23
Bootstrap

I also wish to thank the many friends who suggested names more colorful than *Bootstrap*, including *Swiss Army Knife*, *Meat Axe*, *Swan-Dive*, *Jack-Rabbit*, and my personal favorite, the *Shotgun*, which to paraphrase Tukey, "can blow the head off any problem if the statistician can stand the resulting mess."

Bradley Efron[1]

23.1. Introduction

Throughout this book, we have used Monte Carlo simulations to demonstrate statistical properties of estimators. We have simulated data generation processes on the computer and then directly examined the results.

This chapter explains how computer-intensive simulation techniques can be applied to a single sample to estimate a statistic's sampling distribution. These increasingly popular procedures are known as bootstrap methods. They can be used to corroborate results based on standard theory or provide answers when conventional methods are known to fail.

When you "pull yourself up by your bootstraps," you succeed – on your own – despite limited resources. This idiom is derived from *The Surprising Adventures of Baron Munchausen* by Rudolph Erich Raspe. The baron tells a series of tall tales about his travels, including various impossible feats and daring escapes. Bradley Efron chose "the bootstrap" to describe a particular resampling scheme he was working on because "the use of the term bootstrap derives from the phrase *to pull oneself up by one's own bootstrap* . . . (The Baron had fallen to the bottom of a deep lake. Just when it looked like all was lost, he thought to pick himself up by his own bootstraps.)" [Efron and Tibshirani (1993), p. 5].

In statistics and econometrics, bootstrapping has come to mean to resample repeatedly and randomly from an original, initial sample using each

[1] Efron (1979, p. 25).

P1: IYP
0521843197c23 CB962B/Barretto 0 521 84319 7 September 25, 2005 1:15

bootstrapped sample to compute a statistic. The resulting empirical distribution of the statistic is then examined and interpreted as an approximation to the true sampling distribution.

The tie between the bootstrap and Monte Carlo simulation of a statistic is obvious: Both are based on repetitive sampling and then direct examination of the results. A big difference between the methods, however, is that bootstrapping uses the original, initial sample as the population from which to resample, whereas Monte Carlo simulation is based on setting up a data generation process (with known values of the parameters). Where Monte Carlo is used to test drive estimators, bootstrap methods can be used to estimate the variability of a statistic and the shape of its sampling distribution.

There are many types of bootstrapping because there are many ways to resample, and there are a variety of ways to use the bootstrapped samples. The next section introduces the bootstrap by returning to the free-throw shooting example used to explain Monte Carlo simulation. We then apply the bootstrap with regression analysis, using data presented by Ronald Fisher. Section 23.4 demonstrates how the Bootstrap Excel add-in can be used on your own data to obtain bootstrapped SEs. We conclude our introduction to bootstrapping by exploring how the bootstrap can be applied to get a measure of the variability of the R^2 statistic.

23.2. Bootstrapping the Sample Percentage

Workbook: PercentageBootstrap.xls

We introduce the bootstrap with a simple example. Suppose you had a single sample of 100 free throws and computed the percentage made. If you did not know the true, underlying accuracy of the free-throw shooter, your best estimate of the shooter's probability of making a free throw would be the sample percentage made.

Of course, there is variability in the percentage made out of 100 free throws. The standard error of the sample percentage can be estimated via conventional methods by dividing the sample SD (an estimate of the unknown population SD) by the square root of the number of free throws. This is not the exact SE because the true SD is unknown.

With the estimated SE of the sample percentage and taking advantage of the central limit theorem, we can generate confidence intervals and compute P-values. This relies on the sampling distribution being approximately normal.

A bootstrapping approach to the problem of estimating the SE and finding the sampling distribution of the sample percentage treats the original sample as a population from which to sample, with replacement, 100 free throws. By repeatedly sampling 100 free throws from this artificial population, we generate a list of bootstrapped sample percentages. The length of the list is equal to the number of bootstrap repetitions. Each number in the list is the percentage of 100 free throws made from a bootstrapped sample. Just as in a Monte Carlo simulation, the spread in the list approximates the SE of the sampling distribution, and the empirical histogram of the repetitions mirrors the probability histogram of the sample percentage.

The PercentageBootstrap.xls workbook puts these ideas into action. From the *Introduction* sheet, click the [Draw a Sample] button. A new sheet, called *OriginalSample*, appears in the workbook. Columns A and B contain the results of 100 free-throw attempts. The workbook is set up so that the shooter will have a true probability of success between 65 and 75 percent. The best estimate of this unknown probability is the sample percentage. Cell D15 reports the estimated SE using the conventional approach, and cells D17 and D18 display the lower and upper bounds of a 95-percent confidence interval (relying on the normal distribution).

To understand how the bootstrap method works, click the [Draw One Bootstrap Observation] button several times. Each click draws a new observation for the bootstrapped sample (from the 100 free throws in the original sample) and places it in columns H and I. The sampling is done with replacement, and each observation in the original sample is equally likely to be drawn. To obtain a complete bootstrapped sample, we need 100 observations, the same size as the original sample.

Instead of drawing the bootstrapped sample one observation at a time, you can simply click the [Draw One Bootstrap Sample] button to draw 100 observations. Click the [Draw One Bootstrap Sample] button repeatedly. Each click draws a bootstrapped sample. The bootstrapped sample percentage is displayed in cell I1. Each new bootstrapped sample generates a new bootstrapped sample percentage.

A particular observation may appear more than once in a bootstrapped sample, whereas another may not be drawn at all. Cell K1 displays the number of times a particular observation, number 27 in the original sample, appears in the bootstrapped sample. Click the [Draw One Bootstrap Sample] button repeatedly and keep your eye on cell K1. Sometimes observation number 27 does not appear at all, but usually it is drawn at least once. As you repeatedly draw a new bootstrapped sample, you will probably see it appear between zero and three times.

The bounce in the bootstrapped sample percentage is the sampling variation we want to capture. We need to resample repeatedly, keeping track of the sample percentage in each bootstrapped sample. Click the [Bootstrap Simulation] button to access a new sheet, *Bootstrap*, from which a bootstrap analysis can be carried out.

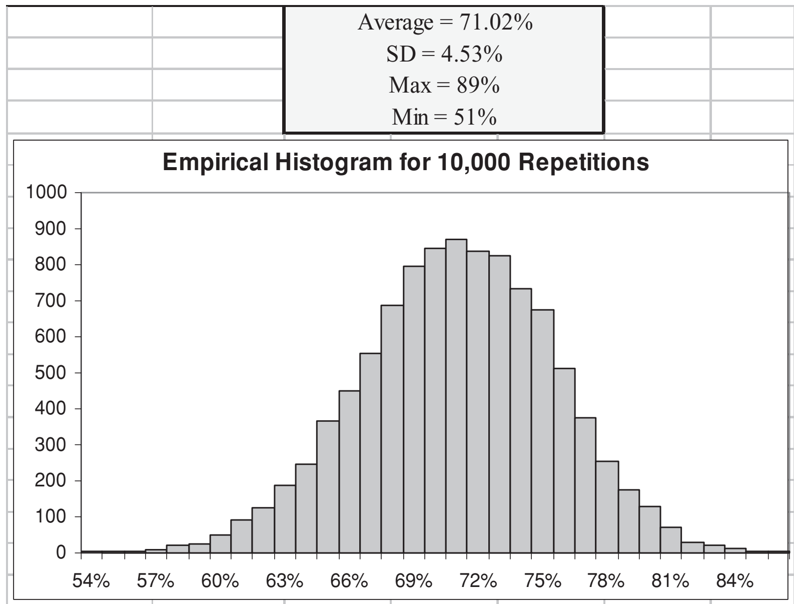

Figure 23.2.1. Bootstrapping the sample percentage.
Source: [PercentageBootstrap.xls]Bootstrap.

When we drew a sample, our Original Sample had 71 free throws made. Using conventional methods – that is, the sample SD divided by the square root of *n* (the number of observations)– the estimated SE is 4.56 percentage points. Our bootstrap results are displayed in Figure 23.2.1. The bootstrapped SE, the estimate of the exact SE based on bootstrapping, is 4.53 percentage points. How did you do?

When estimating the SE of the sample percentage of 100 free throws, the bootstrap and conventional approaches are in substantial agreement. This makes sense because both are using the same information from the original sample. The conventional approach uses the sample SD to construct the estimated SE via a formula. The bootstrap treats the sample as a population and resamples from it. The bootstrap converges to the conventional result as the number of repetitions increases.

The two methods differ in estimating the sampling distribution itself. Instead of relying on the normal distribution to approximate the unknown shape of the sampling distribution, the bootstrap uses the empirical histogram from the simulation as an estimate of the sampling distribution. Brownstone and Valleta clearly stake out the issues:

This bootstrap method described above will only give accurate estimates if the original sample is large enough to reflect the true population accurately. The traditional analytic approach approximates the sampling distribution by a normal distribution

P1: IYP
0521843197c23 CB962B/Barretto 0 521 84319 7 September 25, 2005 1:15

centered at the sample mean with variance equal to the sample variance. This traditional approximation requires that the sample be large enough for the central limit theorem to apply to the sample mean. If the sample size is small and the true population is not normally distributed, then the bootstrap approximation should be more accurate. Brownstone and Valleta (2001, p. 130).

In other words, the bootstrap will do a better job of answering questions that involve the shape of the sampling distribution when its profile is not normal. Suppose, for example, that we wanted to know the chances that a 95-percent free-throw shooter will make 16 or less out of 20 free throws. The standard approach will fare badly because the sampling distribution of the sample percentage for this case is not very normal.

Summary

This section has introduced the bootstrap by showing how it can be used to estimate the SE and sampling distribution of the sample percentage. By sampling with replacement from an original sample, we generate an artificial sample. We use the artificial, or bootstrapped, sample to compute a statistic of interest. By repeating this procedure many times, we obtain an approximation to the sampling distribution of the statistic. The next section shows how the bootstrap can be applied to regression analysis.

23.3. Paired XY Bootstrap

Workbook: PairedXYBootstrap.xls

In the 1940s, "although digitalis had been a standard medication for heart disease for more than a century, there were still no reliable methods for evaluating its potency. Biological assays (bioassays) were performed on frogs, pigeons, and cats, but none were totally satisfactory" (Scheindlin, 2001, p. 88). In too high a dose, digitalis is deadly. Doctors needed to know the right dosage for different patients. Experiments on laboratory animals were undertaken in an attempt to determine toxicity levels.

In 1947, Ronald Fisher published an article that analyzed the data from digitalis assays from 144 cats. The data set had the sex, heart weight (in grams), and body weight (in kilograms) of each cat. Fisher's Table 1 (see Figure 23.3.1) displayed salient summary characteristics.

Fisher noted that the "heart as a fraction of the entire body" was remarkably similar for female and male cats. Could the optimal digitalis dose be determined simply as a function of the patient's body weight? After all, if given body weight, heart weight is simply a constant fraction, then from body weight we can infer heart weight and administer the correct dosage.

Fisher's Original TABLE 1		
	Females	Males
Number	47	97
Total Body Weight	110.9 Kg.	281.3 Kg.
Total Heart Weight	432.5 g.	1098.2 g.
Heart as fraction of entire body	.3900%	.3904%

Figure 23.3.1. Fisher's cat data for digitalis study.
Source: [PairedXYBootstrap.xls]Data.

Unfortunately, closer inspection revealed that the correspondence between body and heart weight broke down. Fisher reported that the slope coefficients from regressions of heart weight on body weight for each sex differed: "namely .2637% for females and .4313% for males." A 1-kg increase in body weight led, on average, to a 4.313-g increase in heart weights for males but only a 2.637-g increase in heart weights for females. Figure 23.3.2, which Fisher did not include in his published article, shows the two individual regressions.

Fisher suspected that male and female cats in his sample had different relationships between body and heart weight. The next step required a decision on whether this difference was real. It could be that the difference observed in the sample was simply due to chance error in the selection of the particular cats chosen for the study.

Fisher used the data to illustrate how the analysis of covariance method can be used to determine if the coefficient estimates from the two regressions are statistically significantly different from each other. He concluded that "the close agreement between the sexes in the average percentage of the body taken up by the heart seems to mask a real difference in the heart weight to be expected for a given body weight" (Fisher 1947, p. 68).

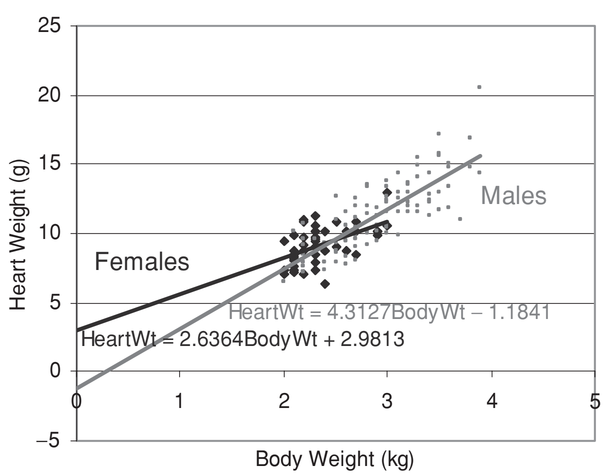

Figure 23.3.2. Individual regressions on female and male cats.
Source: [PairedXYBootstrap.xls]Data.

	Dependent Variable: Heart Weight (g)			
	Model 1	Model 2	Model 3	Model 4
	Females	Males	Both	Both
Intercept	2.981	–1.184	–0.497	–1.184
	(1.485)	(0.998)	(0.868)	(0.925)
Body Weight (kg)	2.6364	4.3127	0.082	4.313
	(0.625)	(0.340)	(0.304)	(0.315)
Female			4.076	4.165
			(0.295)	(2.062)
Female*BodyWeight (kg)				–1.676
				(0.837)
N	47	97	144	144
RMSE	1.162	1.557	1.457	1.442
R^2	0.28	0.63	0.65	0.66
	SEs in parentheses			

Figure 23.3.3. Uncovering gender differences via regression.
Source: [PairedXYBootstrap.xls]Data.

Although he chose to use the analysis of covariance method, Fisher could have explored the effect of sex on the relationship between heart and body weight with multiple regression analysis. Figure 23.3.3 compares the results from four models. Models 1 and 2 treat females and males separately. Model 3 is a multivariate model that forces the slopes to be equal but allows the intercepts to be different for female and male cats. The interaction term, Female*BodyWeight (kg), in Model 4 relaxes the restriction on the slopes. The coefficient on the interaction term has a *P*-value of 4.7% when testing the null that it is 0. We would conclude that the slopes are statistically significantly different from each other.

The hypothesis test of the null that the coefficient on Female*Body Weight (kg) is zero relies heavily on the estimated SE. In turn, the computation of the estimated SE is based on the estimate of the spread of the errors, the RMSE. Ordinary least squares regression requires homoskedastic errors and uses a single number to estimate the spread of the errors. Unfortunately, the RMSEs from the individual regressions are worrisome because it looks like the male cats have much greater spread around the regression line (RMSE = 1.557) than the female cats (RMSE = 1.162). This is evidence of heteroskedasticity. Fisher was aware of this problem and ended the paper with the following observation: "It may be noted that the estimated variance of heart weight for given body weight in males, 2.424 g.^{2}, is considerably greater than the value for females, 1.351 g.^{2} The greater residual variance for males possibly was related to their larger size. The heaviest female weighed 3.0 Kg. while nearly 40 percent of the males exceeded this weight" (Fisher 1947, p. 68).

Heteroskedastic errors pose serious problems for OLS regression analysis. Although estimates remain unbiased, OLS is no longer the best linear unbiased estimator, and the reported OLS estimated SEs cannot be trusted.

P1: IYP
0521843197c23 CB962B/Barretto 0 521 84319 7 September 25, 2005 1:15

Regression Statistics for Heart Weight (g)				
Number of observations	144	Number of missing observations = 0		
Mean of Dep Var	10.631			
RMSE	1.442			
Coefficient Estimates				
Variable	Estimate	SE	Robust SE	
Intercept	−1.184	0.925	1.166	
Female	4.165	2.062	1.854	
Female*BodyWeight (kg)	−1.676	0.837	0.735	
Body Weight (kg)	4.313	0.315	0.414	

Figure 23.3.4. Robust SEs of regression coefficients.
Source: [PairedXYBootstrap.xls]Data.

Because we use the estimated SE to compute the *t*-statistic and *P*-value, the hypothesis test conducted on the Female*Body Weight (kg) coefficient is flawed.

The conventional solution is to estimate SEs that are robust to the presence of heteroskedasticity. Figure 23.3.4 shows the results of this approach (using the OLS Regression add-in described in detail in the chapter on heteroskedasticity). The estimated SE falls by 12 percent from 0.837 to 0.735. The *P*-value on the null that the slope is zero falls by half from 4.7 to 2.4%.

Another approach to estimating the SE is to use the bootstrap. For regression analysis, several different resampling schemes are possible. We will demonstrate the most popular one, called paired XY or case resampling. Using the original sample with 144 observations, three independent variables (Female, Female*Body Weight, and Body Weight), and the dependent variable (Heart Weight), we generate each bootstrap sample by randomly drawing 144 rows from the data.

Scroll over to column AK in the *Data* sheet of PairedXYBootstrap.xls. Click the Draw One Bootstrap Observation button several times. Each click draws a new observation for the bootstrapped sample and places it in columns AK, AL, AM, and AN. Each click takes an entire row or record (which accounts for the names paired XY or case resampling). The sampling is done with replacement, and each observation in the original sample is equally likely to be drawn. To get a complete bootstrapped sample, simply click the Draw One Bootstrap Sample button to draw 144 observations. Click the Draw One Bootstrap Sample button repeatedly. Each click draws a bootstrapped sample. Regression results for the artificial bootstrapped sample are displayed in cells AP2:AS6 of the *Data* sheet. Each new bootstrapped sample generates a new bootstrapped regression line. The cell highlighted in yellow (AQ2) is the coefficient for the interaction term.

The bootstrapped SE of the slope of Female*Body Weight (kg) is the standard deviation from the list of coefficients generated by repeatedly resampling. The *Bootstrap* sheet enables you to run your own analysis by simply clicking the Run Bootstrap button. Figure 23.3.5 shows our results.

P1: IYP
0521843197c23 CB962B/Barretto 0 521 84319 7 September 25, 2005 1:15

Slope Female*BodyWt Estimates		Average - Original Sample Slope Female*BodyWt Estimate
Average	−1.6903	−0.0140
SD	0.7224	
Max	0.9250	
Min	−4.7243	

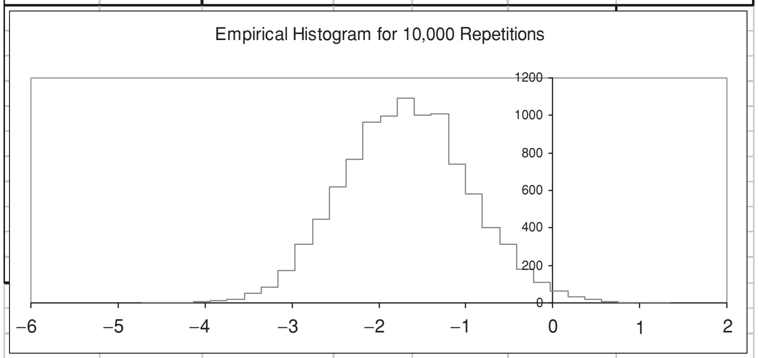

Figure 23.3.5. Bootstrapping the interaction term.
Source: [PairedXYBootstrap.xls]Bootstrap.

The bootstrapped SE, the spread of the 10,000 bootstrapped coefficients, is about 0.72 or 0.73. This agrees with the estimated SE via robust methods, 0.735. By resampling the entire row, or case, the paired XY bootstrap correctly handles the heteroskedasticity.

To construct a confidence interval or conduct a test of significance via the bootstrap, we have the possibility of two approaches. First, one can simply use the bootstrapped SE as the estimated SE in a conventional computation. For example, for a hypothesis test of the null that the coefficient on the interaction term is zero, we use the bootstrapped SE to compute the *t*-statistic:

$$\frac{\text{observed} - \text{expected}}{\text{estimated SE}} = \frac{-1.676 - 0}{0.722} = -2.32.$$

This *t*-stat produces a *P*-value of about 2.2 percent.

There is an alternative to marrying the SE generated via the bootstrap to the conventional approach. By directly using the bootstrapped approximation to the sampling distribution, we can compute confidence intervals and conduct hypothesis tests. A 95-percent confidence interval for the interaction term coefficient is simply the 2.5th to the 97.5th percentile of the 10,000 bootstrap repetitions. Scroll over to column AJ of the *Bootstrap* sheet to see that this interval is from roughly −3.0 to −0.2. Because the interval does not cover 0, you would reject the null that the true parameter value is 0.[2]

[2] Efron and Tibshirani (1993) discuss the connection between confidence intervals and hypothesis tests. The simple approach to bootstrapping canfidence intervals presented here, the percentile method, is not used very often. For a review of better alternatives, see DiCiccio and Efron (1996).

P1: IYP
0521843197c23 CB962B/Barretto 0 521 84319 7 September 25, 2005 1:15

It appears Fisher was right. There is a statistically significant difference in the relationship between body and heart weight for male and female cats. Using the usual estimated SEs from OLS, however, is an inappropriate way of obtaining the variability in the estimated coefficients because heteroskedasticity is present. Robust SE methods and the bootstrap are two alternative, better approaches.

Summary

In the previous section, we bootstrapped the SE of the sample percentage by generating an artificial sample, finding the sample percentage for the artificial sample, and repeating the procedure many times. This section has done the same thing. From an original sample, we generated a pretend sample, ran a regression on the pretend sample, and repeated the procedure 10,000 times. The heart of bootstrapping is to generate artificial samples and construct the same statistic on each sample as the statistic of interest in the original sample.

In both examples thus far, the spreadsheet has been set up for you. Can you run a bootstrap analysis on your own data? Yes, you can, and the next section shows you how.

23.4. The Bootstrap Add-In

Workbooks: PairedXYBootstrap.xls; Bookstrap.xla (Excel add-in)

The previous sections introduced bootstrapping using workbooks especially designed for that purpose. This section shows how to use an Excel add-in packaged with this book that enables you to run a bootstrap from any Excel workbook. Thus, the add-in allows you to use bootstrapping methods on your own data and your own statistic of interest.

The first step is to install the Bootstrap add-in. The software is in the Basic Tools\ExcelAddIns\Bootstrap folder. Open the Bootstrap.doc file in that folder for instructions on how to install and use the add-in.

Having installed the Bootstrap.xla file, open the PairedXYBootstrap.xls workbook to test drive the Bootstrap simulation add-in. The *Female* sheet shows the OLS estimated SE on Body Weight is about 0.625. Let us use the Bootstrap add-in to find the Paired XY bootstrap SE of Body Weight.

Begin by inserting a sheet into the workbook (Insert: Worksheet) and renaming it *BootFemale* and then go to the *Data* sheet and copy the body and heart weight data for the female cats (cell range C1:D48) to the A1:B48 range of the *BootFemale* sheet. We use the data in the *BootFemale* sheet as our original sample and the same range as the place in which we will write our bootstrapped resamples. This will destroy the original sample in

P1: IYP
0521843197c23 CB962B/Barretto 0 521 84319 7 September 25, 2005 1:15

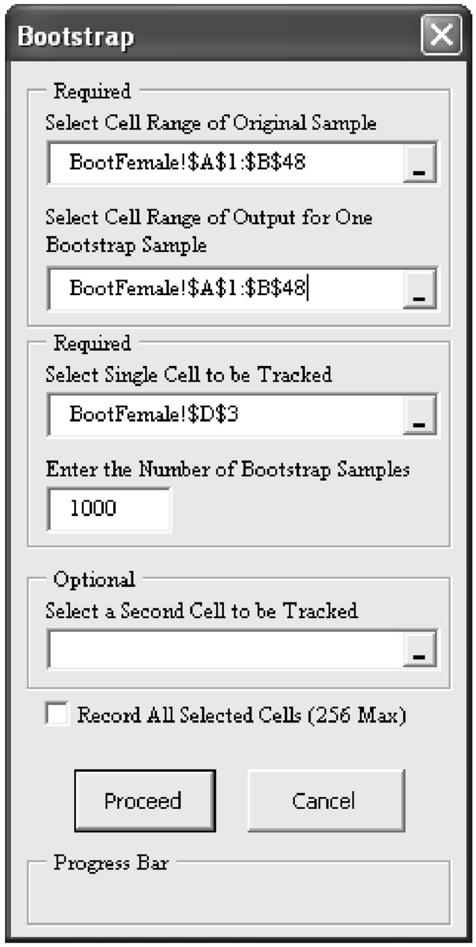

Figure 23.4.1. Preparing to run a bootstrap.

the *BootFemale* sheet, but we have it in the *Data* sheet and thus this is not a problem.

We need, however, to compute the statistic of interest (the OLS estimated SE of Body Weight) for each bootstrapped sample. We can use Excel's LINEST function for this. In the *BootFemale* sheet, select a 5 × 2 cell range and use LINEST to regress Heart Weight on Body Weight. With the data in the *BootFemale* sheet in cells A1:B48, the LINEST formula should look like this: "= LINEST(B2:B48,A2:A48,1,1)." The LINEST results (especially the OLS estimated SE for Body Weight) should be exactly equal to the regression results in the *Female* sheet.

With LINEST available to recompute the slope coefficients as we repeatedly put down new samples in the worksheet, we are ready to bootstrap. Execute Tools: Bootstrap . . . to bring up the bootstrap dialog box.

P1: IYP
0521843197c23 CB962B/Barretto 0 521 84319 7 September 25, 2005 1:15

Summary Statistics			Notes
Average	2.634		**1000 repetitions**
SD	0.6082		
Max	4.911		
Min	0.940		

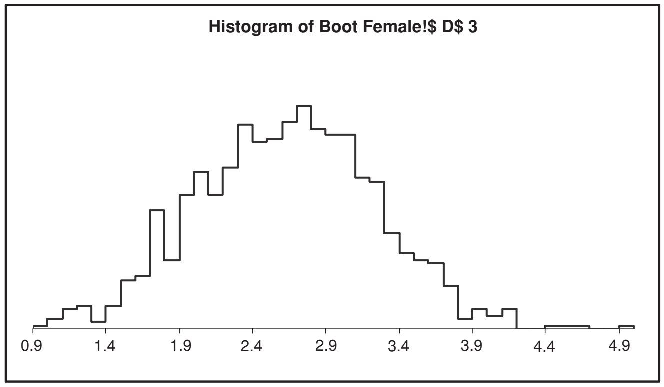

Figure 23.4.2. Results from the bootstrap add-in for the heart weight coefficient.

Enter the same cell range for the Original and Bootstrap Sample input boxes and select the coefficient on Body Weight as the cell to be tracked. Figure 23.4.1 shows how the dialog box should look. The BootFemale! $A$2:$B$48 range contains the data, and we selected cell D3 as the tracking cell because we put Excel's LINEST array function in cells D3:E7, which reports the slope coefficient in cell D3. We obtain a bootstrapped approximation of the slope coefficient's sampling distribution by repeatedly resampling and keeping track of the slope coefficient from each bootstrapped sample.

When you click the Proceed button, the add-in immediately warns you that the original sample data will be overwritten. The Bootstrap add-in reads the original sample, samples from it (with replacement), and then writes the bootstrap sample (temporarily) to the spreadsheet. It records the tracked cell and then repeats this procedure for as many repetitions as you request. Because the original sample is used as the place in which bootstrapped samples are written, a warning is issued. In this case, we can safely proceed because the original female cat data is in the *Data* sheet.

When the bootstrap simulation finishes its last repetition, a worksheet is added to the workbook that displays the first 100 repetitions along with summary statistics and a histogram of the complete results (see Figure 23.4.2).

For female cats, the paired XY bootstrapped SE and OLS estimated SE on Body Weight are almost the same. This is not true for the male cats (as the *Q&A* sheet in the PairedXYBootstrap.xls workbook asks you to show).

P1: IYP
0521843197c23 CB962B/Barretto 0 521 84319 7 September 25, 2005 1:15

Summary

You may have noticed that the bootstrap built into the workbook is much faster. Unlike the simulations in the *Bootstrap* sheet, the Bootstrap add-in spends a great deal of time writing each sample to the spreadsheet. Using this add-in on a large data set may be impractical (although the authors have let the Bootstrap add-in run over night).

The Bootstrap add-in is ideal, however, for exploring problems on your own. Any statistic you can compute on the spreadsheet, no matter how complicated, can be bootstrapped. The next section shows how to apply bootstrap methods to a statistic for which no conventional method exists for estimating its sampling distribution.

23.5. Bootstrapping R^2

Workbook: BootstrapR2.xls

In this section, we apply the bootstrap to a statistic for which there is no standard analytical means of estimating its variability. We also introduce a new resampling scheme called the residuals bootstrap.

The coefficient of determination, commonly abbreviated and reported simply as R^2, is often used as a measure of the overall goodness of fit of a regression. This coefficient ranges from 0 to 1: 0 signifies that the regression explains none of the observed variance in the dependent variable, and 1 denotes a perfect fit. Chapter 5 explains the R^2 statistic in detail and shows how it is calculated. Excel reports R^2 through its Data Analysis: Regression add-in and in the third row and first column of the LINEST array function.

Like the sample slope, estimated SE, and other sample-based statistics, R^2 is a random variable. If you draw a new sample, a new R^2 will result. Ohtani (2000) points out that the sampling properties of R^2 have been investigated. Researchers, however, rarely, if ever, report a measure of the precision of the R^2 value because the sampling distribution of R^2 is complex and depends on the particular values of the X variables. Thus, although we know R^2 is a random variable, without an SE, confidence intervals and hypotheses tests using R^2 are simply ignored.

The bootstrap offers a way to estimate the SE of R^2 and its sampling distribution. The bootstrap, in this case, is conducted by repeatedly resampling from the original sample and keeping track of the R^2 of each artificial sample. Just like any other sample-based statistic, we can approximate the sampling distribution of R^2 via the empirical histogram generated by the bootstrap simulation and use the SD of the bootstrapped R^2 values as an estimate of the exact SE of R^2.

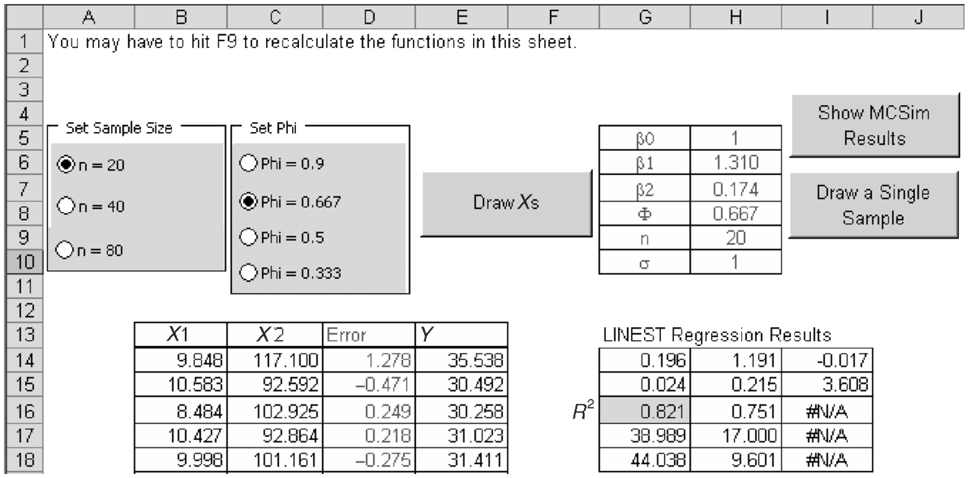

Figure 23.5.1. The data sheet.
Source: [BootstrapR2.xls]Data

Open the Excel workbook BootstrapR2.xls and go to the *Data* sheet. Both the Monte Carlo Simulation and Bootstrap add-ins will be applied to this workbook, and so you need to have them available.

Let us begin with a tour of the *Data* sheet, a portion of which is displayed in Figure 23.5.1. Hit F9 to recalculate the sheet and confirm that R^2 is a random variable.

The data generation process meets all of the classical model's requirements. The *X's* are fixed in repeated sampling (and thus do not change when you hit F9); the errors are independently and identically distributed (and, in addition, drawn from a normal distribution); and each Y is generated by $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$.

The *Data* sheet allows you to control two crucial parameters, the sample size and Φ, by clicking on the buttons. The Greek letter Φ is the parent coefficient of determination. This parameter controls the position and shape of the sampling distribution of R^2. In Figure 23.5.1, and on the spreadsheet in cell G16, notice that Φ (set at 0.667) does not equal the R^2 obtained from the 20 observation sample. This is due to chance error, which is also responsible for the deviation of the sample coefficients (in the first row of the LINEST Regression Results table) from their respective parameter values (the betas in cells H5:H7).

Unlike the sample slope coefficients, whose expected value is equal to the parameter value, R^2 is a biased estimator of Φ. The R^2 statistic is consistent, however, and thus, as the sample size increases, its expected value does converge to its parent parameter value (and the SE converges to 0). You can quickly get a sense of the sampling distribution of R^2 by running a Monte Carlo simulation. Execute Tools: MCSim . . . and select cell G16 as the tracking variable. The average of your Monte Carlo repetitions is an approximation to the expected value, and the SD is an estimate of the exact SE. Your results should be similar to the *MCSimN20Phi0.667* sheet (available by clicking the Show MCSim Results button). Note the bias of R^2 as an estimator of Φ (the average of the 10,000 R^2 values is not close to 0.667) and the nonnormal shape of the histogram.

Now that the properties of the sampling distribution for R^2 for this data generation process are known, we are ready to proceed to the bootstrap. Instead of using the paired XY Bootstrap, we introduce a different resampling scheme. The residuals bootstrap uses the residuals as a stand-in for the errors and produces a bootstrapped sample by shuffling the residuals and creating a bootstrapped Y observation according to the equation

$$\textit{Bootstrapped } Y = b_0 + b_1 X_1 + b_2 X_2 + \textit{residual.}$$

Note that the coefficients are not the β's (because the true parameter values are unknown) but the original sample-estimated coefficients.

Some preparatory work is needed to run the residuals bootstrap, but we have set up the spreadsheet for you. Click the Draw a Single Sample button in the *Data* sheet to obtain an Original Sample and regression results. Click on the Y data cells in column AE to see that the cells contain numbers (not formulas) that represent a single realization from the data generation process. Column AD is blank because you cannot observe the errors.

Cell range AG14:AI18 of the *Data* sheet reports the regression results for your Original Sample. Cell AG16 displays the R^2 value for which we want to find the SE. In column AK, we have computed the residual for each observation. Click on cell AK14 to see the usual actual minus predicted formula for the residual.

The data next to the residuals column are labeled "Adj residuals." By multiplying the residuals by an adjustment factor, we improve the performance of the bootstrap.[3] The Adj Residuals represent the errors and are our artificial population. By sampling with replacement from the Adj Residuals, we can create artificial dependent variables and bootstrapped regression results. Click on one of the Boot Y cells in column AQ and examine the formula. It uses the Original Sample coefficients along with a randomly sampled Boot Residual to form Boot Y.

We will use the Bootstrap add-in to write the Boot Residuals in column AP and track the R^2 for each bootstrapped sample in cell AS16. Figure 23.5.2 shows how the Bootstrap add-in should be configured.

Click Proceed to obtain a bootstrap estimate of the variability of R^2 and the shape of its sampling distribution. You now have Monte Carlo and Bootstrap

[3] For more on rescaling the residuals, see Wu (1986, p. 1281).

P1: IYP
0521843197c23 CB962B/Barretto 0 521 84319 7 September 25, 2005 1:15

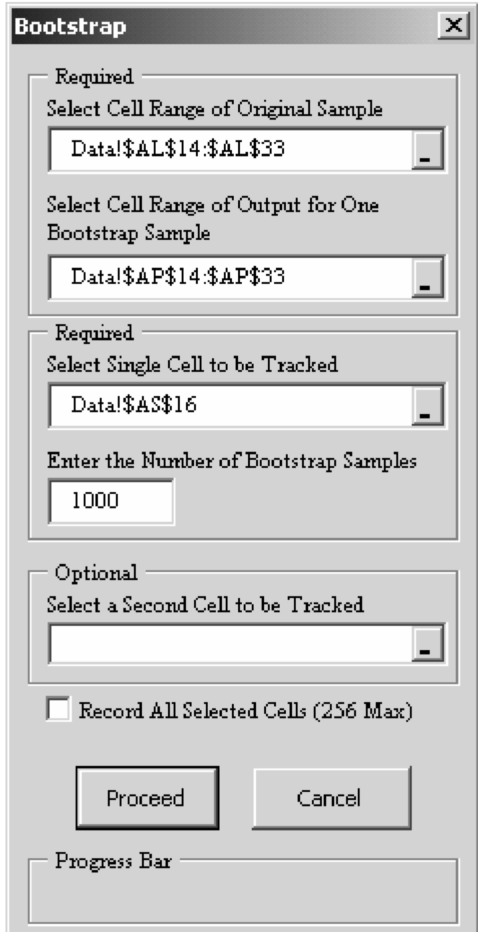

Figure 23.5.2. Setting up the bootstrap.
Source: Bootstrap.xla add-in.

simulation results. It is time to figure out what all of this means. Figure 23.5.3 compares the Monte Carlo with the Bootstrap for $n = 20$ and $\Phi = 0.667$.

The Monte Carlo results, on the left, are a good approximation to the true sampling distribution of R^2. The average of the 10,000 repetitions is 0.706, which is close to the exact expected value (reported by Ohtani) of 0.7053. Similarly, the SD of the 10,000 repetitions, 0.0945, is a good approximation to Ohtani's exact SE of 0.0955. The Monte Carlo simulation is based on knowing the data generation process and simply repeating it and directly examining the results. Your Monte Carlo results should be quite close to ours. It should not be surprising that the Monte Carlo with 10,000 repetitions does a good job of reflecting the true sampling distribution.

The Bootstrap results, the right panel in Figure 23.5.3, are not as good as the Monte Carlo results. With 1,000 bootstrap repetitions, we had an average R^2 of

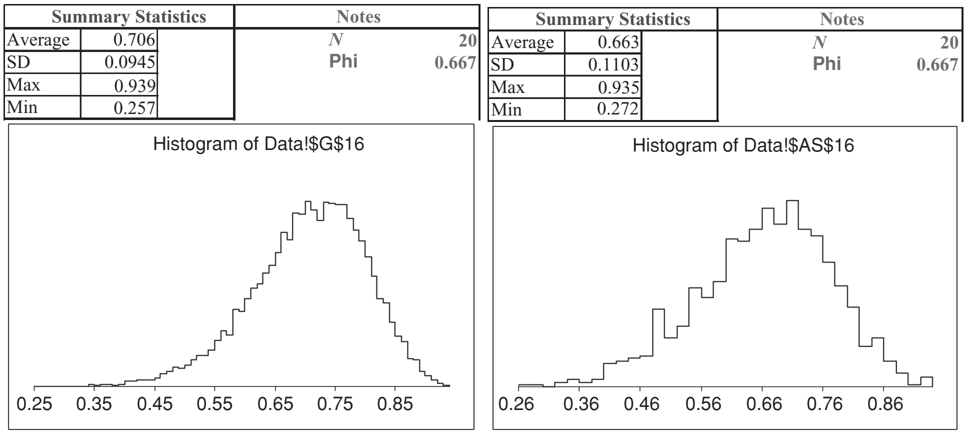

Figure 23.5.3. Monte Carlo and bootstrap simulation results.
Source: [BootstrapR2.xls]Data.

0.663 with an SD of 0.1103. Your bootstrap results may be markedly different from ours (available by clicking the Show Bootstrap Results button). To understand the inferiority of the bootstrap compared with the Monte Carlo, remember that the latter is based on knowing and running the true data generation process. The bootstrap, however, takes one Original Sample – one realization of the DGP – and treats it as a population from which to resample. The bootstrap relies on the premise that the Original Sample will closely mirror the population. The sample size, however, is merely 20 in this case, and so it is quite possible that the Original Sample differs substantially from the true population.

In fact, seen in this light, it is actually quite remarkable that the bootstrap does as well as it does. After all, the bootstrapped and Monte Carlo sampling distributions are reasonably similar, and our bootstrap approximate SE (0.1103) is not that far off the true mark (0.0955).

Of course, to run a full test of the bootstrap, we would have to nest simulations. In other words, take an Original Sample, bootstrap it (like we did), then take another Original Sample, bootstrap it, and repeat this many times. The Advanced Thinking button allows you to do exactly this, but Ohtani (2000) has done the hard work for us. His experiments show that the Residuals Bootstrap has an expected value of 0.7089 with a spread of 0.0899. This shows that, for this case, the Residuals Bootstrap does a good job of approximating the sampling distribution of R^2.

Once a bootstrap approximation of the variability of the statistic has been obtained, we have two options: (1) use the Bootstrapped SE in conventional ways to construct confidence intervals and conduct tests of significance or (2) use the bootstrapped values themselves for these purposes. Note that we are using the bootstrap to estimate the variability of R^2, not the statistic itself.

Could we have used the paired XY instead of the residuals bootstrap on this problem? Yes, and the *Q&A* sheet in BootstrapR2.xls invites you to do so. Remember that, unlike Fisher's cat data, the DGP in the BootstrapR2.xls workbook exactly follows the classical econometric model. If you know that the errors are identically, independently distributed, then the Residuals Bootstrap is appropriate. On the other hand, if the DGP is based on sampling X and Y from a population, then use the paired XY Bootstrap. In general, the bootstrap procedure adopted should mimic the DGP as closely as possible.

Unlike the paired XY Bootstrap, if the residuals bootstrap is applied to Fisher's cat data (in PairedXYBootstrap.xls), you will not correctly estimate the sampling distribution. You could use a modified residuals bootstrap, tying the size of the residual to whether the cat was male or female.

As Efron and Tibshirani make clear, "perhaps the most important point here is that bootstrapping is not a uniquely defined concept" (Efron and Tibshirani 1993, p. 113). In other words, within the realm of "resample from an original sample," there are a great many possibilities in the resampling scheme. Research in bootstrapping methods focuses on the properties of alternative resampling plans.

Summary

Unlike previous sections in this chapter where we used bootstrapping methods to reproduce results obtained with conventional techniques, this section showed how the bootstrap can be used to estimate the variability of R^2, a statistic with a sampling distribution whose analytical solution is beyond the reach of traditional statistical practice. This example also allowed us to introduce the idea that there is more than one way to resample. The next section concludes our introduction to the bootstrap by highlighting a few of the points in the debate about the role of bootstrap methods.

23.6. Conclusion

The heart of the bootstrap is not simply computer simulation, and bootstrapping is not perfectly synonymous with Monte Carlo. Bootstrap methods rely on using an original sample (or some part of it, such as the residuals) as an artificial population from which to randomly resample.

Because the bootstrap utilizes resampling, advances in computing power have facilitated the development of the bootstrap. Bradley Efron is recognized as the inventor of the bootstrap – not because he was the first to conceive of replacing an unknown population with a single sample but because he realized that the explosion in computing would permit a wide variety of resampling schemes.

The method, however, is still in its infancy, and many questions remain unanswered.

> Grand claims sometimes have been made for bootstrap analysis. For instance, Efron and Tibshirani (1993) and Vinod (1998) envision the bootstrap as part of a strategy to find universally applicable methods for estimation and inference, which can be implemented with very little effort or analysis by researchers. This vision is tempting, especially given the ease and speed with which bootstrap estimates for many models can be obtained using modern desktop computers. However, Manski (1996) argues that this vision is flawed due to the inherent ambiguity of statistical theory in comparing alternative estimation procedures.
>
> Brownstone and Valleta (2001, p. 139)

The fundamental requirement of the bootstrap is that the resampling be faithful to the data generation process. This can be difficult to do in practice. Consider the two bootstrap methods used in this chapter: paired XY and residuals bootstraps. These are two of many possible resampling schemes. The paired XY Bootstrap handled the heteroskedasticity in Fisher's cat data, but it is not always clear which resampling strategy is best for a particular case.

But critics have not been able to slow the advance of bootstrap methods. Modern data analysis software includes commands for bootstrapping, and the latest research papers report bootstrap results. Econometrics textbooks increasingly devote space to explaining the bootstrap.

The allure of the bootstrap is due to the weakness of its competition as much as its own inherent advantages. Remember that conventional statistical theory relies heavily on large-sample asymptotic theory. With finite sample sizes, we know for a fact that using the limiting distribution (for example, the normal distribution for a regression coefficient) is merely an approximation to the exact sampling distribution. Research is showing that bootstrapping outperforms the conventional approach in areas in which the shape of the sampling distribution is crucial such as confidence intervals.

In addition, bootstrap methods force you to confront the data generation process directly. You must describe the way the dependent variable is generated and the role of the *X's* (for example, fixed or stochastic) to construct a resampling scheme that mimics the DGP. Once so described, the bootstrap can quickly approximate the sampling distribution of complicated statistics that would require difficult (and sometimes impossible) mathematical derivations.

Finally, no restrictive distributional assumptions are required to use the bootstrap. The sample data simply are what they are. Are the errors normally

P1: IYP
0521843197c23 CB962B/Barretto 0 521 84319 7 September 25, 2005 1:15

distributed? This is a crucial question for anyone wishing to apply a conventional *t*-test correctly, but the answer is irrelevant for a bootstrap analysis.

"Bootstrap methods, and other computationally intensive statistical techniques, continue to develop at a robust pace.... The twenty-first century may or may not use different theories of statistical techniques, but it will certainly be a different, better world for statistical practitioners" (Efron and Tibshirani 1993, p. 394).

23.7. Exercises

1. In Section 23.2, the text claims that, "in other words, the bootstrap will do a better job of answering questions that involve the shape of the sampling distribution when its profile is not normal. Suppose, for example, that we wanted to know the chances that a 95-percent free throw shooter will make 16 or less out of 20 free throws. The standard approach will fare badly because the sampling distribution of the sample percentage for this case is not very normal."
   a. Use the normal approximation to estimate the chances that a 95-percent free-throw shooter will make 16 or less out of 20 free throws. Describe your procedure and show your work. HINT: You need to find the SE of the sample percentage and use the endpoint correction (calculating the area under the normal curve up to 16.5, instead of just 16).
   b. Suppose you had an original sample of 19 out of 20 free throws made. Use the Bootstrap add-in to find the chances that the shooter will make 16 or less out of 20 free throws. Describe your procedure and take a screenshot of your results.
   c. Given your work in parts a. and b., what do you conclude about the claim that the bootstrap will do better than the standard approach (using the normal approximation)?
2. Suppose you had an original sample from a 95-percent free shooter in which he or she made all 20 free throws. How would the bootstrap work in this case?
3. Use the *Bootstrap* sheet in PairedXYBootstrap.xls to estimate the SE and sampling distribution of the coefficient on BodyWeight in Model 4. Take a screenshot of your bootstrap results.
4. The OLS estimated SE for the coefficient on BodyWeight in Model 4 is 0.315. Does your bootstrap SE substantially agree with the OLS estimated SE? Explain the reason for the difference or agreement.
5. Use the Bootstrap add-in to run a residuals bootstrap of the coefficient on BodyWeight in Model 4. Take a screenshot of your bootstrap results.
6. Compare the paired XY and residuals bootstraps for this case. Which one do you prefer? Why?

References

For general introductions into bootstrapping methods, we recommend the following:

Brownstone, David and Robert Valleta (2001). "The Bootstrap and Multiple Imputations: Harnessing Increased Computer Power for Improved Statistical Tests," *Journal of Economic Perspectives* **15**(4): 129–141.

Chernick, Michael (1999). *Bootstrap Methods: A Practitioner's Guide* (New York: John Wiley & Sons). This includes an extensive bibliography.

Davison, Anthony C. and David V. Hinkley (1997). *Bootstrap Methods and Their Application* (Cambridge, UK: Cambridge University Press). Part of the cat data from Fisher (1947) is also used in an exercise in this book.

Efron, Bradley and Robert Tibshirani (1993). *An Introduction to the Bootstrap* (New York: Chapman and Hall).

Leisch, Friedrich and A. J. Rossini (2003). "Reproducible Statistical Research" *Chance* **16**(2): 41–45. This easily accessible article is critical of bootstrap methods because each simulation leads to different results. This article cites Fisher (1947) and provided the inspiration for our use of Fisher's data in Section 23.3.

More advanced articles on bootstrapping include the following:

Andrews, Donald W. K. and Moche Buchinsky (2000). "A Three Step Method for Choosing the Number of Bootstrap Repetitions" *Econometrica* **68**(1): 23–51.

DiCiccio, Thomas J. and Bradley Efron (1996). "Bootstrap Confidence Intervals," *Statistical Science* **11**(3): 189–212.

Ohtani, Kazuhiro (2000). "Bootstrapping R^2 and Adjusted R^2 in Regression Analysis," *Economic Modelling* **17**(4): 473–483.

The following are early works in the bootstrapping literature:

Efron, Bradley (1979). "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics* **7**(1): 1–26. This is Efron's original article on the bootstrap.

Freedman, David (1981). "Bootstrapping Regression Models," *The Annals of Statistics* **9**(6): 1218–1228.

Wu, C. F. J. (1986). "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis," *The Annals of Statistics* **14**(4): 1261–1295.

Additional sources for this chapter include:

Fisher, R. A. (1947). "The Analysis of Covariance Method for the Relation between a Part and the Whole," *Biometrics* **3**(2): 65–68.

Project Gutenberg at <www.gutenberg.net/> has the full text of *The Surprising Adventures of Baron Munchausen* by Rudolph Erich Raspe.

Scheindlin, Stanley (2001)."A Brief History of Pharmacology," *Modern Drug Discovery* **4**(5) <pubs.acs.org/subscribe/journals/mdd/v04/i05/html/05timeline.html>.