Modeled Insight Deliverables in Data-Driven Operations

Five new-age questions to ask and answer of our operations

Richard G. Lamb, PE, CPA
Tel: 832-710-0755; Email: rchrd.lamb@gmail.com
Website (educational): https://analytics4strategy.com/

This work is licensed by Richard G. Lamb under a Creative Commons Attribution 4.0 International License (CC BY).

Agenda:

- ❑ **Preamble, purpose and approach.**
- ❑ **Essential bearings.**
- ❑ **Context variables as questions.**
- ❑ **The five new-age questions.**
  - **Relationship.**
  - **Difference.**
  - **Time series.**
  - **Duration.**
  - **Apparency.**
- ❑ **Reference library.**

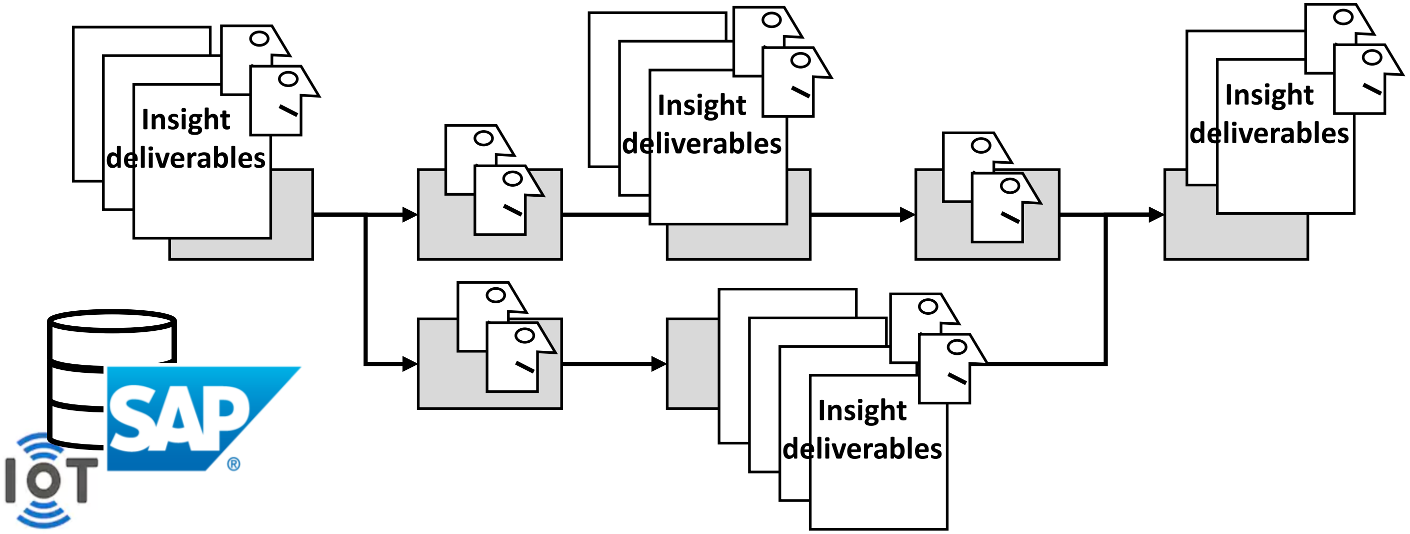
Data-driven defined:

A "Data-driven" operation is defined as one that harnesses its operational data to augment the experience and judgement of its operatives, managers, analysts and engineers as they plan, organize, conduct and control their processes.

A data-driven organization simply **improves** its processes to include all augmenting **“insight deliverables”** that will make a difference

At some places along any process, the “**best outcomes**” can only be realized when experience and judgement are augmented with “**insight deliverables**.”

> At each such place, a suite of system reports, tables, charts and models is recognized, built and worked to realize the best of outcomes.



Of four types of insight deliverables, this training session will focus on "**Modeled**" insight deliverables

Category	Description
System reports	Taken from operating systems as standard reports.
Know-thy-data	Data is explored in descriptive, graphic and statistical perspectives.
Recountive	Insight direct from data—without processing through analytics—to ask and answer questions of who, what, when, where, how much and metrics.
Modeled	**Insight gained upon data flowing through ML/AI models that ask and answer questions of relationship, difference, time series, duration and apparency.**

See

- **Know-thy-data is explained in the training session for the framework to data-drivenness and often part and parcel to modeled insight deliverables (https://analytics4strategy.com/train-frststpdtdrvnops).**
- **Recountive insight is inherent to the practices explained by the session to build super tables (https://analytics4strategy.com/train-builddatatables).**

It's not the analytics of modeled insight that are so important, but the questions we can newly ask and answer to augment the experience and judgement of role holders to a data-driven operation.

But we will not ask them unless we know to ask them.

The purpose of the session is to enable the questions to rise in your mind as you look at your processes.

Papers on data analytics with respect to the five types questions they allow to be asked and answered: https://analytics4strategy.com/new-age-five-questions

(CC BY) Richard Lamb 2019

This session will unfold as follows:

- **Essential bearing points to the discussion of modeled insight.**
- **Establish the nature of context variables as the initial question to the five modeled insights.**
- **For each of the five question types:**
  - **Introduce the analytic concepts and models to the five question-types—why we can work the questions.**
  - **Imagine representative questions we would ask in the domain of maintenance and reliability—domain specific questions.**

Agenda:

- ❑ Preamble, purpose and approach.
- ❑ **Essential bearings.**
- ❑ Context variables as questions.
- ❑ Five types of new-age questions.
    - Relationship.
    - Difference.
    - Time series.
    - Duration.
    - Apparency.
- ❑ Reference library.

(CC BY) Richard Lamb 2019

All analytic models can be classified by which five types of questions we can ask of our operations and assets

Question Type	Generic Question	Answering Model (1)
Relationship	Which context variables are most strongly related to a performance of interest?	Linear, logistic and Poisson regression
Difference	How do slice-dice combinations of context variables comparatively effect a performance of interest?	One-way and multi-way ANOVA, ANCOVA, repeated-measures and mixed ANOVA.
Time series	What are the components that underlie the summary-level-only history that operating systems are limited to providing?	Holt-Winter, series regression, ARMA and ARIMA.
Duration	What is the probability a condition will hold for some time and then what is the probability the condition will end?	Cox regression, Cox proportional hazard, Cox mixed-effects, cumulative incidence, proportional hazard regression, Weibull and Crow-AMSAA.
Apparency	Are there hidden context variables to the performance of assets and processes?	Decision tree, regression tree, model tree, naïve Bayes K-mean and principle component analysis, MANOVA.

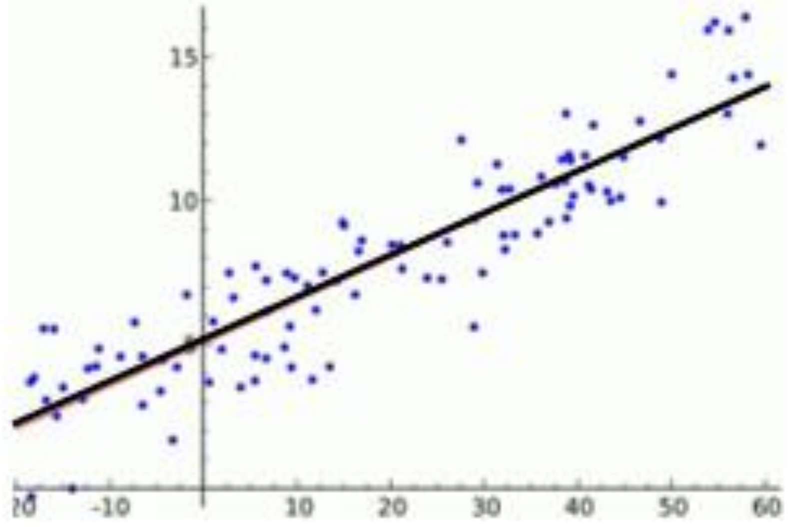
Note: The term "context variables" reflects that any aspect of the universe of the performance of interest is available as variables much wider than operational structure, process and assets.

(1) A set of explanatory papers to the models can be downloaded at https://analytics4strategy.com/new-age-five-questions

(CC BY) Richard Lamb 2019

Let's establish the place of **machine learning**, **artificial intelligence** and **algorithms**—using regression analysis as a frame of reference

- **Machine learning** takes place when we feed the data of variables to the set-up regression and its gut **algorithm** conducts a trial-and-error calculation until "learning" the best fit.
- Most often our interest is with the returned coefficients for each variable and other inferences telling us how much, if any, the variable plays in predicting the outcome.
- In contrast, **Artificial intelligence** feeds new data to the fitted model of coefficients to predict or forecast outcomes upon the "learned" coefficients.
- AI does not distinguish the model—all model types entail machine learning and most can be deployed as AI.



Until recently we just called everything a model—and still should.

When the model is intended to be used as AI, the learning process will entail an additional stage

- A portion of the original data set is held out from building the model—e.g., one third—to be a test set.
- The remaining portion is fed to the model for learning.
- The test set is fed to the learned model to evaluate for how accurately the “trained” model estimates or calls the actual outcome of each case in the test set.
- If accuracy is acceptable to the intended use, model is deployed to make the learned judgement—or augment human judgement.

(CC BY) Richard Lamb 2019

Condition-based maintenance and IIoT must not be mistaken as the whole of data-driven operations—but to understand their place it is necessary to understand the framework of data-driven operations

- **Condition-based maintenance (CBM)** is a taking data through mobile and fixed sensors, and analyticis revealing changes to an asset's performance upstream (potential failure) to the functional failure we are watching for.

- **IIoT-based** CBM automates or makes continual the monitoring, data processing and insight—requiring a great deal of infrastructure.

- IIoT can be applied to any operational process by placing sensors on the process and building the infrastructure to process the data and generate insight.

- The design issue is whether an IIoT solution is "worth it" for a particular case of CBM that, without consideration of IIoT, has already been found to be technically feasible and worth it?

- The reality is that for non-CBM insight, almost all modeled operational insight is technically feasible, but does not require an IIoT-based solution to get at it—an IIoT-based solution is not worth it—and the "sensors" are our operational systems.

Note:
There is a distinction between on-condition and condition-based as one of four ways to seek potential failures to a functional failures.

(CC BY) Richard Lamb 2019

Agenda:

- ❑ Preamble, purpose and approach.
- ❑ Essential definitions.
- ❑ **Context variables as questions.**
- ❑ Five types of new-age questions.
  - Relationship.
  - Difference.
  - Time series.
  - Duration.
  - Apparency.
- ❑ Reference library.

(CC BY) Richard Lamb 2019

Almost all models are based on the same structure as a formula

"~" equivalent to "="

Performance of interest.

"X's" are the context variables we select from our super table (https://analytics4strategy.com/train-builddatatables).

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

"β's" reveal the strength and direction of effect on the outcome.

"α" is the intercept to the outcome dimension.

- The measure of how well a model allows us to question our operations is how much of the variance around Y is explained by the to the terms of the variables.
- Too few variables and much of the variance is left unexplained; too many and it is difficult to interpret the findings.

(CC BY) Richard Lamb 2019

In statistic-speak, the "X" terms are called predictor variables, but the slides will call them context variables

- **Variables in a model can be anything from anywhere.**
  - **Financial.**
  - **Operation.**
  - **Process.**
  - **Asset.**
  - **Environment.**
  - **Seasonal.**
  - **Cultural.**
- **Whatever they are, they make up the context of the performance of interest.**

(CC BY) Richard Lamb 2019

Upon our subject-matter expertise in the subject of analysis, we determine which context variables are practically significant to the performance of interest

Exam ~ $\alpha + \beta_1 Anxiety + \beta_2 Revise$

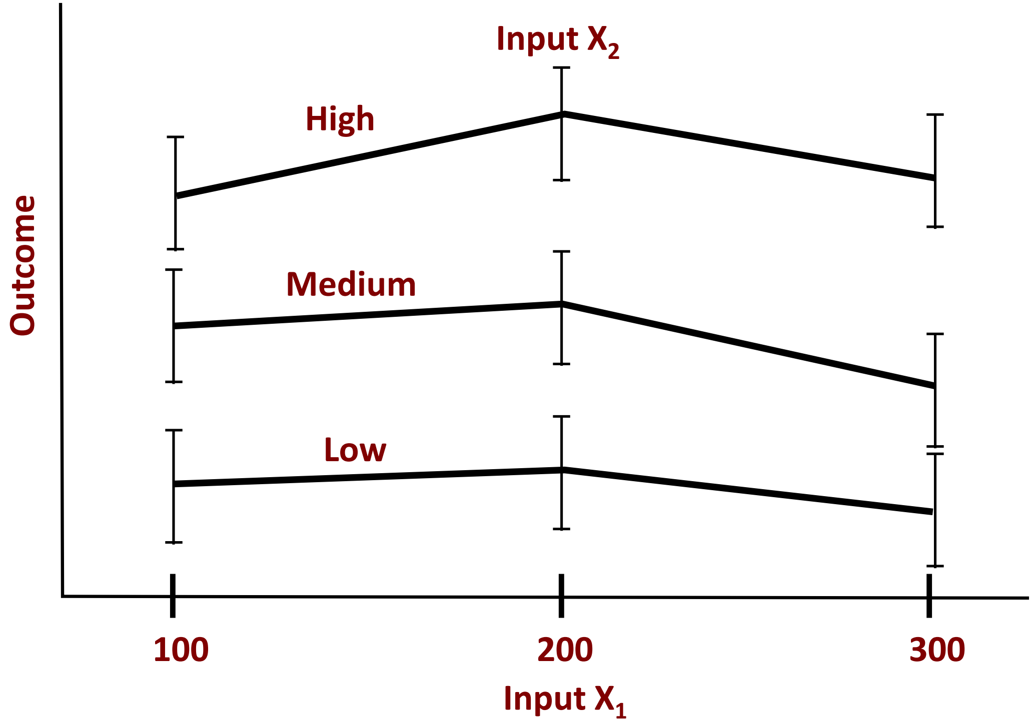
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  87.8326    17.0469   5.152  1.3e-06 ***
Anxiety      -0.4849     0.1905  -2.545   0.0124 *
Revise        0.2413     0.1803   1.339   0.1837
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Revise time is NOT practically significant—p-value >0.05—predictor of exam performance, but Anxiety is.

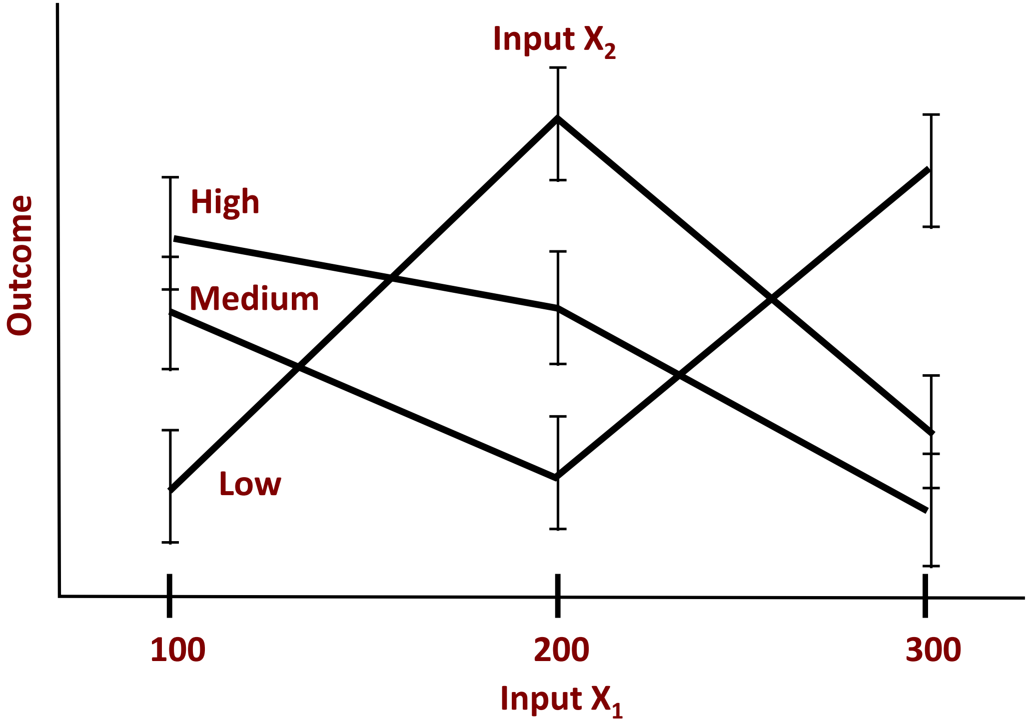
(CC BY) Richard Lamb 2019

It is important to determine if there is interaction between context variables—the effect of a context variable on the outcome depends on the value of one or more other context variables in the formula

Shows insignificant interaction



Shows interaction



(CC BY) Richard Lamb 2019

Recognizing all interactions between context variables becomes very difficult very quickly—something that cannot be readily recognized by a simple graphic such as the previous slide

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

To the initial formula we have added three terms as all 2-way combinations and a 3-way term.

(CC BY) Richard Lamb 2019

Using the model itself, we inspect our context variables for interaction —possibly something we were not aware of

$$\text{Exam} \sim \alpha + \beta_1 Anxiety + \beta_2 Revise + \underbrace{\beta_1 Anxiety \boxed{*} \beta_2 Revise}$$

- We will be made aware of whether or not there are interactions for each term coded into the formula.
- The interaction Anxiety:Revise is not significantly relevant—p-value is >0.05.
- As descending X-ways interactions are found to not be significant, they are eliminated from the model and the model rerun.

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)    100.959573  19.031389   5.305 6.89e-07 ***
Anxiety         -0.683887   0.230505  -2.967  0.00377 **
Revise           0.139686   0.309041   0.452  0.65226
Anxiety:Revise   0.007343   0.004854   1.513  0.13352
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the fit of the model is evaluated, we may discover there are non-linear dynamics between context and outcome variables—possibly something we were not aware of

Transformed variable

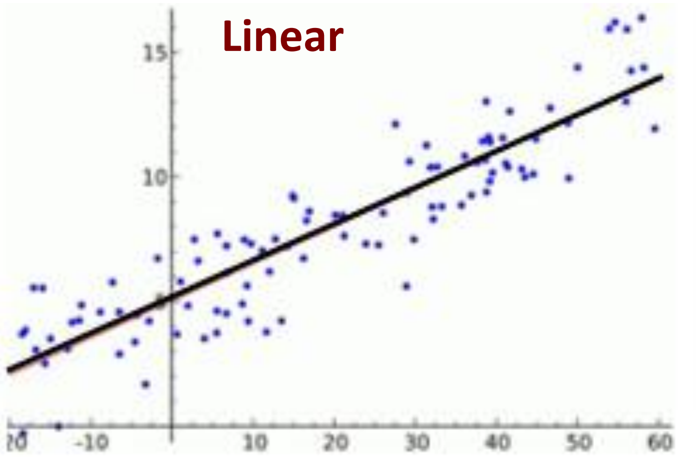
$$\boldsymbol{Y \sim \alpha + \beta_1 X_1 + \underbrace{\beta_2 X_2 + \beta_4 X_2}_{\text{Variable with polynomial relationship to outcome}} + \beta_3 Log(X_3)}$$

(CC BY) Richard Lamb 2019

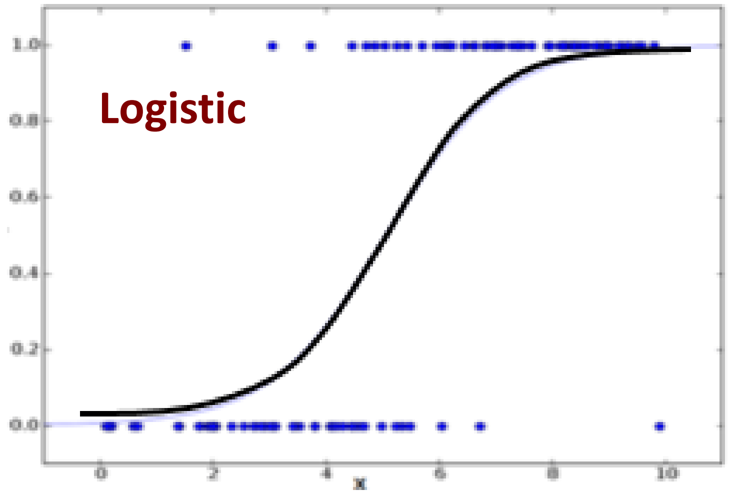
Agenda:

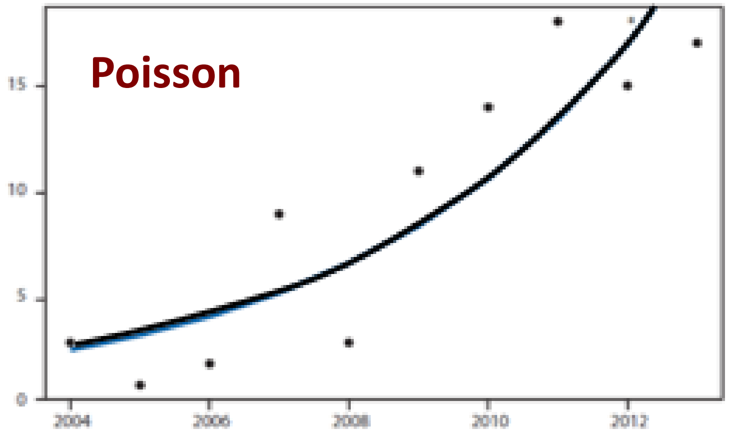
- Preamble, purpose and approach.
- Essential definitions.
- Context variables as questions.
- Five types of new-age questions.
  - **Relationship**.
  - Difference.
  - Time series.
  - Duration.
  - Apparency.
- Reference library.

Relationships questions: Which context variables are most strongly related to a performance of interest?



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





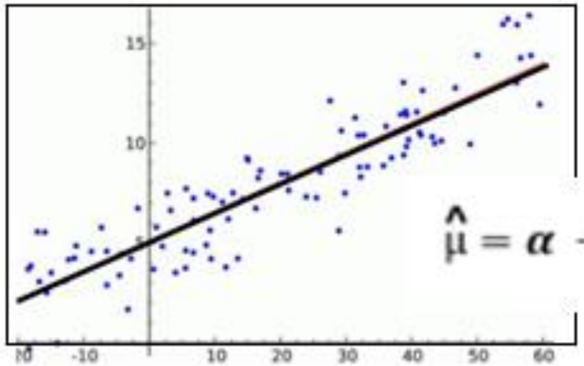
Note:
We can visualize two-dimensional models, but need analytic read-outs to understand three or more.

See

See paper, "Find What Matters with Relationship Questions of Operations," https://analytics4strategy.com/relatqstoci

For completeness, know that the context side of the formula is the same, but the outcome side (Y) of is a linking expression from which the answer is computed of the machine-learned model

Linear Regression



$$\hat{\mu} = \alpha + \beta X$$

Without interactions (main effects)

$$\alpha + \beta_1X_1 + \beta_2X_2 + \beta_3X_3$$

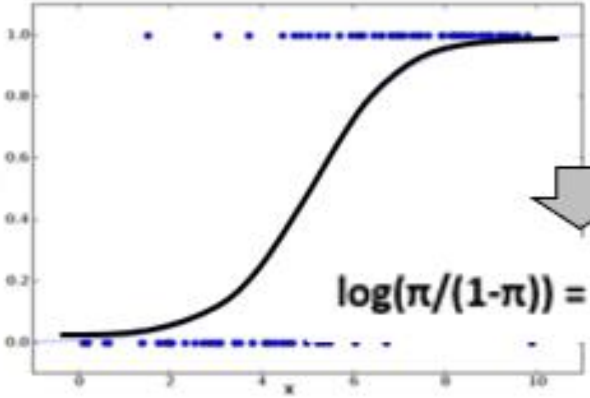
With polynomial and transformed variables

$$\alpha + \beta_1X_1 + \underbrace{\beta_2X_2 + \beta_4X_2^2}_{\text{Variable as polynomial}} + \overset{\text{Transformed variable}}{\beta_3Log(X_3)}$$

With two-way and three-way interactions

$$\alpha + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \underbrace{\beta_4X_1X_2 + \beta_5X_1X_3 + \beta_6X_2X_3}_{\text{2-way}} + \underbrace{\beta_7X_1X_2X_3}_{\text{3-way}}$$

Logistic Regression



$$\log(\pi/(1-\pi)) = \alpha + \beta X$$

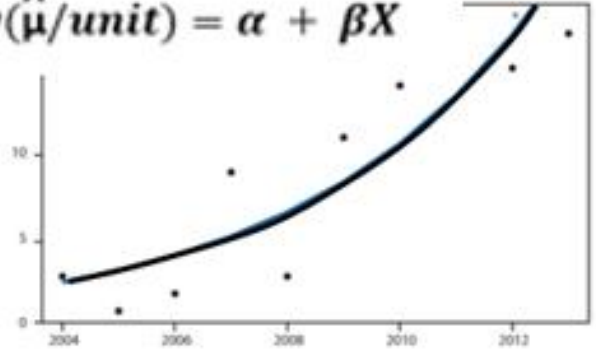
Poisson Regression

- Count

$$Log(\hat{\mu}) = \alpha + \beta X$$

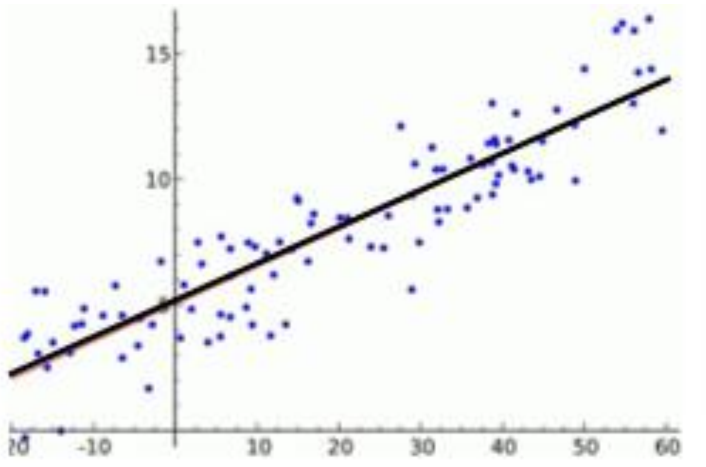
- Rate

$$Log(\hat{\mu}/unit) = \alpha + \beta X$$



(CC BY) Richard Lamb 2019

The questions asked with linear regression help us isolate what matters most to realizing standards of operational and asset performance



Linear regression:

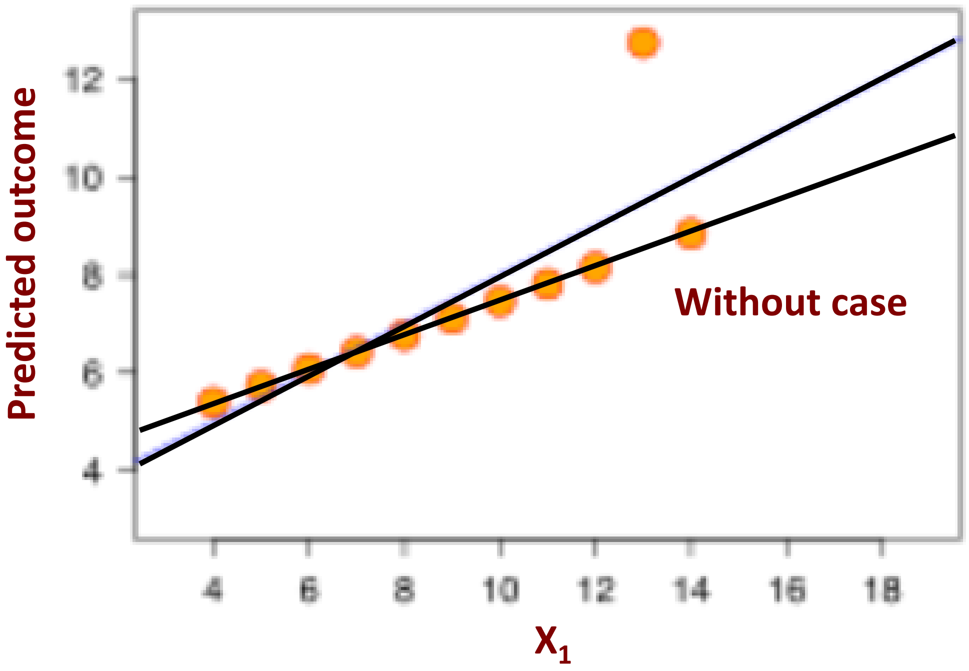
- Numeric score—e.g., costs, hours, productivity and KPIs.
- Fit is linear.

Example maintenance and reliability questions:

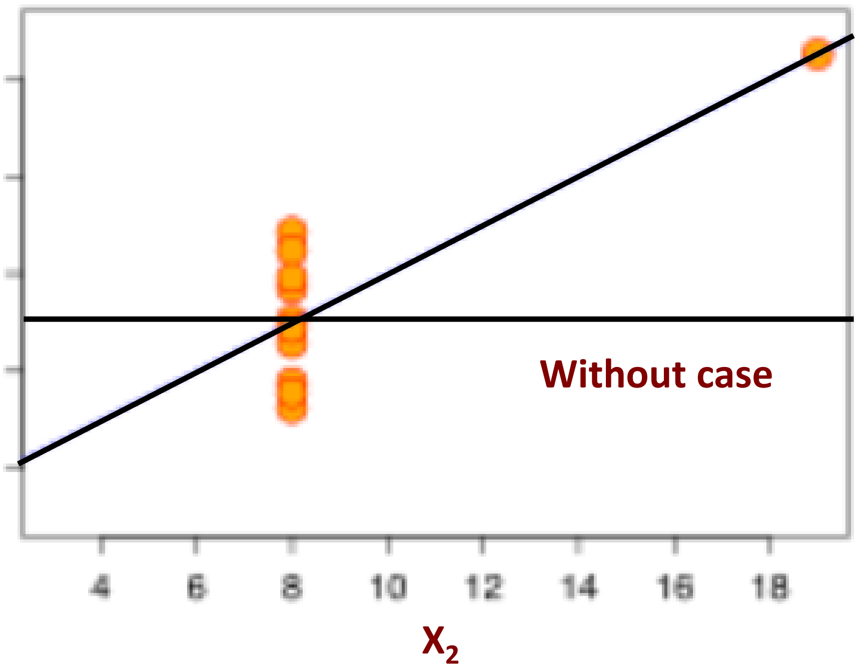
- Which context variables are most related to work order cost, hours and productivity?
- Do the components to maintenance KPIs directly relate, as thought, to enterprise and operational performance?
- Given the variables to a work order to be planned, what is the expected and range of cost and if, once planned, is out of range?
- Which variables are most related to the difference between work plan hours and actual time sheet hours?
- Are there variables with which it is possible to flag pending shifts in operational performance due to stresses of anticipated business cycle?

With the linear model, we can seek hidden outliers in the context of a modeled performance of interest—unexpected and overly influential—using tests on the model at the time of its validation

Unexpected



Influence



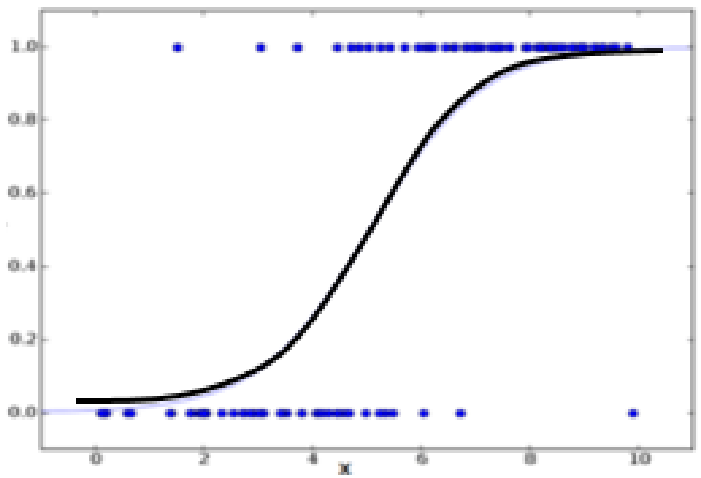
Outlier because if model were applied, the case would never be reasonably predicted.

Outlier because the case is so extreme in its location that it significantly decides the slope of the curve.

Note: The example is two-dimensional, allowing visual recognition, but the analytic tests of validity are needed to spot the outlier case for two or more context variables.

(CC BY) Richard Lamb 2019

Not all performances of interest present numerically, thus, questions that classify outcomes require logistic regression



Logistic regression:

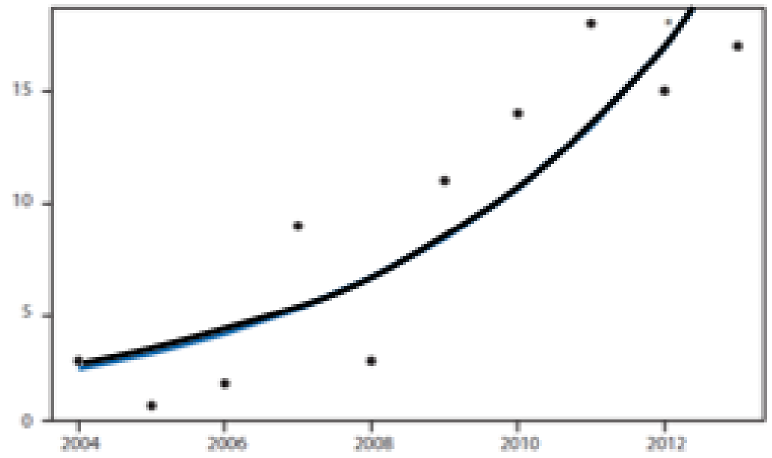
- Classifies between possible categorical outcomes.
- Binomial: Yes/No and multinomial: low, middle, high.
- Classification based on the probability of an outcome between two or more possible mutually exclusive outcomes.

Example maintenance and reliability questions:

- Which context variables most foretell the probability of finding crafts and crews engaged in value or non-value work?
- Which variables most foretell the probability that routine work orders will be classified as emergency, breaker or scheduled work?
- Are some classifications in the month's data questionable such that we want to investigate for correct classifications?
- Are there variables with which it is possible to flag an asset failure through loss of an outcome functional performance?

(CC BY) Richard Lamb 2019

The Poisson regression allows us to better understand the occurrence of events and understand the context variables which play in the counts and rates



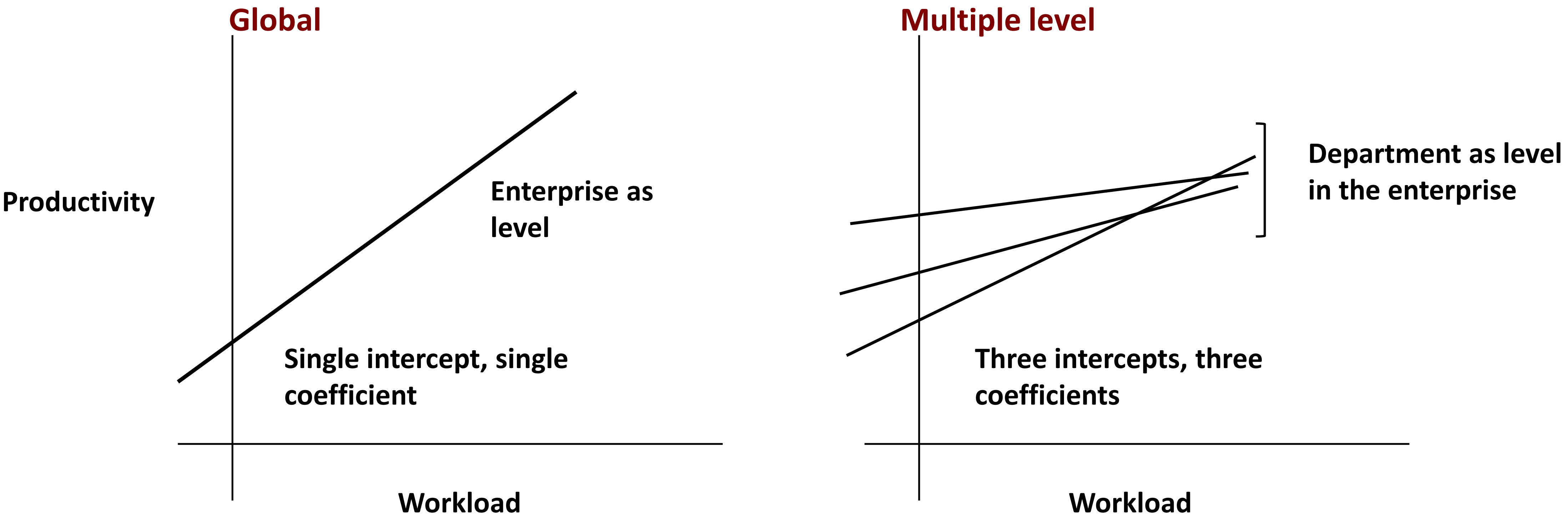
Poisson regression:

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

Example maintenance and reliability questions:

- Which organization, process and asset variables most seem to be present to occurring bad actors?
- Which variables are most related to the count and rate of asset failures, functional failures and process non-compliances?
- Does over-dispersion of the model suggest that there are variables related to the counts that we are not aware of?

With multilevel regression we can question whether outcomes of linear and logistic regression apply across the across the enterprise as if a single level are dependent on organization levels or groupings



- Shown depicts the linear regression case, but also applies to the logistic case.
- We are acting upon misinformation if we accept the left-most model, with the right-most model represents the true case.

(CC BY) Richard Lamb 2019

To this point we've assumed that the questions we ask of our operations do not significantly vary as the categories or numeric range of context variable fall into distinctive groups

- **Question: Do the monthly KPIs associated with the plant's cost centers show significantly different strength and direction of relationship to craft productivity?**

  For these cases we must model for random measures as compared to "fixed" as shown to this point.

- **Question: Are the plant's KPIs actually associated with plant unplanned downtime?**

  For these cases we build a mixed model—fixed and random.

  As compared to the membership of KPIs to cost center, downtime "may be" independent of cost center.

- **Question: Which variables are independent (fixed) and dependent (random) in the envisioned consequences?**

  The approach is to start with a basic model, . .

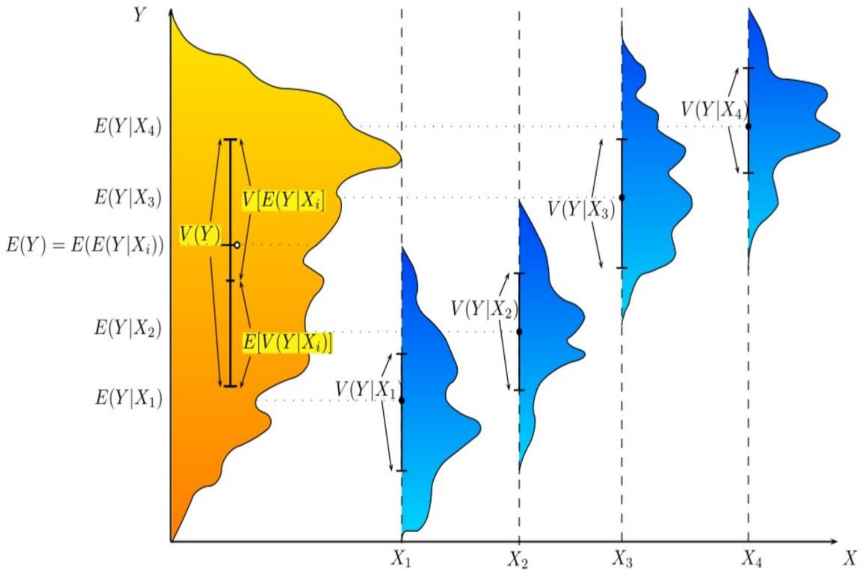
  structure a suspected random context variable in the model and compare its fit to the basic model. . .

  add one context variable at a time as a level to the KPI and compare to previous model until there is no further improvement in fit.

Agenda:

- ❑ Preamble, purpose and approach.
- ❑ Essential definitions.
- ❑ Context variables as questions.
- ❑ Five types of new-age questions.
  - Relationship.
  - **Difference.**
  - Time series.
  - Duration.
  - Apparency.
- ❑ Reference library.

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



- The data set made of all context variables in a model has a single mean as its "basic model."
- Grouping on context variables results in a mean for each such group.
- Difference models are necessary because simple pair-wise statistical comparisons of means will be misinformation due to the familywise math of error.

Actual confidence is 0.95^4 = 0.81 < Desired confidence = 0.95

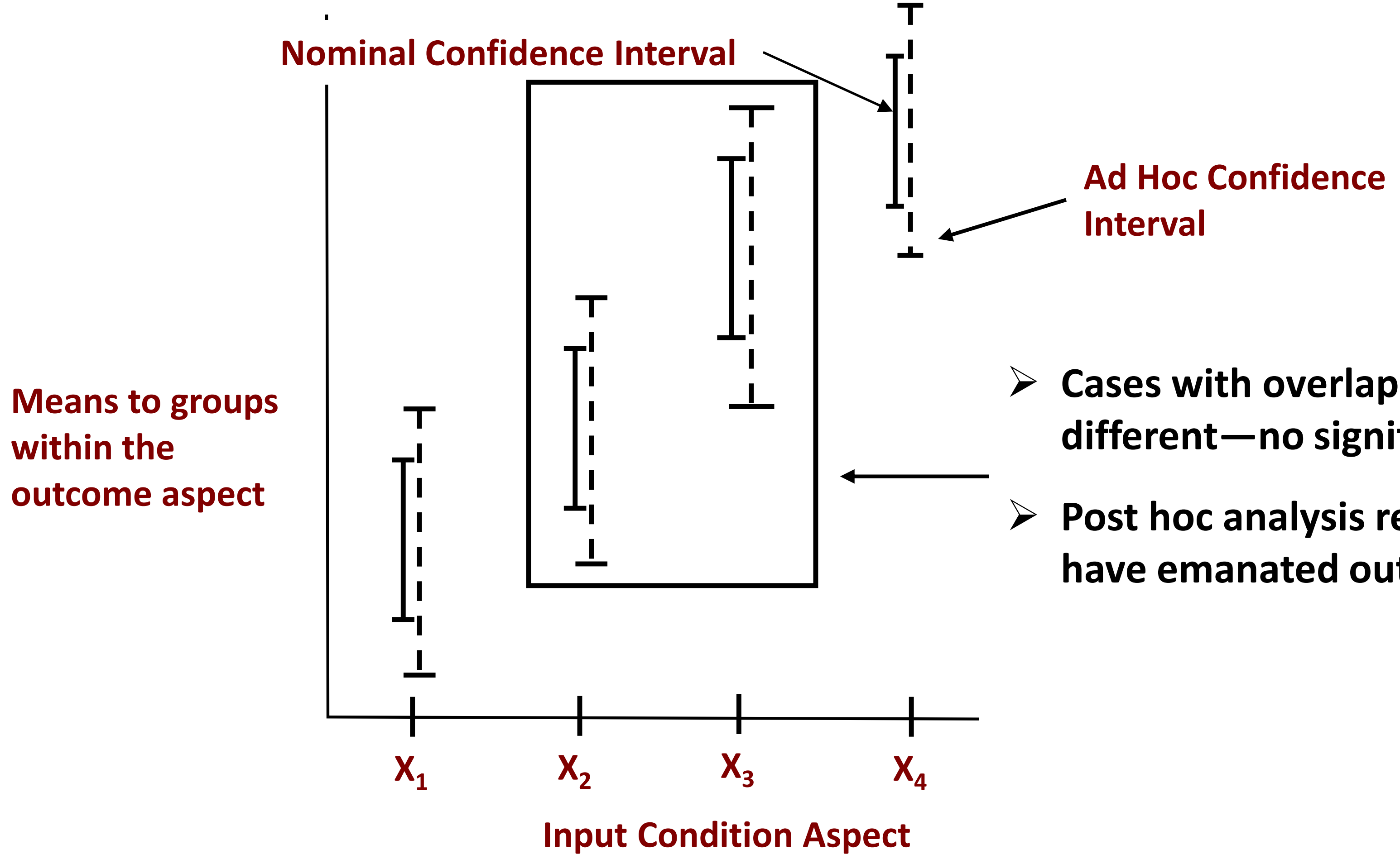
- The methods of comparison are the power of difference models—post hoc and contrasts.

See

See paper, "Know that Improvements Work by Asking Difference Questions," https://analytics4strategy.com/viveladifference

(CC BY) Richard Lamb 2019

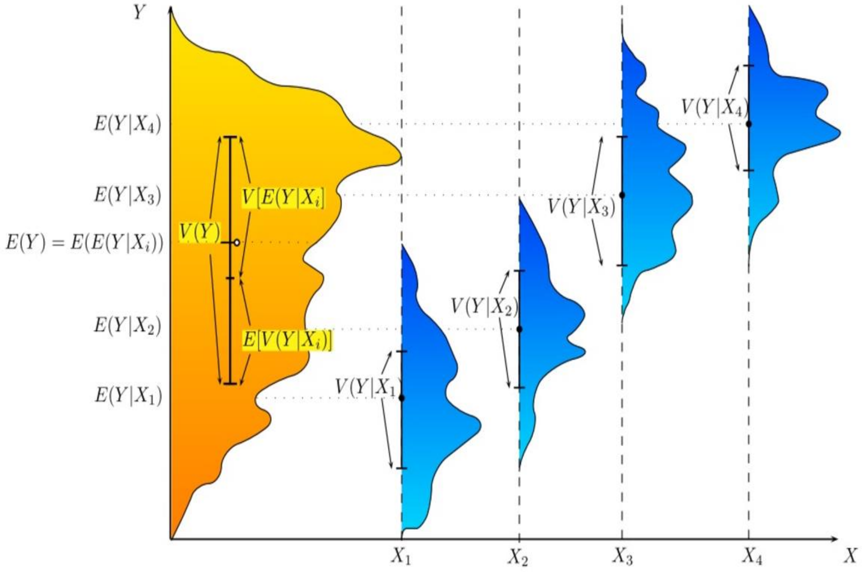
The **post hoc** method adjusts the confidence intervals to reflect the familywise math error such that the overall error remains within the intended overall confidence—e.g., 95 percent



- Cases with overlapping confidence intervals are not different—no significant difference between pairs.
- Post hoc analysis reveals one false conclusion would have emanated out of pair-wise comparison.

(CC BY) Richard Lamb 2019

The most powerful of methods is to construct contrasts or comparisons upon the sum of the squares (1) of the groups in the model—eliminating the family error from the comparison



Contrasts, essentially reflecting questions, are decided by the analyst with respect to the insight being sought.

SSt in the data to the mean of the data

SSm: Explained by the model — **Separated from** — **SSr: Unexplained by the model**

SSx1x2: Explained by the groups X1 & X2 — **Compared to** — **SSx3x4: Explained by the groups X3 & X4**

SSx1: Explained by group X1 — **SSx2: Explained by group X2**

Compared to

SSx3: Explained by group X3 — **SSx4: Explained by group X4**

Compared to

(1) For each group, sum of the squares (SS) is the sum of each case less basic model mean, squared.

(CC BY) Richard Lamb 2019

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?**

To this point we've assumed that different entities contribute to different means, but in operational analytics same entities contribute to different means—multilevel answers to the previous questions

- **Question: Do the three-year monthly KPIs for the plant's cost centers reveal that a program has envisioned consequences?**

  For these cases we must use the "repeated" measures method of ANOVA compared to "fixed" as shown to this point.

- **Question: How have the envisioned movement for the KPIs been associated with plant unplanned downtime?**

  For these cases we build a mixed ANOVA—fixed and repeated.

  As compared to KPIs to cost center, downtime "may be" independent of cost center.

- **Question: Which variables are independent (fixed) and dependent (repeated) in the envisioned consequences?**

  The approach is to start with a basic model, . .

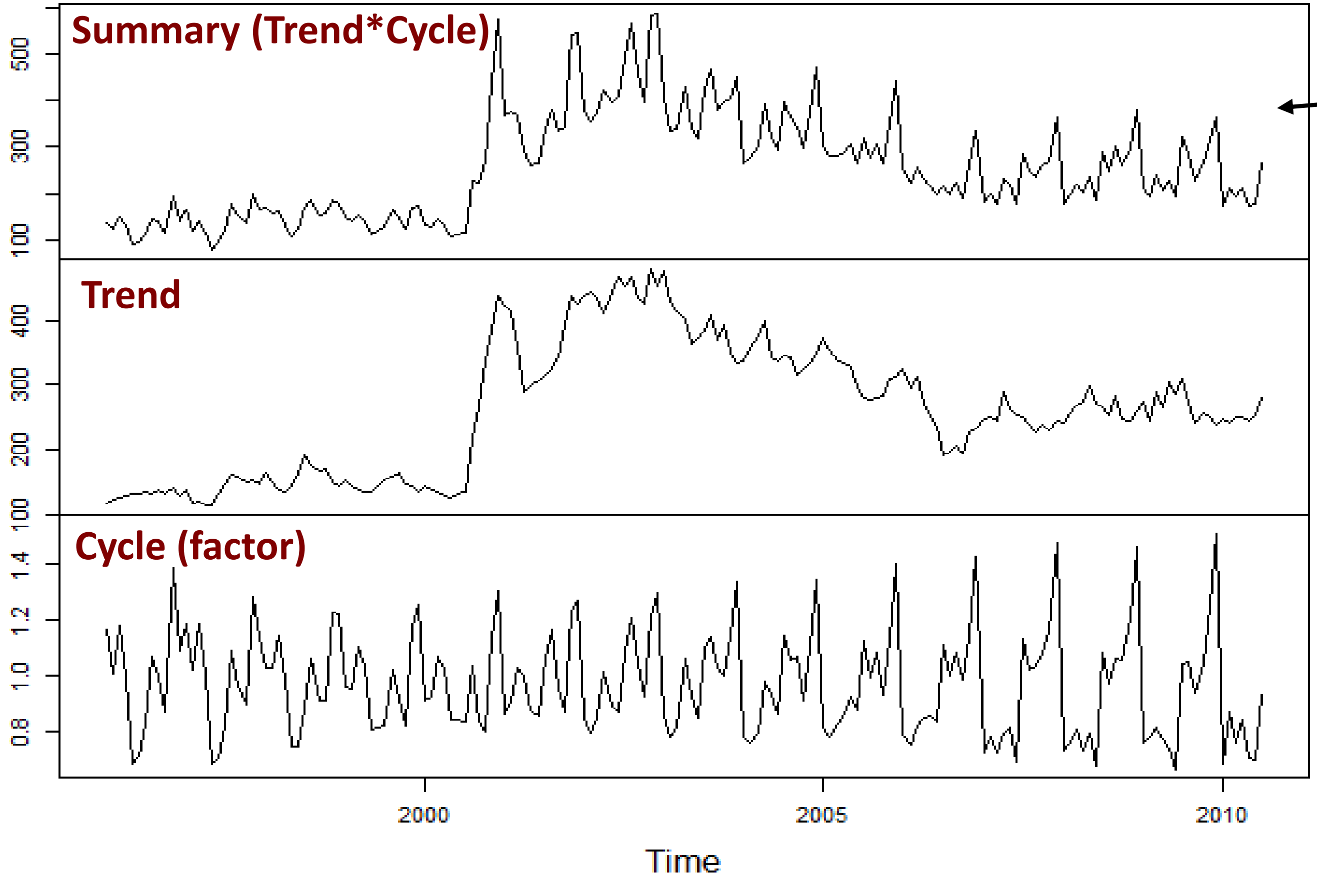
  structure the repeated measure and compare its fit to the basic model. . .

  add one variable at a time to as a level to the repeated KPI and compare to previous model until there is no further improvement in fit.

Agenda:

- ❑ Preamble, purpose and approach.
- ❑ Essential definitions.
- ❑ Context variables as questions.
- ❑ Five types of new-age questions.
  - Relationship.
  - Difference.
  - **Time series.**
  - Duration.
  - Apparency.
- ❑ Reference library.

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.

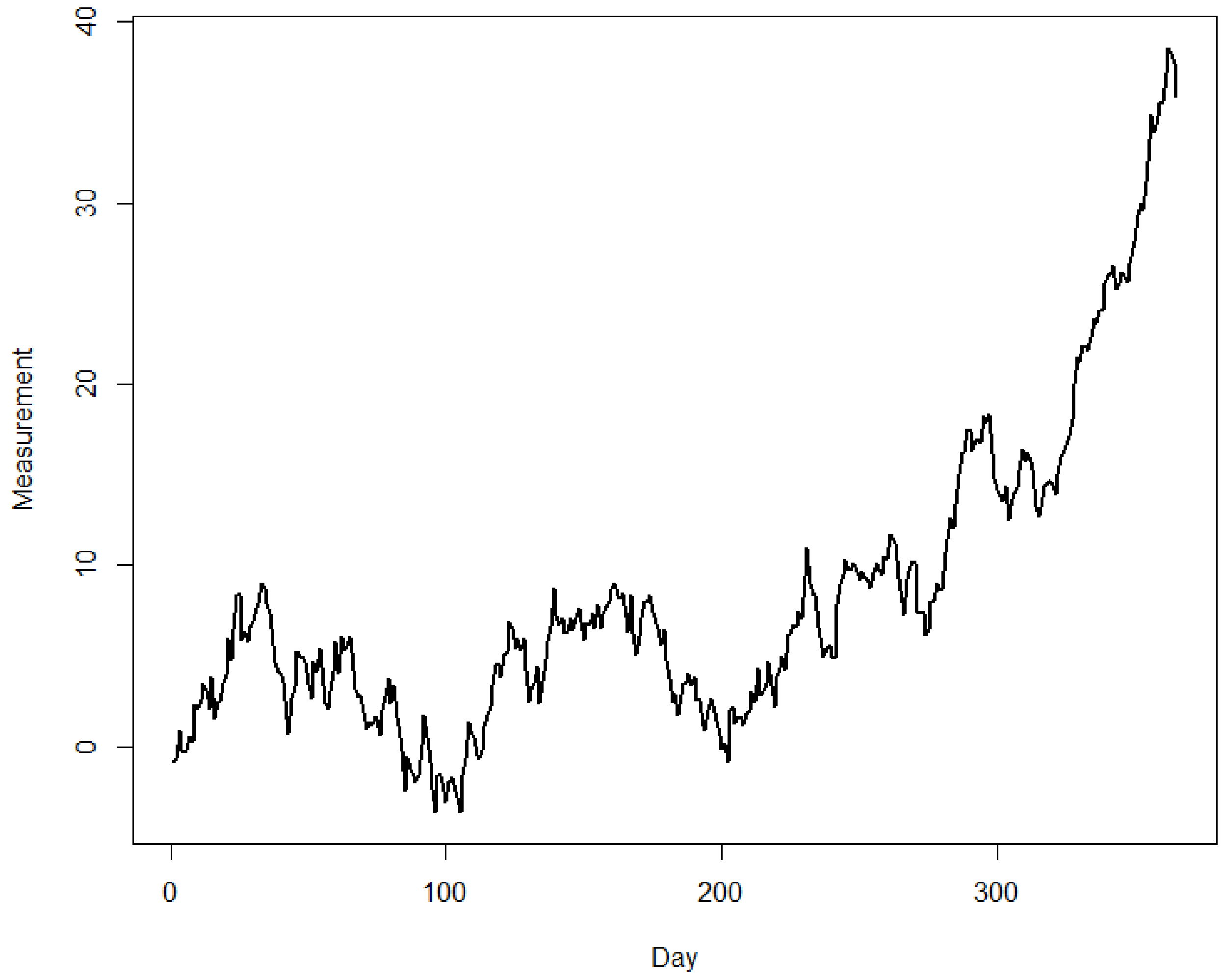
The true story lies behind the summary-level data as the history of constituent trend and cycle.

See

See paper, "Explore What Did and May Happen with Time Series Questions," https://analytics4strategy.com/timeseriesqs

(CC BY) Richard Lamb 2019

After extracting the cycles from the summary-level, we must determine if the remaining trend is deterministic, random walk or random?



- **It is immediately obvious that the trend is not random.**
- **Because of the shape, we would love to assume this trend is deterministic—making us happy.**
- **Times series modeling would find the trend is a random walk—making us wishful fools in paradise.**
- **The implied question is whether we can influence the trend?**

(CC BY) Richard Lamb 2019

There are countless questions to ask of the trends we determine to be deterministic

- 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 are the smoothed scheduled and non-scheduled workloads 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 maintenance policies changing—scheduled and non-scheduled, and between scheduled types?

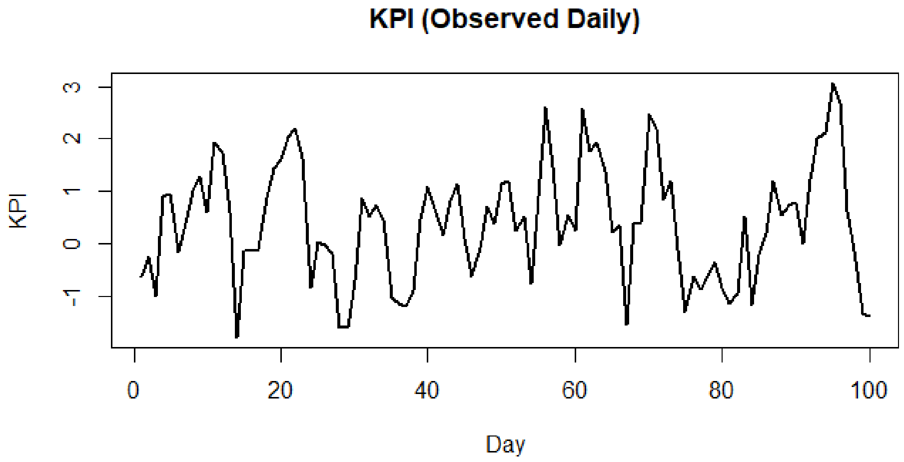
Note:
When the trend shows change, we may want to confirm and much more deeply explore the change through ANOVA upon the context variables to the change.

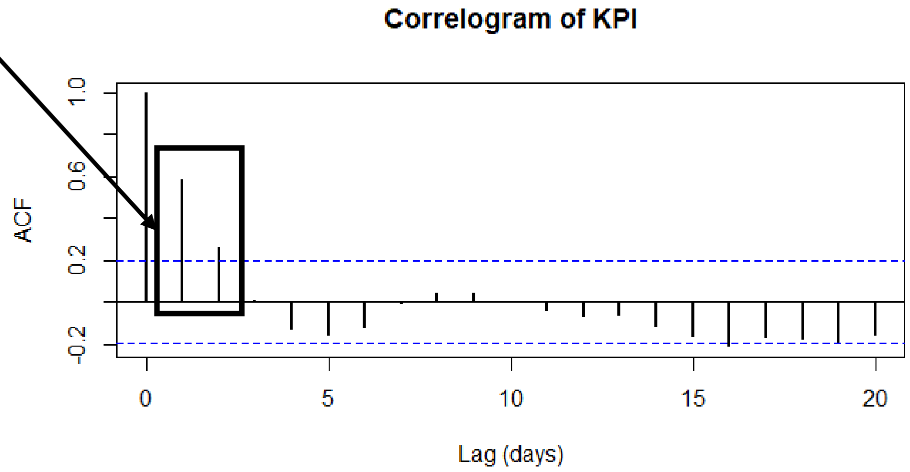
(CC BY) Richard Lamb 2019

It is possible that what is being reported currently also reflects previous periods--autocorrelation

In the time series, it is very difficult to discern autocorrelation.

In the correlogram, it is apparent that today's reported performance also reflects the previous two days.

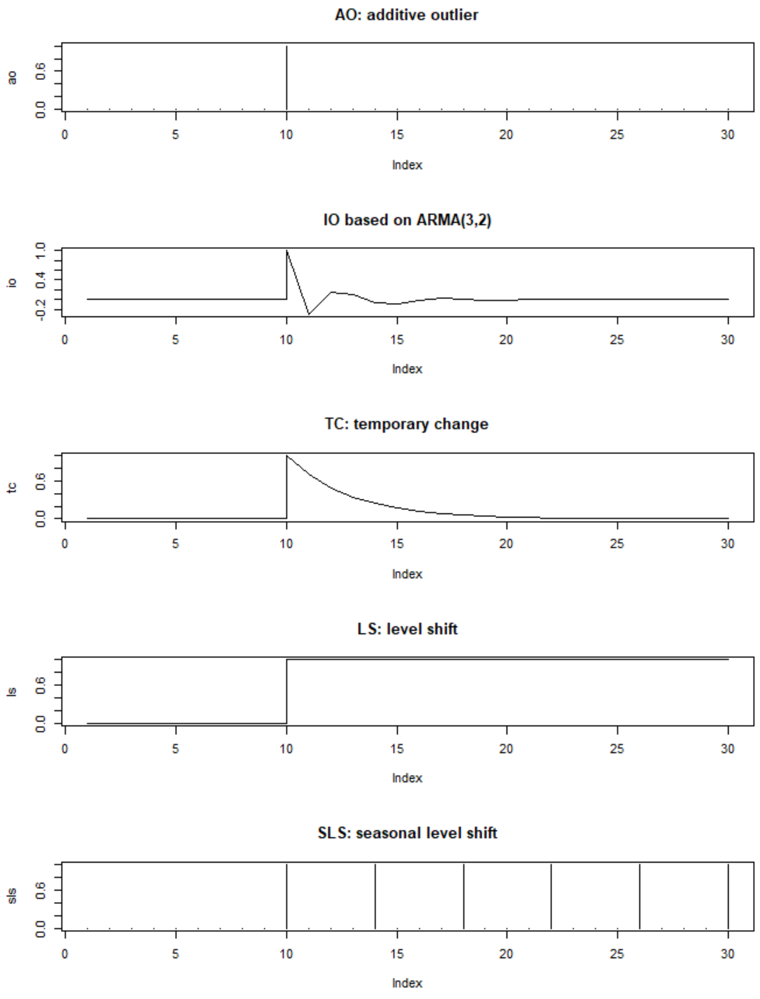




Example maintenance and reliability questions:

- Which of the KPIs to maintenance are distorted by autocorrelation?
- What spacing must we give to reporting KPIs to reveal the true story between periods?

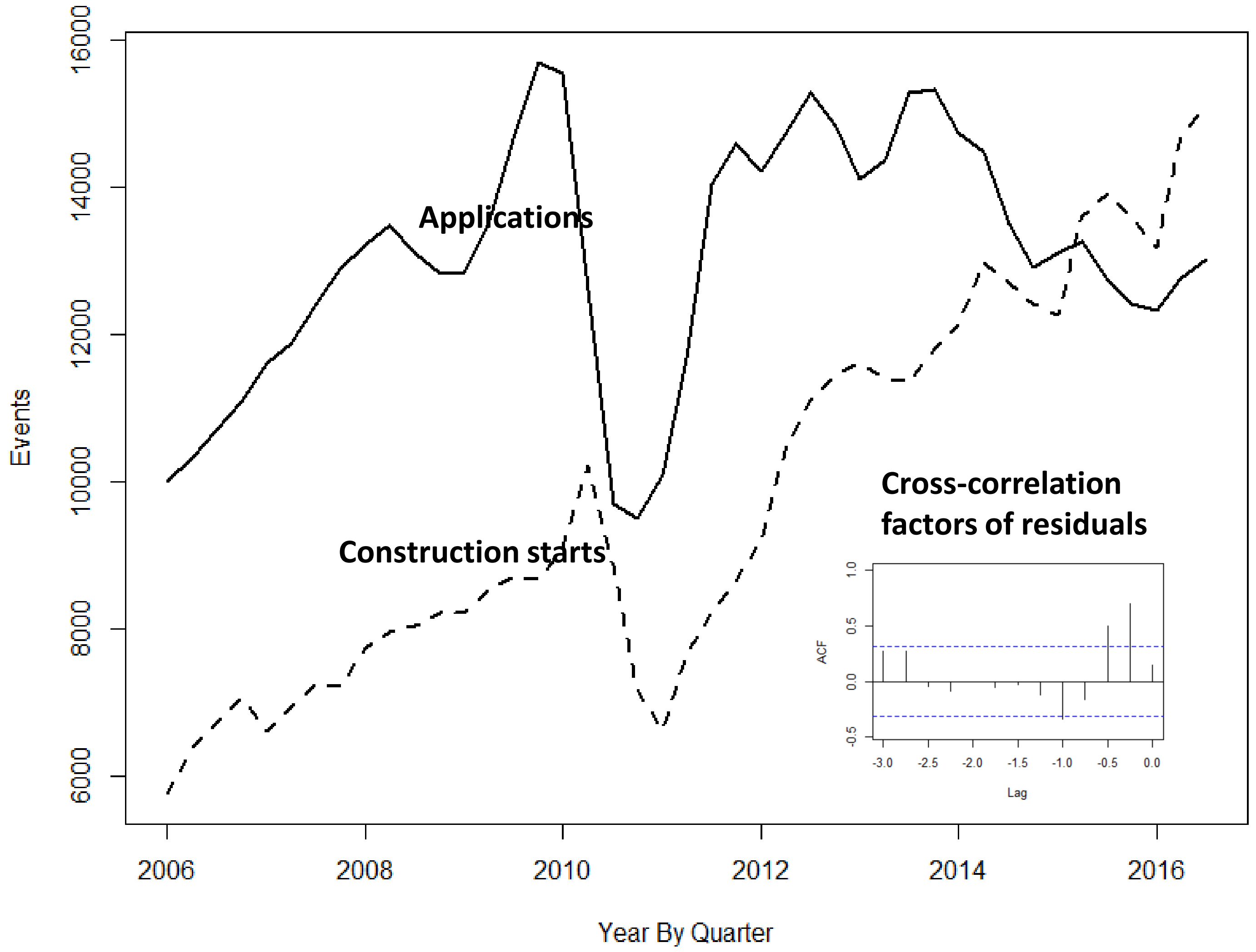
When we review performance after removing cycles, we can look for patterns that will lead us to conduct explorations to find what is driving them



Type	Description
Additive Outlier	Appears as a surprisingly large or small value occurring for a single observation. Subsequent observations are unaffected by an additive outlier.
Innovational Outlier	Appears after an initial impact with effects lingering over subsequent observations. The influence of the outliers may increase as time proceeds.
Temporary Change	Effect of the outlier diminishes exponentially over the subsequent observations. Eventually, the series returns to its normal level.
Level Shift	All observations appearing after the outlier move to a new level. In contrast to additive outliers, a level shift outlier affects many observations and has a permanent effect.
Seasonal level shift	A seasonal additive outlier appears as a surprisingly large or small value occurring repeatedly at regular intervals

(CC BY) Richard Lamb 2019

Will an analysis of cross-correlation reveal lead-lag series we can use to prepare or know to prepare for coming periods?

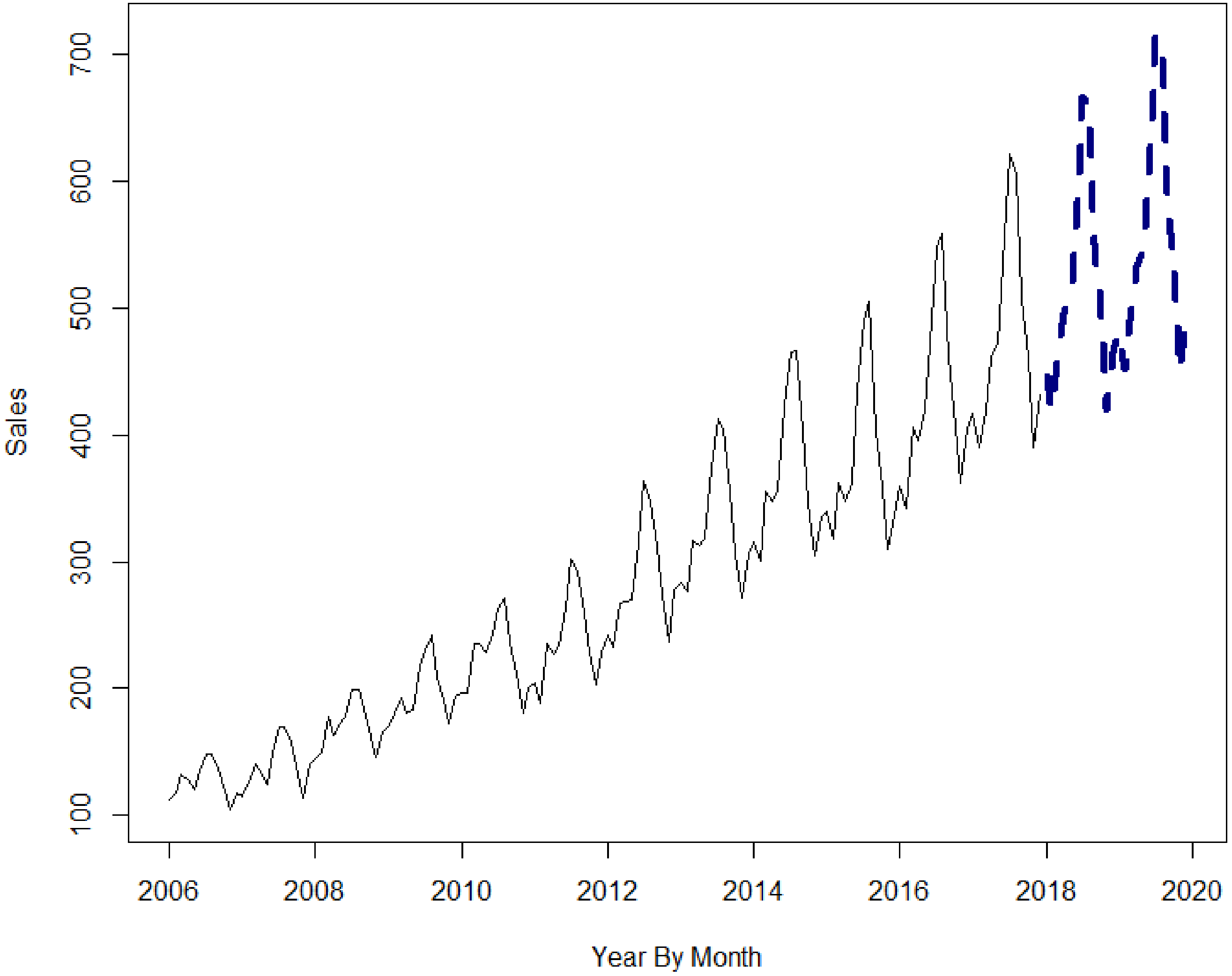


Example maintenance question:

Management wants to know, for the massive credence given to the insight derived from KPIs, can we actually observe the lead-lag relationship we have until now intuitively accepted as true?

(CC BY) Richard Lamb 2019

If the relationship of context variables to a performance of interest can be assumed to continue we can forecast that performance



Example maintenance and reliability questions:

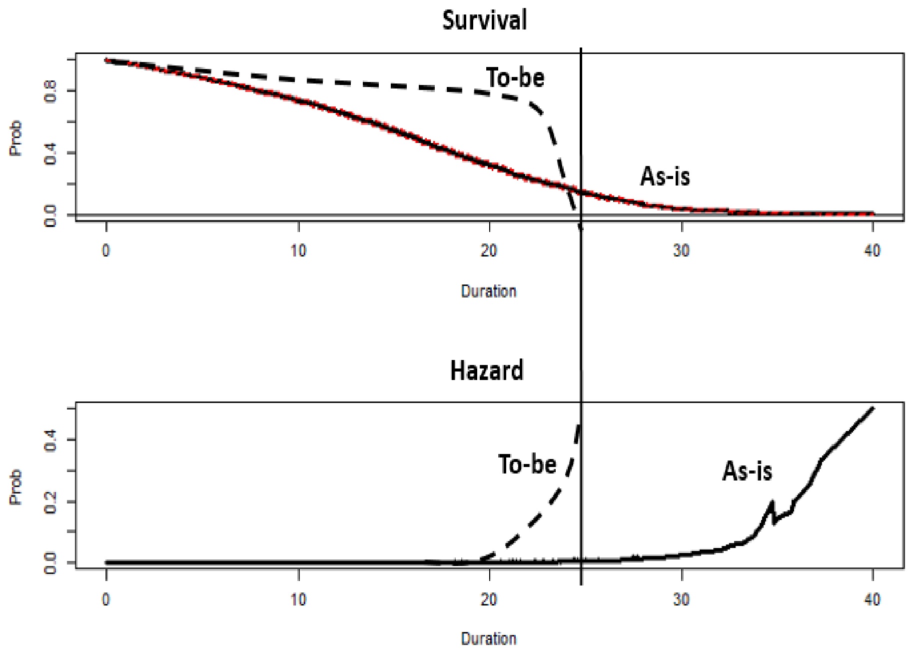
- Does history show that the workload for non-schedule work will increase in the coming year in response to production plan?
- What is the smoothed workload upon which to size the craft work force in the planned production year?

Agenda:

- ❑ Preamble, purpose and approach.
- ❑ Essential definitions.
- ❑ Context variables as questions.
- ❑ Five types of new-age questions.
  - Relationship.
  - Difference.
  - Time series.
  - **Duration.**
  - Apparency.
- ❑ Reference library.

(CC BY) Richard Lamb 2019

Duration questions: What is the probability an asset or process condition will hold for some time (AKA reliability) 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?

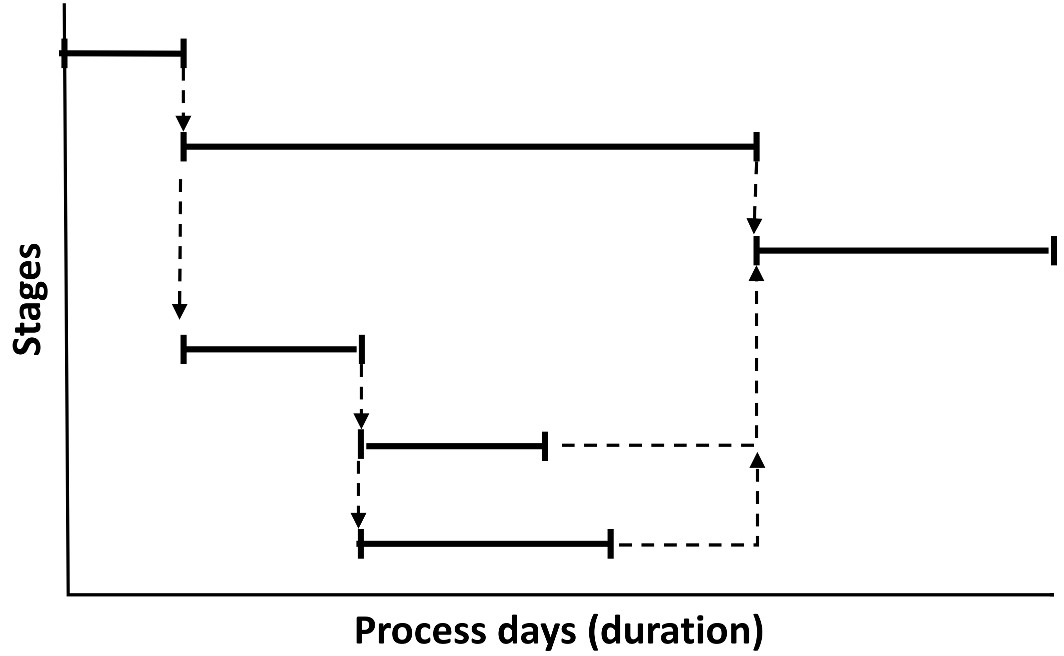
- **Actual occurring shapes tell the story compared to the shape we expect or want it to be—as-is and to-be.**
- **Point of reference: The "six shapes of failure" are hazard plots.**

See

See paper, "Find the Time That is Money by Asking Duration Questions," https://analytics4strategy.com/tmismnyqstns

(CC BY) Richard Lamb 2019

An issue to production management is the collective duration through the critical-path stages of maintainability in excess of the time needed to sustain a smoothed workload for craft productivity



- In our operational systems, e.g. CMMS, we have tremendous amounts of data with which to see clearly the patterns through each stage from notification to completion.
- Status history from the CMMS is the core variable to analysis of maintainability interval.

See

See paper, “Anatomy of SAP Statuses History for Routine Maintenance,” https://analytics4strategy.com/statusanatomy

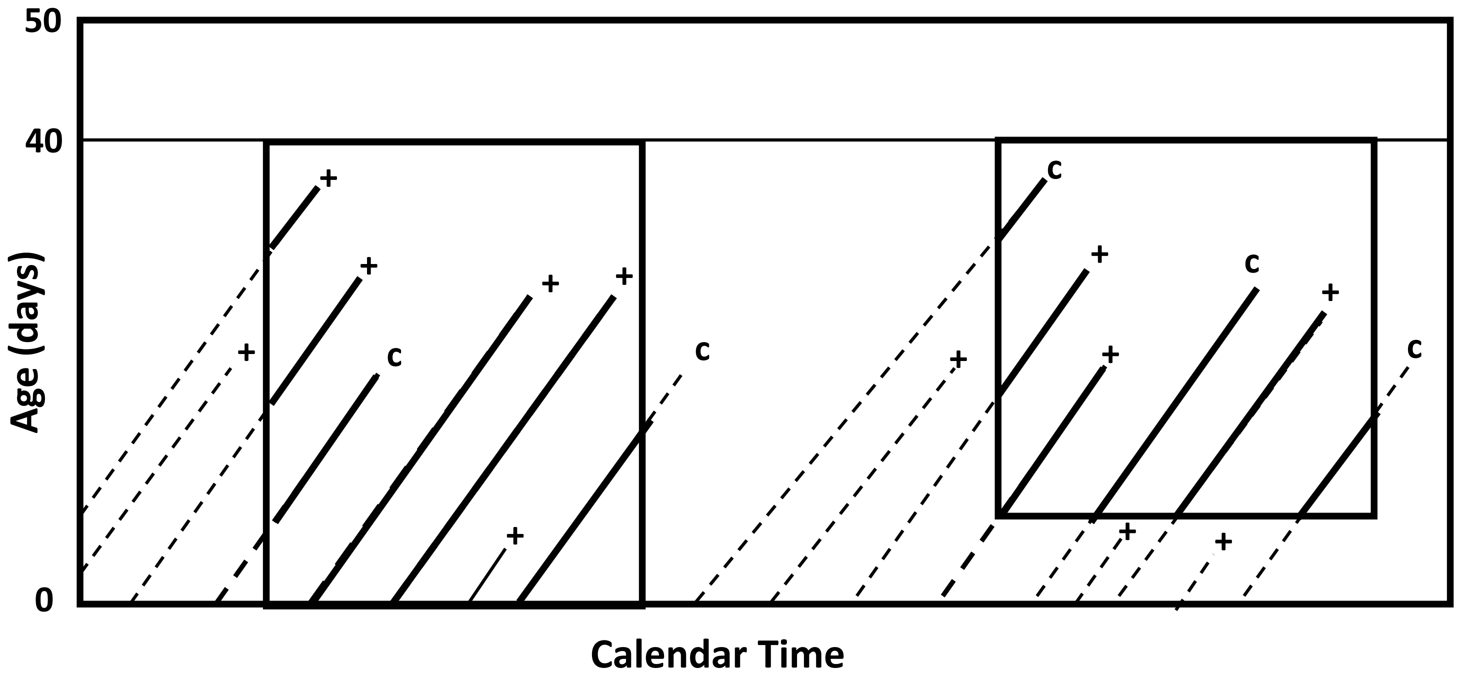
The questions orbit around the shape of the survival and hazard curves for process stages and asset reliability

Example maintenance and reliability questions:

- Are the shapes of the curves acceptable for each stage from work notification to completion—maintainability?
- Do the curves 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 unnecessary exposure to functional and multiple failures, or compensating shorter intervals?
- Is there evidence that very old work normally hides in some stages of the backlog—distorting the perception of true backlog?
- What RCM-based maintenance policy is indicated by the shape of a particular failure mode as the exiting event?

We will spend some slides to overview the Kaplan Meir method which is parameterized by the Cox models per Weibull or alternative distributions

The Kaplan Meir method with the two decisions of analysis—age and calendar



The plot allows us to inspect and better understand the history of cases before making age-calendar decisions suitable to our analysis.

Legend:

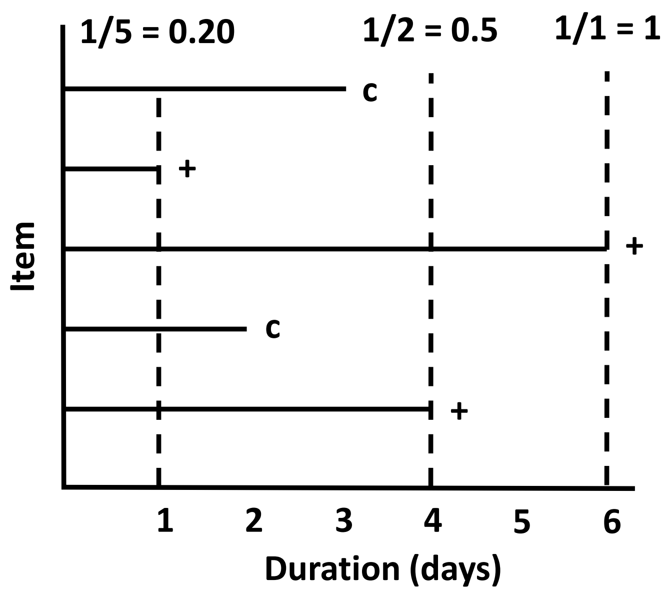
✚ Exit event to the analysis.

C Censored (suspends in Weibull literature)—point at which the case experienced something other than the subject exit event or had not yet happened at the end of the subject interval.

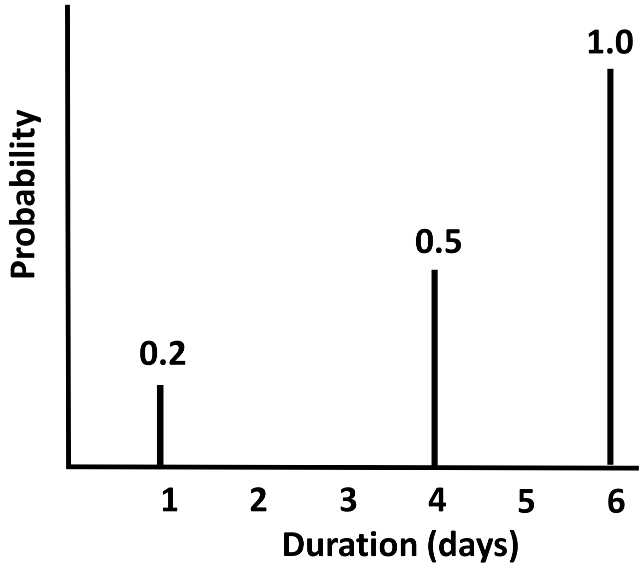
(CC BY) Richard Lamb 2019

The hazard function is built empirically—as shown—with the cases as a composite variable of enter, exit and event variable for the falling in the age-calendar plot

Hazard events and censoring

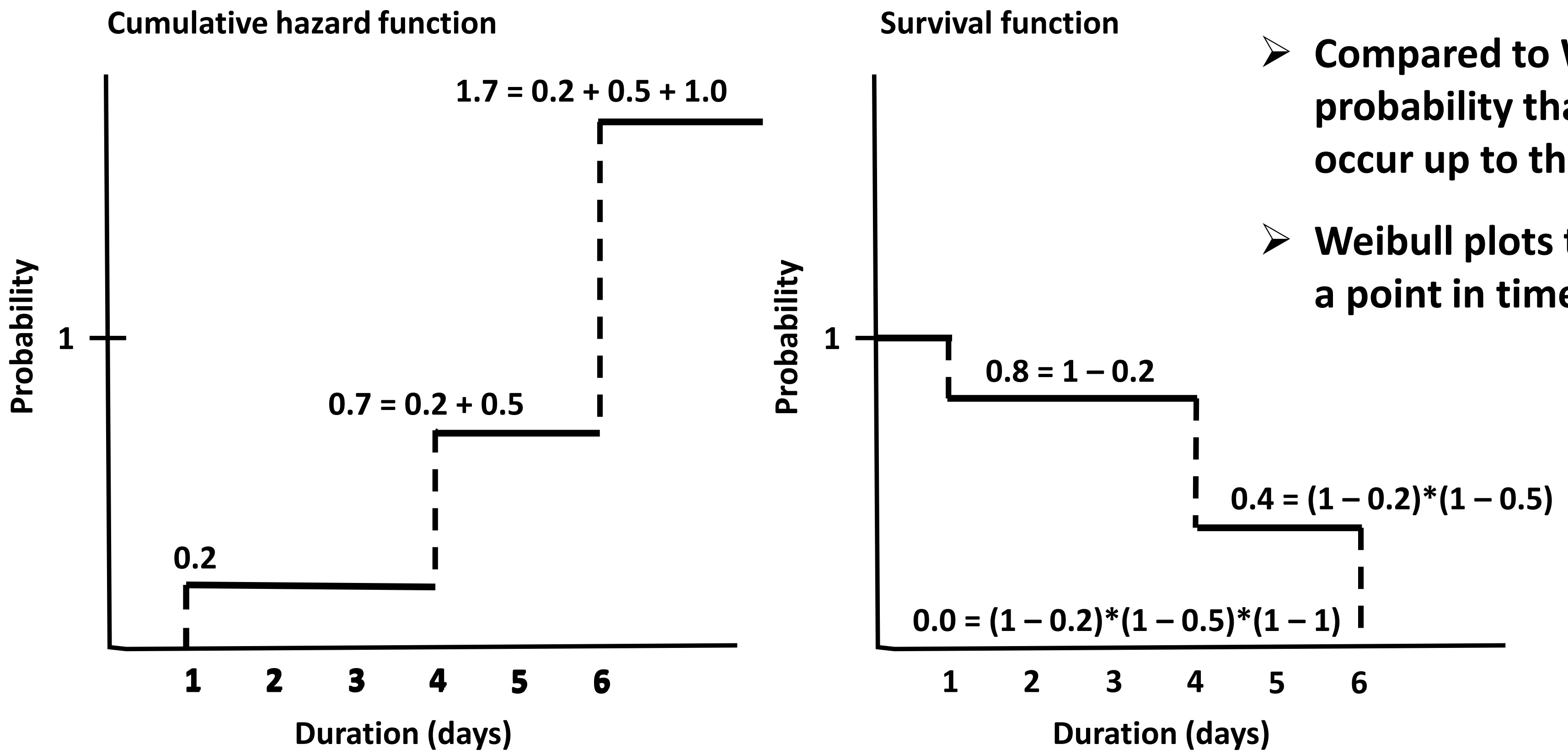


Event hazard function



- "Hazard" is the chance of an exit event from the cases that have survived to just before that time.
- In Weibull, hazard is called "instantaneous failure."

Through hazard, the cumulative hazard of an event over a period of time and the survival curves are plotted as shown



- Compared to Weibull, the calculation is the probability that an event (failure) will not occur up to the point in time (R(t)).
- Weibull plots the probability of failures up to a point in time (1 – R(t)).

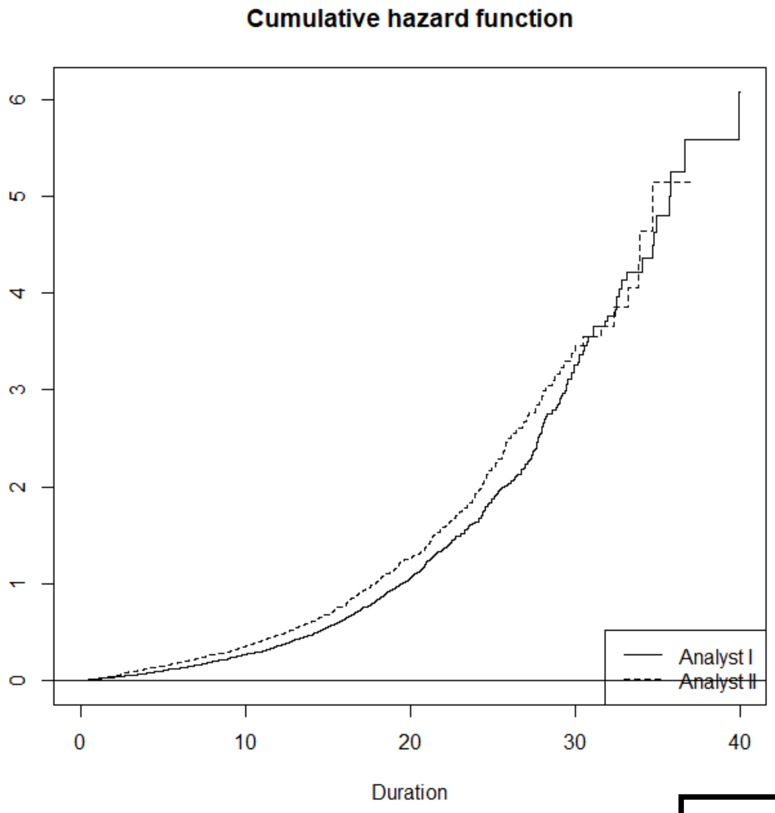
The Kaplan Meir calculation, can be parameterized through Cox model which fits the event history to Weibull and returns the eta and beta parameters

```
> (para.fit<- phreg(Surv(enter-60, exit-60, event)~1, data=om) ) ##Returns beta and eta
Call:
phreg(formula = Surv(enter - 60, exit - 60, event) ~ 1, data = om)

Covariate          W.mean      Coef Exp(Coef)  se(Coef)    Wald p
log(scale)                    2.887             0.013       0.000
log(shape)                    0.541             0.018       0.000

Events                     1971
Total time at risk          37824
Max. log. likelihood       -7437.5

> #plot(para.fit)##Will return the four-set to weibull
> #str(para.fit)
> (eta.exp<- exp(para.fit$coefficients[1]))##Convert ln of eta to eta.
log(scale)
  17.93449
> (beta.exp<- exp(para.fit$coefficients[2]))
log(shape)
  1.718258
```

- Confirms the goodness of the fit (< 0.05 to the default Weibull.
- Alternatives are extreme value, "Gompertz", piecewise constant hazards, "loglogistic" and "lognormal."

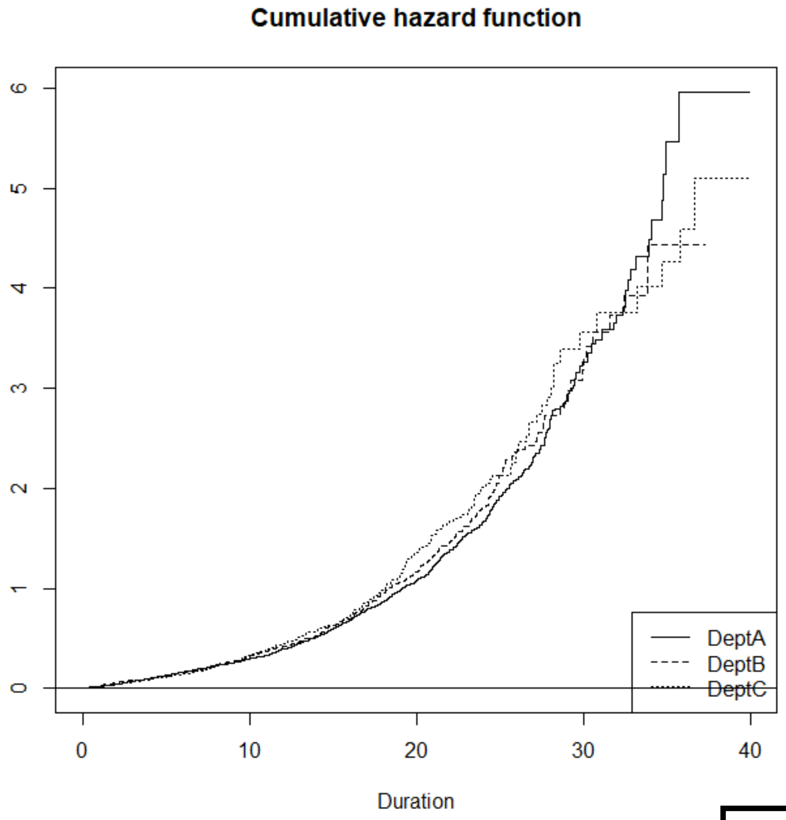
The eta and beta parameters of Weibull to fit the data to a Weibull family.

A frequently powerful insight is which context variables have the strongest relationship to exiting events, thus, survival in the status—giving focus to advancing operational performances?




```
                 coef exp(coef) se(coef)     z Pr(>|z|)
GradeAnalyst II 0.1930    1.2128   0.0456 4.232 2.32e-05

Score (logrank) test = 17.96  on 1 df,   p=2.254e-05
```

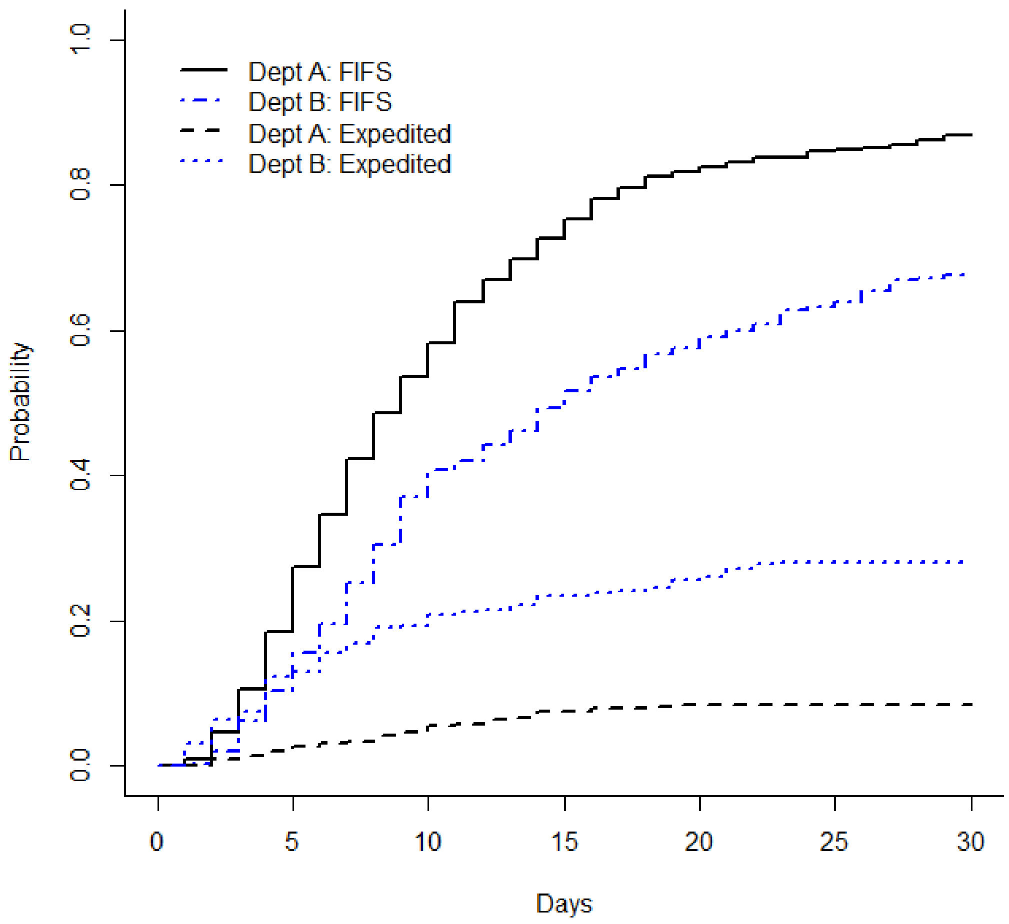
P-value < 0.05 reveals that analyst level has significant relation to exit events.




```
             coef exp(coef) se(coef)     z Pr(>|z|)
DeptDeptB 0.0594    1.0612   0.0552 1.076   0.2819
DeptDeptC 0.1008    1.1060   0.0595 1.694   0.0903

Score (logrank) test = 3.28  on 2 df,   p=0.1939
```

P-values > 0.05 reveals that department does not have a significant relation to exit events.

We can explore for competing events, if one event occurs, none of the others can occur.

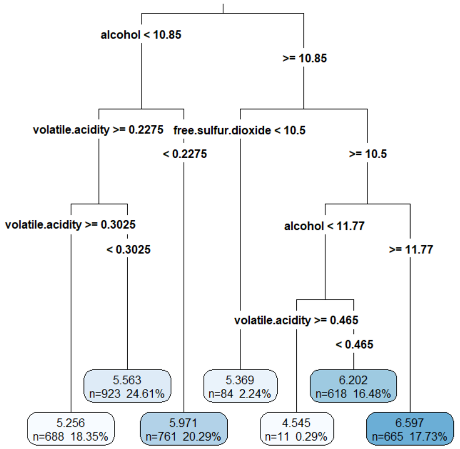


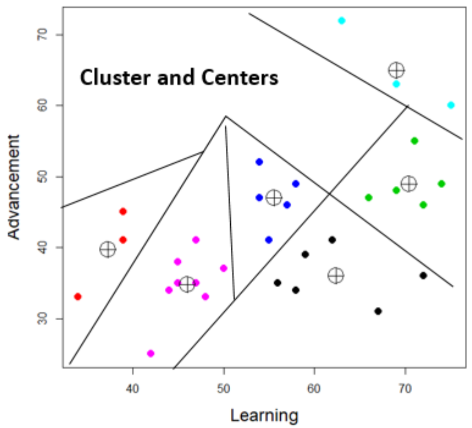
- There are two competing events—first-in, first-served (FIFS) and expedited. The events are also grouped by department.
- Questions we can ask:
  - Are there a significant number of expedited cases as compared to the mandated first-in, first-served policy?
  - Is the level of expedited cases acceptable for each department?
  - Is there a pattern of rush exits in the first two days?
  - Are the first-in, first-served cases exiting in a timely manner?
- From the answers, we may want to look into the plots singly or in combinations around single events, thence, assess the curves as baselines, multiple survival distributions, interactions and mixed effects.

Agenda:

- ❑ Preamble, purpose and approach.
- ❑ Essential definitions.
- ❑ Context variables as questions.
- ❑ Five types of new-age questions.
  - Relationship.
  - Difference.
  - Time series.
  - Duration.
  - **Apparency.**
- ❑ Reference library.

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





Decision models:

- ➢ Hidden **rules** as variables associated with numeric and categorical outcomes.
- ➢ An outcome variable is given to the model—supervised (directed).
- ➢ Models include decision tree, regression tree, model tree, naïve Bayes and more.

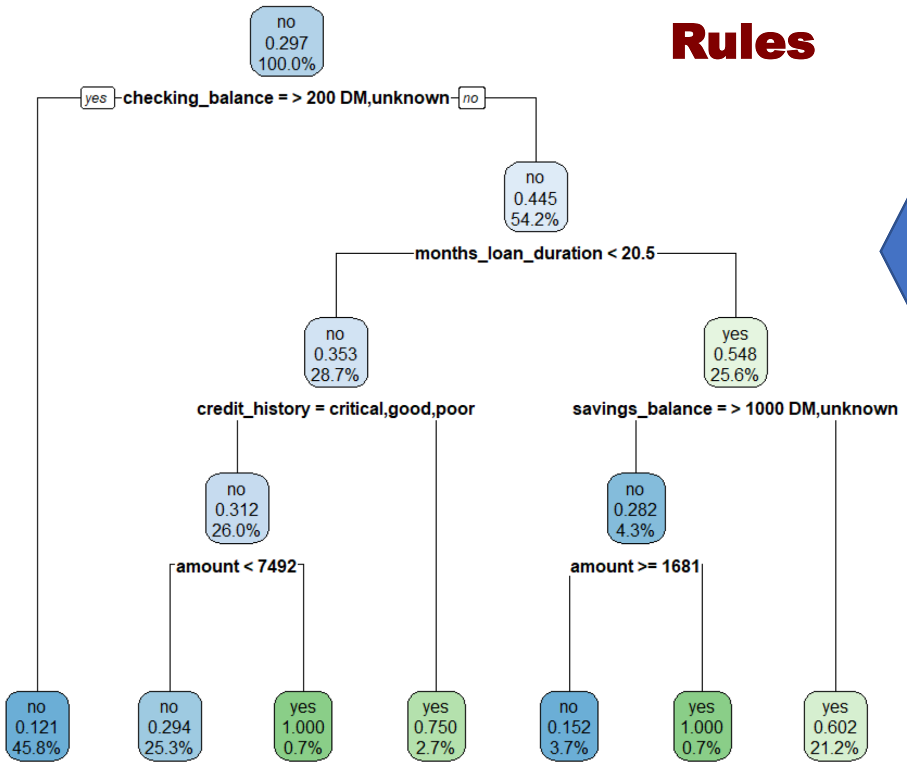
K-Means:

- ➢ Hidden **variables** are learned by slicing-dicing existing predictor variables until clusters of similarity emerge.
- ➢ No outcome variable is given to the model—unsupervised (undirected).
- ➢ Models include K-mean, principle component analysis, MANOVA and more.

See

See papers, "Dive Below the Surface of Process Functioning with Apparency Questions," https://analytics4strategy.com/apprqsblwfnctng and for MANOVA see, "Know that Improvements Work by Asking Difference Questions," https://analytics4strategy.com/viveladifference

A decision tree model establishes rules for classifying cases along with the probability that each case is as it has been classified



Rules

THEN

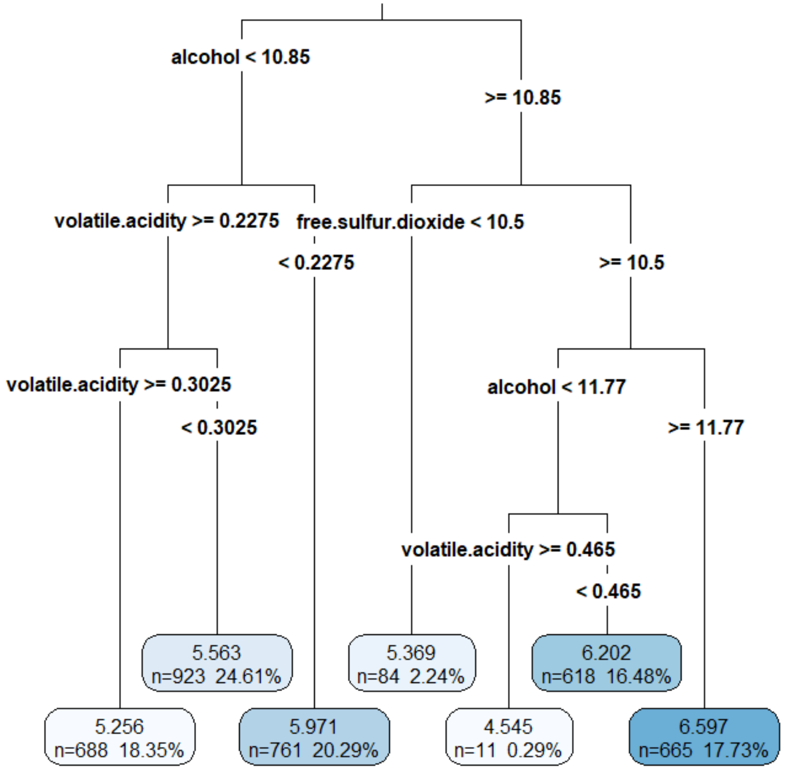
IF:
(checking_balance = < 0 DM) and (percent_of_income >= 3) and
(amount >= 3051) => default=yes (53.0/15.0)
ELSE IF:
(checking_balance = 1 - 200 DM) and (savings_balance = < 100 DM)
and (months_loan_duration >= 24) => default=yes (58.0/19.0)
ELSE IF:
(checking_balance = < 0 DM) and (percent_of_income >= 4) and
(housing = rent) => default=yes (20.0/5.0)
ELSE IF:
(checking_balance = < 0 DM) and (months_loan_duration >= 18) and
(amount <= 2473) and (phone = no) => default=yes (25.0/6.0)
ELSE IF:
(age <= 29) and (months_loan_duration >= 33) => default=yes
(32.0/14.0)
ELSE:
=> default=no (712.0/138.0)

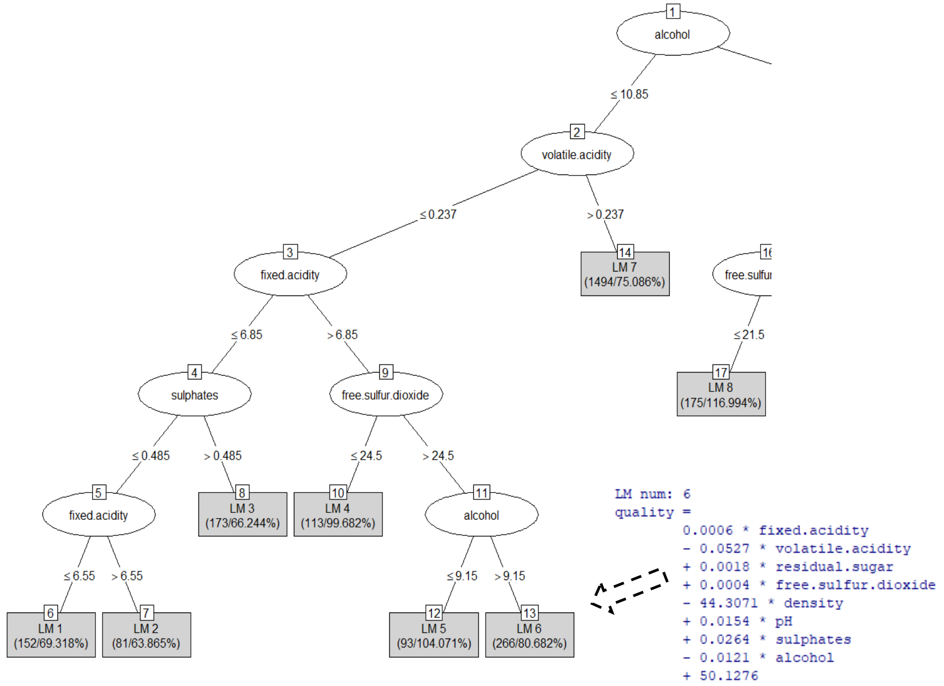
Number of Rules : 6

Case	Probability of default	
	no	yes
865	0.8786408	0.1213592
448	0.7061404	0.2938596
199	0.8484848	0.1515152
861	0.8786408	0.1213592
621	0.3979058	0.6020942
645	0.7061404	0.2938596

(CC BY) Richard Lamb 2019

Two types of regression tree models

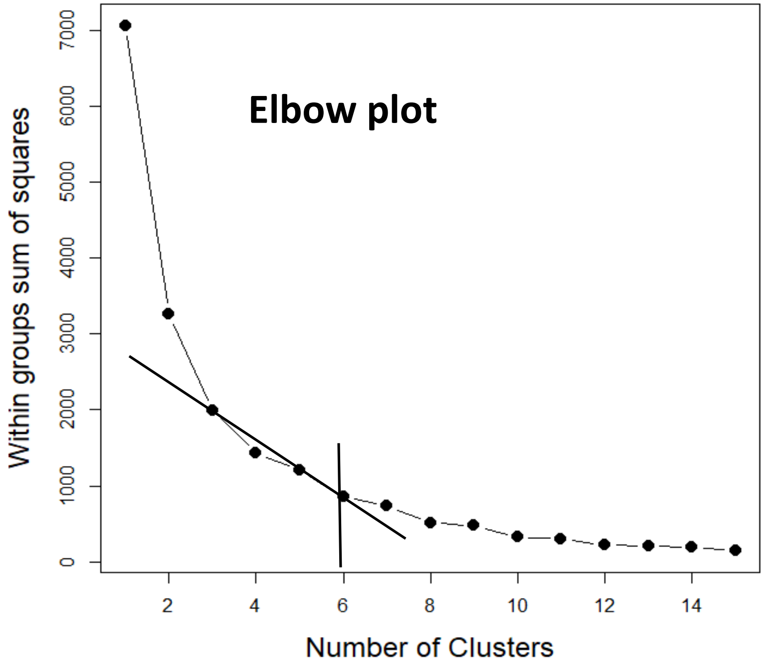


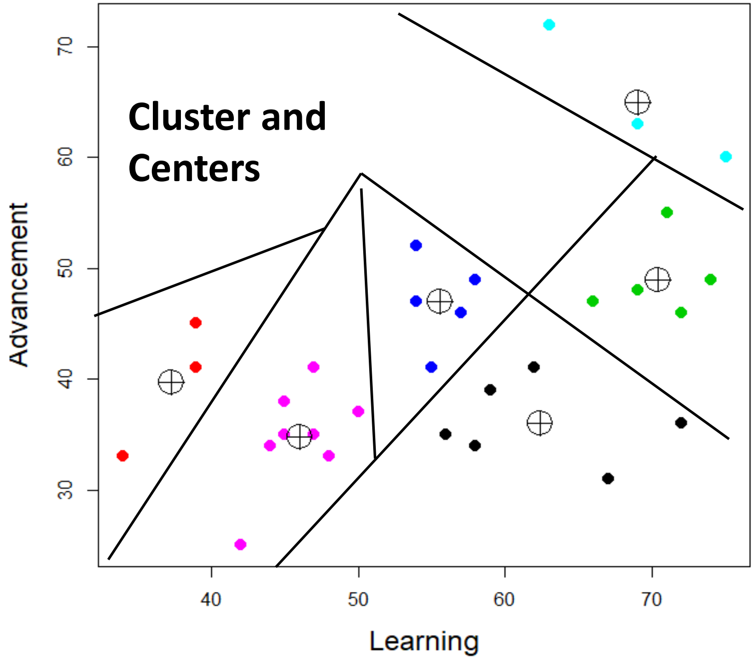


A **regression** tree model returns the mean of the outcomes at each leaf.

A **model** tree returns a regression model at each shown leaf—suggest further investigation.

We can slice-dice context variables to reveal clusters we cannot otherwise see in the data





- **Example: Six hidden variables are teased out of two associated context variables.**
- **The elbow plot analytic gives us a sense of how many clusters to see from the K-Means model.**
- **Experts in the operational context would interpret the meaning of the clusters.**
- **However, without interpretation, clusters can be placed in the super table to strengthen analyses from the table.**

(CC BY) Richard Lamb 2019

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

Example maintenance and reliability questions:

- **Hidden Rules:**
  - Is the process operating to the established rules of conduct for decision-making and classifications?
  - Are there situations—evidence by a revealed rule—in the maintenance process we have not before recognized as existing?
  - Can outcomes for hours, costs, productivity and KPIs be 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 context variables point to an outcome we have never before recognized?

Agenda:

- ❑ Preamble, purpose and approach.
- ❑ Essential definitions.
- ❑ Context variables as questions.
- ❑ Five types of new-age questions.
  - Relationship.
  - Difference.
  - Time series.
  - Duration.
  - Apparency.
- ❑ **Reference library.**

The considerations for variations to each are explained by the following referenced papers

Question	Paper	Link
Relationship	Find What Matters with Relationship Questions of Operations	https://analytics4strategy.com/relatqstoci
Difference	Know that Improvements Work by Asking Difference Questions	https://analytics4strategy.com/viveladifference
Time series	Explore What Did and May Happen with Time Series Questions	https://analytics4strategy.com/timeseriesqs
Duration	Find the Time That is Money by Asking Duration Questions	https://analytics4strategy.com/tmismnyqstns
Apparency	Dive Below the Surface of Process Functioning with Apparency Questions	https://analytics4strategy.com/apprqsblwfnctng

(CC BY) Richard Lamb 2019

Library: "What-to" and "how-to" for data-driven development and functioning

Knowledge and skills		Papers, presentations, training sessions	Texts or equivalents
Data-drivenness	Framework	▪ First Step to Becoming a Data-Driven Operation ▪ Data-Driven Maintenance Operations	None available
R	Coding	None available	• R for Dummies, de Vries, Meys, 2015. • Art of Programing R, Matloff, 2011. • Manual at https://r-project.org.
Data tables	Super tables	▪ Build Super Tables from Operational Data ▪ Purge the Fused Spreadsheets That Undermine Data-Drivenness	Access 2016 Bible, Alexander and Kusleika, 2016, Chapters 8 – 13.
Data preparation	Cleansing	▪ Build Super Tables from Operational Data	Rstudio for R Statistical Computing Cookbook. Andrea Cirillo, 2016, Chapter 2
Pivot tables, graphs	Pivot dashboards	None available	Pivot Tables In-Depth for MS Excel 2016, Oesko, 2017.
	Layered charting	None available	ggplot2, Elegant Graphics for Data Analysis, Wickham, 2016

Continued. . .

Library: Continued

Knowledge and skills		Papers and presentations	Texts or equivalents
Five analytic questions	Relationship	Find What Matters with Relationship Questions of Operations	▪ Discovering Statistics Using R, Field and Miles, 2012 ▪ Multilevel Modeling Using R, Holmes, 2014
	Difference	Know that Improvements Work by Asking Difference Questions	
	Time series	Explore What Did and May Happen with Time Series Questions	▪ Introductory Time Series with R, Cowpertwait and Metcalfe, 2009 ▪ R Package "tsoutliers," Javier López-de-Lacalle, 2017
	Duration	Find the Time That is Money by Asking Duration Questions	▪ Event History Analytics with R, Bostrom, 2012 ▪ New Weibull Handbook, Abernathy, 2007 ▪ R Package "WeibullR" Weibull Analysis for Reliability Engineering, Silkworth & Symynck, 2018
	Apparency	Dive Below the Surface of Process Functioning with Apparency Questions	Machine Learning with R, Lantz, 2015
Machine learning, AI	Methodology	None available	