Taylor & Francis
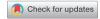Taylor & Francis Group

INTERDISCIPLINARY

Check for updates

Analytical Problem Solving Based on Causal, Correlational and Deductive Models

Jeroen de Mast[a,b], Stefan H. Steiner[a], Wim P. M. Nuijten[c], and Daniel Kapitan[b]

[a]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada; [b]Jheronimus Academy of Data Science, Den Bosch, Netherlands; [c]Eindhoven Artificial Intelligence Systems Inst., Eindhoven University of Technology, Eindhoven, Netherlands

ABSTRACT
Many approaches for solving problems in business and industry are based on analytics and statistical modeling. Analytical problem solving is driven by the modeling of relationships between dependent (Y) and independent (X) variables, and we discuss three frameworks for modeling such relationships: cause-and-effect modeling, popular in applied statistics and beyond, correlational predictive modeling, popular in machine learning, and deductive (first-principles) modeling, popular in business analytics and operations research. We aim to explain the differences between these types of models, and flesh out the implications of these differences for study design, for discovering potential X/Y relationships, and for the types of solution patterns that each type of modeling could support. We use our account to clarify the popular descriptive-diagnostic-predictive-prescriptive analytics framework, but extend it to offer a more complete model of the process of analytical problem solving, reflecting the essential differences between causal, correlational, and deductive models.

1. Introduction

1.1. Three Frameworks for Analytical Problem Solving

The World Economic Forum (2020) lists problem solving as one of the essential skills that current generations of professionals in business and industry need to master. Problem solving in its broad sense is the task of figuring out how to go from an unwanted to a wanted situation. Solving problems is the crux of many tasks in management, engineering and science, and methods, techniques and procedures for solving business and engineering problems are studied in various fields.

Many approaches for solving problems are based on analytics and statistical modeling, and many techniques in statistics are intended to support problem solving (Hoerl and Vining 2020). In business and industry, the traditional framework for statistical problem solving is based on cause-and-effect modeling. Highly influential is Juran's (1998) distinction between a *Diagnostic Journey*, where the causes of a problem are established, and a subsequent *Remedial Journey*, where the identified causes drive the design of solutions. The notion that solving problems is based on the modeling of causal relations between dependent (*Y*) and independent variables (*X*'s) permeates Six Sigma's DMAIC method (De Mast and Lokkerbol 2012), Shainin's Statistical Engineering system (Steiner, MacKay, and Ramberg 2008), and scientific studies of the problem-solving process (Smith 1988, 1998; MacDuffie 1997; De Mast 2013). Trying to design a solution without first establishing a problem's causes is often portrayed as bad practice and relying on make-shift or stopgap solutions.

In recent years, this causal framework has gotten more and more competition from an alternative framework for solving problems based on analytics, which on the surface seems to have many similarities, but which, we believe, is essentially different. This is the framework of purely predictive, nonexplanatory modeling popular in data science and machine learning. Also in this framework, problems are solved by modeling relations between dependent and independent variables, but the relationships are not claimed to be causal, and usually are instead correlational. Although both frameworks use partly similar modeling techniques, such as regression, the distinction between causal and correlational models has crucial ramifications for the type of studies needed to model them and the types of solutions that they allow. Correlational predictive models and causal models are contrasted in Breiman (2001), Shmueli (2010), and Hernan, Hsu, and Healy (2019).

Both in cause-and-effect models and machine-learning-style correlational models, relationships between dependent and independent variables are modeled using data. There are, however, many problems where these relationships can be derived from theory and universal laws by mathematical deductions, and this constitutes a third framework for analytical problem solving. Many problem-solving methods in operations research (OR) and business analytics are not based on data-driven modeling of relationships, but instead, relationships are derived from mathematics or probability theory, either analytically or by means of simulation (Den Hertog and Postek 2016). In engineering, relationships are often derived from the laws of physics, and such *first-principles* or *white-box* models are contrasted to

empirical or *black-box* models, where relationships are established by fitting functions with data (Estrada-Flores et al. 2006).

1.2. *Purpose of the Article*

All three frameworks for problem solving—referred to as causal modeling, correlational modeling and deductive modeling—try to solve problems by modeling relationships between dependent (Y) and independent variables (the X's). These $Y=f(X)$ relationships predict the effects of interventions in the X's and are the basis for designing a solution. In this article, we seek to understand, contrast and integrate these three frameworks for analytical problem solving. Various fields have a traditional predisposition towards one of the three frameworks: causal modeling in business and industrial applications of statistics, correlational predictive modeling in machine learning, and deductive, nondata-driven modeling in business analytics and OR. But these predispositions are by no means exclusive, as many models in statistics are noncausal, machine learning has recently started to embrace causal inference, and the OR community is integrating data-driven modeling more and more as an alternative to deductive models. These integrative ambitions would be greatly assisted by a clarification of the differences and commonalities between them. An attempt at integration is Gartner's Maturity Model for Data and Analytics (discussed in Lepenioti et al. 2020), which discerns *descriptive*, *diagnostic*, *predictive*, and *prescriptive* analytics. The popularity of Gartner's model attests to its helpfulness, but we believe that it needs further development. For example, diagnostic analytics refers to causal modeling, and the Gartner model presents it as an antecedent to predictive modeling, but the model does not clarify whether predictive analytics should be based on a causal model, and it does not acknowledge the important differences between predictions based on a causal model versus predictions based on a correlational model.

The distinction between *predictive* and *prescriptive* analytics is popular in business analytics. Den Hertog and Postek (2016), however, note a gap between these two subfields, where predictive analytics typically involves data-driven model building as done in statistics and machine learning, while prescriptive analytics seems to focus on the mathematical optimization of nondata-driven, first-principles models. The authors state that "(…) the deep relation between predictive and prescriptive analytics is neither understood nor exploited" (Den Hertog and Postek 2016).

Since the methods and notation systems used in the three frameworks are so similar, the three types of models are easily confused, or their relations misunderstood. Clarifying the different types of modeling in analytical problem solving is the first purpose of this article. The differences between causal, correlational and deductive models have ramifications for data collection, study design, and the types of solutions that can be derived from them. These ramifications are underappreciated, resulting in flawed analyses and conclusions. Fleshing out these ramifications is the second purpose of the article.

1.3. *Analytical Problem Solving*

Following Ackoff and Vergara (1981), a problem is a choice situation where the problem owner is dissatisfied with the current state of affairs, and is in doubt about which course of action to take. *Analytical problem solving*, as we take it, are approaches to solving problems driven by the modeling of X/Y relationships. Based on Ackoff and Vergara (1981), the elements of a problem are:

1. Controllable variables X^C in the system, which are under the control of the problem solver.
2. Uncontrollable variables X^U, governed by a probability distribution F_U.
3. The outcome, which are one or more variables Y that depend on the X^C and X^U.
4. A problem owner with her value system, who attaches positive or negative value $V(Y, X^C, F_U)$ to a state (X^C, F_U) and its implied outcome Y.
5. Constraints, to which some of the X^C may be subject.

Strategies for analytical problem solving work out how the outcomes Y depend on the X^C and X^U variables, and then use the modeled $Y=f(X^C, X^U)$ relationships as a predictive device to assess the results of various interventions. The task for the problem solver is to determine an intervention in terms of the controllable X^C variables that maximizes the problem owner's value $V(Y, X^C, F_U)$.

All three frameworks discussed above use analytical models $Y=f(X)$ that relate Y to X, but the relations can be causal, correlational or mathematical-deductive. We discuss each of these frameworks, in turn, in the next three sections, partly aiming to capture how they work as problem-solving approaches, and partly aiming to identify relevant differences. Each section begins with two examples, and then discusses how X variables are discovered, how the $Y=f(X)$ relationship can be modeled, and what sort of solutions can be based on the modeled $Y=f(X)$ relationship.

2. Problem Solving Based on Cause-and-Effect Relations

In problem solving based on cause-and-effect modeling, the design of a solution is driven by identified X variables that causally affect Y. Rubin (1974, 2005), Holland (1986) and Pearl (2009) offer formal definitions of causal relations. A less precise definition capturing the essential idea in simpler terms is this: X has a causal effect on Y if modifying X (keeping everything else constant) implies that the probability distribution of Y will change as a consequence. Below, we describe problem solving driven by causal modeling. After introducing two stereotypical examples, we discuss how potential causes can be identified, how their effects can be modeled, and what sort of solutions can be based on the modeled relationships.

2.1. *Stereotypical Examples*

2.1.1. *Paper Helicopters*

A famous example in statistics are George Box's paper helicopters (Box and Liu 1999). The problem is to improve the

design of paper helicopters such that the predicted flight time Y is maximized. In analytical problem solving, this is done by establishing a model $Y=f(X_1^C, X_2^C, X_3^C)$ that relates the flight time to various design parameters such as the wing length (X_1^C), paper weight (X_2^C) and tail length (X_3^C). The relationship is causal, as modifying, say, the wing length will change the flight times as a consequence. Typical analytical strategies involve statistically designed experiments, resulting in a 1st- or 2nd-order polynomial model that approximates $Y=f(X_1^C, X_2^C, X_3^C)$. Based on this model, the problem solver determines settings for the X's, within given constraints, that maximize the predicted mean flight time Y, and thus, $V\left(Y, X^C, F_U\right) = \mu_{Y|X_1^C, X_2^C, X_3^C}$ (the mean flight time for given choices of wing length, paper weight and tail length).

2.1.2. Variation Reduction

In modern manufacturing, tolerances on product characteristics such as dimensions are tight, and thus, a common type of problems in manufacturing is that of excessive variation in product characteristics (MacKay and Steiner 1997). Such variation typically has a multitude of causes, but it is also typical that a few of these causes have disproportionate influence. Identifying what these dominant causes of variation are, can be a challenging detective process. The relevant causal relationship here is $Y=f(X_1^U, X_2^U, X_3^U, \ldots)$, where Y is a product characteristic such as a dimension, and the X_i^U are uncontrollable factors in the production process, whose variability causes variation in Y. Typical value functions are inversely related to the variance of Y, that is, $V(Y, X^C, F_U) \propto \left(\sigma_Y^2\right)^{-1}$, or to the probability that the characteristic Y is within its tolerances: $V(Y, X^C, F_U) \propto P(t_l \leq Y \leq t_u)$. The modeled $Y=f(X_1^U, X_2^U, X_3^U, \ldots)$ relation shows how σ_Y^2 depends on the variances $\sigma_{X_i}^2$ of the uncontrollable variables X_i^U, and it allows one to evaluate various solution scenarios, such as enforcing narrower tolerances on inputs in the production process. Other solution scenarios, such as a feedback or feedforward control mechanism, will be discussed below.

2.2. Identifying the X's

For solving the two problems above, namely, finding optimal settings in the design of paper helicopters or reducing the variability in a product characteristic transmitted from various variation sources, the problem solver may use a statistically designed experiment or observational data to model the $Y=f(X)$ relationships. Before such study can be undertaken, however, it is often necessary to first identify potential causal influence factors (X's) that should be included as factors in the experiment or whose values should be measured in an observational study. Various strategies for discovering potential causes are presented in Table 1 (largely derived from De Mast 2013).

The first approach, *Exploratory data analysis*, has the problem solver identify potential X's from datasets, where correlations between X and Y variables flag the X's as potential influence factors. Other salient patterns in data, such as clusters, could also help the problem solver to discover potential X's (De Mast and Kemper 2013).

Restricting oneself to such data-driven approaches, however, one is likely to miss potential causes for which such data are not available. One will also miss X's that do not vary during the study, and therefore, leave no traces of their effects in the dataset. For example, when paper helicopters are exclusively built from 80 g/m² paper, paper weight will not emerge as a potential influence factor for flight time in studies driven by observational data. The second approach in Table 1, labeled *Experiential knowledge*, therefore complements data-driven approaches. This approach capitalizes on earlier experiences with similar problems, and causal influence factors identified in those earlier problems may be taken as candidate X's for the problem under study. Domain experts are likely to recognize analogies with earlier problems, and the problem solver herself may try to match the problem at hand with earlier cases documented in a library, such as the expert literature, taxonomies of known problems, or even an internet search.

Finding the causal influence factors of a problem can be challenging when the search space is extensive, complex or ill-defined. It's easy, then, to get overwhelmed by the multitude of possibilities or to get bogged down in the wrong part of the search space. In various fields, efficient strategies have been proposed for causal diagnosis that try to narrow down the search by a sequence of studies designed to eliminate whole classes of potential causes at once. Such strategies are called *hierarchical diagnosis* in AI (Chittaro and Ranon 2004), *branch-and-prune* in De Mast (2011), and *eliminate-and-home-in* by Shainin (1993). Table 1 lists such approaches under the name *Sequence of hierarchical elimination studies*. Approaches numbered 4. and 5. in Table 1 will be discussed in subsequent sections.

2.3. Modeling the $Y = f(X)$ Relationship

Given a list of candidate X's, data from an experiment (i.e., an interventional study) are ideal for modeling their causal effects on Y. Well-designed, randomized and controlled experiments directly demonstrate cause-and-effect relations (Holland 1986; Pearl and MacKenzie 2018). Experiments range from simple A/B tests to complex, multi-factor experiments in irregular experimental regions designed by optimal design algorithms (Goos and Jones 2011). Response-surface methodology advises against model building on the basis of single, one-shot experiments, because the design of an experiment hinges on many assumptions concerning the relevant ranges of variables, the presence

Table 1. Approaches for identifying candidate X's.

1. Exploratory data analysis	X's are identified from correlations, clusters and other salient patterns in data
2. Experiential knowledge (domain experts, case libraries)	X's are identified as they were relevant in similar problems previously studied
3. Sequence of hierarchical elimination studies	X's are identified by a narrowing-down process, driven by a sequence of studies aimed at ruling out classes of potential X's
4. Feature engineering	X's are created by merging and transforming features in data sources
5. Deduction from theory	X's are derived from (axiomatic) theory

of higher-order effects, and the reliability of the Y measurements (De Veaux, Hoerl, and Snee 2016). Instead, a sequence of smaller experiments is suggested, where the design of later studies is driven by the findings of earlier studies (Myers, Montgomery, and Anderson-Cook 2009).

Observational, as opposed to experimental, data are a more challenging basis for demonstrating causal effects, as they primarily show correlations. The introduction of additional assumptions and premises sometimes allows one to substantiate a cause-and-effect claim from observational data (Lederer et al. 2019; Hernan, Hsu, and Healy 2019). Pearl's structural causal models, which integrate and generalize such approaches as path analysis, structural equations modeling and Rubin's potential-outcome framework, allow problem solvers to do this in a systematic manner (Pearl 2009, Pearl and MacKenzie 2018). In this approach, the problem solver reasons from domain knowledge to speculate about potential causal relations and their structure, which she visualizes by means of directed graphs. Causal structures are composed of three types of junctions, namely chains, forks and colliders, that imply different conditional probabilities, which may be observable in the system under study. Therefore, it may be possible to rule out alternative potential causal structures from observational data and thereby corroborate a single causal interpretation of the correlations observed between the variables.

2.4. Designing Solutions Based on Causal $Y = f(X)$ Relationships

Once candidate X's have been identified and their effects (if any) onto Y have been modeled, the $Y=f(X)$ model is then used as a basis to design a solution. The literature in industrial engineering and applied statistics describes many standard solution patterns based on $Y=f(X)$ models (e.g., Myers, Montgomery, and Anderson-Cook 2009; MacKay and Steiner 1997). Five solution patterns are listed in Table 2.

Fitted models are often 1st- or 2nd-order polynomial approximations of the general form

$$Y = \beta_0 + \sum \beta_j X_j + \sum \beta_{jj} X_j^2 + \sum \sum \beta_{ij} X_i X_j + \epsilon$$

where ϵ is the model's error term. Such models are called response surfaces (Box and Draper 1987; Myers, Montgomery, and Anderson-Cook 2009), which are prediction devices that predict Y or characteristics of its probability distribution given $\mathbf{X} = (X_1, \ldots, X_k)'$. The first standard solution pattern, named *Response-surface optimization* in Table 2, is where you use the final model to find settings for the X's giving the desired predicted mean Y value $\hat{\mu}_Y$. In the helicopters case, the objective is to find $\mathbf{X}^0 = \arg\max_{X_1,\ldots,X_k} \hat{\mu}_Y(X_1, \ldots, X_k)$, that is, the combination of X settings that maximizes the predicted mean flight

time $\hat{\mu}_Y$. Since response surface models are lower-order polynomials, such optimization problems, given the final model, are often rather straightforward (Derringer and Suich 1980).

The second solution pattern, called *Robust design* in Table 2, or *desensitizing the process to input variation* in MacKay and Steiner (1997), aims to solve a variation problem, as in the second example introduced above (*Variation Reduction*). The concept of robust design is that some of the independent variables are *controllable variables*, which have a fixed setting under the control of the problem solver, and other independent variables are *uncontrollable variables* that randomly vary. Since uncontrollable variables have an effect on Y, their variability may partly be transmitted to Y. Consider a simple case with only one controllable and one uncontrollable variable:

$$Y = \beta_0 + \beta_1 X_1^C + \beta_2 X_2^U + \beta_{12} X_1^C X_2^U + \epsilon$$

with $X_2^U \sim N(\mu_2, \sigma_2^2)$ and $\epsilon \sim N(0, \sigma^2)$ (all random variables independent). Then, $\sigma_Y^2 = (\beta_2 + \beta_{12} X_1^C)^2 \sigma_2^2 + \sigma^2$, and the first term on the right-hand side is the variation transmitted from X_2^U to Y. Note that the chosen setting for the control variable X_1^C affects how much of the variance of X_2^U is transmitted to Y. The setting $X_1^C = -\frac{\beta_2}{\beta_{12}}$ would completely eliminate variation from X_2^U transmitted to Y, and is called a *robust setting*. This exploitation of interaction effects for the reduction of variation transmitted to Y is the essential element of Taguchi's robust-design procedure (Kackar 1985; Myers, Montgomery, and Anderson-Cook 2009), which was highly influential in industrial engineering in the 1980s and 90s.

The third strategy, *Tolerance design*, also aims at variation reduction. Where robust design reduces σ_Y^2 by selecting robust settings for the control variables X^C, in tolerance design we reduce variation in the uncontrollable variables X_j^U themselves. Let

$$Y = \beta_0 + \beta_1 X_1^U + \beta_2 X_2^U + \epsilon,$$

with $X_1^U \sim N(\mu_1, \sigma_1^2)$, $X_2^U \sim N(\mu_2, \sigma_2^2)$ and $\epsilon \sim N(0, \sigma^2)$ independent. Then $\sigma_Y^2 = \beta_1^2 \sigma_1^2 + \beta_2^2 \sigma_2^2 + \sigma^2$. This shows us which of the X_j^U have the largest contribution to the variation in Y, which depends on the variance σ_j^2 of the X_j^U and on their effects β_j. It is these dominant causes of variation that are prime candidates for efforts to reduce variance. Variation in the X_j^U is often controlled by imposing tolerance limits on them. Models such as the one above are used to evaluate a set of tolerance limits for each of the X_j^U and adjust them such that the resulting variance of Y is acceptable.

Table 2 lists a fourth and a fifth standard solution pattern, *Decision optimization* and *Predictive control*, which are not primarily based on a causal $Y=f(X)$ model, and for that reason will be discussed in the next section.

3. Prediction Machines Based on Correlations

When $Y=f(X)$ relations are modeled based on observational (as opposed to experimental) data, they often do not make a causal claim. Such models reflect the correlation structure among variables. They can predict Y for the population from

Table 2. Generic solution patterns based on $Y=f(X)$ relationships.

1. Response-surface optimization	Find settings for the X^C yielding satisfactory $\hat{\mu}_Y$
2. Robust design	Find settings for the X^C minimizing σ_Y^2
3. Tolerance design	Reduce variation σ_j^2 in one or more of the X_j^U
4. Decision optimization	Use predicted Y to guide decisions
5. Predictive control	Compensate for drifts in Y by adjusting a control variable X^C

which the data were sampled, but they cannot predict the effects of interventions (as we discuss below). Since most of the solutions in Table 2 are based on interventions, the options for solving problems based on correlational models are more limited.

In machine learning and AI, models and algorithms that develop a correlational predictive $Y=f(X)$ model from observational data are called supervised learning. They include random forests, support-vector machines, regularized regression and neural networks. Typically, these models are much more complex than the models used in traditional cause-and-effect modeling, both in the number of X's that they have and in the number of model parameters. Random forests and neural networks may have thousands or millions of parameters. This allows them to handle high-dimensional data streams that have emerged as sensors, storage and computing power have become easily available. Such models are also better in handling non-linear relationships, which makes them effective for modeling $Y=f(X)$ relationships where the X's are not numerical variables, but less structured information such as images and natural-language text.

3.1. Stereotypical Examples

3.1.1. Predicting Passenger Numbers
A railway company develops an app that helps passengers find empty seats in the train. At the heart of the application is a correlational predictive model, that predicts the number Y of passengers in a train compartment from measured CO_2 levels in the air and other X's. CO_2 is a good predictor for the number of passengers, but it is not a causal influence factor: increasing CO_2 levels by artificially releasing extra CO_2 would not increase the number of passengers. The cause-and-effect mechanism is actually the reverse, with passengers as causal influence and CO_2 as the effect.

3.1.2. Predictive Maintenance
One of the challenges in maintenance of systems like trains is timing when a component should be replaced. If the replacement is scheduled too late, the component breaks down in the field, which may result in very high costs. Therefore, components are preferably replaced preventively, when this can be done at a convenient moment and in a convenient place. However, premature replacement wastes useful lifetime of the component, so the preventive replacement should be scheduled as late as possible. Predictive maintenance is based on a model that predicts when a breakdown is imminent, for example, by monitoring a condition such as vibration patterns or the concentration of ferrous particles in lubrication (Carden and Fanning 2004). In our framework, the solution strategy is driven by a correlational predictive model $Y=f(X^U)$, with Y the probability of a breakdown in the next epoch, and X^U a predictor such as a vibration characteristic or particles concentration. The controllable variable X^C is the timing of the replacement. We try to minimize the total expected cost, which is a weighted average of the cost of a breakdown in the field and the cost of a preventive replacement (weighed by the probability that a breakdown occurs).

3.2. Identifying the X's

Before a model $Y=f(X)$ can be fitted, we first need to identify potential predictors. Often in machine learning practice, the data for fitting a model are assumed *given*, and the collection of new data is deemed impossible or not needed (De Veaux, Hoerl, and Snee 2016). Identifying the X's, then, boils down to selecting, merging, transforming and rescaling features recorded in the available data sources into variables to be used in the model-building effort, a task, that is, described as *Feature engineering* in Table 1. Feature engineering is done by human analysts, but the promise of deep-learning techniques is that stacked neural networks can be trained to recognize structure and features in raw data automatically, at least within certain domains such as image recognition and natural-language processing.

In addition to feature engineering, frameworks for data science such as CRISP-DM (Chapman et al. 2000) recommend to consult domain experts, which is similar to approach number 2 in Table 1 (*Experiential knowledge*). Domain experts may help identify X's not yet represented in available data sources, for which it may yet be possible to obtain data either by initiating new measurements or by acquiring data sources in which they are represented. Other approaches suggested in machine learning are techniques such as principal-components analysis, t-SNE and auto encoders, which facilitate the discovery of candidate X's by revealing structure in data through clustering and dimensionality reduction (James et al. 2013). Such techniques for *unsupervised learning* are similar to *Exploratory data analysis* (approach number 1 in Table 1).

3.3. Modeling the $Y = f(X)$ Relationship

Random forests, neural networks, support-vector machines and other models popular in machine learning do not make a causal claim. The predictions are essentially based on the correlation (association) structure of X and Y variables: high CO_2 levels co-occur with large numbers of passengers, but they do not cause them, and ferrous particles may predict a breakdown, but they do not cause it.

This has important ramifications for inference. First, the model should be fitted on a representative dataset, in which the correlation structure is representative for the correlations in the target population (the universe of future observations that the model is claimed to predict). This implies that, in general, experiments are unsuited, since by deliberately setting the levels of the X's, they break the correlation structure among the X's and potentially the Y. For example, a randomized controlled experiment studying the *effect* (sic) of CO_2 on the number of passengers would have the experimenter manipulate CO_2 levels according to an experimental design, and then measure the corresponding passenger numbers (note how inappropriate the term *effect* is in this context of correlational modeling). By manipulating the CO_2 levels, their relationship with passenger numbers is perturbed, and the algorithm cannot be used to predict passenger numbers in normal situations. Second, the fitted model can only be trusted within the population from which the dataset was sampled. The algorithm predicting passengers from CO_2 levels does not give reliable predictions for passenger

numbers in different types of train compartments (or for aircraft cabins, boat cabins, etc.).

Instead of causality, machine learning focuses on predictive accuracy and generalizability (James et al. 2013). The latter refers to the model's predictive accuracy for a new dataset generated by the same data generating mechanism. Machine learning uses a variety of metrics for expressing predictive accuracy, such as the mean-squared error MSE, the coefficient of determination R^2, or the precision and recall pair. To guard against the detrimental effect overfitting has on generalizability, machine learning uses the train-test split method and cross-validation to regulate a model's complexity.

3.4. Designing Solutions Based on Correlational $Y = f(X)$ Models

Correlation models do not support solutions based on interventions in the X's: the fact that lower CO_2 levels co-occur with lower passenger numbers does not imply that reducing the CO_2 level would make the compartment less crowded. Correlation models can only make predictions where all X's were generated by the same data-generating mechanism as in the target population, with the same correlation structure among the X's. Making predictions about situations where the X's are not generated by the data-generating mechanism, but by an intervention by the experimenter, is called counterfactual prediction (Hernan, Hsu, and Healy 2019) or interventional inference (Pearl 2009). Pearl (2009) also introduced the "do" notation to discriminate between $P(Y|X)$ (where both Y and X are generated by the data-generating mechanism of the target population) and $P(Y|do(X))$ (where X is set by an intervention, such as a randomization procedure in an experiment, or a problem solver setting it to a specific value $X=x_0$). Causal models allow the calculation of $P(Y|do(X))$, whereas correlational models only allow the calculation of $P(Y|X)$. For this reason, correlational models are not a sound basis for the first three solution patterns in Table 2 (response-surface optimization, robust design, and tolerance design).

Business analytics and data science often conceptualize the utility of predictive models in decision-theoretic frameworks, where the utility of a predictive model is in informing decision making. Thus, knowing the passenger numbers (and, consequently, the availability of seats) in train compartments allows passengers to make more informed decisions on where to look for a seat. Let Y_i be the number of available seats in compartment i, which is predicted by $\hat{Y}_i = f(X_i^U)$, with X_i^U the (uncontrollable) CO_2 in that compartment. The controllable variable X^C, here, is in which compartment a passenger will look for a seat. A simple algorithm that tells passengers to go to the compartment with the highest predicted number of available seats ($X_C = \arg\max_i \hat{Y}_i$) maximizes the probability to find an available seat. Similarly, for maintenance planning, the value function $V(Y, X^C, F_U)$ is the inverse of the expected total cost. The decision variable X^C is the timing of the replacement, which is optimized based on the predicted probability Y of a breakdown in the next epoch, which is predicted from a predictor X^U such as a vibration characteristic or particles concentration. In Table 2, such solutions are called *Decision optimization* (pattern no. 4).

The final pattern in Table 2 is *Predictive control*. The setting here is that we aim to reduce variation in Y, for example, a process variable that continues to drift away from its target value. Predictive control combines a model that forecasts drifts in Y with a causal model that informs us how to intervene and compensate for the drift. Let

$$Y = \beta_0 + \beta_1 X_1^U + \beta_2 X_2^C + \epsilon,$$

with X_1^U an uncontrollable random variable that causes Y to drift away from target, and X_2^C a controllable variable that we use to readjust the process. In a feedforward control mechanism, we predict Y from the measured uncontrollable variable X_1^U. If X_1^U deviates by $\Delta_1 = 1$ from its nominal value, then the model predicts that Y will deviate by $\Delta_Y = \beta_1$ from its target, which we compensate by an adjustment of $\Delta_2 = -\frac{\beta_1}{\beta_2}$ in X_2^C of the controllable variable. Note that, while the relationship between X_1^U and Y does not need to be causal, but can be correlational, the relationship between the adjustment variable X_2^C and Y needs to be one of cause and effect.

Also *feedback* control is based on such adjustments by a controllable variable X_2^C, but now the predicted Δ_Y is not based on a model relating it to variation in an uncontrollable variable X^U. Instead, the predicted Δ_Y is based on an ARIMA or other time-series model (Box, Jenkins, and Reinsel 1994). The output value Y_t at time t is predicted from output values at previous time instants $Y_{t-1}, Y_{t-2}, \ldots$. Equations for optimal adjustment schemes have been derived in statistics (Box and Kramer 1992) and control engineering (Åström and Hägglund 2001).

4. Deductive Relationships $Y = f(X)$

Optimizing the design of paper helicopters by empirically modeling the $Y = f(X)$ relations is a popular example in statistics. Engineers, however, may object to such "black-box" approach, and point out that much of the behavior of paper helicopters could be derived from the laws of aerodynamics and other physics theory. *White-box* or *first-principles* models are not based on data, but are derived from theory. Also, *gray-box* or *hybrid* models are partly derived from theory, and partly based on empirical model-building. Annis (2005), for example, models the behavior of paper helicopters by combining a deductive analysis based on aerodynamics theory with empirical models based on statistically designed experiments.

First-principles models follow the logic of Hempel's deductive-nomological theory (Hempel and Oppenheim 1948; Woodward 2019). This model was the culmination of philosophy of science in the first half of the 20th century, and championed by Karl Popper and Carl Hempel. It was an attempt to capture scientific explanation in a framework that avoids the concept of causality, but yet allows interventional prediction. In the deductive-nomological framework, we start with universal laws, such as the laws of physics, the laws of probability, or the axioms of mathematics. Given a set of particular circumstances, we then derive the phenomena to be explained or predicted from the universal laws by mathematical deduction.

Laws typically do not specify a causal direction, but instead, they state that the ratios of certain variables are constant, as in the general gas law $pV = cT$ (the product of the pressure and

volume of an ideal gas is equal to its absolute temperature times a constant). Since laws are universal (i.e., exceptionless), they can be used, however, to predict the effect of an intervention. From the general gas law, we can predict that if we heat up a gas contained in a cylinder with a frictionless piston, its volume will increase (the pressure remaining constant), thus preserving the equality stated in the law. From the same law, we can also predict that if we compress the gas, the temperature and/or the pressure will go up (depending on the boundary conditions imposed on the system). Note that in the second intervention the causal direction is the reverse of the direction in the first intervention, thus demonstrating that laws themselves do not specify a causal direction. When used in analytical problem solving, such conservation or differential equations models are rewritten to reflect that some characteristics are outcomes (Y), taking the remaining characteristics as independent variables (X).

After the 1960s, the deductive-nomological model was seen more and more as an idealized rather than realistic version of scientific explanation, although it seems to apply rather well to modern physics. Both causal and deductive models are *explanatory models*, as they give understanding by showing how observed phenomena were to be expected as consequences of either a cause-and-effect mechanism or a universal law (Kitcher 1998; Woodward 2019).

4.1. Stereotypical Examples

4.1.1. Optimizing Appointment Schedules

Too tight scheduling of appointments (e.g., in outpatient clinics) results in congestion and waiting queues. If the appointment time slots are too long, however, the service provider may be idle too often. The goal is to achieve a good balance between expected waiting time for customers and idle time for service providers by optimizing the appointment times (Kuiper et al. inpress). The relation between expected waiting and idle time (the Y's) and the appointment times (X^C), given unpredictability in service times, no-shows and random walk-ins (the X^U), could be determined empirically by running an experiment (causal modeling) or by training a machine-learning algorithm on a historical dataset (correlational model). But given the distribution of the service times, the expected waiting and idle time implied by any combination of appointment times can be deduced from probability theory, and this is what most OR experts would do. This is an explanatory model in the deductive-nomological framework, where the mean waiting and idle times are derived from queueing theory (universal laws) and the service-time distribution and a set of appointment times (particular circumstances).

4.1.2. Bin-Packing Problems

The objective in bin-packing problems (Korte and Vygen 2006) is to pack N items of various, known volumes into K bins, whose volumes are also given, using as few bins as possible. In this classic OR problem, a solution is an allocation $X_i^C \in \{1, \ldots, K\}$ the bin number to which item i is assigned. A feasible solution is a solution where the sum of the volumes of the assigned items per bin does not exceed the volume of that bin. The Y variable is the number of bins used in a solution, $Y\left(X_1^C, \ldots, X_N^C\right) = $ n-unique$(X_1^C, \ldots, X_N^C)$, and the objective is to find $\left(x_1^0, \ldots, x_N^0\right) = \arg \min Y(X_1^C, \ldots, X_N^C)$. Note that in this problem, the identification of the X's and the modeling of their relationship with Y are not driven by data. Instead, it is rather straightforward to derive the $Y = f(X)$ relationship from mathematics. The nontrivial part of this problem is, given $Y = f(X)$, finding the optimum $\left(x_1^0, \ldots, x_N^0\right)$. Where finding an optimal set of appointment times requires the optimization of a convex function (Kuiper et al. in press), the computation of an optimal allocation of items to bins is NP-hard, which means that solving the problem is at least as hard as the hardest problems in NP. This in turn means that there is no known algorithm that computes an optimal feasible solution in polynomial time. The challenge is to find an algorithm that finds an optimal feasible solution in acceptable computation time for as many instances as possible and a near-optimal feasible solution for the other instances.

4.2. Identifying the X's and Modeling the $Y = f(X)$ Relationship

The $Y = f(X)$ relationship is not empirically modeled from a dataset consisting of observed (X, Y) tuples. Instead, the relationship is derived by deductive reasoning from an axiomatic theory, such as queueing theory, the laws of physics, or combinatorial mathematics.

Sometimes it is possible to fully derive the $Y = f(X)$ function analytically from premises, mathematics and laws. When this becomes intractable, however, one approach is to resort to approximations. For example, the relation between waiting and idle times and appointment times can be determined by approximating the probability distributions of the service times by phase-type distributions (which are convolutions of the exponential distribution and, therefore, analytically convenient; Kuiper et al. in press). Alternatively, one could develop a simulation model that relates Y to X values. Discrete-event simulation is often used to obtain the expected waiting and idle time for given appointment times (Ahmadi-Javid, Jalali, and Klassen 2017).

Note that data are not used to model the relationship between X and Y, although you may need some data to estimate some constants in the equations. For example, laws that relate waiting and idle time to appointment times typically have constants such as the coefficient of variation of the service times, which you would need to estimate from data. Nondata-driven model building is standard practice in OR and business analytics. On the one hand, if it is possible to reduce the behavior of a system to a logical consequence of a set of universal laws, you could say that this gives a superior sense of understanding than an ad-hoc empirical model fitted in isolation. This thought may capture the uneasiness sometimes expressed by mathematicians and engineers about models fitted empirically. Also, Schölkopf et al. (2021) argue that models derived from first principles generalize better across different environments and tasks. On the other hand, Den Hertog and Postek (2016) and Simchi-Levi (2014) appeal to the OR community to become more data driven, and they argue for the advantages of empirical model building, especially when it is difficult to derive complex and

hidden constraints and relations from theory alone, or when the complexity of the system under study makes deductive model building unwieldy.

4.3. Designing Solutions Based on Deductive $Y = f(X)$ Models

Although deductive models may not directly appeal to causality, their being derived from universal laws warrants using them to make interventional predictions. Therefore, they are a solid basis for any of the solution patterns in Table 2.

In response-surface modeling, the final model is usually a lower-order polynomial. Finding the optimum $\mathbf{X}^0 = \arg\max_{X_1,\ldots,X_k} \hat{\mu}_Y(X_1, \ldots, X_k)$, then, can be done in polynomial time. Finding an optimal allocation of items to bins in the bin-packing problem is an example of a much harder optimization problem, where there is no known polynomial-time algorithm to compute an optimal solution. Business analytics and OR discern several classes of computational complexity of solving such optimization problems, such as *P*, *NP* and *PSPACE* (Arora and Barak 2009). When there is no known polynomial-time algorithm, the challenge is to identify an algorithm that finds an optimal or near-optimal feasible solution in acceptable computation time. A widely employed approach to meet this challenge uses heuristics. For example, for the bin-packing problem the literature offers a rich edifice of heuristics and approximation algorithms such as the Harmonic-*k* algorithm (Lee and Lee 1985). More general methods like Mixed Integer Programming (Wolsey 1998), Constraint Programming (Baptiste, Le Pape, and Nuijten 2001), Local Search, and Large Neighborhood Search (Godard, Laborie, and Nuijten 2005) are used to solve NP-hard problems.

Models deduced from first-principles theory may be fully deterministic, as in the bin-packing problem. This means that the model's terms do not include random variables, and the model's "predictions" (for lack of a better word), therefore, are a fixed value for *Y*. Other models do contain random components, either because some of the *X*'s are random variables or because the model contains a random error term. The model, then, does not return a fixed *Y* value, but parameters of its distribution F_Y such as the mean μ_Y, the variance σ_Y^2 or the probability $P(t_l \leq Y \leq t_u)$ that *Y* is within its tolerance limits. In the appointment-scheduling problem, the service times as well as the arrivals of clients are random variables, and consequently, the model does not predict fixed waiting and idle times, but their means. The classical bin-packing problem treats the items' volumes as fixed and given, but a more realistic version of the problem acknowledges that the volumes are subject to random measurement error, and this could be incorporated as a random component in the problem's modeling.

5. Discussion and Conclusions

We have discussed three frameworks for analytical problem solving in business and industry: cause-and-effect modeling, correlational predictive modeling, and deductive modeling. Below, we recapitulate relevant differences between these three types of modeling, and flesh out their implications for the problem-solving process.

5.1. Essential Differences Between the Three Types of Models

The differences between causal, correlational and deductive models have consequences for the design of the analysis and for the solution strategy.

5.1.1. Appropriate Study Design for the Analysis

The first essential difference is the type of study or data needed to establish the $Y = f(X)$ relation. For causal models, experiments are ideal, and the validity of models based on such studies hinges on the proper application of the principles of experimental design, such as randomization, blocking and replication (Box and Draper 1987). Observational studies could be used provided that these can be augmented to warrant causal inference, for example, using structural causal modeling (Lederer et al. 2019; Pearl 2009).

For correlational predictive models, experiments are generally unsuited, as they disrupt the correlation structure, and you need observational data sampled from a data-generating mechanism that is representative for the target population. Careful definition of the target population and an appropriate sampling mechanism make or break the validity of such models (MacKay and Oldford 2000; De Veaux, Hoerl, and Snee 2016).

Mathematical deductive models are not based on statistical studies, whether experimental or observational, but instead are derived from axiomatic theory. The challenge here is in finding a model that is simple enough to be tractable, and at the same time complex enough to capture essential characteristics and produce useful predictions. Typical strategies to make complex relations tractable include approximation and simulation. When the complexity of the system under study makes analytical modeling problematic, or when theory and first principles are not complete enough to derive complex constraints and relations, problem solvers should consider resorting to experimental or observational studies to model all or part of the relationships (Den Hertog and Postek 2016).

5.1.2. Type of Solution Strategy

The second essential difference between the three types of models, is which solution strategies they support. Causal models and deductive models allow interventional predictions (i.e., predictions in "what if" scenarios), and therefore are a good basis for optimizing settings for the X^C parameters, such as determining an optimal combination of settings for the design parameters of paper helicopters, or optimal appointment times in a schedule. The first three solution patterns in Table 2 (response-surface optimization, robust design, and tolerance design) need a causal or deductive model.

Solution patterns 4 and 5 (decision optimization and predictive control) can be based on either a causal or deductive model, or a correlational model. Correlational models, such as a supervised learning model from machine learning, may often have an edge here. Namely, when it comes to predictive accuracy, causal and deductive models are often outperformed by the more

Table 3. Overview of the process of analytical problem solving.

1. Find Y ("What are the relevant outcome variables?")

2. Find X ("What are relevant independent variables?")

Causal model	Correlational model	Deductive model
• Exploratory data analysis • Experiential knowledge • Hierarchical elimination studies	• Exploratory data analysis (unsupervised learning) • Experiential knowledge • Feature engineering	• Deduction from theory

3. Find F_Y and F_U ("What is the current state?")

4. Find $Y=f(X)$ ("How are Y and X related?")

Causal model	Correlational model	Deductive model
• Randomized controlled experiments • Structural causal modeling based on observational data	• Observational study (supervised learning), based on representative sample from target population	• Deductive derivation from axiomatic theory • Approximation • Simulation

5. Find $V(Y, X^C, F_U)$ ("What is the objective?")

6. Find max $V(Y, X^C, F_U)$ ("How will we solve the problem?")

Causal model	Correlational model	Deductive model
• Response-surface optimization • Robust design • Tolerance design • Decision optimization • Predictive control	• Decision optimization • Predictive control	• Response-surface optimization • Robust design • Tolerance design • Decision optimization • Predictive control

complex predictive models used in machine learning, which can use any measurable feature as a potential predictor, whether its relationship with Y is causal or not.

Thus, the choice between a causal, correlational or deductive model could be driven by the type of solution strategy that one aims for, which in turn will then determine what sort of study is needed to establish the $Y=f(X)$ model. Reversely, it could also be the (im)possibilities to collect data that drive the choice for a type of model, which then in turn may limit the options for a solution strategy. Namely, if randomized controlled experiments are not feasible, and if it is not possible to collect observational data suitable for causal inference, this rules out causal modeling. If the system under study is too complex to be approximated well by deductions from first principles, this makes a deductive approach unwieldy. And if representative data are unavailable and cannot be gathered by doing experiments, deriving the $Y=f(X)$ model deductively from theory may be the only option.

Table 3 places the three types of $Y=f(X)$ models and their essential differences in the context of the process of problem solving. The literature proposes many models for the process of analytical problem solving and the tasks that it comprises. These models range from practical models, such as Six Sigma's DMAIC model in operations improvement and industrial engineering (De Mast and Lokkerbol 2012) and CRISP-DM in data science (Chapman et al. 2000), and models in the academic literature, including Smith (1988) and Marksberry, Bustle, and Clevinger (2011). The online supplemental material discusses the problem-solving tasks enumerated in Table 3 in more detail, linking the table to the academic literature.

5.2. Descriptive, Predictive and Prescriptive Analytics

Gartner (Lepenioti et al. 2020) and Davenport and Harris (2017) popularized the widely embraced categorization of problems into descriptive, predictive and prescriptive analytics. We believe

that our account has the potential to clarify what sort of analytics tasks these categories involve, and where this framework should be augmented in order for it to reflect essential differences between various types of analytics.

Descriptive analytics, characterized by Davenport and Harris (2017) by questions such "What happened?," "How many, how often, when?" and "What exactly is the problem?," are problems involving some or all of the tasks 1, 2, and 3. in Table 3, and they seek to describe the current state, summarize or identify relevant outcomes, or link them to independent variables. In the process of analytical problem solving, this is a preliminary stage, where a problem is identified and defined in terms of X and Y variables.

Predictive analytics involve the establishment of a $Y=f(X)$ model as in task 4. in Table 3. This may be a correlational predictive model, which bases predictions on correlations between predictors and outcomes. It could also be an explanatory model, which grounds predictions in an understanding of why things happen by relating outcomes either to their causes (cause-and-effect model) or to the laws from which they can be derived (deductive model). Establishing an explanatory $Y=f(X)$ model as a prelude to prediction is called diagnostic analytics in the Gartner model.

Prescriptive analytics refers to tasks 5. and 6. in Table 3 (determine the objective and design a solution strategy). Davenport and Harris (2017) mainly seem to have interventional solutions in mind, described as "experimental design" and "optimization" by them, which are equivalent to our response-surface optimization, robust design and tolerance design. Note that solutions may also be reactive, as in decision optimization and predictive control, and these strategies do not require a causal or deductive model.

5.3. Conclusions

The discussion in this article showcases the diversity of the analytical and statistical disciplines, and we have emphasized their

complementarity. The statistical sciences bring to the table concepts and techniques for modeling and inference under uncertainty and stochasticity. Machine learning and statistical learning offer powerful algorithms for correlational predictive modeling. Applied statistics has traditionally focused on causal modeling and sophisticated experimentation strategies (although also many models in applied statistics are correlational). Business analytics and OR bring a rich edifice of techniques for "hard" optimization tasks, as well as frameworks for structuring the problem-solving process.

The word *analysis* means: an investigation of the component parts of a whole and their relations in making up the whole. Analytical problem solving revolves around the idea to relate the behavior of an outcome Y to the factors that cause or at least predict it and then design a solution informed by these $Y=f(X)$ relationships. This mirrors the traditional motto of empirical research, the purpose of which is often described as to explain empirical phenomena, which allows us to predict and therefore control them (see typical introductions into empirical research, such as Kerlinger and Lee 2000). This traditional notion has gotten competition from purely correlational, nonexplanatory models, which are unsuited for interventional solution strategies, but which allow solutions driven by decision optimization and predictive control.

Funding

References

Ackoff, R. L., and Vergara, E. (1981), "Creativity in Problem Solving and Planning: A Review," *European Journal of Operational Research*, 7, 1–13. [52]

Ahmadi-Javid, A., Jalali, A., and Klassen, K. (2017), "Outpatient Appointment Systems in Healthcare: A Review of Optimization Studies," *European Journal of Operational Research*, 258, 3–34. [57]

Annis, D. H. (2005), "Rethinking the Paper Helicopter: Combining Statistical and Engineering Knowledge," *The American Statistician*, 59, 320–326. [56]

Arora, S., and Barak, B. (2009), *Computational Complexity: A Modern Approach*, Cambridge: Cambridge University Press. [58]

Åström, K. J., and Hägglund, T. (2001), "The Future of PID Control," *Control Engineering Practice*, 9, 1163–1175. [56]

Baptiste, P., Le Pape, C., and Nuijten, W. (2001), *Constraint-Based Scheduling: Applying Constraint Programming to Scheduling Problems*, New York: Springer. [58]

Box, G. E. P., and Kramer, T. (1992), "Statistical Process Monitoring and Feedback Adjustment: A Discussion," *Technometrics*, 34, 251–267. [56]

Box, G. E. P., and Liu, P. Y. T. (1999), "Statistics as a Catalyst to Learning by Scientific Method, Part I – An Example," *Journal of Quality Technology*, 31, 1–15. [52]

Box, G. E. P., and Draper, N. R. (1987), *Empirical Model-Building and Response Surfaces*, London: Wiley. [54,58]

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994), *Time Series Analysis*, New Jersey: Prentice Hall. [56]

Breiman, L. (2001), "Statistical Modelling: The Two Cultures," *Statistical Science*, 16, 199–231. [51]

Carden, E. P., and Fanning, P. (2004), "Vibration Based Condition Monitoring: A Review," *Structural Health Monitoring*, 3, 355–377. [55]

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. R., and Wirth, R. (2000), *CRISP-DM 1.0: Step-by-step Data Mining Guide*, The CRISP-DM Consortium. [55,59]

Chittaro, L., and Ranon, R. (2004), "Hierarchical model-Based Diagnosis Based on Structural Abstraction," *Artificial Intelligence*, 155, 147–182. [53]

Davenport, T., and Harris, J. (2017), *Competing on Analytics: The New Science of Winning* (updated ed.), Boston, MA: Harvard Business School Press. [59]

De Mast, J. (2011), "The Tactical Use of Constraints and Structure in Diagnostic Problem Solving," *Omega*, 39, 702–709. [53]

——— (2013), "Diagnostic Quality Problem Solving: A Conceptual Framework and Six Strategies," *Quality Management Journal*, 20, 21–36. [51,53]

De Mast, J., and Kemper, B. P. H. (2013), "Principles of Exploratory Data Analysis: What Can We Learn From a Well-Known Case?," *Quality Engineering*, 21, 366–375. [53]

De Mast, J., and Lokkerbol, J. (2012), "An Analysis of the Six Sigma DMAIC Method from the Perspective Of Problem Solving," *International Journal of Production Economics*, 139, 604–614. [51,59]

De Veaux, R. D., Hoerl, R. W., and Snee, R. D. (2016), "Big Data and the Missing Links," *Statistical Analysis and Data Mining*, 9, 411–416. [54,55,58]

Den Hertog, D., and Postek, K. (2016), "Bridging the Gap Between Predictive and Prescriptive Analytics—New Optimization Methodology Needed," Research Report, Tilburg University. [51,52,57,58]

Derringer, G., and Suich, R. (1980), "Simultaneous Optimization of Several Response Variables," *Journal of Quality Technology*, 12, 214–219. [54]

Estrada-Flores, S., Merts, I., De Keterlaere, B., and Lammertyn, J. (2006), "Development and Validation of "Grey-Box" Models for Refrigeration Applications: A Review of Key Concepts," *International Journal of Refrigeration*, 29, 931–946. [52]

Godard, D., Laborie, P., and Nuijten, W. (2005), "Randomized Large Neighborhood Search for Cumulative Scheduling," *Proc. 15th International Conference on Automated Planning and Scheduling* (ICAPS 05), pp. 81–89. [58]

Goos, P., and Jones, B. (2011), *Optimal Design of Experiments: A Case Study Approach*, London: Wiley. [53]

Hempel, C., and Oppenheim, P. (1948), "Studies in the Logic of Explanation," *Philosophy of Science*, 15 135–175. [56]

Hernan, M. A., Hsu, J., and Healy, B. (2019), "A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks," *Chance*, 32, 42–49. [51,54,56]

Hoerl, R. W., and Vining, G. G. (2020). "The Journey to Establish the Discipline of Statistical Engineering," *Applied Statistical Models in Business and Industry*,1–12. [51]

Holland, P. W. (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81, 945–960. [52,53]

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013), *An Introduction to Statistical Learning*, New York: Springer. [55,56]

Juran, J. M. (1998), "The Quality Improvement Process," in *Juran's Quality Handbook* (5th ed.), eds. J. M. Juran, A. B. Godfrey, New York: McGraw-Hill, pp. 51–74. [51]

Kackar, R. N. (1985), "Off-Line Quality Control, Parameter Design and the Taguchi Method" (with discussion), *Journal of Quality Technology*, 17, 176–209. [54]

Kerlinger, F. N., and Lee, H. B. (2000), *Foundations of Behavioral Research* (4th ed.). Fort Worth, TX: Harcourt College Publishers. [60]

Kitcher, P. (1998). "Explanation," in *Routledge Encyclopedia of Philosophy*, ed. E. Craig, (vol. 3), London:Routledge. [57]

Korte, B., and Vygen, J. (2006), "Bin-Packing," in *Combinatorial Optimization: Theory and Algorithms* (chap. 18), New York: Springer. [57]

Kuiper, A., Mandjes, M., De Mast, J., and Brokkelkamp, R. (in press), "A Flexible and Optimal Approach for Appointment Scheduling in Healthcare," *Decision Sciences Journal*. DOI: 10.1111/deci.12517. [57]

Lederer, D. J., Bell, S. C., Branson, R. D., Chalmers, J. D., Marshall, R., Maslove, D. M., Ost, D. E., Punjabi, N. M., Schatz, M., Smyth, A. R., and Stewart, P. W. (2019), "Control of Confounding and Reporting of Results in Causal Inference Studies," *Annals of the American Thoracic Society*, 16, 22–28. [54,58]

Lee, C. C., and Lee, D. T. (1985), "A Simple On-Line Bin-Packing Algorithm," *Journal of the ACM*, 32, 562–572. [58]

Lepenioti, K., Bousdekis, A., Apostolou, D., and Mentzas, G. (2020), "Prescriptive Analytics: Literature Review and Research Challenges," *International Journal of Information Management*, 50, 57–70. [52,59]

MacDuffie, J. P. (1997), "The Road to 'Root Cause': Shop-Floor Problem-Solving at Three Auto Assembly Plants," *Management Science*, 43, 479–502. [51]

MacKay, R. J., and Oldford, R. W. (2000), "Scientific Method, Statistical Method, and the Speed of Light," *Statistical Science,* 15, 254–278. [58]

MacKay, R. J., and Steiner, S. H. (1997), "Strategies for Variability Reduction," *Quality Engineering*, 10, 125–136. [53,54]

Marksberry, P., Bustle, J., and Clevinger, J. (2011), "Problem Solving for Managers: A Mathematical Investigation of Toyota's 8-Step Process," *Journal of Manufacturing Technology Management*, 22, 837–852. [59]

Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M. (2009), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments* (4th ed.), London: Wiley. [54]

Pearl, J. (2009), "Causal Inference in Statistics: An Overview," *Statistical Surveys*, 3, 96–146. [52,54,56,58]

Pearl, J., and MacKenzie, D. (2018), *The Book of Why—The New Science of Cause and Effect*, New York: Basic Books. [53,54]

Rubin, D. (1974), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701. [52]

——— (2005), "Causal Inference Using Potential Outcomes," *Journal of the American Statistical Association*, 100, 322–331. [52]

Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., and Bengio, Y. (2021), "Towards Causal Representation Learning," *Proceedings of the IEEE,* 109, 612–634. [57]

Shainin, R. (1993), "Strategies for Technical Problem Solving," *Quality Engineering*, 5, 433-448. [53]

Shmueli, G. (2010), "To Explain or to Predict?," *Statistical Science*, 25, 289–310. [51]

Simchi-Levi, D. (2014), "OM Research: From Problem-Driven to Data-Driven Research," *Manufacturing and Service Operations Management*, 16, 2–10. [57]

Smith, G. F. (1988), "Towards a Heuristic Theory of Problem Structuring," *Management Science*, 34, 1489–1506. [51,59]

——— (1998), "Determining the Cause of Quality Problems: Lessons from Diagnostic Disciplines," *Quality Management Journal*, 5, 24-41. [51]

Steiner, S. H., MacKay, R. J., and Ramberg, J. S. (2008), "An Overview of the Shainin System for Quality Improvement," *Quality Engineering*, 20, 6–19. [51]

Wolsey, L. A. (1998), *Integer Programming*, New York: Wiley. [58]

Woodward, J. (2019), "Scientific Explanation," in *The Stanford Encyclopedia of Philosophy* (Winter 2019 ed.), ed. E. N. Zalta, Stanford, CA: Stanford University, available at *https://plato.stanford.edu/archives/win2019/entries/scientific-explanation*. [56,57]

World Economic Forum (2020), *The Future of Jobs Report 2020*, Geneva: WEF. [51]

Supplemental material to the paper:

Analytical problem solving based on causal, correlational and deductive models

Jeroen de Mast, Stefan H. Steiner, Wim P. M. Nuijten and Daniel Kapitan

A. The process of analytical problem solving

The paper discusses three frameworks for analytical problem solving: causal modelling, correlational modelling and deductive modelling. Table 3 in the paper summarizes the discussion by placing the three types of modelling in the process of analytical problem solving. In this supplemental material, we explain the structure of Table 3 and relate it to the literature.

Section 5.1 of the paper discusses essential differences between the three types of modelling. Table 3 organizes these in the problem-solving process. The literatures in various applied fields abound in models of the problem-solving process, typically organized in step-wise procedures. Table 4 collates four such procedures:

- Smith (1988), one of the best researched models in the management literature.
- Six Sigma's DMAIC model (De Mast and Lokkerbol 2012), widely used and discussed in quality engineering and statistics-driven operations improvement.
- The CRISP-DM model in data science (Chapman et al. 2000).
- PDCA and Toyota's A3 problem-solving procedure (Marksberry, Bustle and Clevinger 2011), which is representative for many similar problem-solving procedures in practitioner books.

Table 4. Four models for the process of problem solving

Smith	DMAIC	CRISP-DM	PDCA / A3
1. Goal setting	1. Define	1. Business understanding	1. Clarify problem
2. Description of current situation	2. Measure		2. Break down problem
3. Diagnosis (causal explanations)	3. Analyze	2. Data understanding	3. Set targets
4. Generation of alternative solutions	4. Improve	3. Data preparation	4. Root cause analysis
5. Prediction of the effectiveness of solutions	5. Control	4. Modelling	5. Develop countermeasures
6. Evaluation of alternative solutions		5. Evaluation	6. See countermeasures through
7. Implementation and gaining acceptance for chosen solution.		6. Deployment	7. Monitor results and processes
			8. Standardize successful project

Even though most of such models order problem-solving tasks in a numbered sequence of steps, they acknowledge almost universally that actual problem-solving processes are more iterative, with much going to and fro between tasks in cycles in which understanding of the problem becomes more and more refined.

Table 3 in the paper breaks down the task of finding improvements based on $Y = f(X)$ relations into six generic tasks, which are the rows of the table. Below, we explain each of these tasks in more detail.

1. Find Y ("What are relevant outcome variables?")
Driven by domain expertise and the problem owner's needs, the problem solver identifies measurable dependent (Y) variables that may be relevant for the problem. Typical challenges include clarifying implicit, vague or incoherent problem definitions and finding a reliable method for measuring Y. Typical techniques include problem structuring (Mingers and Rosenhead 2004), flowdown diagrams (De Koning and De Mast 2007) and measurement system assessment (Wheeler and Lyday 1989).

2. Find X ("What are relevant independent variables?")
The problem solver identifies candidate X variables that may be useful for predicting Y or understanding its behavior. This task echoes *Causal Diagnosis* in Smith's (1988) model (see Table 3) and *Root cause analysis* in the PDCA / A3 model, but identified X's can be any kind of predictors, causal or not. Some variables will be under the control of the problem solver (X^C), while other are uncontrollable variables (X^U). A typical challenge is avoiding tunnel vision, where the problem solver only identifies X's of a similar kind, and misses entire classes of potential predictors. Table 1 in the paper lists various approaches for identifying potential X's:

- Exploratory data analysis (unsupervised learning, graphical techniques, dimension reduction)
- Experiential knowledge (domain experts, case libraries)
- Sequence of hierarchical elimination studies (eliminate-and-home-in, branch-and-prune)
- Feature engineering
- Deduction from axiomatic theory

3. Find F_Y and F_U ("What is the current state?")
The problem solver investigates the current state and establishes the distributions $F_Y(y) = P(Y \leq y)$ and $F_U(x) = P(X^U \leq x)$ of the Y and X^U variables (including their means, variances, and other relevant statistical properties). A good characterization of the current state helps the problem solver refine her understanding of the problem and achieve focus by considering questions such as: Does the problem manifest itself in the mean or in the variance of Y? Does the problem evolve over time? Can the problem be broken down into subproblems, and which subproblem is the most severe? To what degree can the X's be controlled? Which X's have fixed values, and which have substantial variance? Knowing the means and variances of the Y and X^U variables also helps assess whether the problem owner's perception of the problem and desired end state are realistic. Typical techniques include descriptive statistics and estimation, the process-capability analysis (Wu, Pearn and Kotz 2009) and baseline studies (Steiner and MacKay 2005).

4. Find $Y = f(X)$ ("How are X and Y related?")
The problem solver builds a model that predicts Y (or F_Y) from the X's. This can be a causal, correlational or deductive model. The paper discusses suitable ways of building a predictive model:

Causal model	**Correlational model**	**Deductive model**
• Randomized controlled experiments • Structural causal modelling based on observational data	Observational study (supervised learning), based on representative sample from target population	• Deductive derivation from axiomatic theory • Approximation • Simulation

Typical challenges include causal inference, the bias/variance trade-off, and the analytical tractability of deductive models.

5. Find $V(Y, X^C, F_U)$ ("What is the objective?")
The problem solver translates the problem owner's objectives into a question about relationships between dependent variables Y, and controllable and uncontrollable independent variables X^C and X^U. The problem owner's desired end state is captured in a value function $V(Y, X^C, F_U)$ that is to be maximized. The problem solver also establishes constraints on the X's, and whether they should be treated as a controllable or uncontrollable variable.

6. Find $\max V(Y, X^C, F_U)$ ("How will we solve the problem?")
The various types of predictive models support various solution patterns, as listed in Table 2, which involve that the problem solver maximizes value for the problem owner by intervening in some of the controllable X^C variables. Interventions can be one-time optimizations (as in response-surface optimization) or continual dynamic adjustments as in predictive control.

Causal model	**Correlational model**	**Deductive model**
• Response-surface optimization • Robust design • Tolerance design • Decision optimization • Predictive control	• Decision optimization • Predictive control	• Response-surface optimization • Robust design • Tolerance design • Decision optimization • Predictive control

References

CHAPMAN, P., CLINTON, J., KERBER, R., KHABAZA, T., REINARTZ, T., SHEARER, C.R. and WIRTH, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. The CRISP-DM Consortium.

DE KONING, H. AND DE MAST, J. (2007). The CTQ flowdown as a conceptual model of project objectives. *Quality Management Journal* **14**(2) 19–28.

DE MAST, J. and LOKKERBOL, J. (2012). An analysis of the Six Sigma DMAIC method from the perspective of problem solving. *International Journal of Production Economics* **139**(2) 604–614.

MARKSBERRY, P., BUSTLE, J. and CLEVINGER, J. (2011). Problem solving for managers: A mathematical investigation of Toyota's 8-step process. *Journal of Manufacturing Technology Management* **22**(7) 837–852.

MINGERS, J. AND ROSENHEAD, J. (2004). Problem structuring methods in action. *European Journal of Operational Research* **152** 530–554.

SMITH, G.F. (1988). Towards a heuristic theory of problem structuring. *Management Science* **34** 1489–1506.

STEINER, S.H. and MACKAY, R.J. (2005). *Statistical Engineering: An Algorithm for Reducing Variation in Manufacturing Processes*. Quality Press: Milwaukee.

WHEELER, D.J. and LYDAY, R.W. (1989). *Evaluating the Measurement Process*, 2nd edition. SPC Press: Knoxville.

WU, C.W., PEARN, W.L. and KOTZ, S. (2009). An overview of theory and practice on process capability indices for quality assurance. *International Journal of Production Economics* **117**(2), 338–359.