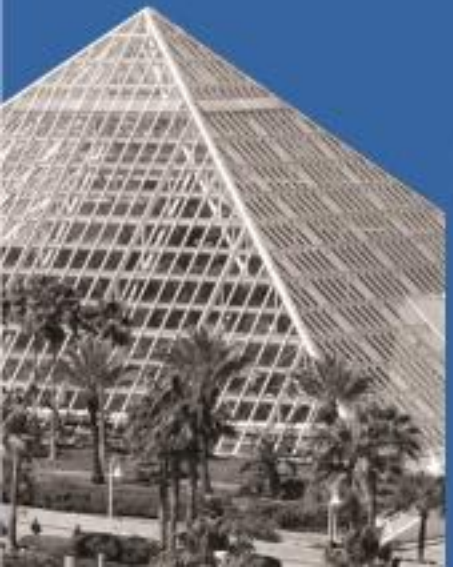
SMRP | Houston Chapter

Maintenance & Reliability Symposium

GALVESTON, TX



AUG 15-17

2018

SMRP Houston Chapter

BM2
Data Analytics as Five New-Age Questions to Ask of Maintenance and Reliability Operations

Slides available at: https://analytics4strategy.com/library

Richard G. Lamb, PE, CPA
Tel: 832-710-0755
Email: rchrd.lamb@gmail.com
Website (educational): https://analytics4strategy.com
Books:
- Availability Engineering and Management for Manufacturing Plant Performance
- Maintenance Reinvented and Business Success: Everything is about business

Richard Lamb, PE, CPA.

As the final contribution to his profession, Richard Lamb is striving to get started the new age of achievable operational excellence. The new age entails tapping into a firm's grassroot culture of systems, software and skills to become "data-driven-capable" and then weave into its data-driven functioning the insight deliverables of advanced analytics to ask and answer five essential performance questions (relationship, difference, time series, duration and apparency) we could not before.

Accordingly, Richard guides organizations, through seed operatives, managers and subject matter experts, to act upon discovering that they already have all that is necessary in their operating systems, software and skills to quickly and cost free become data-driven-capable and, in turn, weave the five questions into its data-driven operation.

This quest shares the outcome of Richard's own new-age strategy to learn and weave data science into the acumen of almost 40 years as an adviser in business strategy, finance and operations; preceded by 6 years as a manager of complex civil engineering projects.

Richard is a Certified Public Accountant, Registered Professional Engineer and Certified Lean Six Sigma Black Belt. Has published two books presenting new ideas for reliability and maintenance management framed in business strategy: Availability Engineering and Management for Manufacturing Plant Performance, and Maintenance Reinvented for Business Performance. He has a Bachelor of Science in Civil Engineering, Bachelor of Business Administration and Master of Business Administration from the University of Houston and a graduate-level Applied Statistics Certificate from the Texas A&M University, Department of Statistics.

- **We are being bombarded by exotic terminology: machine learning, artificial intelligence, big data, algorithms and mystical models (e.g., neural networks).**

- **What matters is that data analytics allow us to ask five types of questions of our maintenance and reliability operations we never could before:**

  - **Relationship: Which asset and process variables are most strongly related to a performance of interest?**
  - **Difference: How do slice-dice combinations of asset and process variables comparatively effect a performance of interest?**
  - **Time series: What are the components that underlie the summary-level-only history that operating systems are limited to providing?**
  - **Duration: What is the probability an asset or process condition will hold for some time and then what is the probability the condition will end?**
  - **Apparency: Are there hidden predictor variables to the performance of assets and processes?**

- **We maintenance and reliability operatives and experts must learn to work the questions because data scientists do not have our domain expertise.**

To separate the hype from the possibilities of data analytics for maintenance and reliability excellence, this presentation will progress as follows:

- **Separate hype from what we actually do with data analytics.**
- **For each of the five question types:**
  - **Introduce the analytic method and concept of the five question-types—why we can work the questions.**
  - **Imagine the questions we would ask in the domain of maintenance and reliability—domain specific questions.**

    Note: Domain-specific questions will be much more nuanced than what is imagined herein because domain-specific questions always reflect the site.
- **Introduce the resources you need to upskill your personal and organization's ability to step into the questions—software and literature.**

All of us know of “machine learning” and many of us have done it—coefficients of a linear regression

- It begins with a table of data we have built from data extracted and integrated from our CMMS, ERP, Excel, etc.
- The columns are “variables” and the rows are “cases.”
- For a model, some variables are selected as “predictors” and another as “outcome.”
- The chosen variables are fed to the model and its gut algorithm does trial-and-error until learning the answer—the best fit.
- Because our computer is a machine, the trial-and-error constitutes “machine learning.”
- The “answers” to the subject question lie in the details of the “fit” and what the model would predict from the fit.

We have also engaged in “artificial intelligence,” but typically without the steps of formal evaluation

- A portion of the original data set is held out from building the model.
- The largest part is fed to the model for “machine learning.”
- The hold-out set is fed to the learned model to evaluate for how closely the “trained” model estimates or calls the known actual outcome of each case.
- “Artificial intelligence” happens when the model is put to work predicting the outcome to expect from cases for which there is yet a result for the outcome variable.

Finally, let's get out of the notion of big data—you and your management will be very happy

- Everything already explained takes place in "big data"—but that does not define what big data is.
- We tend to naturally think of big data in a colloquial sense—thousands or millions of rows is "BIG" compared to what we work with in Excel.
- Media tends to title articles about the use of data analytics as "big data."
- "Big data" is when the data is so massive and unstructured it cannot be processed on a standard computer and, instead, requires high-tech systems and associated expenses.
- What you need to know is that there is little chance that you will ever play in the big data arena, but will engage in the same principles and models.

All data analytic models map to a type of question

Question Type	Answering Models	Explanatory Article (1)
Relationship	Linear, logistic and Poisson regression	Find What Matters With Relationship Questions of Operations
Difference	One-way and multi-way ANOVA, ANCOVA, repeated-measures and mixed ANOVA, and MANOVA.	Know that Improvements Work by Asking Difference Questions
Time series	Holt-Winter, linear regression, ARMA and ARIMA.	Explore What Did and May Happen with Time Series Questions
Duration	Cox regression, Cox proportional hazard, Cox mixed-effects, cumulative incidence and proportional hazard regression.	Find the Time That is Money by Asking Duration Questions
Apparency	Decision tree, regression tree, model tree and K-mean.	Dive Below the Surface of Process Functioning with Apparency Questions

(1) The articles are available as mobile friendly and pdf at the educational website https://analytics4strategy.com/new-age-five-questions

Relationships questions: Which asset and process variables are most strongly related to a performance of interest?

Without interactions (main effects)

$$\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

With polynomial and transformed variables

$$\alpha + \beta_1 X_1 + \underbrace{\beta_2 X_2 + \beta_4 X_2^2}_{\text{Variable as polynomial}} + \beta_3 Log(X_3)$$

Transformed variable

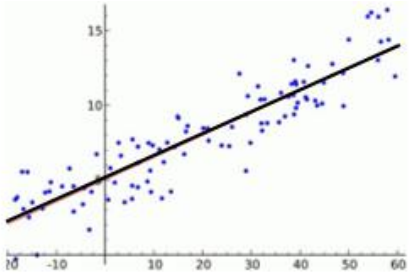
With two-way and three-way interactions

$$\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \underbrace{\beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3}_{\text{2-way}} + \underbrace{\beta_7 X_1 X_2 X_3}_{\text{3-way}}$$

There are three types of regression outcomes, but the mathematical component of "what matters" is common to all.

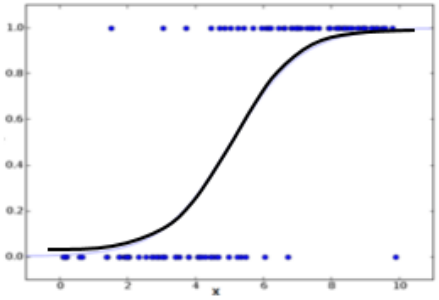
- The story of what matters is told by the coefficient β.
  - If zero does not fall within an established confidence limit, then it matters.
  - The greater (absolute) the coefficient, the stronger the relationship.
- The process of regression is to prune the variables (Xs) to the ones that matter and gain insight as we do.

We all know of linear regression, but there are three types distinguished by the nature of the outcome of interest



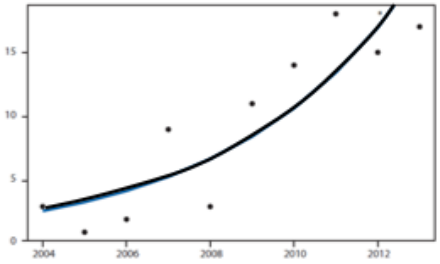
Linear regression:

- Numeric score—e.g., costs, hours, productivity and KPIs.
- Fit is linear.



Logistic regression:

- Classifies between possible categorical outcomes.
- Binomial: Yes/No and multinomial: low, middle, high.
- Classification based on the probability of an outcome between two or more possible mutually exclusive outcomes.



Poisson regression:

- Outcomes of interest are occurrences—e.g., failures and emergency orders.
- Scored as counts (failures) or rates (failures per month).

Domain-specific questions seek out "what matters" in the improvement and conduct of maintenance and reliability operations

Linear Regression:

- Which asset and process variables are most related to work order cost, hours, productivity and KPIs?
- Which variables are most related to the difference between work plan hours and actual time sheet hours—non-value add?

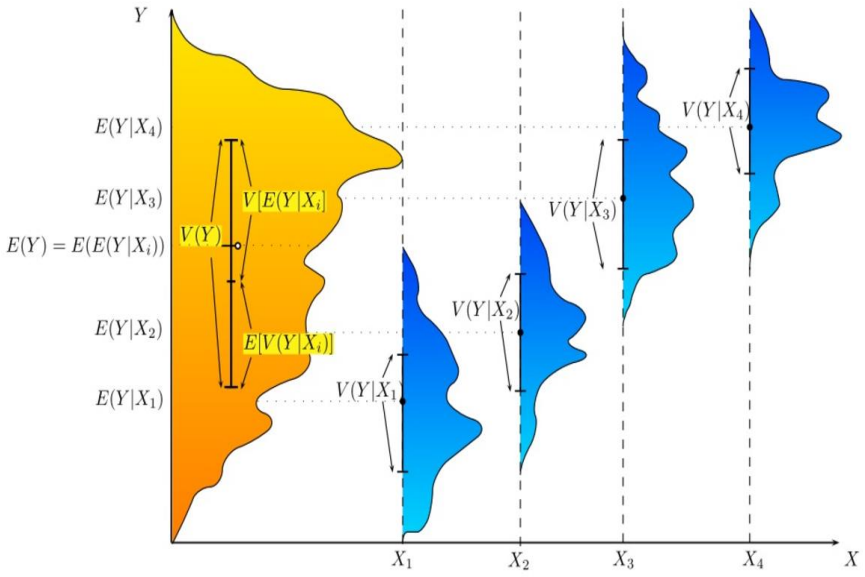
Logistic Regression:

- Which asset and process variables most foretell the probability of finding crafts and crews engaged in either value or non-value work?
- Which variables most foretell the probability of classifying a routine work order as emergency, breaker or scheduled work?

Poisson Regression:

- Which asset and process variables most seem to be present to occurring bad actors?
- Which variables are most related to the count and rate of asset failures and specific process non-compliances?

Difference questions: How do slice-dice combinations of asset and process variables comparatively effect a performance of interest?

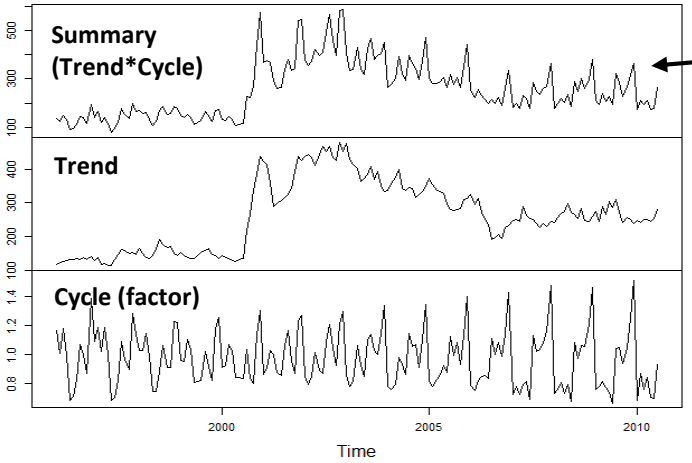


- The data set made up of all variables in a model has a mean (expected value)—left-most.
- Grouping one or more variables results in a mean for each.
- Difference models are necessary because simple pair-wise statistical comparisons are misinformation due to the math of error.
- The methods of comparison—ad hoc and contrasts—are the power of difference models.

Domain-specific questioning reveals grouping for asset and process improvements that will be visible in a performance of interest

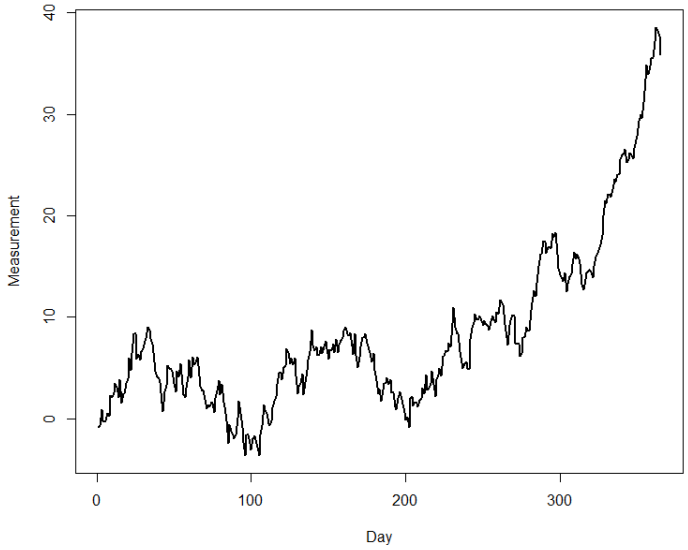
- Are costs, hours, productivity, various occurrences and KPIs different from previous periods?
- In what situations are there the largest gap (non-value) between planned and actual craft hours for work orders?
- Which cost centers geographically within and between operations comparatively excel or falter with respect to the performance of interest?
- Which differences in asset and process performance most explain the advantage and disadvantage between operating entities?
- Which improvements will most widen or close performance gaps?
- Are the improvements made to assets and processes actually working?

Time series questions: What are the components that underlie the summary-level-only history that operating systems are limited to providing?



Operating systems only collect summary-level information.

The true story lies behind the summary-level data as the history of constituent trend and cycle.



Just as important, what is the trend—deterministic, random or random walk?

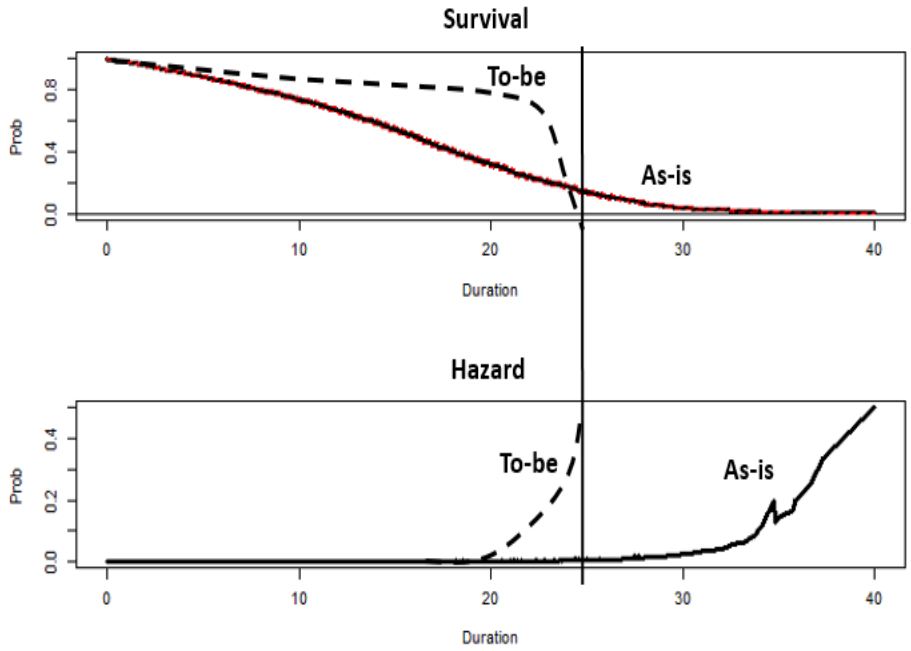
Most of us would assume this trend is deterministic—making us happy.

Times series modeling would find the trend is a random walk—we are in fools' paradise.

Domain-specific questions seek to know what happened with assets and processes, and determine the assumptions on which to plan the future

- Are any of the trends of costs, hours, productivity, failures and other events, and KPIs changing with time?
- What assumptions should we extend into the future for sustaining production assets and staying abreast of deterioration—including the effects of expanded proactive maintenance?
- What is the smoothed proactive and reactive workload on which budget and variance control, and craft and staff force will be based?
- Are there noteworthy trends in compliance to process policies—toward or away—along the maintenance process?
- Are the ratios between certain failure-types changing—proactive versus reactive?
- Can we find a KPI or operational variable that is truly a lead indicator to another of interest and what is the lead interval?

Duration questions: What is the probability an asset or process condition will hold for some time and then what is the probability the condition will end?



What is the probability of a condition lasting up to just before an ending event?

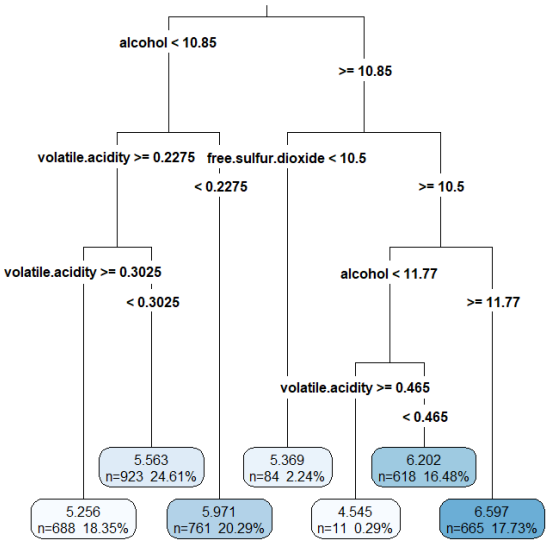
What is the probability of an ending event as a function of how long the condition has existed?

- **The shapes tell the story:**
  - **The as-is case is FIFO with a considerable degree of gaming the process and some orders stagnate in the condition.**
  - **Beside 15 days shorter, the to-be is to show very little slope until probability as which 98 percent of case survive and the probability of exit to increase non-linearly the longer a case has existed.**
- **The mathematics of "hazard" make it possible to determine which asset and process variables matter most to the story.**

Domain-specific questions orbit around the shape of the survival and hazard curves for asset reliability and process stages

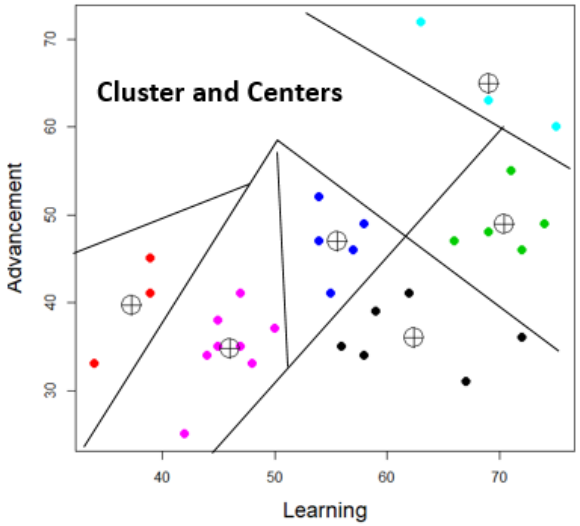
- Are the shapes of the curves acceptable for each stage from work notification to completion—do any show gaming and non-compliance to the work process?
- Is the collective duration through plan, schedule, action and return to readiness in excess of the time needed to sustain a smoothed workload for craft productivity—creating exposure to unforeseen consequences?
- Are the survival and hazard curves for spared assets statistically equal, given the run strategies?
- For condition-based monitoring, where should the P-point on the P-F curve be set—guided by the hazard curve?
- Which asset and process variables are most related to the shape of the survival and hazard curves of interest?
- Is there evidence that very old work normally hides in some stages of the backlog—distorting the perception of true backlog?

Apparency questions: Are there hidden predictor variables to the performance of assets and processes?



Decision tree:

- **Hidden Rules as variables associated with numeric and categorical outcomes.**
- **The learned rules score or classify outcomes.**



K-Means:

- **Hidden variables are learned by slicing-dicing existing predictor variables until clusters of similarity emerge.**
- **Case: Six hidden variables are teased out of two variables, but no outcome variable was given to the model.**
- **It falls to the reliability and maintenance operatives and experts to discern and classify the meaning of the clusters based on their understanding of the assets and process.**

Domain-specific questions explore for subdivisions of outcome and hidden or lost predictors

Hidden Rules:

- Is the process operating to the established rules of conduct?
- Are there situations in the maintenance process we have not before recognized as existing?
- Are outcomes for hours, costs, productivity and KPIs explained by rules?

Hidden variables:

- For a group of assets, are there uncaptured histories such as failure modes and work order types?
- Do the hidden variables point to an outcomes we have never before recognized?

Enhanced questioning:

- To which of the other domain-specific questions would the rules and revealed variables provide new predictor variables?

It is possible to enact data and questions, initially and permanently, with the following software

- Software is virtually free.
  - MS Excel
  - MS Access—most firms already have rights to the software through the MS Office license—very user-friendly tool to build tables (see article, "Building the super tables behind data-driven operations," https://analytics4strategy.com/buildsupertable).
  - The open (free to anyone or organization) top-tier applied statistics software called "R" (https://www.r-project.org/) to build models.
- Software investment is held to virtually zero until your questioning to the design and conduct of operations reveal if there is a rationale for greater systems—often there is not (see article, "First Things First; Become Capable of Data-Driven Operations," https://analytics4strategy.com/datadrivecap).

Literature for self-directed learner is plentiful to explain the principals and interpretations of the models and demonstrated with the R software

Data table skills:

- Access 2016 Bible, Alexander and Kusleika, 2016, Chapters 8 – 13.
- Pivot Tables In-Depth for MS Excel 2016, Oesko, 2017.

Articles explaining the five question as models and how to upskill individuals via upskilling projects (https://analytics4strategy.com/new-age-five-questions):

- DMAIC Done the New-Age Way
- Find What Matters with Relationship Questions of Operations
- Know that Improvements Work by Asking Difference Questions
- Explore What Did and May Happen with Time Series Questions
- Find the Time That is Money by Asking Duration Questions
- Dive Below the Surface of Process Functioning with Apparency Questions

Texts to full explanation of models:

- Relationship and difference: Discovering Statistics Using R, Field and Miles, 2012
  Multilevel Modeling Using R, Holmes, 2014
- Time series: Introductory Time Series with R, Cowpertwait and Metcalfe, 2009
  R Package "tsoutliers," Javier López-de-Lacalle, 2017
- Duration: Event History Analytics with R, Bostrom, 2012
- Apparency: Machine Learning with R, Lantz, 2015
- Optional: ggplot2, Elegant Graphics for Data Analysis, Wickham, 2016