An Introduction to Model Building

1.1 An Introduction to Modeling

Operations research (often referred to as **management science**) is simply a scientific approach to decision making that seeks to best design and operate a system, usually under conditions requiring the allocation of scarce resources.

By a **system,** we mean an organization of interdependent components that work together to accomplish the goal of the system. For example, Ford Motor Company is a system whose goal consists of maximizing the profit that can be earned by producing quality vehicles.

The term *operations research* was coined during World War II when British military leaders asked scientists and engineers to analyze several military problems such as the deployment of radar and the management of convoy, bombing, antisubmarine, and mining operations.

The scientific approach to decision making usually involves the use of one or more **mathematical models.** A mathematical model is a mathematical representation of an actual situation that may be used to make better decisions or simply to understand the actual situation better. The following example should clarify many of the key terms used to describe mathematical models.

EXAMPLE 1

Eli Daisy produces Wozac in huge batches by heating a chemical mixture in a pressurized container. Each time a batch is processed, a different amount of Wozac is produced. The amount produced is the *process yield* (measured in pounds). Daisy is interested in understanding the factors that influence the yield of the Wozac production process. Describe a model-building process for this situation.

Solution Daisy is first interested in determining the factors that influence the yield of the process. This would be referred to as a *descriptive model,* because it describes the behavior of the actual yield as a function of various factors. Daisy might determine (using regression methods discussed in Chapter 24) that the following factors influence yield:

- container volume in liters (V)
- container pressure in milliliters (P)
- container temperature in degrees Celsius (T)
- chemical composition of the processed mixture

If we let A, B, and C be percentage of mixture made up of chemicals A, B, and C, then Daisy might find, for example, that

$$(1)\ \text{yield} = 300 + .8V + .01P + .06T + .001T*P - .01T^2 - .001P^2 + 11.7A + 9.4B + 16.4C + 19A*B + 11.4A*C - 9.6B*C$$

To determine this relationship, the yield of the process would have to be measured for many different combinations of the previously listed factors. Knowledge of this equation would enable Daisy to describe the yield of the production process once volume, pressure, temperature, and chemical composition were known.

Prescriptive or Optimization Models

Most of the models discussed in this book will be **prescriptive** or **optimization** models. A prescriptive model "prescribes" behavior for an organization that will enable it to best meet its goal(s). The components of a prescriptive model include

- objective function(s)
- decision variables
- constraints

In short, an optimization model seeks to find values of the decision variables that optimize (maximize or minimize) an objective function among the set of all values for the decision variables that satisfy the given constraints.

The Objective Function

Naturally, Daisy would like to maximize the yield of the process. In most models, there will be a function we wish to maximize or minimize. This function is called the model's *objective function.* Of course, to maximize the process yield we need to find the values of V, P, T, A, B, and C that make (1) as large as possible.

In many situations, an organization may have more than one objective. For example, in assigning students to the two high schools in Bloomington, Indiana, the Monroe County School Board stated that the assignment of students involved the following objectives:

- Equalize the number of students at the two high schools.
- Minimize the average distance students travel to school.
- Have a diverse student body at both high schools.

Multiple objective decision-making problems are discussed in Sections 4.14 and 11.13.

The Decision Variables

The variables whose values are under our control and influence the performance of the system are called *decision variables*. In our example, V, P, T, A, B, and C are decision variables. Most of this book will be devoted to a discussion of how to determine the value of decision variables that maximize (sometimes minimize) an objective function.

Constraints

In most situations, only certain values of decision variables are possible. For example, certain volume, pressure, and temperature combinations might be unsafe. Also, A B, and C must be nonnegative numbers that add to 1. Restrictions on the values of decision variables are called *constraints.* Suppose the following:

- Volume must be between 1 and 5 liters.
- Pressure must be between 200 and 400 milliliters.
- Temperature must be between 100 and 200 degrees Celsius.
- Mixture must be made up entirely of A, B, and C.
- For the drug to properly perform, only half the mixture at most can be product A.

These constraints can be expressed mathematically by the following constraints:

$$V \leq 5$$
$$V \geq 1$$
$$P \leq 400$$
$$P \geq 200$$
$$T \leq 200$$
$$T \geq 100$$
$$A \geq 0$$
$$B \geq 0$$
$$A + B + C = 1$$
$$A \leq 5$$

The Complete Optimization Model

After letting z represent the value of the objective function, our entire optimization model may be written as follows:

$$\text{Maximize } z = 300 + .8V + .01P + .06T + .001T*P - .01T^2 - .001P^2 + 11.7A + 9.4B + 16.4C + 19A*B + 11.4A*C - 9.6B*C$$

Subject to (s.t.)

$$V \leq 5$$
$$V \geq 1$$
$$P \leq 400$$
$$P \geq 200$$
$$T \leq 200$$
$$T \geq 100$$
$$A \geq 0$$
$$B \geq 0$$
$$C \geq 0$$
$$A + B + C = 1$$
$$A \leq 5$$

Any specification of the decision variables that satisfies all of the model's constraints is said to be in the **feasible region.** For example, V = 2, P = 300, T = 150, A = .4, B = .3, and C = .1 is in the feasible region. An **optimal solution** to an optimization model is any point in the feasible region that optimizes (in this case, *maximizes*) the objective function. Using the LINGO package that comes with this book, it can be determined that the optimal solution to this model is V = 5, P = 200, T = 100, A = .294, B = 0, C = .706, and z = 183.38. Thus, a maximum yield of 183.38 pounds can be obtained with a 5-liter container, pressure of 200 milliliters, temperature of 100 degrees Celsius, and 29% A and 71% C. This means no other feasible combination of decision variables can obtain a yield exceeding 183.38 pounds.

Static and Dynamic Models

A **static model** is one in which the decision variables do not involve sequences of decisions over multiple periods. A **dynamic model** is a model in which the decision variables *do* involve sequences of decisions over multiple periods. Basically, in a static model we solve a "one-shot" problem whose solutions prescribe optimal values of decision variables at all points in time. Example 1 is an example of a static model; the optimal solution will tell Daisy how to maximize yield at all points in time.

For an example of a dynamic model, consider a company (call it Sailco) that must determine how to minimize the cost of meeting (on time) the demand for sailboats during the next year. Clearly Sailco's must determine how many sailboats it will produce during each of the next four quarters. Sailco's decisions involve decisions made over multiple periods, hence a model of Sailco's problem (see Section 3.10) would be a dynamic model.

Linear and Nonlinear Models

Suppose that whenever decision variables appear in the objective function and in the constraints of an optimization model, the decision variables are always multiplied by constants and added together. Such a model is a **linear model.** If an optimization model is not linear, then it is a **nonlinear model.** In the constraints of Example 1, the decision variables are always multiplied by constants and added together. Thus, Example 1's constraints pass the test for a linear model. However, in the objective function for Example 1, the terms .001T*P, $-.01T^2$, 19A*B, 11.4A*C, and -9.6B*C make the model nonlinear. In general, nonlinear models are much harder to solve than linear models. We will discuss linear models in Chapters 2 through 10. Nonlinear models will be discussed in Chapter 11.

Integer and Noninteger Models

If one or more decision variables must be integer, then we say that an optimization model is an **integer model.** If all the decision variables are free to assume fractional values, then the optimization model is a **noninteger** model. Clearly, volume, temperature, pressure, and percentage composition of our inputs may all assume fractional values. Thus, Example 1 is a noninteger model. If the decision variables in a model represent the number of workers starting work during each shift at a fast-food restaurant, then clearly we have an integer model. Integer models are much harder to solve than nonlinear models. They will be discussed in detail in Chapter 9.

Deterministic and Stochastic Models

Suppose that for any value of the decision variables, the value of the objective function and whether or not the constraints are satisfied is known with certainty. We then have a **deterministic model.** If this is not the case, then we have a **stochastic model.** All models in the first 12 chapters will be deterministic models. Stochastic models are covered in Chapters 13, 16, 17, and 19–24.

If we view Example 1 as a deterministic model, then we are making the (unrealistic) assumption that for given values of V, P, T, A, B, and C, the process yield will always be the same. This is highly unlikely. We can view (1) as a representation of the *average* yield of the process for given values of the decision variables. Then our objective is to find values of the decision variables that maximize the average yield of the process.

We can often gain useful insights into optimal decisions by using a deterministic model in a situation where a stochastic model is more appropriate. Consider Sailco's problem of minimizing the cost of meeting the demand (on time) for sailboats. The uncertainty about future demand for sailboats implies that for a given production schedule, we do not know whether demand is met on time. This leads us to believe that a stochastic model is needed to model Sailco's situation. We will see in Section 3.10, however, that we can develop a deterministic model for this situation that yields good decisions for Sailco.

1.2 The Seven-Step Model-Building Process

When operations research is used to solve an organization's problem, the following seven-step model-building procedure should be followed:

Step 1: Formulate the Problem The operations researcher first defines the organization's problem. Defining the problem includes specifying the organization's objectives and the parts of the organization that must be studied before the problem can be solved. In Example 1, the problem was to determine how to maximize the yield from a batch of Wozac.

Step 2: Observe the System Next, the operations researcher collects data to estimate the value of parameters that affect the organization's problem. These estimates are used to develop (in step 3) and evaluate (in step 4) a mathematical model of the organization's problem. For example, in Example 1, data would be collected in an attempt to determine how the values of T, P, V, A, B, and C influence process yield.

Step 3: Formulate a Mathematical Model of the Problem In this step, the operations researcher develops a mathematical model of the problem. In this book, we will describe many mathematical techniques that can be used to model systems. For Example 1, our optimization model would be the result of step 3.

Step 4: Verify the Model and Use the Model for Prediction The operations researcher now tries to determine if the mathematical model developed in step 3 is an accurate representation of reality. For example, to validate our model, we might check and see if (1) accurately represents yield for values of the decision variables that were not used to estimate (1). Even if a model is valid for the current situation, we must be aware of blindly applying it. For example, if the government placed new restrictions on Wozac, then we might have to add new constraints to our model, and the yield of the process [and Equation (1)] might change.

Step 5: Select a Suitable Alternative Given a model and a set of alternatives, the operations researcher now chooses the alternative that best meets the organization's objectives. (There may be more than one!) For instance, our model enabled us to determine that yield was maximized with $V = 5$, $P = 200$, $T = 100$, $A = .294$, $B = 0$, $C = .706$, and $z = 183.38$.

Step 6: Present the Results and Conclusion of the Study to the Organization In this step, the operations researcher presents the model and recommendation from step 5 to the decision-making individual or group. In some situations, one might present several alternatives and let the organization choose the one that best meets its needs. After presenting the results of the operations research study, the analyst may find that the organization does not approve of the recommendation. This may result from incorrect definition of the organization's problems or from failure to involve the decision maker from the start of the project. In this case, the operations researcher should return to step 1, 2, or 3.

Step 7: Implement and Evaluate Recommendations If the organization has accepted the study, then the analyst aids in implementing the recommendations. The system must be constantly monitored (and updated dynamically as the environment changes) to ensure that the recommendations enable the organization to meet its objectives.

In what follows, we discuss three successful management science applications. We will give a detailed (but nonquantitative) description of each application. We will tie our discussion of each application to the seven-step model-building process described in Section 1.2.

1.3 CITGO Petroleum

Klingman et al. (1987) applied a variety of management-science techniques to CITGO Petroleum. Their work saved the company an estimated $70 million per year. CITGO is an oil-refining and -marketing company that was purchased by Southland Corporation (the owners of the 7-Eleven stores). We will focus on two aspects of the CITGO team's work:

1 a mathematical model to optimize operation of CITGO's refineries, and

2 a mathematical model—supply distribution marketing (SDM) system—that was used to develop an 11-week supply, distribution, and marketing plan for the entire business.

Optimizing Refinery Operations

Step 1 Klingman et al. wanted to minimize the cost of operating CITGO's refineries.

Step 2 The Lake Charles, Louisiana, refinery was closely observed in an attempt to estimate key relationships such as:

1 How the cost of producing each of CITGO's products (motor fuel, no. 2 fuel oil, turbine fuel, naptha, and several blended motor fuels) depends on the inputs used to produce each product.

2 The amount of energy needed to produce each product. This required the installation of a new metering system.

3 The yield associated with each input–output combination. For example, if 1 gallon of crude oil would yield .52 gallons of motor fuel, then the yield would equal 52%.

4 To reduce maintenance costs, data were collected on parts inventories and equipment breakdowns. Obtaining accurate data required the installation of a new database-management system and integrated maintenance-information system. A process control system was also installed to accurately monitor the inputs and resources used to manufacture each product.

Step 3 Using linear programming (LP), a model was developed to optimize refinery operations. The model determines the cost-minimizing method for mixing or blending together inputs to produce desired outputs. The model contains **constraints** that ensure that inputs are blended so that each output is of the desired quality. Blending constraints are discussed in Section 3.8. The model ensures that plant capacities are not exceeded and allows for the fact that each refinery may carry an inventory of each end product. Sections 3.10 and 4.12 discuss inventory constraints.

Step 4 To validate the model, inputs and outputs from the Lake Charles refinery were collected for one month. Given the actual inputs used at the refinery during that month, the actual outputs were compared to those predicted by the model. After extensive changes, the model's predicted outputs were close to the actual outputs.

Step 5 Running the LP yielded a daily strategy for running the refinery. For instance, the model might, say, produce 400,000 gallons of turbine fuel using 300,000 gallons of crude 1 and 200,000 gallons of crude 2.

Steps 6 and 7 Once the database and process control were in place, the model was used to guide day-to-day refinery operations. CITGO estimated that the overall benefits of the refinery system exceeded $50 million annually.

The Supply Distribution Marketing (SDM) System

Step 1 CITGO wanted a mathematical model that could be used to make supply, distribution, and marketing decisions such as:

1 Where should crude oil be purchased?

2 Where should products be sold?

3 What price should be charged for products?

4 How much of each product should be held in inventory?

The goal, of course, was to maximize the profitability associated with these decisions.

Step 2 A database that kept track of sales, inventory, trades, and exchanges of all refined products was installed. Also, regression analysis (see Chapter 24) was used to develop forecasts for wholesale prices and wholesale demand for each CITGO product.

Steps 3 and 5 A minimum-cost network flow model (MCNFM) (see Section 7.4) is used to determine an 11-week supply, marketing, and distribution strategy. The model makes all decisions mentioned in step 1. A typical model run that involved 3,000 equations and 15,000 decision variables required only 30 seconds on an IBM 4381.

Step 4 The forecasting modules are continuously evaluated to ensure that they continue to give accurate forecasts.

Steps 6 and 7 Implementing the SDM required several organizational changes. A new vice-president was appointed to coordinate the operation of the SDM and LP refinery model. The product supply and product scheduling departments were combined to improve communication and information flow.

1.4 San Francisco Police Department Scheduling

Taylor and Huxley (1989) developed a police patrol scheduling system (PPSS). All San Francisco (SF) police precincts use PPSS to schedule their officers. It is estimated that PPSS saves the SF police more than $5 million annually. Other cities such as Virginia

Beach, Virginia, and Richmond, California, have also adopted PPSS. Following our seven-step model-building procedure, here is a description of PPSS.

Step 1 The SFPD wanted a method to schedule patrol officers in each precinct that would quickly produce (in less than one hour) a schedule and graphically display it. The program should first determine the personnel requirements for each hour of the week. For example, 38 officers might be needed between 1 A.M. and 2 A.M. Sunday but only 14 officers might be needed from 4 A.M. to 5 A.M. Sunday. Officers should then be scheduled to minimize the sum over each hour of the week of the shortages and surpluses relative to the needed number of officers. For example, if 20 officers were assigned to the midnight to 8 A.M. Sunday shift, we would have a shortage of $38 - 20 = 18$ officers from 1 to 2 A.M. and a surplus of $20 - 14 = 6$ officers from 4 to 5 A.M. A secondary criterion was to minimize the maximum shortage because a shortage of 10 officers during a single hour is far more serious than a shortage of one officer during 10 different hours. The SFPD also wanted a scheduling system that precinct captains could easily fine-tune to produce the optimal schedule.

Step 2 The SFPD had a sophisticated computer-aided dispatch (CAD) system to keep track of all calls for police help, police travel time, police response time, and so on. SFPD had a standard percentage of time that administrators felt each officer should be busy. Using CAD, it is easy to determine the number of workers needed each hour. Suppose, for example, an officer should be busy 80% of the time and CAD indicates that 30.4 hours of work come in from 4 to 5 A.M. Sunday. Then we need 38 officers from 4 to 5 A.M. on Sunday [$.8*(38) = 30.4$ hours].

Step 3 An LP model was formulated (see Section 3.5 for a discussion of scheduling models). As discussed in step 1, the primary objective was to minimize the sum of hourly shortages and surpluses. At first, schedulers assumed that officers worked five consecutive days for eight hours a day (this was the policy prior to PPSS) and that there were three shift starting times (say, 6 A.M., 2 P.M., and 10 A.M.). The constraints in the PPSS model reflected the limited number of officers available and the relationship of the number of officers working each hour to the shortages and surpluses for that hour. Then PPSS would produce a schedule that would tell the precinct captain how many officers should start work at each possible shift time. For example, PPSS might say that 20 officers should start work at 6 A.M. Monday (working 6 A.M.–2 P.M. Monday–Friday) and 30 officers should start work at 2 P.M. Saturday (working 2 P.M.–10 P.M. Saturday–Wednesday). The fact that the number of officers assigned to a start time must be an integer made it far more difficult to find an optimal schedule. (Problems in which decision variables must be integers are discussed in Chapter 9.)

Step 4 Before implementing PPSS, the SFPD tested the PPSS schedules against manually created schedules. PPSS produced an approximately 50% reduction in both surpluses and shortages. This convinced the department to implement PPSS.

Step 5 Given the starting times for shifts and the type of work schedule [four consecutive days for 10 hours per day (the 4/10 schedule) or five consecutive days for eight hours per day (the 5/8 schedule)], PPSS can produce a schedule that minimizes the sum of shortages and surpluses. More important, PPSS can be used to experiment with shift times and work rules. Using PPSS, it was found that if only three shift times are allowed, then a 5/8 schedule was superior to a 4/10 schedule. If, however, five shift times were allowed, then a 4/10 schedule was found to be superior. This finding was of critical importance because police officers had wanted to switch to a 4/10 schedule for years. The city had resisted 4/10 schedules because they appeared to reduce productivity. PPSS showed that 4/10 schedules need not reduce productivity. After the introduction of PPSS, the SFPD went to 4/10 schedules and *improved productivity!* PPSS also enables the department to experiment with a mix of one-officer and two-officer patrol cars.

Steps 6 and 7 It is estimated that PPSS created an extra 170,000 productive hours per year, thereby saving the city of San Francisco $5.2 million per year. Ninety-six percent of all workers preferred PPSS generated schedules to manually generated schedules. PPSS enabled SFPD to make strategic changes (such as adopting the 4/10 schedule), which made officers happier and increased productivity. Response times to calls improved by 20% after PPSS was adopted.

A major reason for the success of PPSS was that the system allowed precinct captains to fine-tune the computer-generated schedule and obtain a new schedule in less than one minute. For example, precinct captains could easily add or delete officers and add or delete shifts and quickly see how these changes modified the master schedule.

1.5 GE Capital

GE Capital provides credit card service to 50 million accounts. The average total outstanding balance exceeds $12 billion. GE Capital, led by Makuch et al. (1989), developed the PAYMENT system to reduce delinquent accounts and the cost of collecting from delinquent accounts.

Step 1 At any one time, GE Capital has more than $1 billion in delinquent accounts. The company spends $100 million per year processing these accounts. Each day, workers contact more than 200,000 delinquent credit card holders with letters, messages, or live calls. The company's goal was to reduce delinquent accounts and the cost of processing them. To do this, GE Capital needed to come up with a method of assigning scarce labor resources to delinquent accounts. For example, PAYMENT determines which delinquent accounts receive live phone calls and which delinquent accounts receive no contact.

Step 2 The key to modeling delinquent accounts is the concept of a **delinquency movement matrix (DMM).** The DMM determines how the probability of the payment on a delinquent account during the current month depends on the following factors: size of unpaid balance (either $<\$300$ or $\geq\$300$), action taken (no action, live phone call, taped message, letters), and a performance score (high, medium, or low). The higher the performance score associated with a delinquent account, the more likely the account is to be collected. Table 1 lists the probabilities for a $250 account that is two months delinquent, has a high performance score, and is contacted with a phone message.

TABLE 1
Sample Entries in DMM

Event	Probability
Account completely paid	.30
One month is paid	.40
Nothing is paid	.30

Because GE Capital has millions of delinquent accounts, there is ample data to accurately estimate the DMM. For example, suppose there were 10,000 two-month delinquent accounts with balances under $300 that have a high performance score and are contacted with phone messages. If 3,000 of those accounts were completely paid off during the current month, then we would estimate the probability of an account being completely paid off during the current month as 3,000/10,000 = .30.

Step 3 GE Capital developed a linear optimization model. The objective function for the PAYMENT model was to maximize the expected delinquent accounts collected during the next six months. The decision variables represented the fraction of each type of delinquent account (accounts are classified by payment balance, performance score, and months delinquent) that experienced each type of contact (no action, live phone call, taped message, or letter). The constraints in the PAYMENT model ensure that available resources are not overused. Constraints also relate the number of each type of delinquent account present in, say, January to the number of delinquent accounts of each type present during the next month (February). This **dynamic** aspect of the PAYMENT model is crucial to its success. Without this aspect, the model would simply "skim" the accounts that are easiest to collect each month. This would result in few collections during later months.

Step 4 PAYMENT was piloted on a $62 million portfolio for a single department store. GE Capital managers came up with their own strategies for allocating resources (collectively called CHAMPION). The store's delinquent accounts were randomly assigned to the CHAMPION and PAYMENT strategies. PAYMENT used more live phone calls and more "no action" than the CHAMPION strategies. PAYMENT also collected $180,000 per month more than any of the CHAMPION strategies, a 5% to 7% improvement. Note that using more of the no-action strategy certainly leads to a long-run increase in customer goodwill!

Step 5 As described in step 3, for each type of account, PAYMENT tells the credit managers the fraction that should receive each type of contact. For example, for three-month delinquent accounts with a small (<$300) unpaid balance and high performance score, PAYMENT might prescribe 30% no action, 20% letters, 30% phone messages, and 20% live phone calls.

Steps 6 and 7 PAYMENT was next applied to the 18 million accounts of the $4.6 billion Montgomery-Ward department store portfolio. Comparing the collection results to the same time period a year earlier, it was found that PAYMENT increased collections by $1.6 million per month (more than $19 million per year). This is actually a conservative estimate of the benefit obtained from PAYMENT, because PAYMENT was first applied to the Montgomery-Ward portfolio during the depths of a recession—and a recession makes it much more difficult to collect delinquent accounts.

Overall, GE Capital estimates that PAYMENT increased collections by $37 million per year and used fewer resources than previous strategies.

REFERENCES

Klingman, D., N. Phillips, D. Steiger, and W. Young, "The Successful Deployment of Management Science Throughout Citgo Corporation," *Interfaces* 17 (1987, no. 1):4–25.

Makuch, W., J. Dodge, J. Ecker, D. Granfors, and G. Hahn, "Managing Consumer Credit Delinquency in the US Economy: A Multi-Billion Dollar Management Science Application," *Interfaces* 22 (1992, no. 1):90–109.

Taylor, P., and S. Huxley, "A Break from Tradition for the San Francisco Police: Patrol Officer Scheduling Using an Optimization-Based Decision Support Tool," *Interfaces* 19 (1989, no. 1):4–24.