

---

# The Use of Microcomputers for Demographic Analysis: An Overview of Options for the Novice User

---

by Diane Crispell<sup>1</sup>  
Associate Editor  
*American Demographics* magazine  
P.O. Box 68  
Ithaca, NY 14851

---



---

## Introduction

As the microcomputer editor and research associate of *American Demographics*, I review demographic-related software and answer research questions. People often ask us about software programs and data sources. I will be focusing on demographic and socioeconomic information since that is my forte, but the same observations apply to any social science data.

---

<sup>1</sup>Presented at the International Association for Social Science Information Service and Technology (IASSIST) Conference held in Washington, D.C., May 26–29, 1988

To give you a little background on my company, *American Demographics* publishes a monthly magazine about demographic and consumer trends. Our subscribers are primarily business people who are learning how to use demographic information. They are not academic demographers, but they have a need to analyze demographic data. At the same time, the use of microcomputers has become nearly universal in business. A natural connection is the use of microcomputers for demographic analysis, which is what my monthly column in the magazine is all about.

The demographic data industry, an industry that followed in the wake of the 1980 Census of Population, has been a forerunner in the release of desktop systems for demographic analysis. You may or may not be familiar with some of these firms – CACI, Claritas, Donnelley Marketing, Market Statistics, and National Planning Data Corporation, to name a few. National Decision Systems came out with the first major stand-alone PC system in 1985 – the Infomark laser disk system. Since then, PC applications have become increasingly sophisticated. Most major vendors now offer desktop systems with huge databases of population estimates and projections, mapping applications, retail sales estimates, geodemographic segmentation systems, and so on.

For the most part, these systems are for retrieval and display of data rather than for true data analysis. Furthermore, they are often complex to learn and require the absolute latest in hardware, such as CD readers, 20x20 Bernoulli boxes, and massive amounts of memory. And finally, the data used with these systems are proprietary and very expensive.

So what about the novice or occasional user? What are the alternatives? You *can* do demographic analysis on a microcomputer without all the bells and whistles.

---

## Spreadsheet Tactics

If you're a number-cruncher of any kind, one of the first things you learn to do with a microcomputer is to use a spreadsheet. Whether it be Lotus 1-2-3 or any of the many similar programs, setting up rows and columns of numbers and doing calculations on them is a basic and a must.

But many of us don't get much beyond the basics. And we may not realize how we can use spreadsheets to calculate complex statistics with survey data, or to generate forecasts. Or we may feel that we can't program the spreadsheet to perform these functions.

There is help. Walonick Associates sells a series of StatPackets, small programs that do statistical analyses using spreadsheet data files. Each StatPacket does a specific statistical procedure, such as descriptive statistics, crosstabulation, or multiple regression. You set up your spreadsheet file in the correct format (cases in rows, variables in columns) and save it, then run StatPackets. Select the procedure, spreadsheet file, some analysis options, and off you go. This can be a sensible alternative to using a full-blown and complex statistical package like Walonick's StatPac Gold. You can also produce rough graphics from spreadsheet programs, as an alternative to learning a whole separate graphics program. Some of the new spreadsheets like Borland's Quattro have pretty nice graphics, better than Lotus 1-2-3 anyway, and they're cheaper, too.

You can use spreadsheets for analysis of data other than survey data, too. We're holding a series of microcomputer workshops this year, and one of the exercises for attendees is to produce population projections with a spreadsheet. We're providing them with templates and some choices of datasets, so they can see the assumptions and formulas involved

in doing this kind of forecasting. One template I've used for a long time is one that calculates the percentages needed to produce a pyramid of the population by sex and age groups.

---

## Bigger Packages

Of course, you can't do a whole lot of raw data processing with just a spreadsheet program. Sooner or later, you have to get involved with a survey or statistical analysis program. The spectrum of software available is broad.

Where you start with microcomputer data processing depends, first and foremost, on how you get your data. Do you collect them yourself? If you do, is it through telephone or personal interviews? Mail-in questionnaires? If you're using secondary data, what format are they in? Are they on tapes, floppy disks? How do you get the data onto your microcomputer anyway?

In my research at American Demographics, I've done a little bit of everything somewhere along the line. We don't do a lot of primary research ourselves, but we do conduct reader surveys occasionally, in the form of mail-in questionnaires. I've analyzed several of these using survey tabulation software.

## Survey Packages

These programs vary from small and simple to vast and complex. At the lower end, you will find programs like Henry Elkins & Associates' SurveyMate, which Dr. Elkins developed primarily for his own use and that of other social scientists out in the field. It is geared towards easy data entry and it is inexpensive (\$145), but this doesn't mean it's a small program. SurveyMate can handle up to 1,000 data fields and 32,000 cases, ample for many

purposes. It doesn't do every kind of statistical analysis, but it does frequencies, crosstabulation, and multiple regression, and you can export the data to ASCII files for presentation purposes or for further analysis. When I talked to Dr. Elkins recently, he told me the program is now used in 42 countries - by academics, government agencies, and market research firms - in places like Nigeria, Singapore, and Papua-New Guinea.

I've looked at a few other survey programs, like SurveyTab, UNCLE, and the Survey System. These more costly commercial programs tend to produce nicer looking tables and may offer more user-friendly features, but I think that you should choose survey software on the basis of what you need. I have a few guides as to what questions to ask about survey software:

1. capacity - This is pretty basic. Can the program handle the number of cases and variables you have?
2. type of data entry - would you rather enter the data field by field on a screen displaying the questions, or would you rather enter data in an 80-column card format, just a string of digits across the screen? The former method is good if you don't have huge amounts of data and less experienced people doing the entry. The latter is faster, but takes more concentration and expertise. And most programs seem to do it one way or the other, so you should decide before you buy.
3. recoding and transformation- how flexible is the software in terms of regrouping data into ranges (like age groups) or other specifications, such as missing values. How does the program deal with multiple-response, open-ended questions, and other tricky survey data quirks?
4. analysis - can the program perform all of the procedures you need? If it can't, how easy is it to transfer the data to another

program that can, and do you want to get involved with this?

### *Statistical Analysis*

In many cases, survey software isn't enough. You may need more complex statistical procedures such as analysis of variance. Or you may want a link data analyses to graphics or mapping programs for presentation purposes. Or you may not need a program to enter raw data, but already have it on a tape or floppy disks. In this case, you might need a full-blown statistical package.

There are a lot of these around, too. A lot of people have called and asked me, "What package should I buy?" I have no pat answer, it depends so much on what you're doing. As with survey software, when choosing a stat program, check its data capacity, ease of use, and analysis options. The big programs are pricier, but of course they do more. Also look out for additional "modules," which a number of programs tack on, at extra cost. For example, SPSS-PC+ is handy for those used to the mainframe version, and the learning curve is smaller, but if you want advanced statistics, that's another few disks (and a few hundred dollars). The same goes for nicely formatted tables, and user-defined data entry modules, as well as separate graphics and mapping programs. It's sort of unavoidable - statistical analysis programs take up a lot of computer space, and it is nice to be able to buy only the parts you need.

---

## Demographic Data and Software Options

I'm not going to go into more detail about these kinds of analysis programs. Many people don't need to analyze raw data. They want to retrieve and manipulate, in relatively simple ways, existing data.

The array of demographic data and software for microcomputers is enormous. Sources range from government agencies to private vendors, as I mentioned earlier. Data media range from datasets on floppy disks to massive databases on optical disks. Software capabilities range from simple retrieval to thematic mapping.

### *Government Sources*

I'm going to give a brief overview of government sources of demographic data for microcomputers. Several federal agencies offer data online and on floppy disks. Stu Weisman will be telling you more about government data available on floppy disks through NTIS later in this session.

The Bureau of Labor Statistics currently offers data on diskettes and online through an electronic news service. The data diskettes are formatted for use with Lotus 1-2-3 spreadsheets and cover topics like Consumer Price Indices for 104 items and 54 U.S. cities, useful for comparing cost of living. Another series has monthly and annual average labor force data by age, sex, and race for current year and 3 prior years. The online news release service lets you download BLS press releases (about 100 a year, which often include data tables) for a minimal fee.

The Census Bureau has an online service, too, called CENDATA. For those of you who aren't familiar with this database, it is offered through DIALOG Information Services, one of the largest online vendors. CENDATA offers a

small amount of the Bureau's voluminous data holdings, including many of their publications such as press releases, the *Monthly Product Announcement* newsletter, and portions of *Data User News*. It also provides other population and economic data from the Bureau.

The Census Bureau offers some data on floppy disks, too. I checked with the Bureau last week to get an update on what's available. They've had the *1986 State and Metropolitan Area Data Book* in disk form for some time, as well as the *1983 County and City Data Book*. I was told that the *1987 County and City Data Book* is on the way, as are 1986 population estimates and 1985 per capita income estimates for governmental units (this was announced as already being available through CENDATA, but apparently there was some problem with the data). The data come in a format that you can load onto spreadsheets and other programs on IBM-type PCs. When you buy disks from the Bureau, they send you a little utility program that basically lets you view the data in a table format, but little else.

The Bureau will also download any data you want on to floppy disks on a custom basis, but of course this costs more and probably has a much longer turnaround time than if you purchase standard offerings.

Government data - whether you download them from an online system or purchase disks - tend to come in the form of printed tables - in other words, they're designed to print out and look at, rather than do any analysis with. It takes a little work to get these data into a spreadsheet format to work with, but it can be done. And if you use a Macintosh rather than an IBM, there are ways to transfer data for use on the Mac. If anyone has any questions about this, ask me afterwards.

### *Private Sources of Demographic Data*

To get more usable data, you often have to go to private sources. There are a number of firms that offer demographic data for microcomputers on floppy disks, often with the software to retrieve and analyze them. As I mentioned earlier, some of these systems are enormous and costly, incorporating massive databases and the latest in hi-tech hardware. But some are lower-end, for use on simpler PCs.

A good example of a self-contained demographic data/analysis package is **Your Marketing Consultant** from Market Statistics in New York. Market Statistics produces the Survey of Buying Power data that appear in *Sales & Marketing Management* each year. For those of you not familiar with this, the Survey of Buying Power is not actually a survey — it is a database of Market Statistics' proprietary estimates and projections of population, income, and retail sales for states, metropolitan areas, and counties. The YMC Advanced Consumer program comes with data for all of these geographic levels, as well as for ADI and DMA market areas (defined by Arbitron Ratings and A.C. Nielsen for the media industry). The company also has a business-to-business version of the program. Your handout includes a copy of a column I wrote for *American Demographics* explaining the capabilities of the earlier version of this software — it has been enhanced since then (I believe the price has gone up slightly, too). Basically, the software can search for areas of a given geographic level based on demographic and economic criteria that you select — for example, you may want to find out which counties in the South have a black population of at least 50,000 and then rank them by median income. You can do this easily with YMC.

Even some hi-tech products are not expensive or hard to learn. Slater Hall Information Products of Washington offers a number of government databases on CD-ROM — the 1982 Census of Agriculture down to the county level,

a county statistics database with demographic and economic data, several business/economic databases, and population statistics including the entire STF-3C computer tape. SHIP sells the CD databases with their software, which is called Searcher. With Searcher, you can retrieve data by geography or numeric criteria, view them as tables and export them to Lotus or other programs for further use. Your handout includes a review of this program, too.

There are a number of other firms offering demographic and economic data on floppy disks, too. I've given you a listing of those we're aware of. Some are specialized, like Analysis and Forecasting in Cambridge, which offers IRS migration data — state-by-state and in- and out flows over time. Or the Center for Continuing Study of the California Economy in Palo Alto, which offers population and employment estimates and projections for the state of California and its counties and cities.

Other companies offer a broader variety of data — such as CACI of Fairfax. CACI is one of what we call the "full-service" demographic data companies. Their data include 1980 census, current-year, and five-year projections of the population by age and sex for virtually any geographic area in the U.S. Other databases include business and retail information, and the ACORN geodemographic segmentation system. (ACORN is an acronym for A Classification of Residential Neighborhoods). These big companies also offer entire databases online or with Bernoulli and CD technology, including the software to analyze them. Smaller companies typically only sell the data — you usually have to figure out how to use them.

#### *Other Software*

The list I've given you of data products doesn't include software to analyze data — survey, statistical, graphics, and mapping software. I mentioned the survey and statistical software earlier. Displaying data with charts or maps is

another important element of demographic analysis, and there are a number of these programs available. Some well-known business graphics programs include Harvard Graphics from Software Publishing, GraphWriter and Freelance from Lotus Development, and ChartMaster from Ashton-Tate. With these programs, you can create bar, line, pie, and other types of charts directly from data series. These programs offer many features in terms of labeling and customizing chart formats, and allow users to create very effective presentations of demographic information.

While you can graph any kind of quantitative data, mapping is a more specialized task - the display of geographic relationships. Since demographic information is often analyzed in terms of geography, demographic data and mapping are a natural combination. And, in fact, mapping software grew out of the demographic industry in the early 1980s. People had been mapping data for a long time, of course, but it was done laboriously by hand or on mainframe computers. Desktop mapping was not a reality until Sammamish Data Systems introduced DIDS - Desktop Information Display System - in 1983, along with demographic data, of course. Since then, a number of mapping programs have come on the market, including MapMaster from Ashton-Tate, Atlas\*Graphics from Strategic Locations Planning, and MAXpc Mapping Analyst from National Planning Data Corporation. There are even a couple of micro mapping programs available for the Macintosh now.

question is what data you need, and why, which is something you have to determine for yourself before you go shopping for demographic data.

□

---

## Conclusion

As you can see, there are a lot of options for demographic analysis in the desktop world. It's a matter of knowing which data are available and in what format. Of course, the real