Turning Numbers into Knowledge

MASTERING THE ART OF PROBLEM SOLVING

Second Edition

JONATHAN G. KOOMEY, PH.D.

Analytics Press
PO Box 20313
Oakland, CA 94620-0313
http://www.analyticspress.com
http://www.numbersintoknowledge.com

CHAPTER

28

MAKE A MODEL

In this book, the term "model" refers to any conceptual (usually mathematical) model that can be used to understand the world. Models can range in scope from the back of an envelope to millions of lines of computer code. Anthony Starfield et al., in their book *How to Model It*, summarize nicely when they state "An explicit model is a laboratory for the imagination." Models allow you to experiment, test, and learn about the world in a systematic way without leaving your office.

SIMPLE MODELS AND BURIED ASSUMPTIONS

A model is an equation or set of equations that describe how something you care about changes when other things change. Below is a useful model that allows you to estimate gasoline consumption as a function of how many miles you drive and the efficiency of your car:

$$\text{Gas used per week (gallons)} = \frac{\text{Miles driven}}{\text{Week}} \times \frac{\text{Gallon}}{\text{Miles}}$$

You can estimate your car's average fuel efficiency based on past experience in a typical week. If you drive 100 miles per week, and your car consumes five gallons of gasoline in that period, your car gets 20 miles per gallon (mpg) on average. For every additional 100 miles you drive, you will consume another five gallons of gasoline.

This model is perfectly adequate for many purposes, but a more detailed model may be required in some cases. Say that you want to estimate your gas consumption for a special week-long, 1,000 mile trip that will involve much more highway driving than would your normal pattern. The following model would then be useful:

$$\text{Gas used} = \frac{\text{City miles driven}}{\text{Week}} \times \frac{\text{Gallon}}{\text{City miles}} + \frac{\text{Highway miles driven}}{\text{Week}} \times \frac{\text{Gallon}}{\text{Highway miles}}$$

Let's say that the car's efficiency is 15 mpg in the city and 25 mpg on the highway, and that your big trip is 95% highway driving, as shown in **Table 28.1** (about 67% of the driving is on the highway in the typical driving pattern embodied in the simpler model). The more complex model indicates that the vehicle would consume a total of 41.3 gallons, while the simpler model (which is based on your typical driving pattern) yields an estimate of 50 gallons when applied to your special trip. The simpler model therefore overestimates gasoline use by 21% in this case.

There are many other subtleties that could be incorporated into such a model, which would be introduced as additional terms in the equation. Neither one of these models is precisely correct in a scientific sense although the second one is more generally applicable. Either can be appropriate in certain situations.

The assumption about how much city and highway driving you do is *buried* in the first model. It's easy to see how someone not aware of the difference in automobile efficiency in city and highway driving could misuse the first model and calculate an erroneous result. This lesson is a more general one about models: assumptions are often buried and can lead you astray.

TABLE 28.1. Simpler and more complex models of automobile fuel use

	Driving Mode	*Simpler model*	*More complex model*	*Ratio of Simpler to More complex*
Miles per gallon	City		15	
	Highway		25	
	Average	20	24.2	0.83
Miles driven	City		50	
	Highway		950	
	Total	1000	1000	1.00
Estimated gallons consumed	City		3.3	
	Highway		38	
	Total	50	41.3	1.21

SIMPLIFY, SIMPLIFY, SIMPLIFY

People are often intimidated by the use of computer models to conduct an analysis, but never forget that these models depend on data and assumptions that are based in large part on analysts' judgment. It is a rare model that does not contain dozens or hundreds of such assumptions; in fact, these data and assumptions determine the outcome in many cases. This heavy reliance on assumptions for parameters that are unknown is most prevalent in models used in business and public policy, but even models of the physical world can rely on assumptions that influence or determine the outcome.

In his real estate practice, Robert J. Ringer encountered sellers with fantastic expectations for the investment value of their buildings:

> As a general but very consistent rule, owners are so unrealistic when it comes to vacancy factors, replacement costs, and other expense items that are not readily ascertainable that I normally considered their projections and operating statements to be virtually meaningless.[63]

Using his electronic calculator and his knowledge of real estate, Ringer was able to determine quickly whether the building was salable at or near the owner's asking price. His simple conceptual model allowed him to create a quick estimate of net cash flow on the proverbial back of an envelope. This screening **29**
technique allowed him to avoid wasting time on deals that would never close because of the huge gap between the reality of the project's net cash flow and the fantasy of the owner's expectations.

The technique of using a simple and transparent model to evaluate the results of a more complex calculation is immensely powerful. Analysts at the International Institute for Applied Systems Analysis (IIASA) found this out to their chagrin when Will Keepin, a Visiting Scholar at IIASA, was able to almost exactly reproduce the results of a multiyear, multimillion dollar study using some of the study's key input assumptions and a hand calculator. Keepin showed that the study's results followed directly from the input assumptions. He concluded that the study's projections of future energy supply "are opinion, rather than credible scientific analysis, and they therefore cannot be relied upon by policy makers seeking a genuine understanding of the energy choices for tomorrow."

Beware of big complicated models and the results they produce. Generally they involve so much work to keep them current that not enough time is spent
26 on data compilation and scenario analysis. Morgan and Henrion, in their book *Uncertainty*, devoted an entire chapter to such models and began by summarizing this fundamental truth:

> There are some models, especially some science and engineering models, that are large or complex because they need to be. But many more are large or complex because their authors gave too little thought to why and how they were being built and how they would be used.[64]

Such large models are essential only for the most complex and esoteric analyses, and a simpler model will usually serve as well (and be more understandable to your intended audience).

The importance of transparency cannot be overestimated. A model that your audience can actually grasp is inherently more persuasive than a "black box" that no one outside of a small circle of analysts understands. Transparent mod-
33 els for which the input data and assumptions are also well documented are even more compelling, but are, sadly, all too rare.

Don't be too impressed by a model's complexity. Instead, ask about the data
26 and assumptions used to create scenarios. Focus on the coherence of the scenarios and their relevance to your decisions, and ignore the marketing doublespeak of those whose obsession with tools outweighs their concern with useful results. Sadly, many modelers fall into this latter camp, and you would do well to avoid their work.

Programming today is a race between software engineers trying to build bigger and better idiot-proof programs, and the universe trying to produce bigger and better idiots. So far, the universe is winning.

— **RICH COOK**

To reference this work, please cite Koomey, Jonathan. 2008. *Turning Numbers into Knowledge: Mastering the Art of Problem Solving*. 2nd ed. Oakland, CA: Analytics Press. <http://www.analyticspress.com>