This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

CHAPTER 10

SPREADSHEET ESSENTIALS

Introduction

Having strong spreadsheet skills is essential for public managers and policy analysts. Spreadsheet programs (such as Microsoft Excel) are not only popular, they are very practical tools with a wide range of applications. This chapter provides an overview of some important uses for spreadsheets in public management. Mostly, spreadsheets involve quantitative, nonstatistical applications (that is, they involve data but not hypothesis testing and associated analyses). Spreadsheets are quite commonly used for making and analyzing budgets, for scenario building and "what-if" analyses, and for developing graphs and charts for presentations. Although spreadsheet applications are often as unique as the contexts in which they are used, the principles and problem-solving methodologies often are readily grasped and easily transferred to new situations.

Sometimes, people who are unfamiliar with statistical analysis wonder why spreadsheets cannot be used for both statistical and nonstatistical analysis. To use a spreadsheet for statistical analysis would be like using a screwdriver to drive a nail into a wall; it can be done, but it is very clumsy, tedious, and prone to error. Although some overlap exists between spreadsheet and statistical analysis programs, the two are actually quite different. Spreadsheets are designed for generating graphs and for making "what-if" calculations. Although spreadsheets do offer some statistical functions, they are not designed to perform complicated statistical analysis (such as Kendall's tau-c or many regression diagnostics—even calculating chi-square requires some work). Spreadsheets are also limited in analyzing subsets of data or in making variable transformations. Often, these data abilities are available in only rudimentary ways. By contrast, statistical programs are designed exactly to perform these kinds of tasks, but they are not very good at what-if analysis or for presenting budget data; using statistical software for these functions is sometimes like using a hammer to drive in a screw. Good analysts appreciate both the spreadsheets and the statistical programs for their specialized purposes in public management. Having both in your toolbox only enhances your abilities.

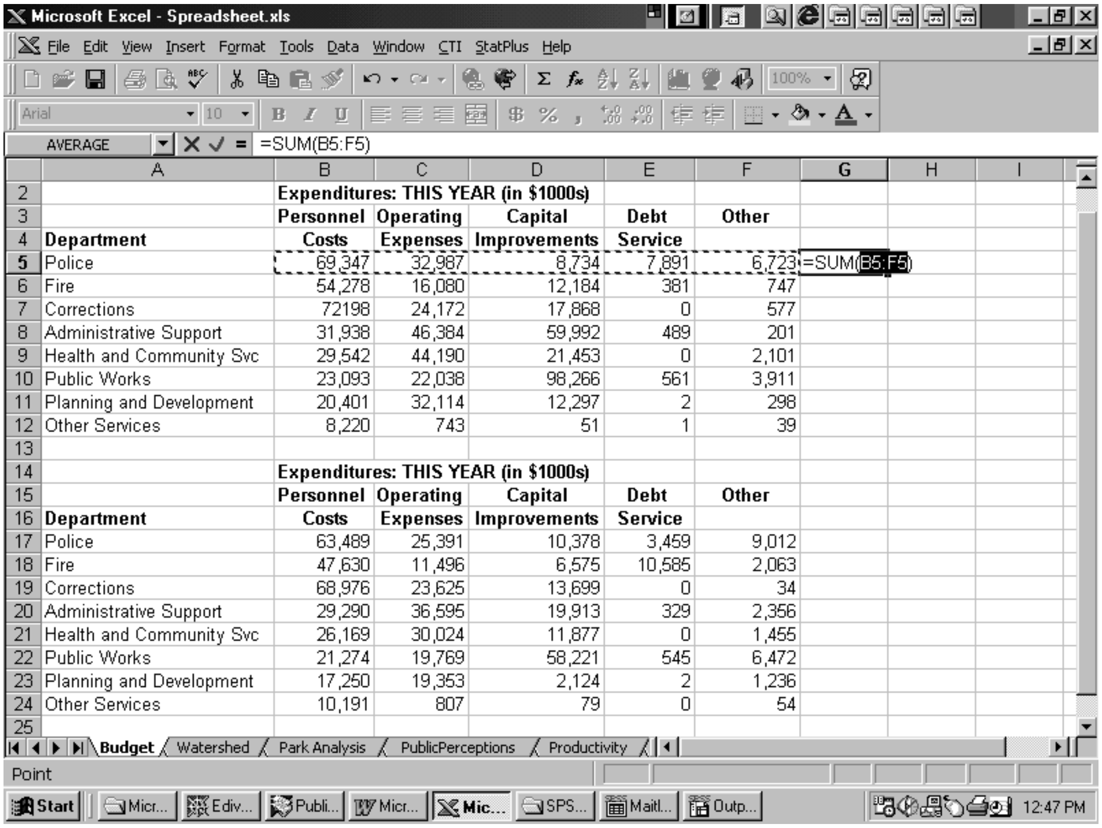
You will be working with a sample spreadsheet file called, conveniently enough, "Spreadsheet.xls." This file contains eleven separate sheets that you will be using as you progress through this guide. These eleven sheets have also been saved as separate files on the accompanying CD-ROM. Each sheet contains data that can be used for different kinds of statistical analysis. If you are uncertain or unfamiliar with any of the functions being performed in a particular sheet, remember to click on individual cells to see the formulas themselves, and be sure to refer to this workbook as you work through the various sheets. The screen shots in this guide will be valuable reference points to make sure that your results are accurate.

In order to get the most from this guide, you must have a working familiarity with spreadsheets. The examples in this guide are based on Microsoft Excel, a widely used spreadsheet program, but any spreadsheet software can be used. (*Note:* The SPSS Data Editor is similar to a spreadsheet; students familiar with SPSS will find it quite easy to obtain such working knowledge.)

How much familiarity is required? Here is a short test to assess your comfort level with spreadsheets. Open the file "Spreadsheet.xls," which has several sheets, shown by the tabs in the lower left corner of the screen. Open the sheet called "Budget." If this sheet does not appear as an option, click on the arrows in the lower left part of your screen until this name does appear. Familiarize yourself with the spreadsheet. Look at the format, the font sizes, and alignment as shown in Screen W10.1. If you are able to design a table, enter the data, change font sizes and alignment, and create the basic "=SUM" function to total the figures in row 5, then you are familiar enough to move on.

This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Screen W10.1 Spreadsheet.xls



	A	B	C	D	E	F	G
2		Expenditures: THIS YEAR (in $1000s)					
3		Personnel	Operating	Capital	Debt	Other	
4	Department	Costs	Expenses	Improvements	Service		
5	Police	69,347	32,987	8,734	7,891	6,723	=SUM(B5:F5)
6	Fire	54,278	16,080	12,184	381	747	
7	Corrections	72198	24,172	17,868	0	577	
8	Administrative Support	31,938	46,384	59,992	489	201	
9	Health and Community Svc	29,542	44,190	21,453	0	2,101	
10	Public Works	23,093	22,038	98,266	561	3,911	
11	Planning and Development	20,401	32,114	12,297	2	298	
12	Other Services	8,220	743	51	1	39	
13							
14		Expenditures: THIS YEAR (in $1000s)					
15		Personnel	Operating	Capital	Debt	Other	
16	Department	Costs	Expenses	Improvements	Service		
17	Police	63,489	25,391	10,378	3,459	9,012	
18	Fire	47,630	11,496	6,575	10,585	2,063	
19	Corrections	68,976	23,625	13,699	0	34	
20	Administrative Support	29,290	36,595	19,913	329	2,356	
21	Health and Community Svc	26,169	30,024	11,877	0	1,455	
22	Public Works	21,274	19,769	58,221	545	6,472	
23	Planning and Development	17,250	19,353	2,124	2	1,236	
24	Other Services	10,191	807	79	0	54	
25							

If you feel uncomfortable using Excel or if the "=SUM" function seems unclear to you, please take a look at an introductory guide to the software itself. The instructions that come with the software, or that can be purchased separately in most bookstores, will do a great job of explaining the basics. Spreadsheets are very straightforward. You should be up and running in no time.

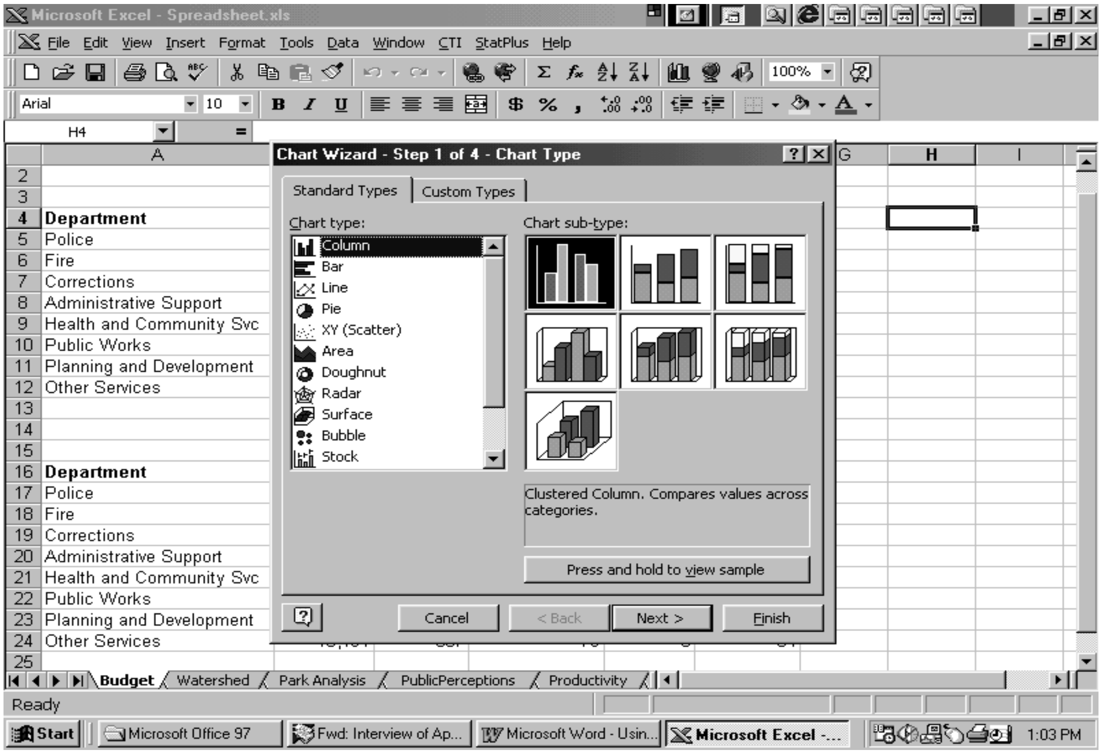
Presenting Your Results

One of the most attractive features of spreadsheet software is its ability to create graphic elements that quickly summarize essential points and findings. Creating these graphs or charts is often the last step in analysis, but it is often the first thing to learn when becoming more familiar with the software's capabilities. Continue on with the spreadsheet called "Budget." If you want to show how personnel costs vary across departments in a year's budget, you can use Excel to produce line graphs, bar charts (histograms), pie charts, scatterplots, and many other useful visual aids. To create these, use the "Chart Wizard," which is obtained by selecting Insert → Chart. The Chart Wizard should appear, as shown in Screen W10.2. Select the type of chart you want, and click on Next. Then, simply walk through the rest of the clearly indicated steps. Please note the importance of correctly specifying the data range, as well as the various options that exist for titles, axes, legends, and data labels. Feel free to play around with the Chart Wizard and its various features to get a good sense of the tools at hand.

After you have produced some graphs and charts, you should edit them because some features in Excel are not part of the Chart Wizard. For example, to change the background color of your chart, you need to place the cursor over that area and then click the right button on your mouse (or select Format → Selected Area). This will bring up a dialogue box called Format Plot Area. If you want to format the bars, place the cursor over the bars, and then right-click the mouse to bring up the dialogue box Format Data Series. Although some of these features are available in other statistical analysis programs, Excel also offers a broad range of options for enhancing the appearance of charts and graphs.

This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Screen W10.2 —— Excel Chart Wizard

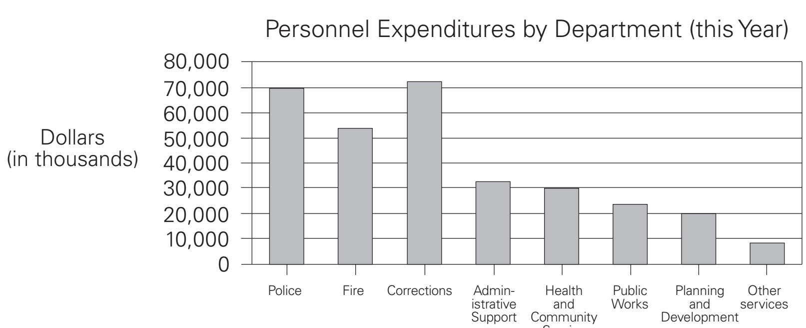


Another added benefit of Excel is the ease of transferring your results to a word processing document. You can copy and paste your graphs directly into a document. Figure W10.1 is a good example of a basic bar chart that can be incorporated into a handout or written presentation. It was copied and pasted from Excel into this document. You should try to reproduce this figure using the "Budget" spreadsheet.

Of course, you already know that many tables have column and row totals, which are created using the =SUM function, and that they are also used in presentations. By way of practice, try modifying the "Budget" spreadsheet so that it looks like Screen W10.3. (For your convenience, this is saved on the "Budget" sheet—scroll down to row 40 to compare your work.)

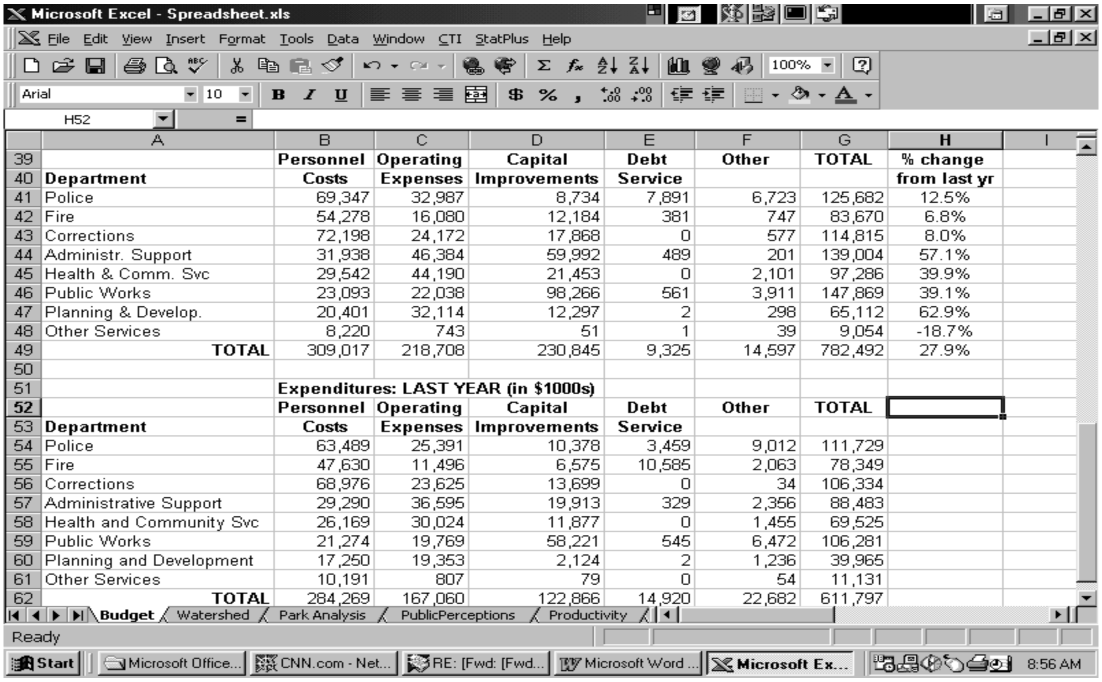
Experiment some more with the different options in presenting your data. For example, change the value of cell B5, and see how your graph changes when you alter the data. Another good exercise is to plot the budget differences between this year and last year, using data on this spreadsheet. You might also explore some of the editing features such as background colors, font sizes, text direction, and so on. Don't be surprised if such practicing turns out to be quite time intensive as you try to perfect all of the various options.

Figure W10.1 —— "Budget" Sample Bar Graph



This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Screen W10.3 “Budget” Spreadsheet Totals



	A	B	C	D	E	F	G	H
39		Personnel	Operating	Capital	Debt	Other	TOTAL	% change
40	Department	Costs	Expenses	Improvements	Service			from last yr
41	Police	69,347	32,987	8,734	7,891	6,723	125,682	12.5%
42	Fire	54,278	16,080	12,184	381	747	83,670	6.8%
43	Corrections	72,198	24,172	17,868	0	577	114,815	8.0%
44	Administr. Support	31,938	46,384	59,992	489	201	139,004	57.1%
45	Health & Comm. Svc	29,542	44,190	21,453	0	2,101	97,286	39.9%
46	Public Works	23,093	22,038	98,266	561	3,911	147,869	39.1%
47	Planning & Develop.	20,401	32,114	12,297	2	298	65,112	62.9%
48	Other Services	8,220	743	51	1	39	9,054	-18.7%
49	TOTAL	309,017	218,708	230,845	9,325	14,597	782,492	27.9%
50								
51		Expenditures: LAST YEAR (in $1000s)						
52		Personnel	Operating	Capital	Debt	Other	TOTAL	
53	Department	Costs	Expenses	Improvements	Service			
54	Police	63,489	25,391	10,378	3,459	9,012	111,729	
55	Fire	47,630	11,496	6,575	10,585	2,063	78,349	
56	Corrections	68,976	23,625	13,699	0	34	106,334	
57	Administrative Support	29,290	36,595	19,913	329	2,356	88,483	
58	Health and Community Svc	26,169	30,024	11,877	0	1,455	69,525	
59	Public Works	21,274	19,769	58,221	545	6,472	106,281	
60	Planning and Development	17,250	19,353	2,124	2	1,236	39,965	
61	Other Services	10,191	807	79	0	54	11,131	
62	TOTAL	284,269	167,060	122,866	14,920	22,682	611,797	

Essential Functions

An important feature of spreadsheets is their ability to analyze data. These very basic functions will help give you a strong foothold before we move on to some more specific applications in later sections. In the preceding section, you used the =SUM function to add up rows and columns. In Excel, other statistical functions are displayed when you select f_x on the standard toolbar (or choose Insert → Function). The dialog box Paste Function appears when you select f_x; it defines each function and shows you how to use it. Take some time to identify functions that you think you will use often in your work.

In most applications =SUM (range) and =AVERAGE (range) are probably most widely used. Just to be clear, the =AVERAGE function calculates the mean. To produce the median, you need to type =MEDIAN. When using the =AVERAGE function, note that =AVERAGE (A1: A4) produces the same result as +(A1+A2+A3+A4)/4 only when each cell has a valid entry. If you use =AVERAGE (A1: A4), cells that have text or missing entries will be ignored and only the valid entries will be used to calculate the mean. By contrast, +(A1+A2+A3+A4)/4 returns “#VALUE!” when any of the indicated cells are invalid.

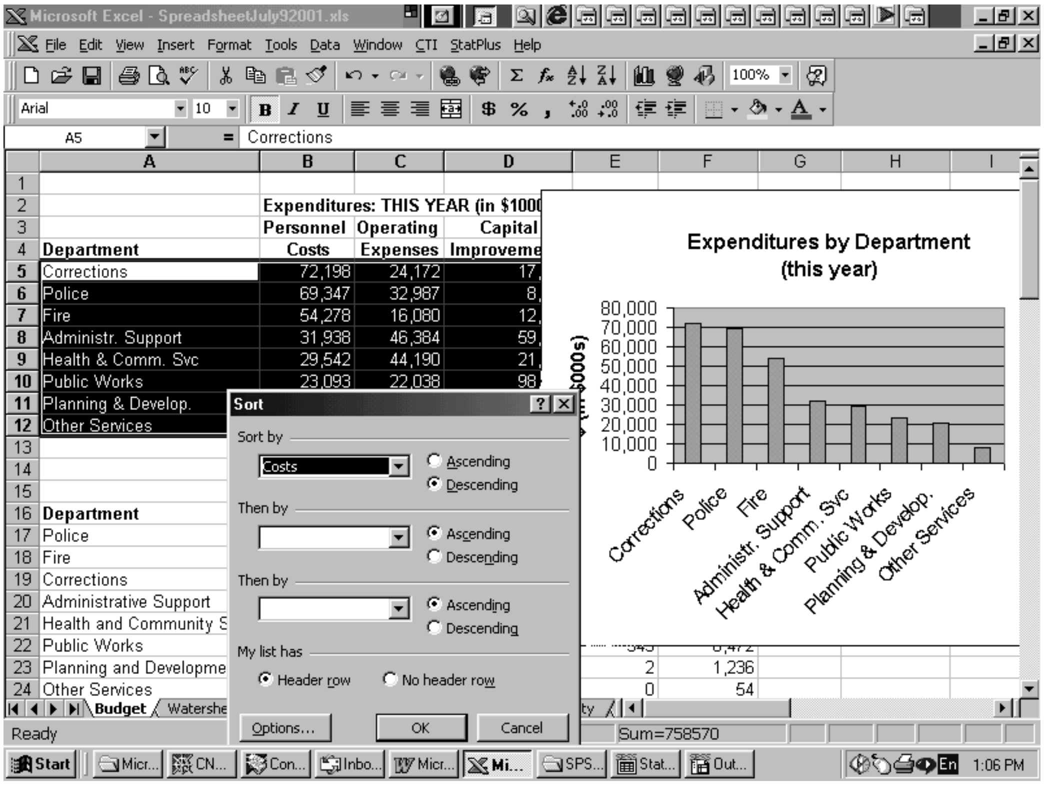
Data Sorting

Another useful feature is data sorting, which is used when an analyst wants to change the order of data presentation. For example, previously you produced a bar chart (Figure W10.1), in which the order of personnel expenditures appeared in the same order as presented on the spreadsheet. But what if you want to show the Corrections’ personnel expenditures before (to the left of) Police personnel expenditures, which would place the bars in descending order? You must then change the order in which the data appear in the spreadsheet; specifically, the data must be sorted in descending order of personnel expenditures.

First, select Data → Sort. Simply put the cursor somewhere in the data table to tell Excel what area you want to sort, or you can specify the exact area by highlighting it. Then identify in the Sort dialogue box (shown in Screen W10.4) whether the highlighted area includes a header row, which is used to identify the different columns and is not part of the data that are sorted. Observe that data can be sorted by more than one variable. After you have sorted the data, you can use the Chart Wizard to produce the new chart. Screen W10.4 shows the result; note that expenditures are sorted and graphed in descending order. Data sorting comes in very handy when dealing with very large data sets. To open a typical, larger data set, click on the Census tab, which contains information about all 398 Florida municipalities. Being able to sort data in

This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Screen W10.4 —— Sort Dialogue Box



ascending or descending order can help an analyst answer such questions as how many cities had populations over 20,000 in the year 2000? (The answer is 90 of the 404 cities.) Or, how many cities lost populations between 1990 and 2000? (Answer, 79.) Answering the latter question requires careful definition of the data ranges to be sorted because the formulas for calculating the change in population extend to an area of the spreadsheet (column G) that is separated by a blank column (column F). You need to identify the entire range, from A8 through G415. Another useful feature for counting is =COUNT (range), which counts the number of observations in a specified range.

When sorting data, there will be times when you are concerned about returning your data to their original format. An easy solution is to *fill* a column with a numeric sequence such as "1,2,3,4 . . ." that is later used to re-sort the data in their original order. To create such a series, highlight a space within a column adjacent to the data that are to be sorted (for example, insert a new column in column A). Then, enter "1" in the first row of the column (shown in Screen W10.5), highlight the appropriate range that is to be "filled" in), and select Edit → Fill → Series. This will create a series that begins with 1 in cell A8 and ends with 408 in cell A12.

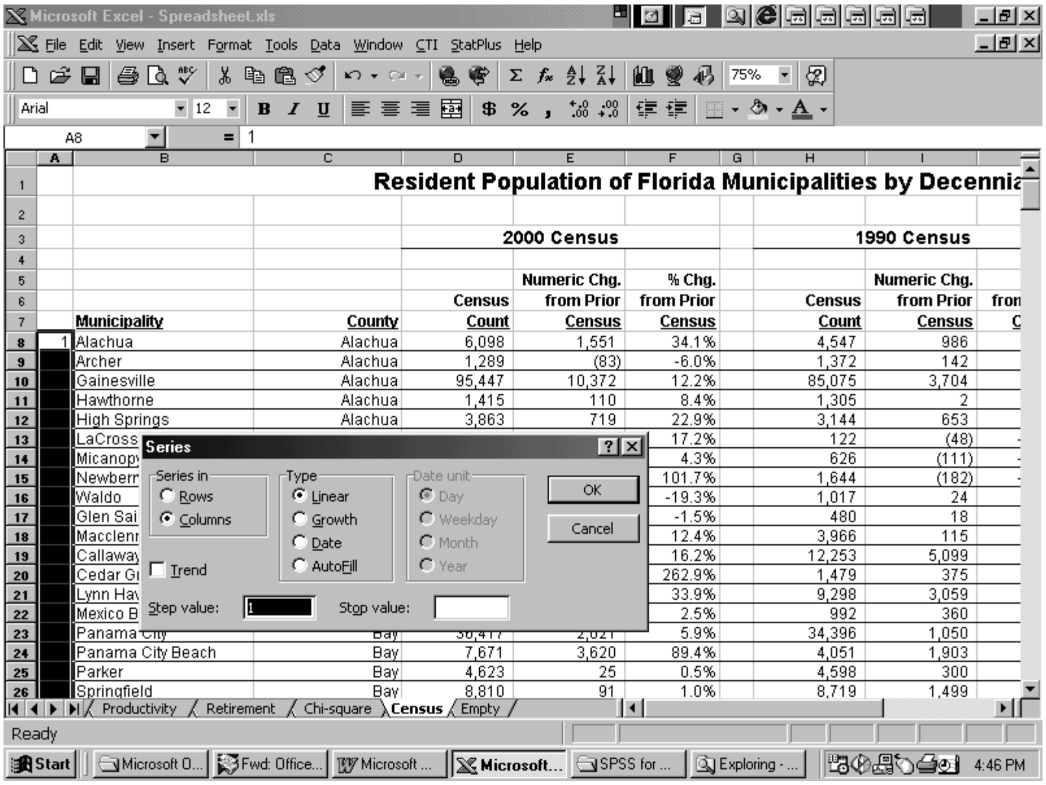
Filtering Data

The filter feature allows analysts to select a particular subset of observations for analysis. If you want to analyze only the ten largest cities in Florida, then you would choose to filter your data. Using the "Census" sheet, place the cursor over the column "Census Count" and select Data → Filter → Auto Filter. New drop-down arrows will appear in each of the columns, as shown in Screen W10.6. Choose the arrow in the "Census" column (column D), and select Top 10. Note that after you select this option, only ten rows of data will appear; the blue filter in the "Census" column indicates that it is now active. You can use just these ten cities to create a chart (Insert → Chart). Or, you can print the spreadsheet, and only these observations will appear.

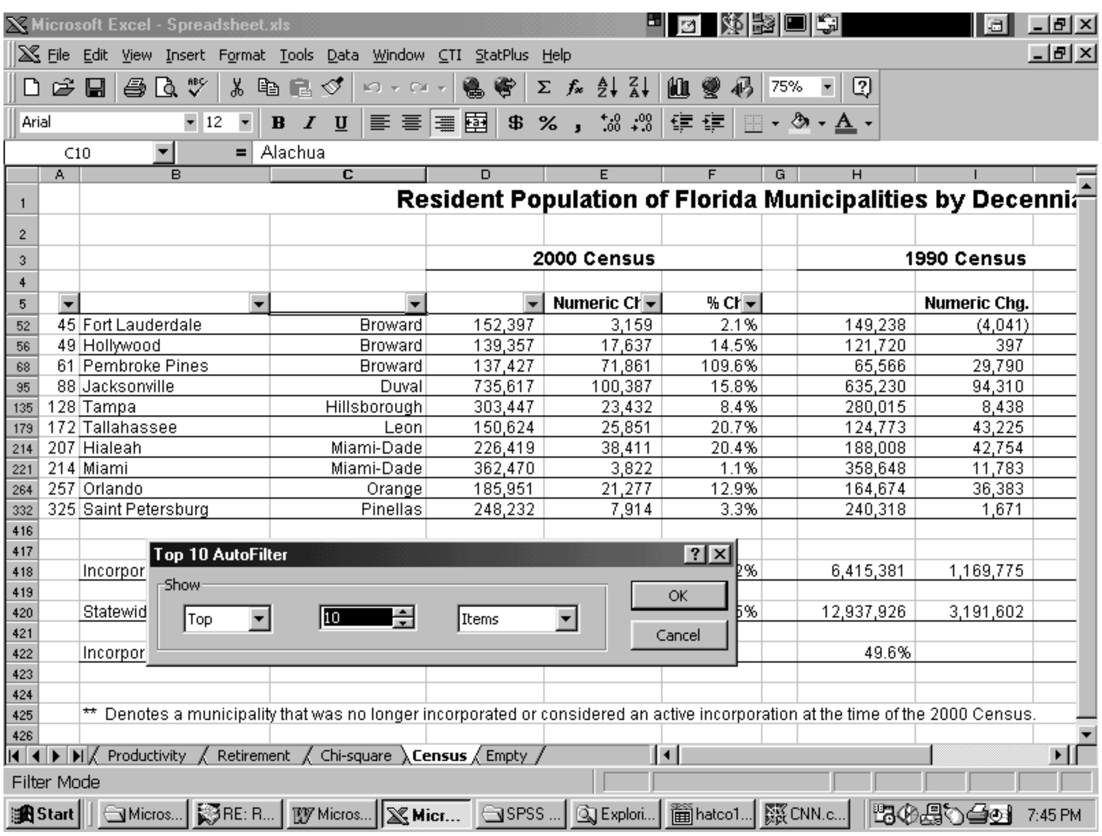
Please note that the Excel filter does have a limitation. If you select a function such as =AVERAGE and attempt to identify these observations as a range, Excel will identify *all* observations that lie between the lowest and highest values of these cities, or D52:D332. Thus, the mean will be calculated for many more

This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Screen W10.5 Series Dialogue Box

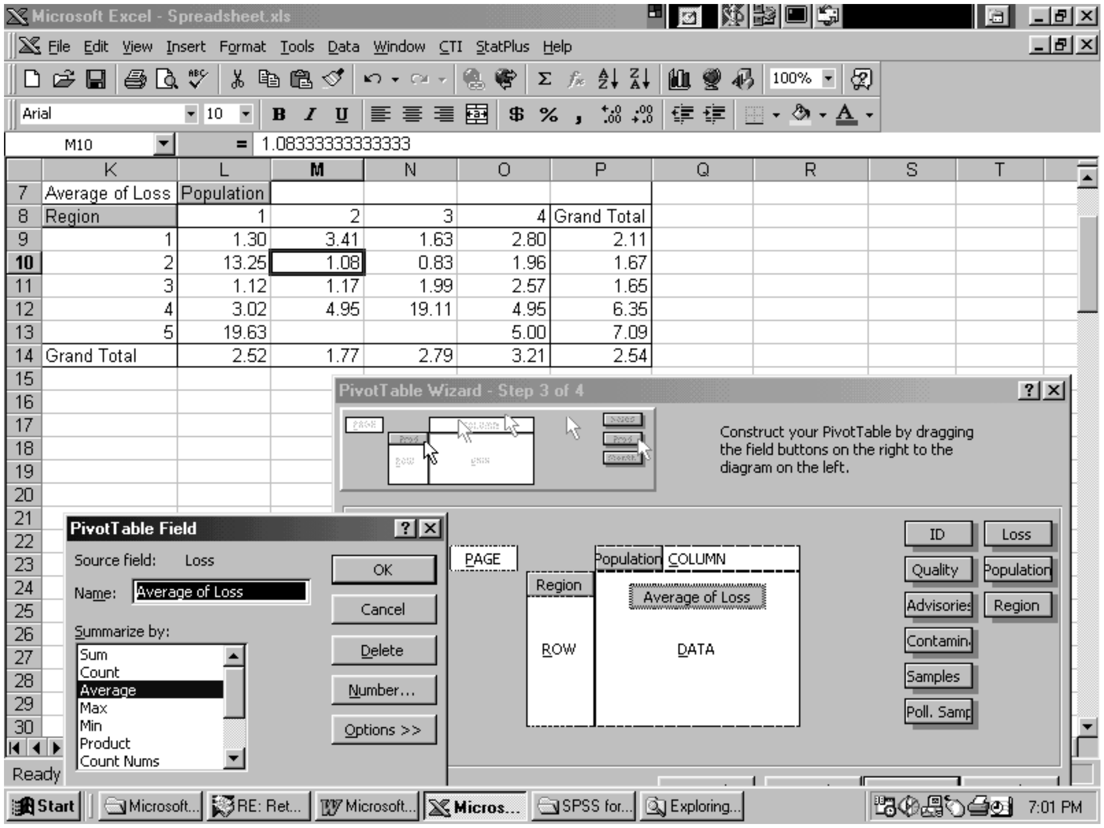


Screen W10.6 Data Filtering



This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Screen W10.7 —— Creating a Pivot Table



than just these ten cities. Instead, analysts will need to specify each of these observations *separately.* That is, =AVERAGE (D52, D56, D68, D95, D135, D179, D214, D221, D264, D332), which is quite cumbersome. To overcome this problem, analysts might take a different approach. They might first sort the data in descending order and then identify the range of the first ten cities (or a statistical analysis program might be used).

Creating Tables from Raw Data

Finally, Excel can also make tables from raw data—another great capability of spreadsheets. Open the sheet "Watershed." The data include various measures of pollution, along with demographic descriptors, such as region. (The variables are selected from the SPSS data set with the identical name, and we refer to the documentation about data sets that is included in Chapter 11, "Data Set Documentation.") To make a summary table (called a Pivot Table in Excel), choose Data → Pivot Table Report. Screen W10.7 shows the result of this analysis below. The Pivot table shows the average loss (in percentages) of watershed wetlands by region and by population size (coded in categories). You might want to create some other Pivot tables from these data to see how other functions appear. Please note that although the default summary function is =SUM, double-clicking variables in the Data field brings up the dialogue box Pivot Table Field where this function can be changed. If you want to show both the counts and the mean of certain variables, you can drag the same variable several times from the right side of the Pivot Table Wizard dialogue box to the Data, Column, or Row fields. Then change the statistical function. (It is worth noting that the manner in which columns and rows are changed in Pivot Tables is similar to that found in SPSS when using Analyze → OLAP cubes and then, in the Output Viewer, double clicking the output table and selecting Pivot → Pivoting Trays.)

DATA ANALYSIS

Other spreadsheet applications are of special interest to public managers and analysts because they can perform valuable data analysis. These applications are quick, easy, and produce attractive results that are easily reproduced for presentations and reports.

This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

What-If Analysis

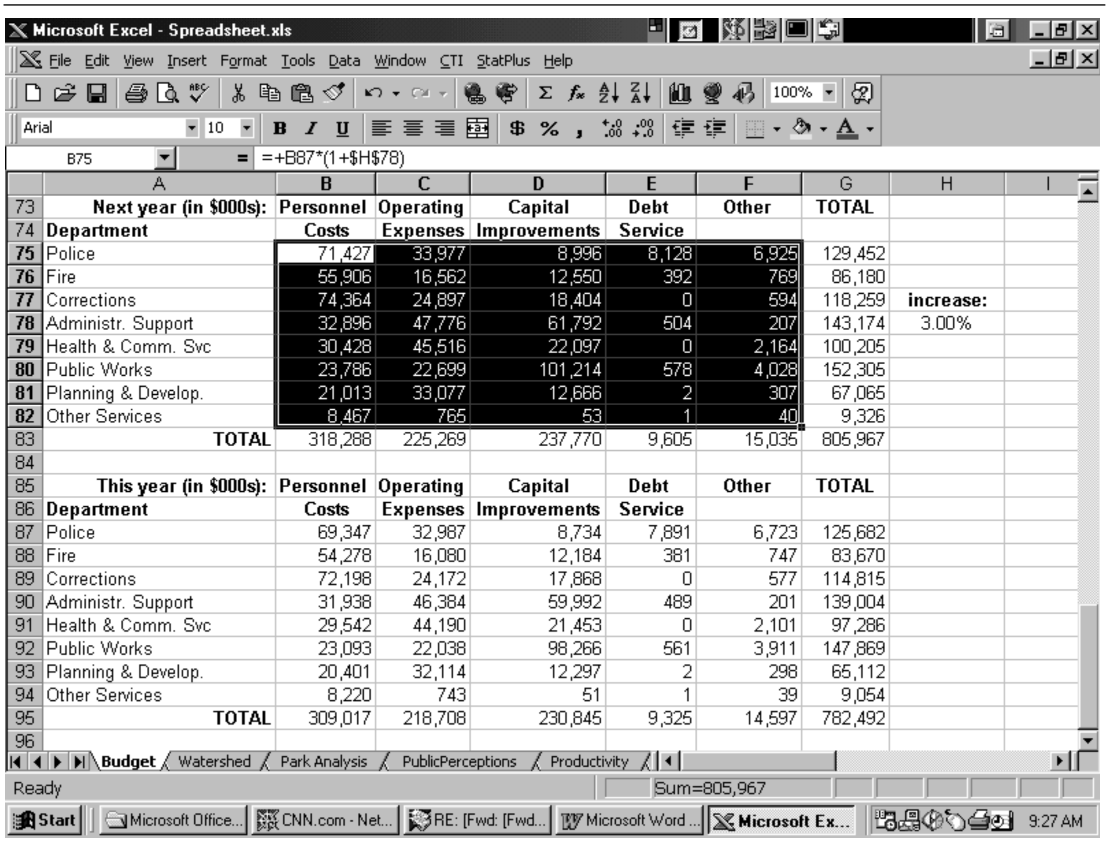
What-if analyses are sequentially performed analyses whereby analysts change one or more parameters in order to see the effects on another variable of interest. For example, you might change population growth in a revenue or pollution-forecasting scenario, in order to see how that affects these variables. By sequentially modifying the same parameter (for example, examining the effects of population decline and growth of varying degrees), what-if analyses can give a sense of the likely range of program or policy outcomes. Although what-if analyses tend to be as specific as the problems they address, a few examples should help illustrate the general idea. Spreadsheets are especially good tools for doing such analyses.

Open the “Budget” sheet again. Scroll down to row 73 so that your computer screen resembles Screen W10.8. Here, the spreadsheet shows how a policy maker could figure out the effect on next year’s budget of increasing all expenditures by 3 percent. Take a look at Screen W10.8, and note the formula in cell B75. It includes a reference to cell H78, which identifies the 3 percent increase. The reference is absolute, $H$78, which allows us to copy the formula to the range of all of next year’s expenditures, B75:F82, while maintaining the same reference to cell H78. Then, next year’s budget will be $805,967,000, as shown in cell G83.

The ease of using spreadsheets to do a what-if calculation lies not in just a single calculation (indeed, the above result can be obtained easily by hand) but rather in the ability to quickly do many calculations. Indeed, you might ask, what if all expenditures are allowed to increase only 2.5, 2.8, 3.2, 4.5, or (a whopping) 6 percent? Then, you need only change the value in cell H78, and the value is automatically shown in cell G83. Furthermore, next year’s budget is shown for all expenditure categories and departments, which could be graphed as well. Using this spreadsheet can save time by quickly doing many calculations otherwise done by hand.

Moreover, what if you want to calculate next year’s budget when personnel expenditures grow 3 percent (except in the Corrections Department, where they stay at last year’s level), all operating expenses are cut by 2 percent, and all other expenses remain at last year’s level? Spreadsheet formulas easily allow for more complexity. Here, you simply need to modify the formulas in each group of cells to reflect this scenario. Be sure to use separate values to reflect the personnel and operating expenses. Try modifying the above spreadsheet

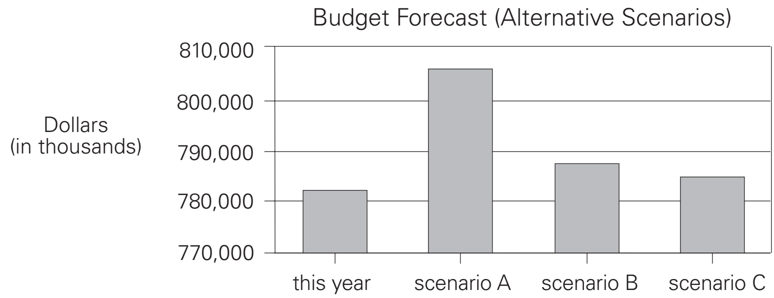
Screen W10.8 What-If Analysis



	A	B	C	D	E	F	G	H
73	Next year (in $000s):	Personnel	Operating	Capital	Debt	Other	TOTAL	
74	Department	Costs	Expenses	Improvements	Service			
75	Police	71,427	33,977	8,996	8,128	6,925	129,452	
76	Fire	55,906	16,562	12,550	392	769	86,180	
77	Corrections	74,364	24,897	18,404	0	594	118,259	increase:
78	Administr. Support	32,896	47,776	61,792	504	207	143,174	3.00%
79	Health & Comm. Svc	30,428	45,516	22,097	0	2,164	100,205	
80	Public Works	23,786	22,699	101,214	578	4,028	152,305	
81	Planning & Develop.	21,013	33,077	12,666	2	307	67,065	
82	Other Services	8,467	765	53	1	40	9,326	
83	TOTAL	318,288	225,269	237,770	9,605	15,035	805,967	
84								
85	This year (in $000s):	Personnel	Operating	Capital	Debt	Other	TOTAL	
86	Department	Costs	Expenses	Improvements	Service			
87	Police	69,347	32,987	8,734	7,891	6,723	125,682	
88	Fire	54,278	16,080	12,184	381	747	83,670	
89	Corrections	72,198	24,172	17,868	0	577	114,815	
90	Administr. Support	31,938	46,384	59,992	489	201	139,004	
91	Health & Comm. Svc	29,542	44,190	21,453	0	2,101	97,286	
92	Public Works	23,093	22,038	98,266	561	3,911	147,869	
93	Planning & Develop.	20,401	32,114	12,297	2	298	65,112	
94	Other Services	8,220	743	51	1	39	9,054	
95	TOTAL	309,017	218,708	230,845	9,325	14,597	782,492	
96								

This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Figure W10.2 Alternative Budget Forecast



in this manner. The final result should be $785,222. To further increase your familiarity, try a few different scenarios, and plot your results on a bar chart such as might be shown in a briefing to decision makers. A hypothetical example is shown in Figure W10.2 (of course, in an actual write-up we would define and explain each scenario).

What-if analyses tend to be quite specific and have much practical application in your own day-to-day life. Open the sheet "Retirement," which shows a simple formula to project your future retirement income. (A little disclaimer: This is for learning purposes only; results or accuracy are not guaranteed!) Although you may not want to think about retirement right now, government leaders and financial planners agree that you should, and the earlier the better. Why is that so? Do your own what-if analyses in which you vary your "years to retirement." Then, change your annual contributions. Can you find a balance between these two that is consistent with your lifestyle and expectations? See how much money you need to save in order to maintain your lifestyle after retirement.

Many situations are appropriate for what-if analyses: calculations to determine whether leasing or buying is better (under different assumptions of depreciation, interest, and inflation), revenue forecasting (under different assumptions about economic growth), service demand (under different assumptions about target population). Wherever there is a spreadsheet with data (that is more than just a listing of names, addresses, and the like), the possibility of doing a what-if analysis is not far away. Can you think of some situations in your current or future work in which what-if analyses might be useful?

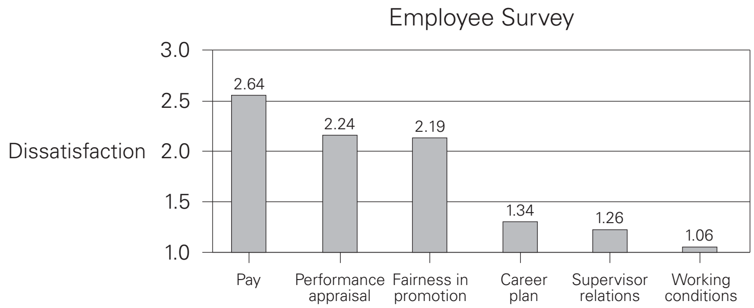
Pareto Charts

You may have participated in brainstorming meetings in which a group attempted to resolve some problem. Typically, a good deal of the discussion focuses on the root sources of the problem. Some time ago, the Italian economist Vilfredo Pareto (1848–1923) observed that 20 percent of the population have 80 percent of the income; decisions that affect the accumulation of wealth can be targeted at the few. Likewise, a few root sources usually account for almost any problem. A Pareto chart is simply a histogram, in which analysts show the importance of different root sources as they contribute to the problem. The idea is to channel energy to those few sources that cause most of the problems, rather than try to solve the myriad number of problems caused by these few root causes.

Service logs, maintenance records, and customer complaints are great sources of data for Pareto charts. Employee and citizen surveys also can be used. Assume that an agency wishes to examine the sources of employee dissatisfaction. Convening groups of staff and managers together in focus groups is apt to provide many possible problems and solutions, but which ones are likely to have the greatest impact? Which problems are most widely and severely experienced? Figure W10.3, based on an actual employee survey, is a Pareto chart because it shows, in descending order, the order of importance of various causes (in this example, factors associated with employee dissatisfaction). Chart Wizard creates the chart after the mean of each response has been determined. Further analysis could examine this result by rank, department, gender, and race to provide additional guidance. Again, the ordering of the importance of a problem's sources helps to move dialogue and decision making forward.

This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Figure W10.3 Pareto Chart



Demand Analysis

Demand analysis often is a very useful productivity improvement tool that can be used to analyze fluctuations in demand. For example, some agencies experience increased demand at some times of the day, weeks of the month, or months of the year. Increased demand not only causes great stress for employees, it also causes increased backlogs and, hence, waiting and complaints among agency clients. At other times, agencies experience very low service demand. Although some "down time" is then used for clearing the backlog and planning for the future, too much downtime is obviously unproductive.

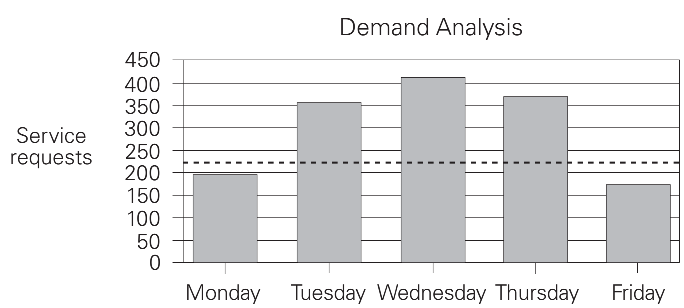
The idea of demand analysis is to align staffing with demand. When service logs are available, an accurate analysis can be made of the average demand. For example, consider the following hypothetical demand for services, which shows the average, daily service demand based on an analysis of the preceding twelve weeks. Figure W10.4 is created by first using Chart Wizard to create the bar chart from available service request data. The figure clearly shows that demand on Tuesdays, Wednesdays, and Thursdays is greater than on other days of the week. Added after the chart has been cut and pasted into a word processing document, the dotted line (---------) shows the normal service load for current staffing levels. (In Microsoft Word, the dotted line is added using the Drawing Tool Bar.) Although staff should be able to handle some above-normal service request loads (for example, during peak periods), the number of requests on Wednesday is considerable. If this unevenness causes problems (such as service that does not meet client expectations), the agency might consider hiring some additional part-time staff for Tuesday, Wednesday, and Thursday. Indeed, it might even consider filling a full-time employee vacancy with two or more part-time employees.

Note that demand analysis charts can analyze service by any time period, such as by the hour, day, week, month, or quarter. For example, a demand analysis chart that analyzes service demand per hour might find that demand in the morning differs from that in the later part of the afternoon, hence causing managers to increase the use of flextime policies.

Allocation Decisions

"Who gets what and how?" is a key research question in the study of public policy and politics. Sometimes, analysts are called upon to help managers and elected officials in making their allocation decisions. For

Figure W10.4 Demand Analysis



This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Screen W10.9 "Allocation" Spreadsheet

			% of budget affected by prioritization:			20%		
					% overall increase:	3%		
Program	Total Points	Priority (%)	Last Year Budget	Last Year Budget plus increase	Share of Budget Unaffected by Ranking	Allocation of Budget affected by Ranking	Total New Budget	Change from last year
1	17	16.2%	3,256	3,354	2,683	336	3,019	-7.3%
2	32	30.5%	2,345	2,415	1,932	632	2,565	9.4%
3	6	5.7%	1,456	1,500	1,200	119	1,318	-9.5%
4	15	14.3%	2,345	2,415	1,932	296	2,229	-5.0%
5	25	23.8%	213	219	176	494	669	214.3%
6	10	9.5%	456	470	376	198	573	25.7%
Total:	105	100%	10,071	10,373	8,299	2,075	10,373	3.0%
1	17	+B18/$B$15	3,256	+D18*(1+$G$3)	+E18*(1-$G$2)	+($E$15-$F$15)*C18	+F18+G18	+(H18-D18)/D18

example, an agency might want to allocate program funds based on a mix of current service demand and population size. Or, in a workshop setting, a city council might ask each of its members to rank different programs according to their priorities. Analysts are then called upon to determine the aggregate ranking and to analyze its impact on funding. They may even show how different criteria and parameters affect overall results.

Open the spreadsheet "Allocation," and look the data over carefully. Assume the following scenario. Five councilpersons are asked to prioritize six existing programs on a scale of 1 to 6 (1 = lowest priority, 6 = highest priority). Assume also that the ranking will affect only 20 percent of available funding; the remaining 80 percent is based on last year's expenditures. Further, the overall budget is increased by 3 percent over the previous year. By ranking each of these six programs, the councilpersons will be able to determine who receives the funding. The spreadsheet "Allocation" (Screen W10.9) shows the analysis used to rank these programs. Note that the total sum of rankings is 105 points, reflecting the sum of all rankings among the five officials [that is, 5*(1 + 2 + 3 + 4 + 5 + 6) = 105]. The formulas for calculating the results for Program 1 are shown in row 18. Although this may look a little confusing at first, please work through each formula, going from left to right across the spreadsheet. The key to understanding the allocation decision is column C, in which the rankings are recalculated to add up to 100 percent, which is then used for subsequent allocation decisions.

Note that the "Allocation" spreadsheet allows analysts to add further value by performing a what-if analysis. How are results affected by limiting that part of the budget that is subject to the prioritization to, for example, 5 percent? Why or why not might officials be interested in such results?

Location Decisions

What is the most efficient location for a fire or police station that makes service calls to different locations with different frequencies? Although specialized software programs exist for this purpose, the general principle is easily illustrated using a spreadsheet, which readily performs the necessary calculations.

This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Figure W10.5 Service Call Locations

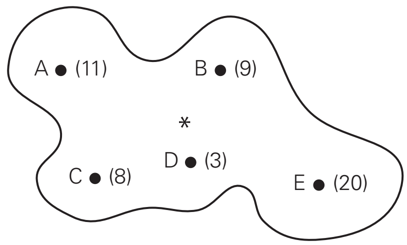


Table W10.1 Service Call Coordinates

Place	Coordinates X	Coordinates Y	Number of trips	Weighted X	Weighted Y
A	1.0	3.5	11	11.0	38.5
B	3.0	3.0	9	27.0	27.0
C	1.5	1.5	8	12.0	12.0
D	3.5	1.5	3	10.5	4.5
E	4.5	1.0	20	90.0	20.0
Total			51	150.5	102.0

Ideal X coordinate: 2.95
Ideal Y coordinate: 2.00

Take a look at Figure W10.5, and assume that service calls were made to each location (number of calls shown in parentheses). You can find the optimal location for the service center by specifying the coordinates of each location and weighting it against the number of visits. The coordinates are identified by overlaying an imaginary grid on the area. Based on such a grid (not shown here), the X and Y coordinates in Table W10.1 are obtained. The weighted value of Y for place A is calculated as 3.5*11 = 38.5, and so on. The ideal X coordinate is calculated as 150.5/51 = 2.95. The ideal Y coordinate is 2.00. The location of the ideal location is identified with an asterisk in Figure W10.5. Of course, the above result does not consider maximum response time and other requirements, which might affect such final decisions. After analyzing your findings, as presented in Table W10.1, the results can be cut and pasted from Excel to a word processing program for presentation. Their quick and thorough computation and ease of presenting results make spreadsheets a great resource when performing this kind of analysis.

Forecasting

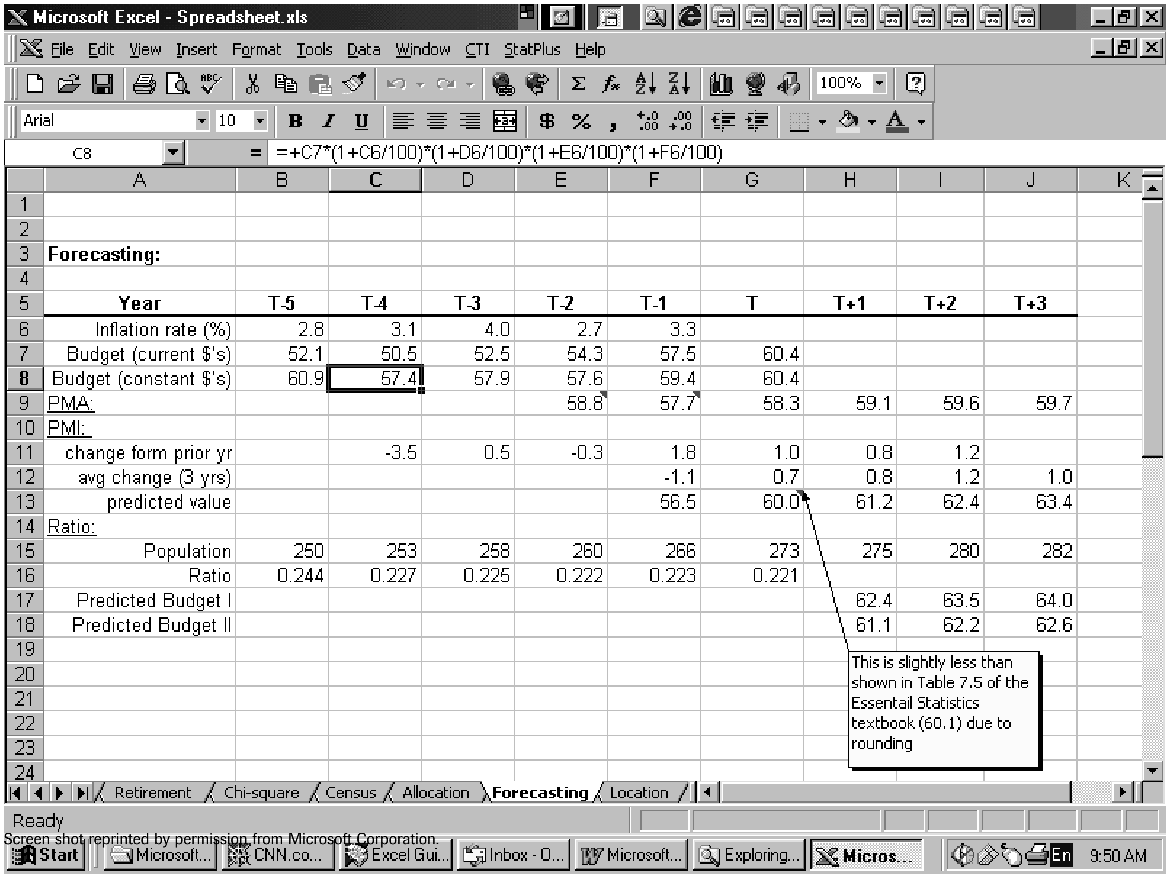
Chapter 7 of the textbook, *Essential Statistics for Public Managers and Policy Analysts,* discusses various forecasting approaches that analysts can use in their work. Excel can help perform these tasks. This section shows how to replicate the textbook forecasting examples using prior moving averages (PMA), prior moving changes, and the method of forecast ratios. Open the "Forecasting" spreadsheet. This spreadsheet also shows the use of pop-up comments that can be attached to each cell. (To create a pop-up window, highlight a cell, then click on the right mouse button and select Insert Comment.) In this example, we are forecasting budget expenditures for the next three years. The first step in forecasting is calculating all expenditures as constant dollars at time = *T;* the spreadsheet includes the formulas you will be using to convert current dollars to constant dollars. For example, the constant dollar value of expenditures at *T* – 4 is calculated as (see highlighted cell in Screen W10.10):

C7*(1 + C6/100)*(1 + D6/100)*(1 + E6/100)*(1 + F6/100).

Then these constant dollar values are used to create forecasts as explained in the textbook.

This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Screen W10.10 "Forecasting" Spreadsheet: Pop-Up Comments



	A	B	C	D	E	F	G	H	I	J
3	Forecasting:									
5	Year	T-5	T-4	T-3	T-2	T-1	T	T+1	T+2	T+3
6	Inflation rate (%)	2.8	3.1	4.0	2.7	3.3				
7	Budget (current $'s)	52.1	50.5	52.5	54.3	57.5	60.4			
8	Budget (constant $'s)	60.9	57.4	57.9	57.6	59.4	60.4			
9	PMA:				58.8	57.7	58.3	59.1	59.6	59.7
10	PMI:									
11	change form prior yr		-3.5	0.5	-0.3	1.8	1.0	0.8	1.2	
12	avg change (3 yrs)					-1.1	0.7	0.8	1.2	1.0
13	predicted value					56.5	60.0	61.2	62.4	63.4
14	Ratio:									
15	Population	250	253	258	260	266	273	275	280	282
16	Ratio	0.244	0.227	0.225	0.222	0.223	0.221			
17	Predicted Budget I							62.4	63.5	64.0
18	Predicted Budget II							61.1	62.2	62.6

The second step involves the forecasting of the constant dollar expenditures. The textbook discusses the three different approaches. PMA uses the mean of the prior three values to forecast the current period value. For example, the formula in cell F9 is =AVERAGE (C8:E8). However, when no more actual values are available, previously forecasted values are used instead. Thus, the forecast for $T + 2$ (cell I9) is calculated as (F8 + G8 + H9)/3.

By contrast, PMI uses the mean of the prior three *changes* to forecast the current period. This average change is added to the previous expenditure in order to calculate the current amount. For example, the predicted value for T (cell G13) is calculated as G12 + F8, where G12 is calculated as (C11 + D11 + E11)/3.

The third approach is forecasting by using known ratios. The ratio approach is simple to calculate; for example, the forecasted value for T is H15*[AVERAGE(B16:G16)]. The results in row 17 ("Predicted Budget I") are based on the average ratio of all known periods ($T - 5$ through T). The results in row 18 ("Predicted Budget II") are based on the average ratio of the three most recent known periods ($T - 2$ through T).

Different forecasting methods produce different results. By comparing alternative methods, analysts can provide decision makers with a range of plausible forecasts. In our example, we would forecast expenditures in $T + 3$ as lying between $59.70 and $64.00 (constant $s).

The "Forecasting" spreadsheet also shows how to incorporate patterns of periodicity, replicating the calculations for Table 7.7 in the textbook. These calculations are shown in the range B27:G50. By working through these formulas, you can see how these forecasts are made that involve periodicity.

STATISTICAL FUNCTIONS

In addition to its analytic abilities, Excel also has a wide range of statistical functions that will be helpful to you. Not only will these functions be of good use, but being aware of these capabilities can help you decide whether you need more specialized statistical analysis software. Statistical functions are found by selecting Tools → Data Analysis (or by selecting f_x on the standard toolbar).

This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Descriptive Statistics

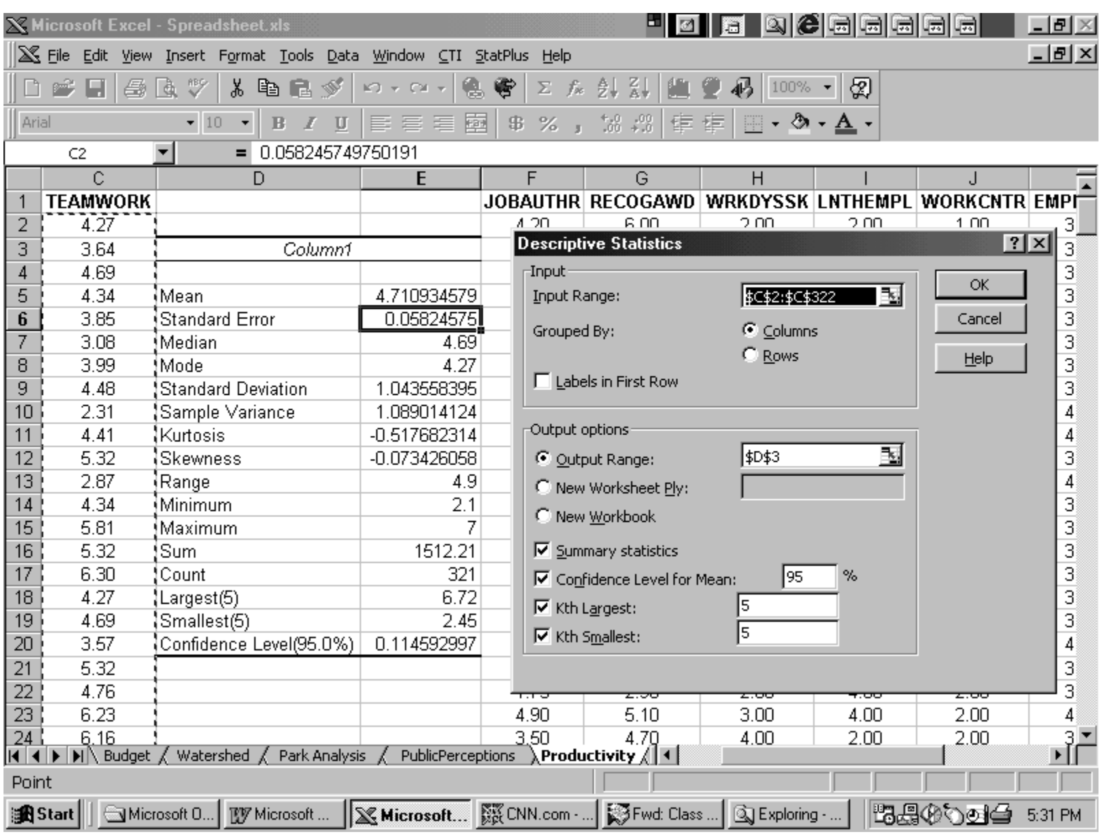
Excel produces a wide range of descriptive statistics, such as those discussed in Chapter 2 of the textbook. To explore these, open the spreadsheet "Productivity." Then, use the Descriptive Statistics Tool to calculate the mean and other measures of the variable Teamwork, an index variable measuring the use of teamwork. Select Tools → Data Analysis → Descriptive Statistics, as shown in Screen W10.11. Although these statistics can also be produced by using separate functions (for example, =STDEV [range] to calculate the standard deviation), the Descriptive Statistics Tool produces them all instantly. As is often the case, users specify the data range that is to be analyzed and the location of the output. Although this is a minor task, it is sometimes inconvenient when the data set is large. You can practice replicating this analysis for another variable on another, larger data set, such as the "Public Perceptions" spreadsheet.

Be aware that Excel does not produce boxplots. To calculate boxplots, Excel users must calculate each individual statistic. For example, to calculate the first quartile of the above variable, use =QUARTILE (C2:C322,1). (This function is fully explained when selecting the function wizard, f_x.) Then select Statistical → Quartile. Next, create your own formulas to calculate the inner and outer fences (see Chapter 2). Using Excel to create boxplots obviously requires a good deal of work. However, some "after-market" products exist for Excel (developed and sold by others) that include add-in statistical procedures that automate the above calculations and produce boxplots faster than would be done by making separate calculations. Comparing Excel's capabilities with those of another software package will help you to determine how you can best meet your software needs.

Chi-Square

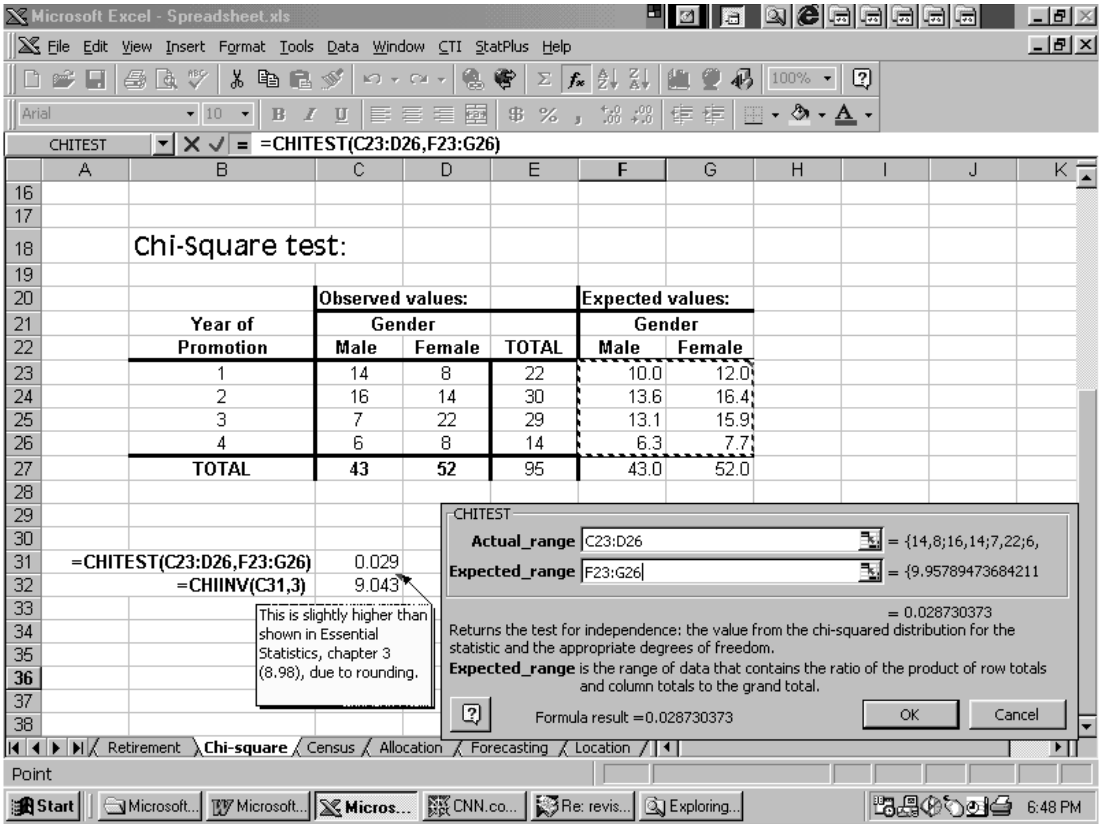
Excel can also assist in calculating chi-square in another step-by-step process. Open the spreadsheet "Chi-square." (This table replicates Table 3.1 in the textbook.) When performing a chi-square analysis in Excel from raw data, the first step is always to produce a table of observed values. Use the Pivot Table feature, as discussed earlier. The second step consists of calculating the expected values, the values that would exist if no differences were found. These are shown in Screen W10.12.

Screen W10.11 Descriptive Statistics Dialogue Box



This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Screen W10.12 CHITEST



The third step is hypothesis testing. The function =CHITEST (observed data range, predicted range) is used to calculate the probability of rejecting the null hypothesis. The dialogue box shown in the lower right-hand corner is obtained by using f_x. Upon calculating the probability, =CHIINV is used to produce the chi-square test statistic used to calculate the before-mentioned probability.

Note that Excel does not produce warnings concerning the chi-square test assumption that cells must have expected values exceeding 5.0. Also, although Chapter 4 of the textbook discusses a wide range of other bivariate statistics for categorical data, Excel has functions only for calculating chi-square. It has no functions for calculating tau-c, Kruskal-Wallis' H, and many other statistics discussed in Chapter 4.

T-Tests

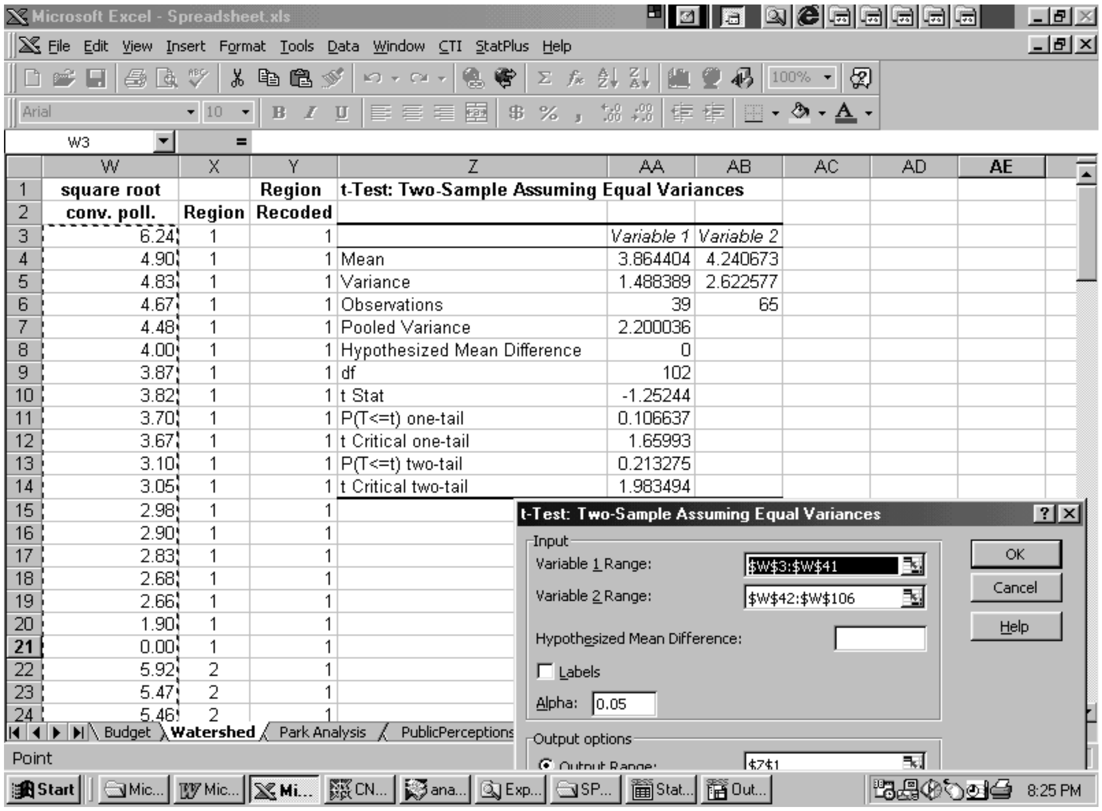
Excel can also produce independent and paired samples t-tests, as well as one-sample t-tests discussed in Chapter 5 of the textbook. However, Excel does not produce tests of normality, and so analysts must determine whether test assumptions of normality are met through visual inspection. Also, because boxplots are not readily produced, outliers are usually identified visually. Thus, the first step requires visual analysis. Fortunately, the Chart Wizard can help produce bar charts and histograms for this step. Visual inspection might also be used to determine whether sample variances are equal because the independent samples t-tests require users to specify, a priori, whether or not variances are equal; tests for the equality of variances are also not available.

The independent samples t-test requires Excel users to specify contiguous input ranges for both variables. Thus, sorting of the data set is sometimes a preliminary step. For example, Chapter 5 of the textbook examines whether mean levels of pollution differ between the East and other regions. To perform that analysis now, open the spreadsheet "Watershed." You should first sort these variables by region, in order to ensure that all of the pollution values associated with the East region are grouped together and that all of the pollution values associated with the other regions are grouped together.

To produce a t-test in Excel, use Tools → Data Analysis → t-test: Two-Sample Assuming (Un)Equal Variances. Look at Screen W10.13 to compare your results. Make sure that the test variable is the trans-

This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Screen W10.13 —— T-Test Results



formed variable Pollution as shown in Figure 5.4 in the textbook. Screen W10.13 and your results should be consistent with those shown in Table 5.2 (also in the textbook). The t-test statistic is 1.252, which is not statistically significant.

Excel can also do paired t-tests and one-sample t-tests. The paired samples t-test is used for evaluating differences between before and after tests (see Box 5.1 in the textbook). To produce it, use Tools → Data Analysis → t-test: Paired Two Sample for Means. The example from the textbook is reproduced on the "Watershed" spreadsheet. Open this spreadsheet and examine the results which are shown in the range AEI:AI15.

The one-sample t-test requires some user-specified formulas. Recall that the one-sample t-test test statistic is defined as $t = (\bar{x} - \mu) / s/\sqrt{n}$, where μ is the population mean (or hypothesized value) against which the sample mean is tested, s = standard deviation, and n = sample size. The first step is to calculate the t-test test statistic by calculating the sample mean, calculating the sample standard deviation, specifying the population mean, and calculating the t-test test statistic. The sample mean and standard deviation are easily calculated using the Descriptive Statistics Tool shown in Screen W10.11. The hypothesized value is user specified, and the t-test test statistic is calculated using the formula shown in Screen W10.14. Here, you want to know where the mean percentage of samples that exceed pollution levels in the East differs from 10.0; test it against $\sqrt{10}$ because the variable has been transformed in that way. The t-test test statistic is 3.59.

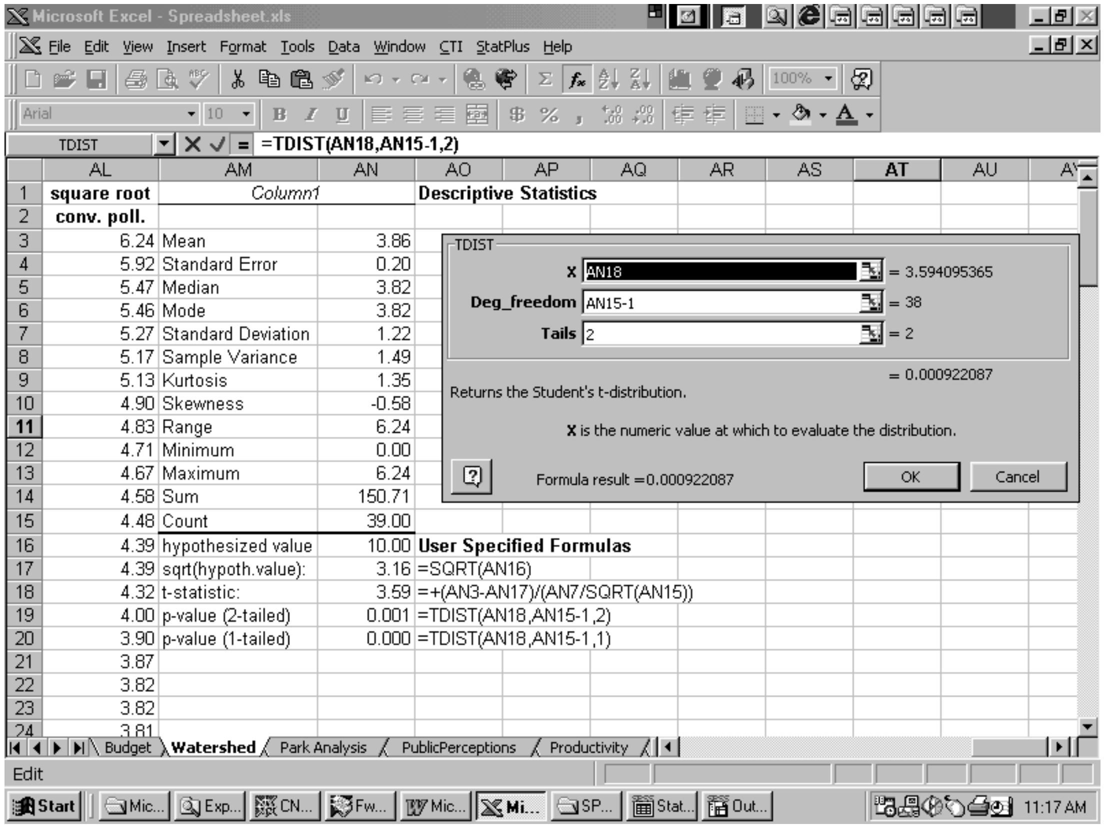
The second step is to determine the probability of obtaining a t-test test statistic of this value. You will need to use the =TDIST function ("TDIST" stands for "t-test distribution") shown in Screen W10.14. As you can see, the two-tailed probability is less than 1 percent; hence, a statistically significant difference exists between the actual and user-specified percentages of samples in the East that exceed pollution standards.

Regression

Other statistical tasks for Excel are calculation of correlation coefficients, performance of simple and multiple regression, and calculation of residuals. For a quick review of these concepts, please look over textbook chapters 6 and 7.

This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Screen W10.14 One-Sample T-Test Results



To calculate Pearson's correlation coefficients, use Tools → Data Analysis → Correlation (the default choice is Pearson's Correlation Coefficients). This dialogue box requires users to specify one contiguous area of all variables and observations. Excel can also calculate Spearman's rank order correlation coefficients. To this end, users must use f_x (or Insert → Function). Then =SPEARMAN is first used to calculate the Spearman's rank order correlation coefficient, and =SPEARMANP is then used to calculate the probability of association. To practice calculating these coefficients, use the data in the "Productivity" data set.

In the following example, you will use the "Productivity" data set and spreadsheet to practice regression analysis. Both simple and multiple regressions can be obtained using Tools → Data Analysis → Regression. The only difference is that in simple regression the input *X* range is only one column wide, whereas in multiple regression it is several, contiguous columns wide. The multiple regression results shown in the textbook (see Table 6.2) have been reproduced in Screen W10.15. The output shows the order of the independent variables, as entered on the spreadsheet. To produce the above output consistent with Table 6.2 in the textbook, the independent variables were copied and pasted in the above order. Use the button "Label" to produce labels for the variables in output range AA16:AA20.

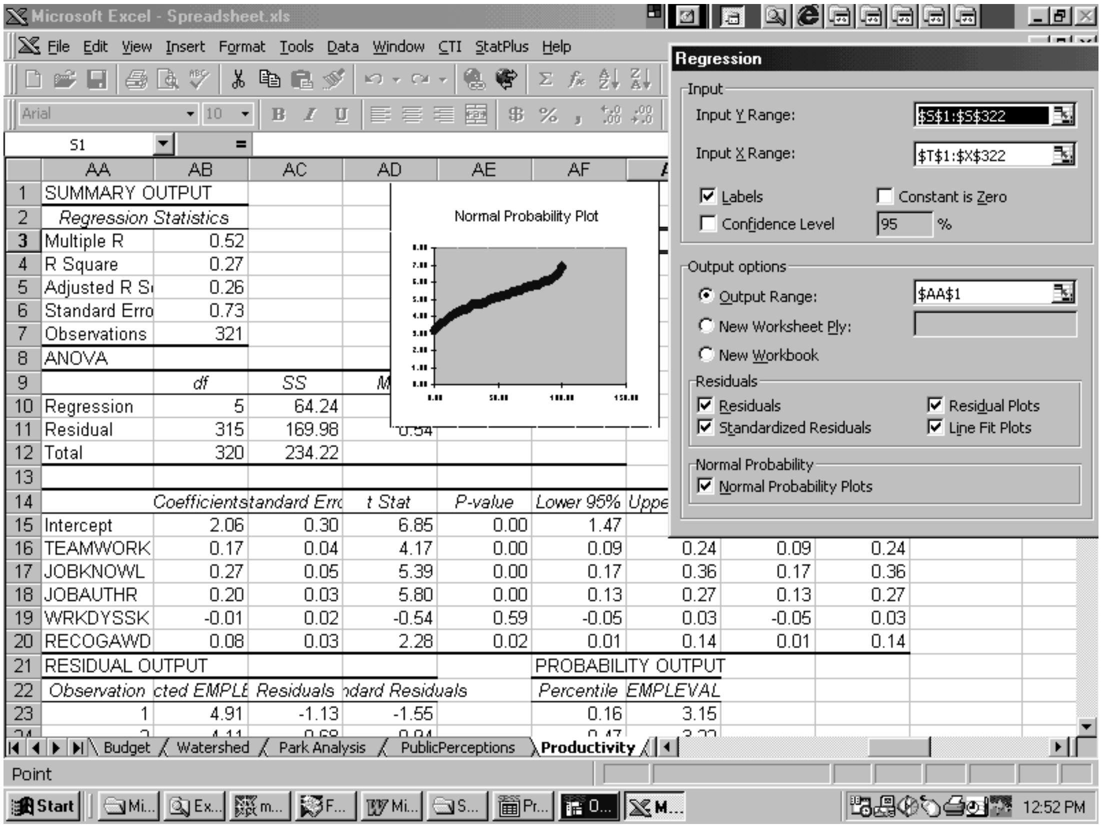
The output also shows predicted values and standardized residuals. The residual and line fit plots are plots of the independent variables; be warned—Excel does not produce the common error term plot shown in the textbook in Figure 6.3. To produce the error term plot, calculate the standardized predicted values. First, you must calculate the mean and standard deviation of the predicted values, and then use the formula $(x - \bar{x})/s$ to transform each value as shown in Screen W10.16. The error term plot in the screen reproduces the example from Box 7.1 in the textbook.

CONCLUSION

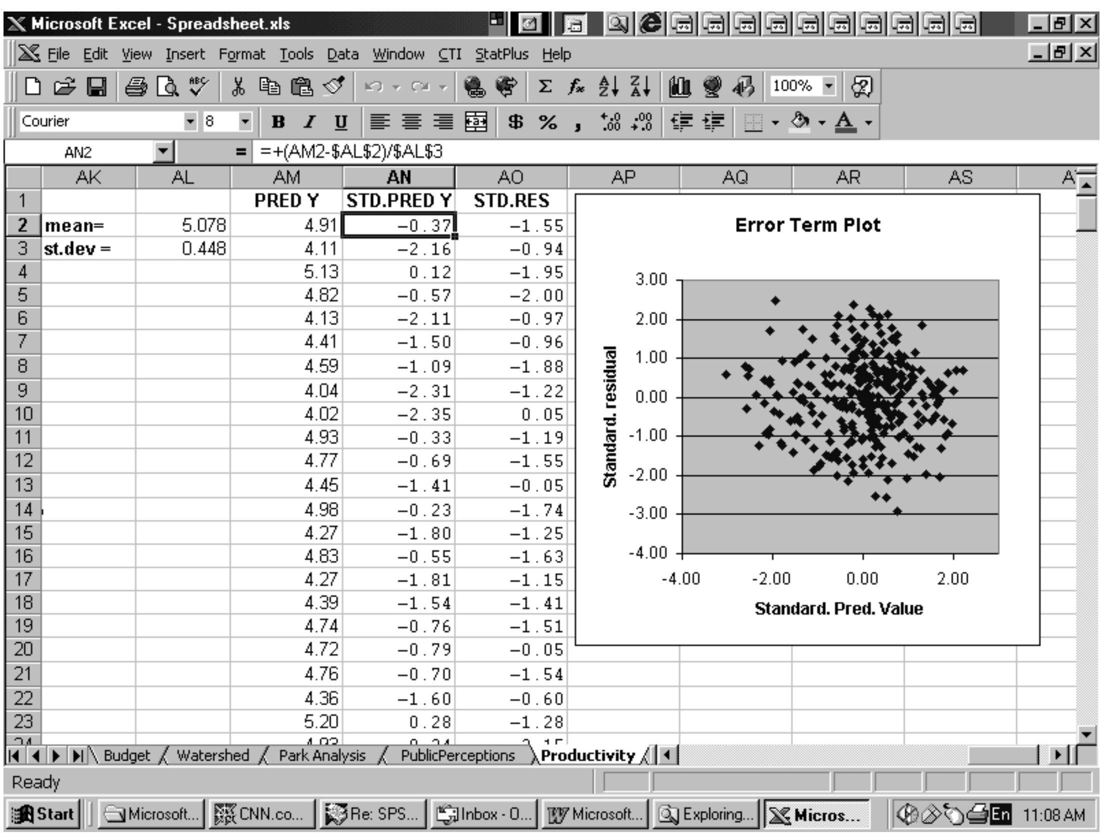
Although Excel does not produce further diagnostic statistics (as discussed in Chapter 7), nor any of the advanced statistical procedures (discussed in Chapter 8), spreadsheets play an important part in the lives of

This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

Screen W10.15 — Multiple Regression Results



Screen W10.16 — Error Term Plot



This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

public administrators and analysts. Accurate calculations, a wide range of analytical functions, and colorful charts and graphs make it easier for analysts to work with their data, to present their findings in a better way, and to undertake a wide range of quantitative, nonstatistical analyses. Spreadsheets do feature some basic statistical functions, which may suit your particular needs. Used in conjunction with specialized statistical software packages, spreadsheet use can round out your skills as an analyst. Through practical and creative use, spreadsheets can allow you to make many valuable and accurate contributions to public administration and policy analysis.

APPENDIX

The Parks Department (Exercise)

The following exercise combines various elements from the "Data Analysis" section. This is your opportunity to practice and further hone your skills. Consider the data in Table W10.2.

Table W10.2 Park Analysis Data

	Maintenance ($)			Estimated capital costs ($)		Avg. use Mon–Fri 9AM–7PM (p.p.hr)	Avg. use Sat–Sun 9AM–7PM (p.p.hr)	Avg. area household income ($1000)
	Year 1	Year 2	Year 3	Year 4–6	Acres			
Park 1	90	100	110	20	30	50	40	25
Park 2	400	425	450	400	200	500	1500	35
Park 3	330	360	390	400	150	100	2000	50
Park 4	30	50	30	5	2	75	50	15
Park 5	70	80	90	25	20	60	60	15
Park 6	150	150	155	45	60	50	700	50

Note: The current year is YEAR 3; Park 4 is a downtown Park. All maintenance and capital costs in thousands, for example, the maintenance cost of park 1 in year 1 is $90,000.

p.p.hr = persons per hour.

Table W10.3 Example Allocation (Hypothetical)

	METHOD A			METHOD B	
	Use	% of use	Allocation of $1,000	Rank use	Allocation of $1,000
Park 1	25	13.9	139	3.5	167
Park 2	40	22.2	222	5	238
Park 3	60	33.3	333	6	286
Park 4	20	11.1	111	2	95
Park 5	10	5.6	56	1	48
Park 6	25	13.9	139	3.5	167
Total	180	100	1,000	21	1,000

Note: Method example: 13.9%*1,000 = 139; 3.5/21*1,000 = 167

1. **What are the projected costs for each park over the next two years? Assume that past and future inflation is 4 percent annually. Assume that capital expenditures are evenly distributed over the three-year period (straight-line method). Calculate future maintenance costs as three-year moving averages; hence, Year 4 forecast is (Year 1 + Year 2 + Year 3)/3, when all expenditures are calculated in constant dollars. Therefore, calculate all of the expenditures in constant dollars *first*; use Year 3 (this year) as the reference (base) year.**

This chapter originally appeared as chapter 10 in Exercising Essential Statistics, first edition.

2. **What is the average maintenance cost per park per user in Year 3? Make a pie chart.**

3. **What is the average maintenance cost per park per acre in Year 3? Make a pie chart.**

4. **Which park is the most expensive? Print a spreadsheet showing the results of Questions 1–4.**

5. **Your total budget for Year 4 is $1,400,000. How would you allocate your budget in order to increase Saturday and Sunday use? Explain your justification. See Table W10.3.**

6. **Because of highway construction, the expected Saturday and Sunday use for parks 2 and 6 is expected to drop by 25 percent in Year 4. How does this affect your result? Explain.**

7. **Your budget just got cut by another 18.5 percent. How do you propose to allocate this shortfall, given your concern that areas with low incomes are affected least?**

Note: **Different answers are possible for Questions 5–7. Explain how different approaches might affect your answer. What do you conclude?**