

Chapter IX

Computer Tools for Public–Sector Management

Carl Grafton

Auburn University Montgomery, USA

Anne Permaloff

Auburn University Montgomery, USA

ABSTRACT

Almost any public-sector task employing a computer can be accomplished more efficiently with a variety of tools rather than any single one. Basic tools include word processing, spreadsheet, statistics, and database-management programs. Beyond these, Web authoring software, presentation software, graphics, project-planning and -management software, decision analysis, and geographic information systems can be helpful depending upon the job at hand.

INTRODUCTION

The use of computer technology in government taps into three sometimes incompatible concepts: government responsiveness to the public, bureaucracy, and technocracy. The tensions between the first two have long been a staple of textbooks

and scholarly work in public administration and organization theory (Blau & Meyer, 1971; Borgmann, 1988; Gullick, 1996; Rosenbloom & Kravchuk, 2002). At first, when all computers were mainframes, the technocratic perspective (rule by experts) appeared to bolster Weberian bureaucracies (Ellul, 1964; Freeman, 1974). Even today, computers are often used by bureaucrats to perform routine tasks efficiently or analysts to rationalize policy, and most of this chapter is taken up by descriptions of some of the tools available to them. However, today's computers are employed in far more ways and by many more members of all parts of government than they were a few years ago. The bureaucracy is less centralized just by virtue of the widespread access of government personnel to information and their ability to process that information.

Changes wrought by computers may go beyond bureaucratic decentralization. Eugene J. Akers (2006) speculates that government

organized along Weberian bureaucratic lines is increasingly out of step with public expectations of a transparent and responsive service-oriented government. Similarly, Carl Grafton and Anne Permaloff (2005) depict what they call Jeffersonian budgeting: understandable government budgets available on the Internet with which the news media and the public can hold public officials accountable. In addition, Christa Slaton and Jeremy Arthur (2004) describe ways to facilitate public participation in government administration using computer technology.

This chapter concerns computer applications and information technology in government other than financial accounting software, which deserves a chapter of its own. Topics covered include Web publishing, spreadsheets, statistics packages, database management, presentation software, project-planning and -management software, decision analysis, graphics for illustrations, and geographic information systems. Since most readers are likely to have substantial word-processing experience, it would be unproductive to devote much space to this topic.

A VARIETY OF TOOLS

To make the most of their time and talents, computer users in the public sector or virtually any other setting should have access to more than one tool for nearly any task that extends much beyond typing a short memo. Access to a variety of tools is usually more productive than having the latest version of a single one.

Word-Processing and Web Authoring Software

Word-processing programs are designed primarily for generating print and graphic images on paper; Web authoring programs do the same thing for the Internet. Web pages are generated using

HTML (hypertext markup language) sometimes in conjunction with supplemental tools such as Java, a programming language.

When a browser such as Microsoft Explorer reads a file containing HTML code (called tags) and text, it displays the file on the computer monitor according to formatting information in the tags (e.g., whether text is centered or in bold face or whether a separate file containing a graphic image is to be merged with the text). The marketplace offers a variety of text editors primarily designed to generate text with HTML tags (see Kent, 2000, for a list). Ordinary word-processing software can do so as well, but specialized HTML editors contain more features that ease the process of Web-page creation or maintenance.

Most government agencies have adopted particular HTML editors that employees are expected to use. Government agencies often appoint individuals responsible for Web-page work partly to preserve consistency in appearance and also out of concerns for security. The majority of employees will submit text generated with an ordinary word processor along with graphic images to accompany the text to these specialists who will then convert these files into Web pages.

The authors' experience in using government Web sites suggests that they are of three basic types: marketing, informational, and interactive. Marketing sites are designed for such purposes as to attract students to universities and visitors to cities. Marketing sites usually contain photographs scanned from prints or downloaded from a digital camera together with attractively arranged text and various decorative graphics. A certain amount of taste is required to make the marketing part of a Web site attractive, and the assistance of someone with a background in art might be helpful. Peter Kent (2000) lists Web sites that collect badly designed Web sites including one called *Web Pages That Suck: Learn Good Design by Looking at Bad Design*. While the bad Web-site collections dramatize how Web designs

can go horribly wrong, fledgling designers will also want to visit sites of organizations similar to theirs to see what works.

Most Web pages presented by government agencies are meant to be entirely informational. Beyond basic neatness, the only significant design considerations are clarity and consistency. For example, the Web sites of most states and cities describe tax structures. These sites typically provide an overview and links that lead to more detailed information such as how taxes can be calculated and paid. These pages are usually quite informative, but sometimes a variety of fonts and formats are employed. For example, in one state, some pages contained buttons while others used simple lists, giving the whole enterprise a somewhat unprofessional appearance. Sometimes information is simply difficult to find, and built-in Web-site search engines require overly precise wording.

Interactive Web sites are the most difficult to build (Kent, 2000). Interactive sites gather information from users. They may take orders for goods or services, maintain an inventory of goods and services available, and accept credit-card information, among other tasks. An example is an online university class-registration system. Interactive sites require specialized programming skills beyond the reach of most public administrators. It is critical that potential users test all Web sites, especially interactive ones. Information and procedures that seem clear to programmers may be sources of confusion and irritation to those unfamiliar with a Web site's logic.

Richard Heeks (2006) observes that failed or flawed Web sites and other IT applications may be found at all levels of government. The central cause of failure is that the planning, construction, implementation, and maintenance of IT systems are important to many actors including agencies with differing responsibilities and agendas, elected members of the executive and legislative branches, and competing vendors, but participation is often limited to engineers, programmers,

and high-level management (Coursey & Killingworth, 2005). Most government employees lack IT backgrounds and tend to surrender Web-site and IT-system development and maintenance to technical personnel. If a system is to serve the needs of all stakeholders including government agency clients, participation in system design must be widespread.

Spreadsheets

No desktop computer tool is more widely used for numerical data storage and analysis than the spreadsheet family of programs. Desktop computers are ubiquitous in homes and workplaces, and spreadsheet programs such as Excel reside on many hard disks as part of a suite of programs; however, a large percentage of public managers know little about spreadsheet operations beyond basic budgeting uses.

Programs such as Excel present the user with a series of worksheets grouped as a book of related information. Each worksheet is a grid of columns and rows. The rectangular intersection of a column and row is a cell. Three major kinds of information may be entered into a cell: text, numbers, and formulas that perform calculations on the numbers or on other formulas. Excel, by far the most widely used spreadsheet program, is also able to utilize Visual Basic, the program language common to the products that make up the Microsoft Office Suite. In addition to performing calculations based on user-inputted formulas, spreadsheet programs can produce graphics such as pie charts, bar charts, and time-series graphs, and they can be used to query a database stored in the spreadsheet. They can also perform statistical analysis with a variety of built-in tools and add-ins (defined below). In addition, the spreadsheet program's basic worksheet is one of the formats used for inputting data into a variety of other programs. It is also used as a format for downloading information (e.g., budget information and election results) from public Web sites.

Benefit-Cost Analysis

Benefit-cost analysis compares the benefits and costs of various choices and selects the choice that yields the highest net benefit (Ammons, 2002; Boardman, Greenberg, Vining, & Weimer, 2001; Stokey & Zeckhauser, 1978). In practice, most benefit-cost problems involve capital expenditures where benefits and costs are received and spent over a period of years.

The first step in performing a benefit-cost analysis is to list sources of benefits and costs associated with a project (Panayotou, n.d.). In this initial phase, the focus is on categories, not dollar amounts. For example, a benefit-cost analysis of a proposed monorail in Seattle listed the following benefits to monorail riders: travel-time savings, parking savings, automobile cost savings, and reliability. Those who continued to drive would enjoy the benefit of lessened traffic congestion (DJM Consulting, 2002; Slack, n.d.).

The second step is to evaluate the relationship between costs and benefits. The analyst is studying the causal relationship (or lack thereof) between expenditures and the benefits they yield (Panayotou, n.d.; Schmid, 1989). For example, the Seattle monorail benefit-cost analysis included an estimate of ridership partly based on comparisons with rail systems in other cities (People for Modern Transit [PMT], n.d.). The three largest benefits (travel-time savings, parking savings, automobile cost savings), which constituted more than 86% of claimed benefits, were heavily dependent on a ridership estimate that could only have been an approximation, a point ignored in a PowerPoint briefing presented by a consulting firm employed by advocates of the monorail. A benefit-cost analysis that fails to provide a range of estimates of critical parameters should be regarded with deep suspicion.

The next step is to establish direct effects of a project: the dollar benefits to project users and costs to build the project. When the market is functioning, project costs are simply off-the-shelf

costs (e.g., concrete for highway construction; Levin, 1983; Nas, 1996). If market failure exists (e.g., leading-edge technology provided by a small number of competitors), cost valuation becomes more difficult (Panayotou, n.d.).

In some cases, project benefits are comparable to products or services sold in the private sector. An example is public housing. The value of a public-housing apartment will be close to that of a similar apartment offered by the private sector. If the government decides to offer units at a discount, the benefit to the renter will be the difference between the private-sector value of the apartment and the rent actually being charged. However, project benefits are often not directly comparable to private-sector market values. Thus, for example, a monorail may offer a faster and safer route. The problem is how to value time saved and deaths averted. In the case of the Seattle monorail, travel time was valued at \$10.10 per hour, an amount that was half the mean wage rate in that area in 2002 (Slack, n.d.). The value of a life saved is more problematic, and the benefit-cost analysis literature offers many ways of doing so, most of which produce widely varying results (Fuguitt & Wilcox, 1999; Nas, 1996). Advocates of the Seattle monorail claimed \$6.3 million in savings from reduced motor-vehicle accidents, but apparently these figures were confined to property damage and not loss of life.

The indirect effects of a project impact individuals or groups who are not users of project output (Schmid, 1989). With the monorail, those who can avoid driving automobiles through Seattle's crowded streets and parking them at high cost are direct beneficiaries. Those whose property values are increased or decreased or who experience other economic impacts because of the monorail are experiencing indirect benefits or costs (Marsden Jacob Associates, 2005; Tindall, 2005). Note that a single individual can enjoy both direct and indirect benefits or pay both direct and indirect costs. The distinction between direct and indirect effects is useful for organizing a benefit-

cost study, but whether a given effect is placed in one or the other category should not have an impact on the results.

The two most common numerical indicators used in benefit-cost analysis are net benefits and the benefit-cost ratio. Net benefits are the difference between benefits and costs calculated on a present-value basis. The benefit-cost ratio is benefits divided by costs also on a present-value basis. The present value is the present-day equivalent of a stream of costs or benefits over time. The Seattle monorail net-benefit figure claimed by DJM Consulting in its 2002 briefing was \$67 million. However, the benefit-cost ratio was only 1.04. In other words, the project would yield only \$1.04 for every \$1.00 spent. The 1.04 ratio represents a very thin margin. With slight cost overruns and tiny mistakes in ridership estimates, a 1.04 ratio could easily have flipped to less than the 1.00 breakeven point, and net benefits would have been negative tens of millions of dollars instead of positive. Again, the failure of this briefing to present ranges of estimates should have raised red flags. In fact, Seattle voters ultimately defeated the project.

The basic perspective of benefit-cost analysis is that of society as a whole (Fuguitt & Wilcox, 1999). So, for example, in a project's planning and construction phase, planners and engineers are benefiting in the form of fees and salaries, but from society's perspective, their benefits are costs. Society does not begin to benefit from a project until it is complete.

Large projects often affect geographical, racial, and other groups differently (Boardman et al., 2001; Nas, 1996; Schmid, 1989). Indeed, benefit-cost analysis' common society-wide perspective not only ignores such differential effects, but it may be biased against the disadvantaged. For example, other things being equal, benefit-cost analyses of alternative routes for a highway will favor a route plowing through economically depressed neighborhoods because property acquisition costs will be relatively low. For that

reason, the federal government discourages the use of benefit-cost analysis by states in making routing decisions (Clinton, 1994; Federal Highway Administration, 1998).

The formula for calculating present value for any benefit or cost stream is:

$$PV = S_0/(1+r)^0 + S_1/(1+r)^1 + S_2/(1+r)^2 + \dots + S_n/(1+r)^n,$$

where S_n represents any sum of money (benefit or cost) in year n , and r is the discount rate. The value of n may be the physical life of a project or its technologically useful life.

Performed on a hand calculator, the benefit-cost analysis of one or two streams of figures using the above formula is of no difficulty, but an actual benefit-cost analysis performed to aid decision making (as opposed to providing decorative support for decisions already made) typically involves sensitivity or "what if" analysis applied to combinations of benefits, costs, and discount rates. The basic question being posed by any kind of sensitivity analysis is how the result will be affected by change in inputs—in this case, benefits, costs, and the discount rate. If the estimates for each of these are all performed using a hand calculator, one is facing a considerable amount of work. With a spreadsheet such a prospect represents no problem.

Figure 1 shows a typical benefit-cost situation. In the first 2 years, benefits are zero because the project in question (e.g., a building) is under design and construction. The design and construction costs are relatively high compared to annual benefits. Because high design and construction costs are incurred in the first few years, they are discounted relatively little by the present value formula. The smaller benefit figures are more and more heavily discounted as they extend further into the future. Thus, the higher the discount rate, which governs the severity of the discounting, the worse the project will look as calculated either by net benefits (benefit minus cost) or the benefit-cost

Figure 1. Benefit-cost analysis using a spreadsheet

	A	B	C	D	E	F	G	H	I	J	
1					Benefit-Cost Estimates						
2					(millions of dollars)						
3											
4											
5					Year						
6				0	1	2	3	4	5	6	
7	Pessimistic benefits		0	0	3	3	3	3	3	3	
8	Most likely benefits		0	0	3.5	3.5	3.5	3.5	3.5	3.5	
9	Optimistic benefits		0	0	3.8	3.8	3.8	3.8	3.8	3.8	
10	Pessimistic costs		1.5	14	0.2	0.2	0.2	0.2	0.2	0.2	
11	Most likely costs		1	12	0.1	0.1	0.1	0.1	0.1	0.1	
12	Optimistic costs		0.9	10	0.08	0.08	0.08	0.08	0.08	0.08	
13											
14											
15											
16					=NPV(C19,\$D\$7:\$I\$7)+\$C\$7						
17											
18											
19	Benefits	Costs	Rate	Benefits	Costs	Net	Present Value				
20	Pessimistic	Pessimistic	0.03	\$13.34	\$15.98	(\$2.64)				=D19-E19	
21	Most likely	Pessimistic	0.03	\$15.56	\$15.98	(\$0.42)					
22	Optimistic	Pessimistic	0.03	\$16.90	\$15.98	\$0.91					
23	Pessimistic	Most likely	0.03	\$13.34	\$13.10	\$0.24					
24	Most likely	Most likely	0.03	\$15.56	\$13.10	\$2.47					
25	Optimistic	Most likely	0.03	\$16.90	\$13.10	\$3.80					
26	Pessimistic	Optimistic	0.03	\$13.34	\$10.96	\$2.37					
27	Most likely	Optimistic	0.03	\$15.56	\$10.96	\$4.60					
28	Optimistic	Optimistic	0.03	\$16.90	\$10.96	\$5.93					

Figure 2. Pivot table output

		Present Value Net Benefits				
		Rate				
Benefits	Costs	0.03	0.035	0.04	0.045	0.05
Most likely	Most likely	2.47	2.24	2.02	1.80	1.59
	Optimistic	4.60	4.36	4.12	3.90	3.68
	Pessimistic	-0.42	-0.63	-0.84	-1.03	-1.23
Optimistic	Most likely	3.80	3.55	3.30	3.06	2.83
	Optimistic	5.93	5.67	5.41	5.16	4.91
	Pessimistic	0.91	0.68	0.45	0.23	0.01
Pessimistic	Most likely	0.24	0.06	-0.12	-0.30	-0.47
	Optimistic	2.37	2.18	1.98	1.80	1.62
	Pessimistic	-2.64	-2.81	-2.98	-3.13	-3.29

ratio (benefit divided by cost). Obviously, results also depend on cost and benefit estimates.

The table at the top of Figure 1 contains six sets of pessimistic, most likely, and optimistic benefits and costs. We are assuming a range of discount rates from 3 to 5% increasing in one half of 1% increments. The table below the top table contains all possible combinations of benefits, costs, and discount rates; it is used to build a pivot table (see below) to display the results in a readable fashion. The formulas in the present value columns for benefits and costs all resemble

the ones in cells D19 and E19. The formulas in Column F all resemble the one in cell F19.

The syntax of the present value (NPV) function is =NPV(rate,range), where rate is the discount rate and range is a range of cells. Column C contains the discount rates. As the Excel's help utility notes, the NPV function begins discounting with the first cell in the range specified in the formula. Thus, if the block D10:I10 was replaced with C10:I10, the NPV function would begin the discounting 1 year too early. The cell address of the cost for year 0 is subtracted separately (cell C10). There

is at least one way to enter the benefit-cost data that would make keying in the formulas more efficient than what we show in Figure 1, but our presentation demonstrates the basic principles more clearly.

The pivot table in Figure 2 summarizes all the combinations of benefits, costs, and discount rates within the specified ranges. Note that along each row, the present value figures decline with increasing discount rates, as they must. This distinctive pattern does not guarantee that the functions in Figure 2 were entered correctly, but it suggests that they were. This table would give decision makers a sense that optimistic or most likely assumptions would produce positive results even at the highest calculated discount rate, but that some inclusion of pessimistic benefit or cost assumptions can generate negative results.

Pivot Tables

We saw that Figure 1 displays part of a worksheet containing four dimensions: benefits characterized as pessimistic, most likely, and optimistic; costs with the same labels; discount rates; and net benefits. There is no way to display all of these data on a single two-dimensional worksheet. It would be even worse if there were an additional dimension such as geographical area. The pivot table in Figure 2 allows us to see benefit and cost categories in rows and rates in columns. Had it been necessary to include geographical categories, they could have been added as well. Potential reader confusion limits the number of variables that can be included in a single pivot table, but it is an efficient way to lay out multidimensional data sets. We lack the space to explain the steps required to build a pivot table, but many reference works (e.g., Berk & Carey, 2004) and Excel help screens explain it clearly. The pivot-table procedure allows us to include and exclude whatever data we want.

The data filter function offers another way to suppress or include data. The data filter function is

invoked in Excel by clicking in succession on Data, Filter, and Autofilter. This choice automatically generates menus at the top of a worksheet—one menu for each column. For example, in Figure 1, the menu for the benefits column (beginning in cell A18 and going downward) contains the three benefits categories (pessimistic, most likely, and optimistic). By selecting, for example, pessimistic benefits and in the next column pessimistic costs, the worksheet would seem to eliminate all other benefits and costs, giving the decision maker a quick look at the worst-case scenarios at various discount rates for this project.

Spreadsheets and Statistics

Spreadsheets possess statistical computing capabilities adequate for many public-administration tasks. Excel can perform basic descriptive statistics, curve smoothing, analysis of variance, multiple regression, random number generation, sampling, *t*-test, and others. These capabilities are part of the Analysis ToolPak, an add-in that comes with all versions of Excel that are likely still to be in use. An add-in is a computer program that when installed merges with a base program (Excel in this case) making the add-in an integral part of the base program. Excel's Analysis ToolPak can be supplemented by another add-in called StatPlus, available with *Data Analysis with Microsoft Excel* (Berk & Carey, 2004). Various editions of this reasonably priced text contain versions of StatPlus designed to work with specific versions of Excel and Windows. This book is perhaps the most clearly written and informative introduction to spreadsheet usage and introduction to statistics on the market. Relatively advanced statistical analysis cannot be done readily or at all with Excel, but for many administrative jobs, it may be adequate. Advanced work is more conveniently performed using statistics packages to be discussed later.

Spreadsheets are better tools than statistics packages for data entry, data editing, data storage,

and printing. A statistics package is a computer program optimized for statistical analysis. Generally speaking, the best statistics packages (e.g., SYSTAT, SPSS, and NCSS) are more suitable for statistical analysis than are spreadsheets, but spreadsheets are superior for printing raw data as well as performing a wide variety of nonstatistical quantitative chores. If spreadsheet data entry follows simple rules specified by all leading statistics package manufacturers, the data may be exported to the statistics package; leading statistics packages can read spreadsheet files almost as easily as they can their native format.

Data to be read by a statistics package should be arranged so that the top row is devoted to variable names and each subsequent row is an observation. For example, if the analyst is examining U.S. coal-mine fatalities over time, data would be arranged as seen in Table 1.

Elsewhere on the worksheet, the definitions of the variables can be entered along with data sources and other information. It is especially convenient to enter such information in the form of comments in cells containing variable names. Some statistics packages allow such comments to be added to a data set, but many do not.

Note that the FATLPERMINER variable in Column D was calculated with a formula that reads =C2/B2. This formula (entered in D2) was copied downward to the last row of data. Someone new to spreadsheet work might think that the copied versions of this formula would all generate the

result 0.002197, but spreadsheet formulas should not be read literally in terms of specific cell locations, but in terms of the relationship among cells. So, for example, the formula in D2 is correctly interpreted to mean to take the number in the Cell One to the left of D2, divide it by the number in Cell Two to the left of D2, and display the result in D2. This is called a relative cell reference formula. When this formula is copied, it maintains this relationship wherever it is put unless the user writes an absolute cell reference formula.

If it is determined that a variable (column) is no longer needed, it can be easily eliminated in a spreadsheet. If a transposition of variables (columns) can aid readability, this is much more easily accomplished with a spreadsheet than a statistics package.

Data sets can easily number 50, 100, 200, or more variables. To avoid mistakes while inputting data, it is important that a key variable, such as YEAR in this example, remain in view. With many statistics-package data-entry screens, when more than approximately five variables have been added, the first column (variable) scrolls to the left out of sight. The invisibility of the YEAR variable could easily result in the user entering data for 1999 when it should be in the 2000 row. With a spreadsheet, the YEAR column may be frozen so that it is always in view. Similarly, with some statistics packages, once data have been entered for roughly the first 20 years, variable names scroll away upward. In a spreadsheet, variable names

Table 1. Example of data arrangement to be read by a statistics package

	A	B	C	D
1	YEAR	MINERS	FATALTIES	FATLPERMINER
2	1935	56,5316	1,242	0.002197
3	1936	58,4494	1,342	0.002296
4
5

can also be frozen in place. One can also view two entirely different parts of the worksheet with a split window, a feature we have not seen in any statistics-package data-input screen.

Many statistics packages lack even elementary print formatting capabilities producing crude-looking or confusing raw-data printouts. The statistics package output may be saved in a format readable by a word-processing program (or a spreadsheet) where the cleanup of layout and fonts may be accomplished. A spreadsheet allows complete control over what appears on which printout page. It is far superior to a statistics package in terms of the appearance of the printout.

Curve Smoothing and Forecasting

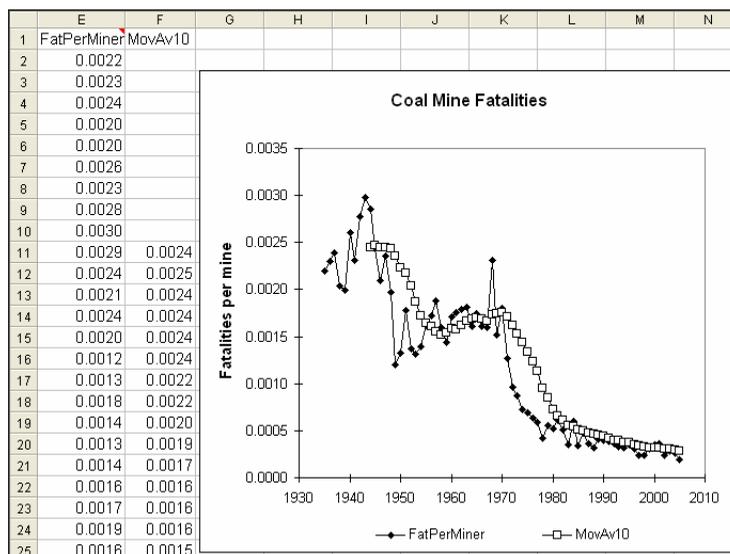
A time-series data set is “a sequence of observations taken at evenly spaced time intervals” (Berk & Carey, 2004, p. 408). Although it is characterized by serious theoretical weaknesses, the single moving average is one of the most common tools of applied time-series analysis, especially among

financial securities analysts. The single moving average is easily calculated and plotted using a spreadsheet.

The time-series data in Figure 3 represent fatalities per coal miner in the years 1935 to 2005. The data are relatively noisy. To highlight trends obscured by year-to-year fluctuations and to perform rudimentary forecasting, it can be smoothed by calculating the mean of data over a number of time periods, in this case 10 years (see Berk & Carey, 2004, pp. 424-426). The first moving average data point (cell F11) is the mean of the first 10 years of raw data; the second moving average data point omits the first year of raw data and includes the 11th, and so forth.

The distance between the moving average curve and the raw-data curve is sensitive to changes in trend. Because each point on the moving average curve represents the past, each point on the moving average will be above raw data when the trend is down as it is in most of the years shown in Figure 3. Note the sharp spike in 1968. This data point might have reflected a

Figure 3. Curve smoothing with spreadsheets



data-entry mistake, but in fact it was due to an increase in fatalities (222 miners in 1967 and 311 in 1968) combined with a decrease in miners (139,312 in 1967 and 134,467 in 1968; Mine Safety and Health Administration, n.d.). The moving average can easily be calculated by entering the formula and copying it or by using the Excel trend line dialog box.

In financial forecasting (e.g., the Dow Jones Industrial Average), moving averages of 30, 60, or even 90 days are common. In such calculations, each day is treated as being equally important; that is, data from a day 30, 60, or 90 days ago is accorded as much weight as yesterday's data. It seems likely that yesterday's data are somewhat more important than data from a day before that, which in turn is more important than data from 3 days ago, and so forth. A family of curve-smoothing techniques known as exponential smoothing takes this likely truth into account by systematically weighting more recent time periods more heavily than earlier ones (Fosback, 1987). Excel also contains a number of exponential smoothing techniques (see Berk & Carey, 2004, pp. 427-429).

Many time-series data sets contain several components. As we have already seen, one such component may be a trend, a relatively long-term tendency to move upward or downward. Another time-series component may be seasonality: In a given month or quarter, the data may generally be average, below average, or above average for the year. City sales-tax revenues are often seasonal with, for example, revenues for November and December being above average for the year because of holiday business. Within a quarter or month, there may also be weekly patterns or other cycles. Finally, time-series data also contain random variation. A family of techniques called complex time-series analysis breaks data of this sort into its component parts (trend, seasons, and cycles) and reassembles those parts into the future generating a forecast. Such forecasts reproduce

the long-term trend of the data set and the ups and downs of seasons and cycles to produce impressive-looking graphs. Of course, the accuracy of such forecasts depends on whether past trends, seasonal patterns, and cycles are maintained. Excel and StatPlus can almost automatically perform complex time-series analysis using trend and season but not cycle. The statistics package NCSS contains one of the most sophisticated time-series analysis modules that we have seen among general-purpose statistics packages (see as follows).

Sampling

Spreadsheets contain functions that can be used to generate random numbers convenient for sampling. In Excel, the function RAND() generates a random number greater than or equal to 0 and less than 1. For example, entering =RAND() into three cells yielded 0.917547614, 0.364495084, and 0.209362913. The numbers are evenly distributed. In other words, there is an equal likelihood that any digit will appear in any of the decimal places.

The Excel help utility advises the following: "To generate a random real number between a and b, use: RAND()*(b-a)+a." The generation of integers (whole numbers) instead of decimals can be accomplished by combining the above technique with the INT function, which rounds down to the nearest integer. For example, random numbers between 1 and 500 are generated by the formula INT(RAND()*500)+1.

The RANDBETWEEN function is an even easier way to generate evenly distributed random integers. The syntax is =RANDBETWEEN(low,high), where low is the smallest integer that will be returned and high is the largest. Worksheets can be designed to generate simple random samples and systematic samples with or without stratification. With the spreadsheet rows representing cases or subjects, each row may be numbered quickly, and those numbers may be used with a formula to

generate a simple random sample. The stratified systematic sample is first developed by sorting the data on one or more variables (for example, sex). Using a formula to generate a skip interval (population size divided by sample size), it is easy to generate another formula for sampling the list once simple random sampling within the first interval generates a random start number. If the data were sorted by sex, the resulting sample will have numbers of males and females proportionate to their numbers on the list.

Spreadsheets contain a calendar with dates represented by integers that can be used to generate random dates between desired limits. The number 1 represents January 1, 1900, something that the reader can confirm by keying 1 into any cell and formatting it as a date (Format, Cell, Number, Date). The generation of random dates is accomplished by determining what integers represent the desired dates and setting up the correct `INT(RAND())` or `RANDBETWEEN` formula.

Linear Programming

Linear programming is a technique for minimizing or maximizing a variable in the face of constraints. We might, for example, be attempting to minimize cost in a government cartography shop by making the most efficient use of equipment and personnel or to maximize the use of city garbage trucks by minimizing travel time between where they start in the morning, pickup locations, and dump locations (both examples are actual applications). Spreadsheets can perform such calculations. Explaining how data for a particular situation are arranged for linear programming solution in a spreadsheet requires far more space than is available here, but many texts on advanced spreadsheet usage provide guidance (Albright, Winston, & Zappe, 2006; Gips, 2002; Weida, Richardson, & Vazsonyi, 2001).

Database Applications

Everything regarding spreadsheet usage up to this point has concerned databases, but there are specific data-storage and -manipulation operations known collectively as database management. The traditional database structure is identical to the column (variable) and row (observation) arrangement discussed above except that in database management, columns are known as fields and rows as records. Inventories, mailing lists, and room assignments in buildings are examples of the kind of information stored and processed as part of database-management operations.

Spreadsheets can be used for this purpose if the number of records does not exceed the maximum number of rows in the spreadsheet. This amount varies depending on spreadsheet manufacturer and version, and it also may be limited by the amount of memory (RAM) in the computer. The practical maximum number of records is also governed by how the user plans to manipulate data; many operations can take additional space. With the spreadsheet software found on most computers of average memory capacity, a database larger than 5,000 records is probably better processed using a database-management program such as Microsoft Access or the easier to use and more reasonably priced Alpha-Five. Database-management programs are better suited to manipulating large data sets than spreadsheets. Database-management programs are also better suited for creating customized data-entry screens.

Statistics Packages

As noted previously, spreadsheets can perform statistical calculations and database-management programs can as well, but if those operations are many in number, involve relatively sophisticated techniques, and/or involve large data sets, statistics

packages such as SYSTAT, SPSS, or NCSS are far superior to spreadsheets and database-management programs. For some calculations, statistics packages are the only tool that can be used by the nonprogrammer.

In public-administration applications, a statistics package might be used for elementary operations such as descriptive statistics. Even for this purpose, a statistics package is more convenient than a spreadsheet if work is to be done on more than a few dozen variables. Some policy analysis work requires more advanced techniques.

Regression Analysis

Regression analysis can be employed to explain the relationships between and among variables, and it can be used alone or in concert with other techniques for forecasting. Regression analysis is one of the most commonly described techniques in the statistics literature, and there is no point in our duplicating the many fine treatments of this subject (Berk, 1998; Berk & Carey, 2004; Garson, 2002; Schroeder, Sjoquist, & Stephan, 1986; Wilkinson, Blank, & Gruber, 1996). An example of one application of regression analysis concerns the relationship between the change in per-capita income and the change in university enrollment for the states. This example is part of a university lobbyist's claim that a state's economic well being was directly proportional to the number of people enrolled in universities. Using the language of statistics, a state's economic health was the dependent variable being acted on by the independent variable, higher education enrollment. Regression analysis allows the user to determine how much, if any, variation in a state's economic health is explained by variation in higher education enrollment.

Regression analysis is used in the fields of human-resource administration and justice and public safety to analyze whether discrimination based on race, sex, or other illegal factors is statistically likely. In this instance, regression

analysis with several independent variables (multiple regression) is often the tool of choice (Berk & Carey, 2004). Regression analysis also can be used for forecasting time-series data, with time serving as the independent variable (horizontal axis). The resulting regression line or the regression equation can be used to forecast the dependent variable, which can be anything from tax revenues to jail inmate populations.

Interrupted Time-Series Analysis

Interrupted time-series analysis, an application of regression analysis, is an especially helpful analytical tool for policy analysis (Welch & Comer, 1988). For example, the authors (Permaloff & Grafton, 1995) used it to test the claim that George Wallace's first term as governor had a positive effect on education spending in Alabama, and Bloom (1999) estimated educational program impacts on student achievement.

With interrupted time-series analysis, data must be available for a number of time periods before and after an event such as the election of a new governor or the enactment of a coal-mine safety law. This extremely useful technique measures whether the event has made a statistically significant difference and the nature of that difference. There are two basic kinds of statistically significant differences that an event can produce: an abrupt increase or reduction and a long-term increase or reduction. In addition, an event may produce no abrupt change and/or no long-term change. These alternatives yield nine possible combinations: no immediate change but a long-term increase, no immediate change but a long-term reduction, and so forth.

As an example, we can apply this technique to the data in Figure 3. Figure 3 shows a sharp spike in fatalities per coal miner that occurred in 1968 followed by a dramatic reduction after passage of the Coal Mine Safety Act of 1969 (CMSA). The question is whether the CMSA improved coal-mine safety. The analysis begins by adding

two variables to the data set. One variable (CV) is a counter variable that begins at zero in 1935 and remains zero for each successive year until 1970 when it becomes one and remains one for the remaining years of the data set. A multiple regression analysis including CV will result in an indication of statistical significance connected with CV if there is an immediate or abrupt shift in fatalities per miner. The second variable is a post-policy intervention counter variable (PPICV). Like CV, it begins at zero in 1935 and remains zero for each successive year until 1970 when it becomes one and, unlike CV, it then increases by one in each successive year. There will be an indication of statistical significance for PPICV if continuing change occurs after CMSA.

Figure 4 shows SYSTAT output for the multiple regression analysis interrupted time series. The P(2 Tail) column shows that the CV variable is statistically significant beyond the 0.001 level, indicating that the drop after 1968 is statistically significant. On the other hand, the PPICV variable is not statistically significant showing that the long-term reduction in fatalities after 1968 is not statistically significant. However, this is only a first cut at the analysis. Many questions

remain. One concerns the 1968 spike. It alone could be the origin of statistical significance for CV. We eliminated the impact of the 1968 spike by replacing it with the mean of fatalities per miner for 1967 and 1969. The resulting analysis still showed CV to be statistically significant but this time at the 0.001 level. Nevertheless, the hypothesis that CMSA resulted in a one-time drop in fatalities is supported. (PPICV remains not statistically significant.)

The lack of statistical significance for PPICV may be in part due to the 1935 starting point that we selected. Soon after 1935, a sharp drop in fatalities began that continued until roughly 1950. The period of 1950 to 1970 was marked by a gradual increase in fatalities. After 1970, fatalities resumed their downward path. What would happen if we began the analysis with 1950 instead of 1935? The result is that both CV and PPICV were statistically significant beyond the 0.001 level suggesting that CMSA had both a short term and long term effect.

Questions regarding how to treat data spikes (or drops) and at what year the data series should be said to begin are very common in interrupted time-series analysis. Answers are matters of

Figure 4. Interrupted time series data output from SYSTAT

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	0.053	0.010	0.0	.	5.203	0.000
YEAR	-0.000	0.000	-0.654	0.120	-5.013	0.000
CV	-0.001	0.000	-0.343	0.250	-3.797	0.000
PPICV	0.000	0.000	0.042	0.187	0.405	0.686

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.000	3	0.000	141.127	0.000
Residual	0.000	67	0.000		

*** WARNING ***
Case 36 is an outlier (Studentized Residual = 3.280)

Durbin-Watson D Statistic 0.814
First Order Autocorrelation 0.590

judgment. The 1968 spike was real and certainly had to be included in a regression run, but it was an unusual figure that invited another run with the spike deleted. Similarly, the original data set from which we took these data began in 1900. We assume that it would not be appropriate to run the analysis that far back; presumably, coal-mine technology improved considerably since 1900, making data of that era meaningless. But at what time point should the analysis begin? If the authors possessed greater knowledge of mining technology, we could have selected a starting point that marked what might be termed contemporary technology. The only certainty is that this data set required multiple analyses and required that they be reported.

It should be noted that we omitted another factor that should be covered in a serious analysis of coal-mine safety: Surface mining is inherently safer than deep mining, and surface mining has increased over the years. A serious analysis would treat them separately. Again, this is the kind of judgment that must be made in analysis of this sort.

Curve Smoothing and Forecasting

SYSTAT, NCSS, SPSS, and other statistics packages contain a wide variety of curve-smoothing and forecasting techniques, some of which have already been discussed in connection with spreadsheets. Generally speaking, this kind of analysis is more conveniently done in a statistics package than a spreadsheet, and statistics packages offer a greater variety of analysis techniques.

Complex Time-Series Analysis

The topic of complex time-series analysis was introduced earlier in the section on spreadsheets. The NCSS statistics package contains one of the easiest to use and most sophisticated complex time-series analysis modules on the market as well as possibly the clearest documentation available

(Armstrong, Collopy, & Yokum, 2005; Jarrett, 1987; NCSS, 2004).

Database Management

Database-management software such as Microsoft Access or Alpha-Five from Alpha Software Corporation is commonly used for personnel records, inventory management, or whenever large numbers of records will be stored. Database-management software is also ideal when data are to be entered by a computer novice. Data-input screens can be designed that include instructions for the user to help ensure that the correct types of information are keyed into particular fields. To some degree, each field can also be designed so that incorrect types of information cannot be entered. A spreadsheet can be modified in a similar fashion, but the expertise required to accomplish this is somewhat greater than is needed with a database-management program, especially one intended for ease of use such as Alpha-Five.

With database-management software, large numbers of records can be sorted and queried and particular records and particular fields can be printed or exported to other software (e.g., statistics, spreadsheet) for further processing. Database-management programs also are ideal for the generation of large numbers of mailing labels and form letters (Schmalz, 2006).

Access, Alpha-Five, and many other database-management programs are relational; they can link and draw information from two or more differently structured databases (as long as the databases have a common field) as if those databases were one (Grafton & Permaloff, 1993). This powerful feature is not shared with the other categories of software represented here.

On a recent project for a state government agency, we received two data files in Excel for each of 19 counties. The files had no common column or field. We used Excel's LEFT and RIGHT functions to extract portions of one column in one data file to create a variable or field common to

one found in the second data file. We then created one merged data file for each county using Alpha-Five. The Alpha-Five file is in dBASE format and readable by Excel and the statistics packages we use most often. We employed Excel pivot tables and sorting commands to generate descriptive data required for the project by the federal government. The databases were exported into SYSTAT for multiple regression and other analysis, including cross-tabulation with control variables and associated statistics. The extraction option in SYSTAT allowed us to create specialized subsets of the data files quickly and easily. Portions of the SYSTAT analysis output were saved in text format and read into WordPerfect for final report writing.

Large organizations sometimes buy specialized relational database-management programs set up for their specific applications. Such programs may cost thousands, if not hundreds of thousands, of dollars to buy, install, and maintain. Often a training component is included in the installation fee. Database-management programs are frequently used to store personnel records. Another function is the processing of government regulatory agency records. For example, a state regulatory board may have multiple responsibilities: licensing applicants to a profession, disciplinary functions including the receipt of complaints and their investigations, monitoring of continuing-education providers, and oversight of educational training programs. Each area would be served by different staff personnel with differing record-keeping needs. As long as each area has at least one common field such as Social Security number as an identifier linking the databases, one set of records could be queried for information needed by another area of the organization. For example, the licensing section may need to know the current status of disciplinary actions to determine if a renewal should be granted. With relationally linked databases, licensing personnel can query the separate components of the database as if they were a single database. At the same time, access

to specific fields can be granted or blocked on a need-to-know basis.

We have observed three pitfalls with database management in government. If standard database software is adapted by a government agency employee for use by the agency (all such software is designed to be tailored to fit specific needs), it is important that at least one other employee be taught how the customization was implemented and backups be made of the complete system, and, of course, all data. If the employee who performs the customization leaves the agency and the computer on which it is running crashes, an agency can be left without access to its records or worse yet the records may disappear. In fact, this has happened. Events of this sort are less likely to occur if the program is sold and maintained by a commercial developer assuming that the developer remains in business.

Another database-management pitfall is the mutation of the database over a period of years with no records kept of changes made. The original database structure may be clearly described by paper or online documentation, but incremental modifications over a period of years with no accompanying editing of the documentation can produce a database difficult for newcomers to use.

A third pitfall is a database that does not reject grossly incorrect data entries. No software can detect slight numerical mistakes in a field intended to hold numbers such as a 1 instead of a 2, for example, but any modern database-management system ought to reject such mistakes as two-digit numbers entered where there should be three or letters entered where numbers are needed. Also, a database-management program should force input of data into a field containing mandatory information before allowing the user to complete follow-on fields. Database-management systems that permit gross mistakes will produce a database full of inaccuracies, especially when hundreds of people are keying in data.

Data Mining

Data mining is a collection of techniques primarily intended to handle extremely large secondary data sets. A secondary data set is by definition not collected with analysis as the primary objective, so the way this transactional information is gathered and structured may not be ideal for traditional statistical analysis. The Transportation Security Administration (TSA) provides an example of a large secondary data set. This agency uses credit reports and more than 100 commercial and government databases to identify potential terrorists among airline passengers (Rotenberg & Hoofnagle, 2003). Other examples of governmental use of data mining come from law enforcement, higher education, and tax administration (SPSS, n.d.).

Large data sets require enormous computer capacity and they sometimes are in the form of multiple related files (the TSA example above fits this characteristic). These factors mean that traditional statistical techniques cannot be applied without complicated sampling (Hand, 1998). Another problem with large secondary data sets is that they almost certainly contain contaminated data requiring mechanized (because they are so large) data cleaning.

The secondary nature of the data sets also means that they may be characterized by selection bias (some geographical regions may be over-represented) and nonstationarity (the population may be changing in subtle ways). These problems can also arise in smaller data sets, but they are more difficult to detect and correct in large data sets (Hand 1998).

Another complication with data mining is that it is often exploratory in nature (Statsoft, 2003). Analysts are often not testing relationships derived from theory as they do (or should) in academic research. Instead, they are frequently searching for unknown clues that lead to answers to vaguely defined questions (Hand, 1998). The sheer size of data-mining data sets combined with the frequent

lack of theory also makes analysts vulnerable to discovering spurious relationships.

Data mining requires data-mining software, large computers, and specialists. Our sense is that an army of particularly hungry dealers and consultants awaits any government agency venturing into this field.

Project Planning and Management

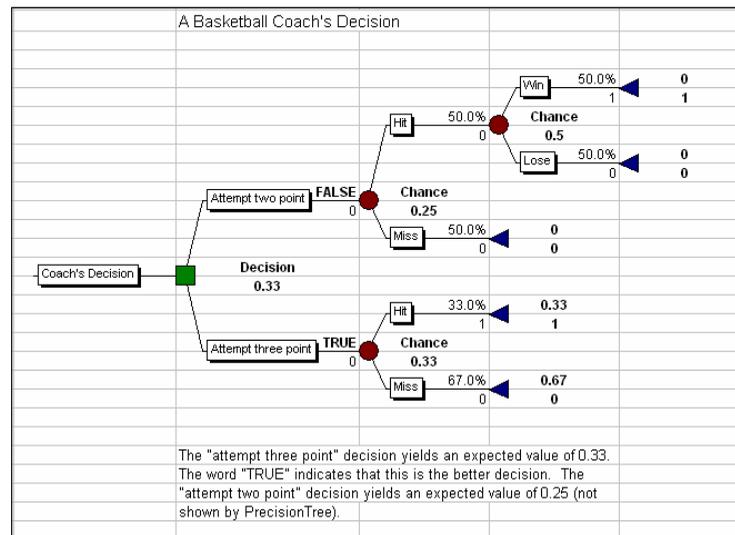
Project planning and management programs are designed to assist in planning and managing complex projects in terms of time schedules as well as budgets, personnel, and other resources. This category of software can be used to schedule and estimate personnel requirements and budgets for construction jobs, landscape projects, conference planning, scheduling activities in a complex lawsuit, or any other enterprise with a clearly defined beginning and end. Over the years, the widely used Primavera Project Planner (P3) has received some of the most positive reviews, but Microsoft Project is easier to use. A Google search on “project planning software” will yield many other products, some of which offer free trial downloads.

Project planning requires that a project be broken down into building blocks called activities or tasks. An activity or task is the responsibility of one person, involves one kind of work, can be scheduled, and uses the same resources from beginning to end.

Gantt charts are one of two major graphical devices used by this family of programs. A Gantt chart is a bar graph with a horizontal bar representing each task in a project. Each bar’s length is proportional to the time the task requires. The beginning of a bar is located at the start date of a task, and the end of the last bar is located at the date the last task should be complete.

The PERT (Project Evaluation Review Technique) chart is the other major graphical tool used by project-planning and -management programs. It is a better tool for highlighting relationships

Figure 5. A decision tree



among tasks, but the Gantt chart is more space efficient, allowing more tasks to be displayed on screen or on a single sheet of paper. Programs of this sort can also display bar charts and tables showing personnel plans and budgets.

Information is entered in these programs in forms. The user keys in task names, duration, personnel, and other resources used to complete the task. The software calculates start and finish dates, personnel plans, budgets, and so forth.

Frequently, the first draft of a project plan results in an end date that is too late. These programs automatically highlight the longest path (the critical path) in a plan where a number of tasks are being executed simultaneously. In concentrating changes on the critical path, an administrator can reduce the completion time of the overall project by focusing attention where it will do the most good. Project planning and management programs are also used to monitor progress once a project is under way. Project-wide implications of early or late task completions are quickly observed.

Decision Analysis

Decision analysis is a graphical form of applied probability theory useful for analyzing situations in which uncertainty plays a part. The most common situation to which this technique is applicable has a decision maker trying to choose between two or more policies, each of which is associated with a variety of benefits and costs and probabilities that those benefits and costs will occur. In decision-analysis terminology, a benefit is a positive payoff and a cost is a negative payoff.

As an example, imagine that a basketball team is behind by two points, and it has time for only one shot. Assume that there is a 50% chance of hitting a two-point shot that would tie the game and a 33% chance of hitting a three-point shot that would produce a victory. Also, assume that the team has a 50% chance of winning in overtime. Which shot should the coach call (Clements, 2006)?

Figure 5 shows a decision tree that represents all possible futures in this situation. This decision

tree was entered and calculated in an Excel add-in called PrecisionTree, manufactured by Palisade Corporation. It is available either from the manufacturer or in a trial version in an excellent and reasonably priced (for what it contains) textbook called *Data Analysis and Decision Making with Microsoft Excel* (Albright et al., 2006). This volume also comes with a version of the powerful and easy-to-use simulation tool @RISK.

The decision tree begins on the left with the rectangle (a choice node) representing the coach's choice. The choice to take the two-point shot leads to a chance node (a circle) offering a 50% chance of hitting the shot and a 50% chance of missing. We use 0 as the value of a game loss and 1 as the value of a victory. A miss at the two-point shot produces a payoff of zero. If the two-point shot is made, there follows another chance node offering a 50:50 chance of a win and a loss. If the coach's basic two-point choice could be made repeatedly as an experiment or simulation, the game would be won only 25% of the time. Instead of running this choice as an experiment or simulation, PrecisionTree can calculate the probabilities and payoffs. In this case, the coach should go for the three-point shot, which offers a 33% chance of victory.

Data such as the names of the branches, probabilities, and pay-offs are entered in the PrecisionTree program one form at a time. PrecisionTree draws the tree and calculates the result. The program also allows the user to conduct sensitivity analyses (with some difficulty at first) to determine the effect on the result of differing pay-off values (which would not arise in this case) and probabilities. Although this example is extremely simple, decision trees can be quite complex.

Graphics for Illustrations

Illustrations often play an important part in reports, presentations, budgets, and other docu-

ments. Specialized software can be of great assistance, augmenting the graphics features built into the software discussed previously (Permaloff & Grafton, 1996).

Flow Charts and Organization Charts

A family of programs such as RFFlow and Visio is used to create flow charts, organization charts, and similar diagrams featuring boxes (of all shapes) and lines (solid, dotted, with or without arrowheads). Diagrams are created by the user selecting box shapes from menus, keying in labels (the boxes expand to fit labels), and connecting the boxes with lines also selected from menus. Once placed on the screen, boxes and lines can be rearranged as aesthetics and logic dictate.

The choice of programs depends on features needed, budget, and time available for learning. For example, comparing RFFlow and Visio, the former is the least expensive (\$49 vs. \$199) and by far easier to learn, but it is limited to creating flow charts and organization charts. Visio can render a greater variety of shapes, and it can be tied to powerful Visual Basic programs. Visio, like all Microsoft products, subjects the user to an intrusive online product-registration process lacking in RFFlow. As with most other products discussed in this chapter, a Web search yields many other flow-chart and organization-chart generation programs, many of which can be downloaded.

We do not recommend that most public administrators attempt to employ a computer-aided design (CAD) program for occasional diagram work unless the individual is already trained in its use. CAD programs can do far more than specialized flow-chart programs, but they exact an even higher price than Visio in terms of learning time, and they require many steps and much more time to complete tasks (see Permaloff & Grafton, 1996.)

Scanned Graphics

Documents can sometimes be improved with the inclusion of scanned material such as photographs. Scanners range in price from \$50 to approximately \$650, but even the least expensive are sufficient for many applications. Flatbed scanners are probably the easiest to use. The flatbed scanner resembles a small photocopy machine. The original image is placed in the scanner that generates a digitized version of the image that can be saved on disk and inserted into application programs including word processing, spreadsheet, database management, and project planning. In some cases, the image can be dragged directly from the scanning program into an application program without the need to save on disk.

Presentation Programs

The use of computers and computer projection hardware or large monitors for presentations is gaining in popularity as the cost of projection equipment declines. Although 35mm slides, the old mainstay of presentations, are virtually gone, computer presentations are built on what might be termed virtual slides. Transparencies, another kind of slide, are still used, but they appear to be regarded as passé despite their continued practicality, readability on a large screen, and low cost.

A computer slide may contain any combination of text and graphics, including sound and motion. A single slide may contain elements from several sources. Menu choices or built-in templates determine the background on which text and/or graphics will appear. That background can be any color or texture. Similarly, each slide's basic layout (whether there is a centered heading, bulleted text, a graphic, columns, etc.) is chosen from menus or templates.

It is easy to include so many clever graphics and sound effects in a computer presentation that the audience is distracted from content or even irritated by excesses. Indeed, a former chairman

of the Joint Chiefs of Staff became so annoyed with elaborate PowerPoint presentations he reportedly attempted to ban the program's use in the military.

Another drawback of slide presentations is audience boredom, a major reason some presenters employ what they hope are entertaining effects. One clever and subtle solution to the boredom problem is the use of two projectors and two screens. Image 1 appears on Screen A and then Image 2 on Screen B (with Image 1 still seen on Screen A). Image 3 then replaces Image 1 on Screen A, and so forth. This idea's effectiveness must literally be seen to be believed. It reduces boredom and eliminates the need for gimmicky dissolves from slide to slide.

Geographic Information Systems

Geographic (or geographical) information system (GIS) programs assemble, store, and display data that are specific to geographical areas such as countries, states, counties, or zip codes. Data might include information regarding population, economics, numbers of users of government agency services, and crime rates. Data are typically displayed in the form of maps in which geographical areas are colored or patterned to represent various levels of whatever is being measured (Garson & Vann, 2001; Orford, Harris, & Dorling, 1999; Vann & Garson, 2001; Vann & Garson, 2003).

Some GIS programs such as MapInfo and BusinessMAP are free standing while others are attached to software such as Excel or Visio. GIS packages vary widely in terms of file formats that they can read, data-manipulation features, display capabilities, and price, so considerable research is required to make an informed purchasing decision. A list of resources including GIS programs, boundary files, and data files is available at the Web site of the Department of Geography, University of Edinburgh (<http://www.geo.ed.ac.uk/home/agidict/welcome.html>).

LONG-TERM STORAGE

Editable data files (e.g., word processing or worksheets) and scanned images of printed pages can be stored digitally for archival purposes. But on what medium should archived documents be stored? One convenient but unsafe possibility is CD-R (compact disk-recordable) or CD-RW (compact disk-rewritable) disks. Unfortunately, these media may not be suitable for long-term storage. CD-R and CD-RW disks are not the same as CD-ROMs (compact disk-read only memory), which are commonly used for commercial software and music. CD-ROMs are manufactured by molding polycarbonate with pits that constitute digital information. The pits are then coated with a reflecting layer of metal. That layer is then coated with protective plastic. Even these relatively tough disks can be rendered unreadable by scratches, but with careful handling, they can probably last for decades.

Recordable or rewritable disks probably have a much shorter life span because they are physically more delicate. CD-Rs use transparent dye, not pits, as their data-storage method. When the disks are “burned,” the computer’s laser darkens the dye. CD-RWs use an alloy that changes state when the recording process occurs (Ancestry.com, 2000). CD manufacturers claim that recordable and rewritable disks can last for decades, but Kurt Gerecke, an IBM storage specialist, estimates that inexpensive and high-quality disks only last approximately 2 to 5 years, respectively (Blau, 2006). Gerecke also cautions against relying on a hard drive for archival purposes. Here the problem is not surface degradation, but the longevity of the disk bearings. Gerecke recommends magnetic tapes for long-term storage. Tape life spans range from 30 to 100 years depending on their quality (Blau; Practical PC, n.d.).

Storage media formats represent another and different kind of problem for archiving. Consider

how difficult it would be to locate a disk drive that can read a 5¼-inch floppy disk much less the larger format that preceded it. The safest materials for long-term storage may still be paper and microfilm or microfiche.

CONCLUSION

Most tasks in government employing a computer can be accomplished more efficiently with a variety of tools. Happily, most of those tools are under \$500. Of the software discussed here, only customized database-management programs, data-mining software (which often require highly trained staff or hired consultants), and high-end geographic information systems go past the \$1,000 mark.

REFERENCES

- Akers, E. J. (2006). *A study of the adoption of digital government technology as public policy innovation in the American states*. Unpublished doctoral dissertation, Auburn University, Auburn, AL.
- Albright, S. C., Winston, W. L., & Zappe, C. J. (2006). *Data analysis and decision making with Microsoft Excel* (3rd ed.). Stamford, CT: Duxbury, Thomson Learning.
- Alpha-Five [Computer software]. (n.d.). Burlington, MA: Alpha Software Corporation. Retrieved from <http://www.alphasoftware.com>
- Ammons, D. N. (2002). *Tools for decision making*. Washington, DC: CQ Press.
- Ancestry.com. (2000). *The life span of compact disks*. Retrieved March 26, 2006, from <http://www.ancestry.com/learn/library/article.aspx?article=2131>

- Armstrong, J.S., Collopy, F., & Yokum, J.T. (2005). Decomposition by causal forces: A produce for forecasting complex time series. *International Journal of Forecasting*, 12 (pp. 25-36).
- Berk, K. N. (1998). *Introductory statistics with SYSTAT*. Upper Saddle River, NJ: Prentice-Hall.
- Berk, K. N., & Carey, P. (2004). *Data analysis with Microsoft Excel: Updated for Windows XP*. Belmont, CA: Brooks/Cole-Thomson Learning.
- Blau, J. (2006, January 10). Storage expert warns of short life span for burned CDs: And don't count on hard disk drives for long-term storage, either. *Computerworld*. Retrieved March 20, 2006, from <http://computerworld.com/hardwaretopics/storage/story/0,10801,107607,00.html>
- Blau, P. M., & Meyer, M. (1971). *Bureaucracy in modern society*. New York: Random House.
- Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2001). *Cost-benefit analysis: Concepts and practice*. Upper Saddle River, NJ: Prentice-Hall.
- Borgmann, A. (1988). Technology and democracy. In M. E. Kraft & N. J. Vig (Eds.), *Technology and politics* (pp. 54-74). Durham, NC: Duke University Press.
- BusinessMAP 4 [Computer software]. (n.d.). Dallas, TX: CMC International. Retrieved from <http://www.businessmap4.com>
- Clementine [Computer software]. (n.d.). *Chicago: SPSS*. Retrieved March 26, 2006, from <http://www.spss.com/downloads/Papers.cfm?ProductID=Clementine&DLT>
- Clements, J. (2006, March 15). Net gains: How watching basketball can improve your approach to investing. *Wall Street Journal*, p. D1.
- Clinton, W. J. (1994). *Executive Order 12898: Federal actions to address environmental justice in minority populations and low-income populations*. Retrieved March 19, 2006, from http://www.fhwa.dot.gov/legsregs/directives/orders/6640_23.htm
- Coursey, D. H., & Killingsworth, J. (2005). Managing e-government in Florida: Further lessons from transition and maturity. In G. D. Garson (Ed.), *Handbook of public information systems* (2nd ed., pp. 331-343). Boca Raton, FL: CRC Press, Taylor & Francis Group.
- DJM Consulting. (2002). *Benefit-cost analysis of Seattle monorail proposal: Elevated Transportation Company Board PowerPoint briefing*. Retrieved March 20, 2006, from http://archives.elevated.org/documents/Cost_Benefit_Analysis_DJM.pdf
- Ellul, J. (1964). *The technological society*. New York: Vintage.
- Federal Highway Administration. (1998). *FHWA actions to address environmental justice in minority populations and low-income populations* (Directive 6640.23). Retrieved March 19, 2006, from http://www.fhwa.dot.gov/legsregs/directives/orders/6640_23.htm
- Fosback, N. (1987). *Stock market logic*. Fort Lauderdale, FL: Institute for Econometric Research.
- Freeman, D. M. (1974). *Technology and society*. Chicago: Markham.
- Fuguitt, D., & Wilcox, S. J. (1999). *Cost-benefit analysis for public sector decision makers*. Westport, CT: Quorum Books, Greenwood Publishing Group, Inc.
- Garson, G. D., & Vann, I. B. (2001). Resources for computerized crime mapping. *Social Science Computer Review*, 19, 357-361.
- Gips, J. (2002). *Mastering Excel: A problem-solving approach* (2nd ed.). New York: John Wiley.
- Grafton, C., & Permaloff, A. (1993). Statistical analysis and data graphics. In G. D. Garson & S. S. Nagel (Eds.), *Advances in social science and*

computers (Vol. 3, pp. 267-284). Greenwich, CT: JAI Press.

Grafton, C., & Permaloff, A. (2005). Analysis and communication in public budgeting. In G. D. Garson (Ed.), *Handbook of public information systems* (2nd ed., pp. 463-488). Boca Raton, FL: CRC Press, Taylor & Francis Group.

Gullick, L. (1996). Notes on the theory of organization. In J. M. Shafritz & J. S. Ott (Eds.), *Classics of organization theory* (pp. 86-95). Fort Worth, TX: Harcourt Brace College Publishers.

Hand, D. J. (1998). Data mining: Statistics and more? *American Statistician*, 52, 112-118. Retrieved March 26, 2006, from <http://www.amstat.org/publications/tax/hand.pdf>

Heeks, R. (2006). *Implementing and managing eGovernment*. London: Sage.

Jarrett, J. (1987). *Business forecasting methods*. Oxford, England: Basil Blackwell.

Levin, H. (1983). *Cost-effectiveness: A primer*. London: Sage.

Marsden Jacob Associates. (2005). *Frameworks for economic impact analysis and benefit-cost analysis*. Retrieved March 20, 2006.

MapInfo [Computer software]. (n.d.). Troy, NY: MapInfo Corporation. Retrieved from <http://www.mapinfo.com>

Mine Safety and Health Administration. (n.d.). *Coal fatalities from 1900 through 2005*. Retrieved March 23, 2006, from <http://www.msha.gov/stats/centurystats/coalstats.htm>

Nas, T. F. (1996). *Cost-benefit analysis*. London: Sage.

NCSS [Computer software]. (n.d.). Kaysville, UT: Number Cruncher Statistical Systems. Retrieved from <http://www.ncss.com>

Orford, S., Harris, R., & Dorling, D. (1999). Geography information visualization in the social sciences: A state of the art review. *Social Science Computer Review*, 17, 289-304.

Panayotou, T. (n.d.). *Basic concepts and common valuation errors in cost-benefit analysis*. Cambridge, MA: Harvard University, Institute for International Development. Retrieved August 14, 2001, from <http://www.eepsea.org/publications/specialp2/ACF2DB.html>

Permaloff, A., & Grafton, C. (1995). *Political power in Alabama: The more things change*. Athens, GA: University of Georgia Press.

Permaloff, A., & Grafton, C. (1996). Computer tools for crafting clear and attractive diagrams in social science. *Social Science Computer Review*, 14, 293-304.

People for Modern Transit. (n.d.). *Break-even on operations: A myth*. Retrieved April 24, 2006, from <http://www.peopleformoderntransit.org/pmtadmin/home.nsf/e93bad-dc3ba8bf0a88256d42000f7a73/58502>

Practical PC. (n.d.). *CD-R media lifespan*. Retrieved March 26, 2006, from <http://www.practicalpc.co.uk/computing/storage/cdr Lifespan.htm>

Primavera Project Planner (P3) [Computer software]. (n.d.). Bala Cynwyd, PA: Primavera Systems, Inc. Retrieved from <http://www.primavera.com>

RFFlow [Computer software]. (n.d.). Loveland, CO: RFF Electronics. Retrieved from <http://www.rff.com>

Rosenbloom, D. H., & Kravchuk, R. S. (2002). *Public administration*. Boston: McGraw Hill.

Rotenberg, M., & Hoofnagle, J. (2003). *Memo to representatives Adam Putnam and William Clay, House Government Reform Subcommittee on*

- Technology, Information Policy, Intergovernmental Relations, and the Census, March 25, 2003.* Retrieved March 10, 2006, from <http://www.epic.org/privacy/profiling/datamining3.25.03.html>
- Schmalz, M. (2005). *Integrating Excel and Access*. Cambridge, UK: O'Reilly Media.
- Schmid, A. A. (1989). *Benefit-cost analysis*. Boulder, CO: Westview.
- Schroeder, L. D., Sjoquist, D. L., & Stephan, P. E. (1986). *Understanding regression analysis: An introductory guide*. Beverly Hills, CA: Sage.
- Slack, B. (n.d.). *Cost/benefit analysis in practice*. Retrieved March 20, 2006, from <http://people.hofstra.edu/geotrans/eng/ch9en/appl9en/ch9a3en.html>
- Slaton, C. D., & Arthur, J. L. (2004). Public administration for a democratic society: Instilling public trust through greater collaboration with citizens. In M. Malkia, A. Anttiroiko, & R. Savolainen (Eds.), *eTransformation in governance* (pp. 110-130). Hershey, PA: Idea Group.
- SPSS. (n.d.). Chicago: Author. Retrieved from <http://www.spss.com>
- Statsoft. (2003). *Data mining techniques*. Retrieved March 27, 2006, from <http://www.statsoft.com/textbook/stdatmin.html>
- Stokey, E., & Zeckhauser, R. (1978). *A primer for policy analysis*. New York: W. W. Norton & Company.
- SYSTAT [Computer software]. (n.d.). Richmond, CA: Systat Software, Inc. Retrieved from <http://www.systat.com>
- Tindall, C. (2005). *The monorail episode redux*. Retrieved March 20, 2006, from http://uchicago.law.typepad.com/faculty/2005/11/the_monorail_ep.html
- Vann, I. B., & Garson, G. D. (2001). Crime mapping and its extension to social science analysis. *Social Science Computer Review*, 19, 417-479.
- Vann, I. B., & Garson, G. D. (2003). *Crime mapping: New tools for law enforcement*. New York: Peter Lang.
- Weida, N. C., Richardson, R., & Vazsonyi, A. (2001). *Operations analysis using Microsoft Excel*. Pacifica Grove, CA: Duxbury.
- Welch, S., & Comer, J. (1988). *Quantitative methods for public administration: Techniques and applications*. Pacific Grove, CA: Brooks/Cole.
- Wilkinson, L., Blank, G., & Gruber, C. (1996). *Desktop data analysis with SYSTAT*. Upper Saddle River, NJ: Prentice-Hall.

This work was previously published in Modern Public Information Technology Systems: Issues and Challenges, edited by G. Garson, pp. 239-264, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).