I’ve read several books on cost analysis and worked through decision analysis problems in some of my college classes. Now that I own my own business, I realize that there was one important thing that I always took for granted in doing those problems. We were always given the data. Now I know that doing the analysis once you have the data is the easier part. How are the costs determined? How do I know if they are fixed or variable? I am trying to decide whether to open a new store and I need answers to these questions.

I thought about the importance of being able to determine fixed and variable costs after reading an article about, of all things, the costs of text messaging [see the In Action item “The Variable Cost of a Text Message” on the next page]. The article talked about the low variable costs of sending text messages and the implications for pricing services. Although I am in a different industry, the basic principles still apply.

Charlene Cooper owns Charlene’s Computer Care (3C), a network of computer service centers located throughout the South. Charlene is thinking about opening a new center and has asked you to help her make a decision. She especially wants your help estimating the costs to use in the analysis.

**Why Estimate Costs?**

When managers make decisions, they need to compare the costs (and benefits) among alternative actions. Therefore, managers need to estimate the costs associated with each alternative. We saw in Chapter 4 that good decisions require good information about costs; the better these estimates, the better the decision managers will make. In this chapter, we discuss how to estimate the cost data required for decision making. Cost estimates can be an important element in helping managers make decisions that add value to the company.

**Basic Cost Behavior Patterns**

The most important characteristic of costs for decision making is how they behave—how they vary with activity is the key distinction for decision making. Therefore, the basic idea in cost estimation is to estimate the relation between costs and the variables affecting costs, the cost drivers. We focus on the relation between costs and one important variable that affects them: activity level. Activities can be measured by volume (for example, units of output, machine-hours, pages typed, miles driven), by complexity (for example, number of different products, number of components in a product), or by any other cost driver.

You already know the key terms for describing cost behavior: **variable costs** and **fixed costs**. You also know that variable costs change proportionately with activity levels but fixed costs do not. Building on that, the formula that we use to estimate costs is the familiar cost equation:

\[ TC = F + VX \]

where \( TC \) refers to total costs, \( F \) refers to fixed costs that do not vary with activity levels, \( V \) refers to variable costs per unit of activity, and \( X \) refers to the volume of the activity.

In practice, we usually have data about the total costs incurred at each of the various activity levels, but we do not have a breakdown of costs into fixed and variable components because accounting records typically accumulate costs by account, not by behavior. What we need to do is to use the information from the accounts to estimate cost behavior.
Text messaging is a common add-on service to mobile phones, but how profitable is it for the phone companies? In September 2008, the chairman of the Senate Antitrust Committee sent letters to four major telecommunications companies asking for information about prices and costs. His interest was prompted by a price increase from $.10 to $.20 for the pay-per-use service.

Although the companies did not discuss the costs of text messaging in their responses, the variable cost can be estimated by the engineering method. First, how does a text message use the carriers’ resources?

A text message initially travels wirelessly from a handset to the closest base-station tower and is then transferred through wired links to the digital pipes of the telephone network, and then, near its destination, is converted back into a wireless signal to traverse the final leg, from tower to handset.

How does sending a text message impact the network?

In the wired portion of its journey, a file of such infinitesimal size is inconsequential. Srinivasan Keshav, a professor of computer science at the University of Waterloo in Ontario, said, “Messages are small. Even though a trillion seems like a lot to carry, it isn’t.”

What does this mean for the costs?

Professor Keshav said that once a carrier invests in the centralized storage equipment—the cost of storing a terabyte now is only $100 and dropping—and the staff to maintain it, its costs are basically covered. “Operating costs are relatively insensitive to volume,” he said. “It doesn’t cost the carrier much more to transmit a hundred million messages than a million.”

In other words, the variable costs are close to zero. What are the implications for pricing? With no incremental fixed or variable costs associated with the texting product, carriers profit from offering unlimited messaging at an affordable rate.

Once one understands that a text message travels wirelessly as a stowaway within a control channel, one sees the carriers’ pricing plans in an entirely new light. The most profitable plan for the carriers will be the one that collects the most revenue from the customer: unlimited messaging, for which AT&T and Sprint charge $20 a month and T-Mobile, $15.


What Methods Are Used to Estimate Cost Behavior?

We will study three general methods to estimate the relation between cost behavior and activity levels that are commonly used in practice:

- Engineering estimates.
- Account analysis.
- Statistical methods, such as regression analysis.

Results are likely to differ from method to method. Consequently, it is a good idea to use more than one method so that results can be compared. Large differences in cost estimates suggest it is worthwhile to conduct additional analysis. If the estimates are similar, you may have more confidence in them. In practice, operating managers frequently apply their own best judgment as a final step in the estimation process. They often modify the estimate submitted by the controller’s staff because they have more knowledge of the process and, more important, they bear ultimate responsibility for all cost estimates. These methods, therefore, should be seen as ways to help management arrive at the best estimates possible. Their weaknesses as well as their strengths require attention.

**Engineering Method**

How might you begin to help Charlene estimate the cost of a new center? One approach is to start with a detailed step-by-step analysis of what needs to be done, that is, the activities the store staff would conduct to operate the center. Probably the first thing you would want to know is the size of the center. Because this is a service firm, the size can be easily
Chapter 5  Cost Estimation

represented by the time it takes employees to provide repair service. Charlene estimates that the new center will average about 480 hours monthly.

Once you determine the size of the center, you can turn to the other necessary activities. Examples might be renting the office where the repairs will take place, using lights and other utilities, providing administrative support, or using supplies such as gloves and screws. You would then estimate the times or costs for each of these activities. The times required for each step requiring labor (administrative support, for example) would be multiplied by an estimated wage rate. Other costs, such as office rent, would be estimated from local market information. The estimate you just made is an

**engineering estimate.**

In practice, labor time estimates might come from a time and motion study. Engineering estimates of the supplies required for typical repairs can be obtained from manufacturers’ manuals and the experience of computer technicians. Other costs are estimated similarly; for example, the size and cost of the building needed to house the reception and service operation can be estimated based on area rental costs and space requirements.

One advantage to the engineering approach is that it can detail each step required to perform an operation. This permits comparison with other centers in which similar operations are performed and enables the company to review its productivity and identify specific strengths and weaknesses. Another advantage to this approach is that it does not require data from prior activities in the organization. Hence, it can be used to estimate costs for totally new activities.

A company that uses engineering estimates often can identify where “slack” exists in its operations. For example, if an engineering estimate indicates that 4,000 square feet of floor area are required for an assembly process, but the company has been renting 6,000 square feet in other centers, the company might find it beneficial to rearrange the plan to make floor space available for other uses or look for smaller rental space.

A difficulty with the engineering approach is that it can be quite expensive to use because it analyzes each activity involved in the business. Another consideration is that engineering estimates are often based on optimal conditions. Therefore, when evaluating performance, bidding on a contract, planning expected costs, or estimating costs for any other purpose, it is important to recognize that the actual work conditions will be less than optimal.

**Account Analysis Method**

One approach to estimating costs that includes the realities of downtime, missed work, machine repair, and the other factors that often cause engineering estimates to be less than realistic is to look at results from existing activities. For example, accountants often use the **account analysis** approach to estimate costs. This method calls for a review of each cost account used to record the costs that are of interest, and the identification of each as fixed or variable, depending on the relation between the cost and some activity.

Identifying the relation between the activity and the cost is the key step in account analysis. For example, in estimating the production costs for a specified number of units within the range of present manufacturing capacity, direct materials and direct labor costs are generally considered variable, and building occupancy costs are generally considered fixed. The identification depends on the accountant’s judgment and experience.

Exhibit 5.1 shows a typical schedule of estimated overhead costs per month for operating an average 3C location, where the average center operates at 360 repair-hours.

Following this approach, each major class of overhead costs is itemized and then divided into its estimated variable and fixed components. 3C typically signs a rental
The total cost for the coming period is the sum of the estimated total variable and total fixed costs. For 3C, assume that accounting personnel have relied on the judgment of a number of people in the company and have estimated fixed costs at $6,216 and the total variable costs at $4,410 for 360 repair-hours, as shown in Exhibit 5.1.

Because the variable costs are directly related to the expected activity, we can state the variable overhead per repair-hour as $12.25 (\( \frac{\$4,410}{360 \text{ repair-hours}} \)) and the general cost equation as

\[
TC = F + VX
\]

Overhead costs = $6,216 per month + ($12.25 per hour \times \text{Number of repair-hours})

For 360 repair-hours

\[
\begin{align*}
\text{Overhead costs} & = \$6,216 \text{ per month} + (\$12.25 \text{ per hour} \times 360 \text{ repair-hours}) \\
& = \$6,216 + \$4,410 \\
& = \$10,626
\end{align*}
\]

Recall that the proposed center was expected to operate at an average of 480 repair-hours. To estimate overhead costs for the new center, we substitute that figure for the 360 repair-hours in the previous equation, resulting in

\[
\begin{align*}
\text{Overhead costs} & = \$6,216 \text{ per month} + (\$12.25 \text{ per hour} \times 480 \text{ repair-hours}) \\
& = \$6,216 + \$5,880 \\
& = \$12,096
\end{align*}
\]

This is simpler than reestimating all overhead cost elements listed in Exhibit 5.1 for the different activity levels that management might wish to consider. Moreover, management’s attention is drawn to the variable cost amount as the cost that changes with each change in volume.

Account analysis is a useful way to estimate costs. It uses the experience and judgment of managers and accountants who are familiar with company operations and the way costs react to changes in activity levels. Account analysis relies heavily on personal judgment, however. This may be an advantage or disadvantage, depending on the bias of the person making the estimate. Decisions based on cost estimates often have major economic consequences for the people making them. Thus, these individuals might not be entirely objective. More objective results are often used in conjunction with account analysis to obtain the advantages of multiple methods.
1. Brown's Baskets makes decorative baskets for sale at local craft shops. Mary Brown, the owner and founder, has collected the following information on costs based on two years of operations and has asked you to help her analyze the behavior of her overhead costs. Mary summarized monthly data as two-year totals:

<table>
<thead>
<tr>
<th>Cost</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect materials</td>
<td>$27,200</td>
</tr>
<tr>
<td>Indirect labor</td>
<td>$44,300</td>
</tr>
<tr>
<td>Lease</td>
<td>$56,000</td>
</tr>
<tr>
<td>Utilities (heat, light, etc.)</td>
<td>$19,200</td>
</tr>
<tr>
<td>Power to run machines</td>
<td>$18,500</td>
</tr>
<tr>
<td>Insurance</td>
<td>$16,400</td>
</tr>
<tr>
<td>Maintenance</td>
<td>$14,500</td>
</tr>
<tr>
<td>Depreciation</td>
<td>$9,000</td>
</tr>
<tr>
<td><strong>Total overhead</strong></td>
<td><strong>$205,100</strong></td>
</tr>
</tbody>
</table>

Direct labor-hours . . . . . . . . . . . . . . . . . . 12,000
Direct labor costs . . . . . . . . . . . . . . . . . . 180,000
Machine-hours . . . . . . . . . . . . . . . . . . . . . 14,400
Units produced . . . . . . . . . . . . . . . . . . . . . 20,000

After visiting the workshop and discussing operations with Mary, you determine that three costs—indirect materials, indirect labor, and the power to run the machines—are variable. All other costs are fixed.

Prepare three analyses of overhead costs that, using the account analysis method, calculate the monthly average fixed costs and the variable rate per (1) direct labor-hour, (2) machine-hour, and (3) unit of output.

The solution to this question is at the end of the chapter on page 197.

**Statistical Cost Estimation**

Engineering estimates and account analysis are valuable approaches to estimating costs, but they have important limitations. Engineering estimates often omit inefficiencies, such as downtime for unscheduled maintenance, absenteeism, and other miscellaneous random events that affect all firms. Account analysis is often based on last period’s costs alone and is subject to managers focusing on specific issues of the previous period even though these might be unusual and infrequent. One approach to dealing with both random and unusual events is to use several periods of operation or several locations as the basis for estimating cost relations. We can do this by applying statistical theory, which allows for random events to be separated from the underlying relation between costs and activities.

Our discussion of statistical methods centers on practical applications rather than underlying statistical theory. We describe the estimation of costs and cost behavior with both single and multiple cost drivers, as well as some important implementation issues.

**Relevant Range of Activity** When using statistical approaches to cost estimation, we need to ensure that the activity levels of the past are relevant for the activity levels estimated. Extrapolations beyond the upper and lower bounds of past observations are highly subjective. Suppose, for example, that the highest activity level observed at any center is 600 repair-hours per month and we wish to predict the cost of a center with 800 repair-hours per month. An estimate may be highly inaccurate simply because the past data do not reflect cost behavior with output of more than 600 repair-hours.

The level of activity for which a cost estimate may be valid is the **relevant range**. It should include only those activity levels for which the assumed cost relations used in the estimate are considered to hold. Thus, when past data are used, the relevant range for the projection is usually between the upper and lower limits of past activity levels for which data are available.

Although the use of past data for future cost estimation has limitations, it works quite well in many cases. In many estimates, past data are adequate representations of future cost relations, even if the forecasted level of activity is somewhat outside the relevant range. Moreover, reliance on past data is relatively inexpensive; it could be the only readily available, cost-effective basis for estimating costs. Past data do show the relations that held in prior periods and at least can be a meaningful starting point for estimating costs as long as their limitations are recognized.
Scattergraphs and High-Low Estimates  When you begin a statistical analysis of costs and activities, it is helpful to begin by graphing the costs against activities using a scattergraph. This visual representation of the data provides a quick indication of the fixed-variable relation of costs and activities and can indicate whether the relation seems to change at certain activity levels. To prepare the graph, we first obtain the relevant data. For example, if estimates of manufacturing overhead are to be based on machine-hours, we must first obtain data about past manufacturing overhead and related machine-hours.

Number of Observations  The number of observations to include depends on the availability of the data, the variability within the data, the relative costs and benefits of obtaining reliable data, and the length of time the current process has been in operation. A common rule of thumb is to use three years of monthly data if the physical processes have not changed significantly within that time. If the company’s operations have changed significantly, however, data that predate the change may be misleading because you will be estimating the relation for two different processes. If cost and activity levels are highly stable, a shorter time period could be adequate.

Data for the past 15 months were collected for a representative center of 3C to estimate variable and fixed overhead. These data are presented and plotted in Exhibit 5.2. Once all data points were plotted, a line was drawn to fit them as closely as possible and was extended to the vertical axis on the scattergraph.

The slope of the line represents the estimated variable costs per unit, and the intercept with the vertical axis represents an estimate of the fixed costs. The slope is referred to as the variable cost per unit because it represents the change in costs that occurs as a result of changes in activity. The intercept is referred to as the fixed cost because it represents the costs incurred at a zero activity level given the existing capacity if the relation plotted is valid from the data points back to the origin. Note that there are no observations of cost behavior around the zero activity level in this example, so the data do not indicate the costs that would be incurred if the activity level were zero. Rather, they provide an estimating equation useful within the relevant range.

Preparing an estimate on the basis of a scattergraph is subject to a high level of error, especially if the points are scattered widely. Determining the best fit is often a matter of “eyeball judgment.” Consequently, scattergraphs are usually not used as the sole basis for cost estimates but to illustrate the relations between costs and activity and to point out any past data items that might be significantly out of line.

High-Low Cost Estimation  A simple approach to estimating the relation between cost and activity is to choose two points on the scattergraph and use these two points to determine the line representing the cost-activity relation. Typically, the highest and the lowest activity points are chosen, hence the name high-low cost estimation. Activity can be defined in terms of units of production, hours of work, or any other measure that makes sense for the problem at hand.

The slope of the total cost line, which estimates the increase in variable costs associated with an increase of one unit of activity, can be estimated using the following equation:

\[
\text{Variable cost per unit } (V) = \frac{\text{Cost at highest activity level} - \text{Cost at lowest activity level}}{\text{Highest activity level} - \text{Lowest activity level}}
\]

The intercept is estimated by taking the total cost at either activity level and subtracting the estimated variable cost:

Fixed cost = Total cost at highest activity level – (Variable cost × Highest activity level)

or

Fixed cost = Total cost at lowest activity level – (Variable cost × Lowest activity level)

Based on the data for 3C in Exhibit 5.2, the highest activity level is 568 repair-hours (RH). At this activity level, total overhead costs are $12,883. The lowest activity level is
200 hours, with overhead costs of $9,054. Substituting these data in the equation for variable costs yields the following:

\[
\text{Variable cost per RH (} V \text{)} = \frac{12,883 - 9,054}{568} RH - 200 RH
\]

\[
= \frac{3,829}{368 RH}
\]

\[
= $10.40 \text{ per RH}
\]

To obtain the fixed cost estimate, either the highest or lowest activity level and costs can be used. Assuming that the highest activity level is used,

\[
\text{Fixed cost} = 12,883 - 10.40 \times 568 RH
\]

\[
= 12,883 - 5,907
\]

\[
= $6,976
\]

An estimate for the costs at any given activity level can be computed using this equation:

\[
TC = F + VX
\]

Total costs = $6,976 + ($10.40 \times \text{Specified RH})
Part II  Cost Analysis and Estimation

For the 480 repair-hours, the estimate of overhead cost is

\[
\text{Total costs} = \$6,976 + (\$10.40 \times 480 \text{ RH}) \\
= \$6,976 + 4,992 \\
= \$11,968
\]

Although the high-low method is easy to apply, use it carefully to ensure that the two points chosen to prepare the estimates represent cost and activity relations over the range of activity for which the prediction is made. This is one reason to prepare the scattergraph. The highest and lowest points could represent unusual circumstances. When this happens, you should choose the highest and lowest points that appear representative.

**Statistical Cost Estimation Using Regression Analysis**

The scattergraph can be used graphically to illustrate cost-activity relations based on past experience and provides a useful visual display of the cost-volume relation. However, because it offers only a rough approximation of the relation, we recommend using the scattergraph in conjunction with other cost estimation methods, especially those that rely on statistical approaches. Although the high-low method allows computation of estimates of the fixed and variable costs, it ignores most of the information available to the analyst.

With computational tools included in many calculators or in spreadsheets such as Microsoft Excel®, the additional cost of using all the data instead of two points is quite small. Regression techniques are designed to generate a line that best fits a set of data points. Because the regression procedure uses all the data points, the resulting estimates have a broader base than those based on a few select points (such as the highest and lowest activity levels). In addition, regression techniques generate information that helps a manager determine how well the estimated regression equation describes the relations between costs and activities. Regression analysis also permits the inclusion of more than one predictor, a feature that can be useful when more than one factor affects costs. For example, variable overhead can be a function of both direct labor-hours and the amount of direct material processed. We leave the description of the computational details and theory to computer and statistics courses; we will focus on the use and interpretation of regression estimates. We describe the steps required to obtain regression estimates using Microsoft Excel in Appendix A to this chapter.

**Obtaining Regression Estimates**

The most important step in obtaining regression estimates for cost estimation is to establish the existence of a logical relation between activities and the cost to be estimated. These activities are referred to as predictors, **X terms, independent variables**, or the **right-hand side (RHS)** of a regression equation. The cost to be estimated can be called the **dependent variable**, the **Y term**, or the **left-hand side (LHS)** of the regression equation.

Although regression programs accept any data for the **Y** and **X** terms, entering numbers that have no logical relation can result in misleading estimates. The accountant or cost analyst has the important responsibility of ensuring that the activities are logically related to costs.

Assume, for example, that a logical relation exists between repair-hours and overhead costs for 3C. A cost analyst starts by estimating the parameters (repair-hours) to use in a simple regression (one with a single predictor) to estimate overhead costs. The analyst enters data on repair-hours as the **X**, or independent variable. Data on overhead costs are entered as **Y**, or the dependent variable. The computer output giving the estimated relation between repair-hours and overhead for this situation is shown in Exhibit 5.3. (The scattergraph for this regression is shown in Exhibit 5.2, along with the data.)

The computer output is interpreted as follows:

\[
\text{Total overhead} = \$6,472 + (\$12.52 \text{ per RH} \times \text{Number of RH})
\]

For cost estimation purposes, when you read the output of a regression program, understand that the intercept term, $6,472, is an estimate of fixed costs. Of course, it should be used with caution because the intercept at zero activity is outside the relevant range.
of observations. The coefficient of the X term (in this example, $12.52 per repair-hour) is an estimate of the variable cost per repair-hour. This is the slope of the cost line. The coefficients are often labeled $b$ or given the variable name (repair-hours) on the program output. Thus, the cost estimation equation based on this regression result is

$$\text{Total costs} = \text{Intercept} + (b \times RH)$$

Substituting 480 $RH$ into the equation yields

$$\begin{align*}
\text{Total costs} &= 6,472 + (12.52 \times 480) RH \\
&= 6,472 + 6,010 \\
&= 12,482
\end{align*}$$

This estimate of cost behavior is shown graphically in Exhibit 5.4.

---

**Exhibit 5.3**  Regression Results for the Overhead Cost Estimation—3C

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SUMMERY OUTPUT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Regression Statistics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Multiple R</td>
<td>0.90981797</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>R Square</td>
<td>0.82776855</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Adjusted R Square</td>
<td>0.81451998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Standard Error</td>
<td>678.27424</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Observations</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>ANOVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>of SS</td>
<td>MS</td>
<td>F</td>
<td>Significance F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Regression</td>
<td>1</td>
<td>28744216.46</td>
<td>287442156</td>
<td>62.47983</td>
<td>2.55004E.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Residual</td>
<td>13</td>
<td>59.807227.272</td>
<td>460055.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Total</td>
<td>14</td>
<td>34724943.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Coefficients</td>
<td>Standard Error</td>
<td>t Stat</td>
<td>P-value</td>
<td>Lower 95%</td>
<td>Upper 95%</td>
<td>Lower 95.0%</td>
<td>Upper 95.0%</td>
</tr>
<tr>
<td>17</td>
<td>Intercept</td>
<td>6472.12784</td>
<td>606.3287798</td>
<td>10.67429</td>
<td>5162.234157</td>
<td>7782.021532</td>
<td>5162.234157</td>
<td>7782.021532</td>
</tr>
<tr>
<td>18</td>
<td>X Variable 1</td>
<td>12.5229407</td>
<td>1.584296313</td>
<td>7.904418</td>
<td>2.55E-06</td>
<td>9.100276656</td>
<td>15.94560483</td>
<td>9.100276656</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Correlation Coefficients In addition to the cost-estimating equation, the regression program provides other useful statistics. The \textit{correlation coefficient} ($R$, referred to as multiple $R$ in Exhibit 5.3) measures the proximity of the data points to the regression line. The closer $R$ is to 1.0, the closer the data points are to the regression line. Conversely, the closer $R$ is to zero, the poorer the fit of the regression line.

The square of $R$ is called $R$-\textit{squared} ($R^2$) or the \textit{coefficient of determination}. $R^2$ is interpreted as the proportion of the variation in $Y$ explained by the right-hand side of the regression equation, that is, by the $X$ predictors.

For 3C, the correlation coefficient and $R^2$ are the following (see the regression results in Exhibit 5.3):

\[
\begin{align*}
\text{Correlation coefficient (} R \text{)} & \quad .910 \\
R^2 & \quad .828
\end{align*}
\]

Because the $R^2$ is .828, it can be said that 82.8 percent of the changes in overhead costs can be explained by changes in repair-hours. For data drawn from accounting records, an $R^2$ of .828 is considered a good fit of the regression equation to the data.

The most commonly used regression technique is called \textit{ordinary least squares regression} (OLS). With this technique, the regression line is computed so that the sum of the squares of the vertical distances from each point to the regression line is minimized. Thus, as a consideration, it is important to beware of including data points that vary significantly from the usual. Because the regression program seeks to minimize squared differences, the inclusion of these extreme points, or “outliers,” can significantly affect the results. Consequently, organizations often exclude data for periods of unusual occurrences such as strikes, extreme weather conditions, and shutdowns for equipment retooling. Plotting data on a scattergraph often reveals such outliers so they can be easily identified and omitted. We discuss the effects of outliers later in this section.

Confidence in the Coefficients In many cases, it can be desirable to determine whether the estimated coefficient on the independent variable is significantly different from zero. For example, when determining fixed and variable costs, if the estimated coefficient is significantly different from zero, we can conclude that the cost is not totally fixed. The \textit{t-statistic} is used to test the significance of the coefficient.

The \textit{t-statistic} is computed as the value of the estimated coefficient, $b$, divided by its estimated standard error ($SE_b$). For the data used in the 3C regression, which is shown in Exhibit 5.3, the \textit{t-statistic} is

\[
t = \frac{b}{SE_b} = \frac{12.5230}{1.5843} = 7.9047
\]

As a general rule of thumb, a \textit{t-statistic} greater than 2.0 is considered significant. The significance level of the \textit{t-statistic} is called the \textit{p-value} and is shown in Exhibit 5.3. For the 3C data, the \textit{p-value} for the estimated coefficient on repair-hours is 0.000000255 (2.55E-06). This means that the probability that the true value of the coefficient is zero, given the data, is virtually zero.

To construct a 95 percent confidence interval around $b$, we add or subtract to $b$ the appropriate \textit{t-value} for the 95 percent confidence interval times the standard error of $b$, as follows:

\[
b \pm t \times SE_b
\]

For the 3C data, $SE_b = 1.5843$. We obtain the value of $t$ for a 95 percent confidence interval from probability tables. This value is $t = 2.160$. Therefore, a 95 percent confidence interval for the coefficient $b$ in the 3C regression is,

\[
b \pm 2.160 \times 1.5843 = b \pm 3.4221
\]

With $b$ equal to $12.52$, we would be 95 percent confident that the variable cost coefficient is between $9.10 (= 12.52 - 3.42)$ and $15.94$. 
Multiple Regression

Although the prediction of overhead costs in the previous example, with its $R^2$ of .828, was considered good, management might wish to see whether a better estimate can be obtained using additional predictor variables. In such a case, they examine the nature of the operation to determine which additional predictors might be useful in deriving a cost estimation equation.

Assume that 3C has determined that parts cost as well as repair-hours can affect overhead. The results of using both repair-hours ($X_1$) and parts cost ($X_2$) as predictors of overhead, $Y$, were obtained using a spreadsheet-based regression analysis. The output from the analysis using repair-hours and parts cost yields the prediction equation:

$$\text{Cost} = \text{Intercept} + b_1 RH + b_2 \text{Parts cost}$$

The statistics supplied with the output (rounded off) are:

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation coefficient ($R$)</td>
<td>.953</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.908</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.892</td>
</tr>
</tbody>
</table>

The adjusted $R^2$ is the correlation coefficient squared and adjusted for the number of independent variables used to make the estimate. This adjustment to $R^2$ recognizes that as the number of independent variables increases, $R^2$ (unadjusted) increases. Statisticians believe that adjusted $R^2$ is a better measure of the association between $X$ and $Y$ than the unadjusted $R^2$ value when more than one $X$ predictor is used.

The correlation coefficient for this equation is .953, and the adjusted $R^2$ is .892. This is an improvement over the results obtained when the regression equation included only repair-hours. Improved results can be expected because some overhead costs may be related to parts cost (for example, administrative support) but not to repair-hours.

Preparing a cost estimate using this multiple regression equation requires not only the estimated repair-hours for the new center but also the estimated parts cost. The additional data requirements for multiple regression models can limit their usefulness in many applications. Of course, in planning for the new center’s activity, 3C probably has already estimated parts cost and repair-hours, and in such a situation the added costs of obtaining data could be quite low.

Charlene estimates that in addition to 480 repair-hours at the new center, parts costs will be $3,500. Using the estimated equation, she estimates the total overhead costs as:

$$\text{Total costs} = 6,416 + (8.61 \times 480) + (0.77 \times 3,500)$$

$$= 6,416 + 4,133 + 2,695$$

$$= 13,244$$
Although our focus in this chapter is on cost estimation, we could also use the regression results to test whether a particular factor is related to cost. In other words, we could test whether the factor is a cost driver. For example, in the analysis above, we could examine the \( t \)-statistics for each of the coefficients to determine if they are both significant. (In Appendix A, where we discuss the use of Excel for estimating the regression, we see that they are both significant.) If the analysis showed, for example, that parts cost was not significant, we would conclude that repair-hours is the better cost driver.

**Practical Implementation Problems**

Advances in easy-to-use computer software, especially spreadsheet software, have greatly simplified regression analysis and made it available to more people. Consequently, regression methods have been increasingly used (and misused). In particular, analysts can be tempted to enter many variables into a regression model without careful thought of their validity. The results can be misleading and potentially disastrous.

Some of the more common problems with using regression estimates include (1) attempting to fit a linear equation to nonlinear data, (2) failing to exclude outliers, (3) including predictors with apparent, but spurious, relations to the dependent variable, and (4) using data that do not fit the assumptions of regression analysis.

**Effect of Nonlinear Relations**

The effect of attempting to fit a linear model to nonlinear data is likely to occur when the firm is operating near its capacity limits. Close to maximum capacity, costs increase more rapidly than activity because of overtime premiums paid to employees, increased maintenance and repair costs for equipment, and similar factors. The linear cost estimate understates the slope of the cost line in the ranges close to capacity. This situation is shown in Exhibit 5.5.

One way to overcome the problem is to define a relevant range of activity, for example, from 25 percent to 75 percent capacity, and use the range for one set of cost-estimating regression equations. A different equation could be derived for the levels between 81 and 100 percent capacity. Another approach is to model the nonlinearity explicitly by including the squared value of an independent variable as well as the variable itself. However, this approach does not provide a constant unit variable cost estimate; the estimate is different at each level of activity.
Effect of Outliers Because regression minimizes the sum of the squared deviations from the regression line, observations that lie a significant distance away from the line could have an overwhelming effect on the regression estimates. Exhibit 5.6 shows a case in which most of the data points lie close to a straight line, but because of the effect of one significant outlier, the computed regression line is a substantial distance from most of the points.

This type of problem can easily arise in accounting settings. Suppose that a year’s worth of supplies was purchased and expensed entirely (but not used) within a single month or a large adjustment was made for underaccruing payroll taxes. The accounting records in such cases are clearly abnormal with respect to the activity measure.

An inspection of the scattergraph can often reveal this problem. When an extreme outlier appears in the data set, scrutiny of the output from the regression analysis will rarely identify it. Instead, a plot of the regression line on the data points is usually needed. If multiple predictors are used, an outlier will be even more difficult to find. The best way to avoid this problem is to examine the data in advance and eliminate highly unusual observations before running the regression.

Effect of Spurious Relations It is sometimes tempting to include many variables in the regression and let the program “find” relations among the variables. This can lead, however, to spurious relations. For example, a relation between variable 1 and variable 2 could appear to exist, when, in fact, variable 3, which was left out of the analysis, explains the situation. An obvious example is estimation of a regression to explain direct materials cost by using, say, direct labor costs as the independent variable. The association will typically be quite high, but both are driven by output.

Effect of Using Data That Do Not Fit the Assumptions of Regression Analysis Regression analysis is a powerful tool for analyzing and estimating costs, but it relies on several important assumptions. If the assumptions are not satisfied, the results of the regression will not be reliable. Two important assumptions that are often not satisfied in estimating costs are that (1) the process for which costs are being estimated remains constant over time and (2) the errors in estimating the costs are independent of the cost drivers.
Businesses today change processes frequently as part of continuous improvement efforts. Regression analysis assumes, however, that the process remains the same. This situation leaves the cost analyst with two choices. The analyst can restrict the data to a short period and thereby assume the process has remained the same. However, the estimates will not be as reliable because there are relatively few observations. Alternatively, the analyst can use a longer period. As long as the process has not changed, the estimates will be more reliable (since they are based on more information), but the analyst then risks using estimates that might not be meaningful if the process has changed.

These trade-offs indicate that using regression analysis for estimating costs requires care in the selection and use of the data. It is not enough to rely on a spreadsheet program to generate the results; the analyst must be assured that the data being used are appropriate for regression analysis.

**Regression Must Be Used with Caution**

A regression estimate is only an estimate. Computerized statistical techniques sometimes have an aura of truth about them. In fact, a regression estimate can be little better than an informal estimate based on plotted data. Regression has advantages, however. It is objective, provides a number of statistics not available from other methods, and could be the only feasible method when more than one predictor is used.

We recommend that users of regression (1) fully understand the method and its limitations; (2) specify the model, that is, the hypothesized relation between costs and cost predictors; (3) know the characteristics of the data being used; and (4) examine a plot of the data.

**Learning Phenomenon**

You might recall the first time that you used a spreadsheet program on a computer. While you might have been slow at first, your speed improved as you gained more experience. In the same way, companies find that experience—or learning—affects labor costs. Specifically, the more experience that workers have performing a task, the less time they spend on it. As we discussed in the previous section, cost estimation methods assume that

**In Action**

**Learning Curves**

From Chapter 4, we know that cost estimates are important in pricing decisions. This is especially true when the product is new and no market price exists. The problem is that companies typically become more efficient as employees learn how to work with the materials and processes required for the product.

Statistical cost estimation can help by estimating the rate at which costs of new products will decline as a function of cumulative output. Plotting unit cost against cumulative production yields a graphical example of the learning curve. Assuming constant unit costs, the steeper the slope, the faster and greater the learning. If unit costs do not vary with cumulative output, then there is no learning effect.

Having an accurate estimate of the rate of learning is especially important for companies that produce large, technologically sophisticated products. For example, new airplane at $100 million or even $50 million. But can it make a profit by pricing them at $25 million each? When will the company reach a break-even level of production, how much will it have lost up to that point, and how much profit will it make on planes built after that? Learning curves can help answer those kinds of questions.

"Today, Boeing uses learning curves for capacity analysis, resource requirements planning, cost-reduction proposals and estimations of production-line performance," says Dwight Miller, director of industrial engineering for Boeing's commercial airplanes group. "We benefit daily from this concept."

Indeed, the equations underlying learning curves can be an essential part of cost estimating, pricing and staff planning. "The potential applications of learning curves far outstrip their current usage," says Charles Bailey, an accounting professor at the University of Central Florida in Orlando.

the process for which costs are being estimated has not changed. If, because of learning, for example, the process has changed, we need to incorporate that change in our estimation methods.

The learning phenomenon refers to the systematic relationship between the amount of experience in performing a task and the time required to perform it. This can occur when companies introduce new production methods, make new products (either goods or services), or hire new employees. For example, the effect of learning on the cost of aircraft manufacturing is well known. Manufacturers of products for the aerospace industry, such as General Electric and Boeing (see the In Action discussion on learning curves), recognize the effect of learning on the production cost of a new product by writing contracts that establish a lower cost for consecutive units produced. For example, the second unit produced has a lower production cost than the first unit, the third unit produced has a lower production cost than the second unit, and so on.

The following example and Exhibit 5.7 show the effect of learning on costs. Assume that the company’s engineers have found a systematic relation between the time required to produce units and the volume of units produced. These engineers estimate that the time required to produce the second unit is 80 percent of the time required to produce the first unit. Further, the time to produce the fourth unit is 80 percent of the time to produce the second unit, and so forth. (What is the time to produce the eighth unit? Answer: 80 percent of the time to produce the fourth unit.)

This is called an 80 percent learning curve. If the time to produce the fourth unit was 70 percent of the time to produce the second unit, then the relationship would be called a 70 percent learning curve. If the time to produce the fourth unit was 90 percent of the time to produce the second unit, then the relationship would be called a 90 percent learning curve. You get the idea.

Now assume that the first unit takes workers 100 hours to produce. Then, given an 80 percent learning curve, the second unit will require 80 hours to produce \((= 80\text{ percent} \times 100\text{ hours})\). The fourth unit will require 64 hours \((= 80\text{ percent} \times 80\text{ hours})\), and so forth, as shown in the table that follows. Appendix B presents the mathematical formula for deriving the learning curve and extends this example.

<table>
<thead>
<tr>
<th>Unit</th>
<th>Time to Produce</th>
</tr>
</thead>
<tbody>
<tr>
<td>First unit</td>
<td>100 hours (assumed)</td>
</tr>
<tr>
<td>Second unit</td>
<td>80 hours ((= 80\text{ percent} \times 100\text{ hours}))</td>
</tr>
<tr>
<td>Fourth unit</td>
<td>64 hours ((= 80\text{ percent} \times 80\text{ hours}))</td>
</tr>
<tr>
<td>Eighth unit</td>
<td>51.2 hours ((= 80\text{ percent} \times 64\text{ hours}))</td>
</tr>
</tbody>
</table>

Exhibit 5.7 shows the relation between volume and the number of labor hours required to produce the last unit in Panel A. The relation between volume and total labor costs appears in Panel B. Assume the labor cost is $50 per hour. Note that the labor cost per unit for the first unit is $5,000, but that the cost drops to $2,560 per unit for the eighth unit, a substantial decrease due to the learning phenomenon \((= 2,560 = 51.2\text{ hours} \times 50)\).

Applications

The learning phenomenon means that variable costs tend to decrease per unit as the volume of activity increases. Thus a linear cost estimate, such as the one shown in Exhibit 5.4, will overstate the variable cost per unit. The learning phenomenon affects

---

1 The approach that we demonstrate is the incremental unit-time learning model. Another approach, which is harder to understand but has the same principles, is the cumulative average-time learning model. For a discussion of this latter model, see Hilton, Maher, and Selto, Cost Management, 4th edition (Burr Ridge, IL: McGraw-Hill/Irwin), Chapter 11.
most professional activities such as consulting, legal, medical, and engineering work, as well as any overhead costs, such as supervision, that are related to labor time.

When estimating costs, decision makers should consider the potential impact of learning. The learning phenomenon can affect costs used in cost management, decision making, and performance evaluation. Failing to recognize learning effects can have some unexpected consequences, as shown in the following examples.

**Decision Making**  Assume that Generic Electric Company is considering producing a new navigational device for NASA. NASA has indicated it will pay $500,000
per unit for the device. Generic Electric engineers and cost management analysts estimate the cost to Generic Electric to produce the first four units of the device to be $600,000 per unit. At first, Generic Electric decides not to produce the device because the unit cost exceeds the unit price. However, NASA assures Generic Electric that it will order 40 units of the device. After considering the learning phenomenon for the device, Generic Electric realizes that the average cost per unit will drop to $400,000 for 40 units. For four units, producing the device is unprofitable. For 40 units, however, it is profitable because the learning phenomenon reduces the time and costs for units 5 through 40 sufficiently to bring the average cost down to $400,000 per unit.

**Performance Evaluation**  Elite State University (not its real name) developed labor time and cost expectations for clerical activities that were subject to the learning phenomenon. For example, employees were expected to answer an inquiry about the status of an application to the university’s law school in one minute. Management observed that time spent on these activities systematically exceeded expectations. Upon investigating the problem, management found high personnel turnover, which meant that the activities were often being performed by inexperienced people. As a result, the university never experienced the expected benefits of learning. After changing personnel practices to reduce turnover, the university had more experienced people in jobs. These experienced people performed the activities faster than less experienced people, and the time spent on activities now met expectations.

**How Is an Estimation Method Chosen?**

Each of the methods discussed has advantages and disadvantages. Probably the most informative estimate of cost behavior results from using several methods discussed because each has the potential to provide information that the others do not.

We have discussed a variety of cost estimation methods ranging from the simple account analysis method to sophisticated techniques involving regression analysis. Which of these methods is best? In general, the more sophisticated methods yield more accurate cost estimates than the simpler methods do. However, even a sophisticated method yields only an imperfect estimate of an unknown cost behavior pattern.

All cost estimation methods make assumptions to simplify the analysis. The two most common assumptions follow:

- *Cost behavior depends on just one cost driver.* (Multiple regression is an exception.) In reality, however, costs can be affected by a host of factors, including the weather and the mood of the employees.

- *Cost behavior patterns are linear within the relevant range.* We know that costs actually follow curvilinear, step, semivariable, and other patterns.

You must consider on a case-by-case basis whether these assumptions are reasonable. You must also decide when it is important to use a more sophisticated, and more costly, cost estimation method and when it is acceptable to use a simpler approach. As with all management accounting methods, you must evaluate the costs and benefits of various cost estimation techniques.

**Data Problems**

If a company’s operations have followed a particular pattern in the past and that pattern is expected to continue in the future, using the relation between past costs and activity to estimate future costs can be useful. Of course, if the relation changes, it could be necessary to adjust the estimated costs accordingly or explicitly consider the changes when developing the estimates.
Analysts must be careful when predicting future costs from historical data. In many cases, the cost-activity relation changes. Technological innovation, increased use of automation, more mechanized processes, and similar changes have made the past cost-activity relations inappropriate for prediction purposes in many organizations. For example, switching to a just-in-time inventory system will alter the relation between materials-handling costs and volume because the intermediate storage step is eliminated.

In other cases, the costs change so dramatically that old cost data are worthless predictors of future costs. Because of the high variation in costs, companies using precious metals or relying on labor in developing countries have found that past cost data are not very helpful in predicting future costs. Although accountants can adjust the data, the resulting cost estimates tend to lose their objectivity as the number of adjustments increases.

No matter what method is used to estimate costs, the results are only as good as the data used. Collecting appropriate data is complicated by the following problems:

- **Missing data.** Misplaced source documents or failure to record a transaction can result in missing data.
- **Outliers.** Extreme observations of cost-activity relations can unduly affect cost estimates. For example, a tornado recently affected operations in Oklahoma businesses, resulting in unusually low volume.
- **Allocated and discretionary costs.** Fixed costs are often allocated on a volume basis, resulting in costs that could appear variable. Discretionary costs also can be budgeted so that they appear variable (e.g., advertising expense budgeted as a percentage of revenue).
- **Inflation.** During periods of inflation, historical cost data do not accurately reflect future cost estimates. Even if inflation remains low in one country, firms with international operations must consider the effects of subsidiary operations when making cost estimates.
- **Mismatched time periods.** The time period for the dependent and independent variables may not match (e.g., running a machine in February and receiving [recording] the energy bill in March).

### Effect of Different Methods on Cost Estimates

Each cost estimation method can yield a different estimate of the costs that are likely to result from a particular management decision. This underscores the advantages of using two or more methods to arrive at a final estimate. The different manufacturing overhead estimates that resulted from the use of four different estimation methods for 3C are summarized in Exhibit 5.8.

<table>
<thead>
<tr>
<th>Method</th>
<th>Total Estimated Costs</th>
<th>Estimated Fixed Costs</th>
<th>Estimated Variable Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account analysis (see page 157)</td>
<td>$12,096</td>
<td>$6,216</td>
<td>$12.25 per repair-hour</td>
</tr>
<tr>
<td>High-low (see page 160)</td>
<td>$11,968</td>
<td>$6,976</td>
<td>$10.40 per repair-hour</td>
</tr>
<tr>
<td>Simple regression ((RH)^a) (see page 162)</td>
<td>$12,482</td>
<td>$6,472</td>
<td>$12.52 per repair-hour</td>
</tr>
<tr>
<td>Multiple regression ((RH\ and parts cost)^b) (see page 165)</td>
<td>$13,244</td>
<td>$6,416</td>
<td>$8.61 per repair-hour + 77% of parts cost</td>
</tr>
</tbody>
</table>

\( ^{a} \) for 480 repair-hours.

\( ^{b} \) for 480 repair-hours and $3,500 in parts costs.