Methodology

The Demographics & Workforce Section of the University of Virginia's Weldon Cooper Center is the agency responsible for producing the Commonwealth's official population estimates for its counties and independent cities.

The population estimates are produced annually for each non-census year. This document explains the data required to make the estimates and the methodology used to generate them.

What is a Population Estimate?

Although population estimates and population projections are both attempts to measure the size of a population, they are actually quite different.
Population projections look to the future. They aim to produce a number that represents the population size one, five, ten, or twenty years from now. As a result, projected quantities like births, deaths, and net migration are an integral part of making a projection.
Population estimates look to the present or the recent past. They are usually much more accurate than projections because they make use of current indicators of population growth and change-data like births or licensed drivers-that are direct measurements, usually derived from government agency records. The range of statistical methods that can be used to make estimates is consequently greater than the methods available to produce projections.
The three most widely used estimating methods are the tax return method, the housing unit method, and ratio-correlation. The Census Bureau uses the tax return method to develop its county/city series and the housing unit method to develop its town series. (In the process of generating the town estimates, the Bureau also generates housing-unit-method estimates for counties and independent cities. These were released to the public for the first time in 2003.) The Weldon Cooper Center uses the third method, ratio-correlation, since this method has proven to be more accurate than the other two in estimating the population of Virginia localities.

Data Requirements

1. The sum of personal and dependent exemptions on state tax returns in the estimating and base years (the base year is the most recent census year), obtained from the Virginia Department of Taxation.

2. A sum of births in the estimating and base year and the two years preceding each, obtained from the Virginia Department of Health, Center for Health Statistics. A multi-year sum of births is used because births vary from year to year, and a three-year sum tends to even out some of this random variation.

3. Public school enrollment in grades 1-8 in the fall of the year preceding the estimating and base years. Public school figures are for fall (September 30) membership and are obtained from the Virginia Department of Education.

4. The number of licensed drivers in the estimating and base years, obtained from the Department of Motor Vehicles.

5. The housing stock in the estimating year. The housing stock is estimated by adding cumulative building permits together with an estimate of the number of mobile homes to the base year housing unit counts. Building permits are obtained from the Bureau of the Census' Construction Division and, in the case of missing data,

from the localities themselves. Mobile home data are derived from 2000 Census counts and Division of Motor Vehicle figures.

6. The civilian group quarters population. The group quarters population is made up of persons permanently or quasi-permanently residing in institutions with populations of approximately 50 or more. Included are long-term health care facilities, correctional institutions, and residents of college and university dormitories The data are obtained from a variety of sources, including the Virginia Department of Corrections (prisons), the State Compensation Board (jails), the Department of Juvenile Justice (juvenile detention facilities), and the institutions themselves (dormitory populations).

The Ratio-Correlation Method

The ratio-correlation method uses a prediction equation generated by multiple-regression analysis, a statistical technique that measures the relationship between one or more indicator variables (X's, or independent variables, or exogenous variables) and a dependent variable (Y). For use in estimating the population, a new equation is generated once every ten years, using the Census Bureau's decennial census population count as the dependent variable. This equation takes the form:

$$\text{Estimated } Y = b1(X1) + \ldots + bn(Xn) + C$$

where each b is a regression coefficient, each X is an indicator variable, C is a constant, and Estimated Y is the dependent variable.

These variables are used in the equation in a double ratio form, in which each locality's share of the state total in the estimating year is compared to its share of the state total in the base year (the year in which the equation was developed). For example, in producing an estimate for 2006, the form of the births indicator is:

[(2006 locality births) / (2006 state births)]

divided by

[(2000 locality births) / (2000 state births)]

where "births" in the numerator are the sum of 2004-2006 and "births" in the denominator are the sum of 1998-2000. It is because all the variables in the multiple-regression equation are expressed as ratios of ratios that the method is called the "ratio" correlation.

These double ratios actually measure change in the locality's share of the state total for the indicator and are the major diagnostic indicator when a question about why an estimate has risen or fallen arises. A double ratio under 1.0 indicates that the locality's share of the state total for that variable has fallen; a double ratio over 1.0 indicates growth in that variable.
The double ratios are also a major diagnostic tool for evaluating errors, either in procedure or, more often, in the data. Each time the estimates are produced, the 134 series of five ratios (one set for every county and city) is reviewed for any instance where one or more of the double ratios seem out of line with the others. In each such case the input data is checked, first against the Cooper Center's own records, then with the data provider. If neither yields a satisfactory answer, Weldon Cooper staff contact the locality for clarification.

The current indicator variables and their coefficients are:
Estimated housing stock 0.429
State tax return exemptions0.172
Three-year sum of births 0.083
Public school enrollment 0.214
Licensed drivers 0.163

Constant (intercept) -0.064

After multiplying each of the double ratios by its coefficient and adding the constant, mathematical adjustments are made to allow for the time elapsed between the base year and the estimating year. The resulting estimate is not a population figure; rather it is the percentage of the state's total non-institutional population that will be allocated to the locality. These percentages are adjusted to equal 100%, and the adjusted percentages are applied to the state's total population to yield a non-group quarters, or household, population estimate.

The final step in producing an estimate is the addition of the institutional population to the estimated household population to obtain an estimate of the total population.

It is extremely important to understand that ratio-correlation does not involve a one-to-one correspondence between each indicator variable and the population estimate. That is, a coefficient of 0.214 in no sense indicates that for every student in grades 1-8, 0.214 of a person will be added to the estimate. Similarly, although it may seem that each person in the institutional population constitutes one person in the estimate, this is not the case. Because the equation is derived using the base year non-institutional population, the amount of change in the institutional population between the base and estimating years is actually more important than the numbers themselves.

The Housing Unit Method

The Cooper Center's 2006 final and 2007 provisional estimates, as well as onward estimates up to 2010, are prepared by combining ratio-correlation estimates with a new series of estimates derived from the housing unit method. This approach incorporates the Cooper Center's newest research on estimates methodology, which was funded by the U.S. Census Bureau in 2007.

The housing unit method of estimating post-census population is straightforward. In order to make a current population estimate, four data items are required, each of which is estimated for July 1 of the estimate year: (1) the current housing stock, (2) the occupancy rate (3) the average number of people living in each occupied unit, and (4) the number of people living in group quarters. The household population is derived by multiplying total housing stock by the occupancy rate and by population per occupied housing unit. An estimate of the group quarter population living in the locality is added to the household population estimate to produce the estimate of the total population.

The housing unit estimates for all Virginia localities are then summed, and raked to the state control total. Finally, ratio-correlation based estimates and housing unit based estimates are evaluated and averaged to produce the most accurate population estimate series.

The State Control

The state control is an estimate of the entire state's population in the estimating year. The current model uses the most recent Census Bureau estimate available at the time the locality estimates are produced.
Census Bureau Estimates and Weldon Cooper Center Estimates
The Census Bureau's population estimates are currently produced by a method that in some cases yields very different results from those produced by the Weldon Cooper Center. The Census Bureau uses a component estimating methodology. It involves updating the population from the date of the last census with births, deaths, and an estimate of net migration derived from a matched sample of federal income tax returns. Due to federal confidentiality requirements these tax data are administratively restricted to use by Census Bureau personnel.

Provisional vs. Final Estimates

Ordinarily the difference between provisional and final estimates for a given year are produced by differences in both the data inputs and the state total. First, at the time the estimates are generated, final figures for births, housing permits, and state tax returns are unavailable and must themselves be estimated. Thus, provisional estimates are just that- population estimates that are made from several estimated data series. Second, the Bureau of the Census periodically revises the state total, and each of the estimates uses the most recent total. As a result, the state totals used for provisional and final estimates almost always differ.

As a result of both these factors, the provisional estimates, though more timely than the final ones, are less accurate.