Chapter 1

Econometrics: The Economist's Approach to Statistical Analysis

In This Chapter

- Discovering the goals of econometric analysis
- Understanding the approach and methodology of econometrics
- Getting familiar with econometrics software

Welcome to the study of econometrics! The Econometric Society, founded in 1930, defines econometrics as a field based on a "theoretical-quantitative and empirical-quantitative approach to economic problems." This mouthful means that, at times, econometricians are mathematicians and use complex algorithms and analytical tools to derive various estimation and testing procedures. At other times, econometricians are applied economists using the tools developed by theoretical econometricians to examine economic phenomena.

In this chapter, you see that a distinguishing feature of econometrics is its development of techniques designed to deal with data that aren't derived from controlled experiments and, therefore, situations that violate many of the standard statistical assumptions. You also begin to understand that, under these circumstances, obtaining good quantitative results depends on using reliable and adequate data as well as sound economic theory.

And because computers and econometric software are now commonly used in introductory econometrics courses, I also devote a section of this chapter to introducing basic commands in STATA (version 12.1), a popular econometrics software program. This software allows you to immediately apply theoretical concepts and enhance your understanding of the material.

Evaluating Economic Relationships

Economics provides the theoretical tools you use to evaluate economic relationships and make qualitative predictions of economic phenomena using the *ceteris paribus* assumption. You may recall from your previous courses that the *ceteris paribus* assumption means that you're keeping everything else constant. Two examples among numerous possibilities are:

- In microeconomic theory, you'd expect economic profits in a competitive market to induce more firms to enter that market, *ceteris paribus.*
- In macroeconomic theory, you'd expect higher interest rates to reduce investment spending, *ceteris paribus.*

Econometrics ties into economic theory by providing the tools necessary to quantify the qualitative statements you (or others) make using theory. Unknown or assumed relationships from abstract theory can be quantified using real-world data and the techniques developed by econometricians.

The following section explains how econometrics helps characterize the future and describe economic phenomena quantitatively, and then I clarify why an econometrician must always make sensible assumptions.

Using economic theory to describe outcomes and make predictions

One of the characteristics that differentiate applied research in econometrics from other applications of statistical analysis is a theoretical structure supporting the empirical work.

Econometrics is typically used to explain how factors affect some outcome of interest or to predict future events. Regardless of the primary objective, your econometric study should be linked to an economic model. Your model should consist of an outcome of interest (dependent variable, *Y*) and causal factors (independent variables, *X*s) that are theoretically or logically connected to the outcome.

Relying on sensible assumptions

A variation of a famous joke about economists goes as follows: A physicist, a chemist, and an economist are stranded on an island with nothing to eat.

A can of soup washes ashore. The physicist says, "Let's smash the can open with a rock." The chemist says, "Let's build a fire and heat the can first." The economist says, "Let's assume that we have a can opener. . . ." Despite the joke, making assumptions about reality can help you construct logical arguments and predict outcomes when specific preexisting conditions apply. In econometrics, however, making assumptions without checking the feasibility of their reality can be dangerous.

WARNING!

Making too many assumptions about preexisting conditions, functional form, and statistical properties can lead to biased results and can undermine the estimation accuracy you're trying to accomplish. Although you have to make some assumptions to perform your econometric work, you should test most of them and be honest about any potential effects on your results from those you can't test.

REMEMBER

Testing predictions from economic theory or logical reasoning is rarely a straightforward procedure. Observed data don't tend to be generated from a controlled experiment, so testing economic theory is challenging in econometric work because of the difficulty in ensuring that the *ceteris paribus* (all else constant) assumption holds. Consequently, in applying econometrics, you need to give considerable attention to the control (independent) variables you include in the analysis to simulate (as closely as possible) the *ceteris paribus* situation.

Applying Statistical Methods to Economic Problems

Most econometrics textbooks assume you've learned all the statistics necessary to begin building econometric models, estimating, and testing hypotheses. However, I've discovered that my students always appreciate a review of the statistical concepts that are most important to succeeding with econometrics. Specifically, you need to be comfortable with probability distributions, the calculation of descriptive statistics, and hypothesis tests. (If your skills are rusty in these areas, make sure you read the material in Chapters 2 and 3.)

Your ability to accurately quantify economic relationships depends not only on your econometric model-building skills but also on the quality of the data you're using for analysis and your capacity to adopt the appropriate strategies for estimating models that are likely to violate a statistical assumption. The data must be derived from a reliable collection process, but you should also be aware of any additional limitations or challenges. They may include, but aren't limited to

- **Aggregation of data:** Information that may have originated at a household, individual, or firm level is being measured at a city, county, state, or country level in your data.
- **Statistically correlated but economically irrelevant variables:** Some datasets contain an abundance of information, but many of the variables may have nothing to do with the economic question you're hoping to address.
- **Qualitative data:** Rich datasets typically include qualitative variables (geographic information, race, and so on), but this information requires special treatment before you can use it in an econometric model.
- **Classical linear regression model (CLRM) assumption failure:** The legitimacy of your econometric approach always rests on a set of statistical assumptions, but you're likely to find that at least one of these assumptions doesn't hold (meaning it isn't true for your data).

REMEMBER

Econometricians differentiate themselves from statisticians by emphasizing violations of statistical assumptions that are often taken for granted. The most common technique for estimating an econometric model is ordinary least squares (OLS), which I cover in Chapter 5. However, as I explain in Chapters 6 and 7, a number of CLRM assumptions must hold in order for the OLS technique to provide reliable estimates. In practice, the assumptions that are most likely to fail depend on your data and specific application. (In Chapters 10, 11, and 12, you see how to identify and deal with the most common assumption violations.)

In the following sections, I describe how familiarity with certain characteristics of your data can help you build better econometric models. In particular, you should pay attention to the structure of your data, the way in which variables are measured, and how quantitative data can be complemented with qualitative information.

Recognizing the importance of data type, frequency, and aggregation

The data that you use to estimate and test your econometric model is typically classified into one of three possible types (for further details on each type, see Chapter 4):

- **Cross sectional:** This type of data consists of measurements for individual observations (persons, households, firms, counties, states, countries, or whatever) at a given point in time.

- **Time series:** This type of data consists of measurements on one or more variables (such as gross domestic product, interest rates, or unemployment rates) over time in a given space (like a specific country or state).
- **Panel or longitudinal:** This type of data consists of a time series for each cross-sectional unit in the sample. The data contains measurements for individual observations (persons, households, firms, counties, states, countries, and so on) over a period of time (days, months, quarters, or years).

REMEMBER

The type of data you're using may influence how you estimate your econometric model. In particular, specialized techniques are usually required to deal with time-series and panel data. I cover time-series techniques in Chapter 12, and I discuss panel techniques in Chapters 16 and 17.

TIP

You can anticipate common econometric problems because certain CLRM assumption failures are more likely with particular types of data. Two typical cases of CLRM assumption failures involve heteroskedasticity (which occurs frequently in models using cross-sectional data) and autocorrelation (which tends to be present in models using time-series data). For the full scoop on heteroskedasticity and autocorrelation, turn to Chapters 11 and 12, respectively.

In addition to knowing the type of data you're working with, make sure you're always aware of the following information:

- **The level of *aggregation* used in measuring the variables:** The level of aggregation refers to the unit of analysis when information is acquired for the data. In other words, the variable measurements may originate at a lower level of aggregation (like an individual, household, or firm) or at a higher level of aggregation (like a city, county, or state).
- **The *frequency* with which the data is captured:** The frequency refers to the rate at which measurements are obtained. Time-series data may be captured at a higher frequency (like hourly, daily, or weekly) or at lower frequency (like monthly, quarterly, or yearly).

WARNING!

All the data in the world won't allow you to produce convincing results if the level of aggregation or frequency isn't appropriate for your problem. For example, if you're interested in determining how spending per pupil affects academic achievement, state-level data probably won't be appropriate because spending and pupil characteristics have so much variation across cities within states that your results are likely to be misleading.

Avoiding the data-mining trap

As you acquire more data-analysis tools, you may be inclined to search the data for relationships between variables. You can use your knowledge of statistics to find models that fit your data quite well. However, this practice is known as *data mining,* and you don't want to be seduced by it!

REMEMBER

Although data mining can be useful in fields where the underlying mechanism generating the outcomes isn't important, most economists don't view this approach favorably. In econometrics, building a model that makes sense and is reproducible by others is far more important than searching for a model that has a perfect fit. I reinforce the importance of building sensible models in Chapter 4 and provide some specific examples of common economic models in Chapter 8.

Incorporating quantitative and qualitative information

Economic outcomes can be affected by both quantitative (numeric) and qualitative (non-numeric) factors. Generally, quantitative information has a straightforward application and interpretation in econometric models.

Qualitative variables are associated with characteristics that have no natural numeric representation, although your raw data may code qualitative characteristics with a numeric value. For example, a U.S. region may be coded with a 1 for West, 2 for South, 3 for Midwest, and 4 for Northeast. However, the assignment of the specific values is arbitrary and carries no special significance. In order to utilize the information contained in qualitative variables, you'll usually convert them into *dummy variables* — dichotomous variables that take on a value of 1 if a particular characteristic is present and 0 otherwise. I illustrate the use of dummy variables as independent variables in an econometric model in Chapter 9.

Sometimes the economic outcome itself is qualitative or contains restricted values. For example, your dependent variable could measure whether or not a firm fails (goes bankrupt) in a given year using various firm characteristics as independent variables. Although standard techniques are sometimes acceptable with qualitative and noncontinuous dependent variables, usually they result in assumption violations and require an alternative econometric approach. Flip to Chapters 13 and 14 to discover the appropriate techniques for situations when your dependent variable isn't continuous.

Using Econometric Software: An Introduction to STATA

Specialized software makes the application of econometric techniques possible for anyone who's not a computer programming genius. Keep in mind that several good software options are available to you and that, as a good economist, you should weigh the cost and benefits of each. Of course, the type of software you ultimately end up working with in your introductory econometrics course depends on what your professor uses for his research or finds to be the easiest to integrate into the course. I rely on STATA extensively in my academic research and use it exclusively in my econometrics courses, but your professor may employ EVIEWS, SAS, or some other program.

Because I find STATA to be the best software, it's what I use exclusively in this book. It provides an excellent combination of a user-friendly interface, consistent structure in syntax, and simple commands to implement all the techniques you learn about in econometrics, and it's available for a variety of platforms or operating systems.

STATA can be used as a *point-and-click* software (like you would use Excel or most other software these days). With point-and-click, you can use the icons and menu bar at the top to execute tasks. However, over time, you're likely to prefer using STATA as a *command-driven* program because it's faster and easier. When used in this manner, you perform tasks by providing STATA with specific syntax on the command line (using lowercase letters for the commands). In this chapter, I explain both methods, but in the later chapters, I rely almost exclusively on the command-driven approach.

The following sections show you some STATA commands that allow you to get started with the software. (Note that I introduce STATA commands as needed in other chapters.)

REMEMBER

My coverage of STATA is not exhaustive. The supporting documentation consists of a User's Guide and several Reference manuals (thousands of pages), so clearly I can't cover every facet of STATA that you may use in econometrics. However, if you run into an obstacle, the manuals are easy to use and provide good examples. With STATA running on your computer, you also have access to the Help menu and online documentation.

Getting acquainted with STATA

In this section of the chapter, I show you how to open data files, log your modifications to data, and save your data files.

Creating and saving STATA datasets

In order to begin doing any exploratory data analysis or econometric work, you need a dataset that's in STATA format (*.dta). If you're downloading data from an online source, you may be able to obtain the data in STATA format. Many econometrics textbooks also give you access to data files in STATA format. In addition, the STATA program is preloaded with examples that you can use to familiarize yourself with the basic commands.

After opening STATA, you can access the sample datasets by selecting **File ➪ Example Datasets…** If you want to open any other dataset that's already in STATA format, select **File ➪ Open** and then choose the file you want to work with. On the command line, you can open a STATA dataset by typing "use *filename*" and hitting return.

If you're inputting data manually or downloading it in a non-STATA format, then you can use one of two methods to read it into STATA:

- Select **File ➪ Import:** This option can be used if the data is in Excel, SAS XPORT, or Text format. You select the appropriate format of your raw data, and then you're prompted to select the file you'd like to import into STATA.
- Select **Data ➪ Data Editor:** This option opens an editor that resembles a spreadsheet. You can paste columns of data into the editor or input data manually.

If you import a dataset that wasn't originally in STATA format, you need to save the dataset in STATA format in order to use it again, particularly if you inputted data through the editor and want to avoid replicating all your efforts. Also, if you made any changes to an existing STATA dataset and want to retain those changes, you need to save the revised dataset. I recommend you select **File ➪ Save As** (or type "save *new filename*" on the command line) and choose a new name for the modified file. That way if you accidentally delete a variable or drop observations, you can always go back to the original data file.

Viewing data

Before you begin doing econometric analysis, make sure you're familiar with your data. After all, you don't want to estimate an econometric model with data that's mostly incomplete or full of errors.

In version 12.1 of STATA, the default setting allows you to open a dataset as large as 64 megabytes (MB) and containing up to 5,000 variables. If your dataset is larger than 64MB, you need to increase the memory allocated to STATA by typing "set memory #m" on the command line, where # is the size of your dataset in MB. Similarly, if your dataset contains more than 5,000 variables, you need to type "set maxvar #" on the command line, with # being the number of variables in your dataset.

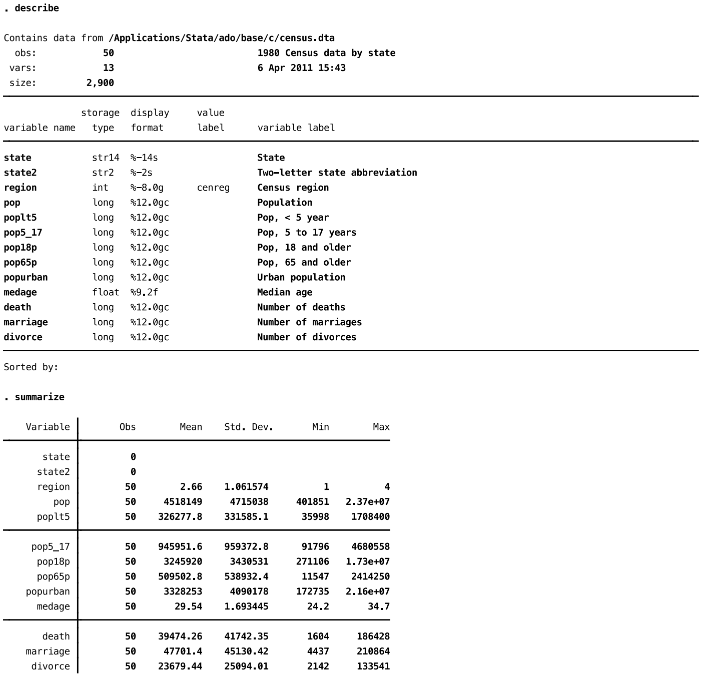
The **Data** tab in the menu bar contains most of the elements you need in order to get acquainted with your data. After opening a STATA dataset, you'll regularly use the following commands:

- Select **Data** ➪ **Describe data** ➪ **Describe data in memory** or type "describe" on the command line and hit return: STATA shows you how many observations and variables are contained in the dataset. In addition, it lists the names and types (numeric or string) of all the variables.
- Select **Data** ➪ **Describe data** ➪ **Summary statistics** or type "summarize" on the command line and hit return: With this command, STATA provides you with basic descriptive measures for all the numeric variables in your dataset. Specifically, you get the number of observations with nonmissing values, mean, standard deviation, minimum value, and maximum value for each variable. ***Note:*** The string variables contain letters, names, or phrases, so no mean or standard deviation can be calculated for them.

In Figure 1-1, I use the "describe" and "summarize" commands to view the fundamental characteristics of my dataset.

The **Data** tab or "describe" and "summarize" commands provide the basic information you use for your econometric analysis. Examine the tables containing the descriptive information and make sure that all the values are sensible. In other words, make sure that the minimum, maximum, and mean values are feasible for each variable in your dataset.

You can also use the "list" command on occasion, but be careful with it because it displays the value for every variable and every observation. In other words, it displays the entire dataset. With a large dataset (thousands of observations and dozens of variables), this list isn't likely to help you find errors unless you refine the list to a specific observation using an "if" statement or by subscripting (I discuss this in the later "Creating new variables" section).



```
. describe

Contains data from /Applications/Stata/ado/base/c/census.dta
  obs:            50                          1980 Census data by state
 vars:            13                          6 Apr 2011 15:43
 size:         2,900
------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
state           str14  %-14s                  State
state2          str2   %-2s                   Two-letter state abbreviation
region          int    %-8.0g      cenreg     Census region
pop             long   %12.0gc                Population
poplt5          long   %12.0gc                Pop, < 5 year
pop5_17         long   %12.0gc                Pop, 5 to 17 years
pop18p          long   %12.0gc                Pop, 18 and older
pop65p          long   %12.0gc                Pop, 65 and older
popurban        long   %12.0gc                Urban population
medage          float  %9.2f                  Median age
death           long   %12.0gc                Number of deaths
marriage        long   %12.0gc                Number of marriages
divorce         long   %12.0gc                Number of divorces
------------------------------------------------------------------------------
Sorted by:

. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       state |         0
      state2 |         0
      region |        50        2.66    1.061574          1          4
         pop |        50     4518149     4715038     401851   2.37e+07
      poplt5 |        50    326277.8    331585.1      35998    1708400
-------------+--------------------------------------------------------
     pop5_17 |        50    945951.6    959372.8      91796    4680558
      pop18p |        50     3245920     3430531     271106   1.73e+07
      pop65p |        50    509502.8    538932.4      11547    2414250
    popurban |        50     3328253     4090178     172735   2.16e+07
      medage |        50       29.54    1.693445       24.2       34.7
-------------+--------------------------------------------------------
       death |        50    39474.26    41742.35       1604     186428
    marriage |        50     47701.4    45130.42       4437     210864
     divorce |        50    23679.44    25094.01       2142     133541
```

Figure 1-1: Examining data two ways in STATA.

Keep in mind that the results section of STATA, by default, displays approximately one page of output. STATA then prompts you with the "-more-" message. Hitting the return key allows you to see an additional line of output, and hitting the spacebar shows another page of output. If you don't want STATA to pause for "-more-" messages, type "set more off" on the command line. Subsequent output is then displayed in its entirety.

Interpreting error messages

If you make a mistake with a command, STATA responds with an error message and code. The error message contains a brief description of the mistake, and the code has the format r(#), where # represents some number. Reading the error message and carefully examining the command that resulted in the error usually helps you arrive at a solution. If not, the codes, known as a return codes, are stored in STATA, and clicking on the code allows you to obtain a more detailed description of the error.

TIP

The outcome of a command can be identified quickly by looking at the colors of the text in the results area (the middle portion of STATA's interface). If you see the color red, it means something has gone wrong and you should correct your mistake before moving on.

Stopping STATA

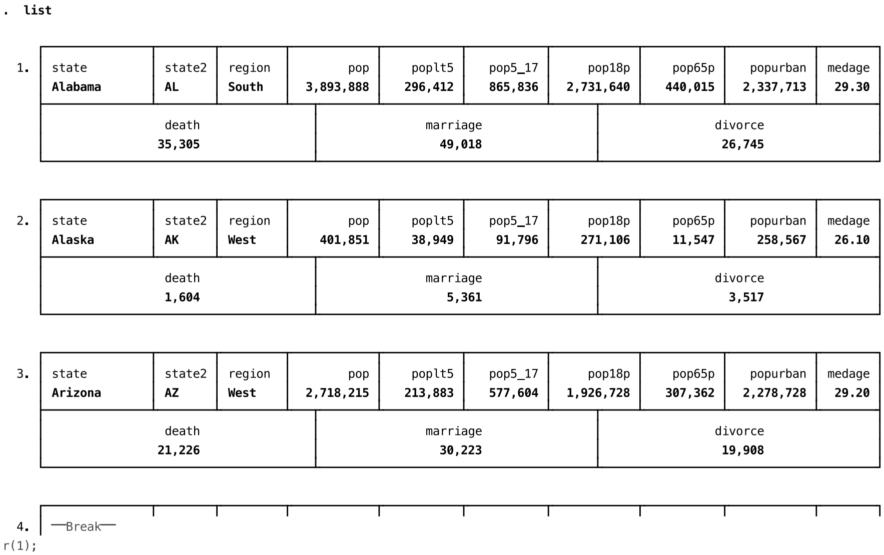
When you occasionally want to terminate a process in STATA, you can just click the **Break** button on the toolbar (right below the menu bar). Stopping STATA may be appropriate if an estimation procedure doesn't converge to a result or you change your mind about the command you'd like to execute and don't want to wait until the process is complete. After you stop STATA, your data remains in memory, and you can continue with any command.

In Figure 1-2, I use the "list" command to see each observation in the dataset. However, after I see a few of the observations, I decide that I don't need to see more observations one by one. I click the **Break** button to stop the command.

Preserving your work

REMEMBER

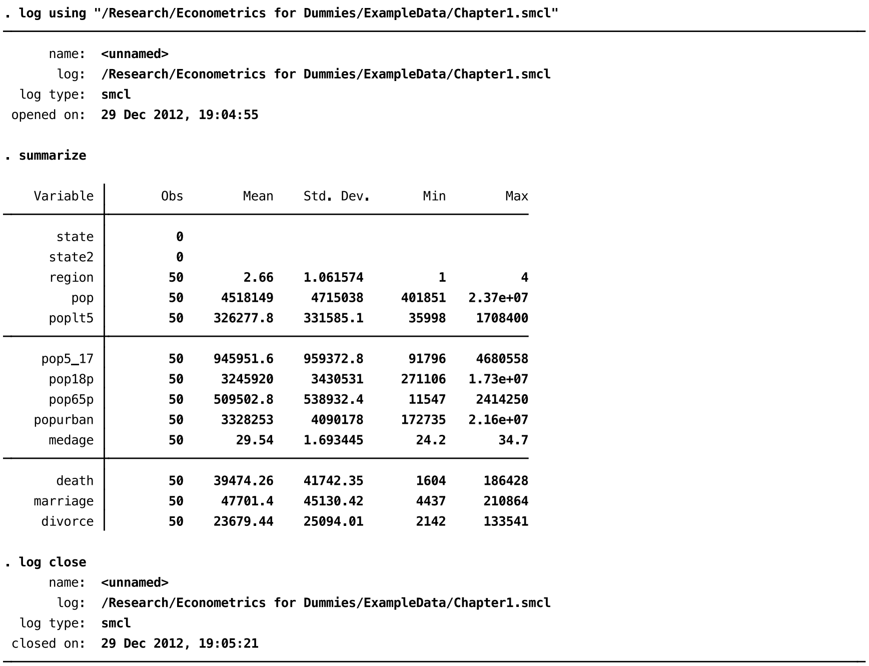
Saving your commands and resulting output in a log file is one of the most essential things you can get into the habit of doing while using STATA. You can do it by selecting **File ⇨ Log ⇨ Begin…** from the menu bar and then assigning the file a name or by typing "log using *filename*" on the command line and hitting return. After you complete the work you want to save, select **File ⇨ Log ⇨ Close** or type "log close" on the command line and hit return. Your log files are given a .smcl file extension.



```
. list

     +--------------------------------------------------------------------------------------------------------+
  1. | state     | state2 | region |       pop |  poplt5 | pop5_17 |    pop18p |  pop65p |  popurban | medage |
     | Alabama   | AL     | South  | 3,893,888 | 296,412 | 865,836 | 2,731,640 | 440,015 | 2,337,713 |  29.30 |
     |--------------------------------------------------------------------------------------------------------|
     |            death            |              marriage               |               divorce              |
     |           35,305            |                49,018               |                26,745              |
     +--------------------------------------------------------------------------------------------------------+

     +--------------------------------------------------------------------------------------------------------+
  2. | state     | state2 | region |       pop |  poplt5 | pop5_17 |    pop18p |  pop65p |  popurban | medage |
     | Alaska    | AK     | West   |   401,851 |  38,949 |  91,796 |   271,106 |  11,547 |   258,567 |  26.10 |
     |--------------------------------------------------------------------------------------------------------|
     |            death            |              marriage               |               divorce              |
     |            1,604            |                 5,361               |                3,517               |
     +--------------------------------------------------------------------------------------------------------+

     +--------------------------------------------------------------------------------------------------------+
  3. | state     | state2 | region |       pop |  poplt5 | pop5_17 |    pop18p |  pop65p |  popurban | medage |
     | Arizona   | AZ     | West   | 2,718,215 | 213,883 | 577,604 | 1,926,728 | 307,362 | 2,278,728 |  29.20 |
     |--------------------------------------------------------------------------------------------------------|
     |            death            |              marriage               |               divorce              |
     |           21,226            |                30,223               |                19,908              |
     +--------------------------------------------------------------------------------------------------------+

     +--------------------------------------------------------------------------------------------------------+
  4. | --Break--
r(1);
```

Figure 1-2: The break action in STATA.

In Figure 1-3, I open a log file, execute a “summarize” command, and close the log file. I can examine the contents of the log file by selecting **File ➪ View...** from the menu bar and then choosing my log file.



```
. log using "/Research/Econometrics for Dummies/ExampleData/Chapter1.smcl"
------------------------------------------------------------------------------------------

      name:  <unnamed>
       log:  /Research/Econometrics for Dummies/ExampleData/Chapter1.smcl
  log type:  smcl
 opened on:  29 Dec 2012, 19:04:55

. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       state |         0
      state2 |         0
      region |        50        2.66    1.061574          1          4
         pop |        50     4518149     4715038     401851   2.37e+07
      poplt5 |        50    326277.8    331585.1      35998    1708400
-------------+--------------------------------------------------------
     pop5_17 |        50    945951.6    959372.8      91796    4680558
      pop18p |        50     3245920     3430531     271106   1.73e+07
      pop65p |        50    509502.8    538932.4      11547    2414250
    popurban |        50     3328253     4090178     172735   2.16e+07
      medage |        50       29.54    1.693445       24.2       34.7
-------------+--------------------------------------------------------
       death |        50    39474.26    41742.35       1604     186428
    marriage |        50     47701.4    45130.42       4437     210864
     divorce |        50    23679.44    25094.01       2142     133541

. log close
      name:  <unnamed>
       log:  /Research/Econometrics for Dummies/ExampleData/Chapter1.smcl
  log type:  smcl
 closed on:  29 Dec 2012, 19:05:21
------------------------------------------------------------------------------------------
```

Figure 1-3: Saving log files in STATA.

Using STATA’s viewer, you can always go back to your log file to see how you modified the data or any statistical estimates you may have previously calculated. You can also copy and paste from your log file to any other file, or you can simply print your log file.

Don’t forget to close your log file when you’re done with the work you want to retain. Otherwise, everything you do in STATA continues to be written to the log file you opened, which may create an unnecessarily huge file.

Creating new variables

After you compile your data, you’ll likely want to create new variables for the analysis. Your econometric model may specify that a variable should be measured in logs, or you may need to use a squared term for a quadratic function

(I cover these types of econometric models in Chapter 8). Your data may also contain qualitative variables that you want to convert into dummy variables (turn to Chapter 9 for guidance on using dummy variables). These examples are just a couple of the many instances in which creating a new variable is in your best interest.

You can create new variables in STATA by selecting **Data ⇨ Create or change data ⇨ Create new variable** from the menu bar or by typing "generate *new variable* = *exp* [*if*] [*in*]" on the command line, where *new variable* is the name you choose to assign the new variable, *exp* specifies how the new variable is created, and the terms in brackets are optional expressions that can be used to restrict the subsample over which you'd like to define the new variable.

A number of arithmetic, relational, and logical operators have been programmed into STATA and can be used to create new variables. You can browse through them in the STATA manuals or the electronic documentation.

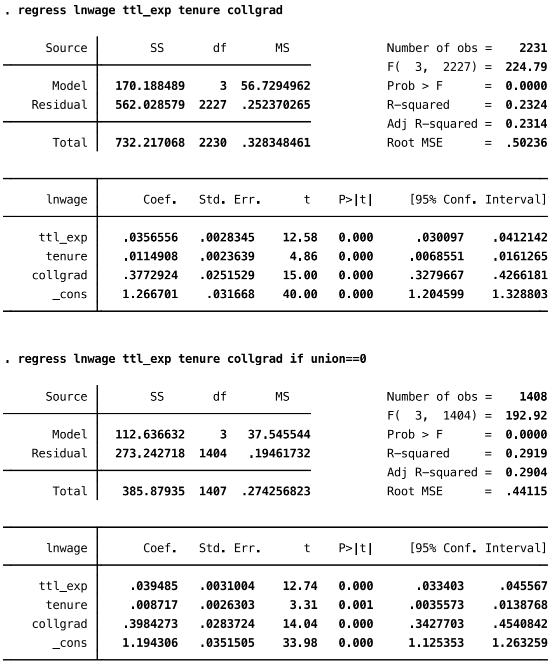
I recommend using the "summarize" command after you create new variables. Doing so allows you to confirm that your new variable doesn't contain errors and that its values are in line with what you intended.

Estimating, testing, and predicting

After you collect your data and create any additional variables necessary for analysis, you're ready to estimate your econometric model and perform hypothesis tests.

The appropriate estimation technique depends on the nature of your econometric model. All the model estimation commands can be found by selecting **Statistics** from the menu bar. If you use the command line, you use similar syntax for all estimation techniques; the syntax is "command variable1 variable2 . . . [*if*] [*in*] [*weight*] [, *options*]" followed by hitting return, where variable1 is the dependent variable in your model.

In Figure 1-4, I estimate a multiple regression model using a sample of workers. The natural log of the hourly wage *(lnwage)* is my dependent variable, and I use years of work experience *(ttl_exp),* years with the same employer *(tenure),* and a dummy variable indicating whether the individual graduated from college *(collgrad)* as my independent variables. I also estimate the same model using the subsample of nonunionized workers.



```
. regress lnwage ttl_exp tenure collgrad

      Source |       SS       df       MS              Number of obs =    2231
-------------+------------------------------           F(  3,  2227) =  224.79
       Model |  170.188489     3  56.7294962           Prob > F      =  0.0000
    Residual |  562.028579  2227  .252370265           R-squared     =  0.2324
-------------+------------------------------           Adj R-squared =  0.2314
       Total |  732.217068  2230  .328348461           Root MSE      =  .50236

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     ttl_exp |   .0356556   .0028345    12.58   0.000      .030097    .0412142
      tenure |   .0114908   .0023639     4.86   0.000     .0068551    .0161265
    collgrad |   .3772924   .0251529    15.00   0.000     .3279667    .4266181
       _cons |   1.266701    .031668    40.00   0.000     1.204599    1.328803
------------------------------------------------------------------------------

. regress lnwage ttl_exp tenure collgrad if union==0

      Source |       SS       df       MS              Number of obs =    1408
-------------+------------------------------           F(  3,  1404) =  192.92
       Model |  112.636632     3   37.545544           Prob > F      =  0.0000
    Residual |  273.242718  1404   .19461732           R-squared     =  0.2919
-------------+------------------------------           Adj R-squared =  0.2904
       Total |   385.87935  1407  .274256823           Root MSE      =  .44115

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     ttl_exp |    .039485   .0031004    12.74   0.000      .033403     .045567
      tenure |    .008717   .0026303     3.31   0.001     .0035573    .0138768
    collgrad |   .3984273   .0283724    14.04   0.000     .3427703    .4540842
       _cons |   1.194306   .0351505    33.98   0.000     1.125353    1.263259
------------------------------------------------------------------------------
```

Figure 1-4: A STATA regression estimation.

STATA also has a number of postestimation commands for hypothesis testing, obtaining residuals, and predicting the dependent variable. You can explore them in the STATA manuals or electronic documentation. However, throughout the book, I also provide several examples of postestimation commands alongside the relevant econometric model estimates.