Making Statistical Comparisons: An Application of The Bootstrap Technique in AIS Research

Jon Woodroof, (jwoodroof@mtsu.edu), Middle Tennessee State University

Abstract

In academic research, the classical approach to constructing confidence intervals and testing for significance has relied on the assumption of distribution normality. Non-parametric techniques do not require such assumptions, but have traditionally been difficult to implement. One such nonparametric technique is the bootstrap. In order to estimate the shape of a statistic's sampling distribution, bootstrapping uses large numbers of repetitive computations rather than strong distributional assumptions and analytic formulas. Instead of imposing (by assumption) a shape on a statistic's distribution, bootstrapping, through a resampling simulation, empirically estimates a statistic's entire distribution. With the development and widespread use of the electronic spreadsheet, researchers are now able to easily and powerfully implement the bootstrap. This paper demonstrates how to create a bootstrapping template in a Quattro Pro for Windows spreadsheet package that will allow academic researchers to compare various instruments designed to measure user satisfaction.

Introduction

Researchers must have confidence in the effectiveness of their measuring instruments; thus existing instruments must be constantly and rigorously evaluated to insure that they remain appropriate and effective. Evaluating, testing, and comparing measurement instruments is a very important task. Wanous and Lawler (1972) argue that there is a need for measuring constructs in more than one way and comparing the results to provide convergent validity. It is important to be able to statistically test which instrument taps into the construct of interest more effectively -- which instrument is "better".

Comments and questions concerning this article should be sent directly to the authors via e-mail.

For a number of reasons, such an evaluation has been difficult in accounting information systems research, especially as it relates to those instruments developed to measure a user's satisfaction with an information system. First, many of these instruments attempt to be all-encompassing measures of user satisfaction, applicable to a variety of systems. Users are being asked to provide generic responses to a potpourri of dimensions and attributes of user satisfaction. Zmud, Byrd, Sampson, Lenz, and Reardon (1994) make the following observation:

Many, if not most, of the measurement scales developed by information systems researchers tend to have been constructed in a generic fashion in order to increase the extent to which the results

of a study can both be generalized to contexts other than that examined in the particular study for which the scale was developed and be readily interpreted as informing the full body of theory to which the study was directed. All constructs, however, can be viewed as being located on a continuum reflecting the extent to which a construct is intellectually linked with experientially-based phenomena. (p. 6)

The generic nature of many of these instruments may have contributed to other methodological problems, such as weak construct validity which, in turn, may have been responsible for the lack of progress and the mixed results reported in user satisfaction research (Jarvenpaa, Dickson, DeSanctis, 1985; Kim, 1989; Klenke, 1992; Melone, 1990). There is concern in accounting information systems about whether researchers studying user satisfaction are measuring what they purport to be measuring (Tinker and Yuthas, 1995). It is very difficult to interpret analytical findings using such a broad, generic research approach.

Second, traditional parametric techniques utilized in statistical inferencing have relied on the general assumption that distributions of the statistics of interest are normal or student's *t* shaped. The central limit theorem assures the normality of the distributions of certain commonly used statistics as sample size increases, and the distributions for other statistics are known. However, in many research situations (especially in the social sciences), distribution normality assumptions may not be appropriate. Mooney and Duval (1993) state that there are:

...many statistics that might be of interest to social scientists that have unknown sampling distributions, no analytic formulas available for the parameters of these sampling distributions, and/or formulas that are intractable in certain cases (pp. 43,44).

They continue, giving the following example of a statistical procedure that is most commonly understood in the social sciences to be affected by its reliance on the assumption of dis-

tributational normality -- the ordinary least squares (OLS) estimation of regression parameters:

*In order to make inferencial statements using OLS, we need to assume that the random error in the model is normally distributed. This assumption is required because the sampling distribution of the OLS estimator is based on the random error of the model. If this parametric assumption holds, we can accurately develop confidence intervals for these coefficients using the *z* or *t* table. If we assume normality when the error is in fact not normal, our confidence intervals and hypothesis tests could have a greater than nominal probability of error (p. 55).*

Third, comparing different measurement instruments may require a statistical comparison of different populations. Statistical techniques for addressing these problems are limited and have, until recently, been difficult to implement. However, there are now simple but powerful nonparametric techniques available that (1) allow for a more effective testing of construct validity; (2) do not rely on the assumption of distribution normality; and (3) allow for statistical comparisons to be made between instruments taken from different populations.

The Bootstrap -- An Overview

The bootstrap is a nonparametric statistical technique for making probability-based inferences about a population characteristic (Efron, 1982). It differs from the traditional parametric approach to statistical inferencing in that,

Bootstrapping relies on an analogy between the sample and the population from which the sample was drawn. The central idea is that it may sometimes be better to draw conclusions about the characteristics of a population strictly from the sample at hand, rather than by making perhaps unrealistic assumptions about that population (Mooney and Duval, 1993, p.1).

Like traditional techniques of parametric statistical inference, bootstrapping allows a researcher, with limited information, to estimate

the underlying sampling distribution of a statistic, and from this distribution, to make inferences, with some degree of confidence, about a population parameter. Bootstrapping, however, employs a different approach from traditional techniques for doing this. To estimate the shape of a statistic's sampling distribution, bootstrapping uses large numbers of repetitive computations rather than strong distributional assumptions and analytic formulas. Instead of imposing (by assumption) a shape on a statistic's distribution, bootstrapping, through a resampling simulation, empirically estimates a statistic's entire distribution.

Thus, bootstrapping provides researchers with an important advantage: it is an excellent tool for analyzing statistics that have an unknown sampling distribution and no analytic formulas for calculating the parameters of these sampling distributions. In fact, this procedure is one of the only ways of obtaining such knowledge (Mooney and Duval, 1993). In cases where no parametric alternative exists, the fact that bootstrapping provides a method for making statistical inferences is an important contribution.

Bootstrapping was designed to be a non-parametric technique where normality is not assumed. Thus, bootstrapping can be employed to estimate the underlying sampling distributions of statistics that cannot be assumed to be normal (Mooney and Duval, 1993). By estimating the distributions of a statistic through a resampling simulation, bootstrapping enables confidence intervals to be derived. These confidence intervals allow a statistical comparison to be made between statistics from different instruments.

Bootstrapping treats the sample as if it is the population, empirically estimating the entire sampling distribution of the estimator of a population characteristic by examining the variation of the statistic within the sample. Here, is how it works. From the original sample, a Monte Carlo-style random resampling (with replacement) is performed so that each resample has the same number of elements as did the original sample. For each iteration of this resampling

simulation, the statistic of interest is recalculated and stored in a vector. This is done at least 1000 times (Efron, 1986). The resulting vector of statistic estimates is then ranked (ascending order), confidence interval endpoints of a particular alpha level of confidence are identified (for $\alpha = .05$ and # of samples = 1000, the endpoints are the 25th lowest value and the 25th highest value), and inferences about the population parameter are made.

There are four variations of the bootstrap: normal approximation method, percentile method, bias-corrected method, and percentile-t method. The percentile method was selected for this investigation because (1) it is the most straight-forward and probably the most widely used among applied statisticians; (2) the asymmetrical distribution of a statistic does not in theory adversely affect the accuracy of the percentile method's confidence interval; and (3) this method allows the bootstrapped sampling distribution to conform to any shape the data suggests (Mooney and Duval, 1993).

Example Problem: Generic Vs. Context-specific

The example problem used to apply the bootstrapping technique described in this paper is an experiment in which subjects were asked to use a software application and to evaluate it using two instruments: a generic instrument and a context-specific instrument. Since the focus of this paper is on a statistical technique for evaluating and comparing two measurement instruments rather than on the experiment itself, details regarding the experiment will be limited.

According to Zmud et al. (1994), many of the traditional measurement scales developed by information systems researchers have been constructed in a broad and generic fashion in order to increase both the variety of satisfaction dimensions and attributes able to be measured by an instrument and the extent to which the results of a study can be generalized. As a result, these generic instruments may be inadequate when used to measure the affective responses of users

of a specific application. Zmud and his colleagues argue that a context-specific instrument would provide a better measure of users' responses. Therefore, to test whether a context-specific instrument is a better measure of user satisfaction than is the more traditional generic instrument, the following null hypothesis is proposed:

H1: There is no difference between the amount of variance explained by a context-specific instrument and the amount of variance explained by a generic instrument.

Instrument Type

The generic instrument used in this study was the instrument developed by Doll and Torkzadeh (1988) shown in Appendix A. In this instrument, although targeted to be able to evaluate specific applications, the nature of the items is somewhat generic and lacks context specificity (Zmud, et al., 1994). The context-specific instrument (shown in Appendix B) consists of items suggested by Zmud, et al. (1994), Shneiderman (1992), and Bailey and Pearson (1983) that are comparatively more specific in nature.

Subjects were first allowed to use a particular software application. None of the subjects had seen or heard of the software prior to the experiment. Then, each subject was asked to evaluate the software application by responding to both the generic and context-specific instruments. Because subjects answered items on both the generic and context-specific instruments, the responses are correlated. In this situation, the resampling procedure of the bootstrap technique preserves the dependencies between the responses of the two instruments.

In addition to the generic and context-specific questions, subjects responded to behavioral intent questions. Turnover intentions have been shown to be consistently good indicators of turnover behavior (Vandenberg and Scarpello, 1990). Because of this link between intention and behavior, a technique was adopted for this

experiment similar to what is followed in job satisfaction research where subjects' responses to job satisfaction questions are used to predict turnover (Wanous, 1980). To establish the link in information systems between reuse intention and reuse behavior and to provide a common dependent variable, all subjects answered the following behavioral intent questions: "Will you use the system again?" and "Will you recommend this system to others?" Finally, subjects rated their overall satisfaction.

Approach

The bootstrap can be used on any statistic in which the researcher is interested. In this study, for reasons stated earlier, the statistic of interest is the R². Because comparing the results of different measures of the same construct is important (Wanous and Lawler, 1972), it was determined that a comparison of R²s would be made to determine which instrument (generic or context-specific) was a better measurement of user satisfaction. However, because the R²s are from different instruments, it must first be determined that the two instruments are measuring the same construct, and thus are comparable.

To provide a justification for being able to compare the instruments, the correlations between the response variables and the common construct of a user's behavioral intent were tested using the Fisher Z transformation (Neter, Wasserman, and Kutner, 1989). The same correlation was made between the response variables and users' overall satisfaction with the software application. The results of these tests indicated that there was no statistically significant difference between the abilities of these instruments to predict a user's behavioral intent or a user's overall satisfaction with the software application. Therefore, since the different response variables equally predicted a common behavioral variable as well as a user's overall satisfaction, the instruments that produced the response variables were considered comparable, and a comparison was made between the various R²s to determine which instrument explained the most variance.

Bootstrapping In A Spreadsheet

Although bootstrapping for statistical inference is not a new technique, the ability to perform this technique in a user-friendly electronic spreadsheet is relatively new (Willemain, 1994). Many researchers in accounting information systems are more comfortable with a spreadsheet than they are with a dedicated statistics package. The bootstrapping technique can be performed in a spreadsheet environment with relative ease and without sacrificing power. The following sections demonstrate how to create a bootstrapping template in a spreadsheet. Each page of the spreadsheet is presented in detail. Table 1 summarizes the spreadsheet formulas. The particular spreadsheet package utilized in this study to implement the bootstrapping technique was Quattro Pro for Windows, although any spreadsheet having a macro language would work.

Generic Response Page

The generic response page is shown in Figure 1. Here, the R^2 statistic is calculated for the responses collected on the generic instrument. The **Index** column in Figure 1 merely numbers the subject responses. This column will be used later as an index from which to randomly select a residual. The **Cell** column indicates from which experimental cell the response came. The **Response** column is the actual user satisfaction response from the subjects. Formulas in the **Cell Mean** column look up the mean for the particular cell indicated in the **Cell** column in a table on page CellMeans (not shown). The **Cell Mean Residual** column is derived by subtracting the associated cell mean from the original user satisfaction response of the subject. The **Cell Mean Residual 2** column is the square of the cell mean residual (the combination

Quattro Pro for Windows - BOTH.WB1

File Edit Block Data Tools Graph Property Window Help

Generic.D3 @AVG(D8:D264)

	A	B	C	D	E	F	G	H	I	J	K
1				Grand							
2				Mean						R-square:	
3				3.397			130.347		198.417	0.34307	
4											
5											
6											
7				Index	Cell	Response	Cell	Cell	Grand	Grand	
8							Mean	Mean	Mean	Mean	
9							Residual	Residual ²	Residual	Residual ²	
10				1	1	4.583	4.092	0.492	0.242	1.186	1.407
11				2	1	4.250	4.092	0.158	0.025	0.853	0.727
12				3	1	4.500	4.092	0.408	0.167	1.103	1.216
13				4	1	4.083	4.092	-0.008	0.000	0.686	0.471
14				5	1	4.083	4.092	-0.008	0.000	0.686	0.471
15				6	1	4.583	4.092	0.492	0.242	1.186	1.407
16				7	1	2.917	4.092	-1.175	1.381	-0.481	0.231
17				8	1	3.250	4.092	-0.842	0.708	-0.147	0.022
18				9	1	3.667	4.092	-0.425	0.181	0.269	0.073
19				10	1	5.000	4.092	0.908	0.825	1.603	2.569
20				11	2	3.083	3.515	-0.432	0.186	-0.314	0.099
				12	2	3.667	3.515	0.152	0.023	0.269	0.073
				13	2	3.047	3.515	0.402	0.161	0.519	0.270

Generic ContextSpecific ResampleCalc ResampleOutput Macro Frequency Inference

Start My Stuff Quattro Pro for Wind 9:02 AM

Figure 1: Generic Response Page

"^2" squares numbers in a spreadsheet formula). The **Grand Mean Residual** column is derived by subtracting the grand mean (cell D3: the average of all subjects' responses) from the original user satisfaction response of the subject. Finally, the **Grand Mean Residual ^2** column is the square of the grand mean residual. The formulas in the range E8 through I8 (detailed in Table 1) are able to be copied down to as many rows as there are subject responses. The sum of the cell mean residual squares (G3) is then divided by the sum of the grand mean residual squares (I3) and subtracted from 1. This derives the R^2 statistic (.34307) found in K3.

Context-Specific Response Page

The context-specific response page is shown in Figure 2. The layout, logic, and formulas for this page are exactly the same as for

the generic response page.

Resample Calculation Page

The resample calculation page is shown in Figure 3. This page is the heart of the random resampling used in the bootstrap technique. The **Cell** column indicates from which experimental cell the response came. Formulas in the **NewObs** column add a randomly sampled residual to the mean for the particular cell indicated in the **Cell** column (brought over from the generic response page), thereby deriving a new simulated observation. Formulas in the **SampledRes** column use the @RAND, @COUNT, and @VLOOKUP functions to randomly select a residual (see Table 1).

@COUNT counts the number of responses from which @RAND generates a ran-

Quattro Pro for Windows - BOTH.WB1

File Edit Block Data Tools Graph Property Window Help

ContextSpecific.D3 @AVG(D8..D264)

Index	Cell	Response	Cell Mean	Cell Residual	Cell Residual^2	Grand Mean Residual	Grand Mean Residual^2
1	1	4.182	4.109	0.073	0.005	0.835	0.697
2	1	4.182	4.109	0.073	0.005	0.835	0.697
3	1	4.636	4.109	0.527	0.278	1.289	1.662
4	1	4.364	4.109	0.255	0.065	1.017	1.034
5	1	4.182	4.109	0.073	0.005	0.835	0.697
6	1	4.000	4.109	-0.109	0.012	0.653	0.426
7	1	3.273	4.109	-0.836	0.700	-0.074	0.006
8	1	3.727	4.109	-0.382	0.146	0.380	0.145
9	1	3.545	4.109	-0.564	0.318	0.198	0.039
10	1	5.000	4.109	0.891	0.794	1.653	2.732
11	2	3.182	3.347	-0.165	0.027	-0.165	0.027
12	2	2.545	3.347	-0.802	0.643	-0.802	0.642
13	2	3.636	3.347	0.289	0.084	0.289	0.084

Grand Mean: 3.347, Cell Mean: 4.109, R-square: 0.44589

Generic ContextSpecific ResampleCalc ResampleOutput Macro Frequency Inference

Start My Stuff Quattro Pro for Wind 9:25 AM

Figure 2: Context-Specific Response Page

dom number. This random number is then used by @VLOOKUP to select a residual. Every time the spreadsheet is recalculated, a new random number is generated and a new residual selected.

The **NewCellMean** column is derived by averaging the new simulated observations associated with each experimental cell. Formulas in the **NewRes** column calculate a new residual by subtracting the new cell mean from the new simulated observation. The **Y*-Yhat*^2** column is the square of the new cell mean residual. The **Y*-Ybar*** column is derived by subtracting the grand mean (cell B2: the average of all new simulated observations) from the each new simulated observation. Finally, the **Y*-Ybar*^2** column is the square of the grand mean residual. The formulas in the range B4 through I4 (detailed in Table 1) are able to be copied down

to as many rows as there are subject response. Note that Figure 3 shows only the resampling calculation cells for the generic instrument. The resampling calculation cells for the context-specific instrument are exactly the same as they were for the generic instrument and are in columns J through R of this page.

Resample Output Page

The resample output page is shown in Figure 4. This page shows the results of each iteration of the bootstrap, as well as summary statistics on the distributions of the simulated R^2 s from the two instruments. Of special importance are the formulas in cells F10 and G10. The R^2 s shown in these cells are calculated the same way as were the R^2 s in Figures 1 and 2: the sum of the new residual squares (G2) is divided by the sum

	A	B	C	D	E	F	G	H	I
1									
2		3.367	< new overall mean				112.238		190.816
3	Cell	NewObs	CellMean	SampledRes	NewCellMean	NewRes	Y*-Yhat*^2	Y*-Ybar*	Y*-Ybar*^2
4	1	4.501	4.092	0.410	4.259	0.242	0.059	1.134	1.286
5	1	3.978	4.092	-0.114	4.259	-0.281	0.079	0.611	0.373
6	1	5.130	4.092	1.038	4.259	0.870	0.757	1.762	3.105
7	1	4.750	4.092	0.658	4.259	0.491	0.241	1.383	1.911
8	1	4.463	4.092	0.371	4.259	0.203	0.041	1.095	1.200
9	1	3.493	4.092	-0.598	4.259	-0.766	0.587	0.126	0.016
10	1	4.895	4.092	0.803	4.259	0.635	0.403	1.527	2.332
11	1	3.904	4.092	-0.188	4.259	-0.355	0.126	0.537	0.288
12	1	4.114	4.092	0.023	4.259	-0.145	0.021	0.747	0.558
13	1	3.367	4.092	-0.725	4.259	-0.893	0.797	-0.001	0.000
14	2	3.902	3.515	0.386	3.090	0.811	0.658	0.534	0.285
15	2	2.123	3.515	-1.392	3.090	-0.967	0.935	-1.244	1.548
16	2	3.328	3.515	-0.188	3.090	0.237	0.056	-0.040	0.002
17	2	3.811	3.515	0.295	3.090	0.720	0.519	0.443	0.196
18	2	2.061	3.515	-1.455	3.090	-1.030	1.060	-1.307	1.708
19	2	3.538	3.515	0.023	3.090	0.448	0.200	0.170	0.029
20	2	3.048	3.515	-0.467	3.090	-0.042	0.002	-0.319	0.102

Figure 3: Resample Calculation Page

Table 1: Summary of Spreadsheet Formulas

Page	Formulas
<i>Generic</i>	D3: @AVG(D8..D264) G3: @SUM(G8..G264) I3: @SUM(I8..I264) K3: 1-(G3/I3) E8: @VLOOKUP(C8,CellMeans:\$B\$4..CellMeans:\$C\$27,1) F8: +D8-E8 G8: +F8^2 H8: +D8-\$D\$3 I8: +H8^2
<i>Context-Specific</i>	Same as <i>Generic Response</i> page
<i>Resample Calculation</i>	B2: @AVG(B4..B260) G2: @SUM(G4..G260) I2: @SUM(I4..I260) B4: +C4+D4 C4: +Generic:E8 D4: @VLOOKUP(@INT(@RAND*@COUNT(Generic:\$B\$8..Generic:\$B\$264))+1,Generic:\$B\$8..Generic:\$F\$264,4) E4: @DAVG(\$E\$1..\$F\$5,"NewObs",CELL=\$E2) F4: +B4-E4 G4: +F4^2 H4: +B4-\$B\$2 I4: +H4^2
<i>Resample Output</i>	C3: +A3-B3 F4: +Generic:K3 G4: +ContextSpecific:K3 H4: +F4-G4 F6: @STDS(A3..A5002) G6: @STDS(B3..B5002) H6: @STDS(C3..C5002) F7: @MAX(A3..A5002) G7: @MAX(B3..B5002) H7: @MAX(C3..C5002) F8: @MIN(A3..A5002) G8: @MIN(B3..B5002) H8: @MIN(C3..C5002) F10: 1-(ResampleCalc:G2/ResampleCalc:I2) G10: 1-(ResampleCalc:P2/ResampleCalc:R2) H10: +F10-G10
<i>Macro</i>	No formulas
<i>Frequency Distrib.</i>	No formulas
<i>Inference</i>	D7: @VLOOKUP(C7,\$H\$13..\$I\$5012,1) E7: -D8+2*E10 C8: 1-C7 D8: @VLOOKUP(C8,\$H\$13..\$I\$1012,1) E8: -D7+2*E10 E10: +ResampleOutput:H4

of the grand mean residual squares (I2) and subtracted from 1. From these cells, the macro (explained in the following section) outputs the R^2 s simulated with each iteration of the bootstrap to this page, beginning in cell A3. Formulas in the **Difference** column subtract the new R^2 s generated by resampling the context-specific instrument responses from the new R^2 s generated by resampling the generic instrument responses, creating a vector of new R^2 differences.

Macro Page

The macro page is shown in Figure 5. Three range names were created and referenced in the macro: (1) D9: **countloc**, which is incremented with each iteration; (2) D8: **iterations**, where the researcher enters the number of times to resample and recalculate; and (3) D10: **startloc**, the place in the macro where the loop initiates.

In the loop, which initiates in D10, the bootstrap macro performs three tasks: (1) it executes the formulas that randomly sample the original residuals and creates new calculations for the R^2 statistic, (2) it places the results of these calculations (found in cells F10 and G10 on the resample output page) in two vectors on the resample output page, and (3) it increments the counter and loops to perform 1 and 2 the number of iterations specified. In this example, the re-sampling and recalculation was done 5000 times, making the resulting distributions smoother. The macro is activated by a button (here, "Start Bootstrap") attached to the macro starting cell, C5.

Frequency Distribution Page

The frequency distribution page is shown in Figure 6. The R^2 statistics in the generic and context-specific vectors on the resample output

Figure 4: Resample Output Page

page (Figure 4) are assigned to a "bin" or R^2 range. The number of assignments in each bin are then tallied and placed in the **Frequency** column. This is automatically accomplished by a data frequency algorithm in the spreadsheet. An area graph is then generated, based on this frequency distribution, showing the shape of the two simulated distributions.

As can be seen in Figures 1 and 2, the generic instrument was able to explain 34.31 percent of the variance, while the context-specific instrument was able to explain 44.59 percent, an increase in explanatory ability of almost 30 percent. Also, as can be seen in Figure 6, there is a visible difference in the distributions, although the tails of the distributions overlap somewhat. And the distributions seem to approach normality, although they really cannot be

called normal distributions.

Inference Page

The last spreadsheet page, the inference page, is shown in Figure 7. The question remains, "Is the difference between these instruments significant?" Here is where the advantages of using the nonparametric statistical inferencing technique of bootstrapping can really be seen. Assuming that $[(R^2_1)^* - (R^2_2)^* - (-\text{difference between original } R^2\text{'s})]$ has approximately the same distribution as $(R^2_1 - R^2_2) - (\rho^2_1 - \rho^2_2)$, equation (1) was used to construct confidence intervals for the comparison between the generic and context-specific saturated models:

$$P(L - D \leq (R^2_1)^* - (R^2_2)^* - D \leq U - D) = .95, \quad (1)$$

Figure 5: Macro Page

implying equation (2)

$$P((R_1^2) - (R_2^2) - (U-D) \leq \rho_1^2 - \rho_2^2 \leq (R_1^2) - (R_2^2) - (L-D)) = .95, \quad (2)$$

where

L = the new R^2 difference associated with the lower 2.5% quantile,

D = difference between the original R^2 s,

$(R_1^2)^*$ = estimated R^2 for the model based on the generic instrument,

$(R_2^2)^*$ = estimated R^2 for the model based on the context-specific instrument,

U = the new R^2 difference associated with the upper 97.5% quantile,

ρ_1^2 = population R^2 for the generic instrument, and

ρ_2^2 = population R^2 for the context-specific instrument.

The vector of R^2 differences on the re-sample output page is sorted (ascending order) and placed in the **Ranked Sample** column. The **Ascending** column contains the quantile increments (determined by dividing 1 by the number of resampling iterations). In cell C7, the researcher inputs the upper endpoint for the confidence level of interest; here, .975. Cell C8 calculates the lower endpoint. Cells D7 and D8 look up (in the **Ranked Sample** column) the R^2 difference associated with the upper and lower endpoints, respectively.

Formulas in the **Bootstrap** column determine the upper and lower bootstrap range (see Table 1). If zero is between the upper and lower bootstrap ranges, the difference in the distributions is not significant. However, if zero is not between the upper and lower bootstrap ranges, the difference is significant.

Figure 6: Frequency Distribution Page

Interpretation of Results

The interpretation of the results shown in Figure 7 is as follows: if many independent samples are taken where the lot sizes are the same and a 95% confidence interval is constructed for each sample, 95% of the constructed intervals will contain the true difference between the R2s from the models based on the generic and context-specific instruments. Because zero is not within the upper and lower bound of the interval, we can have confidence at the 95% level that the difference between the amount of variance explained by the models based on the generic and context-specific instruments is significant. Based on this finding, **H1**, There is no difference between the amount of variance explained by a generic instrument and the amount of variance explained by a context-specific instrument, can be rejected. The context-specific

instrument is able to explain a greater amount of variance and, thus, is better able to tap into the user satisfaction construct than is the generic instrument.

Summary and Conclusions

Bootstrapping has several advantages over traditional parametric techniques. First, this technique provides an effective way for researchers to make statistical comparisons between constructs measured by different instruments, so that convergent validity can be assessed. Second, this method does not require the researcher to make assumptions about the distribution of a statistic which may inappropriate, but allows the bootstrapped sampling distribution to conform to any shape the data suggests. Third, this technique gives researchers a nonparametric technique for being able to make statistical com-

Figure 7: Inference Page

parisons between measurement instruments from different populations. In fact, this procedure is one of the only ways of obtaining such knowledge.

The problem with many non-parametric statistical techniques (like the bootstrap) is that they have traditionally been difficult both to understand and to implement. This paper demonstrates the use of bootstrapping in a user-friendly spreadsheet environment, providing researchers (who may not be comfortable using a dedicated statistics package) easy access to this powerful nonparametric technique.

Implications for Future Research

This paper helps researchers to better understand the "workings" of the bootstrap technique and become more confident in its use. It is the author's hope and expectation that such understanding and confidence might cause this powerful nonparametric technique to be used more frequently in future research. Many statistics of interest to researchers have an unknown (and unknowable) distribution and no analytic formulas for calculating the parameters of these distributions. Increased use of the bootstrap technique should have a significant impact on future research, since it is able to help researchers analyze and test such statistics.

References

1. Bailey, J.E. and S.W. Pearson, "Development of a Tool for Measuring and Analyzing Computer User Satisfaction," *Management Science*, Vol. 29, 530-545, 1983.
2. Doll, W.J. and G. Torkzadeh, "The Measurement of End-User Computing Satisfaction," *MIS Quarterly*, Volume 12, Number 2, pp. 259-274, 1988.
3. Efron, B. "The Jackknife, the Bootstrap, and Other Resampling Plans," *CBMS-NSF Monograph 38*, Philadelphia: Society of Industrial and Applied Mathematics, 1982.
4. Efron, B. and R. Tibshirani, "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," *Statistical Science* 1: pp. 54-77, 1986.
5. Jarvenpaa, S.L., G. Dickson and G. DeSanctis, "Methodological Issues in Experimental IS Research: Experiences and Recommendations," *MIS Quarterly*, Volume 9, No. 2, pp. 141-156, 1985.
6. Mooney, C.Z. and R.D. Duval, "Bootstrapping: A Nonparametric Approach to Statistical Inference," Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-095, Newbury Park, CA: Sage, 1993.
7. Neter, J., W. Wasserman and M.H. Kutner, *Applied Linear Regression Models*, (2nd edition), Boston, MA: Richard D. Irwin, Inc., p. 531, 1989.
8. Shneiderman, B., *Designing the User Interface*, Addison-Wesley Publishing Company, Inc., 1992.
9. Tinker, T. and K. Yuthas, "There's Nothing Like Success: Cartesian Dualisms in Information Systems Research," *Advances in Accounting Information Systems*, Volume 3, pp. 31-59, 1995.
10. Vandenberg, R.J. and V. Scarpello, "The Matching Model: An Examination of the Processes Underlying Realistic Job Previews," *Journal Of Applied Psychology*, Volume 75, Number 1, pp. 60-67, 1990.
11. Wanous, J.P., "Organizational Entry: Recruitment, Selection, and Socialization of Newcomers," Reading, MA: Addison-Wesley, 1980.
12. Wanous, J.P. and E.E. Lawler, "Measurement and Meaning of Job Satisfaction," *Journal of Applied Psychology*, pp. 95-105, 1972.
13. Willemain, T.R., "Bootstrap on a Shoestring: Resampling Using Spreadsheets," *The American Statistician*, Volume 48, pp. 40-42, 1994.
14. Zmud, Robert W., Terry A. Byrd, James P. Sampson, Jr., Janet G. Lenz, and Robert C. Reardon, "Construct Measurement in Information Systems Research," *Information Technology and People*, Vol. 7 (2), pp. 24-45, 1994.

Appendix A - Generic Instrument

This questionnaire deals with the feelings and attitudes people have about a computer application. Please indicate how much you agree with each of the following statements by using the scale below:

1-----	2-----	3-----	4-----	5-----
Disagree	Disagree strongly	Neither agree nor disagree	Agree	Agree Strongly

The output is presented in a useful format.
 I am satisfied with the accuracy of the system.
 The information is clear.
 The system is accurate.
 The system provides sufficient information.
 The system provides up-to-date information.
 I get the information I need in time.
 The system provides reports that seem to be just about exactly what I need.
 The system is easy to use.
 The system is user friendly.
 The system provides the precise information I need.
 The information content meets my needs.

Appendix B - Context-Specific Instrument

This questionnaire deals with the feelings and attitudes people have about a computer application. Please indicate how much you agree with each of the following statements by using the scale below:

1-----	2-----	3-----	4-----	5-----
Disagree	Disagree strongly	Neither agree nor disagree	Agree	Agree Strongly

The font and size of characters used make the screen more readable.
 The color schemes used in the application are pleasing.
 The layout of fields and text on the screen is thoughtful.
 The message feedback is helpful.
 The system is accurate.
 I get the information I need in time.
 The information content meets my needs.
 The response time is good.
 The time required to learn the application is reasonable.
 It is easy to recover from the errors I make when using the application.
 I will be able to use this application again in about a month without having to be reminded of how to use it.