PROJECTING POPULATION USING SAS SOFTWARE
Larry E. Robinson, Virginia Department of Planning and Budget

Abstract
The Virginia Department of Planning and Budget is legislatively mandated to develop short and long-range population projections for use by other state agencies and the legislature for programs that involve or necessitate population projections. However, other clients include local and regional governments, universities, and business as well as private citizens. Prior to 1983, this task was done primarily by consultants with our agency providing the current data, making model assumption decisions, and adjusting the mechanical projections. In 1984, the decision was made to develop the model in-house and SAS software was chosen as the best computer language to use.

Using SAS software, we developed a model for roughly what the consultant was charging for one update cycle using his model. The marginal cost for running the model is low, allowing us to test many additional assumptions that we wouldn't have otherwise because of prohibitive costs. Several indirect benefits occurred. We are much more proficient now in both SAS software and demography and this added expertise has been useful in other projects as well.

To explain the entire project is beyond the scope of this paper. Instead this paper will present a general overview and discuss the planning and development of the project with the project with emphasis on the aspects of SAS software that were found to be most beneficial.

The following topics are discussed:
- Telephone Survey and Literature Search
- Workplan and Structure Charts
- Data Adjustments
- Coding Scheme
- Macros and Arrays
- Testing
- User's Manual

Overview
Let me briefly explain the objective of the project. The concept of a cohort is central to this discussion and is defined as:

A requirement of our model was that separate projections be developed for seventy-two cohorts per locality. These cohorts include two race categories (white and non-white), 18 five-year age categories (0-4 to 85 and over) and two sex categories. That is, one cohort would be white males 0 to 4 in age while white males 5 to 9 in age would be a separate cohort. For each year projected, 72 individual projections were to be developed for 136 localities translating into nearly 20,000 projections. Additionally, projections were required for multiple years: 1990, 1995, 2000, etc. out of 2020.

Telephone Survey and Literature Search
An initial objective was to find out what currently existed in the field and not to duplicate work that already had been done. We conducted a telephone survey of other states to determine the types of models in use, how successful they had proven to be, and the feasibility of transfer. We also conducted a literature search to determine what new innovations there might be in this area.

Results of the survey of other states indicated that models did not exist that both projected well and that could be transferred to Virginia. For example, adjustments for Virginia's unique migration patterns and racial composition could not be incorporated into some of the more promising models.

A number of different methods are used to project population; we chose the cohort-component method. An area can only grow in two ways, either by natural increase which is the excess of births over deaths or (2) by migration which is persons moving into an area. (Of course, either of these components could be negative contributing to a loss in population). The cohort-component method projects each of the components individually. That is, births, deaths, and migration are each projected separately and then combined to produce population totals.

The literature search indicated that the U.S. Census Bureau had done the most research in this area. Even more importantly, they had published a handbook of methods for doing the type of projections that we felt would be the most suitable for Virginia. This handbook was designed as a guide for doing a projection for one
locality by hand using a calculator. However, with some modifications, it provided the basis by which our model was developed.

Workplan and Structure Charts

A detailed workplan was developed in which all the steps were listed with a description of the programs and data needed. This determined the sequence of program development and data collection, program dependency, and an estimate of the total time required for the project.

The entire process was divided into a number of separate programs which were divided among two research analysts. For each program a structure chart was produced. Structure charts take a process and divide it into boxes on a very macro level. Each of these steps is then subdivided into discrete steps and continue until the level needed. A structure chart was found to be superior to a flowchart for our purposes; it is much easier to visualize and it is easier to make changes. The structure charts served two purposes. It was an ideal way to view both the overall program as well as the detail; and management could keep abreast of the project by reviewing the highest level without becoming involved with the detail.

Data Adjustments

Many of the data had to be adjusted for situations such as annexations or special populations. If the data base were not adjusted for annexations the latest census figures would indicate that some localities had grown at unusually rapid rates while others had declined rapidly and would have resulted in false historical rates. In other words, the data base had to be adjusted as if the annexation had occurred before the time period that we were working with. The adjustments for special populations remove persons in military barracks, colleges, etc. from the data base before birth rates and survival rates were applied. For example, the birth experience for women in college would be different than for women of the same age in the general population thus requiring adjustments.

Coding Scheme

Many variables were used and since we were limited to eight letters, a uniform coding scheme was used. Key attributes that were used numerous times over in many programs such as age, race, or sex were assigned key letters that were reserved. That is, for nonwhite the letter "N" was used, the letter "M" for males, "B" for births, etc. Using this scheme, NBMB4 would mean births of non-white males in the year 1984. The order in which the letters would appear was also predetermined. We attempted, as much as possible, to build a data dictionary initially and list and define all variables; but we realized that additional variables would be needed. This scheme allowed one programmer to create new variables consistent with what the other programmer was developing. PROC CONTENT was used frequently to review variable names and history.

Macros and Arrays

Macros (the old style) were used as extensively as possible. Often the same program was run numerous times with slight variations to test different assumptions (i.e. various birth rates, male/female ratios, etc.). Macros were extremely useful for this; anything that could be varied in the program was placed in a macro even if it was just the name of a data set. For example, a data set containing birth rates would be referred to as BIRTHRTX in the program but BIRTHRTX would be a macro previously defined perhaps as:

```
Macro BIRTHRTX BASE.BIRTRT90 %
```

Then, for each different data set to be used, a two-part data set name would be changed in the macro statement. To make it easier for a user to run the program, the macros were moved to the initial section of the program so that a user would only have to search one place.

SAS arrays were found to be the best way to develop most of the repetitive routines. Often we had arrays two or three levels deep which, not only was more efficient in terms of programming time, but made the program much easier to read.

Testing

PROC MEANS was used frequently to determine totals, averages and ranges to ascertain the validity of the output quickly. PROC PRINT was employed to obtain a view of the output before a new report was formatted. I should mention at this point that we did keep our testing and production libraries separate and only moved a program to the production library after it was tested.

User's Manual

After the programs were developed and tested, a user's manual was developed. The user's
The manual contained the following for each program:

1. Purpose.
2. Methodology.
3. Comments.
4. Data inputs.
5. Data outputs.
6. Macros to be defined.
7. Structure chart.

Also contained in the manual is an overview of the entire process as well as copies of the programs and sample outputs.

Conclusion

In conclusion, the project was quite successful. We initially identified experts in both SAS software and demography in case we ran into some stumbling blocks and needed outside help. Surprisingly, the process flowed smoothly and we finished in the estimated time without the need for outside assistance. Since our agency uses COBOL and other conventional languages for most BP work and with much longer timeframes and the use of more resources, we are convinced that the use of SAS software is the key reason that this project was successful.

An additional advantage of using SAS software is that some of our clients are SAS software users. This allows us to provide them with a tape with the projections already in a SAS data set ready to use with their SAS programs.

Our installation has previously only had Base SAS software but recently other SAS products have been installed. We are especially excited about the possibility of using SAS/GRAPH* to produce population pyramids, density maps, etc. that have previously been produced by hand.

*SAS and SAS/GRAPH are the registered trademarks of SAS Institute Inc., Cary, NC, USA.

Acknowledgement: Development of the model was a joint effort of Debra A. Stephens, Donald P. Lillywhite, and the author. I would also like to thank Michael House for introducing me to SAS software a few years ago.