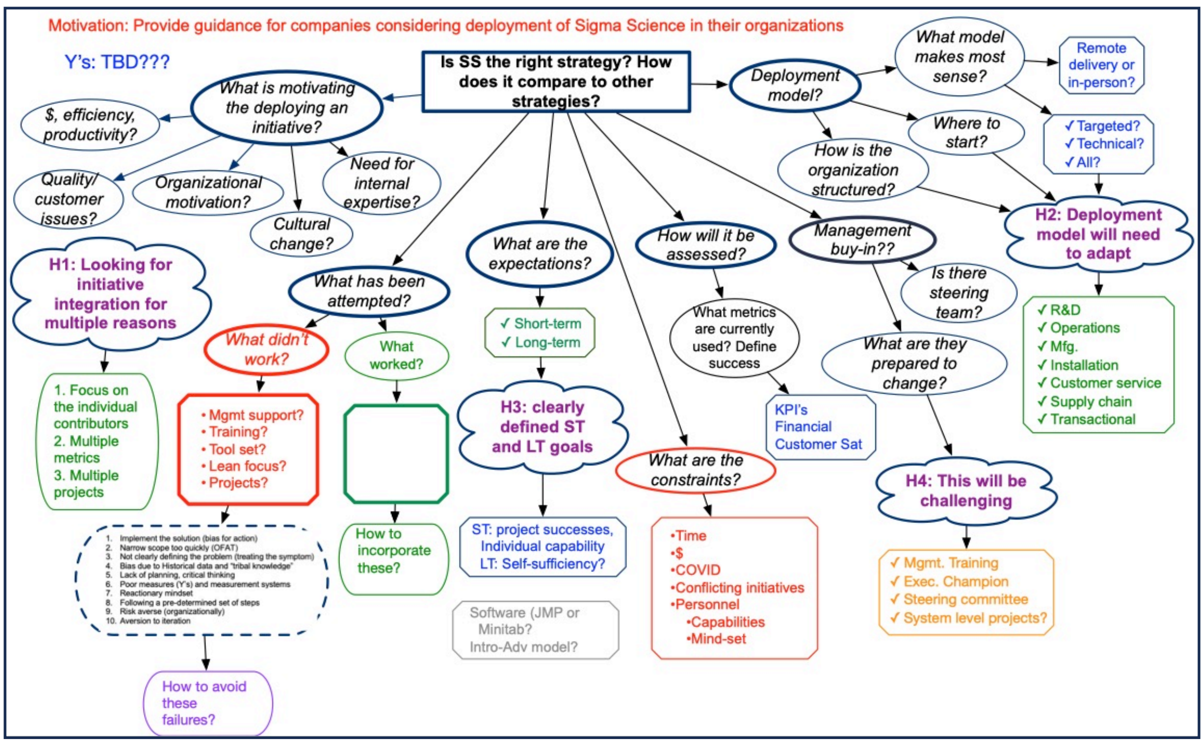


Figure 18.1: Thought Map for Initial Engagement

There are as many deployment strategies as there are companies attempting improvement initiatives. No single strategy can possibly be optimal across all organizations, just as no single tool can be appropriate for all situations. Each organization has its own structure, culture, incentives, resource constraints, leadership style, and technical maturity. Any deployment strategy must align with those realities. Sustainable change requires several elements:

- An organization willing to accept change
- Enlightened and motivated leadership
- Alignment of metrics and goals
- Clear communication of intent
- A critical mass of technically competent individuals
- Credible success stories

Without these elements, even the best methodology will struggle to survive.

Common Failure Modes

Improvement initiatives fail in predictable ways. Many of these failure modes are not technical. They are organizational. One failure mode is lack of management commitment, not in words, but in time and resource allocation. When leaders demand short-term results without allowing for iteration, sequential learning, and occasional failure, the initiative becomes reactive rather than investigative. Another failure mode is multiplicity of initiatives. When improvement efforts compete for the same limited resources, confusion and fatigue develop. If individuals cannot clearly articulate the intent of the initiative or how it aligns with broader objectives, cynicism follows. Poor measures and poor measurement systems are another common constraint. Metrics that are misaligned, misunderstood, or poorly defined drive unintended behavior. When metrics become targets rather than indicators of process behavior, people learn to distort the system or the data. Risk aversion also inhibits progress. Sigma Science is grounded in scientific method, observation, hypothesis generation, testing, refinement. Learning is motivated by understanding variation. Experimentation requires the possibility of being wrong. Organizations that penalize failure prevent discovery. Finally, inadequate education is a persistent issue. When training focuses primarily on vocabulary and tools rather than on reasoning and interpretation, the initiative becomes mechanical. The use of statistical software does not guarantee understanding.

> *"There are three ways to get better figures: Improve the system, Distort the system or Distort the figures"*
>
> *Brian Joiner*
>
> *And the Ross Corollary, the 4th way: Blame someone*

Deployment Is Not Tool Distribution

One of the most common misconceptions is that deployment consists of distributing tools across the organization. Tools are necessary but insufficient. The real objective is to cultivate investigative thinking, disciplined framing of problems, explicit hypotheses, deliberate data acquisition, rational interpretation, and iterative refinement. Sigma Science is not about statistics per se. It is about structured thinking applied in pursuit of causality. Deployment, therefore, must focus on developing individuals who can:

- Perform situational analyses
- Develop and articulate rational hypotheses
- Predict possible outcomes before data are collected
- Select sampling or experimental strategies appropriate to the question
- Interpret outputs through the lens of engineering knowledge
- Iterate in a disciplined manner

The objective is not to produce statisticians. It is to augment engineers and scientists with improved deductive reasoning so they can better use their technical knowledge.

At Least Two Phases of Deployment

In practice, deployment often occurs in two phases, though these phases may repeat over time to build critical mass. The first phase is edification of the management team. Leadership must understand the methodology sufficiently to identify appropriate projects, identify participants, define expectations, allocate resources, and model the behavior required for iterative learning. Managers must understand that progress is not always linear and that learning often precedes visible improvement. The second phase is development of the technical staff through application. Training must be coupled with real projects. The methodology cannot be learned in isolation from practice. Between formal sessions, individuals must apply the concepts to actual work. Project reviews, mentoring, and site visits are integral to developing judgment. This second phase is where critical thinking begins to influence daily operations. Over time, repetition builds competence. Competence builds confidence. Confidence builds credibility. There are, of course, pros and cons to any approach. I have had experiences where the management team was "on board" and actively showing support and the methodology thrived. But when those managers moved on, so went the methodology. I have also had success with organizations where the management team was not fully engaged. Though the deployment was limited to engineers who recognized the benefits.

Deployment Requires Patience and Iteration

Real improvement rarely occurs in a straight line. Early projects may fail to produce dramatic financial returns. Definitions will need to be refined. Hypotheses will be wrong. Sampling plans will need adjustment. Experiments will reveal unexpected results. Failure in this context is not a setback. It is feedback. Organizations must be prepared to iterate, not just technically, but organizationally. Deployment models will need adjustment. Curriculum will evolve. Metrics may need refinement. Project selection criteria will change as maturity increases. The danger lies in abandoning the methodology prematurely because it did not produce immediate, visible gains. Sustainable improvement is cumulative. It builds from disciplined learning applied repeatedly over time. This is particularly a challenge when working in the design world. First there is the ridiculous expectation of "designing it right the first time". Persons who believe this do not understand design. Second, it is difficult to cost justify the application of the methodology. You almost need to design it wrong, determine that cost and then iterate the design to show the savings. Perhaps there is need for a "leap of faith".

I recall a discussion with J.C. Anderson, a Senior Vice President of Operations for a major appliance manufacturer. He was concerned that the organization was not seeing enough "breakthrough" projects emerge from its Operational Excellence program. At the time, participants in the program were required to identify three projects, each with the potential to deliver $200,000 in savings. I suggested that imposing savings thresholds as a prerequisite for project selection could unintentionally discourage many important investigations. Some of the most valuable work begins with uncertainty, and its eventual benefit cannot be predicted at the outset. When projects are filtered primarily by projected savings, opportunities to address deeper systemic issues are often overlooked. He nodded in agreement as we discussed the concern, and our meeting ended. Shortly afterward I saw a memo announcing a change to the program requirements. The projected savings requirement had been increased from $200,000 to $500,000 per project. Surprised, I contacted him and asked if he had heard my advice. He assured me that he had. His reasoning, however, was different from what I had expected. By setting the projected savings requirement so high that very few projects could realistically qualify, he hoped the organization would be forced to reconsider how projects were being defined and selected in the first place. Whether intentional or not, the episode illustrates an important reality: the way organizations structure improvement programs strongly influences the type of work that gets pursued. When financial thresholds dominate project selection, learning-oriented investigations may be displaced by efforts designed primarily to meet numerical targets. Sustainable improvement requires balancing accountability with the freedom to explore problems whose benefits may not be immediately quantifiable.

From Individual Competence to Organizational Culture

Ultimately, deployment succeeds when critical thinking becomes cultural rather than programmatic. When engineers arrive at meetings with thought maps rather than just opinions, the culture shifts. When managers ask about hypotheses rather than just outcomes, the culture shifts. When sampling and experimentation are used deliberately rather than reactively, the culture shifts. When interpretation precedes reaction, the culture shifts. At that point, Sigma Science is no longer an initiative. It is simply how the organization thinks. Deployment is the bridge between method and culture. It is the process of transforming structured reasoning from an individual capability into an organizational norm. It becomes an expectation.

Part III examines how to build that bridge, how to select projects, align leadership, manage metrics, navigate organizational constraints, and sustain learning over time. The methodology is necessary. The tools are necessary. But without disciplined deployment, neither will endure.

Chapter 19
Choosing a Deployment Strategy: Alignment Before Action

There is no universal deployment model. Organizations often ask for a recommended strategy as if there were a best practice independent of context. There is not. Just as no single experimental design is appropriate for every technical question, no single deployment model is optimal for every organization. Structure must align with intent. The first question, therefore, is not *"How should we deploy Sigma Science?"* The first question is *"What are we trying to accomplish?"* Are we attempting to energize the entire organization? Are we attempting to deepen the technical capability of engineers? Are we addressing a specific crisis? Are we trying to build internal self-sufficiency? Without clarity on intent, deployment becomes reactive and fragmented. I often start a dialogue with a new client by interrogating their motivation.

On another occasion I received an inquiry from Marc Calavetti, Vice President of Research and Development at a manufacturing company. He was familiar with my work from a previous employer and contacted me to ask whether I would be available to assist his organization. He also asked if I would submit a proposal outlining how Sigma Science might help them. Rather than immediately preparing a proposal, I responded with a request of my own. I asked him to first prepare a **Thought Map** describing what he hoped to accomplish and why he believed Sigma Science might be relevant to the situation. This request was not intended to delay the discussion, but to clarify it. Too often organizations seek assistance before they have clearly articulated the problem they want to solve or the outcomes they hope to achieve. A Thought Map forces the essential questions to be addressed: What are we trying to accomplish? Why does it matter? What assumptions are being made? What evidence supports the concern? Marc agreed to the request and provided the Thought Map shown in **Figure 19.1**. The map became the starting point for our discussions and helped ensure that any subsequent work was grounded in clearly stated objectives and reasoning rather than vague expectations or general interest in improvement methods. The Thought map made clear what problem the organization believed it had, and just as importantly, what it did not yet understand.

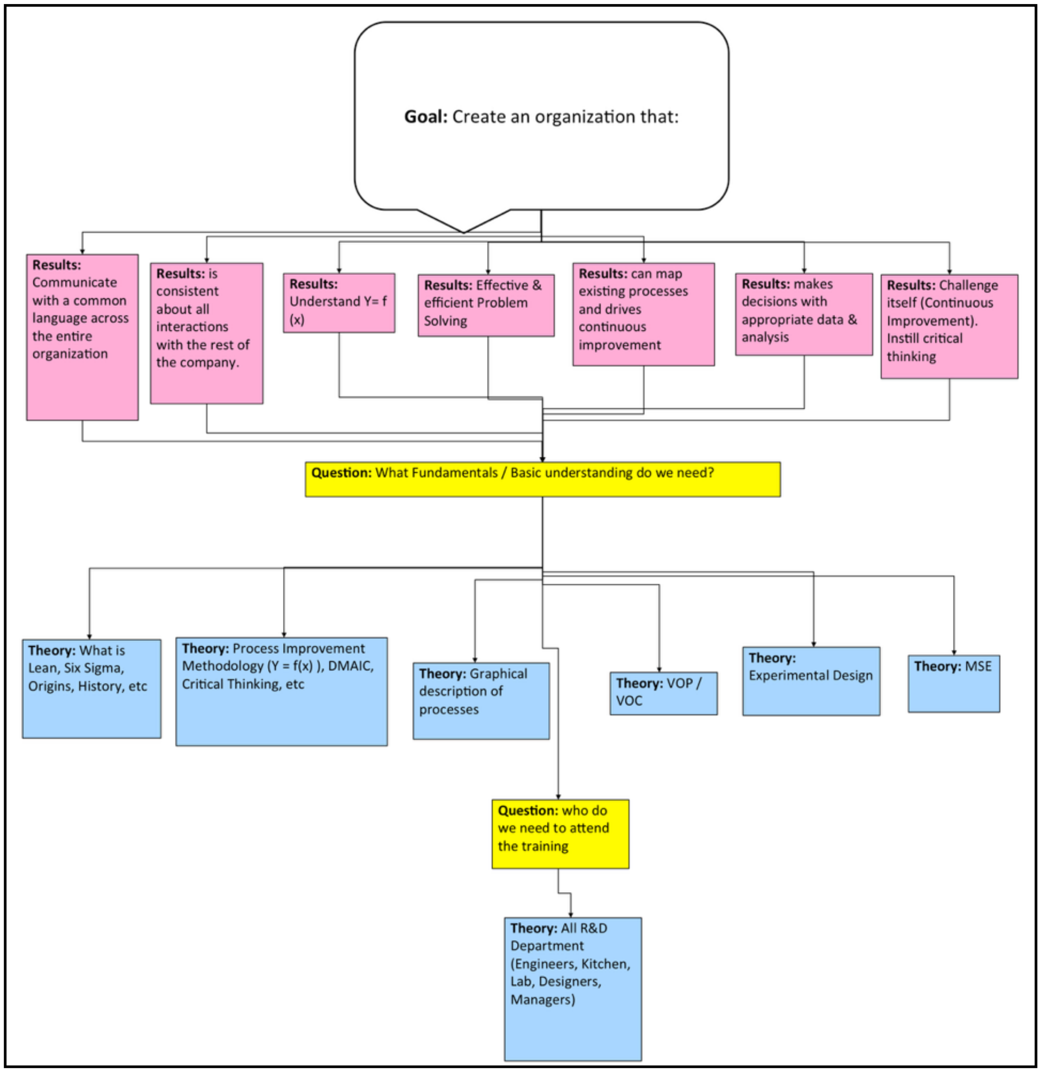


Figure 19.1: Thought Map for Discussion

Four Broad Deployment Archetypes

While real organizations often blend approaches, most deployment efforts resemble one of four archetypes.

1. Company-Wide Involvement

In this model, the initiative is positioned as a unifying improvement framework for the entire organization. A common language is established. Awareness is broad. Many individuals are trained at some level. The strength of this approach is constancy of purpose. A shared vocabulary and philosophy can energize the culture. Cross-functional participation may increase. Improvement becomes visible and organizationally sanctioned. The weakness is dilution. When everyone is trained at a basic level, depth is often sacrificed for breadth. Technical rigor may suffer. Projects are sometimes selected for symbolic reasons rather than for discovery value. Cynicism can develop if results are overstated or if “sigma levels” become targets detached from process knowledge. Company-wide deployment can succeed, but only if leadership resists the temptation to measure progress by training counts rather than by increased understanding.

2. Technical Focus

In this model, the primary objective is to enhance the capability of engineers and technical staff. The emphasis is not motivational, it is developmental. Projects are typically centered around manufacturing variation, product performance, reliability, or design robustness. There is an explicit element of discovery. The curriculum can be deeper because the participants share quantitative backgrounds. The strength of this approach is rapid transfer of knowledge into design and manufacturing work. Because the individuals are technically aligned, more sophisticated methods can be taught and applied effectively. Success stories tend to be technically credible. The weakness is potential isolation. If management and non-technical functions are not educated sufficiently to understand the methodology, resource allocation decisions may conflict with investigative work. Without managerial alignment, even technically sound work can stall. Technical focus deployment works well when leadership understands enough to support iteration and allocate resources intelligently.

3. Application-Specific Focus

This model resembles just-in-time training. Individuals are selected to address specific problems. The curriculum evolves alongside the project. The strength is expediency. When a pressing issue demands resolution, concentrated effort can yield rapid progress. The weakness is fragility. Because the effort is narrow and often temporary, learning may not generalize. Once the specific problem is resolved, momentum can dissipate. There is limited institutional memory unless deliberate effort is made to capture and disseminate learning. Application-specific deployment is useful for crises, but insufficient as a long-term cultural strategy.

4. Becoming Self-Sufficient

In this model, selected individuals are developed deeply and repeatedly. The goal is internal capability, not external reliance. Projects are chosen deliberately to build competence. Repetition is emphasized. Individuals become mentors and catalysts inside the organization. The strength of this approach is sustainability. Knowledge compounds. Judgment develops through iteration. The organization becomes increasingly capable of structured reasoning without external prompting. The weakness is time. This model requires patience and tolerance for learning curves. Short-term financial metrics may not immediately reflect the investment. For organizations committed to long-term cultural change, this model provides the strongest foundation.

Alignment With Organizational Objectives

Deployment strategy must align with organizational objectives. If the objective is to reduce cost rapidly, then project selection and training structure must reflect that urgency, but without sacrificing discovery. If the objective is to enhance product robustness, then the curriculum must emphasize experimentation and robustness methods. If the objective is cultural transformation, then leadership behavior must change alongside technical training. A common error is misalignment: broad training without project clarity, deep technical training without managerial understanding, or aggressive financial targets without tolerance for iteration. Alignment requires explicit discussion of:

- What success looks like
- What metrics will be used

- How resources will be allocated
- What level of iteration is acceptable
- What risks are tolerable
- What is the expectation for deployment time

Without alignment, deployment degenerates into a series of disconnected training events.

Project Characteristics Must Match Deployment Strategy

Deployment strategy influences project selection (see Chapter 20, Project Selection). Projects intended for Sigma Science training should contain an element of discovery. They should require investigation into causal structure (Y = f(x)). Pure implementation projects, where the tasks are known and execution is the primary requirement, are better suited for project management training than for Sigma Science methodology. When deployment is technical in nature, projects should require expanded analytical capability. When deployment is organization-wide, projects may focus on visibility and engagement. When the objective is individual development, projects should allow repetition and iteration. There is no correct project independent of strategy.

The Most Common Strategic Error

The most common error in deployment is confusing activity with progress. Training sessions delivered. Certificates issued. Metrics updated. Savings claimed. None of these activities guarantee improved understanding of causal structure. If individuals cannot articulate hypotheses, predict outcomes, interpret outputs through process knowledge, and iterate intelligently, then the methodology has not been deployed, regardless of how many people were trained. Deployment is successful only when structured thinking becomes habitual.

Before Acting, Decide

Before launching an initiative, leadership should answer:

- What are we trying to accomplish?
- Which deployment archetype aligns with that intent?
- What project characteristics are required?
- Who should be involved in the transformation?
- What level of iteration are we prepared to tolerate?
- How will we measure increased understanding, not just financial outcome?

Alignment before action prevents wasted effort. The methodology described in this book is flexible. It can be deployed in many configurations. But flexibility does not eliminate the need for discipline. In technical work, structure precedes experimentation. In organizational work, structure must precede deployment. Part III continues by addressing the most leverage-rich element of deployment: project selection.

Chapter 20
Project Selection: The Leverage Point of Deployment

Project selection is the single most influential factor in the success or failure of a Sigma Science initiative. Many improvement programs focus heavily on training curriculum, certification structures, and reporting metrics. While these elements can be useful, they are secondary. The real driver of learning and credibility is the nature of the projects people work on. Projects determine what individuals learn, how quickly they learn it, and whether the organization perceives the methodology as useful. Poor project selection can undermine even the most technically competent practitioners. Conversely, well-chosen projects can accelerate learning, demonstrate value, and establish credibility throughout an organization. For this reason, project selection should be deliberate rather than opportunistic. I once had a manager make the following statement in regards to discussing a potential project: "We don't want to turn this into a science project". I asked why not?

> *"The mere formulation of a problem is often far more essential than its solution, which may be merely a matter of mathematical or experimental skills. To raise new questions, new possibilities, to regard old problems from a new angle requires creative imagination and marks real advances in science."*
>
> *Albert Einstein*

Discovery–Execution Continuum (Figure 20.1)
Not all projects are the same. Ultimately, all projects should involve some amount of execution or the project will be abandoned. The question is when do you have sufficient insight to execute. Execution projects involve implementing known solutions. The problem is understood, the causes are known, and the task is primarily organizational or logistical. These projects may be important to the business, but they provide little opportunity to expand understanding of causal structure nor do they provide an opportunity for the engineers to learn. Discovery projects are different. They involve uncertainty about why a process behaves the way it does or how a product can be designed robust to user influence. The objective is to understand relationships between variables, to learn something new about the system. Projects suited to the Sigma Science Methodology should contain a meaningful element of discovery. Notice I do NOT recommend calling them Sigma Science Projects. This title suggests the thinking should only be done on certain projects.

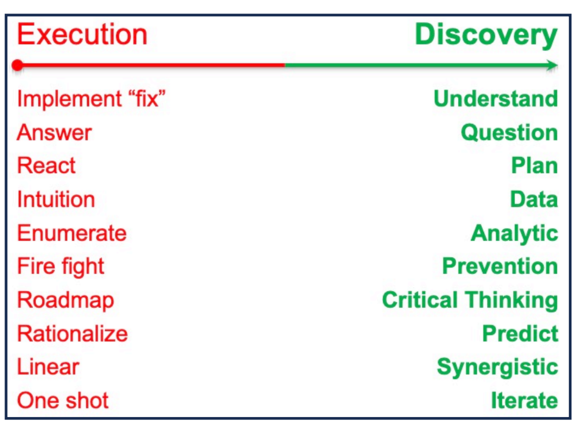


Figure 20.1: Execution and Discovery Project Continuum

This does not mean that execution is unimportant. Organizations must implement improvements to achieve results. However, the methodology described in this book is designed to understand causality. When projects contain no discovery component, the tools and thinking processes become unnecessary. In practice, most real projects exist along a continuum between discovery and execution. At one extreme are projects that are almost entirely execution. At the other are projects that are fundamentally scientific investigations. The most productive training projects oscillate between discovery to execution, investigating causes while pursuing practical improvement.

Why Discovery Matters

Discovery projects are powerful because they develop capability. When engineers and scientists investigate a situation where causal structure is not fully understood, they will be most effective if they apply the entire methodology:

- framing the situation,
- situation diagnostics,
- generating hypotheses,
- predicting data and all possible outcomes,
- acquiring data deliberately,
- interpreting analysis outputs, and
- iterating based on what is learned.

Through this process, individuals develop judgment. They learn to connect data with engineering knowledge. They learn how to structure investigations and interpret results. Execution projects, by contrast, often reinforce procedural thinking rather than causal reasoning. For organizations attempting to develop internal capability, discovery is essential.

Characteristics of Good Projects

Hint: Don't Call Them "Sigma Science Projects"

I have been investigating and collecting data on improving project work for many years. My thought map is in Figure 20.2. Projects suitable for Sigma Science **deployment** typically have several characteristics. First, the situation should involve variation, phenomena or performance that is not well understood. If the cause is already known, investigation adds little value. Second, the system should be observable. Variables can be identified, measured, or manipulated. Without the ability to observe or vary factors, learning is limited. Third, the project should allow iteration. Early hypotheses will often be incomplete or incorrect. The study should allow refinement and follow-up investigations. Fourth, the project should matter to the organization, or at least the management team. When projects address real operational challenges, learning becomes immediately relevant and credibility increases. Finally, the scope should be manageable. Projects that attempt to solve too many problems simultaneously dilute learning and slow progress.

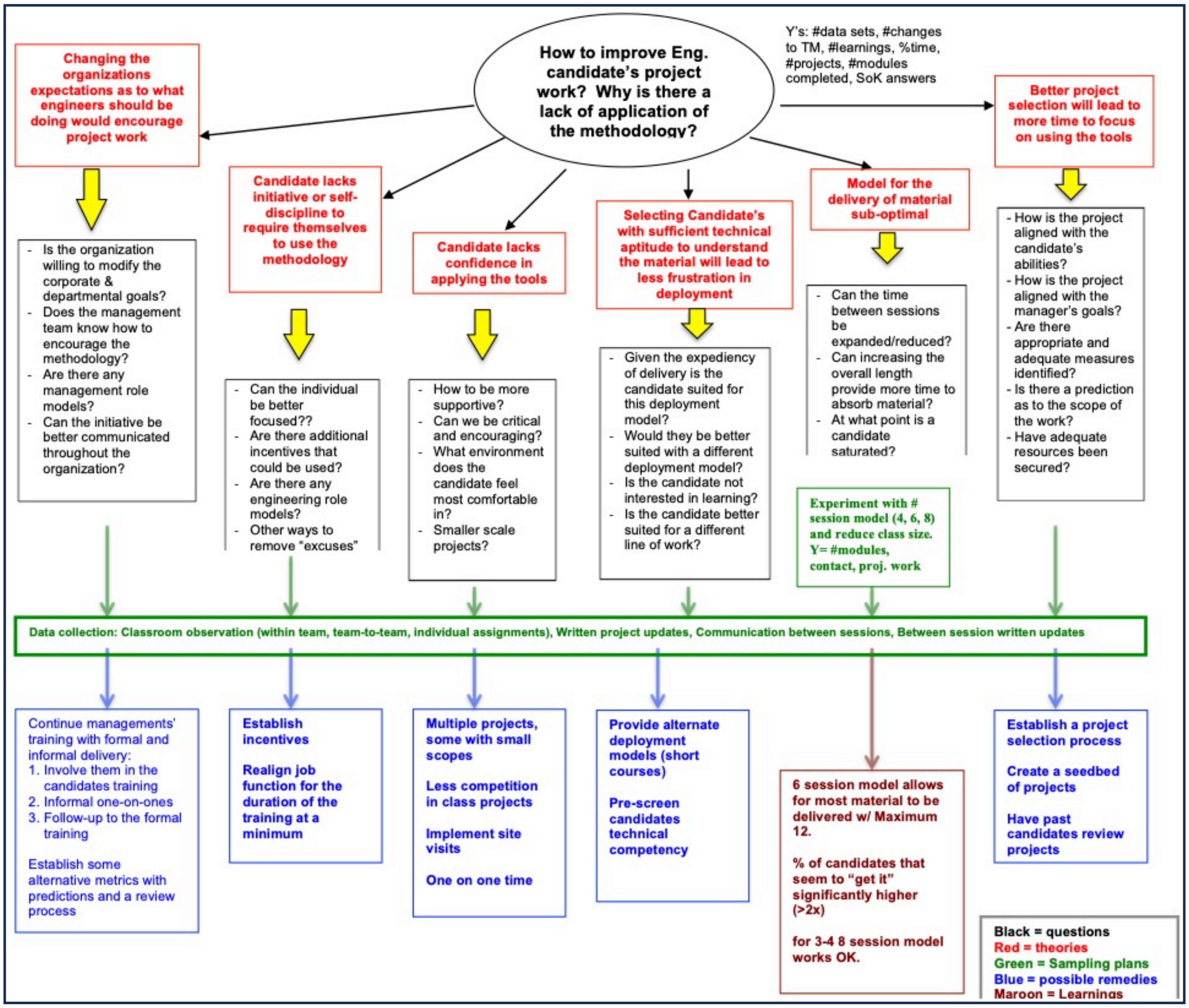


Figure 20.2: My Thought Map Regarding Project Success

Common Project Selection Errors

Several common mistakes appear repeatedly in improvement initiatives. One error is selecting projects based solely on projected financial savings. While financial impact is important, predicted savings are often speculative early in an investigation. When projects are chosen primarily to satisfy accounting targets, they tend to focus on short-term execution rather than discovery. Another mistake is choosing projects that are too large. Large projects often involve multiple subsystems, numerous stakeholders, and competing objectives. While they may be strategically important, they are difficult learning environments for individuals developing new analytical skills. A third mistake is selecting projects with predetermined solutions. When management already believes it knows the answer, investigation becomes a formality. Data collection becomes biased, and learning is suppressed. A fourth mistake is selecting projects that are symptoms of a more systemic issue. Finally, some projects are chosen simply because data already exist, so they are convenient. Historical data can be useful for generating hypotheses, but they rarely provide the context necessary to establish causal relationships. Project selection should be guided by learning potential as well as business importance. A manager, who had at one time taken my class as an engineer, offered a Thought map for project selection (Figure 20.3).

Developing Individuals Through Projects

Projects are also the primary mechanism for developing individuals. I have often been asked how to select candidates to participate in Sigma Science initiatives. My general response is simple: first select the projects, then select the individual in the organization you would bet your job on to work that project. Not the individual who happens to be available. Too often improvement programs begin by identifying available personnel and then assigning them projects. Effective efforts proceed in the opposite direction. The importance of the problem should determine who is assigned to investigate it. Important problems deserve the organization's best thinkers, not simply available resources. Of course, the challenge is the best thinkers are usually the busiest.

Formal instruction introduces concepts, but understanding emerges through application. Engineers and scientists must repeatedly apply the methodology to real problems in order to develop intuition and confidence. This is why deployment often begins with carefully selected projects. These projects provide opportunities for individuals to practice critical thinking, mapping, sampling, experimentation, and interpretation in an integrated way. Repetition matters. Each investigation refines judgment. Over time, individuals begin to recognize patterns in how problems are structured, how data should be collected, and how results should be interpreted. Eventually, this becomes habitual.

Balancing Short-Term Results and Long-Term Capability

Organizations frequently face tension between achieving short-term improvements and developing long-term capability. Projects focused entirely on immediate savings may deliver quick results but fail to build analytical competence. Conversely, projects that are purely investigative may produce valuable learning but limited immediate operational benefit. Effective deployment balances these objectives. Some projects will emphasize discovery. Others will emphasize implementation. Over time, the organization benefits from both improved understanding and tangible results. The key is intentional balance rather than accidental selection.

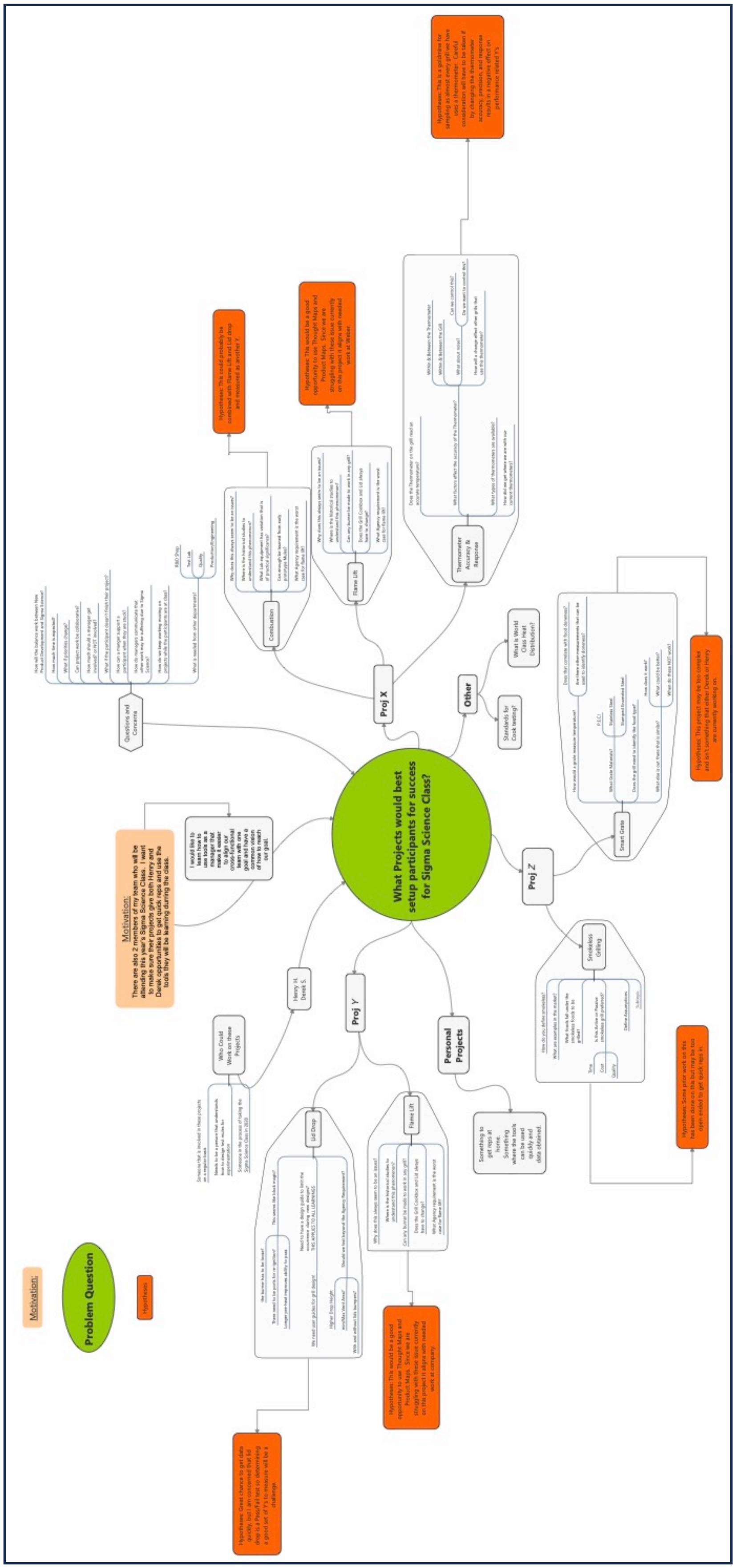


Figure 20.3: Thought Map for Project Selection (Courtesy Paul Hunt)

Creating a Seedbed of Projects

Successful deployment programs develop a “seedbed” of projects, a portfolio of investigations that collectively advance understanding of important processes. These projects should span different systems, products, or operational areas. As individuals complete projects, they develop experience that can be applied to future investigations. The seedbed approach has several advantages. It spreads learning across the organization. It provides multiple opportunities for success. It generates examples that others can learn from. It gradually builds institutional knowledge about causal relationships within the business. Over time, this accumulated knowledge becomes a strategic asset.

The Real Measure of Project Success

A project should not be judged solely by whether it produced a specific improvement. A more meaningful question is whether the project increased understanding of the system. Did the investigation clarify which factors matter and which do not? Did it reveal previously unknown interactions? Did it identify new opportunities for improvement? Did it improve the organization's ability to predict process behavior? And most importantly, did the engineer learn about how to do project work? When projects expand understanding of causal structure, they create lasting value. Results may follow immediately, or they may emerge in later iterations.

Project Selection as Strategic Leverage

Because projects drive both learning and credibility, project selection becomes the primary leverage point in deployment. Well-chosen projects create momentum. They demonstrate the usefulness of critical thinking. They develop capable practitioners. They generate knowledge that spreads through the organization. Poorly chosen projects do the opposite. They frustrate practitioners, produce ambiguous results, and reinforce skepticism about improvement initiatives. For this reason, project selection deserves careful attention at the outset of any deployment effort. The next few chapters examine how leadership, metrics, and organizational dynamics influence the ability of these projects, and the methodology itself, to take root and endure.

Chapter 21
Business Processes

Much of the early application of statistical thinking and scientific method occurred in manufacturing environments. This was natural. Manufacturing processes are visible, measurable, and often highly repetitive. Variation in output characteristics such as dimensions, strength, yield, or cycle time can be readily observed. Because of this visibility, engineers and scientists developed systematic approaches to understanding the causal structure of those processes. However, the same principles apply equally to business and administrative processes. In fact, many organizations contain far more inefficiency and variation in their business processes than in their manufacturing operations. Administrative and service processes often contain excessive complexity, unclear intent, and poorly defined measures of performance. Despite this, they are frequently managed through intuition, tradition, or organizational habit rather than systematic study.

In one large-scale business process deployment initiative, I met with the senior management staff. During the discussion, the Vice President of Marketing mentioned that he managed a marketing budget of nearly $1 billion. I asked how they measured whether their marketing campaigns were successful. He looked at me somewhat quizzically and explained they tracked how many campaigns were executed and whether they were completed on time. I suggested that neither of these measures provided insight into the effectiveness of the campaigns. After the first training class, the improvement projects in marketing produced more savings than the combined applications in manufacturing and product design. The experience illustrated how rarely business processes are examined with the same rigor applied to manufacturing operations.

The Sigma Science methodology applies the same scientific reasoning used in technical processes to the study of business processes. The objective is not simply to make processes faster or cheaper, but to understand how and why they perform as they do, and to improve them through knowledge of their causal structure. These processes are challenging as they often require working on the system rather than in the system. This requires involvement from individuals that have the authority to act on the system and often work cross-functionally in the organization.

The Nature of Business Processes

A business process is any sequence of activities intended to produce a result for a customer. The "product" may be physical, informational, or service-based. Examples include order processing, engineering change control, purchasing, hiring, regulatory reporting, product development, decision making, budgeting, or customer support. Unlike manufacturing processes, business processes are often poorly defined. Many evolve gradually over time through small changes made by individuals or departments responding to local needs. Because these changes are rarely evaluated systematically, the resulting process can become complex and inefficient. It is common to find that the original intent of a business process has been obscured by years of modifications. Steps may be added to solve specific problems or to satisfy managerial concerns, but the cumulative effect of these changes is rarely evaluated. As a result, the process becomes more complex, more costly, and often less effective at achieving its intended purpose. This pattern occurs frequently because business processes are relatively easy to change. Unlike manufacturing equipment, which requires engineering effort to modify, administrative procedures can be altered with little analysis. Over time, these incremental changes can transform a simple process into one that is slow, costly, and poorly aligned with its original purpose.

Inspection versus Process Understanding

Many organizations attempt to achieve quality in business processes through inspection and control rather than through process understanding. In manufacturing, inspection-based approaches are sometimes described as "part control." Quality is achieved by inspecting parts repeatedly and removing defects after they occur. A similar philosophy often exists in administrative systems, where quality is pursued through layers of documentation, approval, and review. This approach might be described as "audit, document, and approve." Processes are designed to ensure that forms are completed correctly, signatures are obtained, and procedures are followed. However, little attention is paid to understanding the causal factors that determine whether the process will produce the desired result. The result is often an expensive and complex system that monitors activity but does little to improve the process itself. Additional oversight and documentation may even increase the cycle time and cost of the process without improving its effectiveness. Sigma Science approaches business processes differently. The emphasis is not on inspection or compliance but on understanding the causal factors that determine process performance. I recall working for an organization and was involved in the budgeting process. I would study previous years expenditures and predict next years. I would submit the budget near the end of the year for next years budget. This would always be rejected and I would be asked to trim the budget. This would go on for several rounds until the budget was ultimately approved, often after the year I was budgeting for began. Then, the first time I spend money to purchase something in the budget, I was required to get 3-levels of management signatures. Improving this budgeting process would be like "polishing the turd".

The Role of Variation

As with manufacturing processes, business processes exhibit variation. Cycle times vary, information varies, resource utilization varies, and outcomes vary. However, this variation is often poorly understood. In many organizations, individual events that do not meet expectations are treated as unique problems requiring explanation and reaction. Managers may attempt to identify a "root cause" for each undesirable outcome, even when the variation is simply part of the underlying process behavior. This is analogous to assuming the event is a special cause when it is actually common cause. Scientific reasoning requires a different perspective. Instead of reacting to individual events, the objective is to understand the structure of variation in the process. This requires identifying the measurable outputs of the process and studying how they vary over time. Once the outputs are defined, the investigator can begin identifying the factors that influence those outputs and developing hypotheses about the causal structure of the process.
The same fundamental model used in technical processes applies here:

$$Y = f(x_1, x_2, \dots, x_n)$$

The output of the process (Y) depends on a set of causal factors (x's). The task of the investigator is to identify those factors, understand their influence, and determine how the process can be improved by managing them more effectively.

Types of Process Change

Organizations attempt to improve business processes in several ways. These approaches can be broadly categorized into three types: motivational approaches, incremental improvement, and redesign.

Motivational Approaches

Motivational approaches assume that processes would perform better if individuals simply worked harder or behaved differently. These approaches emphasize exhortation, training, incentives, or disciplinary measures. While motivation may be useful in certain circumstances, it rarely addresses the underlying causes of poor process performance. In many cases, individuals are already working diligently within a poorly designed system. Asking them to try harder may temporarily increase effort but does not improve the process itself.

Incremental Improvement

Incremental improvement focuses on understanding and reducing variation within the existing process. This approach is closely aligned with the scientific method. The investigator studies the process, identifies the sources of variation, and implements changes that improve consistency, reliability, or efficiency. These improvements are typically achieved through better understanding of the causal factors affecting the process. Incremental improvement is particularly effective when the existing process structure is fundamentally sound but contains sources of variation that reduce its performance.

Process Redesign

In some cases, the existing process has become so complex or misaligned with its purpose that incremental improvement is insufficient. Years of accumulated changes may have produced a process that no longer resembles its original design. Under these conditions, it may be more effective to redesign the process entirely. Process redesign involves rethinking the structure of the process to achieve its intended purpose more effectively. This may include eliminating unnecessary steps, simplifying decision paths, or reorganizing responsibilities. Even in redesign efforts, scientific reasoning remains essential. The redesigned process must be evaluated in terms of its ability to produce the desired outputs consistently and efficiently.

Levels of Process Work

Business processes exist at multiple levels within an organization. Some processes operate within a single function, such as billing or purchasing. Others span multiple functions, such as product development, customer support or order fulfillment. At the highest level, processes may represent entire business systems, such as the customer order system or the product lifecycle management system. The appropriate improvement strategy often depends on the level at which the process operates. Processes within a single function may be improved effectively through incremental methods. Cross-functional systems, on the other hand, often require broader analysis and coordination. At the highest levels of organizational activity, redesign may be necessary to achieve meaningful improvements. Regardless of the level, the objective remains the same: to understand the causal structure of the process and manage the factors that determine its performance.

Measurement in Business Processes

Effective process study requires measurement. Without measurable outputs, it is impossible to evaluate whether a process is performing as intended. Unfortunately, many business processes lack meaningful performance measures. The metrics are typically efficiency metrics rather than efficacy metrics. Metrics may focus on internal activity rather than on the outcomes that matter to customers. For example, a process may measure the number of reports generated, the number of approvals obtained, or the number of forms processed. These metrics describe activity but do not necessarily reflect whether the process is achieving its intended purpose. A more useful approach is to identify the critical outputs of the process and measure them directly. These outputs may include efficiency metrics such as cycle time, lead times or cost, but must also include efficacy metrics such as customer satisfaction, conversion rates, on-time delivery. Once these outputs are measured, the investigator can begin studying the factors that influence them.

Integrating Scientific Method into Business Processes

The Sigma Science methodology applies the same reasoning framework used in technical investigations to business processes. The steps include:

1. Clarifying the purpose and intended outputs of the process
2. Identifying measurable output characteristics
3. Mapping the process to identify potential causal factors
4. Developing hypotheses regarding the factors that influence the outputs
5. Acquiring appropriate data through sampling or experimentation
6. Analyzing and interpreting the data
7. Implementing changes based on the resulting knowledge

This approach emphasizes learning rather than compliance. The objective is not simply to enforce procedures but to understand how the process behaves and how it can be improved.

Why Engineers Often Avoid Business Processes

Despite the large opportunities for improvement in administrative and service processes, engineers often avoid studying these systems. Several factors contribute to this tendency. First, business processes are often poorly defined. In manufacturing environments, processes are typically documented, visible, and constrained by equipment. In contrast, administrative processes frequently evolve through informal practices and undocumented decisions. Because the process boundaries are unclear, engineers may perceive the system as too ambiguous to analyze. Second, measurement is frequently absent. Engineers are trained to study systems through observable data. When meaningful metrics do not exist, the process appears difficult to analyze scientifically. In many business processes, performance is assessed through subjective judgment rather than quantitative measurement, making systematic investigation more difficult. Third, business processes are often highly cross-functional. A single process may involve multiple departments, each with different objectives and management structures. Investigating such a system may require cooperation across organizational boundaries, which can be challenging in hierarchical organizations. Fourth, business processes are often closely tied to existing organizational roles and responsibilities. Studying the process may reveal inefficiencies or unnecessary steps that have developed over time. As a result, individuals may perceive the investigation as a challenge to established authority or practices. Finally, many engineers simply prefer technical systems. Physical processes, equipment, and product performance are familiar territory. Administrative processes, by contrast, may appear less tangible and therefore less appealing to analyze. Despite these challenges, the opportunities for improvement in business processes are often substantial. Administrative systems frequently contain large amounts of hidden variation, unnecessary complexity, and poorly understood causal relationships. Applying the same disciplined reasoning used in technical environments can reveal these opportunities and lead to significant improvements in efficiency, reliability, and cost. The Sigma Science methodology encourages engineers and scientists to extend their analytical thinking beyond traditional technical boundaries. By applying scientific reasoning to business processes, organizations can achieve improvements that are often impossible through intuition or managerial directive alone.

Example: A Document Preparation Process

Consider a document preparation process at an environmental engineering company. This centralized process was responsible for preparing all documents intended for external parties. These included project proposals, technical recommendations, and regulatory reports. Over a period of several months, a series of changes were made to the process in response to complaints about slow turnaround time. These changes were implemented individually and reactively rather than as part of a systematic study of the process. The sequence of events was roughly as follows:

1. Technical staff complained that document preparation was taking too long.
2. Relationships between technical staff and word processing personnel became strained as individuals attempted to influence document priority.
3. One word processing employee was relieved of certain support responsibilities in an attempt to increase output.
4. A manager was hired to supervise the document preparation department.
5. The manager assumed responsibility for scheduling document preparation.
6. Project managers were required to complete document request forms to enter the queue.

Each of these changes was intended to solve a perceived problem. However, none of the changes addressed the underlying structure of the process. When the process was eventually mapped and studied, it became clear that the majority of the total preparation time was not spent actually preparing documents. Instead, the majority of the time was **queue time**, documents waiting to be processed.
On average, the department had sufficient capacity to perform the work. However, during peak demand periods the queue length increased dramatically. Because there was no systematic scheduling mechanism, documents were often prioritized based on personal relationships, persuasion, or perceived urgency rather than objective criteria. As each complaint arose, management responded by adding controls, forms, and supervisory oversight. The result was a process that was more complex, more expensive, slower, and more frustrating for both the technical staff and the document preparation staff.

The underlying issue, variation in workload and queue management, was never studied. This example illustrates a common pattern in business processes. Organizations often respond to problems by adding steps, documentation, or supervision rather than by understanding the causal structure of the process. A Sigma Science approach would have begun with a study of the process itself:

- What is the intended purpose of the process?
- What measurable outputs will provide for understanding causality?
- Are there any intermediate measures?
- How will success be defined?
- How long do documents spend in each stage of the process?
- What factors influence queue length and workload variation?

By collecting data and studying variation in cycle time, the organization could have identified the true drivers of performance and implemented improvements that simplified rather than complicated the process.

Traditional vs Sigma Science Process Improvement

The difference between traditional approaches and the Sigma Science approach to business processes can be summarized as follows in Table 21.1.

Traditional Approach	Sigma Science Approach
Focus on auditing procedures	Study the process itself
Add approvals and documentation to prevent errors	Identify and manage causal factors
Emphasize compliance and reporting	Emphasize understanding and learning
Improvements driven by individual departments	Cross-functional investigation
React to individual problems	Study variation in process performance
Add steps to prevent recurrence	Simplify and streamline processes

Table 21.1: Comparison of Approaches

Traditional approaches often attempt to control behavior by increasing oversight. Sigma Science instead focuses on understanding how the process works and how it produces its outputs. When the causal structure of a process is understood, improvements can often be achieved by simplifying the system rather than adding complexity.

Conclusion

Business processes represent a significant opportunity for improvement in most organizations. These processes often contain substantial inefficiencies, unnecessary complexity, and poorly understood variation. Applying scientific reasoning to these processes allows organizations to move beyond reactive problem solving and toward systematic learning. By identifying measurable outputs, studying variation, and understanding the causal structure of processes, engineers and managers can make informed decisions about how to improve performance. The Sigma Science methodology provides a framework for this work. By combining process understanding, measurement, and statistical reasoning, organizations can transform business processes from opaque administrative systems into predictable, efficient, and continuously improving operations.

Transition to the Rest of Part III

The purpose of this chapter is to illustrate that the principles discussed throughout this book apply not only to technical processes but to all organizational systems. Whether the process produces manufactured parts, engineering designs, regulatory reports, or customer services, the same fundamental questions apply:

- What is the intended purpose?
- What are the measurable outputs?
- What factors influence those outputs?
- How does variation arise?
- How can that variation be understood and managed?

The Sigma Science methodology provides a framework for answering these questions. The remaining chapters in this section discuss how these principles can be applied in the broader organizational environment to guide implementation, and sustained improvement.

Chapter 22
Leadership and Management in a Sigma Science Environment

The success of a Sigma Science initiative depends as much on leadership behavior as it does on technical competence. While engineers and scientists conduct investigations, management determines the environment in which those investigations occur. Resource allocation, priorities, metrics, incentives, and expectations are all set by leadership. These factors strongly influence whether the methodology can take root in an organization. If leadership behavior conflicts with the principles of scientific learning, even the most capable technical staff will struggle to apply the methodology effectively. For this reason, management understanding is essential. Engineers also expressed a consistent message to management: the methodology requires time, iteration, and organizational support. Without those conditions, the tools are often abandoned before they can generate meaningful knowledge. Appendix H includes a collection of these reflections.

The Role of Leadership

Leadership involvement is often misunderstood in improvement initiatives. Some leaders believe their role is simply to authorize the program, allocate a budget, and review results periodically. In practice, leadership behavior must be more engaged than that. Leaders establish priorities and set expectations. They define what types of work are valued. They determine whether individuals have the time and support required to investigate problems properly. They also determine how success is measured and rewarded. If leadership expects immediate answers, discourages experimentation, or penalizes failed hypotheses, individuals quickly learn to avoid investigative work. Sigma Science requires an environment where disciplined inquiry is encouraged rather than suppressed.

Understanding Iteration

One of the most important concepts for leadership to understand is iteration. Scientific learning rarely occurs in a single step. Hypotheses evolve as new information becomes available. Sampling plans may need adjustment. Experiments may reveal unexpected interactions or sources of variation. Early investigations often produce partial answers rather than complete solutions. Organizations that expect immediate certainty often abandon investigations prematurely. When leaders expect linear progress, they misinterpret the natural learning process as inefficiency. Iteration should be viewed as progress rather than failure. Each cycle of investigation refines understanding and reduces uncertainty about the system. I always advise engineers to present management several options for data collection. Each option should explain the potential knowledge it may uncover and the resources required to execute it. There should also be discussion of what the next iteration might be, recognizing that this is a prediction, not a certainty. Finally, I suggest asking the manager a simple question: *Which of these plans should we implement?* This approach engages management in the investigation process while making clear that learning occurs through successive iterations rather than a single predetermined plan. Investigation rarely proceeds exactly as planned; progress comes from learning in stages and adjusting the next step accordingly.

“Most large organizations embrace the idea of invention, but are not willing to suffer the string of failed experiments necessary to get there.”

Jeff Bezos

Tolerance for Learning

Closely related to iteration is tolerance for learning. Investigations occasionally produce results that contradict expectations. Experiments may demonstrate that previously accepted explanations are incorrect. Sampling studies may reveal sources of variation that had been overlooked. These outcomes are valuable because they improve understanding of causal structure. However, if the organizational culture penalizes incorrect predictions or *failed* experiments, individuals become reluctant to explore alternative explanations. Learning slows dramatically. Leadership must reinforce the idea that disciplined investigation, even when it disproves a hypothesis, is productive work.

Resource Allocation

Another critical leadership responsibility is resource allocation. Structured investigations require time. Engineers must observe processes, develop hypotheses, design sampling plans or experiments, analyze data, and interpret results. These activities cannot be performed effectively when individuals are overloaded with routine responsibilities. Organizations sometimes attempt to conduct improvement initiatives without adjusting workloads. When this happens, investigative work is postponed or rushed. Both outcomes reduce effectiveness. Leadership must make explicit decisions about how investigative work will be supported. This may involve temporarily reassigning responsibilities, prioritizing projects carefully, or dedicating specific individuals to improvement activities. Without these adjustments, the initiative competes unsuccessfully with daily operational demands.

Asking Better Questions

Leadership also influences improvement through the questions they ask. When managers focus exclusively on results; “Did the project succeed?” or “How much money did it save?” they unintentionally encourage superficial work. Individuals may rush toward implementation before causal relationships are understood. More productive questions focus on learning. Examples include:

- What are the options for data collection?
- What hypotheses are to be evaluated?
- What are your predictions?
- What sources of variation are to be studied?
- What can we learn about the system?
- What might the next iteration of the study examine?

These questions reinforce the scientific nature of the methodology and signal that understanding is valued.

Avoiding the Target Trap

Another leadership challenge involves the use of numerical targets. Metrics are necessary for managing organizations. However, when numerical targets become the primary objective, individuals often adapt their behavior to meet the metric rather than to improve the system. At least two of Demings’s 14 points are relevant[1]. For example, if a specific sigma level becomes a performance target, individuals may manipulate definitions, redefine defects, or limit the scope of measurement in order to achieve the number. Such behaviors distort the system and obscure the true sources of variation. Leadership must ensure that metrics serve as indicators of process behavior rather than as targets detached from reality.

Enabling Cross-Functional Learning

Many sources of variation originate at the interfaces between functions, for example, between design, manufacturing, supply chain, and customer use conditions. Investigating these interactions often requires cooperation across organizational boundaries. Leadership plays an important role in enabling this cooperation. When departments operate in isolation, important causal relationships remain hidden. Cross-functional collaboration allows individuals to integrate knowledge from different domains. Leaders who encourage information sharing and joint problem solving significantly increase the effectiveness of improvement initiatives. I have worked for a company that was organized by product type. There was the tool department, the accessory department and the power source department. The product, in the hands of the customer, required all three components to function together. However, development was done in silos. The tool development group used old power source d and accessory designs to iterate their tools. The power source department used old accessories and tools and the accessory department used old tools and power source. You can clearly see the sub-optimization.

Modeling the Desired Behavior

Finally, leadership behavior influences culture through example. When managers demonstrate curiosity, ask thoughtful questions, and show interest in understanding processes rather than assigning blame, others tend to follow that example. Conversely, when leaders react to variation with frustration or seek quick explanations without investigation, individuals learn to conceal uncertainty. Culture evolves from repeated behavior. If leadership consistently supports disciplined inquiry, the organization gradually adopts critical thinking as a norm. When I was asked by a manager, “How do I encourage my folks to do critical thinking and use the methodology?”, I responded. Role model the use of the methodology yourself.

Leadership as a Catalyst

Technical capability alone does not create sustained improvement. Organizational systems, incentives, and behaviors must align with the principles of scientific learning. Leadership acts as the catalyst that enables this alignment. By supporting investigation, encouraging iteration, allocating resources thoughtfully, and asking the better questions, leaders create an environment in which Sigma Science can flourish. Without that environment, even the best methodology struggles to survive.

[1] 10-Eliminate slogans and targets, 11-Eliminate numerical quotas

Chapter 23
Communication, Incentives, and Organizational Reality

Even when leadership supports an improvement initiative and capable individuals are trained, success is not guaranteed. Organizations are complex social systems. Decisions are influenced by incentives, competing priorities, and interpersonal dynamics. In technical work we often refer to these influences as “politics,” sometimes with a negative connotation. In reality, these dynamics are simply characteristics of human organizations. They shape how information flows, how decisions are made, and how change is accepted. A Sigma Science initiative must operate within this environment. Understanding organizational dynamics is therefore just as important as understanding statistical methods.

Being Right Is Not Enough

Engineers and scientists often assume that if the analysis is correct, the conclusion will naturally be accepted. Unfortunately, organizational decisions rarely operate that way. Technical correctness is necessary, but it is not sufficient. Individuals must be able to communicate their reasoning clearly and relate their findings to the interests and concerns of decision-makers. Managers often evaluate recommendations through the lens of cost, schedule, risk, and strategic priorities. A technically sound proposal may fail if it does not address these broader considerations. Effective communication bridges this gap. Those presenting results must explain not only *what* was learned, but also *why it matters* and *what action is recommended*. The ability to translate analytical insight into operational relevance is a critical skill. It is also a challenging skill for a technically astute individual, who may tend to be introvert.

Tailoring the Message to the Audience

Different audiences require different levels of detail. Engineers may want to examine the data structure, the hypotheses, and the analytical methods used. Managers are often more interested in the implications for performance, risk, and resource allocation. A common mistake is presenting every audience with the same level of technical detail. When presentations become overly technical, non-specialists may disengage. When they are overly simplified, technical credibility can be undermined. Effective communicators adjust their message to the needs of the audience while preserving the integrity of the analysis. The goal is understanding, not simply presentation.

Aligning Incentives

Incentives strongly influence behavior. If organizational reward systems conflict with the objectives of structured investigation, individuals may unintentionally undermine improvement efforts. For example, when performance evaluations emphasize short-term output or strict schedule adherence, individuals may avoid activities that require experimentation or iterative learning. Similarly, if managers are rewarded solely for meeting quarterly targets, they may resist investigations that temporarily disrupt production. Incentive systems must therefore be examined carefully. Organizations seeking to develop analytical capability should ensure that individuals are rewarded for disciplined investigation, thoughtful experimentation, and collaborative problem solving. These behaviors contribute to long-term performance even if they do not always produce immediate results.

Rewarding Learning

Improvement initiatives often focus on financial results. While financial outcomes are important, they are not the only indicators of progress. Projects that reveal important causal relationships can create value even if they do not immediately reduce cost or increase revenue. By improving understanding of a process, such projects enable more effective decisions in the future. Organizations that recognize and reward learning encourage deeper investigation. Those that focus exclusively on short-term financial metrics risk discouraging the very activities that lead to sustainable improvement. Balanced evaluation criteria are therefore essential.

Managing Early Expectations

Early phases of deployment are particularly sensitive. Individuals are learning new ways of thinking and applying unfamiliar tools. Early projects may produce mixed results. If expectations are unrealistic, skepticism can develop quickly. Leaders should communicate the objective of early projects is learning as much as improvement. Initial investigations help individuals develop competence and build confidence in the methodology. As experience accumulates, the organization becomes more effective at selecting projects, designing studies, and interpreting results. Improvement outcomes typically become more consistent over time. Managing expectations during this learning phase is critical.

Building Credibility Through Demonstration

Nothing builds credibility faster than visible success. When structured investigations produce clear insights and measurable improvements, interest spreads naturally. Other individuals begin to ask how the results were achieved and whether the methodology can be applied to their own challenges. For this reason, early projects should be selected carefully. Projects that allow the methodology to demonstrate its strengths, particularly in understanding variation and identifying causal relationships, can accelerate organizational adoption. These early successes become reference points for future work.

Overcoming Organizational Resistance

Resistance to change is common in any organization and in any individual. Some individuals may be skeptical of new approaches, while others may simply be comfortable with existing methods. Resistance is often reduced through demonstration rather than persuasion. When colleagues observe the methodology producing useful insights, skepticism tends to decline. Practical results speak more convincingly than theoretical arguments. Patience is also important. Cultural change rarely occurs instantly. Individuals adopt new practices at different rates, and organizations evolve gradually. Persistence, clarity of purpose, and consistent application of the methodology help overcome resistance over time.

Communication as a Continuous Process

Communication in improvement initiatives should not be limited to formal presentations. Informal discussions, project reviews, and cross-functional meetings all provide opportunities to reinforce critical thinking. When individuals share their hypotheses, predictions, and interpretations openly, others gain insight into the reasoning process. These interactions spread knowledge and encourage collaborative learning. Over time, communication becomes part of the methodology itself. Ideas are examined collectively. Assumptions are questioned constructively. Learning becomes visible.

Integrating Structured Thinking into Organizational Life

The ultimate objective of deployment is integration or inculcation. Structured thinking should eventually become a natural part of how problems are approached within the organization. Engineers should routinely frame problems in terms of causal relationships. Managers should ask questions about hypotheses, data collection strategies, and interpretation. When this occurs, the methodology is no longer perceived as a special initiative. It becomes the normal way the organization works. The next chapter examines the role of metrics and measurement systems in sustaining improvement and ensuring that organizational behavior remains aligned with the principles of causal investigation.

Chapter 24
Metrics, Measurement, and Behavioral Consequences

Organizations rely heavily on metrics. Performance indicators guide decision making, monitor progress, and communicate priorities throughout the enterprise. When used appropriately, metrics provide valuable insight into how systems behave. However, metrics can also distort behavior. When numbers become targets rather than indicators of system behavior, individuals adapt their actions to satisfy the metric rather than to improve the system itself. This phenomenon is widely observed in organizations and can undermine improvement initiatives if not managed carefully. For this reason, metrics must be designed and interpreted with an understanding of variation and causal structure.

Metrics as Indicators of Process Behavior

Metrics should primarily serve as indicators of how processes behave over time. Most processes exhibit variation. Some of that variation is inherent to the system, what Deming referred to as common cause variation. Other variation arises from identifiable factors that temporarily or permanently change the system, special causes. Understanding this distinction is fundamental to interpreting performance metrics. When variation arises from common causes, reacting to individual fluctuations in the metric rarely improves the system. In fact, such reactions often increase instability by introducing unnecessary adjustments. When variation arises from special causes, investigation can reveal specific factors responsible for the change. Addressing these factors can restore stability or improve performance. Metrics therefore should be interpreted through the lens of variation, not simply as isolated numbers.

The Danger of Numerical Targets

A common organizational practice is to establish numerical targets for performance metrics. While targets may appear to provide clarity and motivation, they often create unintended consequences. When individuals are evaluated primarily on their ability to achieve a numerical target, attention shifts from understanding the process to manipulating the metric. Definitions may be altered, measurement methods adjusted, or activities redirected in ways that improve the reported number without improving the underlying system. This behavior does not necessarily reflect dishonesty. It reflects adaptation to incentives. If the objective becomes achieving the number, individuals will find ways to achieve the number. The result is often a system that appears to improve while the underlying causes of variation remain unchanged.

Metrics and Causal Understanding

Metrics are most valuable when they contribute to understanding causal relationships. For example, a time series of a process characteristic can reveal patterns of variation, shifts in process behavior, or relationships with external factors. Control charts, variability charts, and other graphical methods allow engineers to examine these patterns and generate hypotheses about possible causes. Metrics therefore should be integrated into the investigative process rather than treated as isolated performance indicators. The purpose of measurement is learning. When measurement systems are designed primarily for reporting, their ability to support learning diminishes.

The Importance of Measurement Systems

Another frequently overlooked aspect of metrics is the quality of the measurement system itself. All measurements contain variation. Instruments have limited discrimination, operators may introduce differences in technique, and environmental conditions can influence readings. If measurement variation is large relative to the variation of interest, interpretation becomes difficult or impossible. Before drawing conclusions from data, investigators must consider whether the measurement system is adequate for the intended purpose. Measurement systems analysis helps determine whether measurements are sufficiently precise and consistent to support meaningful investigation. When measurement error dominates the observed variation, improving the measurement system may be the first step in understanding the process.

Aligning Metrics with Organizational Objectives

Metrics should also align with the objectives of the organization. For example, if the goal is to improve product reliability, metrics should reflect reliability characteristics rather than unrelated production indicators. If the objective is to reduce variation, metrics should capture variability rather than simply average performance. Misaligned metrics can lead individuals to optimize aspects of the system that do not contribute to overall objectives. Thoughtful metric selection ensures that measurement supports both operational goals and causal investigation.

Metrics as Feedback for Learning

When metrics are interpreted correctly, they provide valuable feedback for learning. Patterns observed in performance data can suggest new hypotheses about process behavior. Unexpected variation may reveal previously unidentified factors. Changes in metrics following an experiment can confirm or refute predictions about causal relationships. In this way, metrics become part of the iterative cycle of scientific investigation. Observation leads to hypotheses. Hypotheses lead to experiments or sampling plans. Data provide feedback. Understanding improves. This cycle is central to Sigma Science.

Balancing Simplicity and Insight

Metrics must also balance simplicity and insight. Organizations often attempt to summarize complex processes using a single number. While such simplification may be convenient for reporting, it can obscure important information about variation and causal structure. Graphical representations of data frequently provide greater insight than summary statistics alone. Time series plots, variability charts, and other visual methods reveal patterns that may be hidden in aggregated metrics. For this reason, analysts should look at the data, not just the numbers derived from it. As has been humorously observed (Yogi Berra), one can observe a great deal simply by watching.

Sustaining Improvement Through Measurement

Sustaining improvement requires continuous observation. Once a process has been improved, measurement allows organizations to monitor whether the improvement persists. Control charts and related tools help detect shifts in process behavior so that corrective action can be taken when necessary. In this way, measurement supports both improvement and stability. However, measurement must remain connected to understanding. When metrics are interpreted without considering variation, causal relationships, and measurement quality, they become misleading. Metrics are powerful tools. Used wisely, they illuminate system behavior and support disciplined investigation. Used carelessly, they obscure reality and distort decision making. The next chapter of this book reflects on how individuals and organizations can sustain the mindset of causal inquiry that underlies Sigma Science.

Chapter 25
The Engineer in the System

Throughout this book the focus has been on methodology, how to think about problems, how to acquire data, how to interpret analysis outputs, and how to learn about causal structure. These ideas are presented as if investigations occur in a clean, rational environment where individuals can proceed methodically from question to hypothesis to experiment. Real organizations rarely operate that way. Engineers and scientists work within systems shaped by schedules, budgets, personalities, and competing priorities. Decisions are influenced by incentives, past experience, and organizational culture. In this environment, the challenge is not only to understand causality, but also to practice causal investigation within the realities of the organization. The objective of this chapter is to address that practical challenge.

Over the years I have asked engineers who completed Sigma Science training to reflect on their experiences and provide advice to future participants. Their responses were remarkably consistent. Nearly all emphasized the importance of repetition, iteration, and the development of a personal user guide. Several also emphasized that the methodology requires patience and sustained effort before its benefits become clear. A selection of these reflections is provided in Appendix H.

Working Within the System

Most engineers do not control the environment in which they work. Projects are assigned, deadlines are imposed, and priorities shift. In many situations the engineer's role is reactive, responding to problems that have already surfaced. Sigma Science does not change that reality. What it does provide is a way to approach problems more thoughtfully once they appear. Even when the problem is assigned by someone else, the engineer still controls how the situation is framed, how hypotheses are generated, how data are acquired, and how conclusions are interpreted. Critical thinking begins with the individual. The first step is to resist the impulse to move immediately toward solutions. Instead, the engineer should ask the questions introduced earlier in this book: What is the response variable? What factors might influence it? What hypotheses explain the observed behavior? What outcomes would we predict under different conditions? These questions transform reactive work into investigative work.

Starting With the Problems You Are Given

Most improvement opportunities arise from practical concerns: a product fails in the field, a process produces excessive variation, or a system does not perform as expected. These situations rarely arrive with a clear description of the causal structure. Often the explanation provided initially is only a guess. The engineer's task is not simply to confirm the explanation, but to investigate it. This investigation begins with observation and questioning. What is actually happening? Under what conditions does the behavior occur? When did the issue first appear? What has changed? From these observations, hypotheses begin to emerge. Even simple hypotheses can provide structure to the investigation. Once hypotheses exist, predictions can be made and data can be collected deliberately rather than haphazardly. The methodology described in this book provides tools for structuring that investigation, but the process always begins with the individual asking thoughtful questions.

Externalizing Thought

One of the most effective habits an engineer can develop is externalizing thought. Critical thinking is largely internal. Without deliberate effort it remains invisible to others and difficult to refine. Tools such as thought maps and process maps allow engineers to make their reasoning visible. When hypotheses are written explicitly and relationships between variables are drawn visually, gaps in understanding often become apparent. Missing factors can be identified. Assumptions can be examined. Equally important, externalizing thought makes it easier to communicate with others. A well-constructed thought map allows colleagues and managers to see how the investigation is structured. It shifts discussions from opinions toward causal reasoning.

Asking Better Questions

Engineers often influence organizations not by authority, but by the questions they ask. Thoughtful questions can redirect conversations toward understanding rather than speculation. Instead of asking: *"What should we do to fix this?"* the engineer might ask: *"What do we think is causing this behavior?" "What would we expect to see if that explanation were correct?" "How could we collect data to test that idea?"* These questions encourage others to participate in the investigative process. Over time, colleagues begin to adopt similar patterns of reasoning. In this way, disciplined reasoning spreads gradually through conversation.

Building Credibility Through Demonstration

Credibility is rarely established through argument alone. It is established through results. When structured investigations produce useful insights—when they reveal previously unrecognized factors or clarify confusing patterns of variation—others begin to notice. Colleagues become curious about the approach that produced those results. For this reason, early investigations should aim to demonstrate the strengths of the methodology. Projects that clarify variation, identify key factors, or resolve persistent technical questions can build confidence in the methodology. Each successful investigation becomes a reference point for future work.

Navigating Organizational Constraints

Every organization imposes constraints. Engineers must work within resource limitations, reporting structures, and existing processes. Not every investigation can be conducted perfectly. The objective is not perfection but progress. Sometimes the available data are incomplete. Sometimes measurements are imperfect. Sometimes time is limited. Under these conditions the engineer must balance rigor with practicality. Even when the investigation cannot proceed ideally, hypotheses can still be articulated, assumptions can still be documented, and predictions can still be made. Maintaining this discipline preserves the integrity of the learning process.

Winning the Right to Investigate

In many organizations the opportunity to conduct structured investigations must be earned. Managers may initially be skeptical of unfamiliar methods. Colleagues may prefer established routines. Resources may be allocated cautiously. One effective strategy is to begin with small, manageable investigations.

When these studies produce useful insights, they demonstrate the value of deliberate data acquisition and interpretation. Over time, successful investigations create momentum. As confidence grows, larger or more complex studies become possible. In this way, the engineer gradually earns the right to investigate more broadly.

Developing Personal Capability

Applying Sigma Science is ultimately a matter of practice. Each investigation teaches something new about both the system being studied and the process of investigation itself. Engineers begin to recognize patterns: how to frame problems, how to identify variables, how to structure sampling plans, and how to interpret outputs. Judgment develops through repetition. Formal education provides the foundation, but real understanding emerges through experience. The more investigations an engineer conducts, the more natural structured thinking becomes. Eventually, the process becomes intuitive.

Becoming a Catalyst

An engineer who consistently applies critical thinking often becomes a catalyst within the organization. Colleagues seek their perspective when problems arise. Managers begin to involve them in complex investigations. Their approach to problem solving influences how others think about variation and causality. This influence does not require formal authority. It emerges from credibility, curiosity, and disciplined reasoning. Over time, individuals who practice Sigma Science help shape the culture of the organization. Structured thinking spreads not through mandates, but through demonstration.

The Responsibility of the Practitioner

The methodology described in this book provides powerful tools for understanding how and why systems behave as they do. But these tools are only effective when applied thoughtfully. Each engineer or scientist who adopts this approach carries a responsibility: to ask better questions, to investigate problems carefully, and to interpret data with intellectual honesty. When practiced consistently, scientific inquiry not only improves processes and products, it improves the way organizations learn. And ultimately, that is the true objective of Sigma Science.

A statistician boards a 747 in Los Angeles headed to New York. 30 minutes into the flight there is a loud boom. The pilot comes on and says bad news, good news…bad news is we lost an engine, good news is this plane was designed to fly with 3 engines it will just take 20 minutes longer. After another 30 minutes, another loud boom. Bad news, good news…we lost our 2nd engine, but this plane can fly just fine on 2 engines, it's just going to take 45 minutes longer. 30 minutes later another loud boom. Bad news, good news…we lost our 3rd engine, but don't fret, this plane can fly with one engine, it's just going to take 90 minutes longer. The statistician looks over at the passenger in the seat next to him and says, "I sure hope we don't lose that 4th engine or we'll be up here all day."

Chapter 26
Sustaining the Methodology

Launching an improvement initiative includes its own challenges. Sustaining it is far more difficult. Many organizations introduce improvement programs with enthusiasm. Training sessions are conducted, projects are launched, and early successes generate excitement. Over time, however, attention shifts to other priorities. Personnel change roles, leadership changes direction, and the initiative gradually loses momentum. This pattern is common not because the methodology is flawed, but because sustaining change requires deliberate effort. Sigma Science, like any disciplined approach to learning, must become integrated into the normal operation of the organization if it is to endure.

From Initiative to Capability

The objective of deployment is not simply to complete projects. The objective is to build capability. Projects are valuable because they generate learning and demonstrate the usefulness of the methodology. But the long-term benefit emerges only when individuals throughout the organization develop the ability to investigate problems systematically. Capability grows through repetition. Engineers and scientists who repeatedly frame problems, generate hypotheses, acquire data deliberately, and interpret results thoughtfully develop judgment. Over time, structured thinking becomes a natural part of how they approach problems. When this occurs across many individuals, the organization develops a sustainable analytical capability.

Developing Internal Catalysts

Sustained improvement depends on people. Early in deployment, organizations often rely on external experts or a small number of trained practitioners. While this can help initiate the process, long-term success requires internal catalysts, individuals who understand the methodology deeply and can guide others in its application. These individuals play several important roles. They mentor colleagues, help structure investigations, assist with interpretation of results, and reinforce disciplined thinking. They also help maintain continuity as personnel and priorities change. Developing these catalysts takes time. Experience gained through multiple investigations is essential. For this reason, organizations should invest in the continued development of individuals who demonstrate curiosity, technical competence, and a commitment to structured reasoning.

Maintaining Organizational Alignment

Sustaining the methodology also requires alignment between the improvement effort and the broader objectives of the organization. Metrics, incentives, and resource allocation must support investigative work rather than discourage it. If individuals are evaluated solely on short-term output or schedule adherence, they may avoid activities that involve experimentation or iterative learning. Leadership must therefore ensure that the organizational environment supports disciplined inquiry. This includes providing time for investigation, encouraging cross-functional collaboration, and recognizing the value of projects that increase understanding even when immediate financial results are modest.

Protecting the Learning Process

One of the most important leadership responsibilities is protecting the learning process. Scientific investigation occasionally produces results that challenge existing assumptions. Experiments may reveal unexpected interactions or sources of variation. In some cases, previously accepted explanations may prove incorrect. These outcomes are essential to progress. Organizations that punish failed hypotheses or discourage exploration limit their own ability to learn. Sustained improvement requires an environment in which disciplined investigation is respected, even when the results are surprising.

Avoiding Initiative Fatigue

Another threat to sustainability is initiative fatigue. Organizations frequently introduce new improvement programs with different terminology and branding. When initiatives change frequently, employees may begin to view them as temporary management fashions rather than enduring approaches to learning. Consistency of purpose helps prevent this problem. When leaders reinforce the underlying principles, scientific inquiry, causal investigation, and iterative learning, the methodology becomes less dependent on specific program names or labels. The focus shifts from the initiative itself to the way the organization approaches problems.

Building Institutional Knowledge

Each investigation contributes to the organization's understanding of its processes and products. When this knowledge is documented and shared, it becomes a valuable resource. Engineers can build upon previous work rather than repeating the same investigations. Patterns of variation become familiar. Known causal relationships guide future design and process decisions. Over time, this accumulation of knowledge becomes one of the organization's greatest assets. Sustaining the methodology therefore involves not only developing individuals but also preserving and sharing what has been learned.

Continuous Renewal

Even when an organization successfully integrates structured thinking into its culture, renewal is necessary. New employees must be educated. Experienced practitioners must continue refining their skills. Processes evolve, technologies change, and new questions emerge. Continuous learning ensures that the methodology remains relevant and effective. Sigma Science is not a static body of knowledge. It is a framework for ongoing investigation and improvement.

The Long-Term View

Sustained improvement rarely occurs through dramatic breakthroughs alone. More often it results from a steady accumulation of insight gained through disciplined investigation. Each study clarifies some aspect of causal structure. Each iteration refines understanding. Over time these incremental advances compound, producing substantial improvements in performance, reliability, and efficiency. Organizations that adopt this long-term perspective recognize that learning itself is a strategic capability. The value of Sigma Science lies not only in the improvements it produces today, but in the organization's growing ability to understand and improve its systems tomorrow.

Chapter 27
Closing Reflections: Causality, Learning, and the Practice of Sigma Science

The central objective of this book has been to describe a way of thinking. Sigma Science is not simply a collection of statistical techniques. It is not a checklist, a certification program, or a set of templates. At its core, Sigma Science is the deliberate merging of scientific method and statistical thinking in pursuit of understanding causality. Throughout this book the emphasis has been on learning how and why systems behave as they do. The tools described, critical thinking, mapping, directed sampling, and designed experimentation, are mechanisms for organizing that learning. They help investigators structure their thinking, collect meaningful data, interpret analysis outputs, and refine their understanding through iteration. But tools alone do not produce insight. Insight emerges from disciplined reasoning applied repeatedly over time.

Understanding Rather Than Calculation

Modern statistical software is capable of performing extremely sophisticated calculations. Complex models can be generated quickly, and large volumes of data can be processed almost instantly.
These capabilities are valuable, but they do not guarantee understanding. A model that predicts accurately but cannot be interpreted may be useful in limited circumstances, but it provides little insight into the causal structure of the system. Without understanding why a relationship exists, it is difficult to generalize the results, adapt to new conditions, or recognize when the model no longer applies.
For this reason, Sigma Science emphasizes models that contribute to understanding. A simpler model that explains the mechanism behind a phenomenon may be more valuable than a complex model that produces accurate predictions but offers no insight into the underlying causes. Understanding enables prediction. Prediction without understanding is fragile.

Learning Through Iteration

Scientific learning is inherently iterative. Initial hypotheses are rarely complete. Early investigations often reveal unexpected sources of variation or interactions between variables that were not initially considered. As data are collected and interpreted, hypotheses evolve. Each iteration of this cycle; hypothesis, prediction, data acquisition, analysis, interpretation, refines understanding. This process rarely proceeds in a straight line. Investigations may reveal dead ends, incorrect assumptions, or measurement problems that require reconsideration of the study design. These outcomes are not failures. They are part of the learning process. Organizations and individuals who embrace iteration improve their understanding steadily over time.

The Role of Judgment

While statistical tools provide structure, judgment remains essential. Investigators must decide which variables to examine, how to structure sampling plans or experiments, how to interpret unexpected results, and when to pursue further investigation. These decisions cannot be automated. Judgment develops through experience. Each investigation adds to the investigator’s understanding of both the system being studied and the investigative process itself. Over time, patterns emerge and intuition improves. The goal of Sigma Science is not to replace judgment with procedure, but to support judgment with disciplined reasoning and meaningful data.

The Responsibility of the Investigator

Practicing Sigma Science requires intellectual honesty. Investigators must be willing to question their own assumptions, test alternative explanations, and acknowledge uncertainty when the evidence is incomplete. Data should be collected deliberately rather than selectively. Interpretations should be guided by both statistical reasoning and scientific knowledge. Perhaps most importantly, investigators must remain curious. Curiosity drives the questions that initiate investigation. It encourages exploration of unexpected results and motivates the search for deeper understanding of causal mechanisms. Without curiosity, analysis becomes mechanical.

Learning as an Organizational Capability

While this book has addressed the role of individuals, the broader objective is organizational learning. Organizations that cultivate disciplined reasoning develop a deeper understanding of their processes, products, and systems. They become better at predicting outcomes, identifying opportunities for improvement, and adapting to changing conditions. This capability cannot be achieved through tools alone. It requires leadership that supports investigation, metrics that encourage learning rather than manipulation, and practitioners who approach problems with disciplined curiosity. When these elements align, organizations move beyond reactive problem solving toward sustained improvement.

The Continuing Journey

No methodology can eliminate uncertainty completely. Processes evolve. Technologies change. New sources of variation emerge. As systems become more complex, the need for disciplined investigation becomes even more important. Sigma Science does not promise final answers. Instead, it provides a framework for continually refining understanding. The journey toward understanding causality is never complete. But each thoughtful investigation brings us closer.

Final Thoughts

Over the course of a career, engineers and scientists encounter countless problems, anomalies, and opportunities for improvement. Each of these situations offers a chance to learn something new about how the world works. The methodology presented in this book is intended to support that learning.
By merging scientific method with statistical thinking, Sigma Science provides a structured way to investigate problems, interpret data, and refine understanding through iteration. Ultimately, the value of this approach lies not in the tools themselves, but in the mindset they encourage. A mindset focused on asking better questions, on understanding variation, on discovering why systems behave as they do. In the end, Sigma Science is not about statistics. It is about structured thinking in pursuit of causality.

Appendix A
Iterative Thought Map Example
and Additional Thought Maps

A sample of different Thought maps to illustrate the different applications. Thought maps are a function of independent thinking and therefore may take on many structures.

The following set of Figures A-1- A-4 is an example of a sequence of Thought maps. This example is not intended as a template. It illustrates how understanding evolves through repeated cycles of questioning, data collection, and reflection. The structure, content and organization are specific to one investigator and one situation.

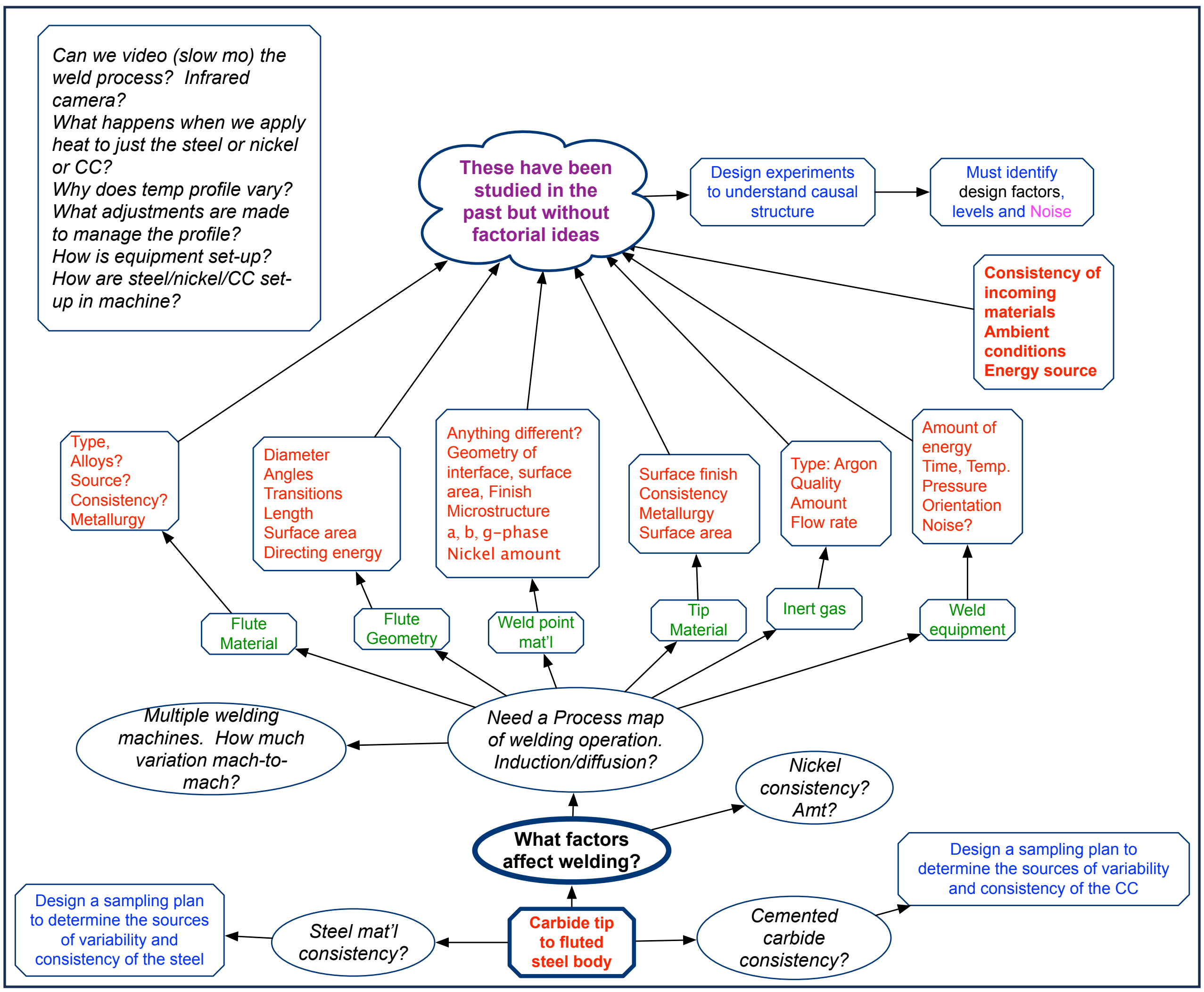


Figure A-1: Initial Thought Map for a Welding Project

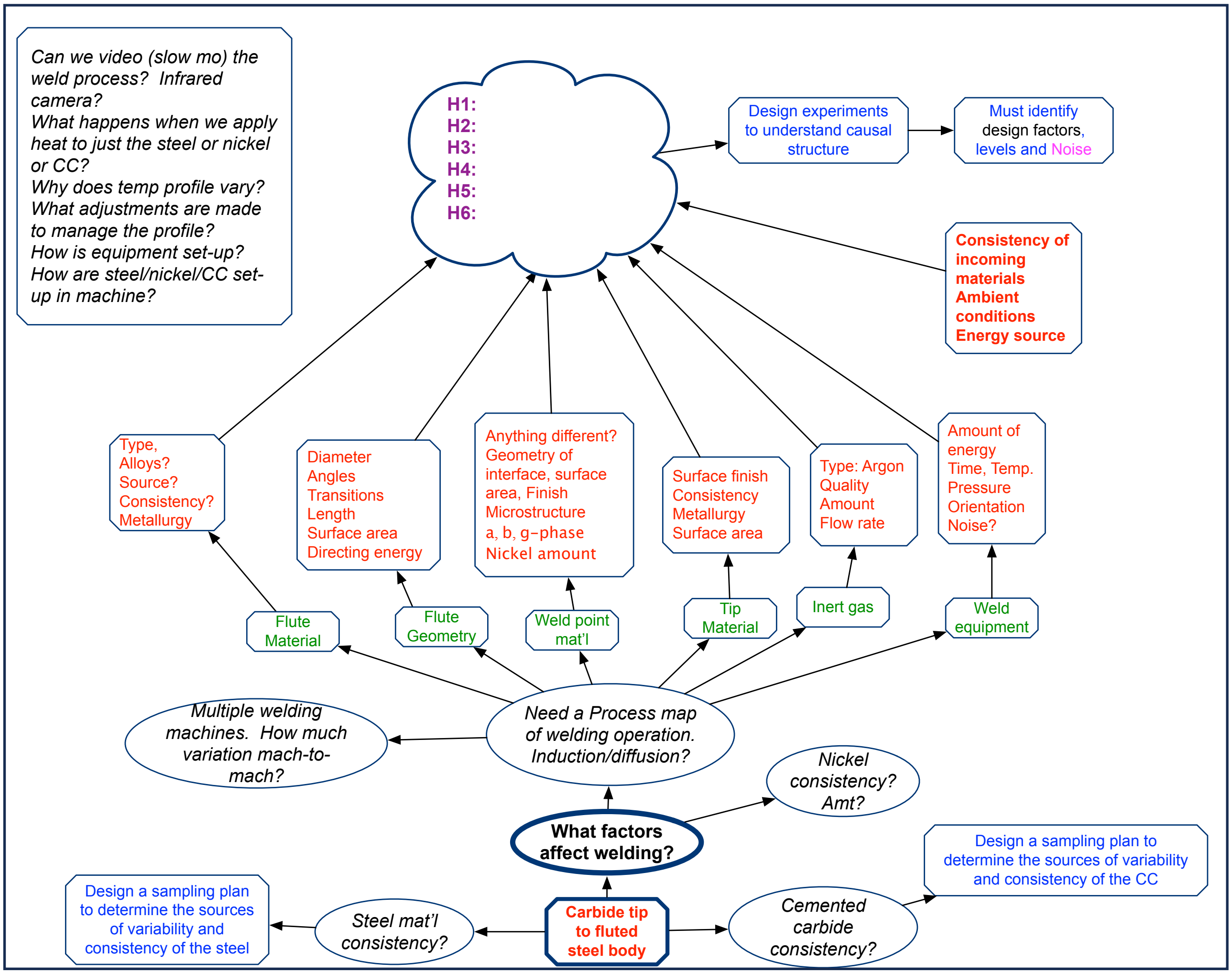


Figure A-2: A Later Iteration of the Welding Thought Map

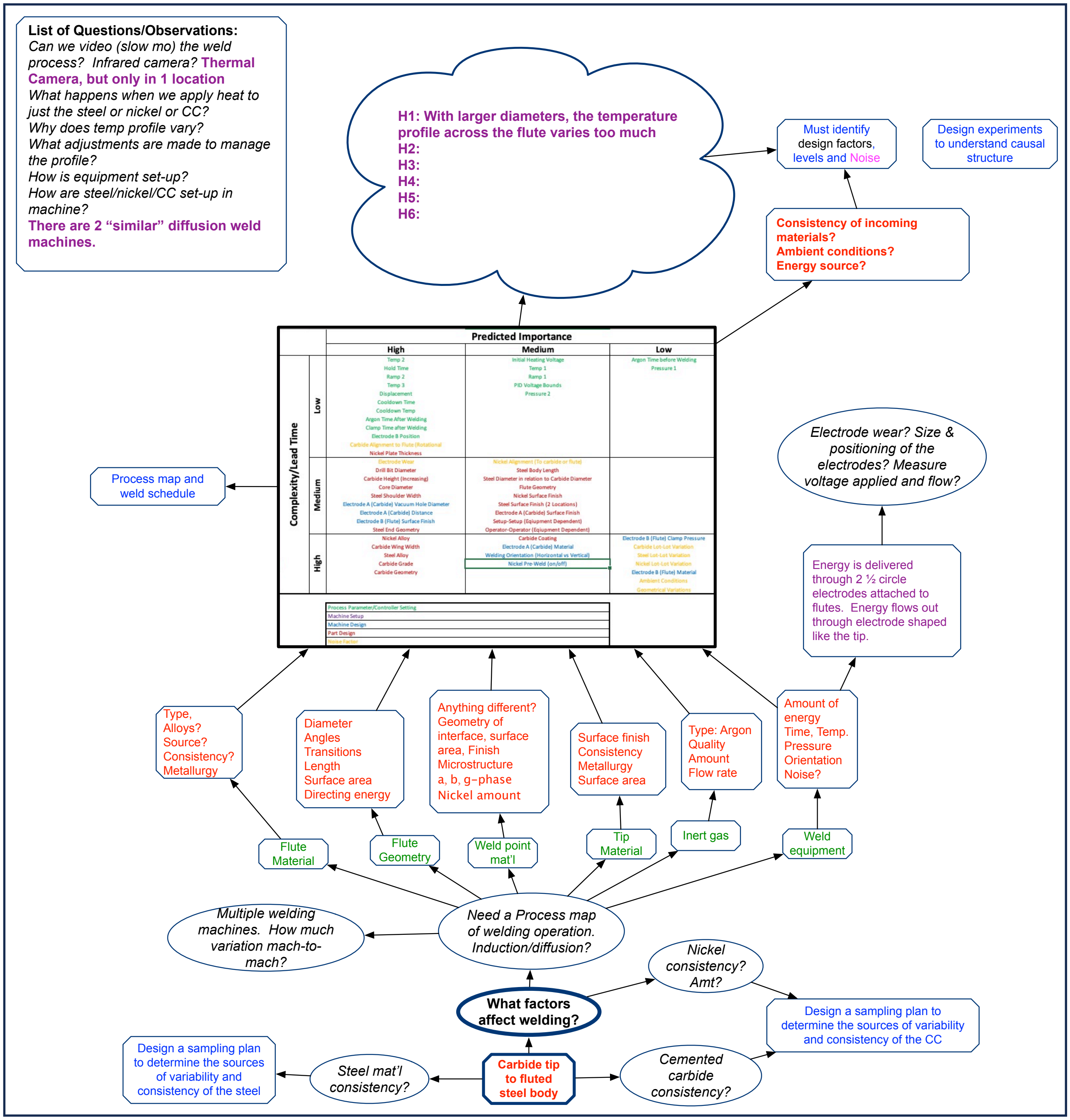


Figure A-3: Continuing Iteration of Welding Thought Map

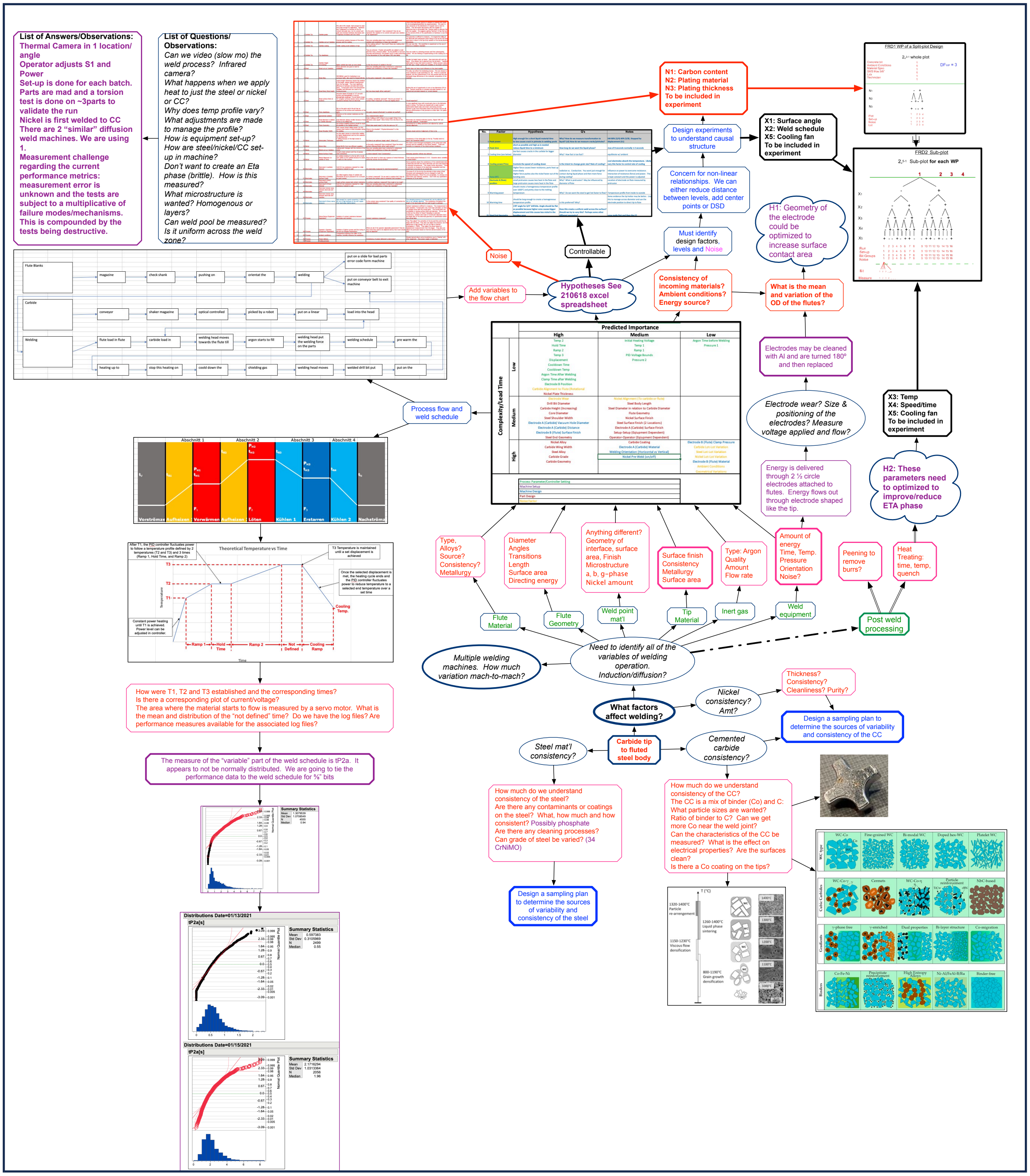


Figure A-4: A Maturing Welding Project Thought Map

Some additional examples to show the diversity of applications and structures. Figure A-5 is a Thought map I started to consider appropriate models for delivering material in a non-manufacturing environment.

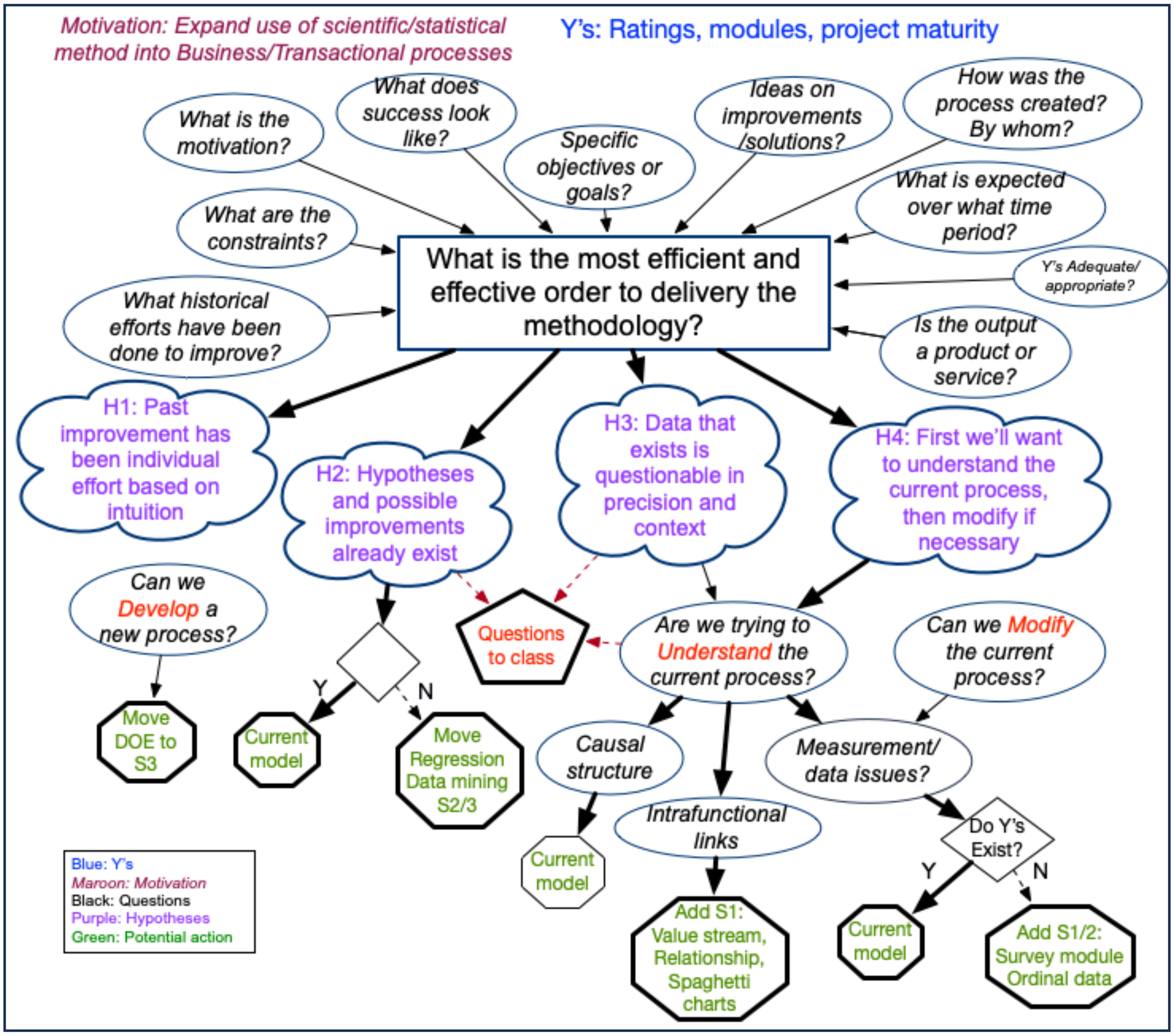


Figure A-5: Thought Map Considering Material Delivery Sequence

Figure A-6 is a Thought map motivated by questioning the metrics used to measure quality and determine customer satisfaction in the appliance industry. Note, SCR is the service call rate.

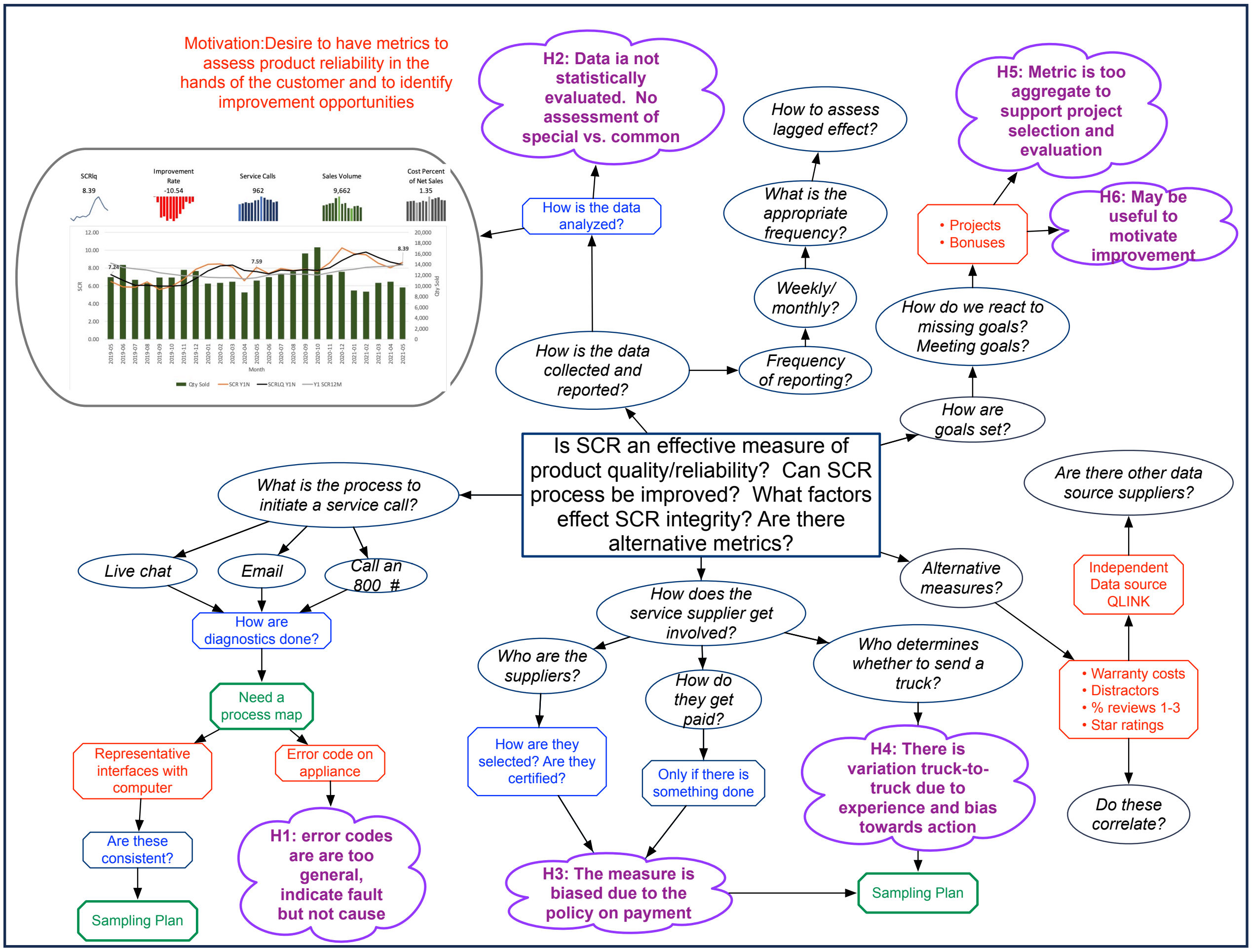


Figure A-6: Thought Map on SCR Metric

When the Covid pandemic hit, I thought about whether it would be possible to deliver the class remotely.

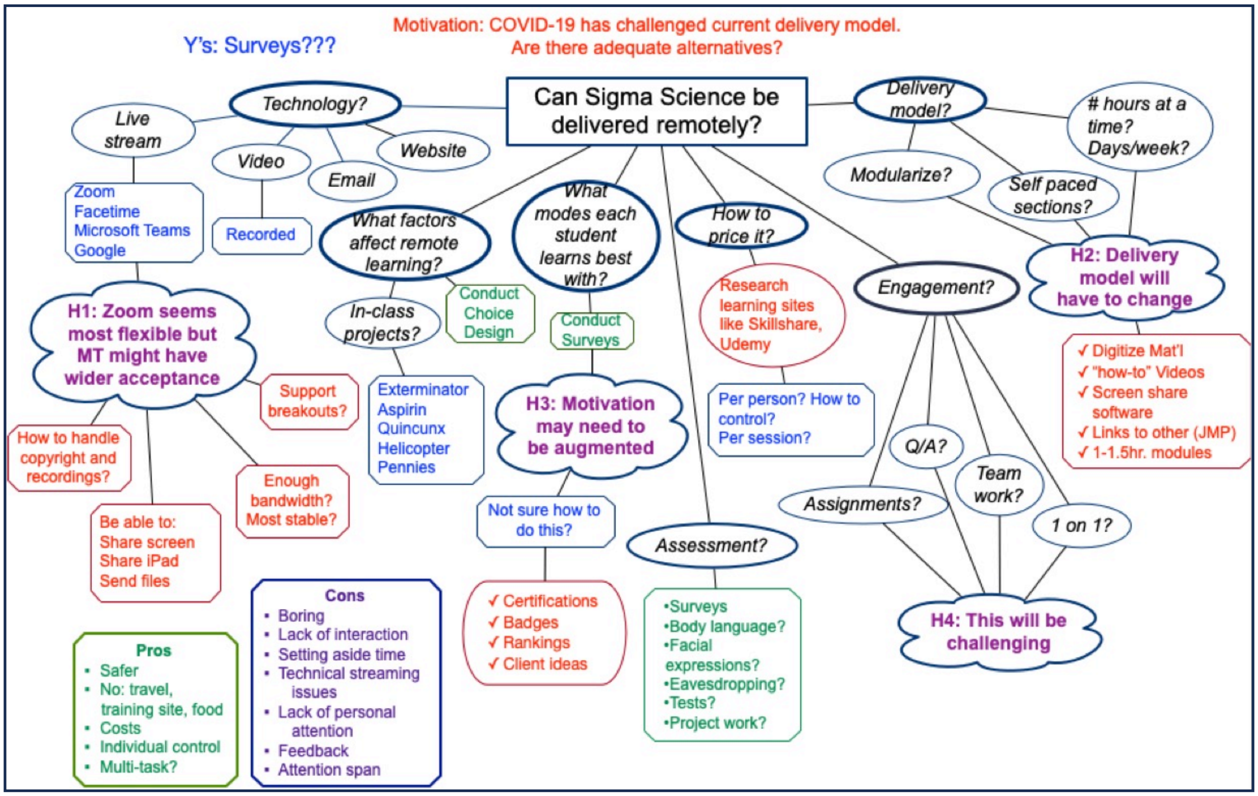


Figure A-7: Thought Map Regarding Remote Training

To develop my own capabilities, I practiced critical thinking on a variety of personal topics.

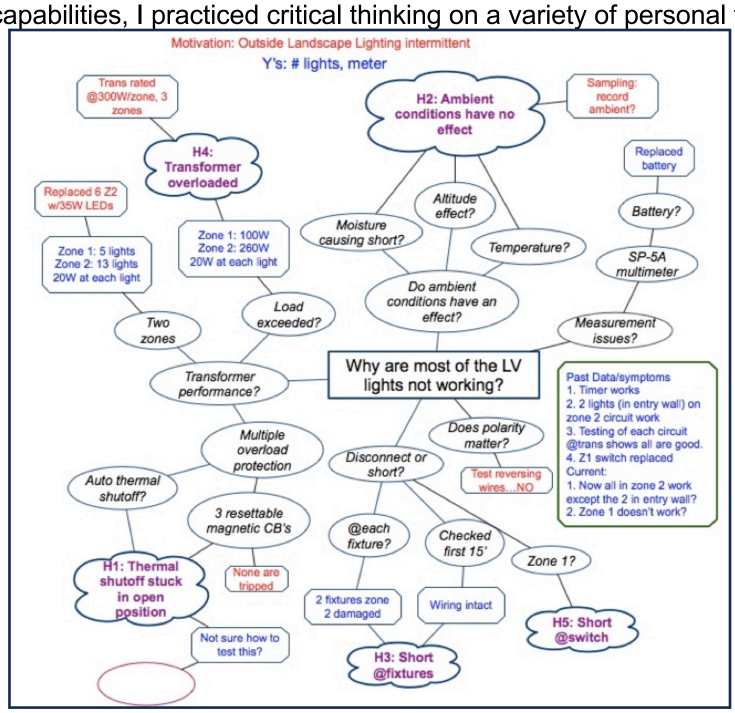


Figure A-8: Thought Map for Landscape Lighting

Figure A-9 is a Thought map comparing different models of automobile.

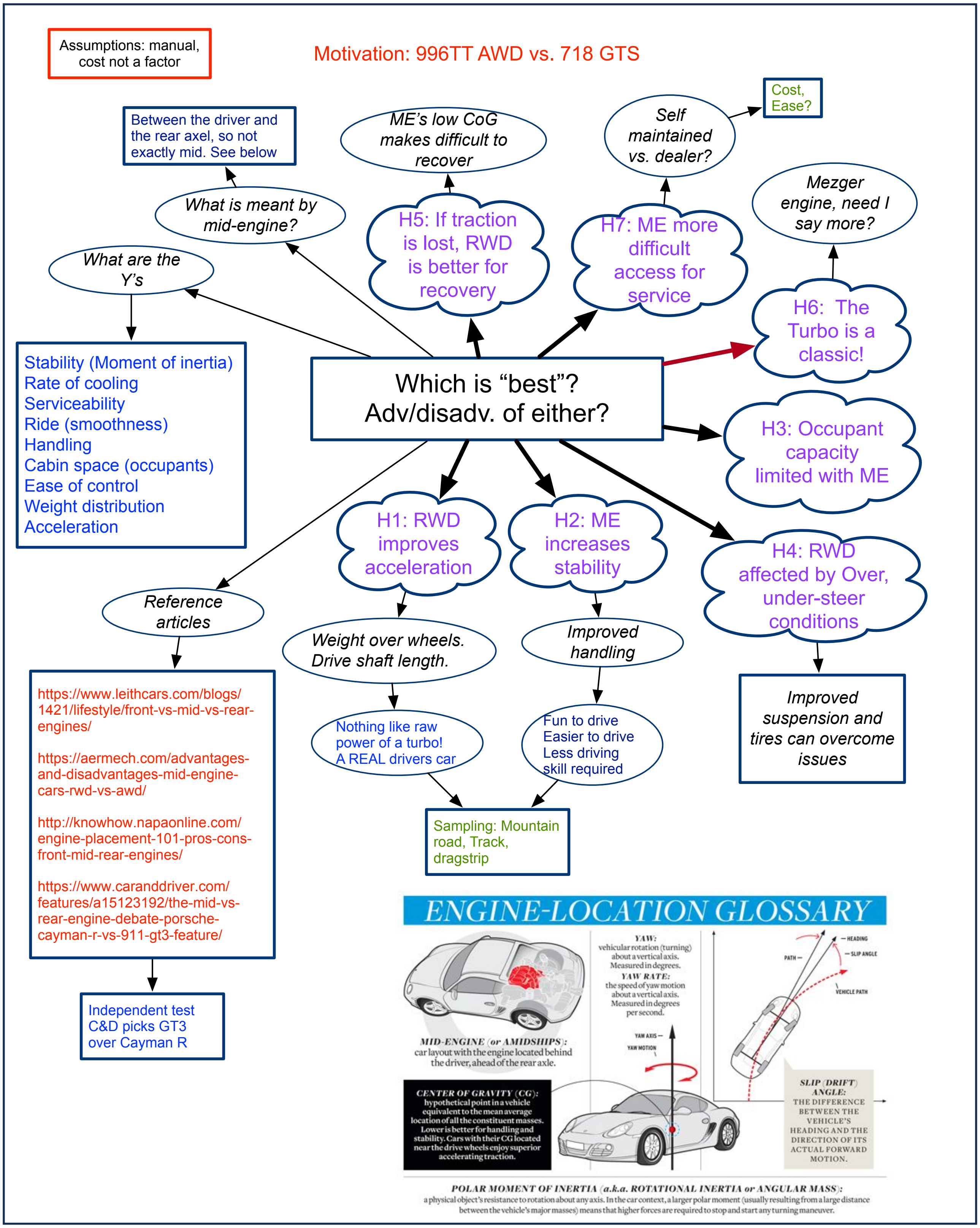


Figure A-9: Thought Map Comparison of Different Automobile models

Lastly, a Thought map on candidate selection for Sigma Science Training, Figure A-9.

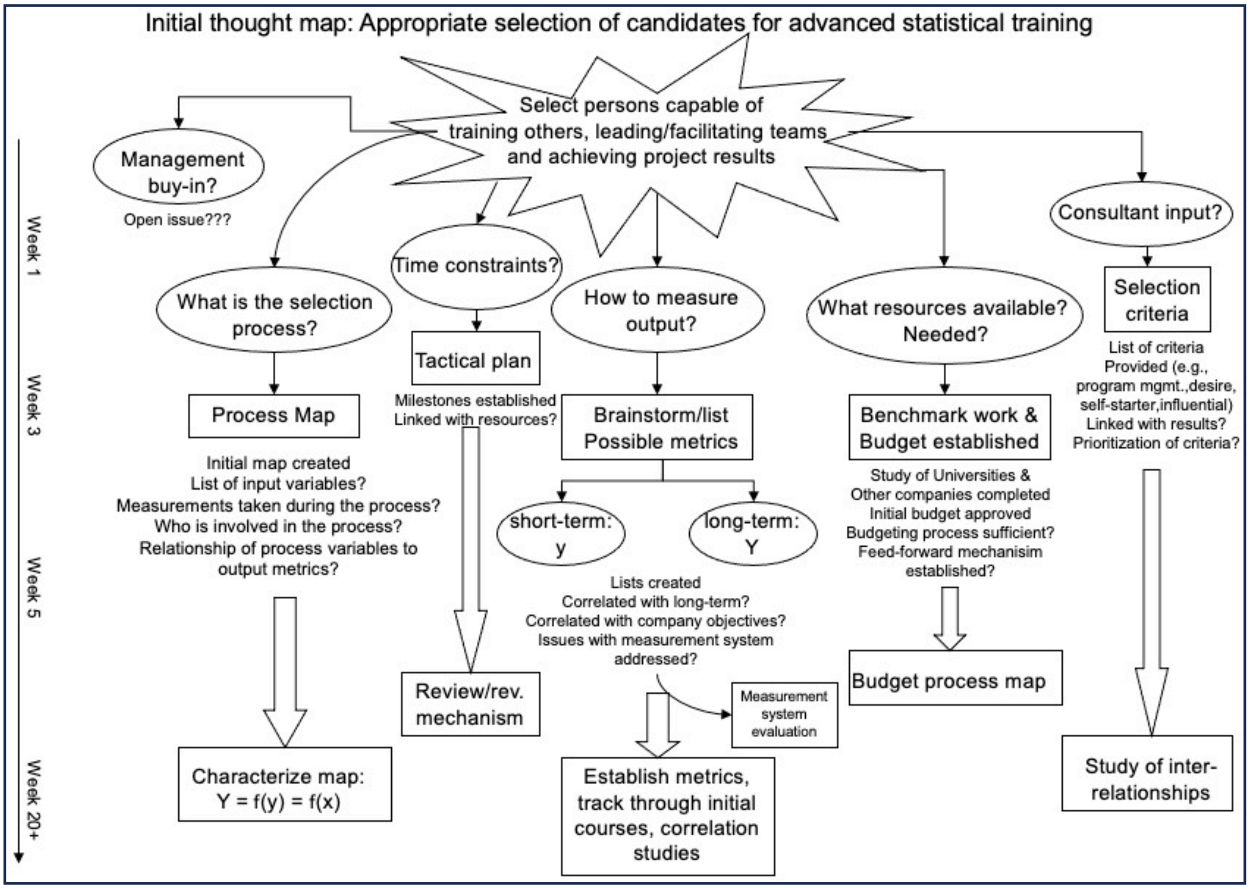


Figure A-9: Thought Map for Candidate Selection

Appendix B
Example Process and Product Maps

The maps in this appendix are not meant to be templates, just illustrating different examples of process and product maps. Figure B-1 is a process map for user interface of an appliance.

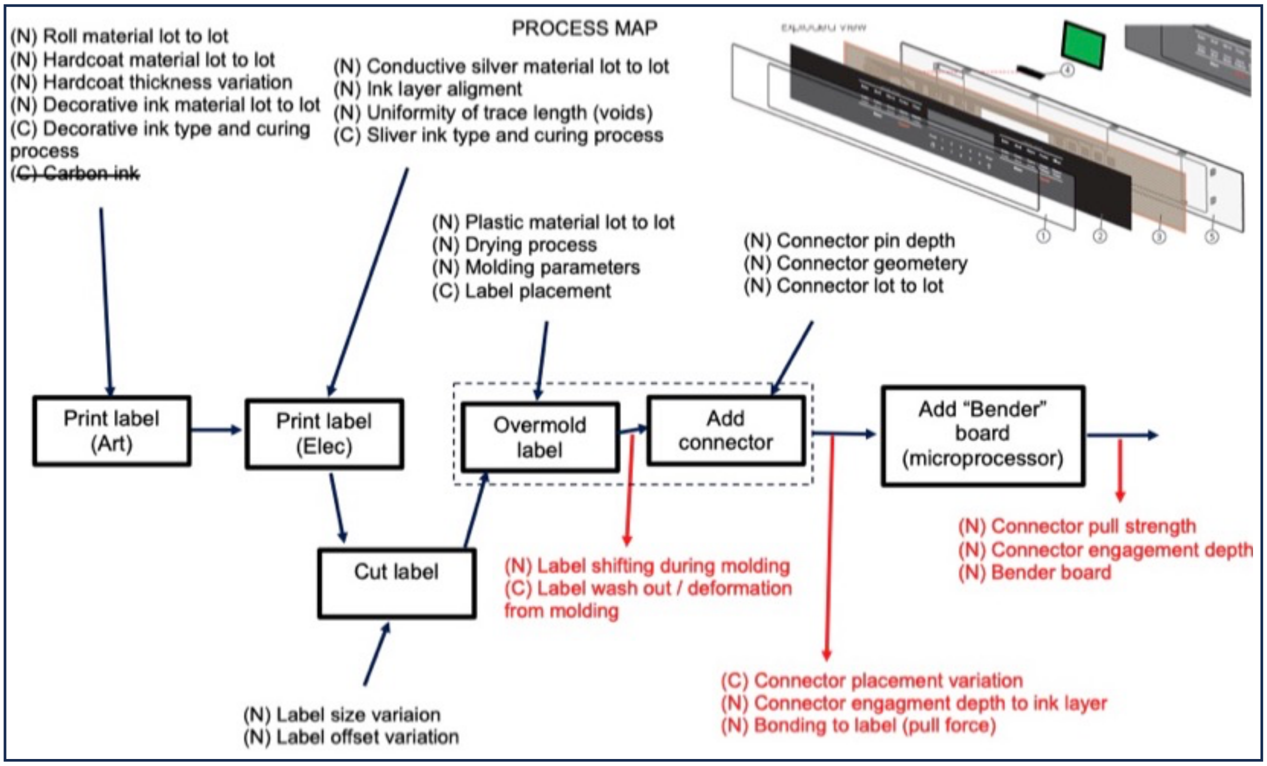


Figure B-1: Process Map for Appliance UI

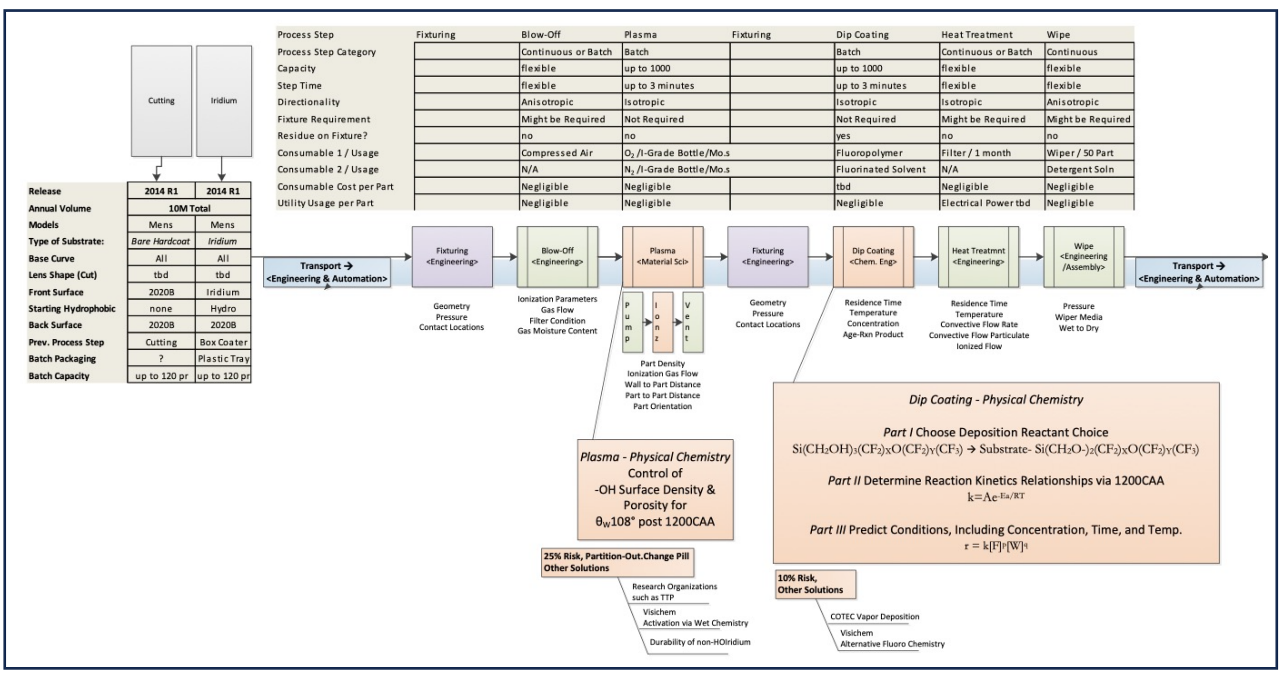


Figure B-2: Process Map Chemical Process

Figures B-3 and B-4 are maps of injection molding and thermal cycling processes.

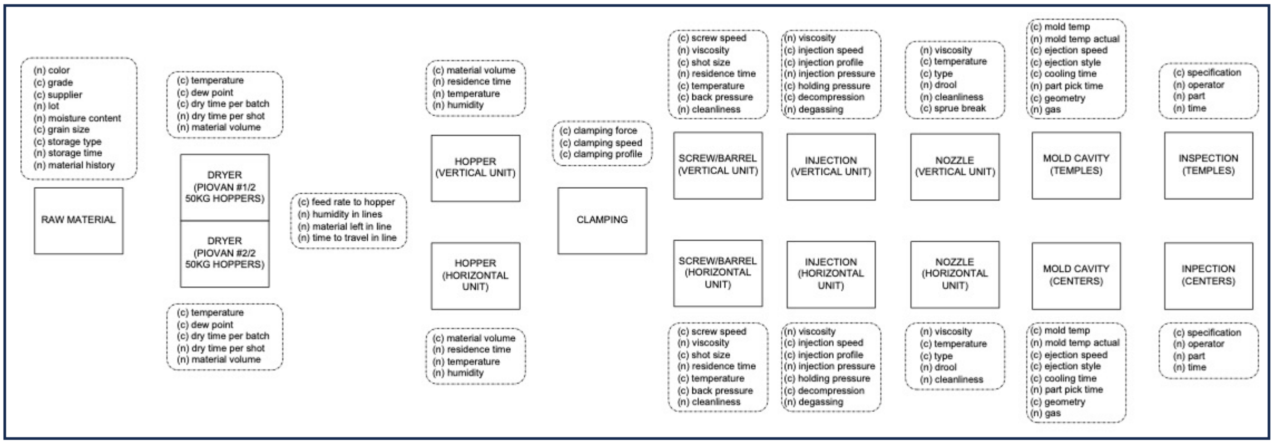


Figure B-3: Injection Molding Process Map

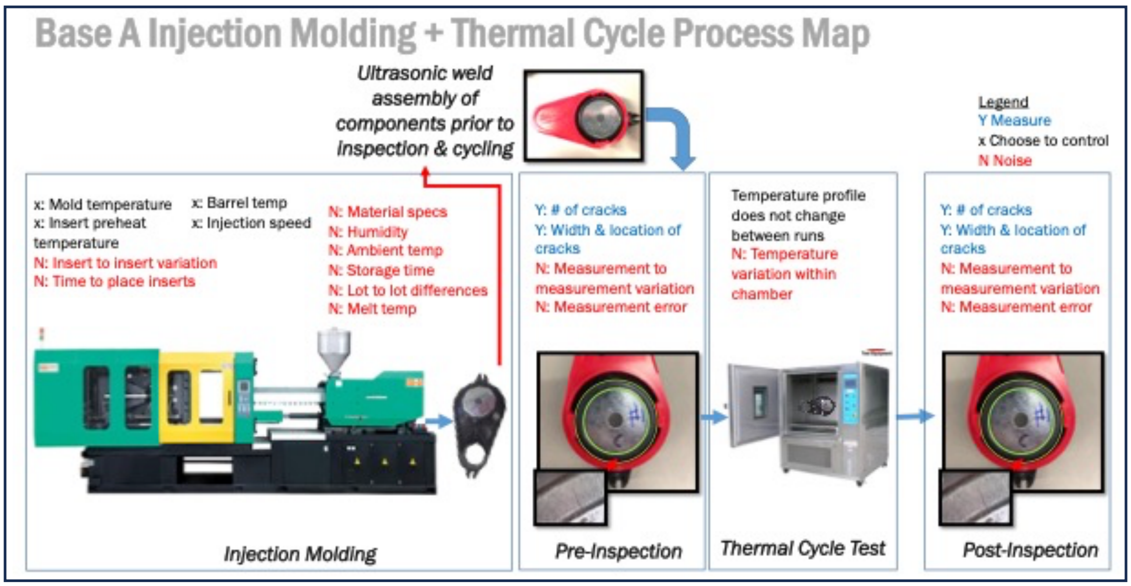


Figure B-4: Process Map Molding and Thermal Cycling

Figure B-5 is a map of the weld strength testing process. Figure B-6 is a non-manufacturing example.

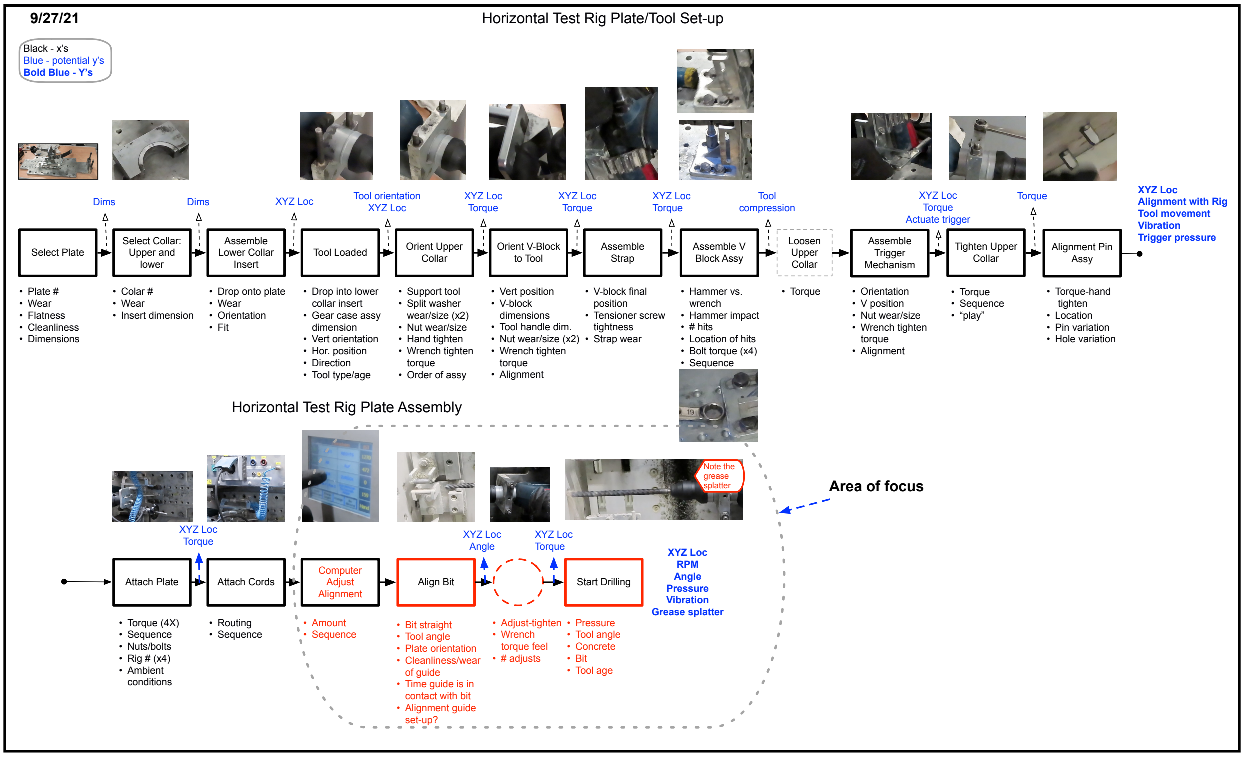


Figure B-5: Process Map Test Rig

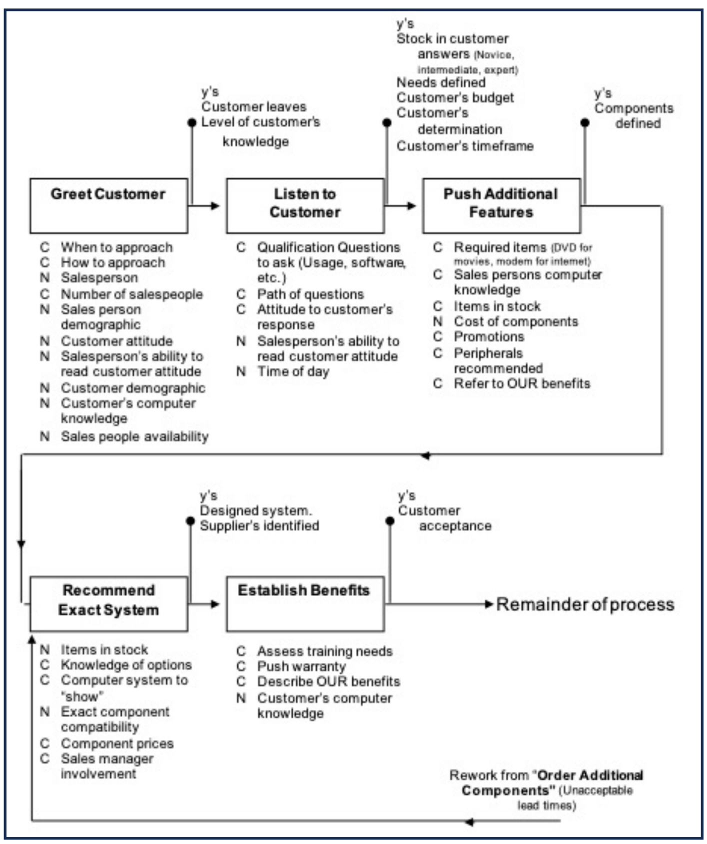


Figure B-6: Process Map Customer Interface Sales

The remaining figures in this appendix are example Product maps. Figure B-7 is a map of an oven cavity.

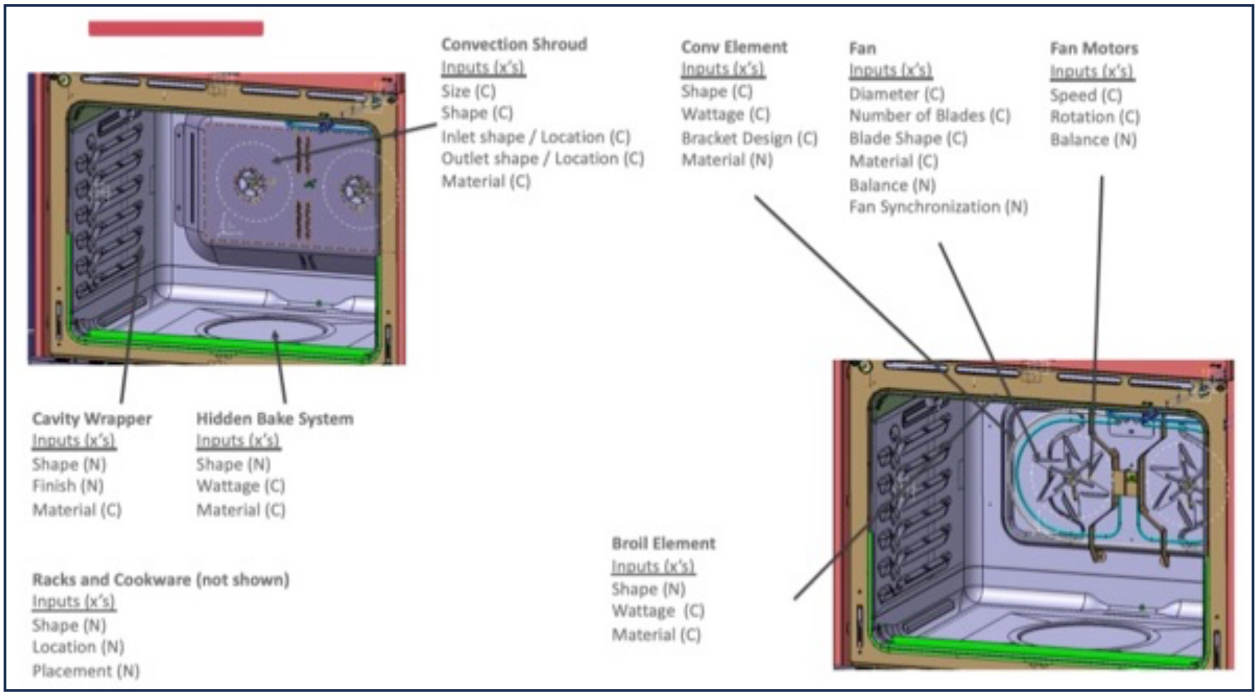


Figure B-7: Oven Cavity Product Map

Figure B-8 contains a map of a hole saw and linear edge technology.

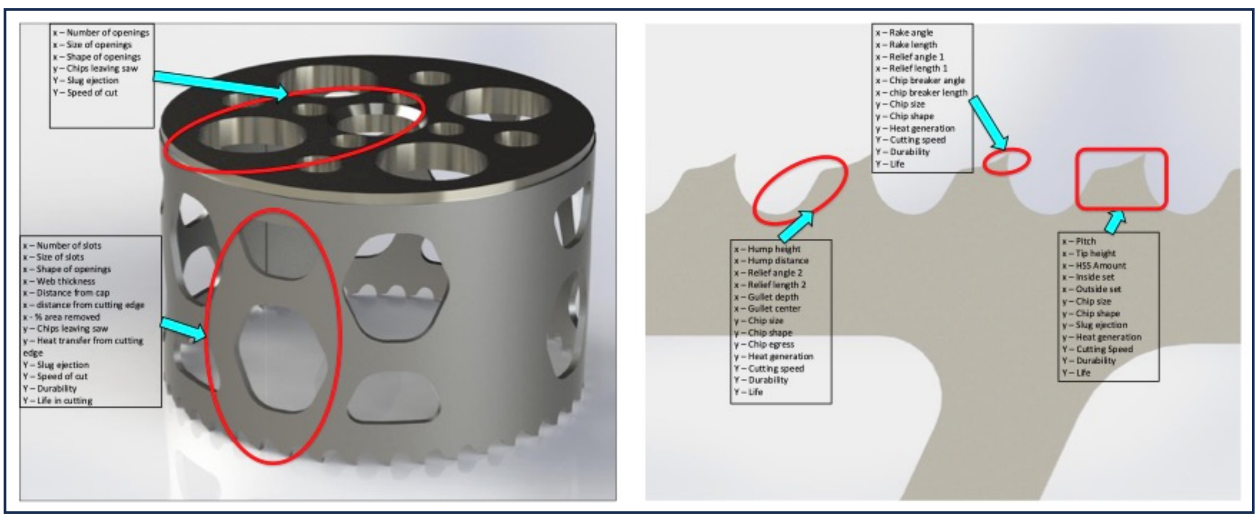


Figure B-8: Product Map: Hole Saw and Linear Edge Blade

Figure B-9 is a map of a grill and Figure B-10 is an oven door.

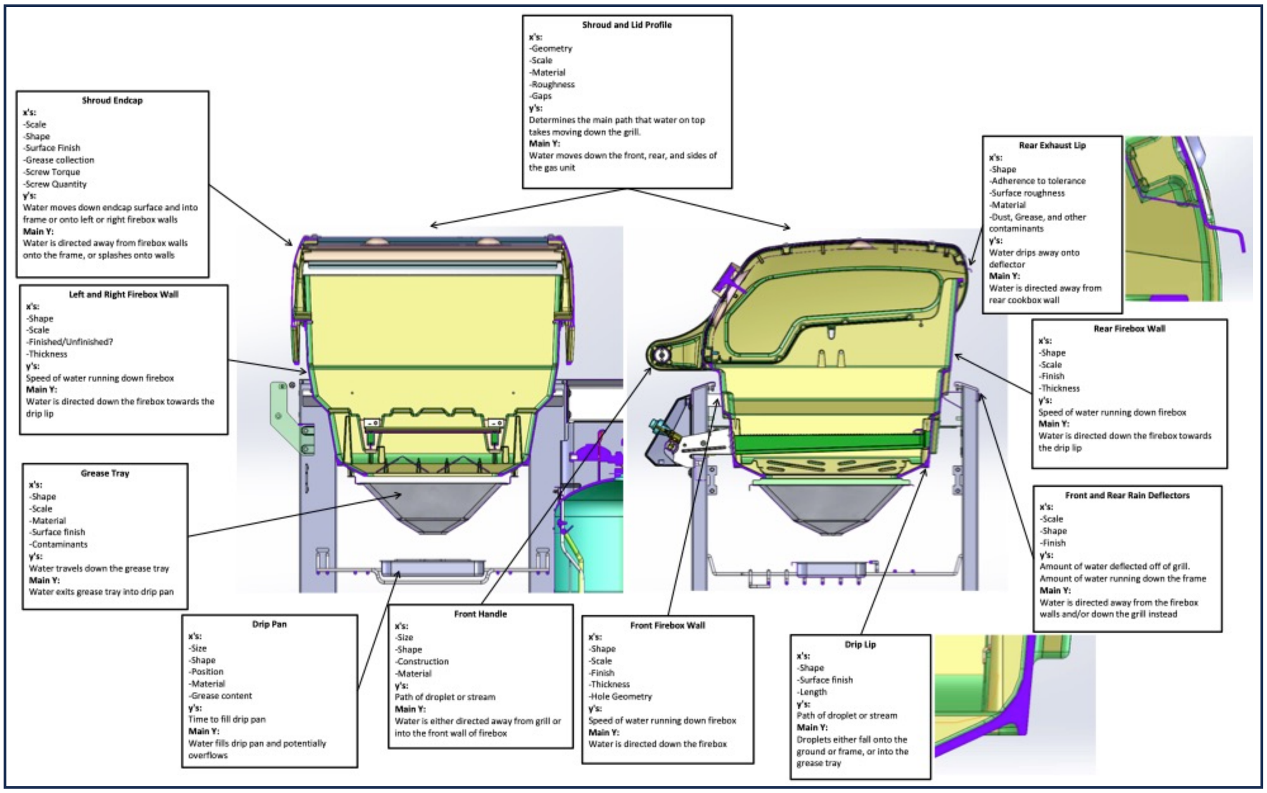


Figure B-9: Product Map Outdoor Grill

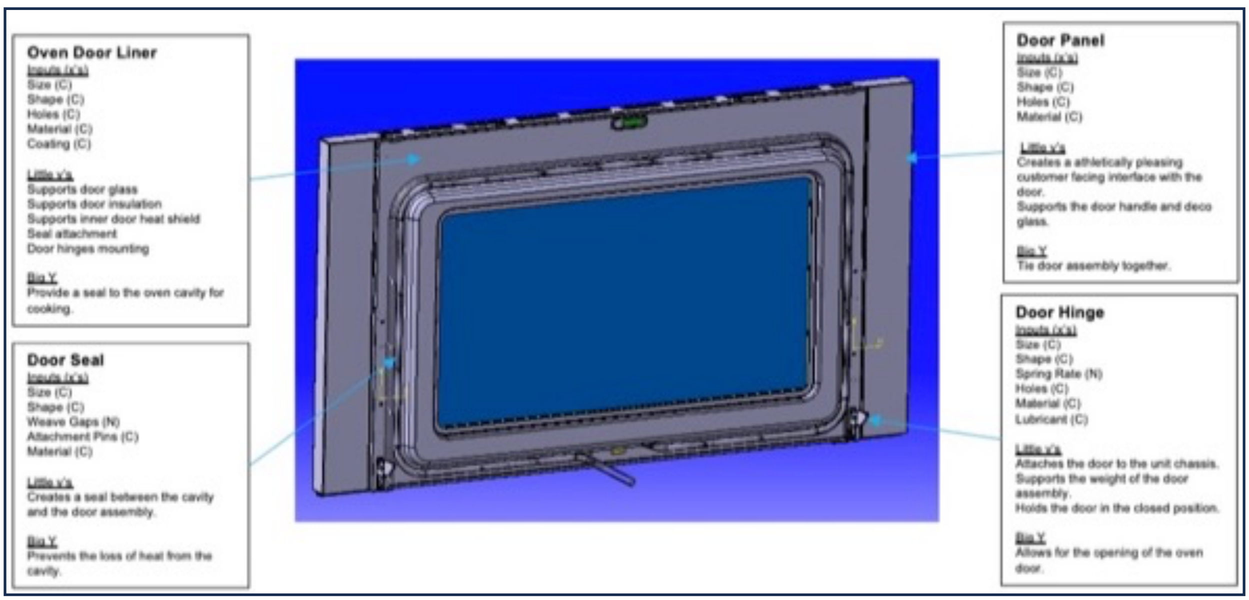


Figure B-10: Product Map Oven Door

Figure B-11 is a map of a steering assembly and B-12 is a writing instrument.

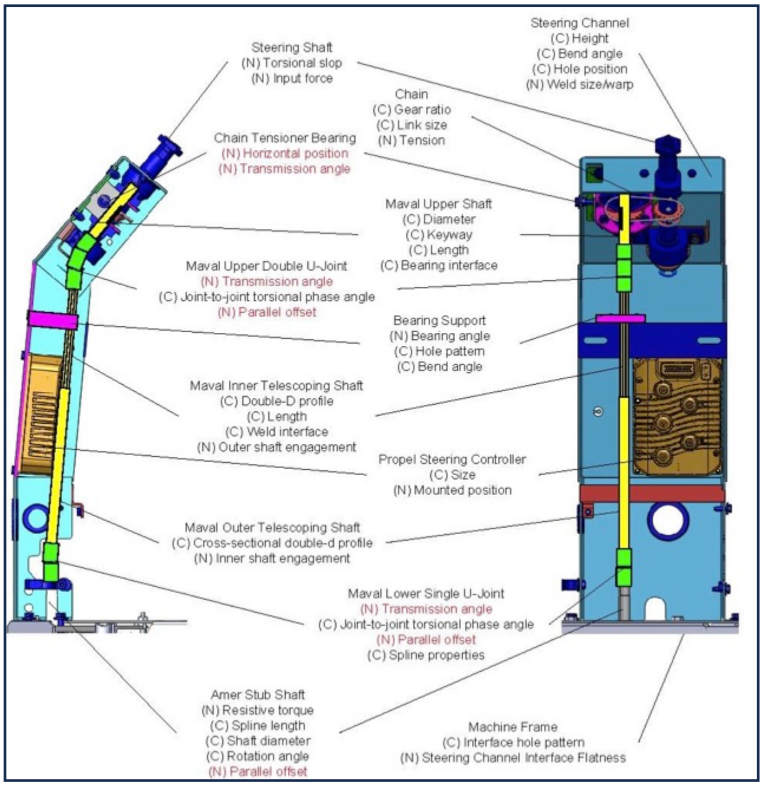


Figure B-11: Steering Assembly Product Map

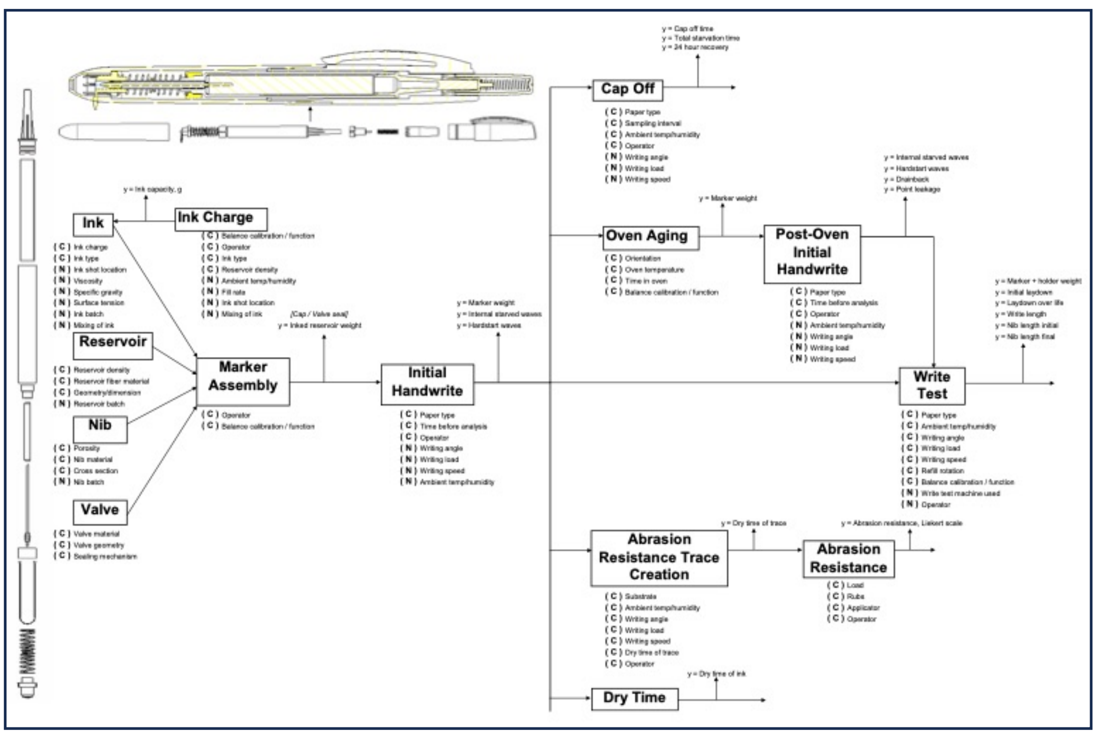


Figure B-12: Product Map Writing Instrument

Appendix C
Sampling and DOE Templates

These templates are intended to support disciplined thinking, not replace it.

Sampling Plan Template

Provide a description of the motivation for the study and any situation diagnostics: What are the symptoms? What are the phenomena of interest? Trying to understand what is happening or implement a solution? Is this product in production? Can the design be changed? What is the frequency of occurrence of the phenomena?

a. Always (common)
b. Occasional (repeated, cyclical)
c. Rare (special)

What is the objective of this Sampling Plan?

What is the sense of urgency? What resources are available?

Has any historical/observational data been analyzed? Has this led to any hypotheses?

Where in the knowledge continuum is current understanding?

Has a list of hypotheses and questions been developed?

- Insert link to Thought map

List, in rank order, the Y's: Has measurement system uncertainty been quantified?

Y#	Description	Unit	Precision	Practical Significance

Do the Y's already exist or do they need to be created? Are the measurement systems available/capable?

Have x's on the process/product maps been color coded to match the layers of the tree?

- Insert link to Maps color coded to match the layers

Rank order these global effects in relation to the specific objective of this study:

Possible Effect	***Rank***
Noise	
Main Effects (Factors)	
Two-Factor Interactions (2nd order linear)	
Noise-by-factor Interactions	
Simple Curvature (2nd order non-linear)	
Complex Non-linear (≥3rd order non-linear)	
≥ Three-Factor Interactions (3rd order linear)	
Stability	
Leverage	
Measurement Uncertainty	
Mean	
Variation	

What Layers are you going to include in the study? How do the plans partition the different sources of variation?

Indicate whether layers are Nested, Systematic or Crossed. Describe links to hypotheses:

Layer	Color	N/S/C	X's Captured	Hypotheses

What are predicted next steps for each possible outcome?

Layer	Special - next steps	Common - next steps

Has a data collection sheet been created in the software? Has data been predicted? Practice analysis.

- Insert link to Data collection spreadsheet

If no practically significant variation is captured in the study, list next steps.

Create multiple Sampling plans. What are the pros and cons for each plan? What is restricted, separated, confounded? Where will it put you on the knowledge continuum? What resources (Equipment time, operator time, raw materials, measurement) are required to acquire the data? Is it reasonable?

- Insert link to Sampling Plans (Trees) here

Plan	Restricted	Separated	Confounded	Resources Required

Steps of analysis.

1. Verify the tree matches the data (if not, modify the tree to match how the data was collected).
2. Has a practical amount of variation or the phenomena of interest been captured in the study?
3. If there are multiple Y's, perform multivariate analysis. Create scatter plots of the correlations. Test for multivariate outliers using Mahalanobis.
4. Create variability plots
   a. Nested Study: Use control charts to partition and quantify the variation due to each layer as appropriate
   b. Systematic: Color code the systematic layers in the variability plots and use control charts to partition and quantify the sources of variation
   c. Crossed: Create variability plots using every permutation of the ordering of layers. Partition and quantify sources of variation using ANOVA.

DOE Planning Template

Provide a description of the motivation for the study.

What are the symptoms? What's the phenomena of interest? Motivation to understand what is happening or implement a solution? Is this product in production? Can the design be changed? What is the frequency of occurrence of the phenomena?

What is the objective of this experiment?

What is the sense of urgency? What resources are available?

Results from analysis of any historical/observational data?

Performed any sampling plans? What did you learn?

- Insert link to Sampling plans

Where are you on the knowledge continuum? Are you ready to perform a DOE?

Have you developed your list of hypotheses and questions you are interested in studying?

- Insert link to Thought Map

List, in rank order, the Y's: Has measurement system uncertainty been quantified?

Y#	Description	Unit	Precision	Practical Significance

Do the Y's already exist or do they need to be created? Are the measurement systems available?

Rank order these global effects in relation to the specific objective of this study:

Possible Effect	***Rank***
Noise	
Main Effects (Factors)	
Two-Factor Interactions (2nd order linear)	
Noise-by-factor Interactions	
Simple Curvature (2nd order non-linear)	
Complex Non-linear (≥3rd order non-linear)	
≥ Three-Factor Interactions (3rd order linear)	
Stability	
Leverage	
Measurement Uncertainty	
Mean	
Variation	

Have x's on the Process/product maps been classified as noise and controllable?

- Insert link to Maps

What x's will be held constant during the experiment (inference space)?

What noise will be changing during the experiment?

What is the rank order of model effects through 2nd order (Main effects and 2 factor interactions)?

What Factors are you going to manipulate? At what levels? Describe links to hypotheses:

Factor	**Code**	**Low (-)**	**High (+)**	**Hypotheses**
	A			
	B			
	C			
	D			
	E			
	F			
	G			
	H			
	I			
	J			
	K			
	L			
	M			
	N			

What are your next steps for each possible outcome?

Effect	No effect - next steps	Strong effect - next steps

Has a data collection sheet been created in the software? Has data been predicted? Practice analysis.

- Insert link to Data collection worksheet

What are some possible special causes?

If no significant variation is created, list next steps.

If there is significant variation, but it is not assignable to the design factors, list next steps.

What are predicted size of effects?

Factor	Prediction	Magnitude

Create multiple FRD's. Include the predicted model.

- Insert link to FRD's here

What are the pros and cons for each plan? What are the pros and cons for each plan? What is restricted, separated, confounded? Where will it put you on the knowledge continuum? What resources (Equipment time, operator time, raw materials, measurement) are required to acquire the data? Is it reasonable?

FRD	Restricted	Separated	Confounded	Resources Required

Steps of Analysis

0. Does the data match the FRD?
   Note 1: If there are repeats, the data must be summarized before analyzing the treatment effects. You can do variability plots for layers nested within treatments. To determine appropriate summary statistics, do a R-chart for each Y (Then summarize data with means and variance estimates if appropriate.
   Note 2: If there are multiple Y's consider correlation and use Mahalanobis to check for outliers
   Note 3: If there are blocks, the complete data set should be analyzed first, then possibly each block separately if block-by-factor interactions exist.
   Note 4: If the experiment is a split-plot design, use split-plot analysis[1].

Practical:

1. Look at the data in ANOG order. Look for:
   a. Total variation and Practical change (ΔY, PSR>1),
   b. Obvious unusual data points,
   c. Patterns associated with factor columns
   d. Compare with predictions
2. Plot MR in ANOG order. Look for:
   e. Special causes/Jumps in the data

Graphical[2]:

3. Analyze standard least squares and rank effects/coefficients
   f. Pareto. Look for:
      - Effects that appear practically significant in the experiment
   g. Normal/half normal Plots. Look for:
      - Effects that appear statistically significant in the experiment (fat pencil test)
4. Main effects/interaction plots.
   h. Main effects for direction
   i. Interaction Plots to interpret interactions.

Quantitative (& Additional Plots):

5. ANOVA / Regression
6. Simplify model by removing unimportant terms (p-values, RMSE, RSq vs. RSq adj Δ)
7. Residuals analysis

Summarize:

8. Summarize findings, pick the predicted action that matches what the data supports
9. Update Thought Map (make sure the findings make "engineering sense")
10. Iterate, execute next sampling plan

[1] For split-plots, separate outputs for each plot (WP & SP) are required.
[2] For nested layers, variability plots may be used to look for patterns (group by treatments or run order)

The following is an example of the DOE template completed for an in-class project trying to optimize flight time for a paper helicopter.

Helicopter DOE Planning

Provide a description of the motivation for the study and any situation diagnostics:
Customers of the helicopters want them to fly longer, land on-target and be more predictable. The project is under time constraints that bias the study towards DOE. Sampling prior to DOE would have been helpful, but was not done.

What is the objective of this DOE?
To plan and run an experiment per Appendix A, Session 4 binder. Run the first of a potential sequence of experiments to understand factors affecting the flight time, accuracy, precision and consistency of paper helicopter flight. This will likely be a screening design to focus subsequent efforts.

What is the sense of urgency? What resources are available?
Must run at least one set of data Thursday before leaving class. Resources available are what were given to the teams to create the different levels for factors to be tested. Additional NOISE variables are to be identified by the team for consideration and investigation in the experiment.

Has any historical/observational data been analyzed? Has this led to any hypotheses?
Hypotheses have been developed and provided to us for evaluation. Additionally the hypotheses have been linked to factors for manipulation in a DOE.

Have any sampling plans been performed? What was learned?
Unknown at the time of this exercise, but no sampling or other data has been provided other than some anecdotal data saying the helicopters typically fall between 1-2 seconds with 15 seconds being the extreme after sequential testing.

Where in the knowledge continuum is current understanding? Is an experiment ready to be performed?
Current level is fairly low. An experiment is ready to be performed to learn more.

Has a list of hypotheses and questions been developed?

1. Length of flight will increase with:
    a. Decrease in weight
    b. Improved aerodynamics
    c. Improved stability
2. Stability in flight will be improved with:
    a. More rigid construction
    b. Added ballast
    c. Consistent dimensional characteristics
3. Making the design robust to noise is essential to a stable, consistent and predictable product:
    a. Air drafts up or down will affect flight time
    b. Air drafts sideways will affect accuracy
    c. Release technique will affect consistency
4. The measurement system lacks precision:
    a. Determining release/landing times
    b. Locating the landing point

List, in rank order, the Y's: Has measurement system uncertainty been quantified?

Y#	Description	Unit	Precision	Practical Significance
1	Flight time	sec	?	0.1
2	Deviation from target	inches	?	2.00
3-4	Variation in both of the above Y's		?	0.1

Do the Y's already exist or do they need to be created? Are the measurement systems available/capable?

Yes, the Y's exist and were chosen for the experiment. Measurement systems (stopwatch app on cellphone, ruler and a tape measure) are available. The precision of the measurement systems is unknown and should be studied.

Rank order these effects in relation to the specific objective of this study:

Possible Effect	***Rank***
Noise	2
Main Effects	1
Two-Factor Interactions	4
Noise-by-factor interactions	3
Simple curvature	5
Complex non-linear (≥cubic)	5
≥3^{rd} order linear	5
Stability	later
Leverage	5
Measurement uncertainty	1
Mean	1
Variation	1

Have x's on the process/product maps been categorized as noise and controllable?

Process and product maps were not provided. The team will create these prior to developing the experiment plans. We are going to do a "dry run" of the building of a copter and a Product map to understand the components of a copter.

We did create an initial list of noise:

-Helicopter construction
-Cuts
-Within constructor
-consistency of cut method
-Constructor to constructor variation (different means)
-alignment to cut line (left, right, center)
-squareness of cuts
-correct length of cuts
-Folds
-Within constructor
-consistency of fold method
-Between constructor (different means)
-alignment to fold line (left, right, center)
-squareness of folds
-crispness of folds
-fold technique (number of folds, firm crease)
-Taping
-Within constructor
-consistency of tape method
-Between constructor (different means)
-amount of tape used (length, layers)
-centering of tape
-crispness of tape crease)
-Material variation
-Within and between paper batch
-paper weight
-paper strength
-paper
-Within and between tape roll
-thickness
-stickiness
-tear to tear squareness
-Execution
-Set-up
-Within operator
-consistency of all factors between operator
-ambient conditions (run to run)
-Between operator (different means)
-Location (XYZ)
-orientation of helicopter
-Release
-Within operator
-consistency of all factors between operator
-Between operator (different means)
-Speed of release
-plumb of release
-follow-through (not disturbing the airflow)
-Timing
-Within operator
-consistency of all factors between operator
-Between operator (different means)
-reaction time to start
-reaction time to stop
-Spotting
-Within operator
-consistency of all factors between operator
-Between operator (different means)
-location of spot (XYZ)
-calling of stop

What noise x's will be held constant during the experiment (inference space)?

See FRD for specifics. Ambient conditions will be restricted as the experiment will take place in one afternoon. We have only 16 sheets of paper from unknown lots of paper (at least 2 for different weights). One ruler, one pair of scissors, one tape measure, one roll of tape. Height of drop will be kept relatively constant. Only one size paper clip will be used.

What noise will be changing during the experiment?

See FRD for specifics: building technique, drop technique, measurement devices, within day ambient (e.g., possibly AC draft), paper templates, within roll of tape.

What is the rank order of model effects (Main effects and 2 factor interactions)?

For all Y's, main effects are ranked above 2nd order effects. Predicted rank order on model effects:

Effect	Code	Time	Accuracy	Variability
Paper Weight	P	4	3	4
Body Width	B	8	8	8
Body Length	L	5	7	7
Wing Length	W	1	4	3
Taped Body	T	7	6	6
Taped Wings	G	6	5	5
Wing Fold	F	9	9	9
Paper Clip	C	2	2	2
Noise/Block	BLK	3	1	1

What Factors are you going to manipulate? At what levels? Describe links to hypotheses:

1. Structural rigidity of the helicopter will affect stability and improve length of flight (type of paper = P).
2. Dimensional characteristics of the body will affect aerodynamics, small dimensions will reduce weight (body length = L), while larger dimensions will increase rigidity (body width = B).
3. Wing length will affect length of flight due to larger wing surface area (wing length = W).
4. Adding ballast will improve stability of flight and aerodynamics, but will reduce flight time (paper clip = C).
5. Adding materials to improve aerodynamics and structure will increase flight time (tape body = T; wings = G).
6. Folding the end of the wings up will increase aerodynamics (wing fold = F).

What are predicted next steps for each possible outcome?

Effect	No effect - next steps	Strong effect - next steps
Paper Weight	Use (-) setting for least $$. Study lot-to-lot variation (sampling)	Investigate other materials to make copter from
Body Width	Use (-) setting for least $$.	Vary body surface to optimize
Body Length	Use (-) setting for least $$.	ibid
Wing Length	Use (-) setting for least $$.	Vary surface area of wing to optimize. More wings?
Taped Body	Use (-) setting for least $$.	Study aerodynamics of copter
Taped Wings	Use (-) setting for least $$.	Study aerodynamics and structural aspects of copter

Wing Fold	Use (-) setting for least $$.	Other modifications to sustain flight. Size of fold?
Paper Clip	Use (-) setting for least $$.	Study location and size of clip

Likely there will be follow-up experiments to further understand factor effects and settings. Also, for significant factors, sampling plans to understand incoming material and manufacturing variation.

Has a data collection sheet been created in the software? Has data been predicted? Practice analysis.

- See Helicopter Data

What are some possible special causes that may occur during the experiment?

Release technique of operator, unexpected airflow changes (AC coming on, doors opening, people movement), starting and stopping watch. Hitting an object during flight.

If no practically significant variation is created, list next steps.

Review identified noise further and see if the appropriate x's have been chosen to be in the experiment. Adjust (broaden) the levels to see if significant variation can be found.

If there is practically significant variation, but it is not assignable to the design structure (factors or interactions), list next steps.

Review items in the unit structure and see what is 'most likely' to be contributing factor(s) and either re-think whether some are controllable or run sampling plans to identify leverage. Determine how the follow-up experiment should be structured/run.

Create multiple FRD's. Include the predicted model and assignment of the degrees of freedom

- See FRD's

Design / Resolution / Plot / Sets x Repeats	**Setup**	**Comments**
Screening / Res IV / Efficiency Split Plot / 3x2	8 factors, 8 treatments, 8 copters, 2 drop operators, 2 operators for measurement, 2 measurements per operator. 64 drops total.	*Pros:* Noise factors with operator and measurement system taken into consideration. *Cons:* Lots of unit structure in sub plot with plenty of space for unit structure to change from drop to drop, which raises the water level. Low resolution on noise. Lots of resources needed. (64 runs)
RCBD / Res IV / Efficiency Split Plot / 1x3	2 Blocks, 7 design factors, 8 copters, 16 treatments, 3 measurements per treatment, 48 drops.	*Pros*: Full resolution on block and block by factor interactions, additional measurement error evaluation. *Cons:* Still resource-intensive, 3 repeats chosen without much rhyme or reason, operator error is not considered and therefore the inference space is too narrow

RIBD / Res IV / Efficiency Split Plot / 2x2 repeats **This became our chosen FRD.**	2 Blocks, 6 Factors (2 of them combined into one), 16 Treatments, 2 drops per treatment, 2 measurements per drop	*Pros:* Noise by factor interactions captured, repeats assess measurement error, repeats on drops assess within operator error, block addresses operator to operator error, least resource intensive.

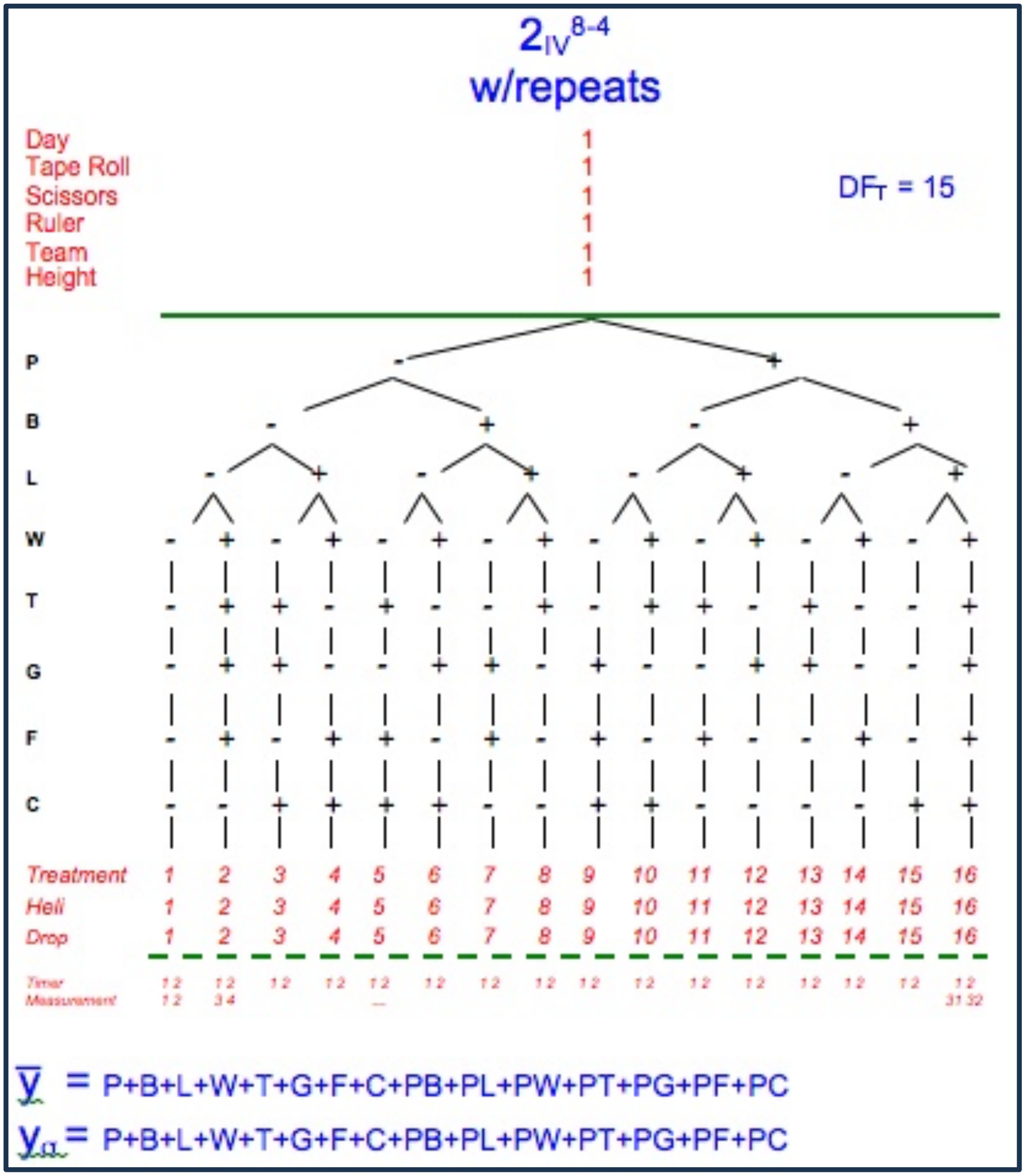
Or something like this (see FRD's following)

FRD	Restricted	Separated	Confounded	Resources Required
1	US	main effects	2^{nd} order	16 treatments, 2 measures
2	US	main effects, all block effects	>2^{nd} order	32 treatments, 16 copters
3	US	main effects, block main effect	>2^{nd} order	16 treatments, 16 copters, 2 measures

What are predicted size of effects?

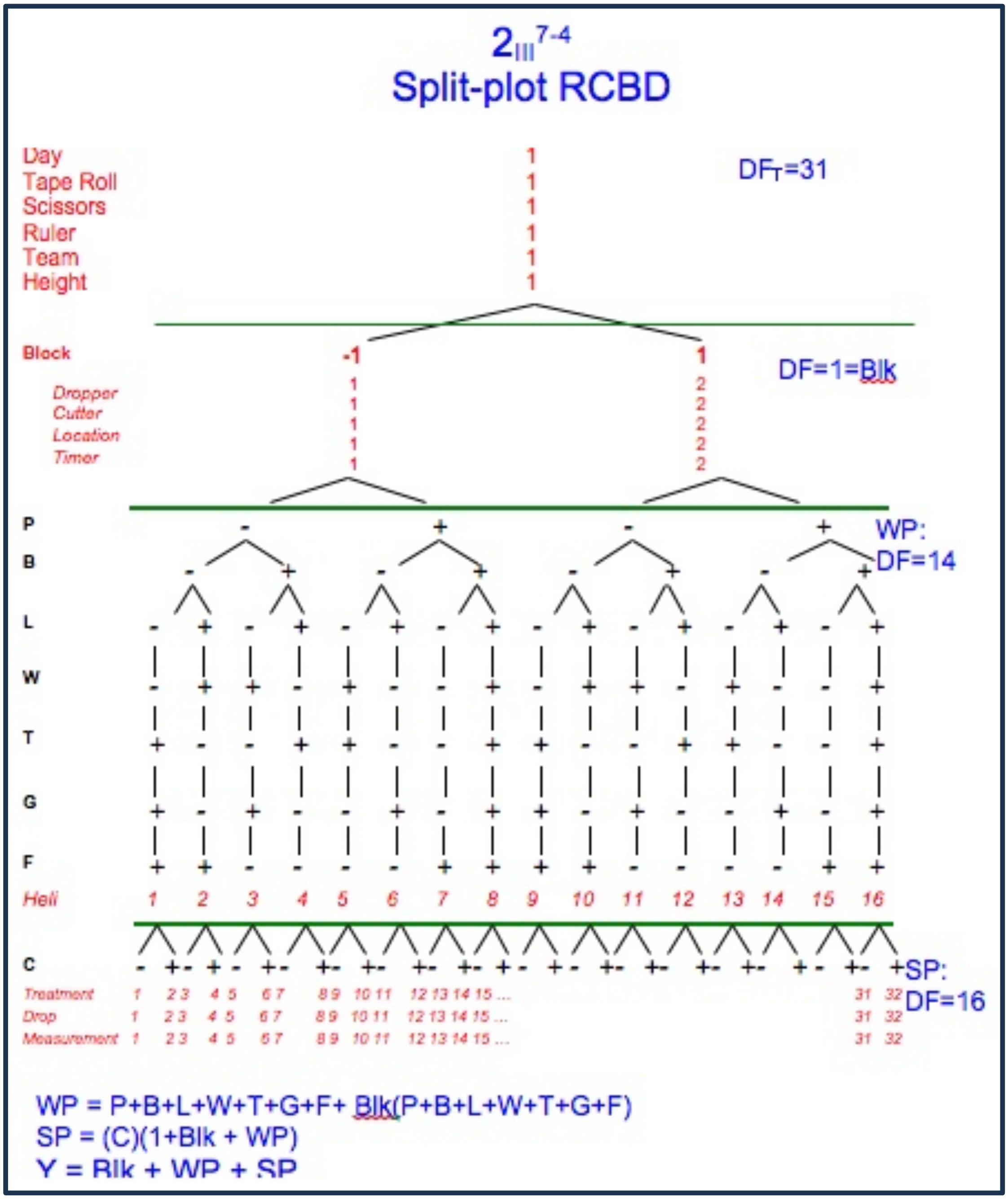
Factor	Prediction	Magnitude
Paper Weight	+P will have effect on flight time (Y1) and slightly reduce deviation from target (Y2).	Y1 = 0 Y2 = -0.25 Y3/4 = -0.1
Body Width	+B will decrease rotation speed/decrease flight time (Y1) and increase deviation from target (Y2).	Y1 = +0.1 Y2 = +0.5 Y3/4 = 0
Body Length	+L will increase rotation speed/increase flight time (Y1) and decrease deviation from target (Y2).	Y1 = +0.2 Y2 = -0.5 Y3/4 = 0
Wing Length	+W will increase rotation speed/increase flight time (Y1) and decrease deviation from target (Y2).	Y1 = +1.0 Y2 = -2.0 Y3/4 = 0.5
Taped Body	+T will have little to no effect on flight time (Y1) and have little to no effect on deviation from target (Y2).	Y1 = 0 Y2 = 0.1 Y3/4 = -0.2
Taped Wings	+G will increase rotation speed/increase flight time (Y1) and decrease deviation from target (Y2).	Y1 = +0.3 Y2 = -1.0 Y3/4 = -0.2
Wing Fold	+F will have little to no effect on flight time (Y1) and decrease deviation from target (Y2).	Y1 = 0 Y2 = -0.5 Y3/4 = 0
Paper Clip	+C will decrease rotation speed/decrease flight time (Y1) and decrease deviation from target (Y2).	Y1 = -1.0 Y2 = -3.0 Y3/4 = -0.2

FRD 1



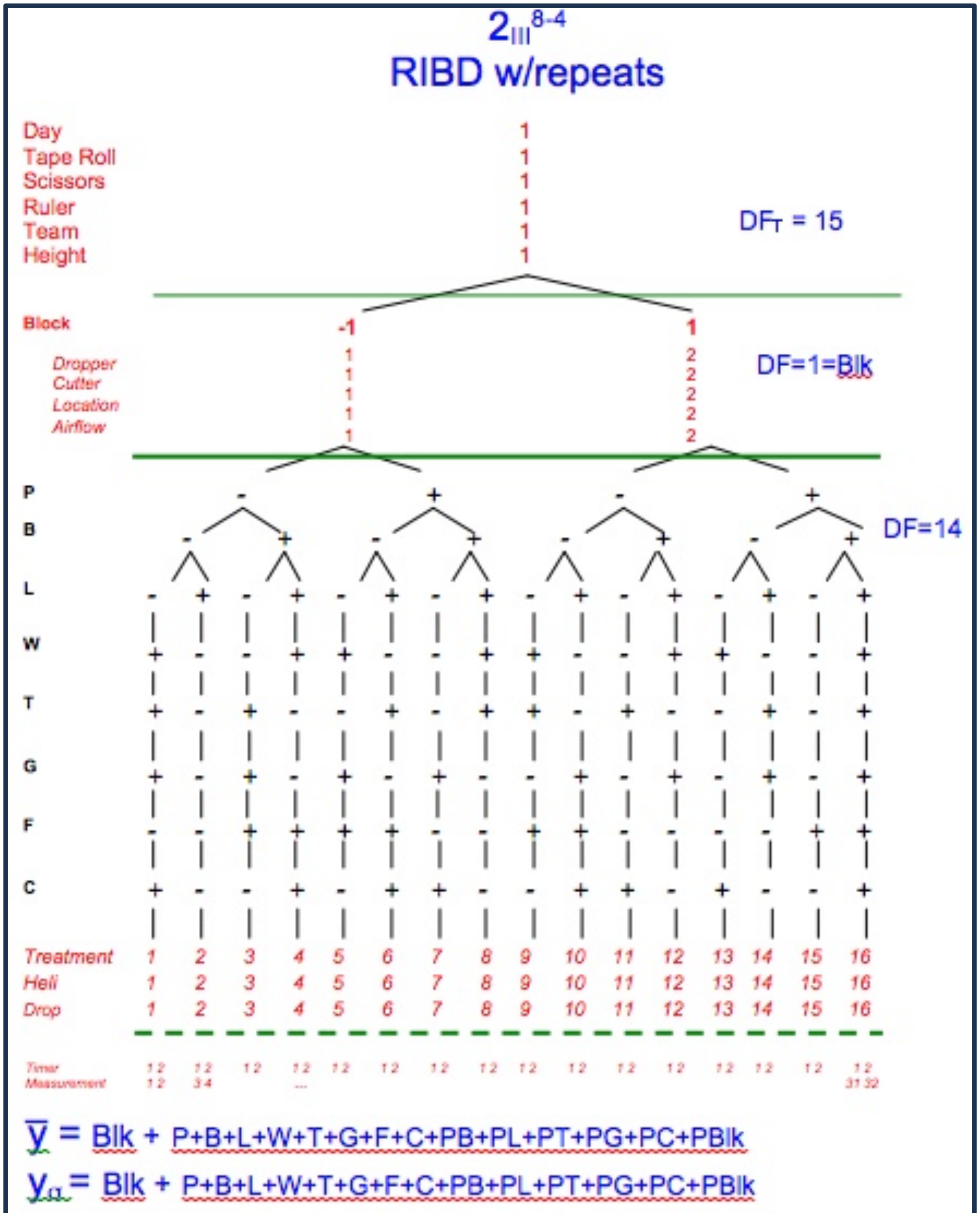
Design / Resolution	Noise strategy	Pros/cons
Screening / Res IV. 8 factors, 16 treatments, 16 copters, 1 drop, 2 measurements	Repeats used to partition measurements and timer from treatments	*Pros:* Res IV screening design, measurement error partitioned increasing precision. A Y of variation could be calculated, but may not be useful? *Cons:* 16 copters of material. Narrow inference space.

FRD 2



Design / Resolution	Noise strategy	Pros/cons
Screening / Res III+. 8 factors, 32 treatments, 16 copters, 2 drops, 1 measurement.	RCBD run in 2 complete blocks	*Pros:* Res III screening design, Increased IS and increased precision. WP and SP increased precision. Full res. on block. *Cons:* 16 copters of material. Low factor res. Full res. on block effects. Measurement error confounded

FRD 3



Design / Resolution	Noise strategy	Pros/cons
Screening / Res III. 8 factors, 16 treatments, 16 copters, 1 drop,	Incomplete blocks / Repeats used to partition measurements and timer from treatments. A DF was used to create the incomplete blocks, 2 measurements	*Pros:* Res III screening design, Wider IS, measurement error partitioned increasing precision. Y of variation. Block effect. *Cons:* 16 copters of material. Low res factors and unit structure

Another example following the welding project Thought Map presented in Appendix A.

Welding DOE Planning
Update and Revision

The original project motivation was to understand the inconsistent performance tests of the cement carbide (CC) tip diffusion bonded to a steel flute. Measurement error in the test process is unknown and cannot be separated from the bit-to-bit performance variation. Hopefully, increase the integrity of the tip-to-flute "bond" and be able to claim product performance and competitive superiority.
The objectives have changed due to resources. Current objectives are:

1. Understand the effects of 3 carbide tip factors (these parts are already at X),
2. Determine and possibly assign effects of the measurement process (via cross-product arrays or blocks), and
3. Iterate what to measure and look for correlation among multiple response variables (Y's), and

Previous experimentation indicated significant noise effects (the "same" experiment when run again produced dramatically different results). In addition, production bits as well as competitive benchmarking tests have produced results that are varying with little understanding of why. We believe we are capturing some of the possible contributors to noise from the measurement process in this study. While we are NOT including multiple powder lots, we are including sintering batch-to-batch variation of tips (which we are exaggerating via experimentation). Also, we are NOT including assigning of the variation in the hardening process.

Historical perspective: No directed sampling has been done either to understand measurement systems, incoming materials or manufacturing variation. Consistency of the steel, CC, nickel, welding or hardening process is unknown. Much of the current knowledge is theoretical. The welding is a hybrid of Diffusion and Induction. The process is *tweaked* to maintain temperature profiles. We are attempting to figure out how to improve the weld integrity while not understanding why there is variation in weld integrity. Is this due to the actual welding variation or due to inconsistent measurement systems? The parallel effort to understand the factors that may affect the consistency of the test process is currently on hold.

A list of factors and hypotheses has been created and prioritized. Thought map attached.

Y	Description	Unit	Precision	Practical Significance
1	**Speed of drilling hole (current test)**	**sec**		
2	**Life drilling concrete (current test)**	**#holes**		
3	**Life drilling rebar (current test)**	**#holes**		
4	**Torsional strength (max axial impact)**	N-m		
5	**Microstructure/cross section Eta phase**			
6	Thermal camera image	ºC		
7	Physical dimensions (tip and flute)	mm		
8	Digital pictures/video of bit			
9	Qualitative observations (failure modes)	ordinal		0.5

Bold indicates measurements we will be taking. None of the measurement systems have been evaluated for capability and consistency.

The following table is my best guess of the ranking of global effects of interest:

Possible Effect	*Rank*
Noise (other than measurement)	4
Main Effects	1
Two-Factor Interactions	4
Noise-by-factor interactions	4
Simple curvature	3
Complex non-linear (≥cubic)	5
≥3rd order linear	5
Stability & Leverage	5
Measurement uncertainty	1
Mean/variation	1

Cursory descriptions and flow charts of the welding process, heat treat process and measurement process have been created, but no explicit delineation of X's or noise.

Inference space

There are two "groups" of bits: relatively small vs. large diameters. **The first study will restrict the study to one diameter for now (5/8").** One welding process/technology There are two different welding machines. We will **focus on one machine (diffusion bonding**), but then possible expand the study to include the second machine (resistance welding). **Current hardening process**. **One powder lot for tip manufacturing.** Within lots of incoming CC designed per the experiment factors. This is predicted to be small. ***Steel lots****…unknown lots of steel flutes and multiple steel suppliers?*

Noise varying during the experiment:

- Ambient.
- Electrode variability (contact area, surface finish, age, etc.). Not sure how to include or if we want to consider this in this iteration. Copious notes on electrode conditions will be taken.
- Hardness process variation is unknown.
- Steel lots variation unknown (lots of steel flutes, multiple steel suppliers)?

The following noise to be manipulated in the whole plot of a split-plot design.

Factor	Code	Low (-)	High (+)	Hypotheses
Magnetic Saturation	**X1**	153	167	Impact on hardness, free C and Eta phase. Value is target, we'll replace with actual value
Coating type	**X2**	Cobalt	Nickel	Nickel better to weld
Coating thickness	**X3**	2.5	5	Different weldability

There are 7 effects possible with 3 factors: X1, X2, X3, X1*X2, X1*X3, X2*X3, X1*X2*X3 where the * indicates interactions. Per our discussion, the SME has prioritized Main Effects (1st order terms) over two-factor interactions (2nd order effects). This results in a Resolution III design. The matrix is captured in the Factor Relationship Diagrams (FRD's). We discussed waiting to consider Electrode wear and contact resistance for a subsequent study or possibly we could include if we do any replication of the design we chose. For level setting, we are starting with 2 BOLD levels and one of those levels be the current conditions. We are concerned with the potential for variation in the testing process. We propose 2 strategies to handle this: Strategy 1 will be to vary certain factors in the measurement process and run these as replicates of the whole plot experiment (this is a split-plot design). The factors could include (should be discussed test dept.):

- Concrete: Hardness, with/without rebar, flat steel
- Power tool: Type, brand, age
- Pressure:
- Cycles: Temperature of the tool/bits
- Orientation: Horizontal/vertical, angles

FOR EXAMPLE:

Factor	Code	Low (-)	High (+)	Hypotheses
Concrete Block	N1	10mm Rebar	20mm Flat	*Inconsistent hitting of rebar*
Power Tool K545	N2	High Impact	Low Impact	*Inconsistent impact energy is supplied*
Pressure (weight)	N3	Low	High	*Inconsistent pressure affects results*

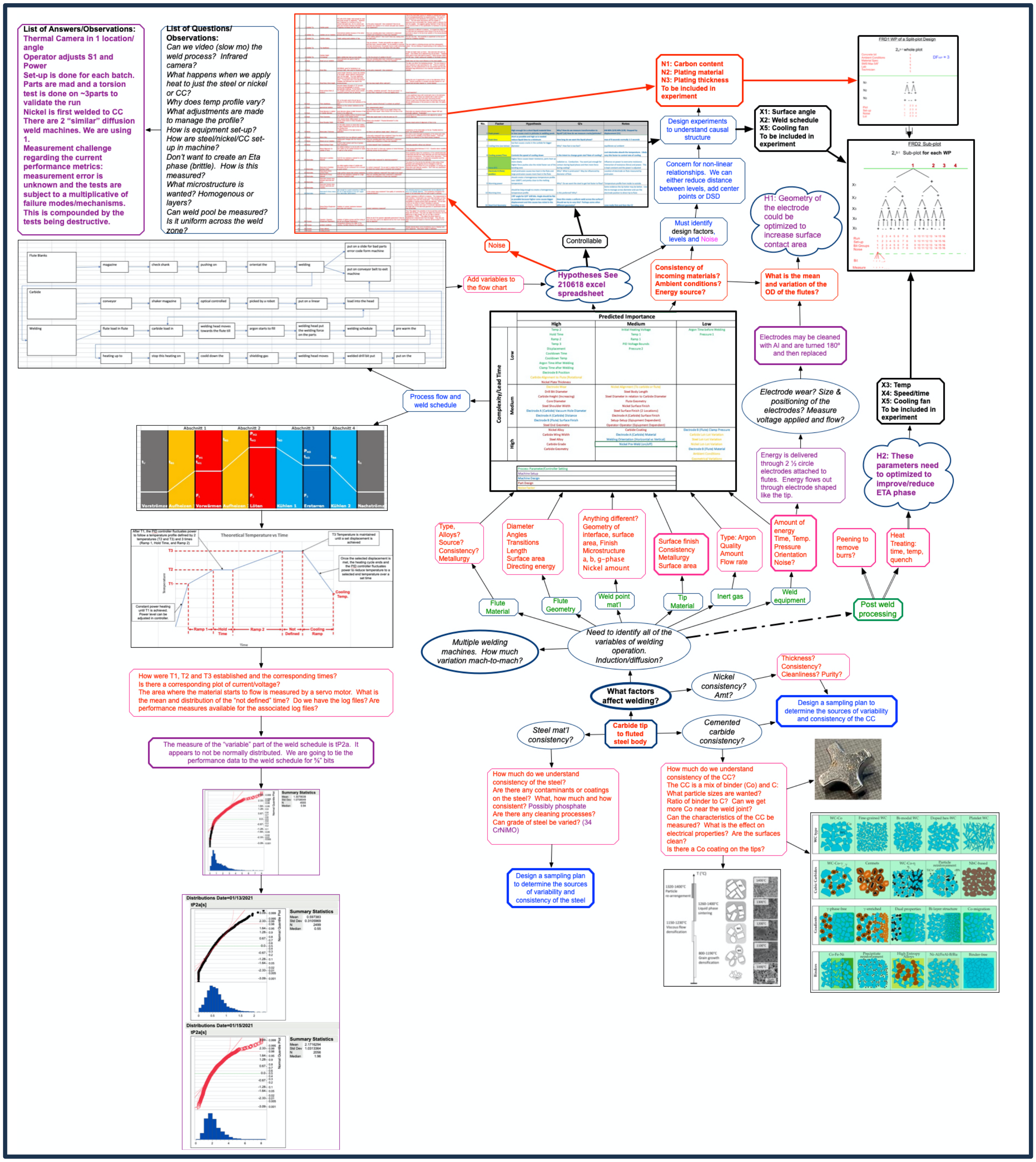
Predicted next steps for every possible outcome:

Effect	Hypothesis – Effect Size	Strong effect action to take
X1	**Large** effect, because the electrical properties change with the amount of carbon	Tighten the specs or find a defined setup for every batch with the knowledge of this value
X2	**Medium** effect, because Nickel coated carbide shown better weld results with Sawzall blades	Use the better coating material or investigate other potential coatings/ratios?
X3	**Low** effect, because the difference of the coating thickness is pretty low	Tighten the specs. Work with supplier?
N1	?? Investigate accuracy of hits (break blocks)	continue using flat steel as primary material? Investigate other materials?
N2	?? Continue using existing parameters? Re-test to find extents of drill bit to parameter interaction. Ex: increase power tool size / decrease bit size	Develop criteria around which tool should be used with what size drill bit. Create schedule for drilling holes to reduce tool fatigue
N3	?? Not controllable in hands of customer	Investigate how to be robust to pressure. Can rotational speed compensate?
X1*N1	*If the following interactions are large, then we have a robustness problem*	
X1*N2		
X1*N3		
X2*N1		
X2*N2		
X2*N3		
X3*N1		
X3*N2		
X3*N3		
No practical variation	Factor level setting not bold enough, measurement system unable to detect, or factors don't matter	
No factor effects significant		Noise has to be more thoroughly examined perhaps through directed sampling

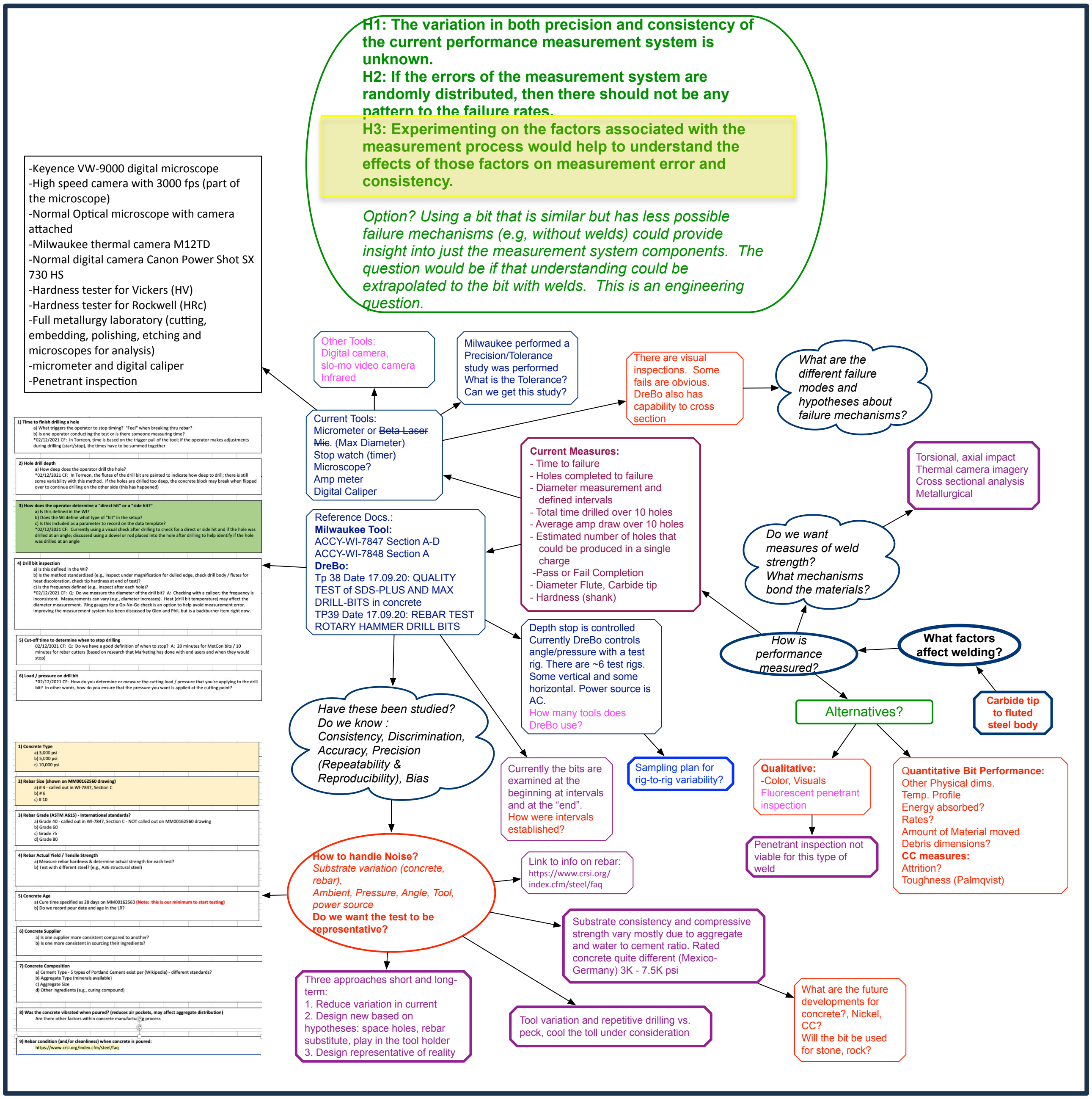
What are some possible unusual events that may occur during the experiment?

We discussed having several bits made at the beginning of each factor combination to ensure the equipment is performing consistently at the levels set. Additional bits for other Y's. One bit for cross-sectioning (assessing microstructure and ETA phase) and one for torsional test (torque).

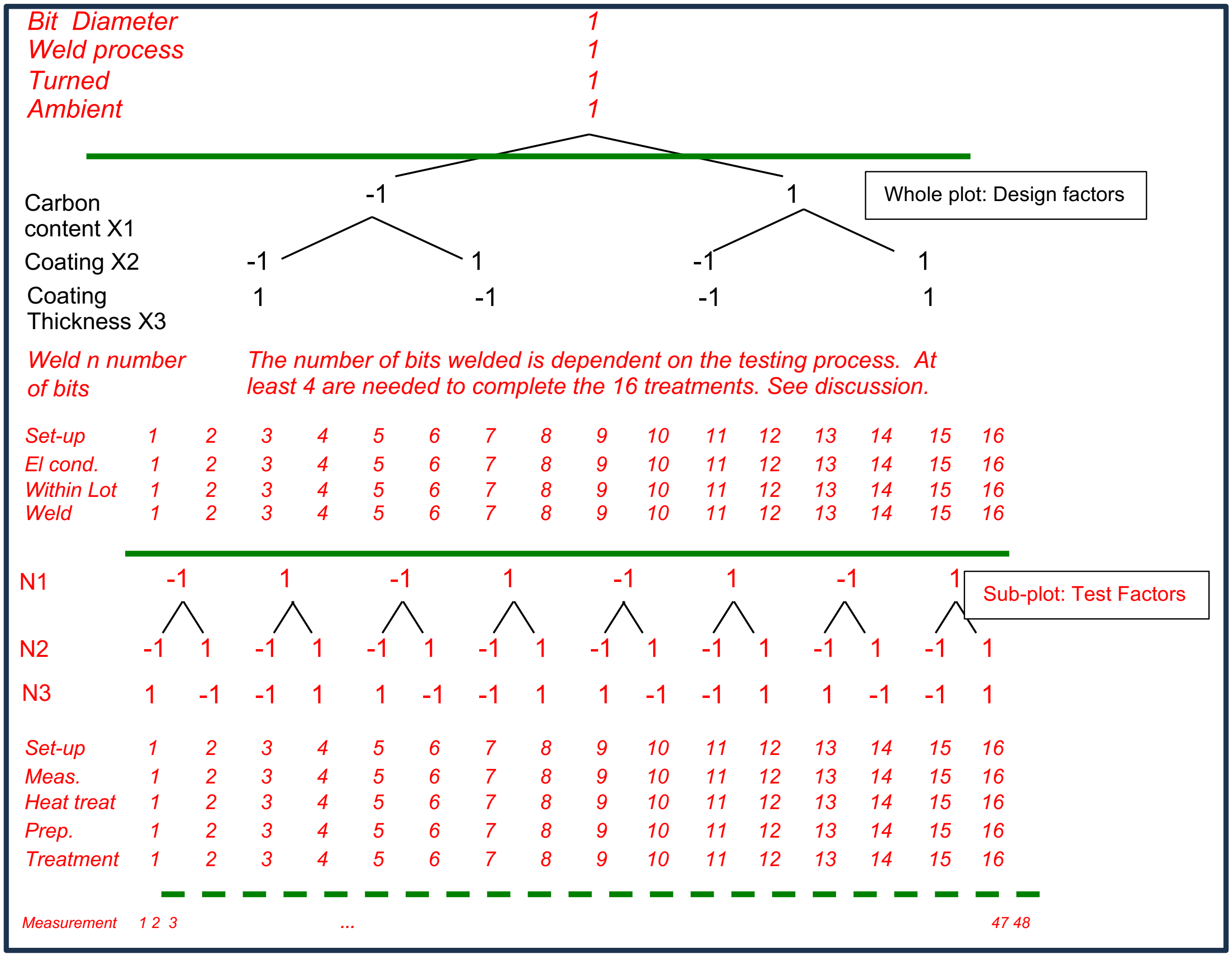
Thought Map Welding



Thought Map: Measure and Test



Option I: Split-plot design with carbide tip factors in the whole plot and test system factors in the sub-plot



Design Resolution/Resources	**Noise strategy**	**Pros/cons**
Res III 3 factors, 1 whole plot. Restriction on size of bit.	Use X-product array to separate and assign measurement noise. Repeated tests for confidence	*Pros:* Low resources, enough resolution given the nature of the factors. Better resolution for test factors *Cons:* Narrow Inference, only main effects.

Effect	Rank	Effect	Rank
Noise	4	Complex Non-linear	6
Main Effects	1	≥ 3rd Order Linear	6
Two-factor Interactions	5	Stability & Leverage	5
Noise-by-factor Interactions	4	Measurement Uncertainty	1
Simple Curvature	6	Mean/Variation	1

- Attempting to implement a possible solution (e.g., nickel instead of cobalt coating) without knowing the problem.
- Resolution of the 3 design factors is low. Main effects of those design factors are confounded with higher order effects.
- Since the measurement process will be experimented on, the measurement errors can somewhat be assigned and accounted for giving greater confidence in the assignment of the design factor effects.

- Lots of potential noise is not being assigned (e.g., powder lots, steel, electrode variation, factors in the heat treat process). This could be included if we choose to replicate this design (Complete block)
- Requires making 4 Batches of carbide tips (already made).

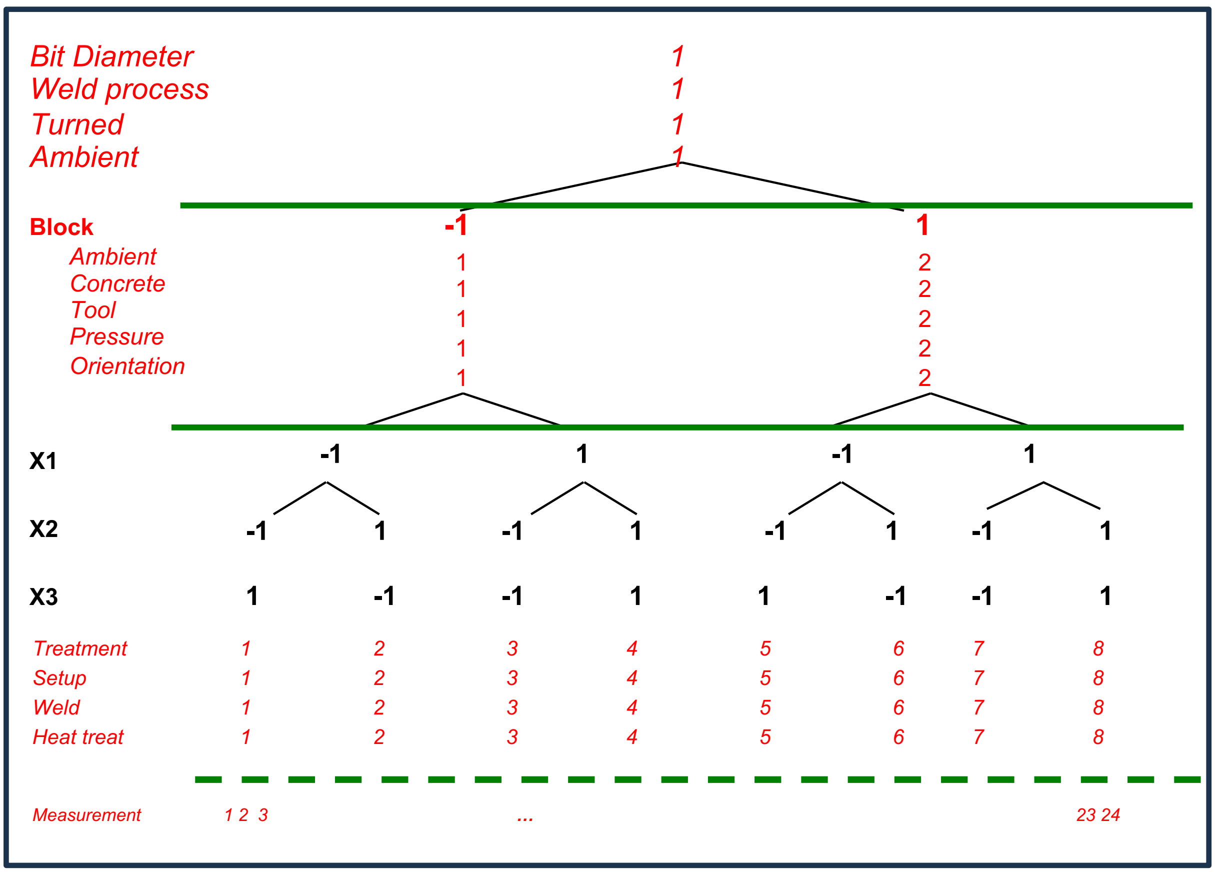
A minimum of 4 bits will need to be welded for each combination in the whole plot. Additional bits may be welded for additional Y's (X-section, torque). In addition, it may be useful to have 3 repeats for each treatment of the test process. In this case, we would need 2 X 4 for x-section and torque for each whole plot and 3 X 16 = 48 for each treatment of the experiment using the test process (drilling concrete holes). We will be able to estimate the following effects for both the average and variance of the repeated test bits:

WP=X1+X2+X3

SP= (N1+N2+N3)(1+WP)

Y= SP+WP

Option II: An alternative option is to confound the test system with a block.



This option confounds all of the factors in the test system. Other noise factors could also be confounded with the block. This allows the test system variation to be assigned to the block, hence improving the precision of the design. However specific assignment of the test factor effects is not possible.

Design Resolution/Resources	**Noise strategy**	**Pros/cons**
Res III 3 factors, Restriction on size of bit.	Use blocking to separate and assign measurement noise. Repeated tests for confidence	*Pros:* Least resources, enough resolution given the nature of the factors *Cons:* Narrow Inference, only main effects.

Effect	Rank	Effect	Rank
Noise	2	Complex Non-linear	6
Main Effects	1	≥ 3rd Order Linear	6
Two-factor Interactions	5	Stability & Leverage	5
Noise-by-factor Interactions	4	Measurement Uncertainty	2
Simple Curvature	6	Mean/Variation	1

- Attempting to implement a possible solution (e.g., nickel instead of cobalt coating) without knowing the problem.
- Resolution of the 3 design factors is low. Main effects of those design factors are confounded with higher order effects (interactions).
- Since the measurement process will be experimented on, the measurement errors can somewhat be assigned and accounted for giving greater confidence in the assignment of the design factor effects.
- Lots of potential noise is not being assigned (e.g., powder lots, steel, electrode variation, factors in the heat treat process). This could be included if we choose to replicate this design (Complete block)
- Requires making 4 Batches of carbide tips (already made).

A minimum of 3 bits would be needed, repeats for each treatment. Additional bits may be welded for additional Y's (X-section, torque). In this case, we would need 2 X 8 for x-section and torque for each treatment and 3 X 8 = 24 for each treatment of the experiment using the test process (drilling concrete holes).

We will be able to estimate the following effects for both the average and variance of the repeated test bits:

Y= X1+X2+X3+Block+Block(X1+X2+X3)

Appendix D
Aliasing Structure for Various Factorial Designs

These tables support design selection decisions; they do not replace the need for hypotheses, prediction, or consideration of inference space.

Designs

This section is designed to aid the experimenter in assigning two-level factors to fractional factorial and factorial experiment designs. These tables address the assignment of up to fifteen factors in up to sixteen run standard order designs and the associated aliasing. For a given number of factors, these designs will produce the greatest amount of information. The tables will identify main effects and two-way interactions appearing in each column of the design (and where appropriate higher order interactions are included, but in most cases the confounding pattern for higher order terms is not shown). The generator at the bottom of the page refers to the interaction that was confounded with the assigned treatment.

Resolution Table

	Number of Treatments	
# of Factors	**8**	**16**
3	V+	V+
4	IV	V+
5	III	V
6	III	IV
7	III	IV
8		IV
9-15		III

2^2

	X_1	X_2	X_3
1	-	-	+
2	+	-	-
3	-	+	-
4	+	+	+

2^3

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	-	-	-	+	+	+	-
2	+	-	-	-	-	+	+
3	-	+	-	-	+	-	+
4	+	+	-	+	-	-	-
5	-	-	+	+	-	-	+
6	+	-	+	-	+	-	-
7	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+

2^4

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
1	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+
2	+	-	-	-	+	-	+	+	-	-	-	-	+	+	+
3	-	+	-	-	+	+	-	+	-	-	+	+	-	-	+
4	+	+	-	-	-	+	+	-	+	+	-	-	-	-	+
5	-	-	+	-	+	+	+	-	-	+	-	+	-	+	-
6	+	-	+	-	-	+	-	+	+	-	+	-	-	+	-
7	-	+	+	-	-	-	+	+	+	-	-	+	+	-	-
8	+	+	+	-	+	-	-	-	-	+	+	-	+	-	-
9	-	-	-	+	-	+	+	+	-	+	+	-	+	-	-
10	+	-	-	+	+	+	-	-	+	-	-	+	+	-	-
11	-	+	-	+	+	-	+	-	+	-	+	-	-	+	-
12	+	+	-	+	-	-	-	+	-	+	-	+	-	+	-
13	-	-	+	+	+	-	-	+	+	+	-	-	-	-	+
14	+	-	+	+	-	-	+	-	-	-	+	+	-	-	+
15	-	+	+	+	-	+	-	-	-	-	-	-	+	+	+
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Three Factors in Eight Treatments

2_V^3

Column	X_1	X_2	X_3
Factor	A	B	C

Column	*Effects Estimated*
X_1	A
X_2	B
X_3	C
X_4	AB
X_5	AC
X_6	BC
X_7	ABC

Four Factors in Eight Treatments

2_{IV}^{4-1}

Column	X_1	X_2	X_3	X_7
Factor	A	B	C	D

Column	*Effects Estimated*
X_1	A
X_2	B
X_3	C
X_4	AB + CD
X_5	AC + BD
X_6	BC + AD
X_7	D

Generator D = ABC

Four Factors in Sixteen Treatments

2^4

Column	X_1	X_2	X_3	X_4
Factor	A	B	C	D

Column	*Effects Estimated*
X_1	A
X_2	B
X_3	C
X_4	D
X_5	ABC
X_6	BCD
X_7	ACD
X_8	ABD
X_9	ABCD
X_{10}	AB
X_{11}	AC
X_{12}	AD
X_{13}	BC
X_{14}	BD
X_{15}	CD

Five Factors in Eight Treatments

2^{5-2}_{III}

Column	X_1	X_2	X_3	X_4	X_5
Factor	A	B	C	D	E

Column	*Effects Estimated*
X_1	A + BD + CE
X_2	B + AD
X_3	C + AE
X_4	D + AB
X_5	E + AC
X_6	BC + DE
X_7	BE + CD

Generator D = AB, E = AC

Five Factors in Sixteen Treatments

$$2^{5-1}_{V}$$

Column	X_1	X_2	X_3	X_4	X_9
Factor	A	B	C	D	E

Column	*Effects Estimated*
X_1	A
X_2	B
X_3	C
X_4	D
X_5	DE
X_6	AE
X_7	BE
X_8	CE
X_9	E
X_{10}	AB
X_{11}	AC
X_{12}	AD
X_{13}	BC
X_{14}	BD
X_{15}	CD

Generator E = ABCD

Six Factors in Eight Treatments

$$2^{6-3}_{III}$$

Column	X_1	X_2	X_3	X_4	X_5	X_6
Factor	A	B	C	D	E	F

Column	*Effects Estimated*
X_1	A + BD + CE
X_2	B + AD + CF
X_3	C + AE + BF
X_4	D + AB + EF
X_5	E + AC + DF
X_6	F + BC + DE
X_7	AF + BE + CD

Generators: D = AB, E = AC, F = BC

Six Factors in Sixteen Treatments

2^{6-2}_{IV}

Column	X_1	X_2	X_3	X_4	X_5	X_6
Factor	A	B	C	D	E	F

Column	*Effects Estimated*
X_1	A
X_2	B
X_3	C
X_4	D
X_5	E
X_6	F
X_7	ACD
X_8	ABD
X_9	AF + DE
X_{10}	AB + CE
X_{11}	AC + BE
X_{12}	AD + EF
X_{13}	AE + BC + DF
X_{14}	BD + CF
X_{15}	CD + BF

Generator: E = ABC, F = BCD

Seven Factors in Eight Treatments

2^{7-4}_{III}

Column	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Factor	A	B	C	D	E	F	G

Column	*Effects Estimated*
X_1	A + BD + CE + FG
X_2	B + AD + CF + EG
X_3	C + AE + BF + DG
X_4	D + AB + EF + CG
X_5	E + AC + DF + BG
X_6	F + BC + DE + AG
X_7	G + AF + BE + CD

Generators: D = AB, E = AC, F = BC, G = ABC

Seven Factors in Sixteen Treatments

2^{7-3}_{IV}

Column	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Factor	A	B	C	D	E	F	G

Column	*Effects Estimated*
X_1	A
X_2	B
X_3	C
X_4	D
X_5	E
X_6	F
X_7	G
X_8	ABD
X_9	AF + BG + DE
X_{10}	AB + CE + FG
X_{11}	AC + BE + DG
X_{12}	AD + CG + EF
X_{13}	AE + BC + DF
X_{14}	BD + CF + EG
X_{15}	AG + BF + CD

Generators: E = ABC, F = BCD, G = ACD

Eight Factors in Sixteen Treatments

2^{8-4}_{IV}

Column	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Factor	A	B	C	D	E	F	G	H

Column	*Effects Estimated*
X_1	A
X_2	B
X_3	C
X_4	D
X_5	E
X_6	F
X_7	G
X_8	H
X_9	AF + BG + CH + DE
X_{10}	AB + CE + DH + FG
X_{11}	AC + BE + DG + FH
X_{12}	AD + BH + CG + EF
X_{13}	AE + BC + DF + GH
X_{14}	AH + BD + CF + EG
X_{15}	AG + CD + BF + EH

Generators: E = ABC, F = BCD, G = ACD, H = ABD

Nine Factors in Sixteen Treatments

2^{9-5}_{III}

Column	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Factor	A	B	C	D	E	F	G	H	I

Column	*Effects Estimated*
X_1	A + FI
X_2	B + GI
X_3	C + HI
X_4	D + EI
X_5	E + DI
X_6	F + AI
X_7	G + BI
X_8	H + CI
X_9	I + AF + BG + CH + DE
X_{10}	AB + CE + DH + FG
X_{11}	AC + BE + DG + FH
X_{12}	AD + BH + CG + EF
X_{13}	AE + BC + DF + GH
X_{14}	AH + BD + CF + EG
X_{15}	AG + CD + BF + EH

Generators: E = ABC, F = BCD, G = ACD, H = ABD, I = ABCD

Ten Factors in Sixteen Treatments

2^{10-6}_{III}

Column	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Factor	A	B	C	D	E	F	G	H	I	J

Column	*Effects Estimated*
X_1	A + FI + BJ
X_2	B + GI + AJ
X_3	C + HI + EJ
X_4	D + EI + HJ
X_5	E + DI + CJ
X_6	F + AI + GJ
X_7	G + BI + FJ
X_8	H + CI + DJ
X_9	I + AF + BG + CH + DE
X_{10}	J +AB + CE + DH + FG
X_{11}	AC + BE + DG + FH
X_{12}	AD + BH + CG + EF
X_{13}	AE + BC + DF + GH
X_{14}	AH + BD + CF + EG
X_{15}	AG + CD + BF + EH + IJ

Generators: E = ABC, F = BCD, G = ACD, H = ABD, I = ABCD, J = AB

Eleven Factors in Sixteen Treatments

2^{11-7}_{III}

Column	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
Factor	A	B	C	D	E	F	G	H	I	J	K

Column	*Effects Estimated*
X_1	A + FI + BJ + CK
X_2	B + GI + AJ + EK
X_3	C + HI + EJ + AK
X_4	D + EI + HJ + GK
X_5	E + DI + CJ + BK
X_6	F + AI + GJ + HK
X_7	G + BI + FJ + DK
X_8	H + CI + DJ + FK
X_9	I + AF + BG + CH + DE
X_{10}	J +AB + CE + DH + FG
X_{11}	K + AC + BE + DG + FH
X_{12}	AD + BH + CG + EF
X_{13}	AE + BC + DF + GH + JK
X_{14}	AH + BD + CF + EG + IK
X_{15}	AG + CD + BF + EH + IJ

Generators: E = ABC, F = BCD, G = ACD, H = ABD, I = ABCD, J = AB, K = AC

Twelve Factors in Sixteen Treatments

2^{12-8}_{III}

Column	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
Factor	A	B	C	D	E	F	G	H	I	J	K	L

Column	*Effects Estimated*
X_1	A + FI + BJ + CK + DL
X_2	B + GI + AJ + EK + HL
X_3	C + HI + EJ + AK + GL
X_4	D + EI + HJ + GK + AL
X_5	E + DI + CJ + BK + FL
X_6	F + AI + GJ + HK + EL
X_7	G + BI + FJ + DK + CL
X_8	H + CI + DJ + FK + BL
X_9	I + AF + BG + CH + DE
X_{10}	J +AB + CE + DH + FG
X_{11}	K + AC + BE + DG + FH
X_{12}	L + AD + BH + CG + EF
X_{13}	AE + BC + DF + GH + JK + IL
X_{14}	AH + BD + CF + EG + IK + JL
X_{15}	AG + CD + BF + EH + IJ + KL

Generators:

E = ABC, F = BCD, G = ACD, H = ABD, I = ABCD, J = AB, K = AC, L = AD

Thirteen Factors in Sixteen Treatments

2^{13-9}_{III}

Column	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
Factor	A	B	C	D	E	F	G	H	I	J	K	L	M

Column	*Effects Estimated*
X_1	A + FI + BJ + CK + DL + EM
X_2	B + GI + AJ + EK + HL + CM
X_3	C + HI + EJ + AK + GL + BM
X_4	D + EI + HJ + GK + AL + FM
X_5	E + DI + CJ + BK + FL + AM
X_6	F + AI + GJ + HK + EL + DM
X_7	G + BI + FJ + DK + CL + HM
X_8	H + CI + DJ + FK + BL + GM
X_9	I + AF + BG + CH + DE + LM
X_{10}	J +AB + CE + DH + FG + KM
X_{11}	K + AC + BE + DG + FH + JM
X_{12}	L + AD + BH + CG + EF + IM
X_{13}	M + AE + BC + DF + GH + JK + IL
X_{14}	AH + BD + CF + EG + IK + JL
X_{15}	AG + CD + BF + EH + IJ + KL

Generators:
E = ABC, F = BCD, G = ACD, H = ABD, I = ABCD, J = AB, K = AC, L = AD, M = AE

Fourteen Factors in Sixteen Treatments

2^{14-10}_{III}

Column	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
Factor	A	B	C	D	E	F	G	H	I	J	K	L	M	N

Column	*Effects Estimated*
X_1	A + FI + BJ + CK + DL + EM + HN
X_2	B + GI + AJ + EK + HL + CM + DN
X_3	C + HI + EJ + AK + GL + BM + FN
X_4	D + EI + HJ + GK + AL + FM + BN
X_5	E + DI + CJ + BK + FL + AM + GN
X_6	F + AI + GJ + HK + EL + DM + CN
X_7	G + BI + FJ + DK + CL + HM + EN
X_8	H + CI + DJ + FK + BL + GM + AN
X_9	I + AF + BG + CH + DE + LM + KN
X_{10}	J +AB + CE + DH + FG + KM + LN
X_{11}	K + AC + BE + DG + FH + JM + IN
X_{12}	L + AD + BH + CG + EF + IM + JN
X_{13}	M + AE + BC + DF + GH + JK + IL
X_{14}	N + AH + BD + CF + EG + IK + JL + IL
X_{15}	AG + CD + BF + EH + IJ + KL + MN

Generators:

E = ABC, F = BCD, G = ACD, H = ABD, I = ABCD, J = AB, K = AC, L = AD, M = AE, N = AH

Fifteen Factors in Sixteen Treatments

$$2^{15-11}_{III}$$

Column	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
Factor	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O

Column	*Effects Estimated*
X_1	A + FI + BJ + CK + DL + EM + HN + GO
X_2	B + GI + AJ + EK + HL + CM + DN + FO
X_3	C + HI + EJ + AK + GL + BM + FN + DO
X_4	D + EI + HJ + GK + AL + FM + BN + CO
X_5	E + DI + CJ + BK + FL + AM + GN + HO
X_6	F + AI + GJ + HK + EL + DM + CN + BO
X_7	G + BI + FJ + DK + CL + HM + EN + AO
X_8	H + CI + DJ + FK + BL + GM + AN + EO
X_9	I + AF + BG + CH + DE + LM + KN + JO
X_{10}	J +AB + CE + DH + FG + KM + LN + IO
X_{11}	K + AC + BE + DG + FH + JM + IN + LO
X_{12}	L + AD + BH + CG + EF + IM + JN + KO
X_{13}	M + AE + BC + DF + GH + JK + IL + NO
X_{14}	N + AH + BD + CF + EG + IK + JL + JL + MO
X_{15}	O + AG + CD + BF + EH + IJ + KL + MN

Generators:

E = ABC, F = BCD, G = ACD, H = ABD, I = ABCD, J = AB, K = AC, L = AD, M = AE, N = AH, O = AG

Blocking

Suggested generators for 2 blocks:

Design	Generator for Block	Resolution
2^{3}	ABC	IV
2^{4-1}	AB	III
2^{4}	ABCD	V
2^{5-2}	BC	III
2^{5-1}	BCD	IV
2^{6-3}	ABC (D = AB, E = AC, F = BC)	III
2^{6-2}	ACD	IV
2^{7-3}	ABD	IV
2^{8-4}	ABCD (E = ABC, F = BCD, G = ACD, H = ABD)	III
2^{9-5}	AB	III
2^{10-6}	AC	III
2^{11-7}	AD	III
2^{12-8}	AE	III
2^{13-9}	AH	III
2^{14-10}	AG	III

Three Factor Box-Behnken Design

Run	X_1	X_2	X_3
1	+	-	0
2	-	-	0
3	-	+	0
4	0	+	+
5	0	0	0
6	0	-	-
7	-	0	+
8	+	0	+
9	0	-	+
10	+	0	-
11	0	0	0
12	-	0	-
13	+	+	-
14	0	+	-
15	0	0	0

Appendix E
Factor Relationship Diagrams

This appendix is a collection of Factor Relationship Diagrams (FRD's). It illustrates how multiple experimental strategies can be developed for the same situation. The objective is not to identify a single "best" design, but to show how different designs balance resolution, noise handling, inference space, and resource requirements.

Each strategy reflects different assumptions about knowledge level, hypotheses, and business constraints. Each design structure provides information about different aspects of the system. The choice of structure determines what questions the data will be able to answer.

Factor Relationship Diagrams:
This appendix will describe multiple potential initial experiment strategies for an injection molding process. These examples are not decision rules. The appropriate structure depends on the specific hypotheses, knowledge level, noise structure, and business constraints. The purpose of these examples is to illustrate tradeoffs, not to prescribe choices.

Plastic parts are made in a multi-cavity mold. In all scenarios, there are 5 design factors to be tested (F1-F5).

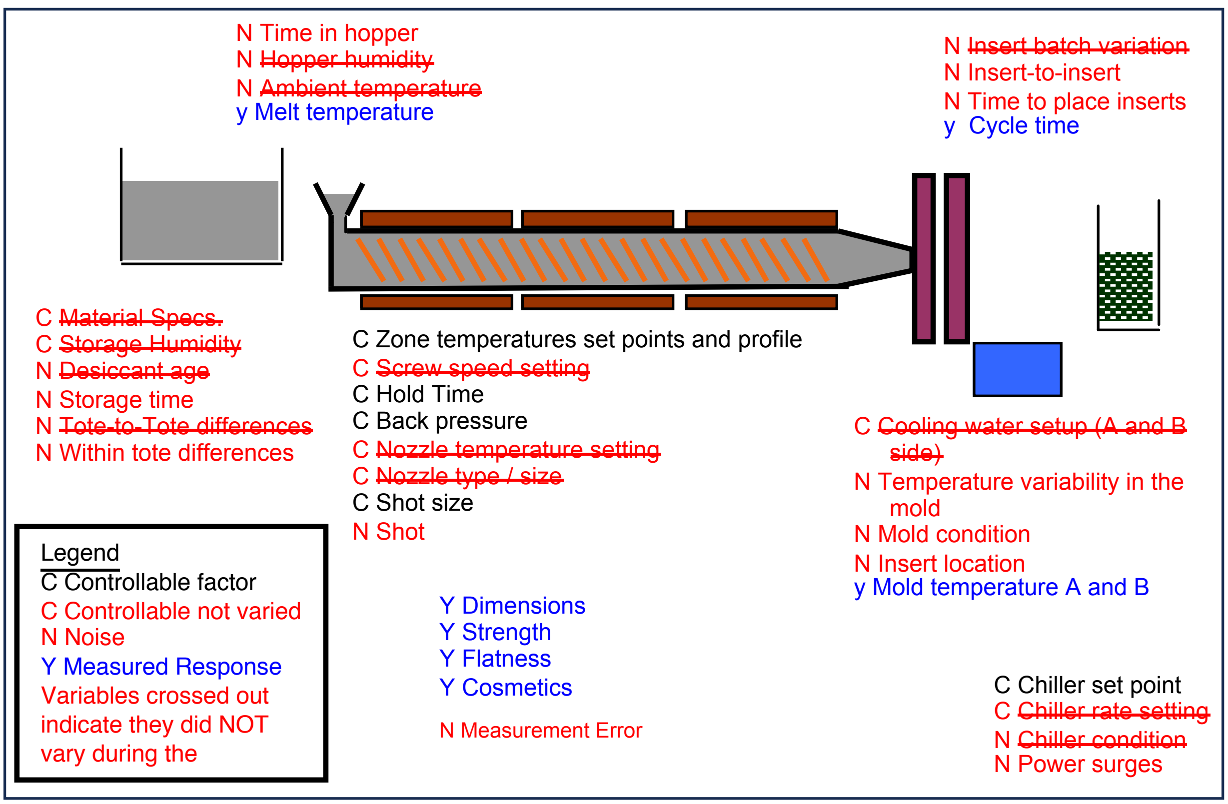


Figure E.1: Process Map for Injection Molding Experimentation

Prior to Selecting an Experimental Plan

The following elements should be considered prior to selecting the design structure:

- ✓ Document the motivation.
- ✓ Specify goals of the experiment. List questions are you trying to answer.
- ✓ Identify response variables, always consider multiple responses. Understand measurement variation of the response, or consider nesting it in the experiment.
- ✓ Identify practical significance of response variables
- ✓ Quantify time and resources available. Understand how the business goals relate to experimental goals. Ask, is this reasonable to run?
- ✓ Identify the team of knowledgeable individuals familiar with the process and situation. Maintain open discussion and dialogue of your thoughts and plans. Make sure to include their input/knowledge.
- ✓ Develop (or gather) necessary Thought maps and Process/product maps to provide context. This should help identify the noise present within the process.
- ✓ Generate hypotheses: These will link to the factors that are varied through the experiment.
- ✓ Identify factors, and predict the rank order of the model effects up to 2^{nd} order.
- ✓ Provide numerical predictions for each Y.
- ✓ Set your factor levels (BOLD, but reasonable).
- ✓ Create multiple FRD's with pros/cons associated with each one.
- ✓ Predict and anticipate every possible outcome and subsequent actions to take.
- ✓ Select the FRD that best balances cost vs. benefits.

Various Initial FRD's Linked to Ranking of Possible Effects

Possible Effect	***Rank***
Noise	3
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	3
Noise-by-factor Interactions	5
Simple Curvature (2nd order non-linear)	5
Complex Non-linear (≥3rd order non-linear)	6
≥ Three-Factor Interactions (3rd order linear)	6
Stability	5
Leverage	5
Measurement Uncertainty	1
Mean	1
Variation	1

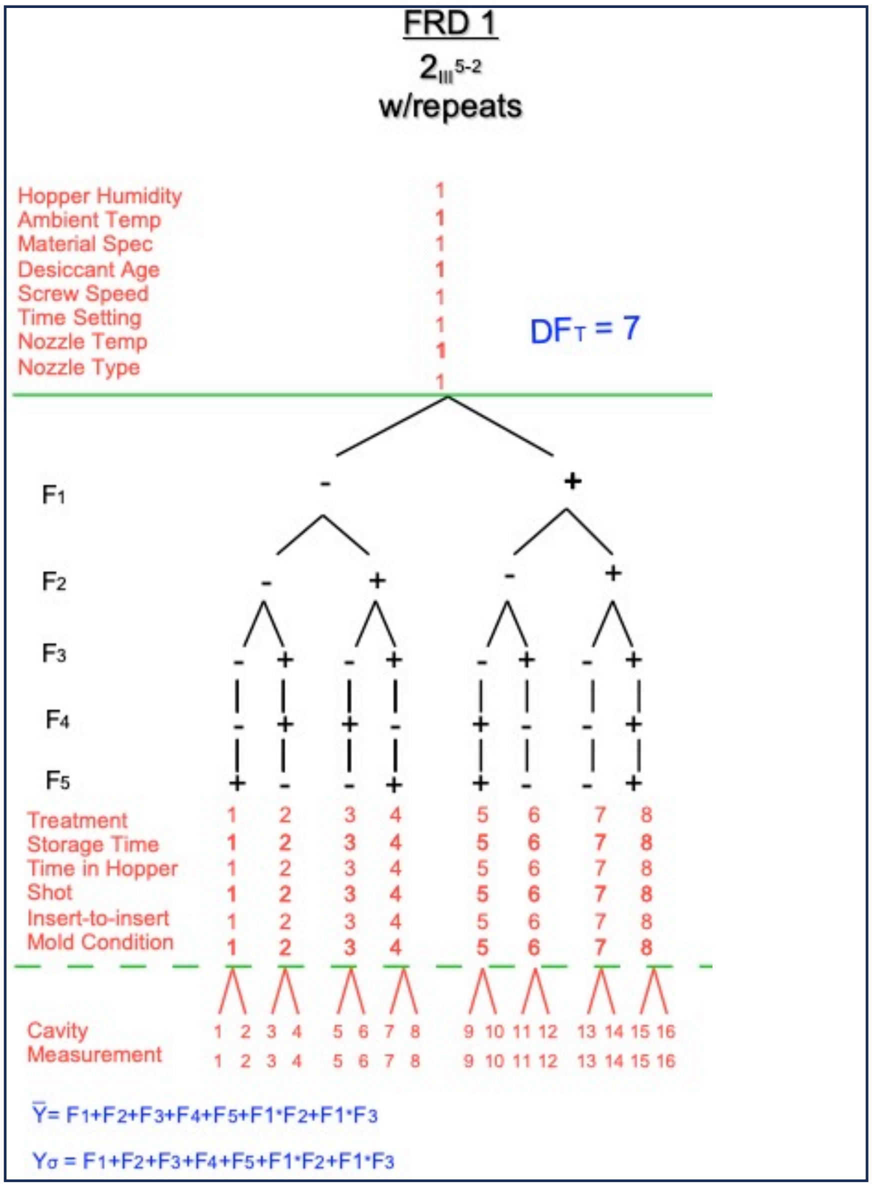


Figure E.2: Fractional Factorial with Repeats

Design Resolution/Resources	Noise strategy	Pros/cons
Res III. 5 factors, 8 treatments, 16 parts measured once	Repeats used to partition part-to-part and measurement error from treatments	*Pros:* Measurement error & part-to-part variation partitioned increasing Design Structure (DS) precision. Y of variation could be analyzed. Data available if one is lost. *Cons:* Low resolution. Narrow Inference Space (IS)

This approach emphasizes:

- Low on the knowledge continuum, main effects are prioritized
- Conserve resources
- Concern for short-term noise effects (e.g., measurement error, within part/batch and part-to-part, batch-to-batch variation)
- Determine factor effects on both the mean and variation

Possible Effect	*Rank*
Noise	1
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	3
Noise-by-factor Interactions	1
Simple Curvature (2nd order non-linear)	6
Complex Non-linear (≥3rd order non-linear)	6
≥ Three-Factor Interactions (3rd order linear)	6
Stability	5
Leverage	5
Measurement Uncertainty	4
Mean	3
Variation	3

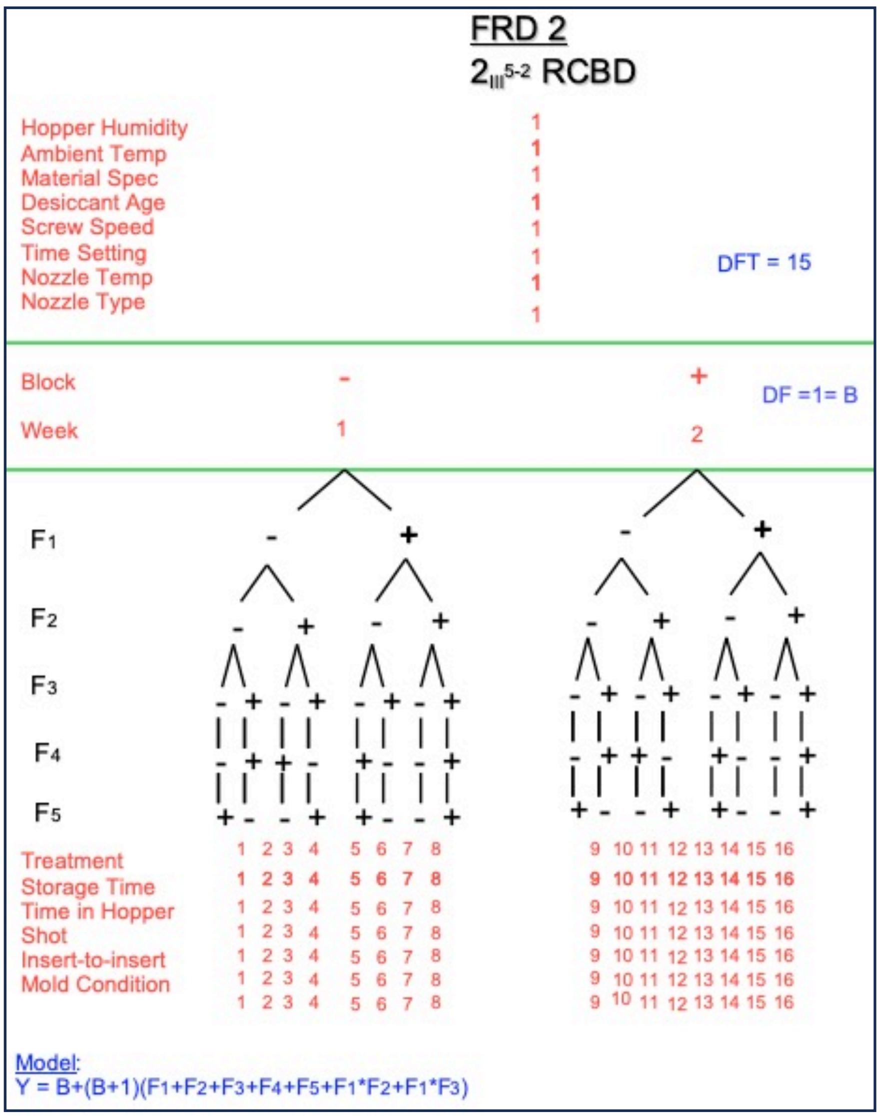


Figure E.3: Fractional Factorial, RCBD

Design Resolution/Resources	Noise strategy	Pros/cons
Res III. 5 factors, 16 treatments, 16 shots, 2 blocks. Full resolution on block effects.	Complete blocks to partition noise and estimate noise-by-factor interactions.	*Pros:* Wider IS with increased DS precision. All block-by-factor interactions estimable. *Cons:* 16 shots of material. Low DS resolution

This approach emphasizes:

- Low on the knowledge continuum, main effects are prioritized
- Concern for noise and factor effect robustness
- Have hypotheses on effect of noise
- Determine factor effects and their robustness to noise
- Early in new design development

Possible Effect	*Rank*
Noise	2
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	3
Noise-by-factor Interactions	4
Simple Curvature (2nd order non-linear)	6
Complex Non-linear (≥3rd order non-linear)	6
≥ Three-Factor Interactions (3rd order linear)	6
Stability	5
Leverage	5
Measurement Uncertainty	4
Mean	3
Variation	3

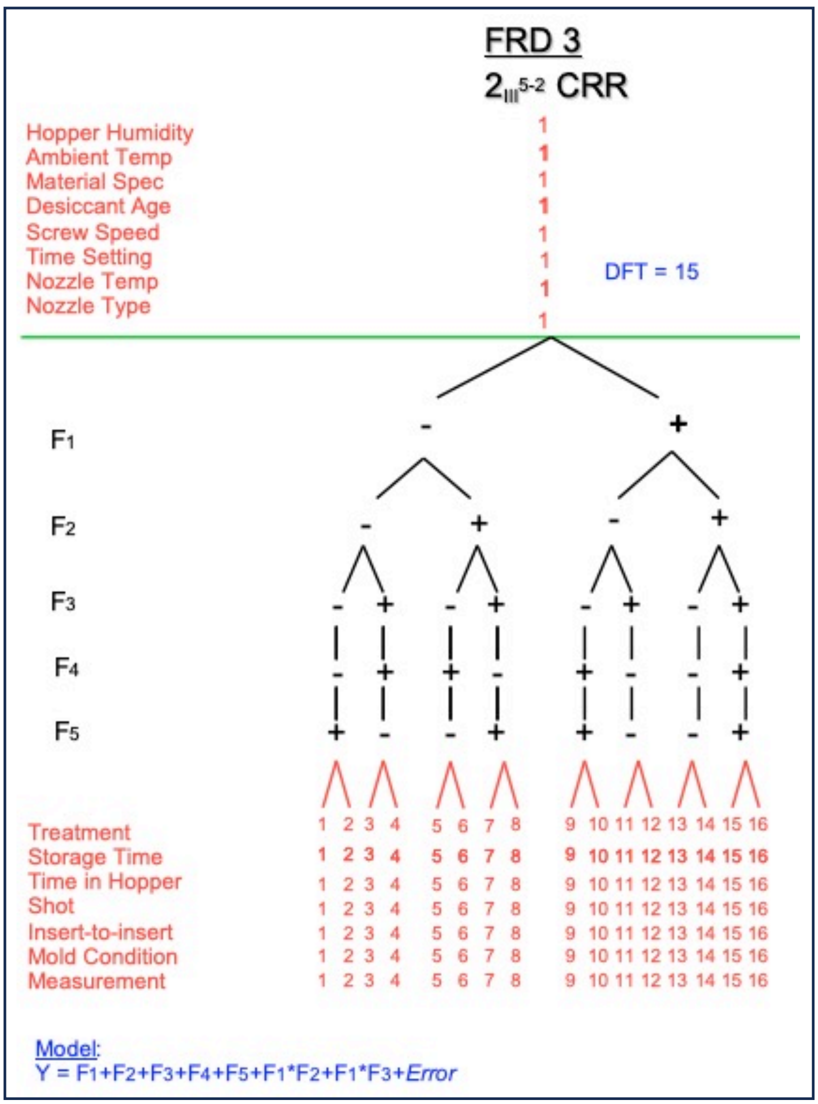


Figure E.4: Fractional Factorial CRR

Design Resolution/Resources	**Noise strategy**	**Pros/cons**
Res III. 5 factors, 16 treatments, 16 shots unassignable, unbiased noise estimate	Theoretically less biased estimate of noise used as a basis for statistical tests.	*Pros:* Wider IS. Better F-test. *Cons:* 16 shots of material. Low resolution for factor effects. Less DS precision

This approach emphasizes:

- Low on the knowledge continuum, main effects are prioritized
- Concern for noise, and the noise has not been identified
- Recognition of noise, but no hypotheses on the effect of noise
- Determine factor effects with a statistical test against random noise effects
- Have available unassigned DF for possible covariates

Possible Effect	***Rank***
Noise	3
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	2
Noise-by-factor Interactions	4
Simple Curvature (2nd order non-linear)	6
Complex Non-linear (≥3rd order non-linear)	6
≥ Three-Factor Interactions (3rd order linear)	6
Stability	5
Leverage	5
Measurement Uncertainty	4
Mean	4
Variation	4

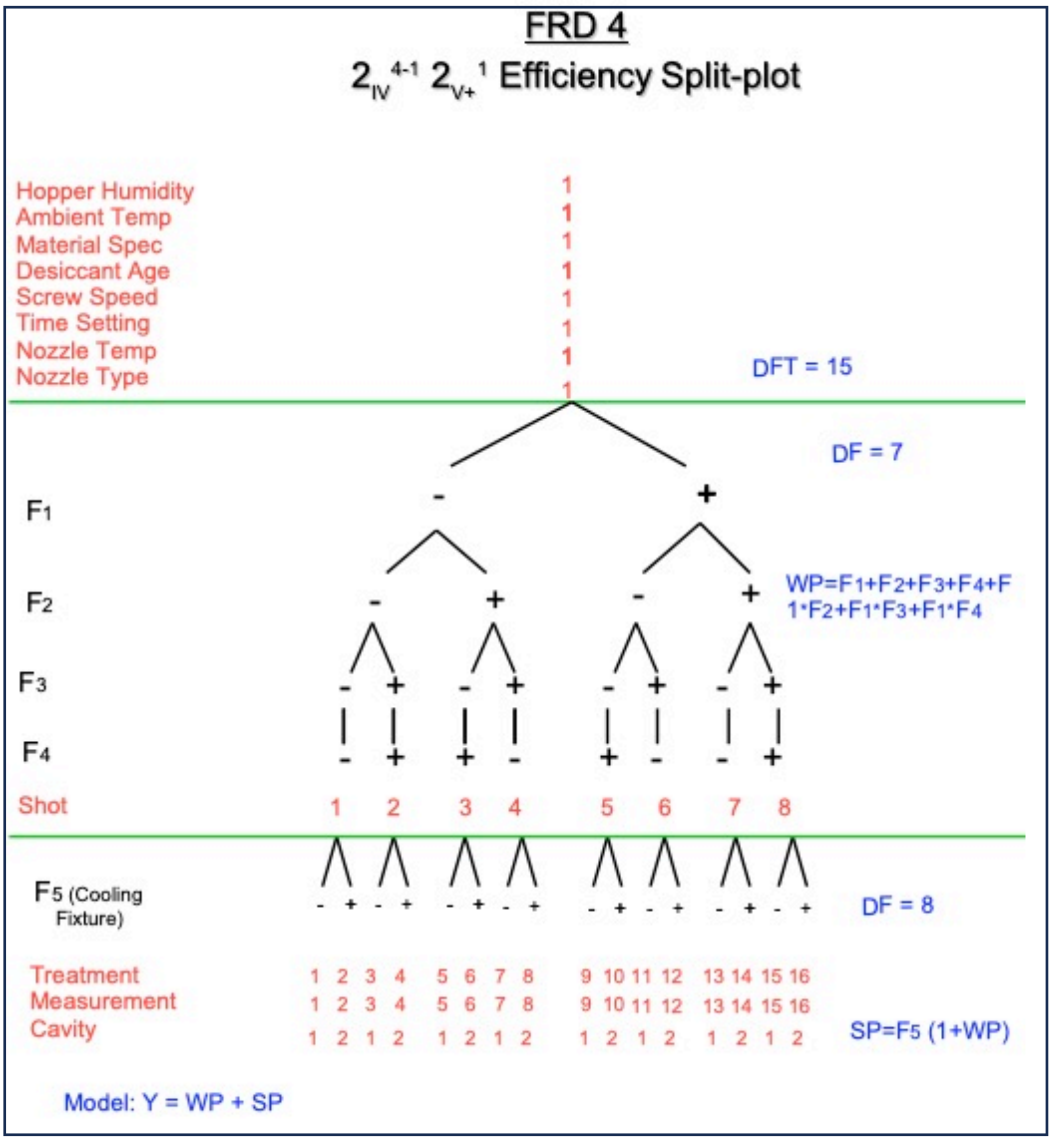


Figure E.5: Efficiency Split-plot

Design Resolution/Resources	Noise strategy	Pros/cons
Res IV Whole Plot (WP), full res on Sub-plot (SP). 5 factors, 16 treatments, 8 shots.	Noise is partitioned by split.	*Pros:* Higher resolution. WP and SP have higher precision. Low cost (8 shots) *Cons:* No assignment of noise effects.

This approach emphasizes:

- Very efficient use of resources
- Desire to separate noise for whole plot from subplot
- Increase precision of WP and SP without increasing resources
- Usually, higher resolution

Possible Effect	*Rank*
Noise	1
Main Effects (Factors)	1
Two-Factor Interactions (2^{nd} order linear)	3
Noise-by-factor Interactions	3
Simple Curvature (2^{nd} order non-linear)	6
Complex Non-linear ($\geq 3^{rd}$ order non-linear)	6
$\geq$ Three-Factor Interactions (3^{rd} order linear)	6
Stability	5
Leverage	5
Measurement Uncertainty	4
Mean	3
Variation	3

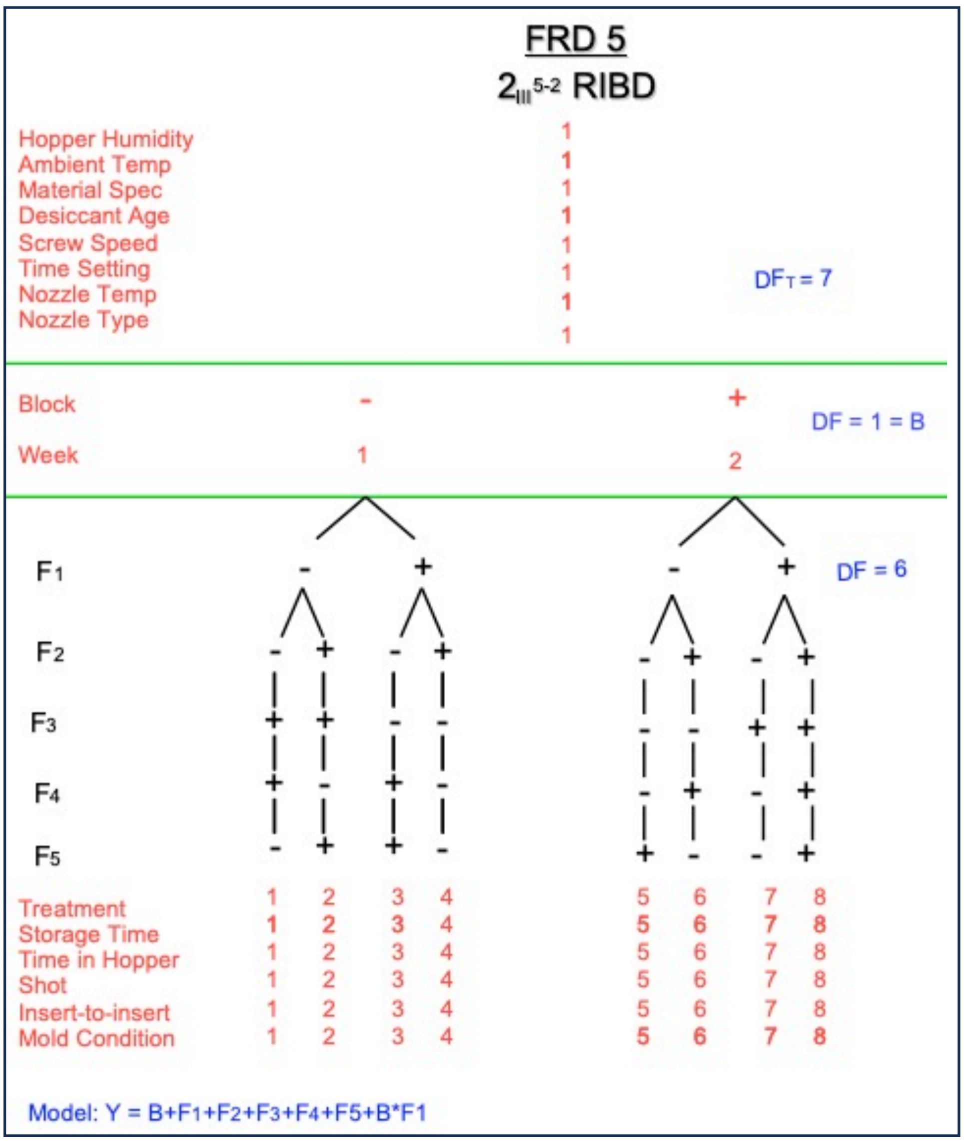


Figure E.6: Fractional Factorial, RIBD

Design Resolution/Resources	Noise strategy	Pros/cons
Res III. 5 factors, 8 treatments, 8 shots, 2 incomplete blocks.	Incomplete blocks to partition noise.	*Pros:* Wider IS with increased precision. Low resources. *Cons:* Low resolution for factor & block effects

This approach emphasizes:

- Low on the knowledge continuum, main effects are prioritized
- Have hypotheses about the main effect of noise
- Looking for clues about the block effect for further iterations (disaggregation)
- Post design applications

Possible Effect	***Rank***
Noise	4
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	1
Noise-by-factor Interactions	5
Simple Curvature (2nd order non-linear)	6
Complex Non-linear (≥3rd order non-linear)	6
≥ Three-Factor Interactions (3rd order linear)	6
Stability	5
Leverage	5
Measurement Uncertainty	4
Mean	3
Variation	3

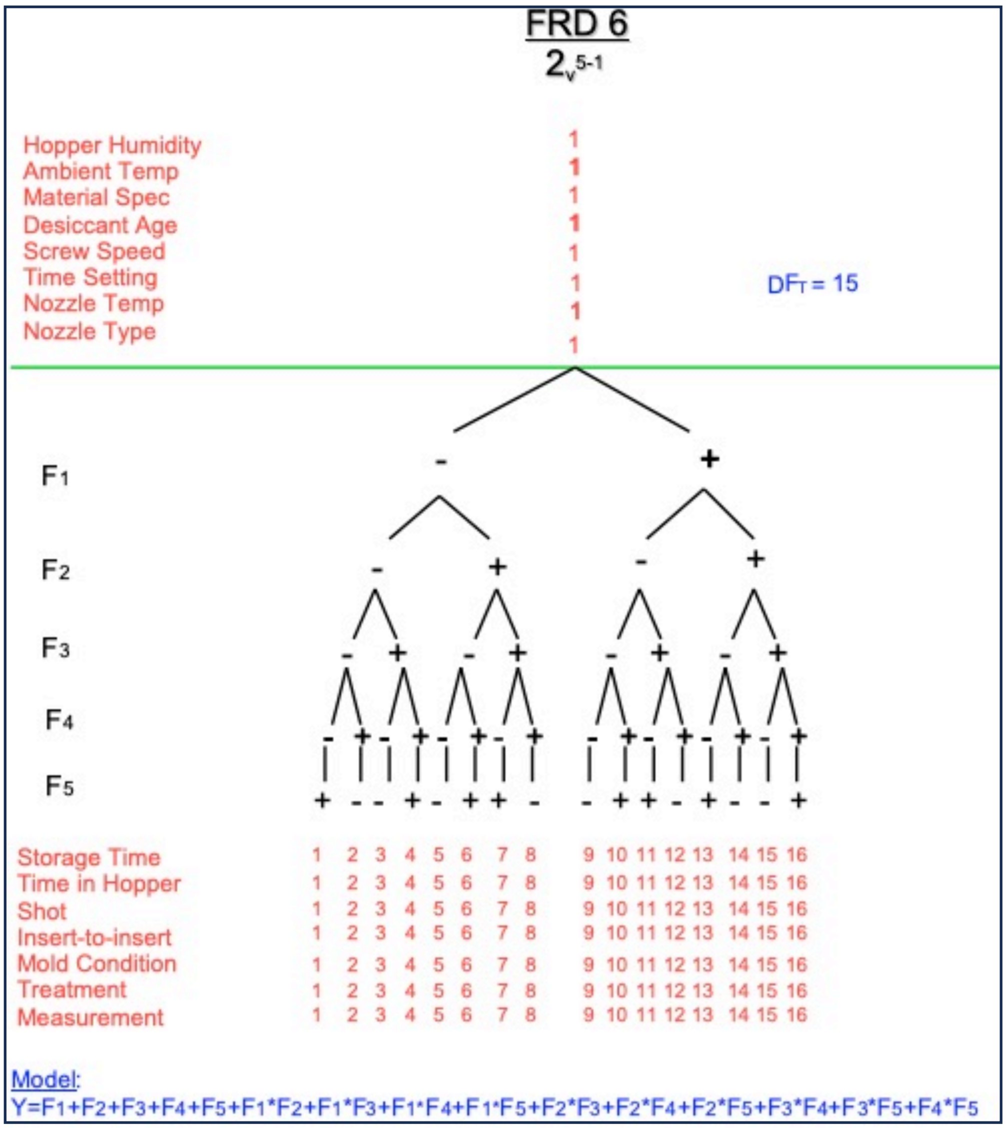


Figure E.7: Higher Resolution Fractional Factorial

Design Resolution/Resources	Noise strategy	Pros/cons
Res V. 5 factors, 16 treatments, 16 shots.	No specific strategy to partition or estimate noise	*Pros:* Res V design, therefore all 1st and 2nd order DS effects are estimable. *Cons:* 16 shots of material. **No noise strategy**. Narrow IS.

This approach emphasizes:

- Theoretically higher on the knowledge continuum, 1st and 2nd order effects separated
- No concern for noise or factor effect robustness

Possible Effect	*Rank*
Noise	3
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	1
Noise-by-factor Interactions	5
Simple Curvature (2nd order non-linear)	6
Complex Non-linear (≥3rd order non-linear)	6
≥ Three-Factor Interactions (3rd order linear)	6
Stability	5
Leverage	5
Measurement Uncertainty	1
Mean	1
Variation	1

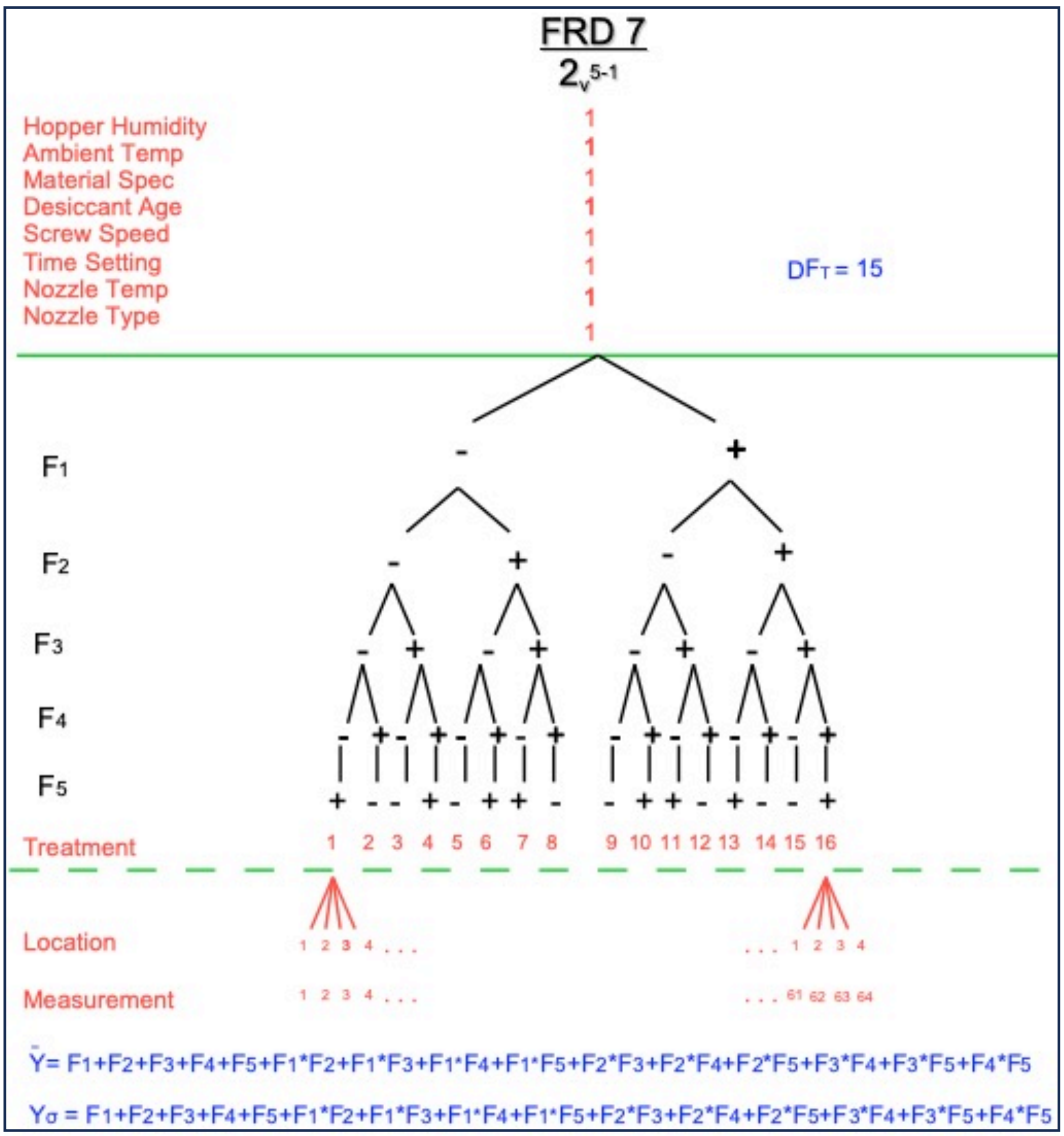


Figure E.8: High Resolution Fractional Factorial with Repeats

Design Resolution/Resources	Noise strategy	Pros/cons
Res III. 5 factors, 16 treatments, 16 shots. Multiple measurements	Repeats to partition the noise systematically within treatment. Ability to detect locational patterns.	*Pros:* Measurement error & within part variation partitioned increasing DS precision. Y of variation could be analyzed. Higher res. for model effects *Cons:* 16 shots of material.

This approach emphasizes:

- Higher on the knowledge continuum, 1st and 2nd order effects separated
- Concern for noise systematically within part and measurement errors
- Have hypotheses on factor effects and within part variation (systematic)

Possible Effect	*Rank*
Noise	3
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	1
Noise-by-factor Interactions	4
Simple Curvature (2nd order non-linear)	6
Complex Non-linear (≥3rd order non-linear)	6
≥ Three-Factor Interactions (3rd order linear)	6
Stability	5
Leverage	5
Measurement Uncertainty	4
Mean	3
Variation	3

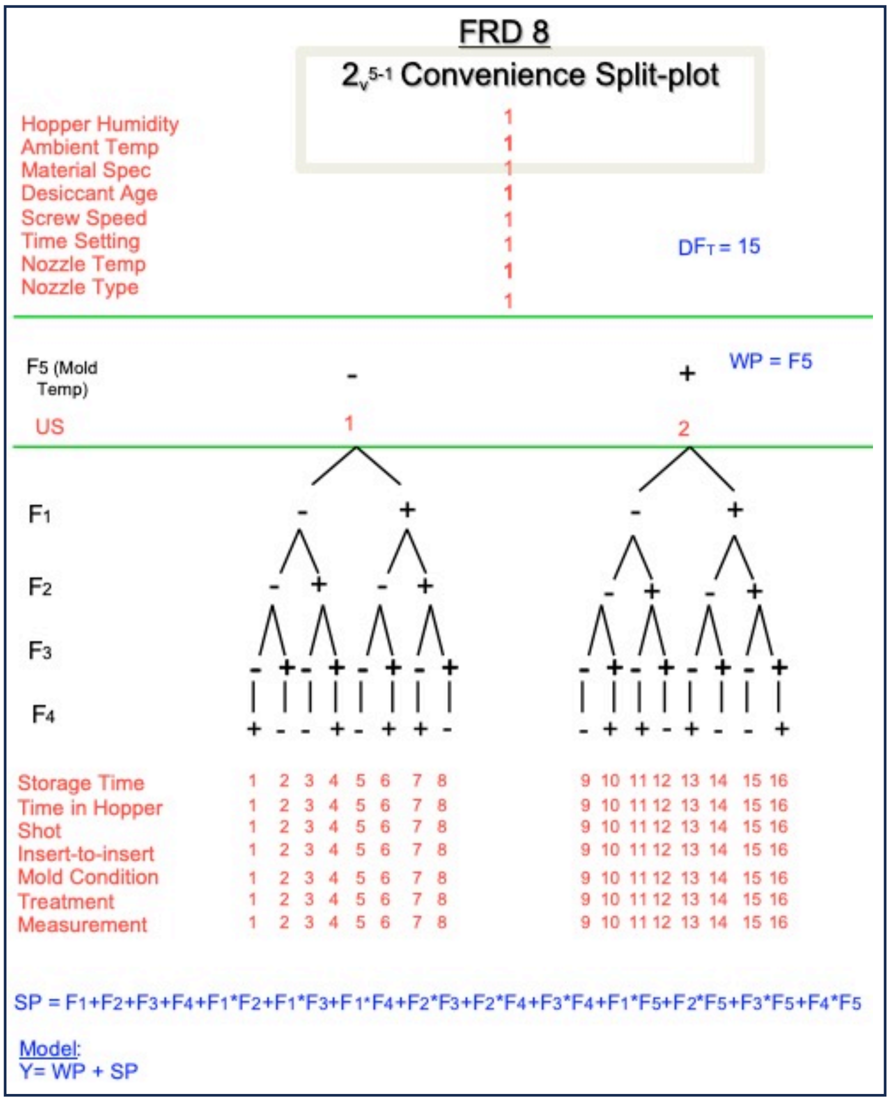


Figure E.9: Convenience Split-plot

Design Resolution/Resources	Noise strategy	Pros/cons
Res III (WP) - V (SP). 5 factors, 16 treatments, 16 shots.	Noise partitioned by split.	*Pros:* SP has increased precision. More convenient to run. *Cons:* 16 shots of material. WP precision is compromised

This approach emphasizes:

- Convenience and resources drive this plan
- Noise is partitioned increasing precision of the SP, but WP precision is compromised
- Increased resolution of the SP

Possible Effect	***Rank***
Noise	1
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	1
Noise-by-factor Interactions	1
Simple Curvature (2nd order non-linear)	6
Complex Non-linear (≥3rd order non-linear)	6
≥ Three-Factor Interactions (3rd order linear)	6
Stability	5
Leverage	5
Measurement Uncertainty	4
Mean	3
Variation	3

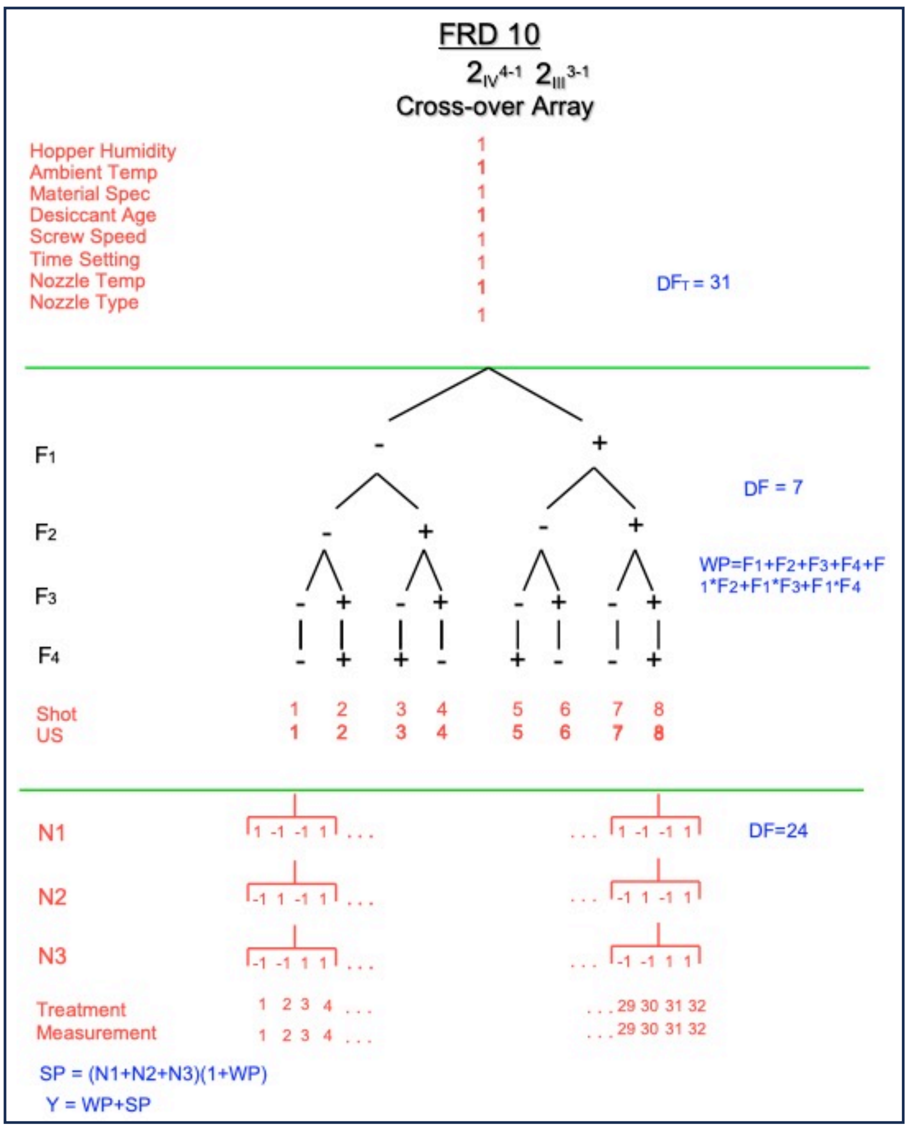


Figure E.10: Cross-Over Array

Design Resolution/Resources	Noise strategy	Pros/cons
Res IV. 5 factors, 32 treatments, 8 shots. Res IV on noise factor effects	Main effects of noise and noise by factor interactions estimable. Splits partition the noise increasing the precision of both the WP and SP	*Pros:* Res Wider IS with increased precision for DS. All noise-by-factor interactions estimable. Low resources *Cons:* Challenge of manipulating noise

This approach emphasizes:

- Specific noise factors identified
- Concern for noise and factor effect robustness
- Have hypotheses on effect of noise and noise-by-factor interactions
- Determine factor effects and their robustness to noise
- Less efficient but highly effective for robust design

Possible Effect	***Rank***
Noise	1
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	4
Noise-by-factor Interactions	3
Simple Curvature (2nd order non-linear)	6
Complex Non-linear (≥3rd order non-linear)	6
≥ Three-Factor Interactions (3rd order linear)	6
Stability	5
Leverage	5
Measurement Uncertainty	4
Mean	3
Variation	3

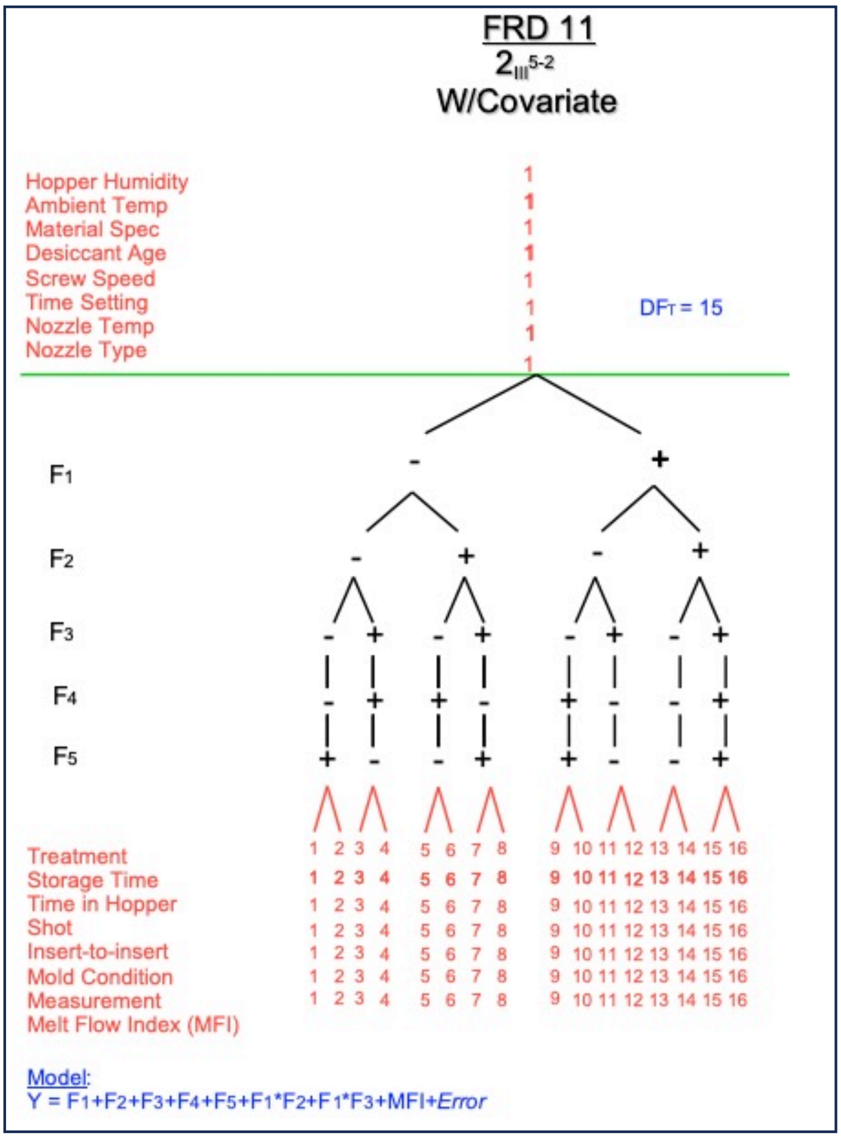


Figure E.11: Fractional Factorial with Covariate

Design Resolution/Resources	Noise strategy	Pros/cons
Res III. 5 factors, 16 treatments, 16 shots	A measured noise variable (random effect) can be estimated	*Pros:* Covariate estimable, increased DS precision *Cons:* 16 shots of material. Low resolution for DS

This approach emphasizes:

- Low on the knowledge continuum, main effects are prioritized
- Concern for a noise variable that is measurable
- Have hypotheses on effect of a specific input noise variable
- Determine factor effects and a model that takes the noise variable into account

Possible Effect	***Rank***
Noise	3
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	1
Noise-by-factor Interactions	3
Simple Curvature (2nd order non-linear)	6
Complex Non-linear (≥3rd order non-linear)	6
≥ Three-Factor Interactions (3rd order linear)	6
Stability	5
Leverage	5
Measurement Uncertainty	4
Mean	3
Variation	3

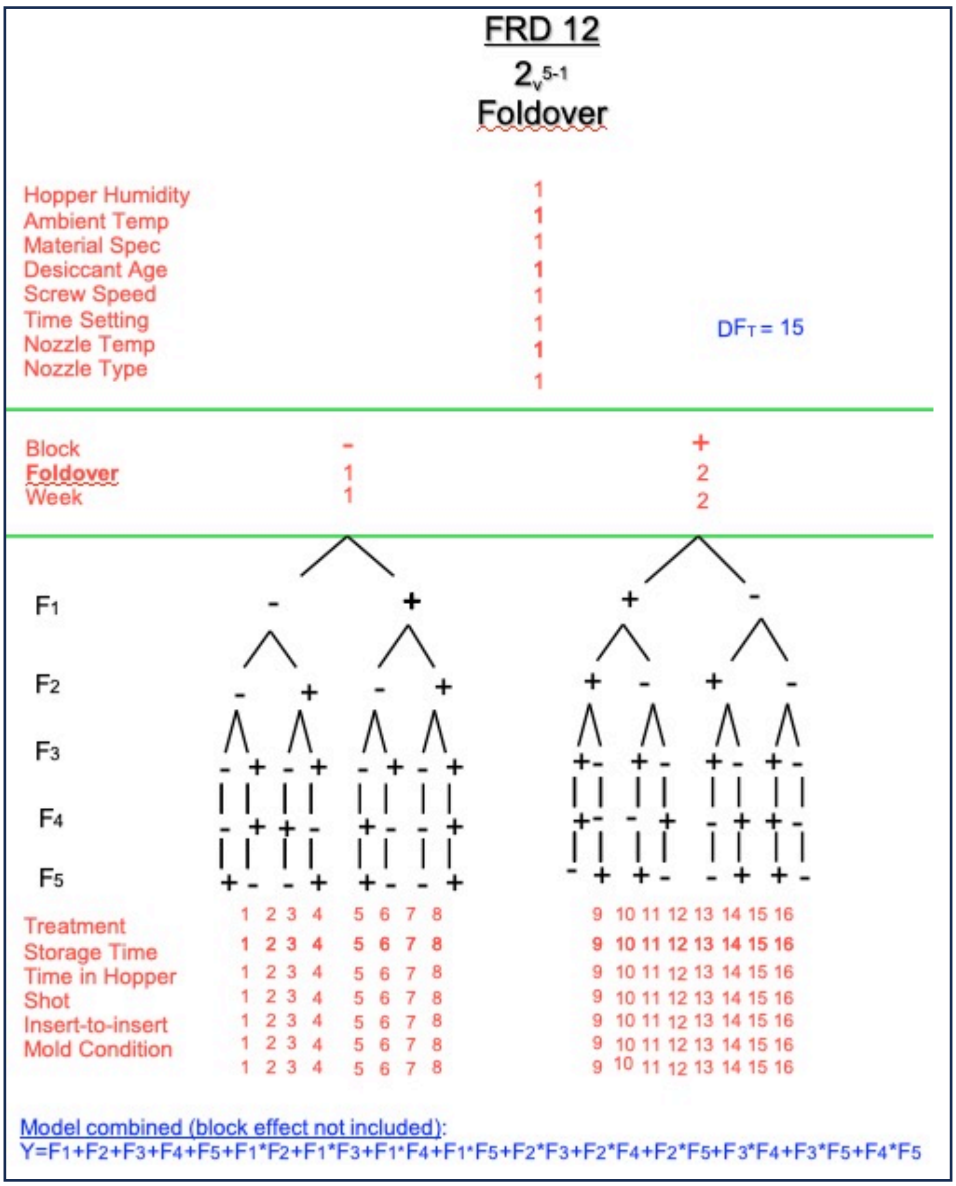


Figure E.12: Fold Over Design

Design Resolution/Resources	Noise strategy	Pros/cons
Res IV. 5 factors, 16 treatments, 16 shots, 2 blocks, but the fold over is done to increase the resolution of the factor effects	While there are complete blocks, their effects are considered negligible.	*Pros:* Increased res. of DS. Wider IS *Cons:* 16 shots of material. Block effects are confounded with fold. (higher resolution)

This approach emphasizes:

- After the first replicate is run, there is concern for confounding of 2^{nd} order effects
- Considered to be running in a reasonable design space
- No concerns for noise (block effect is considered negligible)
- Levels chosen for the design space are near optimum

Possible Effect	***Rank***
Noise	3
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	3
Noise-by-factor Interactions	3
Simple Curvature (2nd order non-linear)	1
Complex Non-linear (≥3rd order non-linear)	5
≥ Three-Factor Interactions (3rd order linear)	4
Stability	2
Leverage	4
Measurement Uncertainty	5
Mean	5
Variation	5

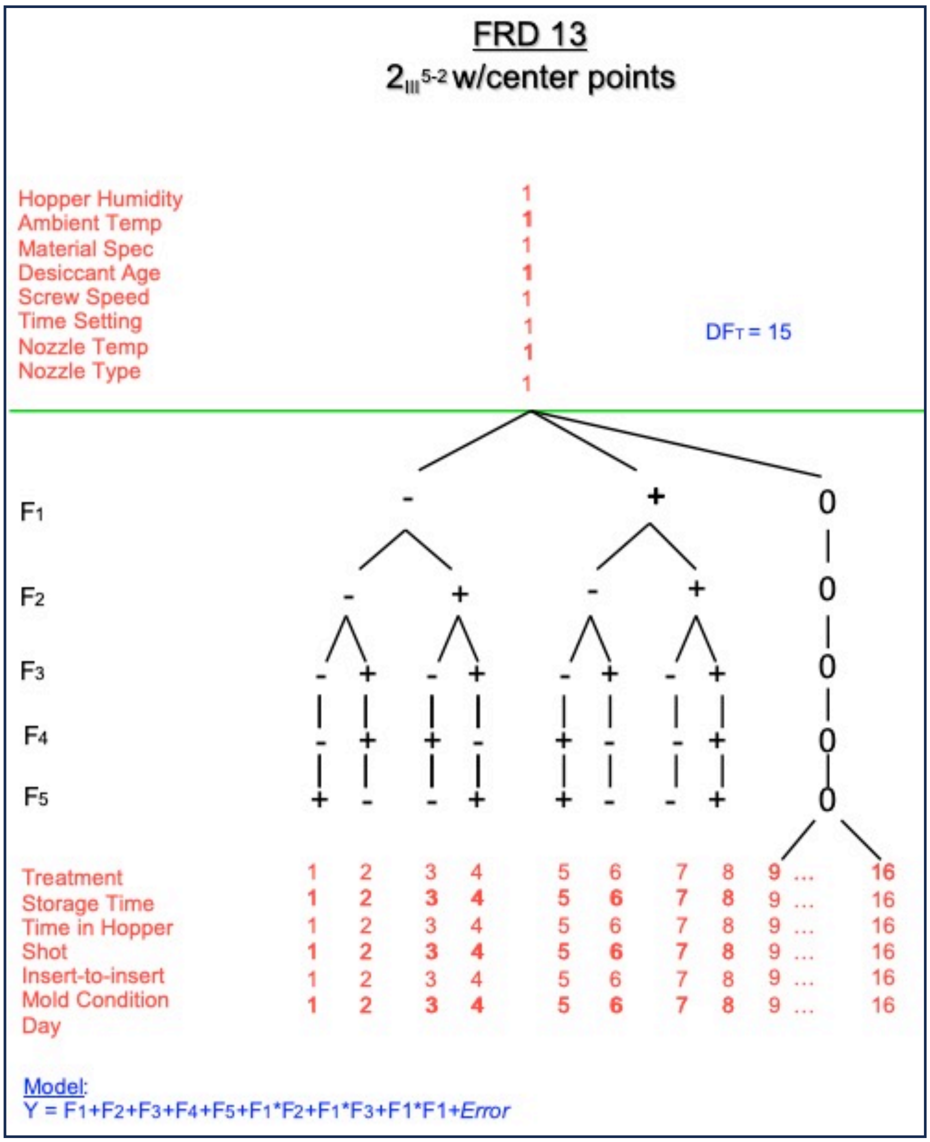


Figure E.13: Center Points

Design Resolution/Resources	Noise strategy	Pros/cons
Res III. 5 factors, 16 treatments, 16 shots. CP added to test linear assumption and estimate consistency in the design space	Noise can be estimated from the replicated CP's.	*Pros:* Wider IS. Test linear assumptions and consistency in design space *Cons:* 16 shots of material. Low resolution for factor effects. Center Point (CP) is not specific to the non-linear effects

This approach emphasizes:

- Medium on the knowledge continuum, main effects are prioritized
- Concern for potential non-linear effects and consistency in the design space

	Noise Strategy	Appropriate Usage	Implementation	Effect	Pros	Cons
1	Hold Constant	Only to be applied when it is known (data supports) that specific noise factors **do not** influence outcomes.	These noise factors are in the inference space.	Little effect as it is already known they have no impact.	Can reduce costs of testing.	Danger that while they have no main effect, they may interact with design factors. Smaller inference space may impact prediction.
2	Randomization	Noise has yet to be identified. Trying to prevent alignment of some noise with a factor in the experiment (hidden confounding).	Noise should be allowed to vary naturally. Randomizing should account for any items desired to be removed from the inference space.	Increases inference space and hopefully prevents hidden confounding of noise with model effects.	Increases the inference space while reducing bias and hidden confounding.	Decreases precision. Noise effects are unassignable. Practically difficult to execute.
3	CRR	Quantify variation between multiple treatments of the same factors. Used when the noise has not been identified.	Multiple experimentyal uints are created for each treatment combination. Run order is randomized. Execution requires complete setup for each replicate.	Theoretically unbiased estimate of error term. Used to estimate mean square error for statistical tests.	Increased inference space with unassignable error term. And an un-biased estimate of variation can be quantified for the basis of a statistical test (F-test)	Decrease in precision. Increase in cost to execute. Practically difficult to execute. Noise unassignable for future disaggregation.
4	Covariates	When noise is measurable.	Record values of the noise factor for each treatment.	Allows for assignment of the noise factor effect in the model. Creates a mixed model of both fixed and random effects.	Increases power of statistical tests. Model accounts for covariate effect.	No noise by factor interactions estimable. Additional measurement error. Only one value per treatment. Potential lage effect missed.
5	Repeats/Nested	Effective when desire to understand variation within treatement & quantify short-term variuation (e.g., reduce measurement system, within part/batch).	Multiple measurements are taken for each experimental unit. Can also be structured for multiple layers (nested designs)	Increased precision, decreases variation due to within-treatment variation via averaging. Quantifies short-term noise for modeling.	Assessment of within treatment stability. Mean and variation can be assessed.	Increases cost relative to cost per measurement. No noise-by-factor interactions.
6	RCBD	Effectively used with product design exposed to high levels of noise. Ability to alias multiple noise factors with block. Exposes noise-by-factor interactions leading to robust design.	Noise factors are confounded with the block, set at 2 bold levels, complete block is replicated for each noise level. Many noise factors can be aliased within each block.	Lowers effective noise present in each independent block, increasing precision. The effect of Block can be estimated. Broadens inference space.	Ability to analyze block and block-by-factor effects. Create increased inference space while increasing precision. Useful for robust design	Doubles the size of an unreplicated experiment.
7	RIBD	Most useful in manufacturing processes where x's are currently unmanaged.	Noise is confounded with block, set at 2 levels, fractional blocks are run for each noise level. Block is created by aliasing with a higher-order effect.	Main effect of Block can be estimated to identify potential future factors for consideration.	Increased inference space, ability to analyze block effects.	Block effects are aliased with higher order effects and block-by-factor effects are aliased with other 2nd order effects.
8	Efficiency Split-plot	Partition noise to improve design precision while economizing resources	Restrictions are associated with sequential experiment or testing steps. Can be done with array of noise factors (Cross-over)	Allows for estimation of the whole plot and sub-plot separately impacting the precision of both.	Increases the precision of the WP and SP with significatly less resources. Specific noise-by-factor interactions identified.	More treatments than RCBD
9	Convenience Split-plot	Make the experiment easier and more economical to execute Paritions noise.	Vary levels of the subplot factors while the difficult to change factor (whole plot) is constant, then change the WP factor and replicate.	Allows for estimation of the whole plot and sub-plot separately impacting the precision of both.	More practical and permits the running of the experiment in situations where randomization would precllude running the experiment. Increased precision of subplot.	WP precision is compromised.

Table E.14: Summary of Noise Strategies for Experimentation

Appendix F
DOE Analysis and Interpretation

This appendix outlines the practical steps used to interpret the outputs of different experimental design strategies referenced in Chapter 12. These steps are software-agnostic and are provided for readers who want a structured workflow. The intent is not to prescribe a specific tool, but to ensure the required outputs are obtained for proper interpretation.

This example illustrates interpreting analysis outputs for several different experiment strategies. The different strategies show multiple ways to partition noise. In all examples, a similar data set is used.

Crazy Cajun Example

An engineer at the Crazy Cajun hot sauce factory has been assigned as a resource to the team working on the *Ultra-Hot* line. The team is composed of managers, operators and food scientists who feel there are four factors that are critical to the 'heat' and taste of the sauce: amount of **V**inegar(V), type of **P**epper(P), **M**ix time(M) & the formula of the **S**ecret ingredient(S). They have designed an experiment to understand the impact of these factors on the 'heat' of the sauce. The experiment is to be run in 1 week with 1 lot of raw materials in 1 vat. Practical significance is 50.

Possible Effect	***Rank***
Noise	3
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	1
Noise-by-factor Interactions	4
Simple Curvature (2nd order non-linear)	4
Complex Non-linear (≥3rd order non-linear)	5
≥ Three-Factor Interactions (3rd order linear)	2
Stability	5
Leverage	5
Measurement Uncertainty	2
Mean	1
Variation	1

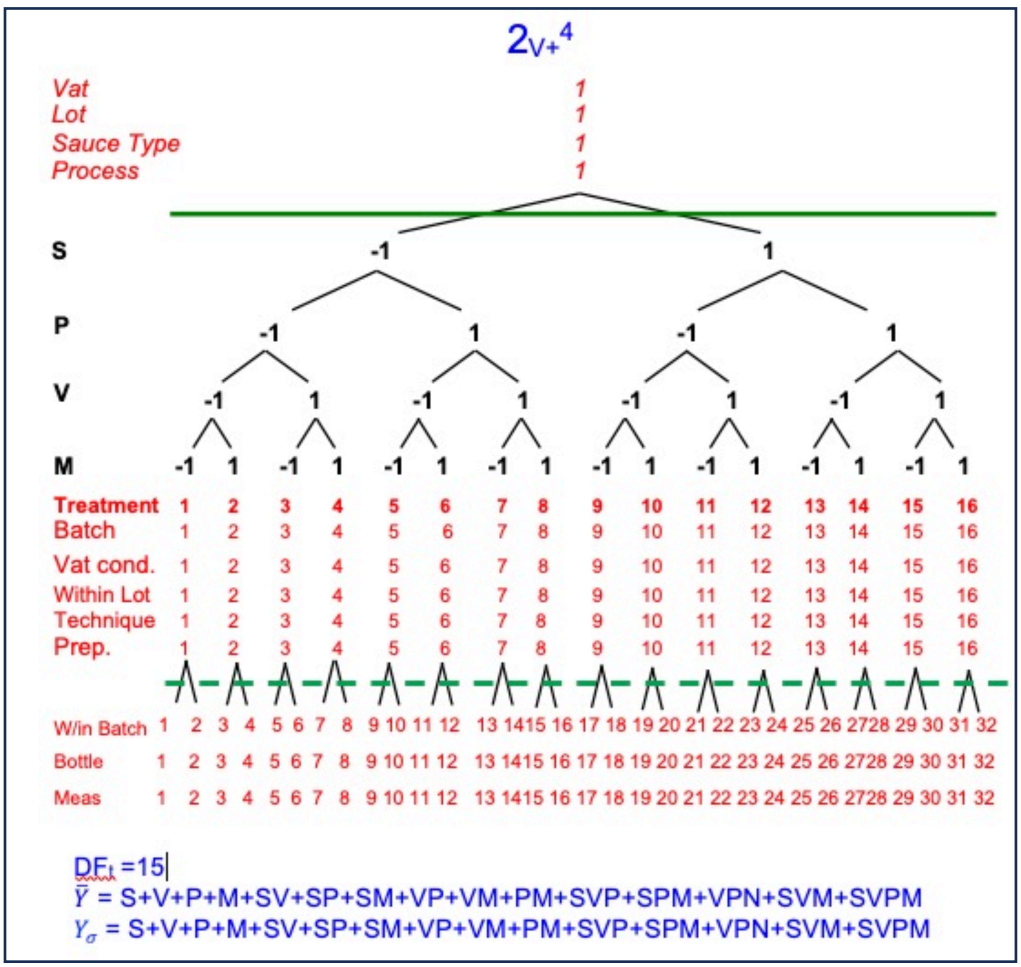


Figure F.1: FRD for Crazy Cajun Displaying Repeats Restriction

Figure F.2 shows a variability plot of the within treatment variation. The vertical lines represent the variation of the repeated measures. Use this graphic to look for unusual data points or any vertical lines that are unusually large or small.

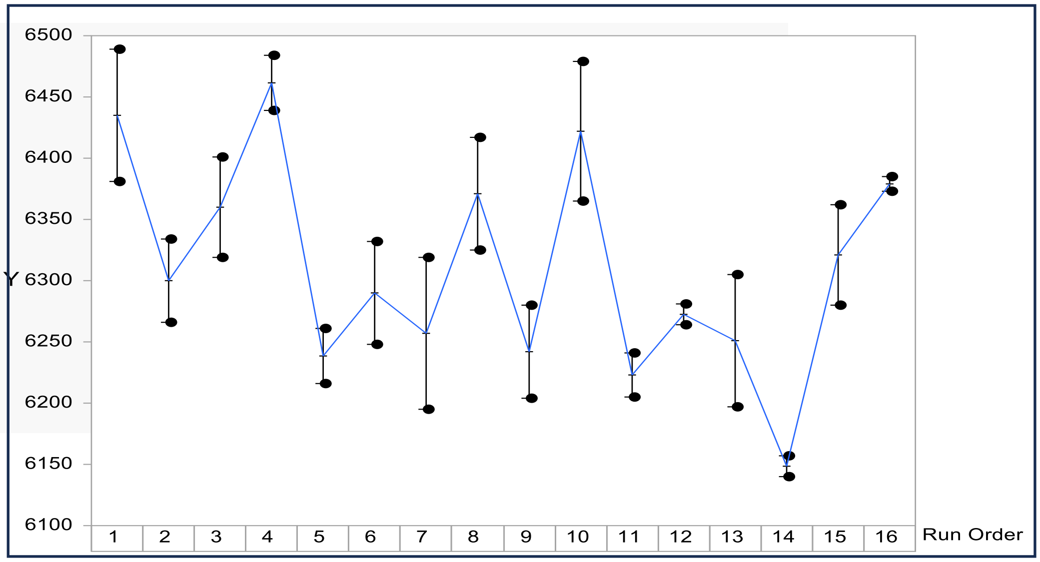


Figure F.2: Variability Plot of Within Treatment Variation

Couple this with a range chart (Figure F.3) to assess statistically significant within treatment data points.

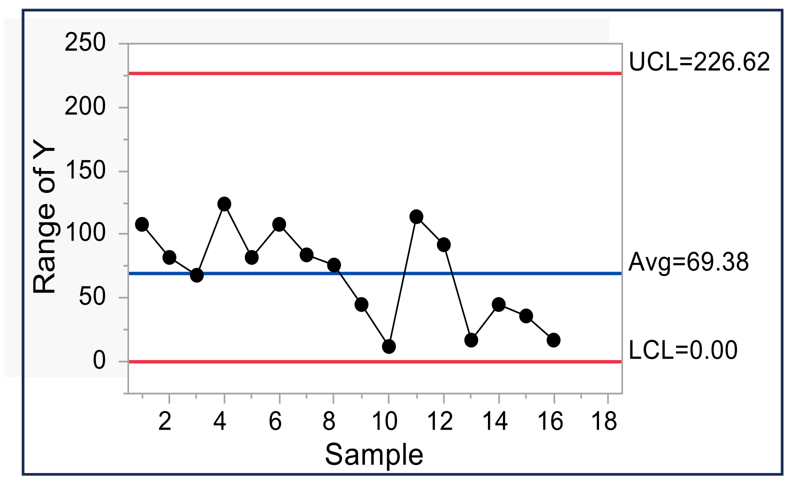


Figure F.3: Range Chart of Data Within Treatment

Since there are no unusual data points, it is acceptable to summarize the within treatment data. An average of the within treatment data reduces the variation within, hence increasing the precision of the experiment (lowers the water level). In addition, a second Y is created, the range of the within treatment data. The range (or any measure of dispersion) is an estimate of the magnitude of the variation due to noise changing within treatment. It cannot be due to the treatments as the treatments are constant when the multiple data points are acquired. This Y can be analyzed separately from the mean. There are now two response variables describing the within treatment effects. It is good practice to look for correlations when you have a multivariate situation, not only to determine if there are relationships between the Y's

(Figure F.4 and Table F.1), but also to test for any outliers (Figure F.6).

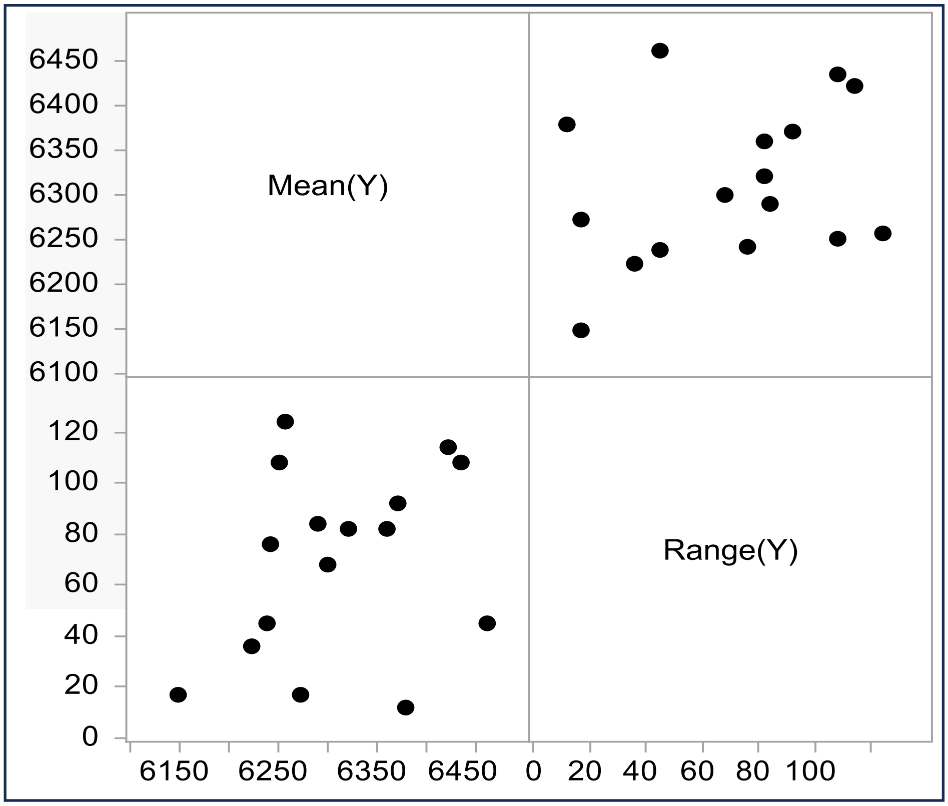


Figure F.4: Scatterplot of Correlation Between the Mean and the Range

	Mean(Y)	Range(Y)
Mean(Y)	1.0000	0.2819
Range(Y)	0.2819	1.0000

Table F.1: Pearson Correlation Coefficients (r)

Since the correlations are small, it is likely different factors affect the mean and the range. Two models will need to be built.

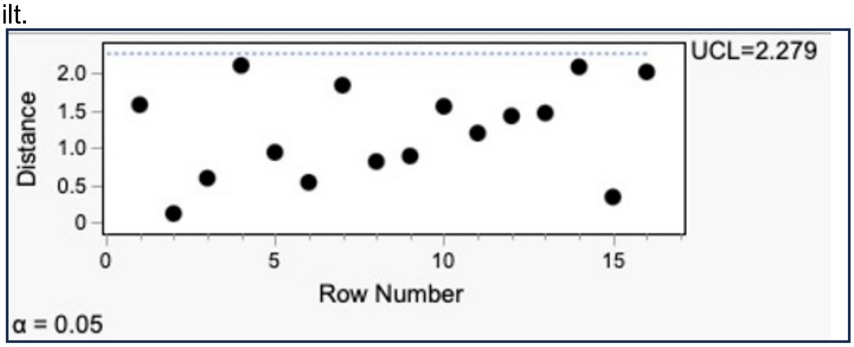


Figure F.6: Outlier Analysis: Mahalanobis

ANOG (Table F.2) is used to reveal practical differences among treatments, patterns and anomalies that guide interpretation. A moving range chart (Figure F.7) is data diagnostic to check for unusual data points.

Run Order	S	V	P	M	S*V	S*P	S*M	V*P	V*M	P*M	Mean(Y)
4	-1	-1	-1	1	1	1	-1	1	-1	-1	6461.5
1	-1	-1	-1	-1	1	1	1	1	1	1	6435
10	-1	1	-1	1	-1	1	-1	-1	1	-1	6422
16	1	-1	-1	1	-1	-1	1	1	-1	-1	6379
8	1	1	-1	1	1	-1	1	-1	1	-1	6371
3	1	-1	-1	-1	-1	-1	-1	1	1	1	6360
15	-1	-1	1	-1	1	-1	1	-1	1	-1	6321
2	-1	1	-1	-1	-1	1	1	-1	-1	1	6300
6	-1	1	1	-1	-1	-1	1	1	-1	-1	6290
12	-1	-1	1	1	1	-1	-1	-1	-1	1	6272.5
7	1	1	-1	-1	1	-1	-1	-1	-1	1	6257
13	1	-1	1	-1	-1	1	-1	-1	1	-1	6251
9	1	1	1	-1	1	1	-1	1	-1	-1	6242
5	1	-1	1	1	-1	1	1	-1	-1	1	6238.5
11	-1	1	1	1	-1	-1	-1	1	1	1	6223
14	1	1	1	1	1	1	1	1	1	1	6148.5

Table F.2: ANOG Order for the Mean

Note the patterns for P, V, S & PM. Look at P, the 6 highest Y's are when P=-1 and the bottom five lowest Y's are when P=1. That is interesting.

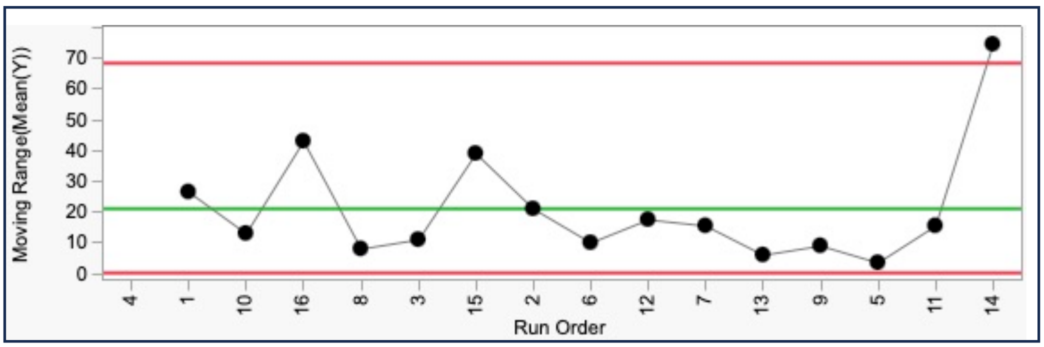


Figure F.7: Moving Range of Mean(Y)

There is evidence run number 14 was unusually small based on the rest of the data from the experiment. What happened there? This is likely due to noise impacting run 14 specifically. For the response of range, the pattern associated with M appears interesting (Table F.3). MR (Figure F.8) does indicate some rather large "jumps" in the data, and those appear to be due to factor effects (vs. noise).

Run Order	S	V	P	M	S*V	S*P	S*M	V*P	V*M	P*M	Range(Y)
16	1	-1	-1	1	-1	-1	1	1	-1	-1	12
12	-1	-1	1	1	1	-1	-1	-1	-1	1	17
14	1	1	1	1	1	1	1	1	1	1	17
11	-1	1	1	1	-1	-1	-1	1	1	1	36
4	-1	-1	-1	1	1	1	-1	1	-1	-1	45
5	1	-1	1	1	-1	1	1	-1	-1	1	45
2	-1	1	-1	-1	-1	1	1	-1	-1	1	68
9	1	1	1	-1	1	1	-1	1	-1	-1	76
3	1	-1	-1	-1	-1	-1	-1	1	1	1	82
15	-1	-1	1	-1	1	-1	1	-1	1	-1	82
6	-1	1	1	-1	-1	-1	1	1	-1	-1	84
8	1	1	-1	1	1	-1	1	-1	1	-1	92
1	-1	-1	-1	-1	1	1	1	1	1	1	108
13	1	-1	1	-1	-1	1	-1	-1	1	-1	108
10	-1	1	-1	1	-1	1	-1	-1	1	-1	114
7	1	1	-1	-1	1	-1	-1	-1	-1	1	124

Table F.3: ANOG for Range

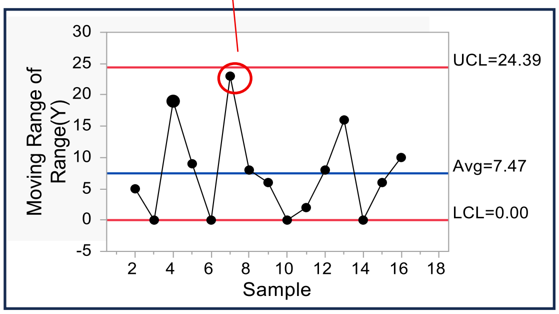


Figure F.8: MR Chart for Range

The patterns observed from ANOG can be substantiated with statistical methods. The Normal/Half normal plots (Figure F.9 and F.11) assist in determining if those patterns are significant. These should only be used for the saturated model. In some software program outputs, there is a blue line on the plot. This is a function of Lenth's pseudo-standard error[1]. In some cases, this line provides guidance interpreting the random errors that are the basis for determining statistical significance (random errors would form a straight line on this plot). Unfortunately, there are instances where Lenth's PSE is not applicable to the effects being plotted. In those cases, interpreting which points make up the random errors must be judged. I use the "fat pencil" test to help interpret. Fit a pencil to the data, centered around an estimate of 0. Points covered by the pencil are random errors. Notice on the normal plot for the mean, P is the farthest from the pencil, then PM, S and V, these are statistically significant. Also notice the significant 3-factor interaction, VPM. This is likely an artifact of the special cause due to run 14 as already identified.

[1] Russel Lenth

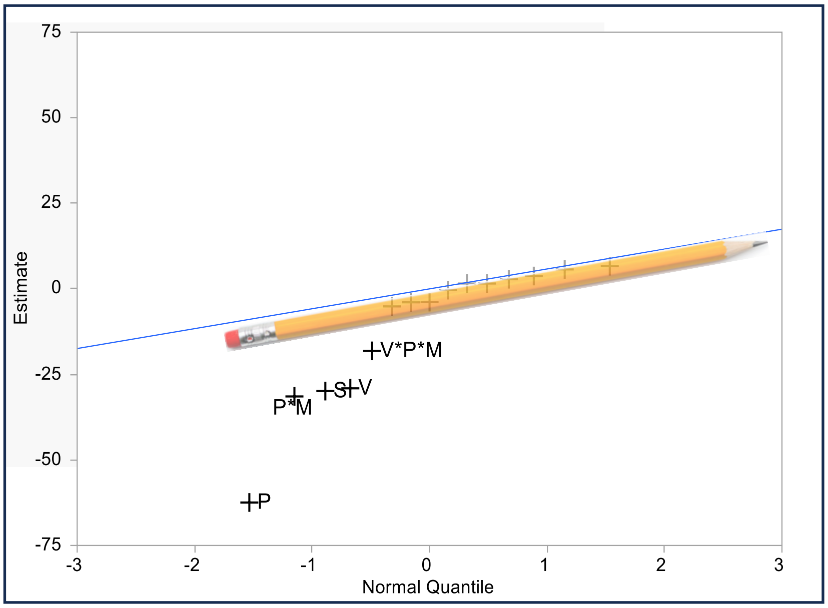


Figure F.9: Normal Plot of Effects for the Mean

The Pareto plot (Figures F.10 and F.12) shows the practical significance of the factors. In this case, practical significance is an effect of 50, on the estimates plot, that is 25 (1/2 the effect).

Term	Estimate
P	-62.43750
P*M	-31.43750
S	-29.87500
V	-29.06250
V*P*M	-18.12500
V*P	6.62500
V*M	5.68750
S*V*P	-5.12500
S*V*M	-3.93750
S*V*P*M	-3.87500
M	3.75000
S*V	2.81250
S*P	1.56250
S*P*M	1.56250
S*M	-0.37500

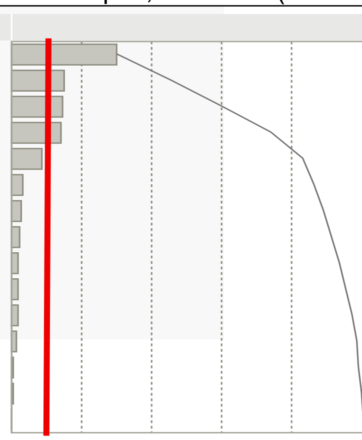


Figure F.10: Pareto Plot of Effects with Practical Significance Reference Line

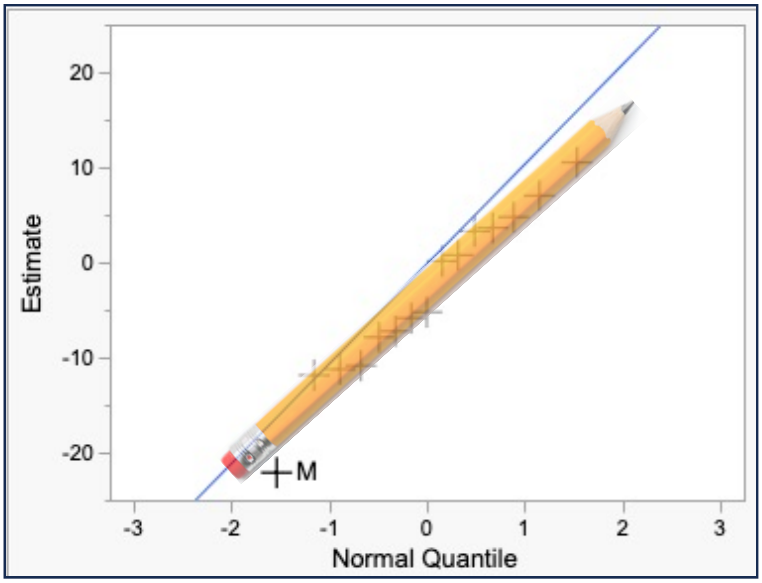
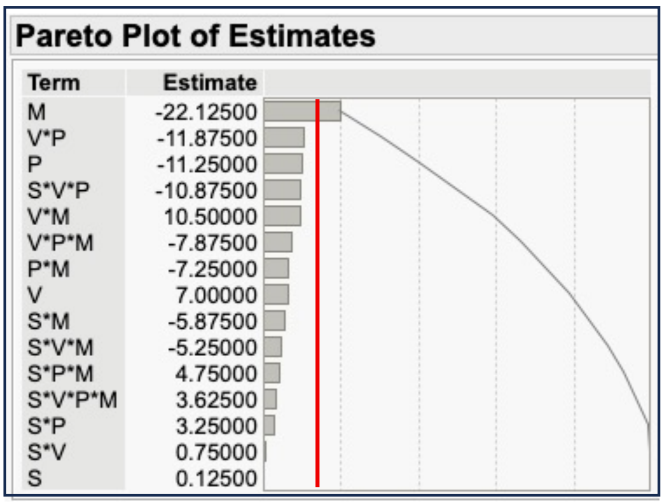


Figure F.11: Normal Plot for Range



Pareto Plot of Estimates

Term	Estimate
M	-22.12500
V*P	-11.87500
P	-11.25000
S*V*P	-10.87500
V*M	10.50000
V*P*M	-7.87500
P*M	-7.25000
V	7.00000
S*M	-5.87500
S*V*M	-5.25000
S*P*M	4.75000
S*V*P*M	3.62500
S*P	3.25000
S*V	0.75000
S	0.12500

Figure F.12: Pareto Plot of Effects for the Range

Once the significant effects (both practical and statistical), are determined, simplify or reduce the model by removing the insignificant terms. There are a number of considerations that guide model reduction (Table F.4 & Figures F.13-16):

- magnitude of the R-square Adj (amount of variation in the data explained by the model),
- Δ between R-square Adj and R-square (test overspecified model),
- root mean square error (standard deviation of the model; smaller is better),
- residuals analysis helps assess quantitative assumptions,
- leverage plots (look for outliers, multicollinearity and confidence limits crossing horizontal line),

RSquare	0.931313
RSquare Adj	0.906335
Root Mean Square Error	26.65702
Mean of Response	6310.75
Observations (or Sum Wgts)	16

RSquare	0.389088
RSquare Adj	0.345452
Root Mean Square Error	29.63769
Mean of Response	69.375
Observations (or Sum Wgts)	16

Table F.4: Summary Statistics for Mean and Range

For the mean model, the 4 terms in the model explain ~91% of the variation in the data. The delta is reasonable and the model has a standard deviation of ~27. For the range, 1 term explains ~35% of the variation, delta is reasonable and the model standard deviation is large at ~30.

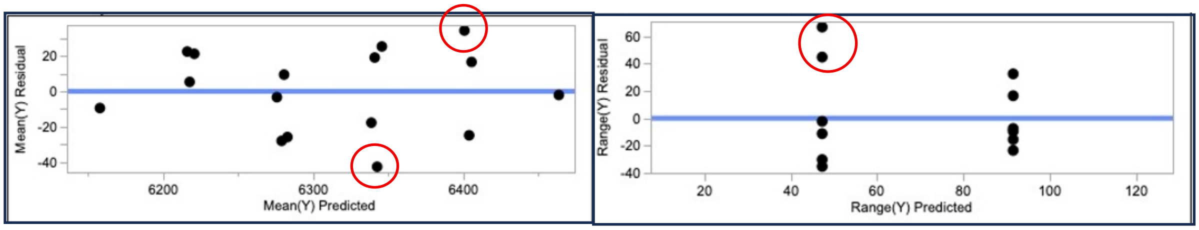


Figure F.13: Residual by Predicted Plots

Residual plots help diagnose special cause and model inadequacy. There appears to be a couple of unusual residuals for both the mean and range (circled). This does not mean the raw data is unusual, it is an indication the model did not predict those actual values well. In some cases, this is due to model insufficiency or noise. Plotting those same residuals in run order may offer clues.

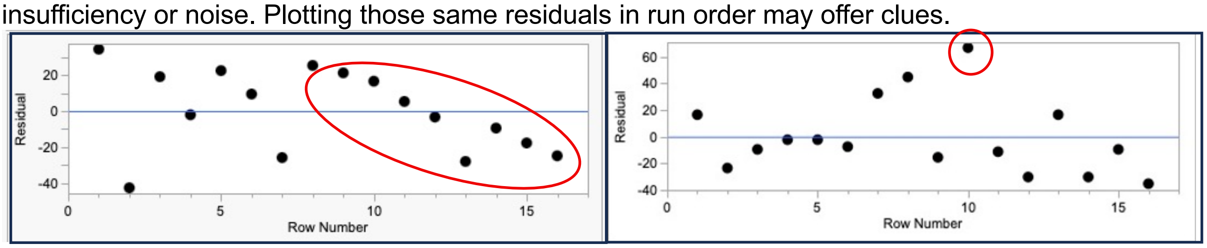


Figure F.14: Residuals Plotted in Run Order (Mean and Range)

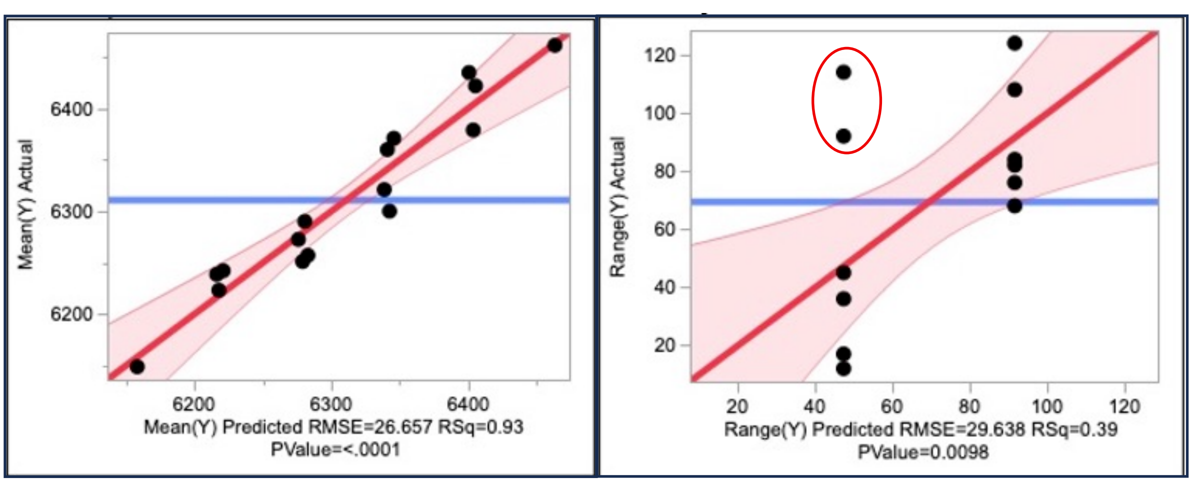


Figure F.15: Leverage Plots, Actual by Predicted

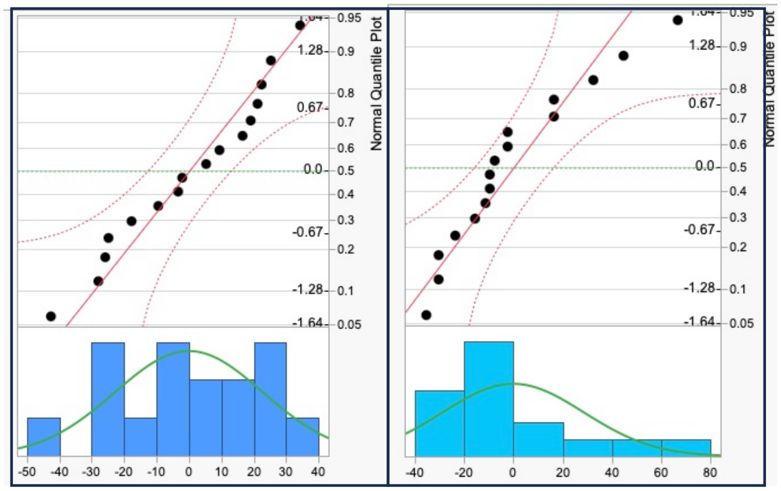


Figure F.16: Normal Plots of Residuals; Mean and Range

The plots do not provide the answers, but stimulate questions about the model adequacy and possible further investigations. Main effects plots (Figure F.17) and interaction plots (Figure F.18) provide directional information.

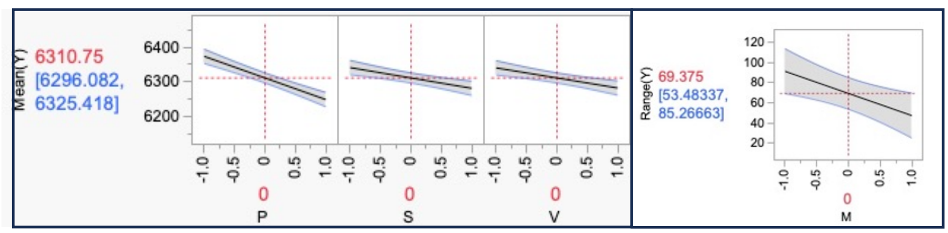


Figure F.17: Main Effects Plots

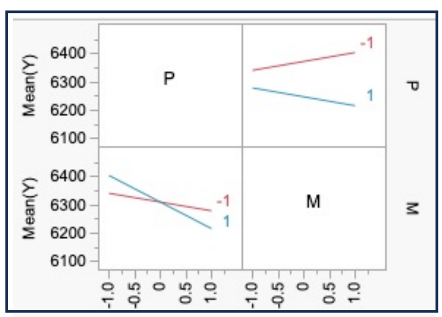


Figure F.18: Interaction Plot

Investigation should continue with a focus on **P**eppers (e.g., type, amount, prep.), **S**ecret ingredient (e.g., formula, amount) and **V**inegar (e.g., type, amount), recognizing the effect of **P** depends on **M**ix time. A longer mix time results in less variation, but with the right peppers, perhaps mix time can be reduced? Or, perhaps changing geometry of mixer blades or speed of mixing or location of blades can reduce mix time?

Strategy for Handling Long-term Noise: Completely Randomized Replicates (CRR):
There are several options for replication. This is NOT the most effective strategy. The statisticians' default recommendation is randomized replicates (Figure F.19). This is an appropriate strategy if noise has not been identified before the experiment is conducted.

Possible Effect	***Rank***
Noise	3
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	1
Noise-by-factor Interactions	5
Simple Curvature (2nd order non-linear)	5
Complex Non-linear (≥3rd order non-linear)	5
≥ Three-Factor Interactions (3rd order linear)	1
Stability	5
Leverage	5
Measurement Uncertainty	5
Mean	5
Variation	5

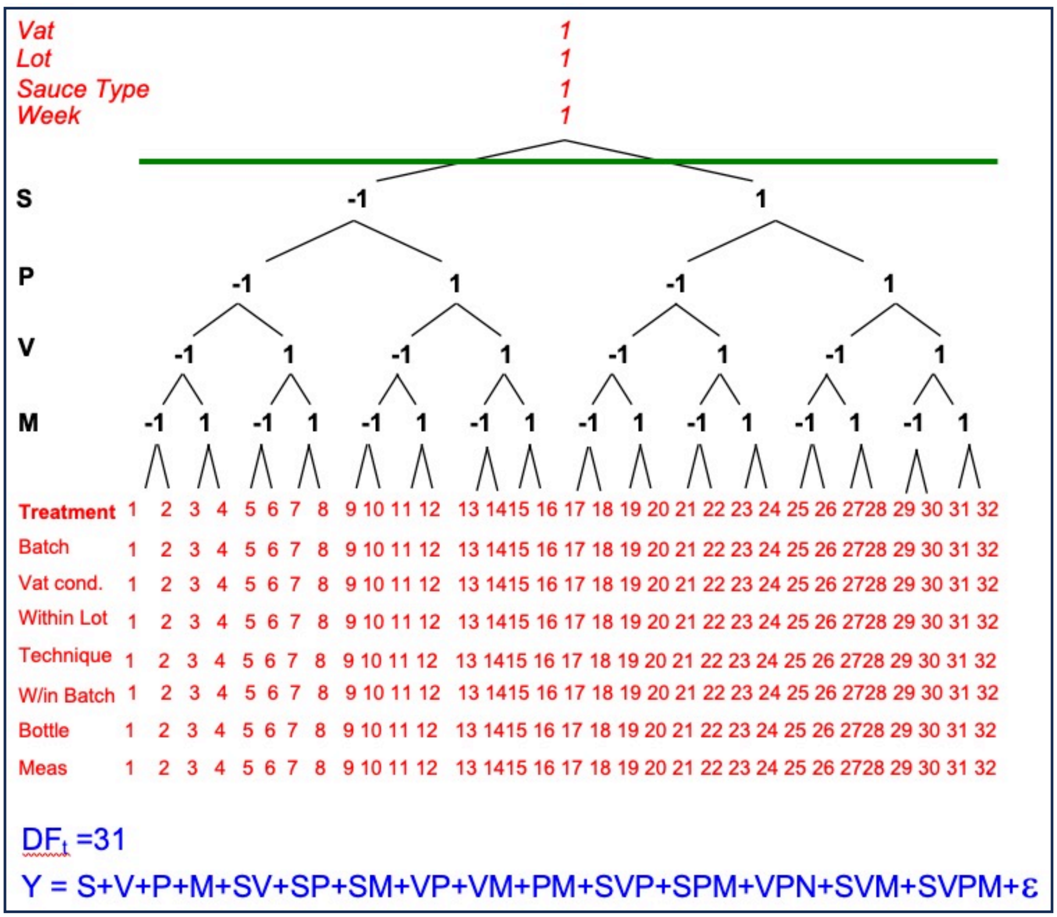


Figure F.19: FRD with Randomized Replicates

Analysis always begins with a practical look at the data. Is there sufficient practical variation in the response to consider further analysis? Recall a change of 50 units was significant. ANOG is used to look for patterns, jumps between treatments and identify any special cause treatments (Table F.5).

S	V	P	M	S*V	S*P	S*M	V*P	V*M	P*M	Y
-1	-1	-1	-1	1	1	1	1	1	1	6489
-1	-1	-1	1	1	1	-1	1	-1	-1	6484
-1	1	-1	1	-1	1	-1	-1	1	-1	6479
-1	-1	-1	1	1	1	-1	1	-1	-1	6439
1	1	-1	1	1	-1	1	-1	1	-1	6417
1	-1	-1	-1	-1	-1	-1	1	1	1	6401
1	-1	-1	1	-1	-1	1	1	-1	-1	6385
-1	-1	-1	-1	1	1	1	1	1	1	6381
1	-1	-1	1	-1	-1	1	1	-1	-1	6373
-1	1	-1	1	-1	1	-1	-1	1	-1	6365
-1	-1	1	-1	1	-1	1	-1	1	-1	6362
-1	1	-1	-1	-1	1	1	-1	-1	1	6334
-1	1	1	-1	-1	-1	1	1	-1	-1	6332
1	1	-1	1	1	-1	1	-1	1	-1	6325
1	-1	-1	-1	-1	-1	-1	1	1	1	6319
1	1	-1	-1	1	-1	-1	-1	-1	1	6319
1	-1	1	-1	-1	1	-1	-1	1	-1	6305
-1	-1	1	1	1	-1	-1	-1	-1	1	6281
-1	-1	1	-1	1	-1	1	-1	1	-1	6280
1	1	1	-1	1	1	-1	1	-1	-1	6280
-1	1	-1	-1	-1	1	1	-1	-1	1	6266
-1	-1	1	1	1	-1	-1	-1	-1	1	6264
1	-1	1	1	-1	1	1	-1	-1	1	6261
-1	1	1	-1	-1	-1	1	1	-1	-1	6248
-1	1	1	1	-1	-1	-1	1	1	1	6241
1	-1	1	1	-1	1	1	-1	-1	1	6216
-1	1	1	1	-1	-1	-1	1	1	1	6205
1	1	1	-1	1	1	-1	1	-1	-1	6204
1	-1	1	-1	-1	1	-1	-1	1	-1	6197
1	1	-1	-1	1	-1	-1	-1	-1	1	6195
1	1	1	1	1	1	1	1	1	1	6157
1	1	1	1	1	1	1	1	1	1	6140

Table F.5: ANOG of CRR Data

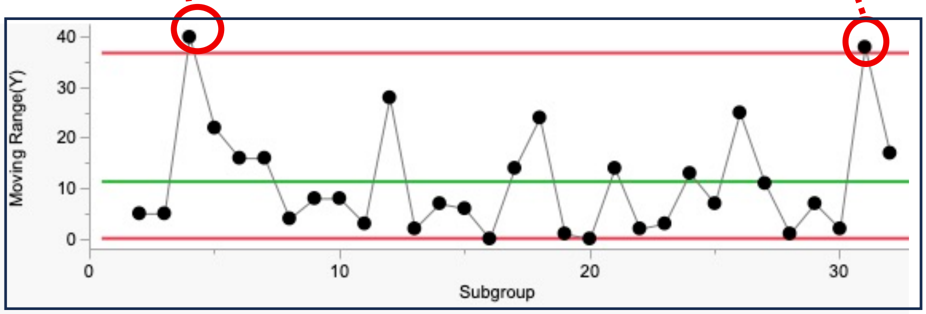


Figure F.20: MR of Data in ANOG Order

The MR chart (Figure F.20) indicates two significant jumps in the response. Where large jumps in the data occur, there is a greater effect. Correlate factors/levels with those jumps. The top 4 rows and the bottom 2 rows of data have been highlighted due to their leveraged impact. Look for factor levels to be the same and opposite top to bottom. P, S, V and PM are interesting. In this strategy, it is not possible to create a saturated model as there are unassignable degrees of freedom. Normal/half-normal and Pareto plots are not appropriate. In this case we rely on quantitative analysis. Model summary statistics contain R-squares and the RMSE (Table F.6). You want the largest R-square Adjusted and the smallest delta between R-square Adjusted and R-square. Watch the change in these statistics as terms are removed from the model.

RSquare	0.824135
RSquare Adj	0.659261
Root Mean Square Error	55.09537
Mean of Response	6310.75
Observations (or Sum Wgts)	32

Table F.6: Summary Model Statistics

Analysis of Variance (ANOVA, Table F.7) is used to assess the over model significance and the model is subsequently disaggregated with effects tests.

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	15	227598.00	15173.2	4.9986
Error	16	48568.00	3035.5	**Prob > F**
C. Total	31	276166.00		0.0014*

Source	DF	Sum of Squares	F Ratio	Prob > F
S	1	28560.50	9.4088	0.0074*
V	1	27028.13	8.9040	0.0088*
P	1	124750.13	41.0971	<.0001*
S*V	1	253.13	0.0834	0.7765
S*P	1	78.13	0.0257	0.8746
V*P	1	1404.50	0.4627	0.5061
S*V*P	1	840.50	0.2769	0.6060
M	1	450.00	0.1482	0.7053
S*M	1	4.50	0.0015	0.9698
V*M	1	1035.13	0.3410	0.5674
S*V*M	1	496.13	0.1634	0.6914
P*M	1	31626.13	10.4188	0.0053*
S*P*M	1	78.13	0.0257	0.8746
V*P*M	1	10512.50	3.4632	0.0812
S*V*P*M	1	480.50	0.1583	0.6960

Table F.7: ANOVA

Initial look at residuals plots may provide insight to model integrity. These display some unusual patterns suggesting issues with assumptions (Figure F.21-24).

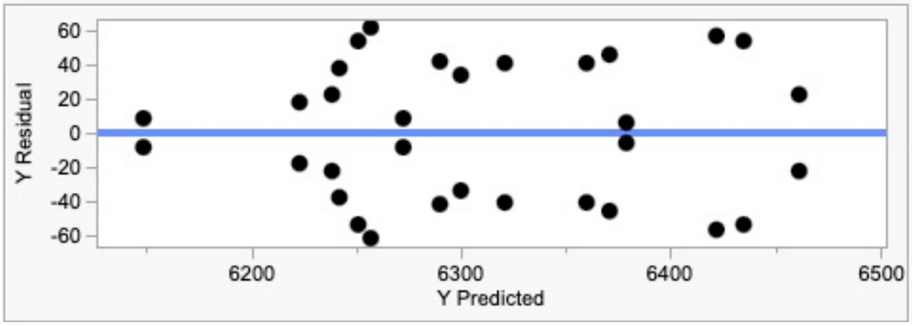


Figure F.21: Residual by Predicted

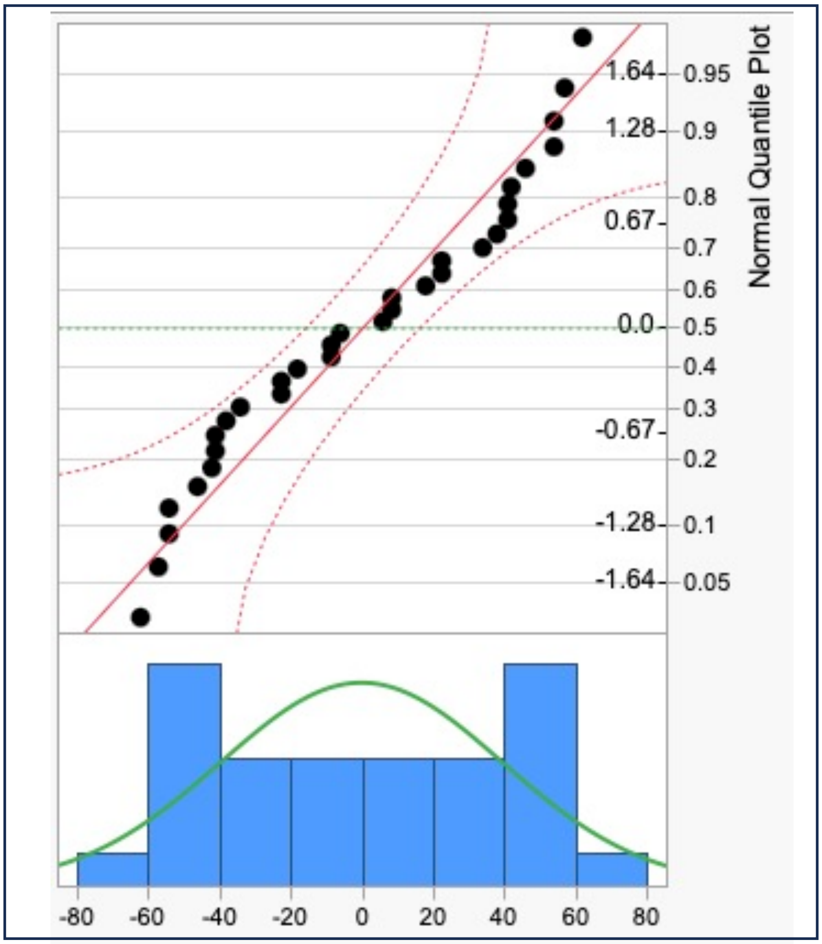


Figure F.22: Normal Probability Plot of Residuals

The model is reduced. The model summary statistics residuals and are re-evaluated.

RSquare	0.767527
RSquare Adj	0.733087
Root Mean Square Error	48.76289
Mean of Response	6310.75
Observations (or Sum Wgts)	32

Table F.8: Model Summary Statistics; Reduced Model

Note the model explains about 73% of the variation in the experiment and the delta between R-square

Adjusted and R-square is reduced. Residuals look better, though there still may be some outliers.

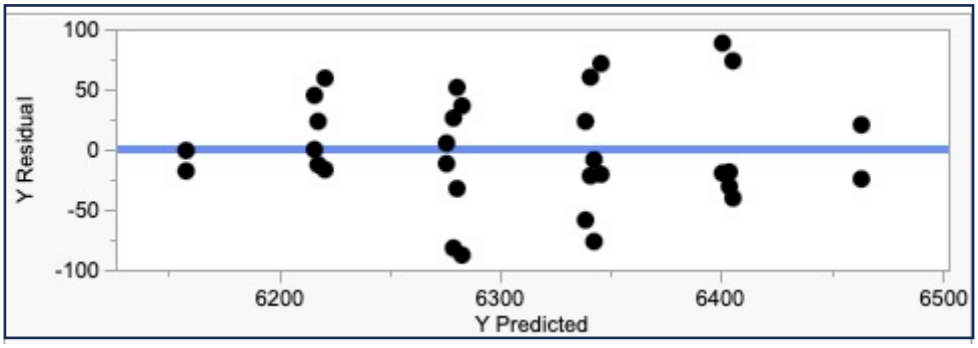


Figure F.23: Residuals for Reduced Model

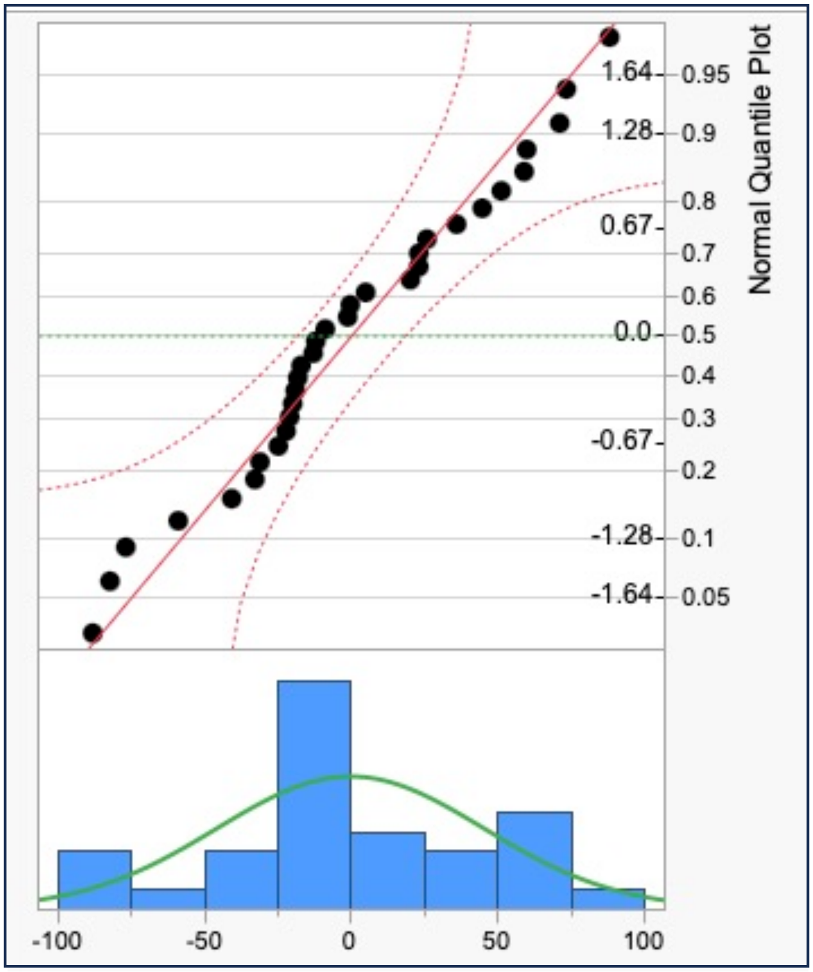


Figure F.24: Probability Plot and Distribution of Residuals

Investigating some of the unusual residuals is likely appropriate. Main effects and interaction plots are provided to provide insight to the next iteration (Figures F.25-26).

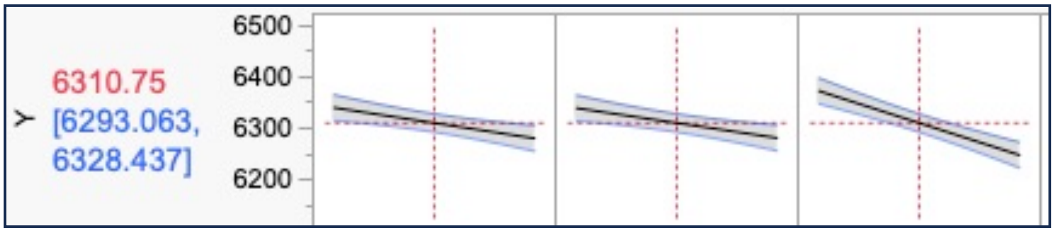


Figure F.25: Main Effects Plot

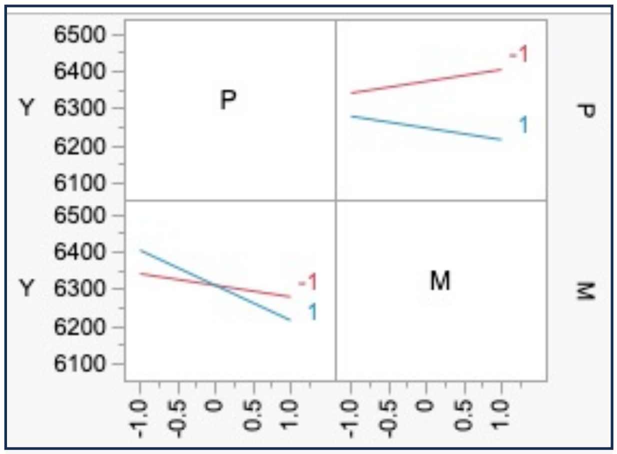


Figure F.26: Interaction Plot

Strategy for Handling Long-term Noise: Randomized Complete Block Design (RCBD):

Possible Effect	*Rank*
Noise	1
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	1
Noise-by-factor Interactions	1
Simple Curvature (2nd order non-linear)	4
Complex Non-linear (≥3rd order non-linear)	5
≥ Three-Factor Interactions (3rd order linear)	4
Stability	3
Leverage	4
Measurement Uncertainty	5
Mean	5
Variation	5

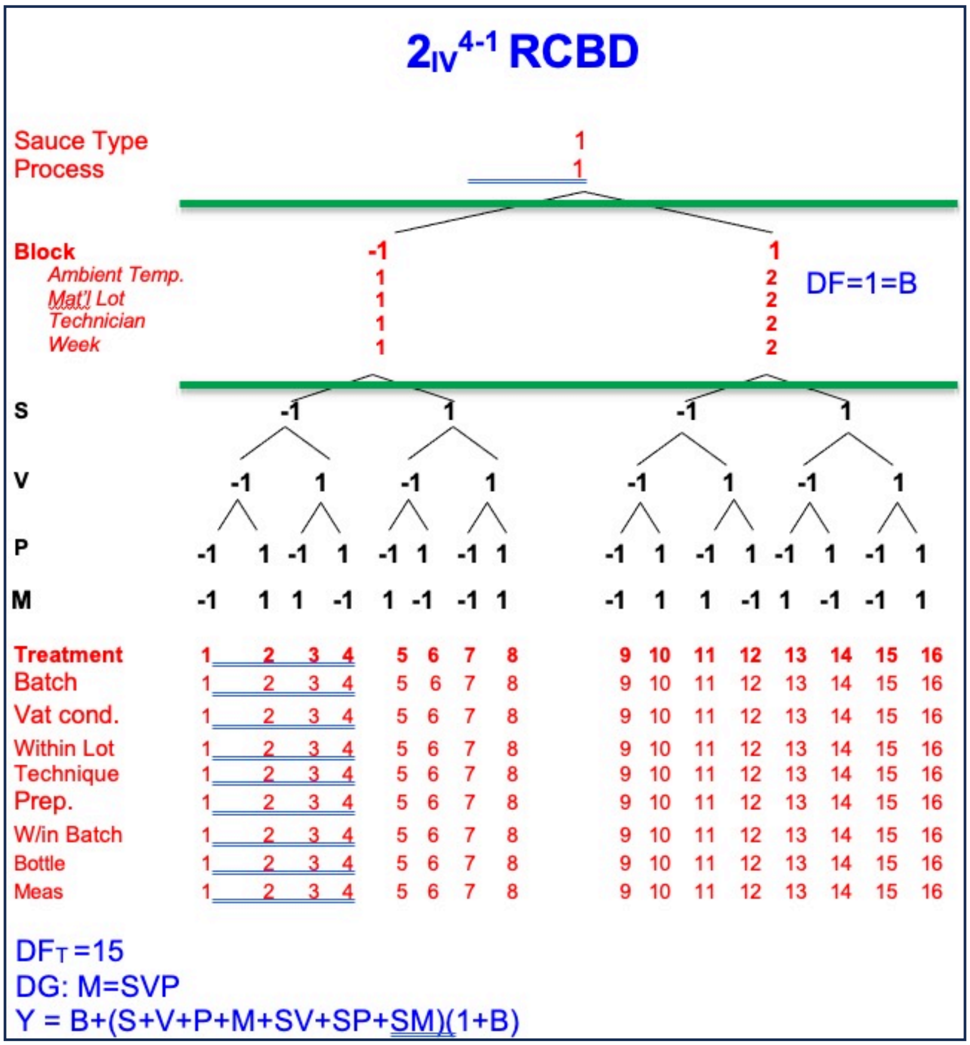


Figure F.27: RCBD

Analysis of RCBD follows the same sequence, PGQ. ANOG is shown in F.9. The ANOG, similar to Table

F.5, identifying the same jumps in the response variable. In this case, the block effect (B) clearly has an associated pattern. To analyze RCBD, start with the saturated model Normal (Figure F.29) and Pareto plots (Figure F.30) are shown.

B	S	V	P	M	Y	B*S	B*V	B*P	B*M	S*V	S*P	S*M	V*P	V*M	P*M
1	-1	-1	-1	-1	6489	-1	-1	-1	-1	1	1	1	1	1	1
1	-1	-1	-1	1	6484	-1	-1	-1	1	1	1	-1	1	-1	-1
1	-1	1	-1	1	6479	-1	1	-1	1	-1	1	-1	-1	1	-1
-1	-1	-1	-1	1	6439	1	1	1	-1	1	1	-1	1	-1	-1
1	1	1	-1	1	6417	1	1	-1	1	1	-1	1	-1	1	-1
1	1	-1	-1	-1	6401	1	-1	-1	-1	-1	-1	-1	1	1	1
-1	1	-1	-1	1	6385	-1	1	1	-1	-1	-1	1	1	-1	-1
-1	-1	-1	-1	-1	6381	1	1	1	1	1	1	1	1	1	1
1	1	-1	-1	1	6373	1	-1	-1	1	-1	-1	1	1	-1	-1
-1	-1	1	-1	1	6365	1	-1	1	-1	-1	1	-1	-1	1	-1
1	-1	-1	1	-1	6362	-1	-1	1	-1	1	-1	1	-1	1	-1
1	-1	1	-1	-1	6334	-1	1	-1	-1	-1	1	1	-1	-1	1
1	-1	1	1	-1	6332	-1	1	1	-1	-1	-1	1	1	-1	-1
-1	1	1	-1	1	6325	-1	-1	1	-1	1	-1	1	-1	1	-1
-1	1	-1	-1	-1	6319	-1	1	1	1	-1	-1	-1	1	1	1
1	1	1	-1	-1	6319	1	1	-1	-1	1	-1	-1	-1	-1	1
1	1	-1	1	-1	6305	1	-1	1	-1	-1	1	-1	-1	1	-1
1	-1	-1	1	1	6281	-1	-1	1	1	1	-1	-1	-1	-1	1
-1	-1	-1	1	-1	6280	1	1	-1	1	1	-1	1	-1	1	-1
1	1	1	1	-1	6280	1	1	1	-1	1	1	-1	1	-1	-1
-1	-1	1	-1	-1	6266	1	-1	1	1	-1	1	1	-1	-1	1
-1	-1	-1	1	1	6264	1	1	-1	-1	1	-1	-1	-1	-1	1
1	1	-1	1	1	6261	1	-1	1	1	-1	1	1	-1	-1	1
-1	-1	1	1	-1	6248	1	-1	-1	1	-1	-1	1	1	-1	-1
1	-1	1	1	1	6241	-1	1	1	1	-1	-1	-1	1	1	1
-1	1	-1	1	1	6216	-1	1	-1	-1	-1	1	1	-1	-1	1
-1	-1	1	1	1	6205	1	-1	-1	-1	-1	-1	-1	1	1	1
-1	1	1	1	-1	6204	-1	-1	-1	1	1	1	-1	1	-1	-1
-1	1	-1	1	-1	6197	-1	1	-1	1	-1	1	-1	-1	1	-1
-1	1	1	-1	-1	6195	-1	-1	1	1	1	-1	-1	-1	-1	1
1	1	1	1	1	6157	1	1	1	1	1	1	1	1	1	1
-1	1	1	1	1	6140	-1	-1	-1	-1	1	1	1	1	1	1

Table F.9: ANOG RCBD

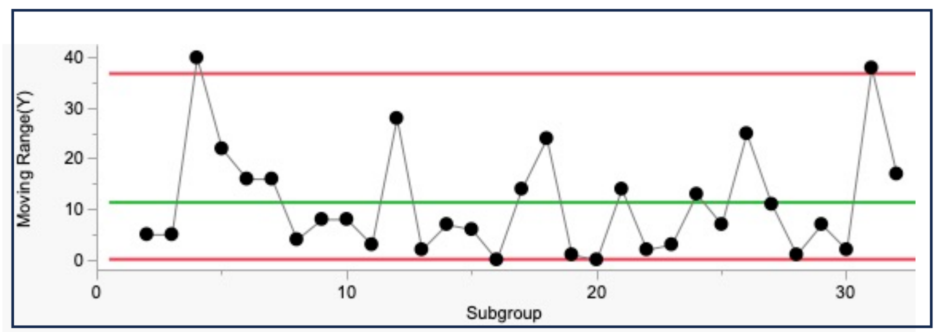


Figure F.28: MR of Data in ANOG Order

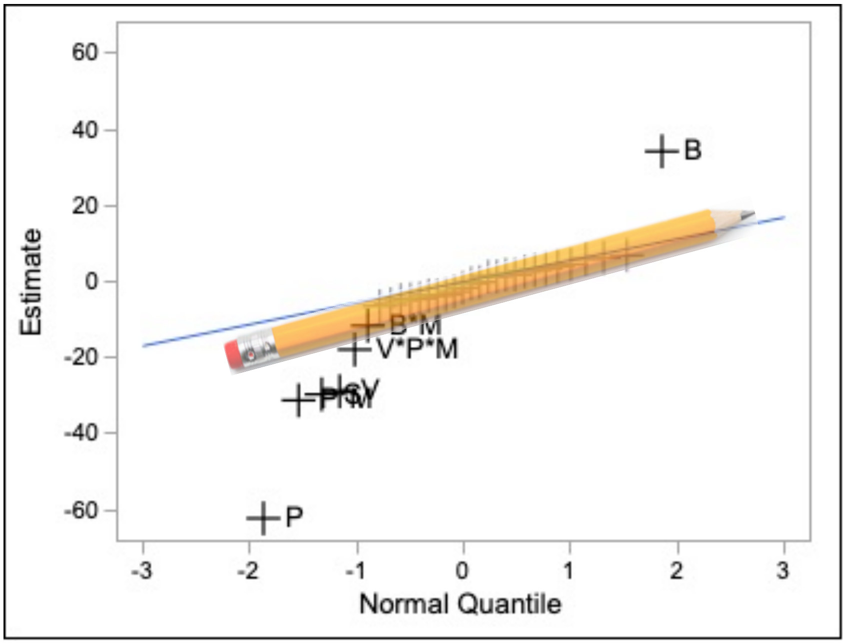


Figure F.29: Normal Plot with Block Effects

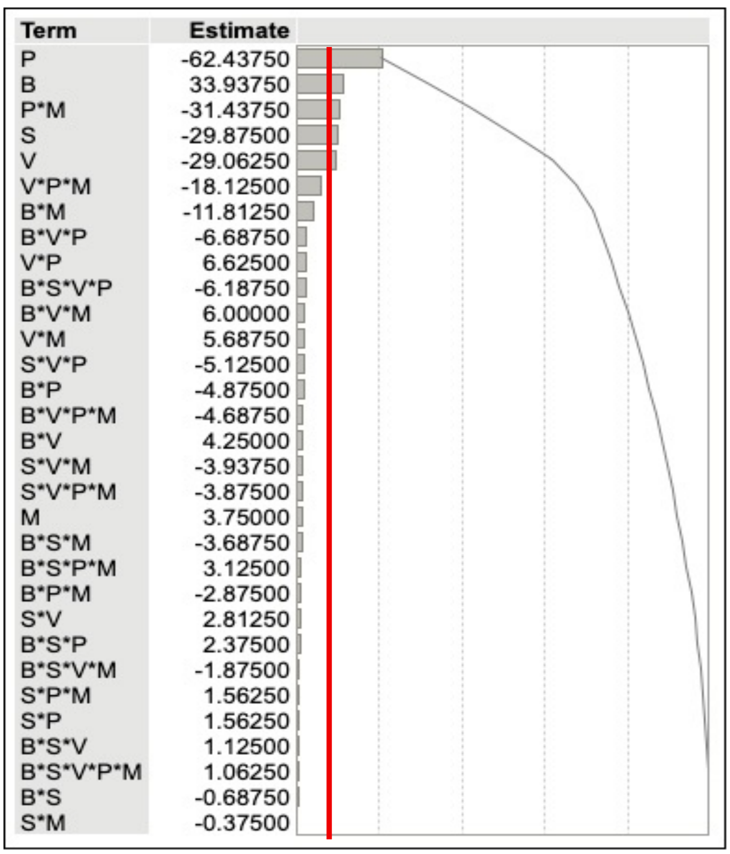


Figure F.30: Pareto Plot with Practical Significance

After reducing the model, Figures F.31-F.32 display the outputs.

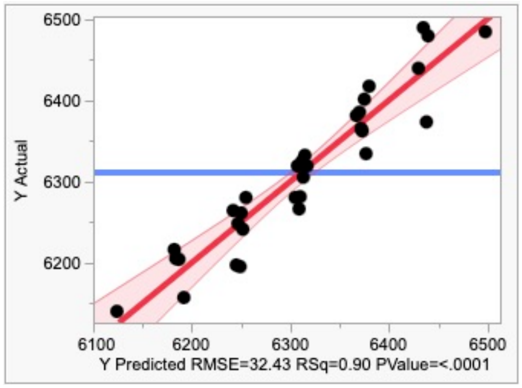


Figure F.31: Actual vs. Predicted

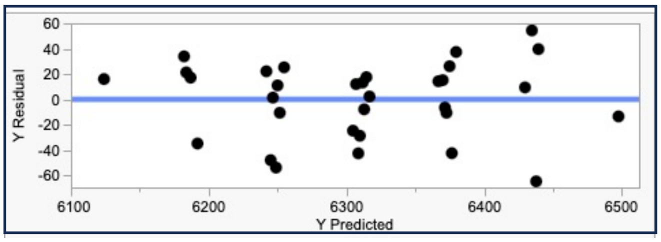


Figure F.32: Residual Plot

RSquare	0.900983
RSquare Adj	0.881942
Root Mean Square Error	32.4304
Mean of Response	6310.75
Observations (or Sum Wgts)	32

Table F.10: Model Summary Statistics

The reduced model explains ~88% of the variation in the data. The delta R-square Adjusted and R-Square is relatively small and the RMSE is smaller than prior to model reduction.

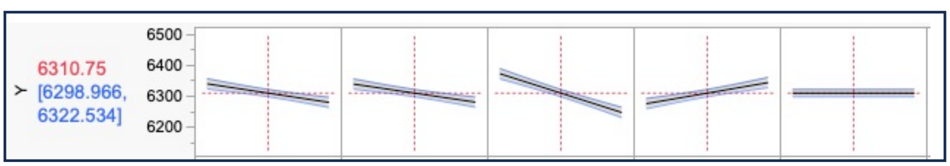


Figure F.33: Main Effects Plot

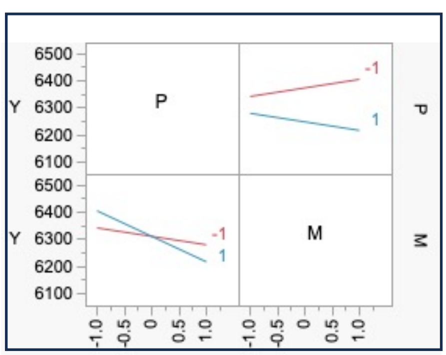


Figure F.34: Interaction Plot

Strategy for Handling Long-term Noise: Randomized Incomplete Block Design (RIBD or BIB): Blocking doubles the size of the experiment. If the investigation is still in the early stages, it may be appropriate to use incomplete blocks to increase inference space improve precision and investigate many confounded x's for further iteration.

Possible Effect	*Rank*
Noise	1
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	3
Noise-by-factor Interactions	3
Simple Curvature (2nd order non-linear)	4
Complex Non-linear (≥3rd order non-linear)	5
≥ Three-Factor Interactions (3rd order linear)	4
Stability	3
Leverage	4
Measurement Uncertainty	5
Mean	5
Variation	5

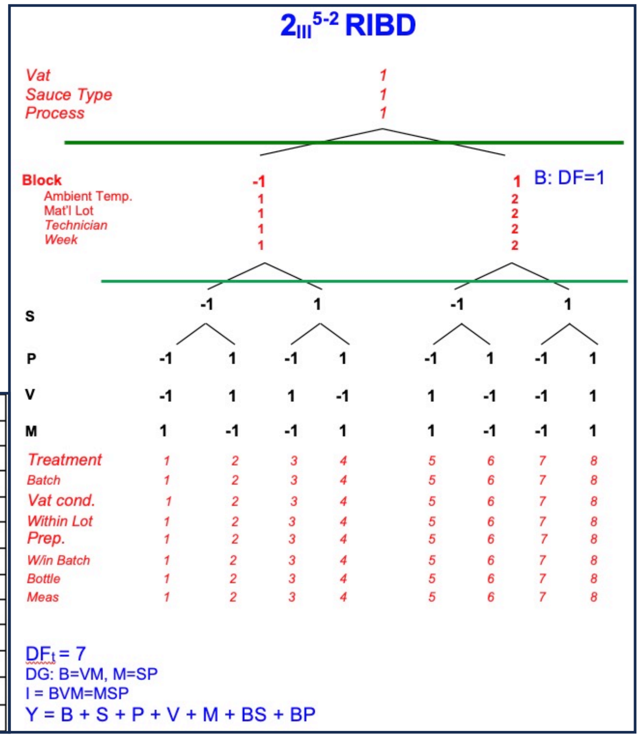


Figure F.35: FRD RIBD

The Printed Circuit Board Screening experiment will provide analysis interpretation.

Strategy for Efficiently Partitioning Noise with Design Factors: Split-plots

Efficiency Split-plot

Possible Effect	*Rank*
Noise	2
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	1
Noise-by-factor Interactions	3
Simple Curvature (2nd order non-linear)	4
Complex Non-linear (≥3rd order non-linear)	5
≥ Three-Factor Interactions (3rd order linear)	1
Stability	3
Leverage	4
Measurement Uncertainty	5
Mean	5
Variation	5

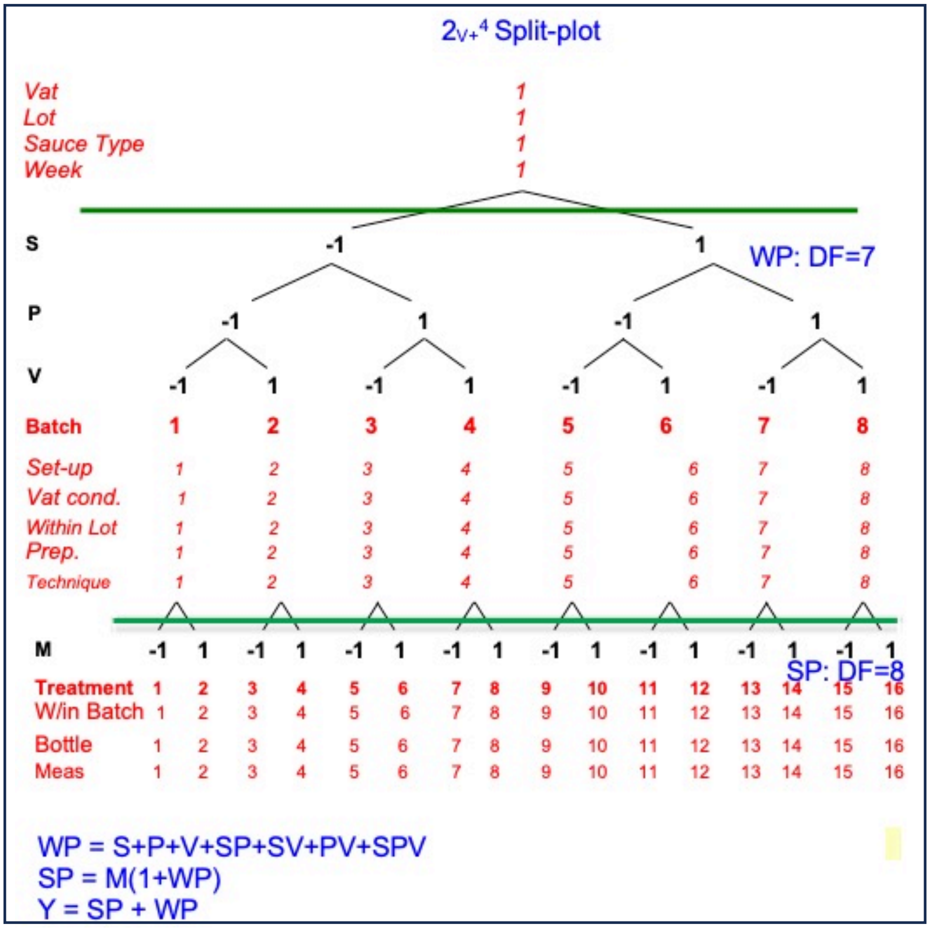


Figure F.36: Efficiency Split-plot

While practical analysis is similar to other examples, graphical and quantitative analyses are quite different. Factors P, V, S and PM are interesting.

S	V	P	M	S*V	S*P	S*M	V*P	V*M	P*M	Y
-1	-1	-1	1	1	1	-1	1	-1	-1	6439
1	-1	-1	1	-1	-1	1	1	-1	-1	6385
-1	-1	-1	-1	1	1	1	1	1	1	6381
-1	1	-1	1	-1	1	-1	-1	1	-1	6365
1	1	-1	1	1	-1	1	-1	1	-1	6325
1	-1	-1	-1	-1	-1	-1	1	1	1	6319
-1	-1	1	-1	1	-1	1	-1	1	-1	6280
-1	1	-1	-1	-1	1	1	-1	-1	1	6266
-1	-1	1	1	1	-1	-1	-1	-1	1	6264
-1	1	1	-1	-1	-1	1	1	-1	-1	6248
1	-1	1	1	-1	1	1	-1	-1	1	6216
-1	1	1	1	-1	-1	-1	1	1	1	6205
1	1	1	-1	1	1	-1	1	-1	-1	6204
1	-1	1	-1	-1	1	-1	-1	1	-1	6197
1	1	-1	-1	1	-1	-1	-1	-1	1	6195
1	1	1	1	1	1	1	1	1	1	6140

Table F.11: ANOG Order

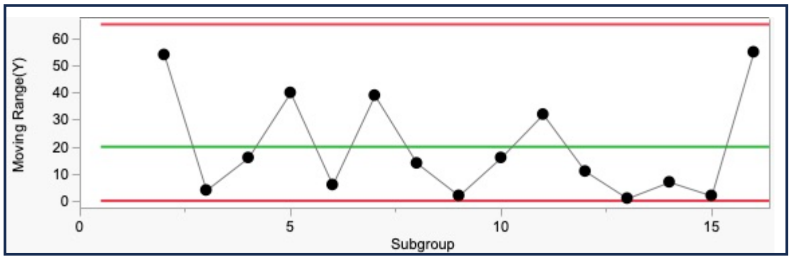


Figure F.37: MR in ANOG Order

Normal and Pareto plots must be created for each split, a separate normal and Pareto plot for the whole plot and the subplot.

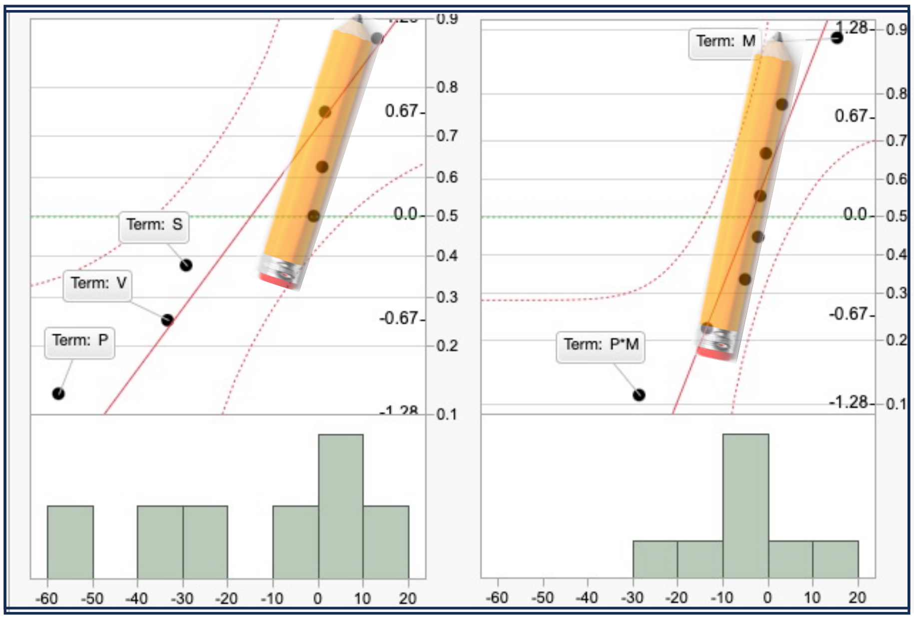


Figure F.38: Normal Plots for WP and SP

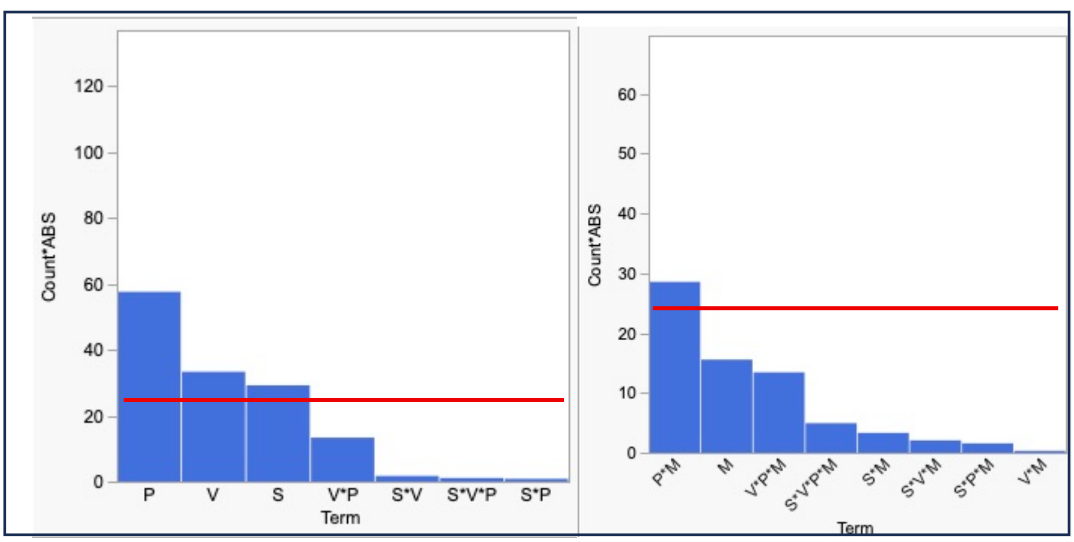


Figure F.39: Pareto Plots for WP and SP

These designs often find more statistically significant effects due to the partitioning of the noise and lowering the water level in both the whole plot and the subplot. Again, it is important to assess practical significance on the Pareto Plot.

Convenience Split-plot

While convenience split-plots do sacrifice information about the whole plot factors, they do increase the precision of the subplot effects. Without replication, the analysis for the whole plot factor is practical.

Possible Effect	*Rank*
Noise	1
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	2
Noise-by-factor Interactions	3
Simple Curvature (2nd order non-linear)	4
Complex Non-linear (≥3rd order non-linear)	5
≥ Three-Factor Interactions (3rd order linear)	4
Stability	3
Leverage	4
Measurement Uncertainty	5
Mean	5
Variation	5

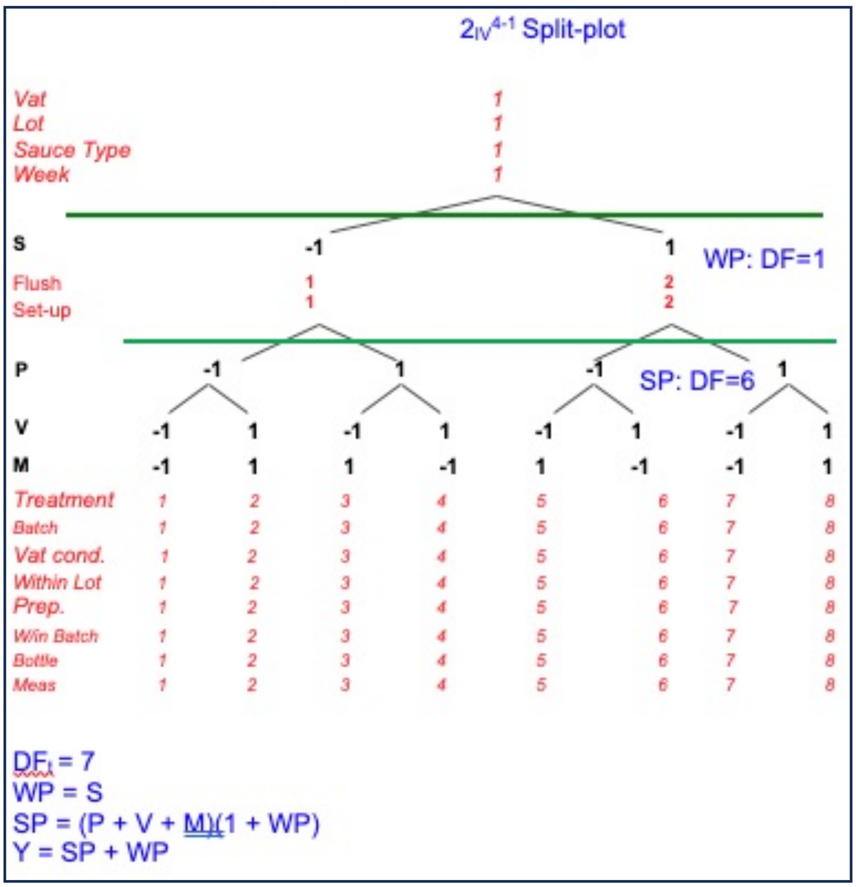


Figure F.40: Convenience Split-plot

S	V	P	M	S*V	S*P	S*M	Y
1	1	-1	-1	1	-1	-1	6385
-1	1	-1	1	-1	1	-1	6365
1	-1	-1	1	-1	-1	1	6319
-1	-1	1	1	1	-1	-1	6280
-1	-1	-1	-1	1	1	1	6266
-1	1	1	-1	-1	-1	1	6264
1	-1	1	-1	-1	1	-1	6204
1	1	1	1	1	1	1	6140

Table F.12: ANOG

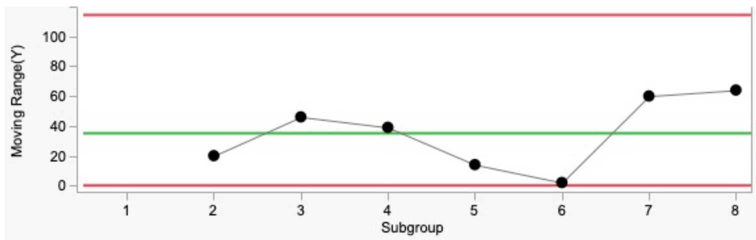


Figure F.41: MR Chart

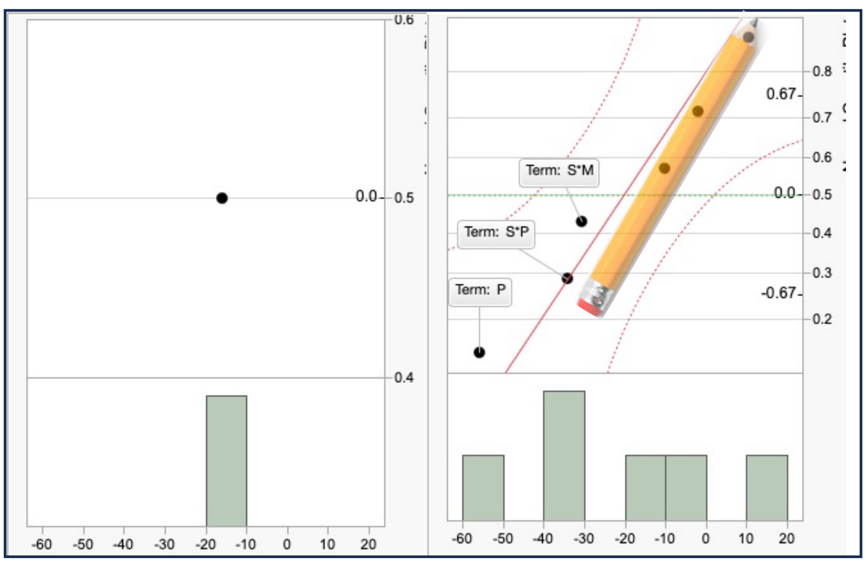


Figure F.42: Normal Plots for WP and SP

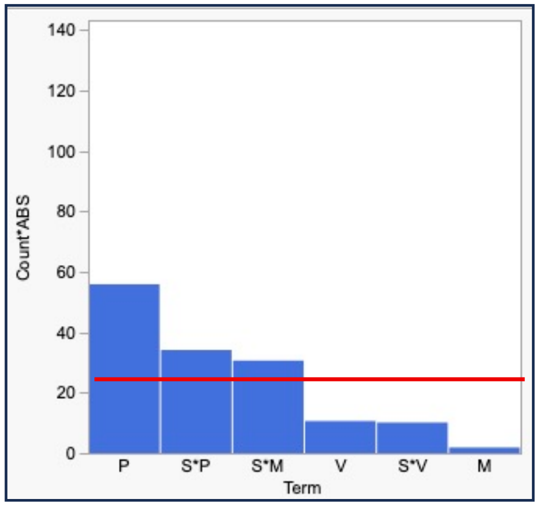


Figure F.43: Pareto Plot of the Subplot Effects

The Pareto plot for the whole plot is not shown as it contains one bar. However practical significance can still be assessed.

Appendix G
Case Study Using Ordinal Data

This appendix includes a case study involving a fractional factorial run using incomplete blocks and a response variable that is ordinal. Ordinal responses are common in engineering studies but are frequently analyzed using methods that assume continuous measurement. Because ordinal scales represent ordered categories rather than equal intervals, interpretation must focus on patterns, consistency, and practical differences rather than precise numerical comparisons.

Figure G.1: Picture of Screen Printing Process

Printed Circuit Board Screening Experiment: highlighted global effects

Possible Effect	***Rank***
Noise	1
Main Effects (Factors)	1
Two-Factor Interactions (2nd order linear)	3
Noise-by-factor Interactions	3
Simple Curvature (2nd order non-linear)	4
Complex Non-linear (≥3rd order non-linear)	5
≥ Three-Factor Interactions (3rd order linear)	4
Stability	3
Leverage	4
Measurement Uncertainty	5
Mean	5
Variation	5

In the ink screening operation, there are a number of responses that affect customer satisfaction:

- Print uniformity
- Print legibility
- Cosmetic characteristics

Although the screening operation was quite mature, the quality of the product coming from the operation seemed more a result of "magic" than from optimum operations. Operator skill & “luck” were required to make acceptable product. A cross-functional team was formed to improve the operation. The team identified over 40 factors that may affect the quality of the product associated with the screening operation. Further discussion and focus on one of the response variables led the team to filter the number of controllable factors down to 10. A scale for the response variable was set-up using ordinal (Likert[1] like) measurement categories. There are a number of actions that can be taken to improve the use of this type of response variable:

1. Involve the appropriate people in the evaluation process (e.g., the customer). ALWAYS have more than one appraiser. Assess consistency of appraiser-to-appraiser using range charts. Then use averages to evaluate the samples. Averaging will both help to reduce the assessor measurement error and expand the ordinal scale resolution. Consider the inference space,
2. Ordinal scales have no absolute value. They are measurements that rely on comparisons. Develop physical specimens as a means of direct comparison (at least 1 sample for each category). This is also useful for reducing bias amongst the evaluators,
3. Be objectively descriptive (operational definitions[2]) of each category,
4. Have each assessor measure the same sample multiple times (incorporate a nested layer with multiple evaluations of the experimental units if performing a DOE),
5. Since it is a visual inspection, consider factors that affect the measurement (e.g., light intensity, lighting source, proximity, magnification, angle, etc.). Perhaps run an experiment to reduce variation and create a consistent evaluation process,
6. Ensure bold level setting for factors in the experiment (bias on creating variation in the Y). The entire scale must be captured during the experiment,
7. Lots of x’s in the initial investigation (e.g., fractional factorials),
8. Develop alternative quantitative Y’s that may correlate with the qualitative Y.

Quantitative Y’s that may, hopefully, correlate to the main response variable to replace the qualitative evaluation were identified:

Y2: Print thickness

Y3: Print width

The 10 factors in the design structure are:

- A. Screen tension
- B. Mesh count
- C. Screen height
- D. Squeegee skew
- E. Squeegee angle
- F. Squeegee speed
- G. Squeegee hardness
- H. Material viscosity
- I. Material type
- J. Squeegee pressure

[1] Rensis Likert, (1961), *New Patterns of Management*, New York: McGraw-Hill Book Co.
[2] Deming, W. Edwards (1986) “*Out of the Crisis*” MIT Press (ISBN 0-911379-01-0)

For this example, there was an input factor to the screen-printing department, height of the surface traces, that was deemed significant by the operators in the department. This was always blamed when things went wrong. Of course, they had no control over this variable so it was noise to them. The team decided to *confound* trace height with an incomplete block. Due to the difficulty in controlling trace height, they decided to group product with similar traces together to simulate the two levels. To facilitate running the experiment, they also decided to run all product with similar traces at the same time. Notice this does not change the integrity of the design as the restriction will be handled with the block.

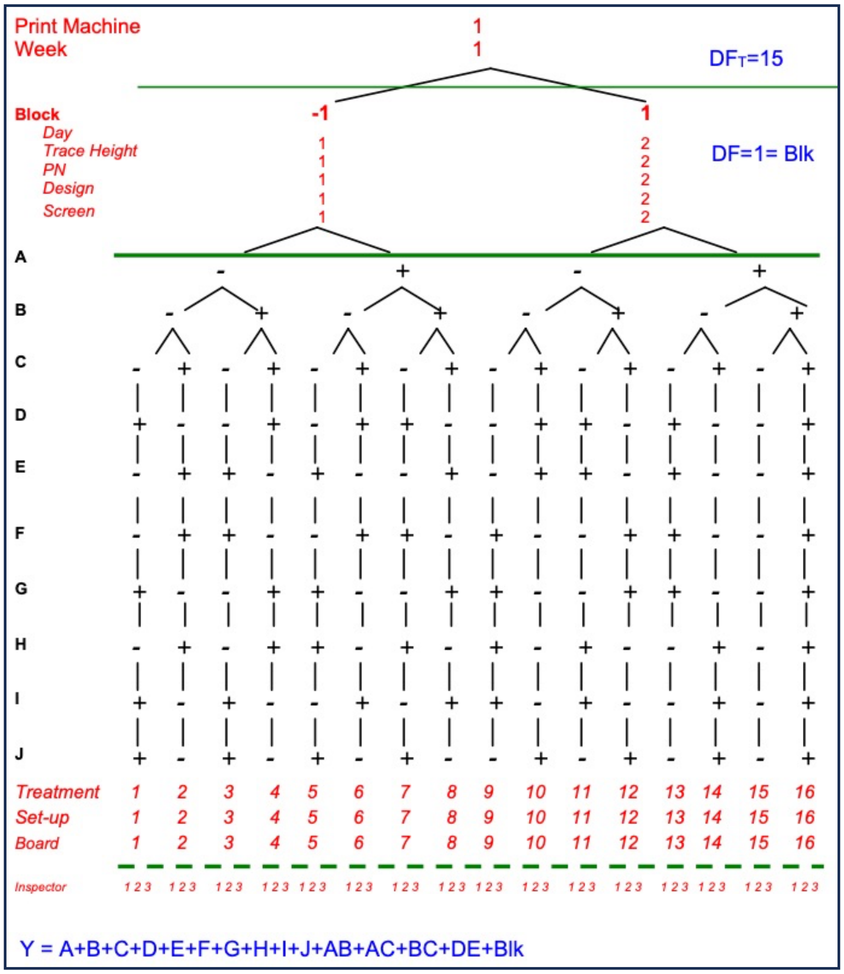


Figure G.1: FRD for Screen Printing Experiment

There are multiple Y's to evaluate. Table G.1 shows the correlation coefficients.

	Y1	Y2	Y3
Y1	1.0000	0.0303	-0.9141
Y2	0.0303	1.0000	0.0268
Y3	-0.9141	0.0268	1.0000

Table G.1: Multivariate Correlations

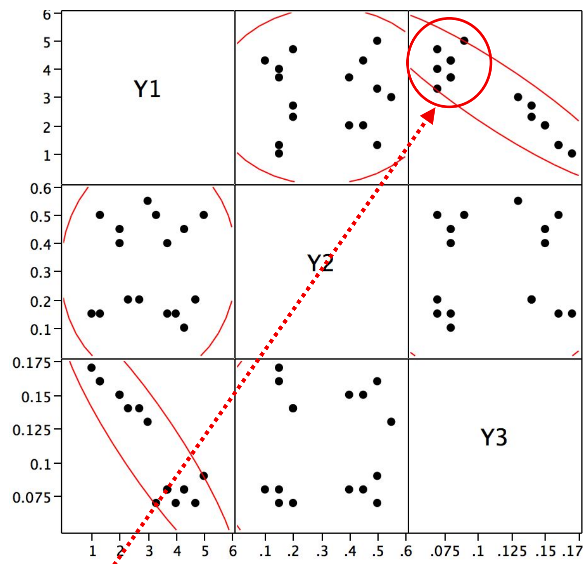


Figure G.2: Scatter Plots

Although the correlation coefficient is quite high for Y1 and Y3 (r=-.91), notice the lack of correlation when Y1 is ≥3, which makes Y3 not useful to replace Y1.

A	B	C	D	E	F	G	H	I	J	Trace	AB	AC	BC	DE	Y1
-1	1	-1	-1	-1	1	-1	1	1	-1	1	-1	1	-1	1	5
1	1	-1	-1	1	1	1	1	-1	-1	-1	1	-1	-1	-1	4.7
-1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	1	1	-1	4.3
1	-1	-1	1	1	1	-1	-1	-1	-1	1	-1	-1	1	1	4.3
1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	1	-1	-1	4
-1	-1	1	1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	3.7
1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	3.7
-1	1	1	-1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	3.3
1	-1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	1	1	3
-1	-1	-1	-1	1	-1	-1	1	-1	1	1	1	1	1	-1	2.7
-1	1	1	1	-1	1	-1	1	-1	1	-1	-1	-1	1	-1	2.3
1	1	-1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	2
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
-1	-1	1	-1	-1	1	1	-1	-1	1	1	1	-1	-1	1	1.3
1	-1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	-1	1.3
-1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	1	-1	1	1

Table G.2: ANOG

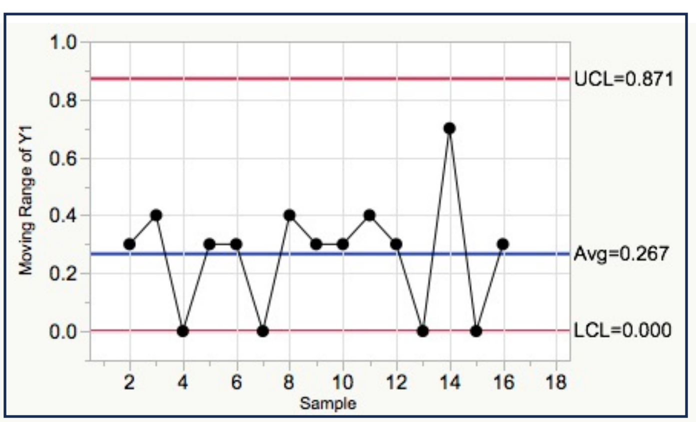


Figure G.3: MR in ANOG Order

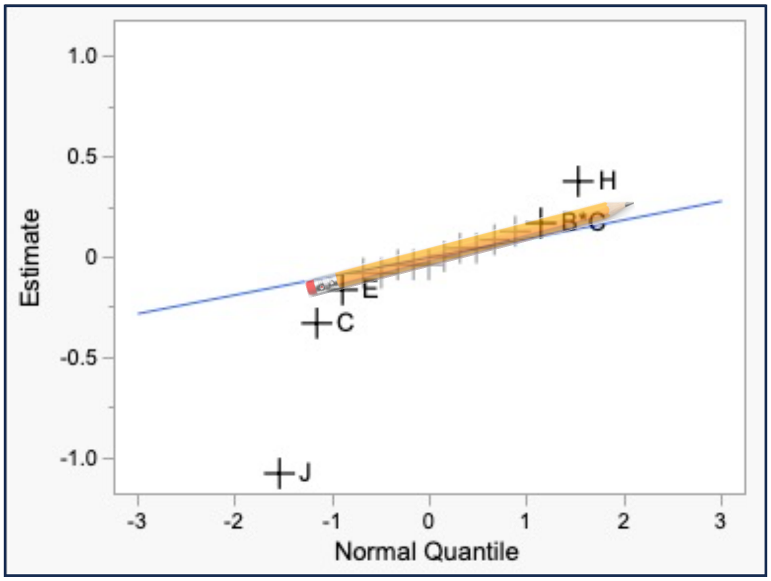


Figure G.4: Normal Plot

Term	Estimate
J	-1.083333
H	0.375000
C	-0.333333
E	-0.166667
B*C	0.166667
F	0.125000
D	-0.083333
A	0.083333
G	-0.083333
D*E	-0.041667
Trace	0.041667
A*C	-0.041667
I	0.041667
B	-0.041667
A*B	2.7756e-17

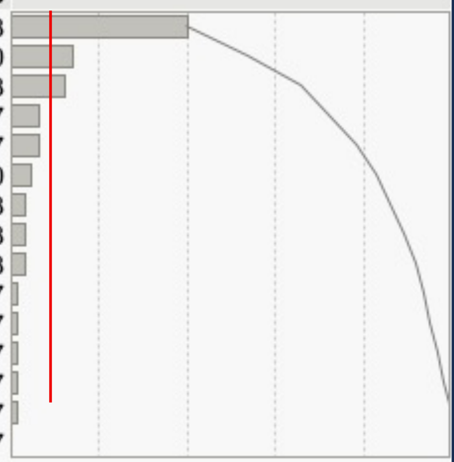


Figure G.5: Pareto Plot

After reducing the model.

RSquare	0.934016
RSquare Adj	0.91752
Root Mean Square Error	0.366414
Mean of Response	3.041667
Observations (or Sum Wgts)	16

Table 15.63: Summary of Model Effects

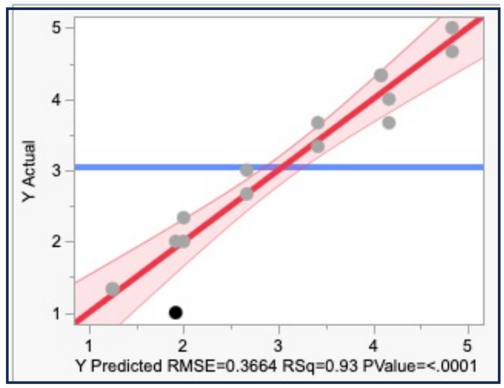


Figure 15.64: Actual vs. Predicted

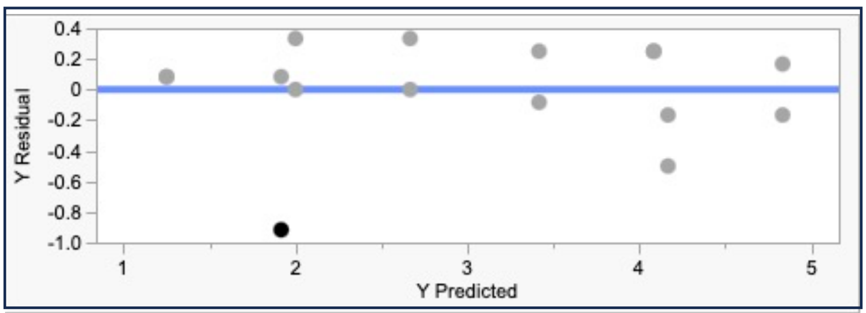


Figure 15.65: Residuals

Note the unusual residual. This occurred when the ratings were 1.

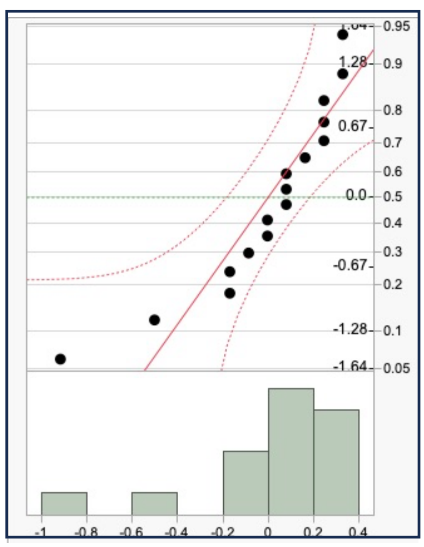


Figure 15.66: Probability Plot of Residuals

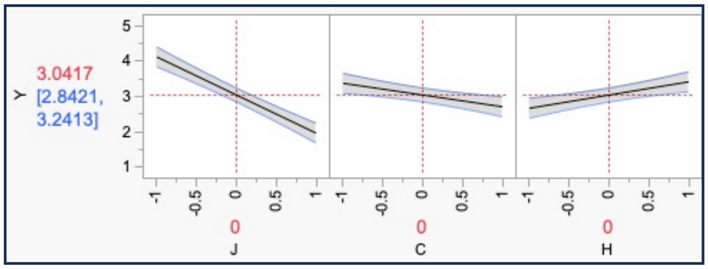


Figure 15.67: Main Effects Plot

The most interesting factor was the squeegee pressure followed by material viscosity and screen height. After considering the results, it became apparent the material viscosity was changing as solvent flashed off as a result of environmental exposure. As the material became more viscous, the operators were increasing the squeegee pressure and varying screen height to compensate. This was, however, a huge over compensation. It was decided to have the material in a sealed container that was constantly blending. The operators could then setup with the correct screen height and pressure without constant adjustment. It was also learned the trace height did not have as much effect as expected.

Appendix H
Voices from Practicing Engineers

Over the years I have asked engineers who completed Sigma Science training to reflect on their experiences and offer advice to future participants. Their responses were remarkably consistent. Although expressed in many different ways, several themes appeared repeatedly. The comments below are representative excerpts from those engineers. Together they reflect the practical realities of learning and applying the Sigma Science methodology in real engineering environments.

H.1 Advice to Engineers, from Engineers, Beginning the Sigma Science Journey

Expect the Learning Curve

Nearly every engineer commented that the early stages of the training can feel overwhelming. The methodology introduces new ways of thinking, and understanding develops gradually through repetition and application. Many engineers described the early sessions as dense and difficult, but noted that the concepts begin to connect after several iterations of applying the methods. One engineer summarized the experience well:

“The first few sessions feel like an eight-hour brain dump. After several sessions it starts to click.”

Another wrote:

“Sometimes you will feel like you understand something, and the next moment feel like you don’t know anything at all. That is normal.”

The key message repeated by many participants was simple: **be patient and stay engaged**.

Ask Questions

A consistent piece of advice from previous participants is to ask questions early and often. Many engineers noted that confusion is common during the learning process and that asking questions is the fastest way to improve understanding. Several engineers also emphasized that most participants share the same questions but hesitate to ask them. As one participant wrote:

“If you don’t understand something, ask. If you still don’t understand, ask again.”

The classroom environment is intended to support questioning and exploration rather than evaluation. Unfortunately, as is often the case, it may be challenging to even phrase a question regarding something you know little about. You may need to iterate your questions.

Expect to Fail and Learn From It

Many engineers emphasized that failure is an essential part of the learning process. Experiments, hypotheses, and early analyses often produce unexpected or inexplicable results. These outcomes are not mistakes; they are part of the discovery process. As one engineer stated:

“Do not be afraid to fail. Failure is how you learn.”

Another participant observed:

“The best way to learn a lesson is to fail miserably.”

The goal of Sigma Science is not to avoid mistakes but to **learn systematically from them**.

Repetition Is Essential

Perhaps the most common theme in the engineers' comments was the importance of repetition. Understanding the tools intellectually is not sufficient. Proficiency develops only through repeated application to real problems. One engineer summarized this clearly:

"Eventually I realized the most important thing is repetitions, using the tools over and over on different projects."

Many engineers recommended applying the methods to small problems whenever possible in order to build familiarity with the tools and reasoning process.

Start Your User Guide Early

Another strong theme was the importance of developing a personal "User Guide" during the training. Participants repeatedly emphasized that the user guide becomes an essential reference for applying the methods after the course is complete. Several engineers wrote that they wished they had started building their guide earlier. Typical advice included:

- Start the user guide on the first day.
- Continuously update it after each session.
- Focus on documenting **how and when to apply the tools**, not just the mechanics of the analysis.

One engineer summarized it simply:

"Your user guide becomes the only textbook that is actually relevant to your work."

Keep an Open Mind

Many engineers noted that the methodology often challenges familiar approaches to engineering problem solving. Participants frequently mentioned the need to temporarily set aside previous habits and assumptions. One engineer wrote:

"Come into the training with a different set of eyes. This course is not about memorizing techniques. It is about learning to think differently."

Another participant advised:

"Leave your assumptions at the door and be prepared to have your brain hurt."

Although the process can feel uncomfortable at first, engineers consistently reported that it ultimately changes the way they approach technical problems.

Work With Others

Many participants also emphasized the value of collaboration. Discussing ideas with classmates, reviewing projects together, and sharing perspectives can accelerate learning and improve understanding. As one engineer observed:

"Many minds are better than one."

Working together also helps normalize the confusion that often accompanies learning a new methodology.

Commit the Time

Finally, many engineers stressed the importance of dedicating sufficient time to the learning process.

Sigma Science requires study, practice, and iteration between training sessions. Engineers noted that it is difficult to develop proficiency without setting aside time to work on projects and review the material. One participant summarized the commitment required:

"This is hard work. There are no shortcuts."

Despite the effort required, nearly every participant concluded that the investment is worthwhile.

H.2 What Engineers Wish Management Understood

Engineers who completed the training also offered advice directed toward management. Their comments reveal common organizational challenges that affect the successful deployment of Sigma Science methods.

Understanding Requires Time

The most common message to management is that the methodology requires time and patience. Engineers frequently reported pressure to deliver immediate solutions, while the Sigma Science approach emphasizes investigation, experimentation, and iteration. As one engineer wrote:

"It requires upfront time in order to save much more time later."

Another commented:

"The goal is understanding the causal structure of the system. That takes time."

Without sufficient time for investigation, engineers are often forced to rely on quick fixes rather than genuine understanding.

Iteration Is Fundamental

Another recurring theme is that experiments and studies rarely provide final answers in a single step. Engineers emphasized that knowledge develops through sequential experimentation and refinement. One participant explained:

"It is almost impossible to run one DOE and have the answer."

Instead, each experiment contributes to a growing understanding of the system. Management support for this iterative learning process is essential.

Avoid the Quick-Fix Mentality

Several engineers expressed frustration with organizational tendencies to pursue immediate solutions rather than systematic investigation. Quick fixes may appear efficient in the short term but often fail to address underlying causes. One engineer described the problem this way:

"Too often we put a band-aid on the problem and move on, only to see the same issue return later."

Sigma Science focuses on identifying and understanding causal relationships so that problems can be solved permanently rather than temporarily.

Invest in Understanding the System

Another important message from engineers is that many organizations underestimate how little they actually understand about their processes and products. Engineers repeatedly emphasized that discovering the causal structure of a system can create significant competitive advantage. As one engineer wrote:

"Finding y = f(x) may take time, but that knowledge can become a serious competitive advantage."
This knowledge often leads to improvements in product performance, reliability, and development efficiency.

Support the Engineers Doing the Work

Engineers also emphasized the importance of managerial support during projects. This includes:

- providing time for experimentation,
- allocating resources for studies,
- removing organizational barriers, and
- encouraging learning rather than demanding immediate results.

One engineer summarized this need succinctly:
"Give engineers the time and resources to do the work."

Ask Better Questions

Finally, several engineers observed that the questions asked during project reviews strongly influence the direction of the work. Questions focused only on schedules, costs, or immediate results can unintentionally discourage scientific investigation. More productive questions focus on understanding and learning. Examples include:

- How much variation exists in the process?
- What did you learn from the experiment?
- What is the next step in the investigation?
- How will this knowledge apply to future work?

When managers ask these types of questions, they encourage deeper understanding rather than superficial answers.

Closing Reflection

Taken together, the voices in this appendix reveal an important lesson. Sigma Science is not simply a collection of statistical tools. It represents a disciplined way of thinking about engineering problems. Engineers who commit to the methodology consistently report that it changes the way they approach experimentation, analysis, and decision making. As one participant summarized:
"It doesn't just give you tools. It changes the way you think about engineering."

Bibliography

Ishikawa, Kaoru (1985) "*What is Total Quality Control: The Japanese Way*" Prentice-Hall (ISBN: 0139524339)

Ishikawa, Kaoru (1984) "*Guide to Quality Control*" Asian Productivity Organization (ISBN: 9283310357)

Mizuno, Shigeru (1988) "*Management for Quality Improvement: The 7 New Tools*" Productivity Press

Hild, C., Sanders, D. and Ross, B. (1999) "*The Thought Map*" Quality Progress, Volume 12, No. 1

Anderson, Virgil and McLean, Robert (1974) Design of Experiments, A Realistic Approach, Marcell Decker (ISBN 0-8247-7493-0)

Barton, Russell R., (1997) "*Pre-Experiment Planning for Designed Experiments: Graphical Method*s", Journal of Quality Technology, Vol. 29, No. 3, July

Box, George, Hunter, William, and Hunter, J. Stuart (1978) Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building, Wiley (ISBN 0-471-09315-7)

Box, George, Hunter, William, and Hunter, J. Stuart (2005) Statistics for Experimenters: Design, Innovation and Discovery, Wiley & Sons (ISBN 0471718130)

Box, George E. P., and Bisgaard, S. (1987) "*The Scientific Context of Quality Improvement*." Quality Progress June

Box, G., S. Bisgaard and C. Fung. *An Explanation and Critique of Taguchi's Contributions to Quality Engineering*. Report No. 28, University of Wisconsin Center for Quality and Productivity Improvement, March 1988.

Box, George E. P., and Draper, Norman (1987) Empirical Model Building and Response Surfaces, Wiley (ISBN 0-471-81033-9)

Box, G.E.P. and N. Draper (1969). Evolutionary Operation. New York: John Wiley and Sons, Inc. (SBN 471 09305 X)

Cox, David R. (1958) Planning of Experiments, Wiley (ISBN 0471574295)

Cornell, John (1990) Experiments with Mixtures, Designs, Models, and the Analysis, Wiley (ISBN: 047152221X)

Daniel, Cuthbert (1976) Applications of Statistics to Industrial Experiments, Wiley (ISBN 0-471-19469-7)

Daniel, Cuthbert and Wood, F. (1980) "*Fitting Equations to Data, Computer Analysis of Multifactor Data for Scientists and Engineers",* Wiley (ISBN: 0-471-05370-8)

Diamond, William (1981). Practical Experiment Designs. New York: Van Nostrand Reinhold Company, Inc. (ISBN 0-534-97992-0)

Gunter, Bert. "*Statistically Designed Experiments",* Quality Progress, December 1989 -August 1990.

Hicks, Charles (1973) Fundamental Concepts in the Design of Experiments, McGraw-Hill (ISBN 0-03-061706-5)

John, P. W. M. (1971) Statistical Design and Analysis of Experiments, Macmillan Company (ISBN: 0898714273)

Jones, Bradley, Christopher J. Nachtsheim (2009) "*Split-Plot Designs: What, Why, and How",* Journal of Quality Technology, Vol. 41, No. 4, pp. 340-361

Kacker, R.N. (1985). "*Off-Line Quality Control, Parameter Design and the Taguchi Method*", Journal of Quality Technology, No. 17, pp. 176-188.

Knowlton, Jeff and Keppinger, Ren (1993) "*The Experimentation Process*" Quality Progress February

Logothetis, N., "*A Perspective on Shainin's Approach to Experimental Design for Quality Improvement*", Quality and Reliability Engineering International, vol. 6, 1990, pp. 195-202.

Milliken, G and Johnson, D (1989) Analysis of Messy Data, Volumes I & II, Van Nostrand Reinhold (ISBN 0-412-99081-4 & ISBN 0-412-06371-9)

Moen, Ronald, Nolan, Thomas, and Provost, Lloyd (1998) Quality Improvement Through Planned Experimentation, 2nd Edition, McGraw Hill (ISBN: 0079137814)

Montgomery, Douglas C. (1991) "*Design and Analysis of Experiments*." Wiley (ISBN 0-471-52000-4)

Sanders, D., Leitnaker M., and McLean R. (2002) "*Randomized Complete Block Designs in Industrial Studies"* Quality Engineering, Vol. 14, Issue 1

Taguchi, G. and D. Clausing. "*Robust Quality*", Harvard Business Review, January-February 1990, pp. 65-75.

Taguchi, Genichi. Introduction to Quality Engineering: Designing Quality into Products and Processes. Tokyo: Asian Productivity Organization, 1986.

Wheeler, Donald (1987) Understanding Industrial Experimentation, SPC Press (ISBN 0-945320-09-4)

Deming, W. Edwards (1981) Japanese Methods for Productivity and Quality, George Washington University Press

Goldratt, Eliyuh M. (1984) The Goal, North River Press (ISBN 0-88427-061-0)

Goldratt, Eliyuh M. (1984) Its Not Luck, North River Press (ISBN 0-88427-115-3)

Hopp and Spearman (1996) Factory Physics, (ISBN 0-25-615464-3)

Neave, Henry R. The Deming Dimension, SPC Press (ISBN 0945320086)

Swartz, James B. (1994) The Hunters and the Hunted: A Non-linear Solution for Reengineering the Workplace, Productivity Press (ISBN 1-56327-043-9)

Wheeler, Donald (1993) Understanding Variation: The Key to Managing Chaos, SPC Press (ISBN 0-945320-35-3)

Gonick, Larry and Smith, Wollcott (1993) A Cartoon Guide to Statistics, Harper Collins (ISBN 0-06-273102-5)

Tufte, E. R. (1983), The Visual Display of Quantitative Information, Cheshire, CT: Graphics Press

Deming, W. Edwards (1986) Out of the Crisis, M.I.T. Center for Advanced Engineering Study, Cambridge, MA (ISBN 0911379010)

Box, G.E.P. and Søren Bisgaard. "*The Scientific Context of Quality Improvement*", Quality Progress, June 1987.

Fukuda, Ryuji (1986) "*Managerial Engineering: Techniques for Improving Quality and Productivity in the Workplace*" Productivity Press (ISBN: 0915299003)

Fukuda, Ryuji CEDAC (1990) "*A Tool for Continuous Improvement*" Productivity Press

Juran, Joseph and F. Gyrna (1970). Quality Planning and Analysis. New York: McGraw-Hill Book Company. (ISBN: 9780070331839)

Moen, Ronald D. and Nolan, Thomas W. (1987) "*Process Improvement: A Step-by-step Approach to Analyzing and Improving a Process*" Quality Progress September 1987

Monden, Yosuhiro (1993) Toyota Production System: An Integrated Approach to Just-in-Time, Industrial Engineering and Management Press (ISBN 0-89806-129-6)

Monden, Yosuhiro (1993) Toyota Production System: Linking the Seven Key Functional Areas, Industrial Engineering and Management Press

Ohno, Taiichi (1988) Toyota Production System: Beyond Large-Scale Production, Productivity Press (ISBN: 0915299143)

Shingo, Shigeo (1989) A Study of the Toyota Production System from an Industrial Engineering Viewpoint, Productivity Press (ISBN 0-915299-17-8)

Womack, James P., Jones, Daniel T., and Roos, Daniel (1990) The Machine that Changed the World, Rawson Assoc. (ISBN: 0892563508)

Wheeler, Donald and Lyday, Richard (1989) Evaluating the Measurement Process, SPC Press (ISBN 0-945320-06-0)

Wheeler, Donald, EMP III Using Imperfect Data, SPC Press (ISBN 0945320671)

AIAG (2002) "Measurement Systems Analysis"

Chawla, Sarita and Renesch (1995) "*Learning Organizations*" Productivity Press

Daniel, Cuthbert and Wood, F. (1980) Fitting Equations to Data, Computer Analysis of Multifactor Data for Scientists and Engineers, Wiley (ISBN 0-471-05370-8)

Myers, Raymond (1992) Classical and Modern Regression with Applications, PWS Publishers (ISBN 0-87150-946-6)

Allen, David P. and Nafius, Robert (1993) "*Dreaming and Doing: Reengineering GTE Telephone Operations*" Planning Review March

Belmonte, Rocco W. and Murray, Richard J. (1993) "*Getting Ready for Strategic Change*" Information Systems Management Summer

Clemmer, Jim (1994) "*Process Reengineering and Process Improvement; Not an Either / Or Choice*" CMA Magazine June

Cotter, John J. (1995) The 20% Solution: Using Rapid Redesign to Create Tomorrow's Organizations Today, Wiley (ISBN 0-471-13278-0)

Davenport, T. H. (1993) "*Need Radical Innovation and Continuous Improvement? Integrate Process Reengineering and TQM*" Planning Review 21(3)

Davenport, T. H. and Short J. E. (1990) "*The New Industrial Engineering: Information Technology and Business Process Redesign*" Sloan Management Review Summer

Furey, Timothy R. (1993) "*A Six Step Guide to Process Reengineering*" Planning Review 48(3)

Guha, Subashish, Kettinger, William J., and Teng, James T. C. (1993) "*Business Process Reengineering*" Information Systems Management Summer

Hammer, Michael (1990) "Reengineering work: don't automate, obliterate" Harvard Business Review July

Hammer, Michael, and Champy, J. (1993) "*Reengineering the Corporation -- A manifesto for Business Revolution*"

Hammer, Michael and Stanton, Steven A. (1995) The Reengineering Revolution: A Handbook, Harper Collins (ISBN 0-88730-736-1)

Harbour, Jerry L. (1994) Process Reengineering Workbook, Quality Resources (ISBN 0-527-76240-7)

Harrington, H. James (1995) "*The New Model for Improvement: Total Quality Management*" Management Decision. 33(3)

Harrington, H. James (1991) "*Improving Business Processes*" TQM Magazine. February

Harrington, H. James (1995) Total Improvement Management, McGraw Hill (ISBN: 0070267707)

Harrington, H. James (1991) Business Process Improvement: The breakthrough strategy for total quality, productivity, and competitiveness, McGraw Hill (ISBN 0-07-026768-5)

Butler, Cynthia and Bryce, G. Rex (1986) "*Implementing SPC with Signetics Production Personnel: A Process Flow Chart Provides a Map for Solving Problems with SPC*" Quality Progress April

Leitnaker, Mary, Sanders, Richard, and Hild, Cheryl (1996) The Power of Statistical Thinking; Improving Industrial Processes, Addison Wesley (ISBN 0-201-63390-6)

Nolan, Kevin M. (1990) "*Planning a Control Chart*" Quality Progress December

Nolan, Thomas and Provost, Lloyd (1990) "*Understanding Variation*" Quality Progress May

Wheeler, Donald, and Chambers, David (1992) Understanding Statistical Process Control, SPC Press (ISBN 0-945320-13-2)

Wheeler, Donald (1995) Advanced Topics in Statistical Process Control, SPC Press (ISBN 0-945320-45-0)

Wheeler, Donald Making Sense of Data: SPC for the Service Sector, SPC Press (ISBN 0-945320-61-2)

Shewhart, Walter A. (1931) Economic Control of Quality of Manufactured Product, D. Van Nostrand Co., NY

Barrett, Derm (1994) Fast Focus on TQM: A Concise Guide to Companywide Learning, Productivity Press

Shiba, Shoji, Graham, Alan, and Walden, David (1993) A New American TQM: Four Practical Revolutions in Management, Productivity Press

Nakamura, Shigehiro (1993) The New Standardization: Keystone of Continuous Improvement in Management, Productivity Press

National Standards for Total Quality Through Statistical Process Control. IPC-PC-90. IPC

EIA Bulletin, Guideline on the Use and Application of Cpk, QB6 (1991)

EIA Standard, Statistical Process Control Systems, EIA-557 (1989)

JEDEC Standard, General Standard for Statistical Process Control, No. 19 (1988)

Box, G.E.P., Woodall, W.H. (2012) *"Innovation, Quality Engineering, and Statistics",* Quality Engineering, 24:20-29, 2012

Ross, W. "*The Basic Seven Quality Tools*" adapted from "*Quality Tools*", Quality Progress, January 2009

Sanders, D., Ross B., and Coleman J. (1999) "*The Process Map - Revised*", Quality Engineering Vol. 11 Number 4

Ross, W. (2015), "*Continuous Improvement Initiatives: The Sigma Science Methodology*", Sigma Science Inc.

Hild, C., Sanders, D., and Ross B., (1999) "*The Thought Map*", Quality Engineering, Vol. 12, Number 1

Ross, William (2009) "*Considerations for Project Selection*", Sigma Science Inc.

Box, George (1997) "*Scientific Method: The Generation of Knowledge and Quality*", Quality Progress, January

Snee, Ronald D., (1990) "*Statistical Thinking and Its Contribution to Total Quality*", The American Statistician, Vol. 44, No.2, May

Utts, Jessica, (2002), "*What Educated Citizens Should Know About Statistics and Probability*", International Association for Statistical Education

Whitney, J.B., and J.C. Young, (1988) "*Statistical Methodology: Prerequisites for Effective Implementation in Industry*", IIQP Research Report, RR-88—01

Chance, Beth L., (2002), "*Components of Statistical Thinking and Implications for Instruction and Assessment*", Journal of Statistics Education, Vol. 10, No. 3

Box, George E. P., and Bisgaard, S. (1987) "*The Scientific Context of Quality Improvement*", Quality Progress, June

Abraham, Bovas, Brajac, M. (1994) "*Variation from Measurement Systems*", The Institute for Improvement in Quality and Productivity

Beachell, Eileen, Monda, M. (1991) "*A Primer for Enumerative Vs. Analytical Studies: Using Caution in Statistical Inferences*", Quality Disciplines

Beachell, Eileen, Monda, M. (1992) "*Managing Analytic Gaps in Industrial Research*", Quality Disciplines

Riley, Matt (2012) "*When Bad Theories Happen to Good Scientists*", Wall Street Journal, July 20, 2012

Wheeler, Don, "*Problems With Gauge R&R Studies*", 46th Annual Quality Congress

Vance, C., Groves, K., Paik, Y., Kindler (2007) "*Understanding and Measuring Linear–NonLinear Thinking Style for Enhanced Management Education and Professional Practice*", Academy of Management Learning & Education, 2007, Vol. 6, No. 2, 167–185.

Wheeler, Donald (2015) "*Rational Sampling*", Quality Digest

Wheeler, Donald (2015) "*Rational Subgrouping*", Quality Digest

Latzko, William (2001), "*The Control Chart as a Diagnostic Tool*", World Conference on Quality Improvement, Volume 55, May, P. 586-592

Escalante, Eduardo, J., (1999) "*Quality and Productivity Improvement: A Study of Variation and Defects in Manufacturing*", Quality Engineering, 11(3), 427-442

Wheeler, Don (2006) "*An Honest Gauge R&R Study*", 2006 ASQ/ASA Fall Technical Conference, No.189

Wheeler, Donald (2020) "*Gauge R&R Methods Compared: How do the ANOVA, AIAG, and EMP approaches differ*?", ASQ Statistics Division Newsletter, Vol.39, No.1

Bisgaard, Søren (2000) "*The Role of Scientific Method in Quality Management*", Total Quality Management, Vol. 11, No. 3, 295-306

Sanders, D., Hild, C., Bayle, P., Farrington, M. Sharp, B., (2001) "*Illustration of Six Sigma Assistance on a Design Project"*, Quality Engineering

Sanders, Doug and Glenn Clark (2009) "*The Role of Thought Maps to Promote Critical Thought in Business Activities*", Sigma Science Inc. Training paper

Czitrom, Veronica, (1999) "*One-Factor-at-a-Time Versus Designed Experiments*", The American Statistician, May, Vol. 53, No. 2

Sanders, D., Sanders, R., and Leitnaker, M. (1994) "*The Analytic Examination of Time-Dependent Variance Components*", Quality Engineering

Magee, Christopher, Frey, D. (2006) "*Experimentation and its Role in Engineering Design: Linking a Student Design Exercise to New Results from Cognitive Psychology*", Int. J. Engineering Ed. Vol. 21, No. 3, 2006

Ross, W. (2015) "*Effective Use of Ordinal Scales*", Sigma Science Inc.

Cochran, William "*The Philosophy Underlying the Design of Experiments*", John Hopkins University

Wood, Stacy, Lynch, J. (2002) "*Prior Knowledge and Complacency in New Product Learning*", Journal of Consumer Research, Inc. Vol. 29, December 2002

DeMast, Jeroen, (2003), "*Quality Improvement from the Viewpoint of Statistical Method*", Quality and Reliability Engineering International, 19:255-264

Balestracci, Davis (2002) "*Statistics and Reality – Part* I", ASQ Statistics Division Newsletter, Vol. 20, No. 3, 14-19

Balestracci, Davis (2002) "*Statistics and Reality – Part* II", ASQ Statistics Division Newsletter, Vol. 21, No. 1, 15-20

Hwan, Marilyn (2000), "*To Randomize or Not to Randomize, That is the Question*", ASQ Statistics Division Newsletter, Vol. 17, No. 1, 26

Abraham, Bovas, Brajac, M. (1995) "*Real Experiments, Real Mistakes, Real Learning*", The Institute for Improvement in Quality and Productivity, RR-95-05, March 1995

MacKay, R.J., "*Continuous Improvement with Control Charts*", The Institute for Improvement in Quality and Productivity, University of Waterloo

Abraham, B, MacKay, R.J "*Designed Experiments for Reduction of Variation*", The Institute for Improvement in Quality and Productivity, University of Waterloo

Sanders, D., Leitnaker M., and McLean R. (2002) "*Randomized Complete Block Designs in Industrial Studies*" Quality Engineering, Vol. 14, Issue 1

Youden, W., (1968) "*Chance, Uncertainty, and Truth in Science*", The Science Teacher, Vol. 35, No. 8, November 1968

Ross, B. (2016) "*Fundamental Strategies to Handle Noise in Experimental Situations*", Sigma Science Inc.

Sanders, D., Cooper, A. & Ross, B. (2010) "*The Role of Sampling Trees in Assisting Industrial Professionals in Developing Sampling Skills and Plans*", Sigma Science Inc.

Box, G.E.P., Patrick Liu (1999) "*Statistics as a Catalyst to Learning by Scientific Method Part I – An Example*", Journal of Quality Technology, Vol. 31, No. 1, January

Barton, Russell, R., (1997) "*Pre-Experiment Planning for Designed Experiments: Graphical Methods*", Journal of Quality Technology, Vol. 29, No. 3

Sanders, D, Hild, C., Clark, G., "*Learnings Via Sequential Applications of Designed Experiments: It's Not Just About the X's*". Sigma Science Inc.

Wheeler, Donald J., (2002), "*Good Data, Bad Data, and Process Behavior Charts*", SPC Press

Wheeler, Donald J., (2002), "*How to Make Process Behavior Charts Work for You*", SPC Press

Wheeler, Donald J., (2021), "*Statistical Process Control: Optimization and Experimentation*", ASQ Statistics Digest, Vol. 40, No. 2

Gruska, Gregory F., Maureen Heaphy, (2001) "*Designing Experiments – An Overview*", ASQ Statistics Division Newsletter, Vol. 19, No. 2

Box, G.E.P., Daniel Meyer, (1993), "*Finding the Active Factors in Fractionated Screening Experiments*", Journal of Quality Technology, Vol. 25, No. 2, April

Montgomery, Douglas, Coleman, D., (1993). "*A Systematic Approach to Planning for a Designed Industrial Experiment*", Technometrics, February 1993, Vol. 35, No. 1

Box, G.E.P., Stephen Jones (1992), "*Split-plot designs for robust product experimentation*", Journal of Applied Statistics, Vol. 19, No. 1

Shoemaker, A. C., Tsui, K. L., & Wu, C. F. J. (1991). Economical Experimentation Methods for Robust Design. Technometrics, *33*(4), 415–427.

Gupta, V.K., R. Parsad, "*Fundamentals of Design of Experiments*". I.A.S.R.I.

Montgomery, Douglas, C., (1999), "*Experimental Design for Product and Process Design and Development*", The Statistician, 48, Part 2, Pg. 159-177

Taguchi, G. and D. Clausing, (1990) "*Robust Quality*". Harvard Business Review, January-February, pp. 65-75

Abraham, B., Brajac, M. (1994) "*Variation Reduction and Robust Designs*", Institute for Improvement of Quality and Productivity

McLean, R., Anderson, V., Bayer, H., McElrath, G., (1973). "*A Scientific Approach to Experimentation for Consulting Statisticians*", Journal of Quality Technology, January, Vol. 5, No. 1

Hahn, Gerald (1977) "*Some Things Engineers Should Know About Experimental Design*", Journal of Quality Technology, January, Vol. 9, No. 1

Jones, Bradley, Christopher J. Nachtsheim (2009) *"Split-Plot Designs: What, Why, and How",* Journal of Quality Technology, Vol. 41, No. 4, pp. 340-361

Colton, James A., (2004), "*Avoiding Mean Square Error Bias in Designed Experiments*", Scientific Computing and Instrumentation, April

Leitnaker, Mary, Antony Cooper (2001) "*Using Statistical Thinking and Designed Experiments to Understand Process Operation",* Sigma Science Inc. Training Paper

Bisgaard, Søren, (2000), "*The Design and Analysis of* 2^{k-p} *X* 2^{q-r} *Split Plot Experiments*", Journal of Quality Technology, Vol. 32, No. 1, January

Bisgaard, Søren, Murat Kulahei, (2001), "*Robust Product Design: Saving Trials with Split-Plot Confounding*", Quality Engineering, 13(3), 525-530

Box, G.E.P., William Hunter, (1965) "*The Experimental Study of Physical Mechanisms*", Technometrics, Vol. 7, No. 1, February

Hinchen, J.D., (1968) "*Multiple Regression in Process Development*", Technometrics, Vol. 10, No. 2, May 1968

Hunter, Stuart (1985) "*Statistical Design Applied to Product Design*", Journal of Quality Technology, Vol. 17, No. 4, October 1985

Bisgaard, Søren, (1989), "*Quality Engineering and Taguchi Methods: A Perspective*", Target, Fall

Daniels, Beverly (2016) "Profound Statistical Concepts: When Theory Collides with reality", ASQ Statistics Division Newsletter, Vol. 35, No. 2, 2016

Bisgaard, Søren, (1996), "*Quality Quandaries Why Three –Level Designs are not so Useful for Technological Experiments*", Center for Quality and Productivity Improvement, University of Wisconsin, Product Report No. 148, August

Hahn, Gerald R., William Meeker (1984), "*An Engineer's Guide to Books on Statistics and Data Analysis*", Journal of Quality Technology, Vol. 16, no. 4, October

Bisgaard, Søren, (1991), "*Process Optimization Going Beyond Taguchi Methods*", Center for Quality and Productivity Improvement, University of WI, Report #70

Ranney, Gipsie (1999) "*Youden Address: Contexts of Statistical Practice*", ASQ Statistics Division Newsletter, Vol. 17, No. 1

Hahn, Gerald, J., (1984) "*Experimental Design in the Complex World*", Technometrics, Vol. 26, No. 1, February

Woodall, William H. (2000), "*Controversies and Contradictions in Statistical Process Control*", Journal of Quality Technology, Vol. 32, No.4 October 2000

Stefan Steiner, Abraham, B. MacKay, J. (1996) "Understanding Process Capability Indices", Institute for Improvement of Quality and Productivity

Kane, Victor (1986) "*Process Capability Indices*", Journal of Quality Technology, Vol. 18, No. 1, January 1986

Ross, B. (2016) "*Guidelines for Effective Application of Statistical Method*", Sigma Science Inc.

Gunter, Berton (1989) "*The Use and Abuse of Cpk*", Quality Progress, January 1989

Kirkpatrick, Auston (1997) "*Lean Manufacturing Principles: A Comprehensive Framework for Improving Production Efficiency*", MIT, February 1997

Wild, C.J., Pfannkuch, M. (1999) "*Statistical Thinking in Empirical Enquiry*", International Statistical Review (1999),67,3, 223-265

Sanders, D., & Coleman, J. (2003). "*Recognition and Importance of Restrictions on Randomization in Industrial Experimentation*". Quality Engineering, *15*(4), 533–543

Sanders, D., Leitnaker, M. G., & McLean, R. A. (2002). Randomized Complete Block Designs in Industrial Studies. *Quality Engineering*, *14*(1), 1–8.

About the Author

Bill Ross (statman@comcast.net) is the founder and CEO of Sigma Science Inc. His work focuses on helping engineers and scientists apply scientific method and statistical thinking to understand and improve processes and products by identifying their underlying causal structure. Early in his career he worked as an engineer supporting supplier improvement initiatives at StorageTek, working with suppliers ranging from Fortune 500 manufacturers to small specialty producers across both low- and high-technology industries. Although his academic training was in geochemistry at the University of Colorado, Ross developed his statistical expertise primarily through self-study and extensive practical application. Over more than forty years he has applied statistical and scientific methods in over 15,000 sampling plans and designed experiments across research and development, manufacturing, and business processes. Ross has studied with and worked alongside several recognized leaders in industrial statistics, including George Box, Robert Hogg, Robert McLean, Søren Bisgaard, and Genichi Taguchi, with whom he co-taught seminars in Japan. He also completed advanced coursework in analytical statistics at the University of Tennessee and the University of Wisconsin. Ross has contributed to national and international standards development, including work with EIA, IPC, ISO, and IEC technical committees focused on quality and reliability engineering. In 1992, following the recommendation of Mikel Harry, Ross founded Six Sigma Associates, becoming one of the earliest consultants dedicated to the emerging Six Sigma methodology. Through Sigma Science Inc., Ross and his colleagues continue to develop and apply the Sigma Science framework to help organizations investigate complex problems and improve technical and business systems. His teams have contributed to operational excellence initiatives in several major corporations, including Whirlpool. Ross's role in these efforts has been as teacher, mentor, protagonist, and facilitator, guiding engineers and managers in the disciplined use of reasoning, experimentation, and data. Website: sigma-science.com