# Guidelines for Effective Application of Statistical Method

for Use in Engineering & Scientific Settings

Bill Ross,
Sigma Science Inc.

## Abstract

This paper is intended to highlight some of the issues and provide recommendations regarding the appropriate application of statistical methods to the fields of engineering and science. The target audience is individuals who are working to improve processes and product designs by understanding causal relationships. This causality is annotated with the symbolic expression, Y=*f* (x). A challenge is to communicate statistical concepts and tools to the engineering community using terminology not universally understood. Definitions at the end of this paper should assist in the translation.

Questions stimulating the discussion below include: What statistics are useful? What does it mean to *apply* statistics? How does one apply statistics to their field of study? How should data be collected? How should data be analyzed? What statistical technique should be used? How are the outputs of statistical analysis interpreted? What does statistical significance mean? How can the data be used to predict?

The application of statistical thinking and methods can greatly enhance the efficiency and effectiveness of any engineering or scientific study. The following list of guidelines provides some foundational "advice" for the appropriate application of analytical statistics. The order is unimportant.

## The Guidelines

### Guideline 1: It All Depends on How the Data is Collected

Sampling may be used to draw conclusions about that which already exists (enumerative statistics[1]) or to understand causality to assist in the prediction of what may happen (analytical statistics). In both cases sampling implies not all items will be measured, only a select few (hence the word sample). Selection of the items to be measured is an **important decision** and will provide the context for data analysis. It affects not only what statistical techniques can be used for analysis, but how to interpret the results.

Dependency on how data is collected:

- The information revealed by data is entirely dependent on how the data is acquired (i.e., the sampling plan). You must first determine an appropriate sampling plan. What is: the sampling location, sampling frequency, subgroup size? Sampling schemes are designed as a function of the hypotheses stated. They link the x's that have been identified to the Y's being measured. Have sufficient x's been captured in the study to provide relevant information?

- The questions that can be answered by data are entirely dependent on the sampling plan. Always be able to answer: What do you want to know? What questions are you trying to answer? What potential sources of variation are captured in your sampling plan? Which sources are exposed within subgroup? Between subgroup? Within treatment of an experiment? Between treatments? Within block of a complete block design? Between blocks? Over what set of conditions (x's) can you make operative interpretations from the data?

- Appropriate actions are <u>entirely dependent</u> on the sampling plan. What actions are appropriate if the range chart is out-of-control? In-control? What actions are appropriate if the Y-bar (x-bar) chart is out-of-control? In-control? What if between block variation in an experiment is greater than within block? Or within level variation of a factor is greater than between level?

- Conclusions drawn from the data and the ability to extrapolate those conclusions are <u>entirely dependent</u> on how the data was acquired. Analysis is dependent on how the data was acquired. The degree of confidence in the study is a function of how representative the study is. How representative of future conditions is the study? Will the conclusions be useful in the future? What has been learned about the causal structure? Has this learning occurred over a wide inference space? Or is it subject to limitations due to restrictions of the study?

*"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. . ."*[2]

## <u>Guideline 2: Investigations are Question, not Tool Driven. Statistical Techniques Better Enable You to Answer Engineering Questions</u>

The difference between inexperienced and skilled users of statistics becomes evident in the planning for the acquisition of data. Component of Variation (COV) studies and Designed Experiments (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.

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 better yet 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 variability charts or control chart techniques. Similarly, in reliability testing, understanding long-term product performance under various conditions (e.g., representative of customer usage or varying installation situations) is one of the goals; making product reliability claims with prediction models is secondary.

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

As Sir Ronald Fisher put 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.*"[3]

**Guideline 3: Understand Variability, Variation Exists in Everything**

All measurements, provided adequate discrimination, exhibit variation. Every number obtained from the data would change somewhat, even if the measurements were repeated on the same sample. If a new lot of material is introduced, there may be an increase in variation due to the natural variability of the raw materials. If data is collected with a different set-up, from a different machine, or under different ambient conditions, there are more potential sources of variability to be accounted for. In the manufacture of coatings, batch effects (e.g., lot-to-lot variation of raw materials) may introduce extra variability. Identifying the potential sources of variability is invaluable for planning the investigation.

Understanding the amount of variability and what creates it is central to the discipline of 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.

It is recommended, if the problem is one of variation, the response variable (Y) be a measure of variation. A range, standard deviation, variance or some transformation may be useful. For experimentation this will likely mean some sort of nested[4] layer of repeats within treatment.

## Guideline 4: Plan and Predict

*"Dans les champs de l'observation le hasard ne favorise que les esprits prepares[5]"*

Roughly, chance favors the prepared mind.

Thoughtful data collection can greatly simplify analysis and make it more precise. 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; what the engineer wants to observe and draw conclusions over? 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 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 (see Thought Maps[6])
- 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 (the nature of the variability).

To assist in planning and linking the x's to the Y's, tools such as Process & Product Maps[7] have proven useful. The maps help to identify and keep track of the plethora of x's captured in any study. Asking questions and making predictions at the design stage can save headaches at the analysis stage. Prediction involves the abstraction of hypotheses to estimate at least three things:

1. The data for each instance of when the data will be collected (treatment of a DOE or sampling point in a sampling plan). This is useful to consider the reasonableness of the study and to get an estimate of your current state of knowledge (actual – predicted), essentially residuals based on engineering predictions rather than a mathematical model[8].
2. ALL possible outcomes. This is a finite set of potential outcomes, and predicting all outcomes helps to mitigate bias in interpreting data.
3. Next steps in the investigation. All scientific and engineering study is iterative. Contemplating the next course of action ensures a thoughtful plan and the appropriate preparation for the data collection process.

Sampling Trees[9] and Factor Relationship Diagrams[10] are excellent tools to assist in planning and subsequent analysis. They link the data collected to the hypotheses described on a Thought Map and the x's identified on a Process Map. They provide context for the analysis of the data.

**Guideline 5: Keep It Simple and Sequential *Stupid* [*sic*]**

Said multiple ways:

*"We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances."[11],*

*"Pluralitas non est ponenda sine necessitate."[12]*

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 do everything 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*."[13] 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 order as necessary (following the Effect Sparsity[14] and Effect Hierarchy[15] Principles and also implied by Taylor series order). Interactions among explanatory x's, nonlinear mechanisms, missing data, confounding, sampling biases, measurement error and so on, can all complicate the ability to create a simple useful model.

It is suggested 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[16]? 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*."[17] Restricting factors in a study provides no opportunity to learn about their effects and constrains the inference space (impacting extrapolation of results). Typically investigations start far from optimum and the initial work is to move in the direction of optimum. 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.

**Guideline 6: Statistical Analysis Requires Interpretation from the Lens of the Engineer or Scientist**

*"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."[18]*

Every accomplished statistical thinker:

- looks at the data for obvious patterns, outliers, special causes, etc.,
- compares the data to the predicted results, and
- then makes appropriate comparisons (e.g., for sampling, between subgroups to within subgroups of x's, for DOE, mean square of the model to mean square error estimates).

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. Once the data have been organized into an appropriate format, have a look!

"*You can observe a lot by just watching*"[19].

**Ross' Rules of Analysis**[20] suggest the following sequence, in this order, of steps to analyze data:

1. **Practical** View the data from the perspective of the engineer. Does it make sense? (e.g., practical significance, patterns, comparisons with predictions),
2. **Graphical** Use charts and plots to look for patterns (e.g., variability, control charts, normal plots, Pareto chart of effects, scatter plots, etc.), then
3. **Quantitative** Augment analysis with statistical procedures and tests (e.g., ANOVA, regression, et. al.)

Analysis for understanding causality is all about the recognition of patterns showing the association between x's and Y's. Statistical analysis is useful for two reasons:

1. Help to distinguishing difficult to see patterns, and
2. Help to prevent accepting patterns that aren't really there (i.e., random, unassignable variation).

Variability and moving range (MR) charts can reveal data quality issues and outliers. Most studies in engineering and science are evolutionary. We are trying to expand the inference of the current set of knowledge. It is this constant iteration that drives continuous improvement. It is also important to acknowledge the specific ways data are selected prior to formal analyses and to consider how such selection might affect conclusions. And to remember using a single set of data to both generate and test hypotheses is inappropriate.

A starting point for many statistical procedures is to introduce theoretical models. These models are functions of engineering hypotheses, rooted in the natural sciences (e.g., first principles, *laws* of physics). For example, an engineer might hypothesize that oven temperature (Y), may be affected by (functional relationship) cavity geometry and air flow (the result of the location and speed of fans): Expressed mathematically, Y = *f*(x1, x2) where in this case:

Y = oven temperature
x1 = cavity geometry
x2 = air flow

Rating the qualitative aspects of a product[21] (e.g., taste tests using an ordinal scale) may vary across the set of evaluators, and the statistical thinker may want to understand whether those ratings are consistent or biased. Graphs such as variability charts or a time series of the data would be useful. A control chart might be used to understand the relationship between the measurements in time series and related changes in the x's and thus provide insight into the underlying causal structure.

Using hypotheses as a basis, the sampling plan specifies the way different sets of x's get combined to create the variation in the Y being measured. Understanding what the x's are doing in the study (i.e., restricted, confounded or separated) is a fundamental step and makes statistical inferences possible. Predicting changes in response variables in terms of theoretical models has proven to be an effective simplification allowing for the variability in data to be captured in order to understand the causal structure.

Deming first differentiated enumerative from analytical problems in statistics[22]. While enumerative statistics may be useful to describe a data set that already exists, it is useless for prediction in and of itself. While enumerative statisticians have largely developed useful statistical software, the **software only provides tools to assist analyses**. The software cannot interpret the outputs, nor can it help to extrapolate those results without context. The context is critical, and the key to principled statistical analysis is to bring analytic methods into close correspondence with scientific or engineering questions and hypotheses. An engineer will likely want to consider the fundamental issue of whether the analytic technique is appropriately linked to the questions being asked. Don't turn engineering off! Engineering heuristic is vital. The outputs of the software need interpretation. How can these outputs be explained by the hypotheses? Do we need to modify, drop or add hypotheses? If there are no rational explanations, perhaps the data should be questioned.

## Guideline 7: Question and Evaluate the Integrity of the Data

The commonest of defects in designed experiments are (paraphrased from Daniel[23]):

1. Oversaturation: too many effects for the number of treatments
2. Overconservativeness: too many observations for the desired estimates
3. **Failure to study the data for bad values**
4. Failure to take into account all of the aliasing
5. Imprecision due to misunderstanding the error variance

There are two aspects of this guideline:

A. Always question the measurement system. Just because it is reporting a number does not mean the number is meaningful. Traditional gage R&R studies are ineffective at providing an answer to the most important question: Is the measurement system capable of providing insight into your hypotheses? This requires due diligence in collecting samples representative of the variation of interest. These form the basis for comparison to the measurement component of variation. Of course, as the study iterates, and the sources of variation of interest change and the question of measurement capability/adequacy returns.

B. Experienced experimenters understand instinctively when it comes to data analysis, "garbage in = garbage out." Performing diagnostics on the data set to determine the extent to which statistical assumptions are violated is imperative for proper quantitative analysis. Further effort to *organize* the data may be needed prior to analysis. This is variously called "cleaning up or scrubbing the data". Hands-on experience can be extremely useful, as data scrubbing often reveals important concerns about data integrity, what was measured was indeed what was intended to be measured (Does it match the data collection plan?) and, therefore, ensuring that appropriate analysis and conclusions are made.

It is imperative to **plan** how the data will be collected and to **predict** the possible results. Why might some data be missing, special or incomplete? Is the study reasonable? Did the data get modified through some relevant mechanism? Understanding such mechanisms can help to avoid some misleading results. For example, in a study to understand what factors affect the performance of a cooking grill, recognition of the noise due to variation in the materials being grilled is critical. When x's change, unplanned or unaccounted for, during the study they may bias the data set resulting in inappropriate analyses and conclusions.

The beauty of practical and graphical analysis is their robustness to the required assumptions for quantitative analysis. If, however you intend to use quantitative analysis techniques, the validity of the assumptions needs to be assessed. The most common quantitative statistical methods involve an assumption of normally and independently distributed residuals with a mean of zero and a constant variance, NID$(0, \sigma^2)$. This seems reasonable once you have a working mathematical model, but may not be if you don't. Randomization attempts to increase the likelihood this assumption is not violated, but randomization may make it more difficult to assign causality. So what do you do when you only have hypotheses regarding what the model might be? Quantitative analysis should be used carefully and when appropriate.

When measurements are made across time, for example, methods appropriate for analysis of the time series need to be considered. In addition to nonlinearity and statistical dependence, missing data, systematic biases in measurements, multicollinearity, and a variety of other factors can cause violations of statistical modeling assumptions. Widely available statistical software makes it easy to perform analyses **without** careful attention to inherent assumptions, and this risks inaccurate, or even misleading, conclusions. ANOVA can be used, inappropriately, to assess *p*-values on a data set that contains special cause variation (e.g., outliers). It is therefore important to understand the assumptions embodied in the methods you are using and to understand and assess those assumptions. At a minimum, you will want to check how well your statistical model fits the data. Does it make sense from a scientific/engineering perspective? Visual displays and plots of data and residuals from fitting are helpful for evaluating the relevance of assumptions and the fit of the model. Basic techniques for assessing model fit (e.g., the $R^2$ and $R^2$ adjusted[24] delta) are available in most statistical software.

## Guideline 8: Have a Strategy to Learn About and Handle Noise

"*Block What You Can, Randomize What You Cannot*" (G.E.P. Box[25])

Robustness is the quality of consistent product or process performance in the face of noise. The identification and understanding of noise is an opportunity to increase the robustness of your products. Noise must be included and varied in studies to understand robustness, not held constant.

There are a number of ways to handle noise during the data collection process. For most of them, the strategy is to partition the noise in some sensible way to allow for better precision in detecting significant effects. This needs to be done while not negatively affecting the inference space of the study. A host of techniques such as blocking, efficiency split-plots, cross-product arrays and nesting (e.g., inside treatments) are affective at accomplishing this.

Leonardo da Vinci

*"Before you make general rule of this case, test it two or three times and observe whether the tests produce the same effects."*[26]

## Conclusion

"*There are three kinds of lies: lies, damned lies, and statistics.*"[27]

It is true data are frequently misused to give arguments a false sense of significance. Knowingly misusing data or concealing important information about the way data have been obtained is, of course, highly unethical. Also insidious are the widespread instances of claims made about hypotheses based on well-intentioned yet faulty statistical reasoning. One of the objectives here has been to emphasize succinctly the problems and ways to avoid them.

The primary task for engineers is to diagnose the situation, critically think about it and apply scientific method. Statistics is a methodology, a way of thinking to assist this process. Principled statistical analysis is critical in grappling with many subtle phenomena to ensure that nothing serious will be lost in translation and to increase the likelihood findings will extrapolate into the future. To achieve full fluency in this methodology requires years of training and practice, but the hope is these guidelines will provide some essential advice.

# Definitions

Analytical Statistics (aka. Inferential statistics): the application of statistical thinking and methods together with principles and laws of the sciences to explain and predict phenomena by understanding the causal structure, thus increasing the confidence in the extrapolation of results.

ANOVA (ANalysis Of VAriance): A general term referring to a calculational procedure for allocating the amount of variation due to each effect in a factorial experiment. The usual objective is to test for differences among factor levels and/or treatment combinations.

Blocks: In industrial experimentation, blocks are frequently a frame where *noise* (x's not explicitly manipulated in the experiment), can reasonably be expected to remain constant (or are held constant) while that part of the experiment takes place. Subsequent replicates are selected so *that noise* changes *between* those blocks. In this manner, there is increased precision for the design factors and information regarding design factors is acquired across changing noise (environmental conditions, variation in raw materials, and other known and unknown noise parameters…)

Control Charts: a graphical technique used to study a process over time. Control charts are used in pairs. One is the Range chart used to determine if the ranges (within subgroup) are consistent. The other is a Y-bar or X-bar chart. These charts are used to compare the between subgroup sources of variation (x's) to the within subgroup sources (x's) to determine which source has greater leverage.

Cross Product Arrays: factorials of design factors run inside of factorial treatments of noise factors (also called inner and outer arrays)

Factor Relationship Diagram (FRD): A graphical description of an experiment showing the relationship between manipulated factors and noise. It consists of design structure, unit structure and line(s) of restriction that depict partitioning of the unit structure.

Physical Science (e.g., physics, chemistry): the study of natural phenomena with the objective of understanding causality to make useful predictions.

Sampling Plan (Sampling Tree): Graphical depiction of the procedure to acquire the units and the relationship of layers to hypotheses (thought map) and x's (process/product map).

Scientific Method: the iterative process of induction and deduction.

Split-plot Designs: a method of handling restrictions on randomization for factorial designs.

Statistics: the science of extracting information from data. This *science* includes the collection, analysis, interpretation and communication of information based on data.

Taylor series: a function often expressed as a polynomial whose order or degree increases left to right.

Treatment or Treatment Combination: A unique experimental condition in a factorial design that is defined by a specific level of each factor.

# References


1. Kass, RE, Caffo BS, Davidian M, Meng X-L, Yu B, Reid N (2016) "Ten Simple Rules for Effective Statistical Practice". *PLoS Comput Biol* 12(6): e1004961.
2. Fisher, RA (1938) Presidential address. Sankhyā 4: 14–17.
3. Hild, Cheryl, D. Sanders (2000) "The Thought Map", *Quality Engineering*, Vol. 12, No. 1.
4. Sanders, Doug, W. Ross, and J. Coleman (2000), "The Process Map", *Quality Engineering*, Vol. 11, No. 4, 2000.
5. Cox, DR, Donnelly CA (2011) *Principles of Applied Statistics*, Cambridge: Cambridge University Press.
6. Tukey, JW (1962) "The future of data analysis", *Ann Math Stat* 33: 1–67.
7. Aschwanden, C (2015) "Science isn't Broken". *FiveThirtyEight.com*, August 11, 2015
8. Yu, B (2015) "Data Wisdom for Data Science", *Departments of Statistics and EECS*, University of California at Berkeley*,* April 13 2015
9. Daniel, Cuthbert (1976) *Applications of Statistics to Industrial Experiments*, Wiley
10. Box, George, Hunter, William, and Hunter, J. Stuart (1978) *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, Wiley
11. Hahn, Gerald, N. Doganaksoy (2008) *The Role of Statistics in Business and Industry*, Wiley
12. Shewhart, Walter (1931) *Economic Control of Quality of Manufactured Product*, D. Van Nostrand Company, Inc., New York
13. Deming, W. E., (1982) *Out of The Crisis,* Cambridge, MA: MIT Center for Advanced Engineering Study
14. Wheeler, Donald and Lyday, Richard (1984) *Evaluating the Measurement Process,* SPC Press
15. Box, George (1997) "Scientific Method: The Generation of Knowledge and Quality", Quality Progress, January
16. Wheeler, Don (2006) "An Honest Gauge R&R Study", 2006 ASQ/ASA Fall Technical Conference, No.189

# Footnotes:

---

[1] Deming, W. E., (1989, October; revised 1990, April) “Foundation for Management of Quality in the Western World”. Paper presented at the meeting of Management Sciences, Osaka, Japan.
[2] Shewhart, Walter (1931) *Economic Control of Quality of Manufactured Product*, Nostrand Company, Inc., New York, p.410-417
[3] Fisher RA (1938) Presidential address. Sankhyā 4: 14–17.
[4] Hierarchical study where one layer of the sampling plan is contingent upon another
[5] Louis Pasteur, 1854
[6] Hild, Cheryl, D. Sanders (2000) “The Thought Map”, *Quality Engineering*, Vol. 12, No. 1.
[7] Doug Sanders, W. Ross, and J. Coleman (2000), “The Process Map”, *Quality Engineering*, Vol. 11, No. 4.
[8] Metric created by Bill Ross known wittily as the Ross Metric meant to quantify one’s level of knowledge
[9] Sanders, Doug, Antony Cooper, Bill Ross (2000) “The Role of Sampling Trees in Assisting Industrial Professionals in Developing Sampling Skills and Plans” Internal Sigma Science Inc. paper
[10] Sanders, Doug and Jim Coleman (1999), “Considerations Associated with Restrictions on Randomization in Industrial Experimentation”, *Quality Engineering*, Volume 12, No. 1
[11] Sir Isaac Newton,( 1687) *Principia*
[12] Occam’s razor, William of Ockham, 14th century
[13] Bill Ross
[14] Box, G.E.P.; Hunter, J.S.; Hunter, W.G. (2005*). Statistics for Experimenters: Design, Innovation, and Discovery*. Wiley. p. 208.
[15] Wu, Jeff and Michael Hamada (2000) *Experiments: Planning, Analysis, and Parameter Design Optimization*. New York: Wiley. p. 112.
[16] Deming, W. E., (1982) *Out of The Crisis,* Cambridge, MA: MIT Center for Advanced Engineering Study, Ch. 11
[17] Ross, advice given for sequential studies provided the confounding has been identified.
[18] Hahn, Gerry, Doganaksoy, N. (2008) *The Role of Statistics in Business and Industry*, Wiley
[19] Yogi Berra
[20] The order of steps for the analysis of data, Bill Ross, Sigma Science Inc. internal training documents
[21] Ross, William (2016) “Effective Use of Ordinal Scales”, *Sigma Science Inc. internal paper*
[22] Deming, W. Edwards (1975), “On Probability As a Basis For Action”, *The American Statistician*, 29(4), 1975, p. 146-152
[23] Daniel, Cuthbert (1976) *Applications of Statistics to Industrial Experiments*, Wiley, p. 205-206
[24] Coefficient of determination used to assess the amount of variation in the study explained by the model.
[25] Box, George, Hunter, William, and Hunter, J. Stuart (1978) *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building,* Wiley.
[26] Isaacson, Walter (2017) *Leonardo Da Vinci,* Simon and Schuster, New York, p. 174
[27] Benjamin Disraeli, also attributed to Mark Twain.