5-1-2010

Another Look at Resampling: Replenishing Small Samples with Virtual Data through S-SMART

Haiyan Bai
University of Central Florida, hbai@mail.ucf.edu

Wei Pan
University of Cincinnati, wei.pan@uc.edu

Leigh Lihshing Wang
University of Cincinnati, lihshing.wang@uc.edu

Phillip Neal Ritchey
University of Cincinnati, phillip.ritchey@uc.edu

Another Look at Resampling:
Replenishing Small Samples with Virtual Data through S-SMART

Haiyan Bai   Wei Pan   Leigh Lihshing Wang   Phillip Neal Ritchey
University of Central Florida   University of Cincinnati

A new resampling method is introduced to generate virtual data through a smoothing technique for replenishing small samples. The replenished analyzable sample retains the statistical properties of the original small sample, has small standard errors and possesses adequate statistical power.

Key words: Small sample, resampling, smoothing, bootstrap.

Introduction

Drawing statistical inferences based upon small samples has long been a concern for researchers because small samples typically result in large sampling errors (Hansen, Madow, & Tepping, 1983; Lindley, 1997) and small statistical powers (Cohen, 1988); thus, sample size is essential to the generalization of the statistical findings and the quality of quantitative research (Lindley, 1997). Unfortunately, sample size is often constrained by the cost or the restrictions of data collection (Adcock, 1997), especially for research involving sensitive topics. Consequently, it is not unusual for researchers to use small samples in their quantitative studies. For example, in *American Educational Research Journal* and *Journal of Consulting and Clinical Psychology*, 155 out of 575 (27.0%) quantitative

Haiyan Bai is an Assistant Professor of Quantitative Research Methodology in the Department of Educational and Human Sciences. Email: hbai@mail.ucf.edu. Wei Pan is an Associate Professor of Quantitative Research Methodology in the School of Education. Email: wei.pan@uc.edu. Leigh Lihshing Wang is an Associate Professor of Measurement and Evaluation in the School of Education. Email: lihshing.wang@uc.edu. Phillip Neal Ritchey is a Professor of Basic Research and Policy Analysis in the Department of Sociology. Email: phillip.ritchey@uc.edu.

research articles published between 2003 and 2007 used relatively small sample sizes ranging from 10 to 100. Therefore, the need for studies addressing the problem of small samples is prominent.

With the development of modern computer science, four commonly used resampling methods emerged as revolutionary techniques to address small sample problems (Diaconis & Efron, 1983), these are:

(a)   randomization test (Fisher, 1935);
(b)   cross-validation (Kurtz, 1948);
(c)   the jackknife (Quenouille, 1949; modified by Tukey, 1958); and
(d)   the bootstrap (Efron, 1979, 1982).

The bootstrap is the most commonly-used resampling method in research with small samples (Efron & Tibshirani, 1998). Not only have resampling methods been applied to basic statistical estimation, such as estimation bias, standard errors, and confidence intervals, but they also have been applied to more advanced statistical modeling, such as regression (Efron, 1979, 1982; Hinkley, 1988; Stine, 1989; Wu, 1986), discriminant analysis (Efron,1979), time series (Hinkley, 1988), analyses with censored data (Efron, 1979; Efron & Gong 1983), missing data (Efron, 1994), and dependent data (Lahiri, 2003).

Existing resampling methods are very useful statistical tools for dealing with small sample problems; however, they have critical limitations (see Bai & Pan, 2008, for a review).

Specifically, the randomization test requires data that can be rearranged, the cross-validation splits an already small sample, and the jackknife usually needs a large sample to obtain good estimates (Davison & Hinkley, 1999). Bootstrap, the most commonly-used method, also has a few limitations. For example, the lack of independent observations is evident due to the possible duplications of the observations in the bootstrap samples that are randomly drawn from the same small samples with replacement (Efron &Tibshirani, 1998); and the bootstrap technique is potentially nonrobust to outliers because every observation in the small sample, including the outliers, has an equal chance of being selected (Troendle, Korn & McShane, 2004).

More importantly, all of the resampling methods have a common problem: the new resamples are limited to being selected from the same original small sample, which leaves a considerable number of data points unselected in the population. The problem with this common resampling process is that the resamples carry no more information than the original small sample. This common limitation of existing resampling methods potentially undermines the validity of the statistical inference. Therefore, if a small sample could be replenished with additional data points to capture more information in the population, it would significantly reduce the common limitation of the extant resampling methods.

The smoothing technique (Simonoff, 1996) made a breakthrough in lessening the limitation of sampling from the original small sample; however, the smoothing procedure in the smoothed bootstrap (de Angelis & Young, 1992, Hall, DiCiccio, & Romano 1989; Silverman & Young 1987) is so complicated that even statisticians ask how much smooth is optimal. In addition, the question of when to use smoothing technique still remains problematic for researchers (Silverman & Young, 1987). Therefore, developing an alternative but simpler resampling method for lessening the limitations of the existent resampling methods is imperative.

The purpose of this article is to develop a new resampling method, namely *Sample Smoothing Amplification Resampling Technique* (S-SMART), for generating virtual data to replenish a small sample based on the information provided by the small sample, both in its own right and as an ingredient of other statistical procedures. The S-SMART technique directly employs kernel smoothing procedures to the original small sample before resampling so that the information carried by the original small sample is well retained. Not only does the S-SMART procedure potentially lessen some limitations of the existing resampling methods, but also it reduces sampling errors and increases statistical power. Therefore, the use of the replenished sample through S-SMART provides more valid statistics for estimation and modeling than does the original small sample; and ultimately, it will improve the quality of research with small samples.

Specifically, this study aims to develop S-SMART as a distribution-free method to produce S-SMART samples which (a) have sufficient sample sizes to provide an adequate statistical power, (b) have stable sample statistics across different replenishing scenarios, (c) have smaller sampling errors, (d) are independent observations, (e) are robust to outliers, and (f) are easily obtained in a single simple operation. This study evaluated the accuracy of the statistics through both a simulation study and an empirical example.

Methodology
The S-SMART Procedure

The S-SMART procedure integrates smoothing technique and distribution theory into a new resampling method. The smoothing parameter, sample size, quantiles, and standard error of the mean of the original small sample are the simulation parameters for generating S-SMART samples.

Figure 1 shows a schematic diagram of the S-SMART procedure. First, the percentile range from 2.5% to 97.5% of the original small sample of size n is evenly divided into k equal intervals for obtaining the corresponding quantiles q_i, $i = 0, 1, …, k$; second, the quantiles are used as means for the small smoothing Gaussian kernels and the standard error of the mean of the small sample are used as a random noise for the Gaussian kernels; third, random samples of size s are generated from the Gaussian kernels as the S-SMART sub-samples

to replenish the small sample; and last, a union of the $(k + 1)$ S-SMART sub-samples is taken to obtain an amplified S-SMART sample with enlarged sample size of $(k + 1) \times s$.

The following algorithm is a more explicit description of the S-SMART procedure for obtaining the amplified S-SMART sample with replenishing virtual data:

1. Let U be a population with an unknown underlying probability density function $f(x)$.
2. Let $X = \{x_1, x_2, \ldots, x_n\}$ be an independent identically distributed (i.i.d.) random small sample of size n drawn from U.
3. Let $f_n(x)$ be the empirical probability density function of X.
4. Let $Q = \{q_0, q_1, \ldots, q_k\}$, $k \geq n$, be a set of quantiles whose corresponding percentile ranks equally divide the 95% range of X's percentile ranks into k intervals (i.e., $q_0 \sim 2.5\%, \ldots, q_k \sim 97.5\%$).
5. Let $f_k^{(q)}(x)$ be the empirical probability density function of Q. By the probability

distribution is the inverse of its cumulative distribution function; therefore, $f_k^{(q)}(x)$ captures the shape of the distribution $f_n(x)$ (Gilchrist, 2000). Using the 95% range of the percentile ranks, instead of using all the percentile ranks, is intended to eliminate possible outliers; and the equal division of the 95% percentile range complies with the principle of smoothing (Simonoff, 1996).

6. Let $X_i^* = \{x_{i1}^*, x_{i2}^*, \ldots, x_{is}^*\}$, $i = 0, 1, \ldots, k$, be an i.i.d. random sample of size s drawn from a Gaussian kernel:

$$G_i(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x - q_i}{h}\right)^2\right);$$

that is, $x_{ij}^* = q_i + h \times \varepsilon_j$, where $\varepsilon_j \sim N(q_i, h^2)$, $j = 1, 2, \ldots, s$; $i = 0, 1, \ldots, k$. The choice for the bandwidth h can be the standard error of $\overline{X}$ or $\overline{Q}$ as suggested by Hesterberg (1999). The kernel estimators center the kernel

Figure 1: A Schematic Diagram of the S-SMART Procedure
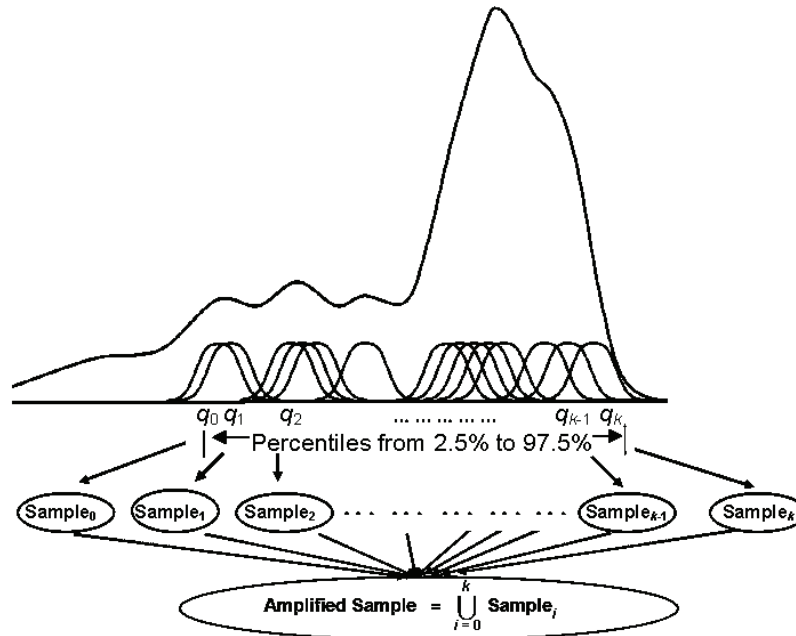
Notes: (1) $q_0, q_1, \ldots, q_k$ are the quantiles whose corresponding percentiles evenly divide the middle 95% percentile range of the original small sample; (2) k is the number of intervals which determine $k + 1$ Gaussian kernels, each uses q_i ($i = 0, 1, \ldots, k$) as its mean and the standard error of the mean of the original small sample as its standard deviation.

function at each quantile point q_i, which smoothes out the contribution of each point over a local neighborhood of that point. X_i^*, $i = 0, 1, \ldots, k,$ serve as the replenishing virtual data to the small sample X.

7. Let $S^* = \bigcup\limits_{i=0}^{k} X_i^*$. Then, the empirical probability density function of S^* can be expressed as a weighted average over the Gaussian kernel functions (Parzen, 1962; Silverman,1986):

$$f_k^*(x) = \frac{1}{kh} \sum_{i=0}^{k} G_i(x)$$
$$= \frac{1}{kh\sqrt{2\pi}} \sum_{i=0}^{k} exp\left(-\frac{1}{2}\left(\frac{x-q_i}{h}\right)^2\right).$$

S^* is called as the amplified S-SMART sample with an enlarged sample size of $(k + 1) \times s$. $f_k^*(x)$ is a consistent estimate of $f(x)$ because the Gaussian kernel $G_i(x)$ is a weighting function and the bandwidth h, specified as the standard error, satisfies the conditions $\lim_{k\to\infty} h = 0$ and $\lim_{k\to\infty} kh = \infty$ (Parzen, 1962).

It is notable that the resamples of all the existing resampling methods are the replica of the original data points in the small sample which is in practice not always representative of the population through the randomization because "random sampling for representativeness is theoretically possible on a larger scale" (Cook & Campbell, 1979, p. 75). On the contrary, this newly-developed S-SMART method intends to recover the randomization through the random noise of the Gaussian kernels in the smoothing technique, rather than striving to achieve a one-mode estimated empirical distribution from the small sample, which is the common goal of smoothing.

Monte Carlo

As with all other resampling methods, the S-SMART method is bound to have some properties that are mathematically difficult to evaluate; therefore, a simulation study is conducted to provide additional knowledge and numerical evidence to address the questions about the method's properties (Johnson, 1987). The statistical behaviors of the S-SMART samples from both the simulated data and empirical data are evaluated in terms of the first two generalized method of moments (GMM), mean and variance, which are commonly used to describe a set of data (Gravetter & Wallnau, 2004). The sample distributions, sampling distributions of the means, and confidence intervals of the means and standard deviations are also studied.

The evaluation of the S-SMART method is conducted for small samples from normal, exponential, and uniform distributions, representing the three families of continuous distributions, which demonstrates that the S-SMART procedure is a distribution-free technique. The Statistical Analysis System (SAS) (SAS Institute Inc., 2001) is employed to both generate the small samples and resample the replenishing virtual data.

To investigate the stability of the S-SMART samples, the first two generalized method of moments of the S-SMART samples amplified from the simulated random small samples were compared across different amplifying times for each different sample size of the small samples. The different sample sizes of the small samples are 10, 15, 20, 25, and 50, which were randomly generated from normal, exponential, and uniform distributions. The small sample sizes were determined according to a power analysis, that is, the smallest sample size for statistical analysis with an adequate power, such as the t-test, is 64 (Cohen, 1988). The corresponding amplified S-SMART samples were simulated with the amplified sample sizes as 10, 20, 50, and 100 times as the original small sample sizes; accordingly, the sample sizes of the S-SMART samples ranged from 100 to 50,000, providing adequate power for a variety of statistical analyses (Cohen, 1988).

Results

Evaluating the Sample Distributions

To understand the properties of the distribution of the S-SMART samples, histograms of the S-SMART samples and the corresponding original small samples were

compared. Three sets of five small samples of size 10, 15, 20, 25, and 50 were randomly generated from three different distributions: standard normal, exponential, and uniform, respectively; then, they were amplified 10, 20, 50, and 100 times, respectively, through the S-SMART procedure.

For illustration purpose, Figure 2 only shows the amplification results for the small sample of size 20 because the amplification results for the small samples of other sizes were the same. In Figure 2, the histograms on the left panel illustrate that at all levels of the amplifying times, the S-SMART samples imitated the original small-sample distribution. That is, all the S-SMART samples amplified from the normally-distributed small sample appear to be also normally distributed. The same phenomenon was also true for both the exponential distribution and uniform distribution (see the middle and right panels in Figure 2).

Evaluating the First Two Moments

To test the stability of the sample statistics of the S-SMART sample across the different S-SMART sample sizes, an ANOVA test for equal means and the Brown and Forsythe's (1974) test for equal variances were conducted. The Brown and Forsythe's test for homogeneity of variances is a more robust and powerful test for comparing several variances under heteroscedasticity and non-normality (O'Brien, 1992; Tabatabai, 1985). It is important to examine the stability of variances because the variance is another essential measure reflecting the statistical properties of data.

The ANOVA results (see Tables 1, 2, 3) indicate that the S-SMART samples statistically did not have mean differences across all levels of the S-SMART samples and also have equal means with those of the original samples with almost all the p-values larger than .90; therefore, they had stable means with different amplifying times. Thus, the S-SMART procedure cloned samples carrying the same most important statistical behavior of the original small sample.

The results of the Brown and Forsythe's test showed that no significant differences existed among the variances across different amplifying times and the small samples from non-normal, exponential, and uniform

distributions (Tables 2, 3). The phenomenon is also generally true for the normal distribution (Table 1) except for the last two cases (see last two p-values for the Brown and Forsythe's test in Table 1) where the sample sizes are very large, which inevitably caused overpowered tests with biased small p-values for testing equal variances between groups.

The significance of the Brown and Forsythe's test for equal variance were generated from the very large sample sizes with excessive power and extremely unbalanced sample sizes. Nonetheless, when all the variances in Tables 1, 2 and 3 were examined, it was found that the magnitudes of the variances did not differ substantially. Taking the most significant group in the three tables, for example, the small sample of size 50 in Table 1 with the p-value of .001, the relative differences (the absolute value of (S-SMART_SD – Small-Sample_SD)/ Small-Sample_SD) are only .033 for the S-SMART of size 500, .041 for the S-SMART of size 1000, and .091 for S-SMART of size 5000. According to a rule of thumb provided by Efron and Tibshirami (1993), if the relative difference is less than .25, the differences can be ignored. As results show, all the relative differences are smaller than .25.

Evaluating Confidence Intervals

To further investigate the properties of the S-SMART samples, three groups of 95% confidence intervals were estimated for comparing the S-SMART samples with the original small samples from the three representative distributions: normal, exponential, and uniform. The current intervals were constructed parametrically because the S-SMART sample sizes were large enough (all > 100) for the sampling distributions to be approximately normal according to the central limit theorem. Figure 3 shows fifty replications of the confidence intervals generated based on the original small samples of size 20 from standard normal, exponential, and uniform distributions. The longest bar for the interval corresponding to the label $S0$ represents the confidence interval of the original small sample from each distribution. The clustered, short intervals corresponding to the labels Li, $i = 1, 2, …, 50$, represent the confidence intervals for the

Figure 2: Histograms of the Original Small Sample ($n = 20$) and the S-SMART Samples

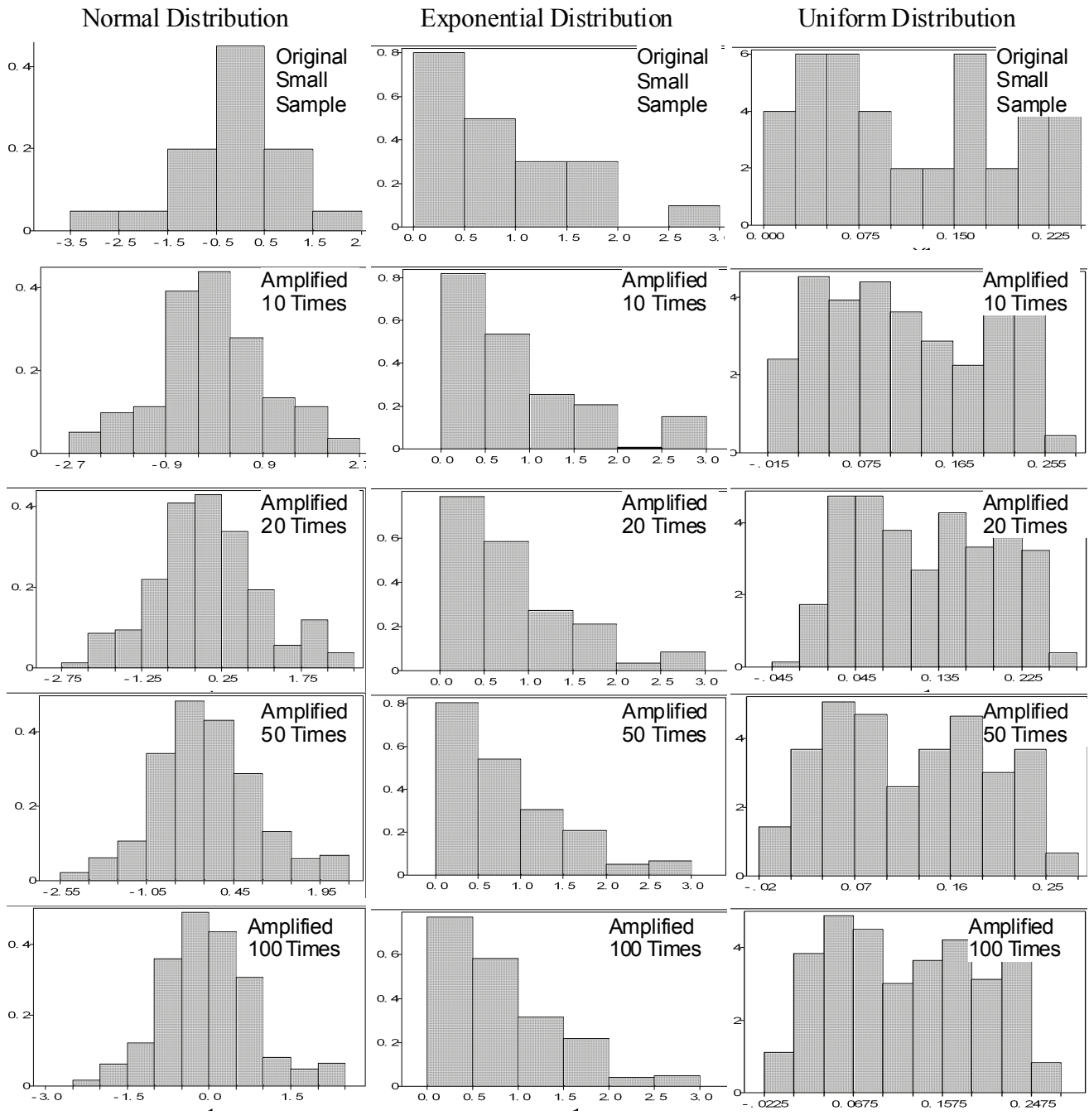Table 1: Statistics Stability Test Results for Normal Data

Small Sample (n)	S-SMART (n)	M	SD	ANOVA		Brown and Forsythe's Test	
				F	p	F	p
10	—	-0.22349274	0.96181667	0.04	0.9971	0.24	0.9143
	100	-0.19069581	0.90544626				
	200	-0.23229805	0.93433640				
	500	-0.21896046	0.89065269				
	1000	-0.21946580	0.88752468				
15	—	0.10073266	1.07928841	0.09	0.9869	1.39	0.2363
	150	0.10655489	1.16529037				
	300	0.09984589	1.10383676				
	750	0.09580675	1.04833640				
	1500	0.10087356	1.02527119				
20	—	0.05855377	1.00343478	0.13	0.9706	0.79	0.5317
	200	0.08582172	1.04418680				
	400	0.06449656	0.98977889				
	1000	0.04226237	0.94223102				
	2000	0.04501257	0.93789847				
25	—	0.06687785	0.95343286	0.21	0.9315	3.64	0.0058
	250	0.06043331	1.04887182				
	500	0.05241659	0.97349343				
	1250	0.08302555	0.89050723				
	2500	0.08860332	0.88523717				
50	—	-0.09631057	0.95323886	0.09	0.9855	4.64	0.0010
	500	-0.08740131	0.98512214				
	1000	-0.08806617	0.91383556				
	2500	-0.07644308	0.87315825				
	5000	-0.07321744	0.86062323				

Table 2: Statistics Stability Test Results for Exponential Data

Small Sample (n)	S-SMART (n)	M	SD	ANOVA		Brown and Forsythe's Test	
				F	p	F	p
10	—	10.7014225	6.89706159	0.11	0.9795	0.14	0.9683
	100	11.1492282	6.82349786				
	200	10.6658978	6.84719726				
	500	10.7544083	6.64931520				
	1000	10.7359794	6.56322669				
15	—	8.99177780	6.68143971	0.10	0.9840	0.16	0.9603
	150	9.29636554	6.81088054				
	300	9.16996847	6.55176270				
	750	9.03166915	6.43617622				
	1500	9.02165396	6.41504747				
20	—	8.48010059	6.72584611	0.05	0.9956	0.09	0.9841
	200	8.57848028	6.73279566				
	400	8.36743416	6.60753442				
	1000	8.42812276	6.50174430				
	2000	8.39118617	6.45948948				
25	—	7.90545531	6.25441137	0.18	0.9465	1.21	0.3050
	250	8.11257241	6.56518979				
	500	7.95441471	6.19530349				
	1250	7.82502044	6.03353920				
	2500	7.82480295	5.94773232				
50	—	8.49794161	7.36349688	0.31	0.8682	1.96	0.0975
	500	8.52943241	7.29392134				
	1000	8.36130973	7.04365418				
	2500	8.24090907	6.73168660				
	5000	8.23225084	6.66201829				

Table 3: Statistics Stability Test Results for Uniform Data

Small Sample (n)	S-SMART (n)	M	SD	ANOVA		Brown and Forsythe's Test	
				F	p	F	p
10	—	0.10577084	0.07134369	0.06	0.9924	0.21	0.9352
	100	0.10730272	0.06681932				
	200	0.10555560	0.06843615				
	500	0.10687757	0.06975516				
	1000	0.10521426	0.06763157				
15	—	0.12313716	0.07294208	0.03	0.9988	0.09	0.9867
	150	0.12102230	0.07446817				
	300	0.12148089	0.07131987				
	750	0.12254982	0.07204738				
	1500	0.12231181	0.07118637				
20	—	0.12922731	0.07208107	0.03	0.9987	0.19	0.9429
	200	0.12916945	0.07348706				
	400	0.12800374	0.06998499				
	1000	0.12892277	0.06975528				
	2000	0.12920833	0.07004994				
25	—	0.13314226	0.06703704	0.03	0.9980	1.24	0.2931
	250	0.13349798	0.06986992				
	500	0.13219283	0.06673523				
	1250	0.13271716	0.06498865				
	2500	0.13314679	0.06537055				
50	—	0.13108854	0.06864826	0.00	1.0000	1.81	0.1238
	500	0.13114757	0.07018810				
	100	0.13115067	0.06762741				
	2500	0.13112839	0.06676879				
	5000	0.13109725	0.06599090				

means and standard deviations of the S-SMART samples with fifty replications for each group. Fifty was the maximum number of replications that could provide a clear enough graphical presentation of the confidence intervals.

As shown in Figure 3, all means and standard deviations of the S-SMART samples were centered at the mean or standard deviation of the original small sample in each group. Furthermore, the confidence intervals of the S-SMART samples covered all the means and standard deviations of the original small samples, except for the standard deviation of the original small sample from the exponential distribution. This problem with the skewed distribution has nothing specifically to do with the S-SMART method. Even for the well-established bootstrap method, the interval estimation for the skewed data also needs adjustment to obtain a better estimation (Hesterberg, 1999).

Evaluating Sampling Distributions

The simulation study shows that the S-SMART sampling distributions closely followed the normal distribution as concluded in the central limit theorem. The histograms in Figures 4 display the sampling distributions of the S-SMART samples based on random small samples from standard normal, exponential, and uniform distributions. For illustration purposes, Figure 4 presents two samples for each distribution to show the property of the S-SMART sampling distributions. The two S-SMART samples were selected based on sample sizes of 10 and 100, representing the sampling distributions of samples with the largest difference among the small sample sizes.

An Empirical Example

S-SMART may be claimed as the easiest resampling method in application because it does not require researchers to learn

Figure 3: 95% CIs of the Mean and Standard Deviation with 50 Replications for the Amplified Samples (Li, i = 1, 2, …, 50; n = 200) and the Original Small Sample ($S0$; n = 20)
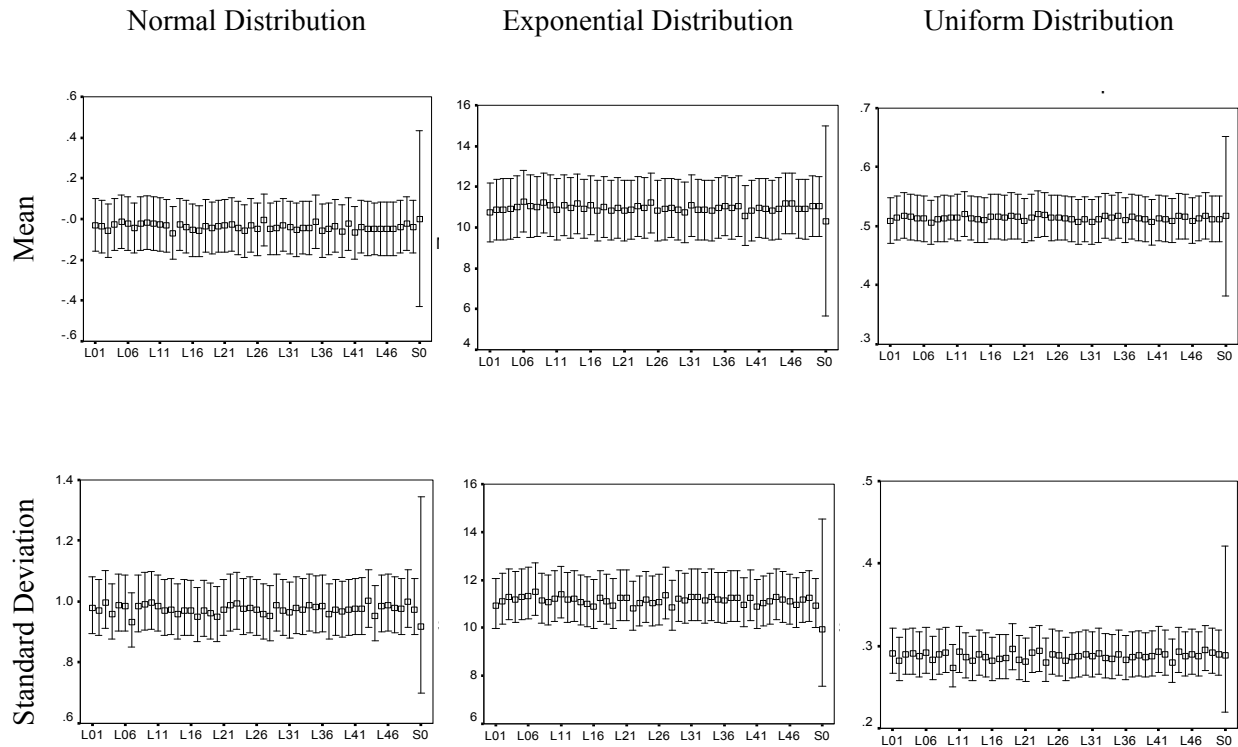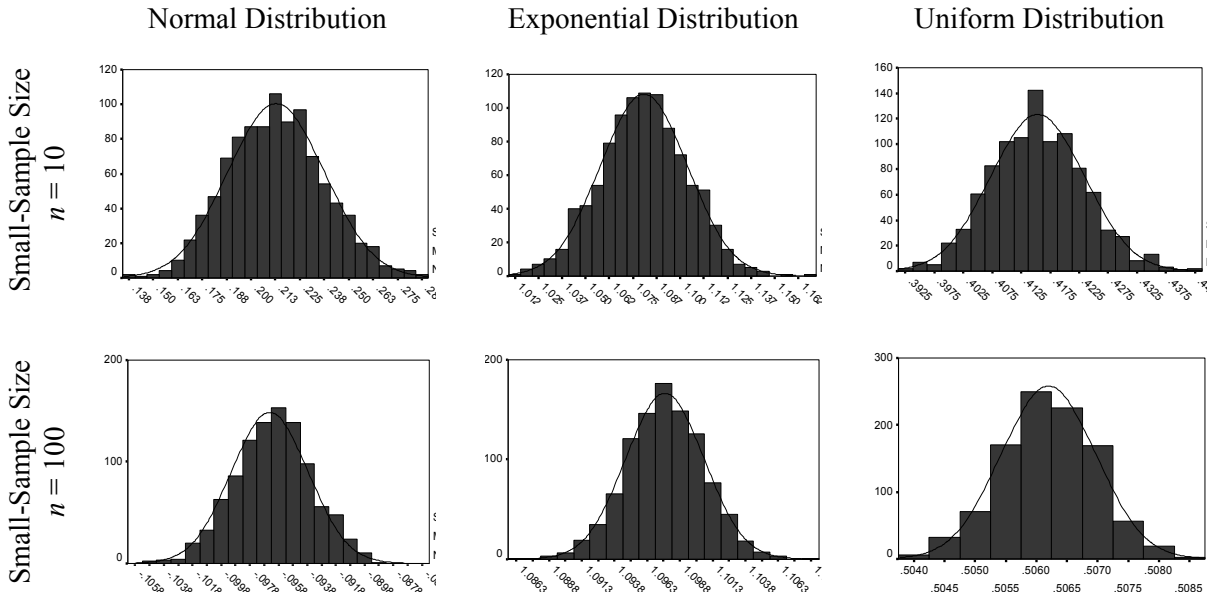
Figure 4: Histograms of 1,000 S-SMART Samples Based on Small Samples Sizes of 10 and 100

any computer programming or model modifications to obtain an adequate sample size to conduct statistical analysis. At the current stage, a SAS macro program is ready for researchers to directly plug in their small sample to get the amplified S-SMART samples. A stand-alone computer program will be available soon. In this article, the SAS macro is presented as an example for the application of the S-SMART method.

To use the SAS macro, researchers need only to specify five macro variables in a short SAS macro: *%S_SMART* (in=, k=, NUM=, Var=, outfile=) to invoke the S-SMART macro *%macro S_SMART* (in=, k=, NUM=, Var=, outfile=) which is available from the first author. Researchers simply plug in the small sample data file after *in =*, the small sample size after *k =*, the times to amplify the small sample after *NUM =*, the name of the variable in the small sample after *var =*, and the output file to save the amplified S-SMART sample after *outfile =*. After running the SAS macro program, the S-SMART sample is ready for statistical analysis.

To illustrate how to obtain an S-SMART sample from an empirical small sample using the SAS macro program, a random sample

(*N* = 269) was drawn from a real, large-scale study of education: the National Educational Longitudinal Study of 1988 (NELS:88) database. A small sample of 20 cases was further randomly drawn from the random sample. An achievement variable, *bytxhstd* (base year history/citizenship/geography standardized test score), was used and was renamed as *BHIST20* (see Table 4).

If it is desired to amplify the 20 cases into 200 (10 times), the small sample data file is C:\NELS20hist.dat, the variable name is *BHIST20*, and the amplified output data file is C:\NELS20_200hist.txt. At this point the amplified S-SMART data can be obtained by plugging in the five macro variables into the SAS macro as follows: %*S_SMART* (in = C:\NELS20hist.dat, k = 20, NUM = 10, Var = BHIST20, outfile = C:\NELS20_200hist.txt).

This SAS macro program invokes the S-SMART macro *% macro S-SMART* (in=, k=, NUM=, Var=, outfile=); then the amplified S-SMART data in a text file is saved as C:\NELS20_200hist.txt. To study the property of the sample distribution of the S-SMART sample from the empirical data, histograms of the S-SMART sample were compared with the

small sample and the random NELS:88 sample. The shape of the small sample distribution (Figure 5, left panel) roughly reflected the random sample of NELS:88 (Figure 5, right panel), but it had a gap between the scores of 33 and 38; while the sample distribution of the S-SMART sample (Figure 5, middle panel) closely followed the shape of the small sample with a similar gap.

To compare the means and variances between the small sample of size 20 from the random sample of the NELS:88 dataset, the corresponding S-SMART sample, and the random sample of NELS:88, ANOVA test and the homogeneity test were conducted over the three sets of data. The test results are shown in Table 5. From the ANOVA results it was found that there were no mean differences among the three groups of data with a p-value of .189 even with the unbalanced group. With two random errors, the sampling error and the Monte Carlo error, the S-SMART sample still reflect the sample mean of the small sample and the

mean of the random sample from which the small sample was selected. When the means of the amplified S-SMART sample were compared with the small sample using the t-test, a p-value of .993 resulted, thus revealing the equality of the two sample means.

The results of the Brown and Forsythe's test revealed that equal variances were assumed among the three groups of data with a p-value of .762. Under this condition, it was found that the mean standard error (.68) of the S-SMART sample is very close to that (.59) of the random sample, and as expected, it is much smaller than that (2.07) of the small sample. This finding demonstrates that the S-SMART method has the potential to reduce sampling errors while maintaining all other statistical properties carried by the small sample.

To explore the property of the interval estimation of the S-SMART sample for the empirical data, the interval estimation of the means among the small sample, the S-SMART sample, and the random sample were compared.

Table 4: The Small Sample of 20 Cases from the NELS:88 Database

No.	*BHIST20*	No.	*BHIST20*	No.	*BHIST20*	No.	*BHIST20*
1	69.508	6	59.132	11	52.907	16	47.009
2	67.761	7	57.385	12	52.907	17	46.026
3	64.266	8	56.402	13	50.396	18	46.026
4	60.770	9	55.747	14	47.883	19	40.020
5	59.460	10	53.781	15	47.883	20	31.392

Figure5. Histogram of the Scores in Base Year History of 20 Cases from NELS:88 and S-SMART Samples
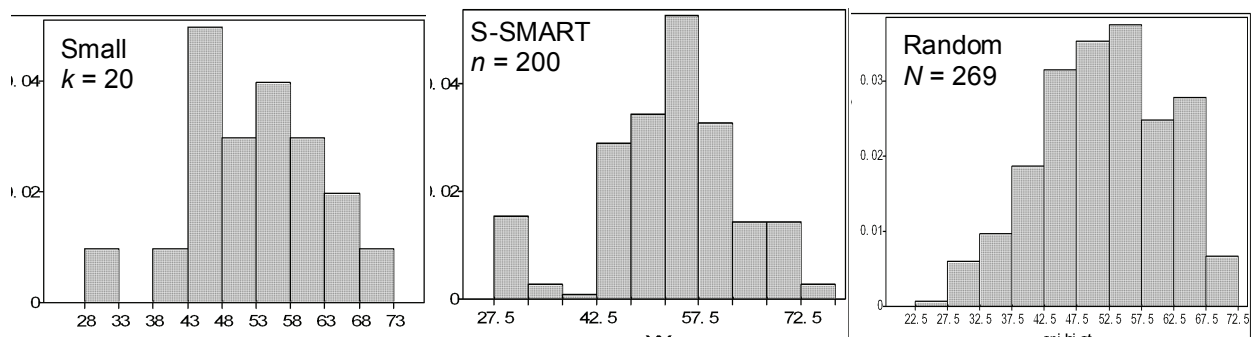
Table 6 shows that S-SMART interval estimation of the mean is much narrower than that of the small sample and very close to the random sample. This result further reveals that if a small sample is randomly selected from a population, the S-SMART method can replenish the small sample to obtain a larger sample with the same statistical properties as those of any random sample of a comparable sample size from the same population.

In sum, the application of the S-SMART method to real-world data demonstrated that the newly-developed resampling method can be utilized in the real-world settings. The evaluation on the quality of the S-SMART sample yielded the same results as those of from the simulated data. In other words, the S-SMART sample generated from the real-world data has the same sample distribution as that of the original small sample; furthermore, the S-SMART method can replenish an original real-world small sample to a larger sample with the same sample distribution, mean and standard deviation, while the standard error is reduced.

Conclusion

This study developed a new resampling method, S-SMART, which can replenish a small sample becoming a larger sample to meet the requirement of sample size for inferential statistics. It allows researchers to use familiar statistical analysis directly on the amplified S-SMART samples.

S-SMART is a Distribution-Free Method

According to the theoretical verification, simulation study, and empirical evidence of distributional characteristics of the amplified S-SMART samples, it was also demonstrated that S-SMART is a distribution-free method. From the distributional study it was found that the S-SMART method can amplify an original small sample from any distributions into a larger sample with the same distribution as that of the original small sample. Regardless of whether the sample distribution of the original small samples are symmetric or asymmetric, the sample distribution of the S-SMART sample follows the same distribution as does the original sample; and the sampling distributions of the S-SMART samples are normal. The use of the Gaussian kernel smoothing over the percentile range from 2.5% to 97.5% of the original small sample captured the shape of the original small-sample distribution.

It may be argued that the S-SMART sample copies the sampling bias caused by the small size of the original sample. However, if a

Table 5: Comparisons of Basic Statistics

Sample	N	M	SD	Std. Error of Mean	ANOVA		Brown and Forsythe's Test	
					F	P	F	P
Small Sample	20	53.3331	9.2380	2.0657				
S-SMART Sample	220	53.3136	10.1527	.6845	1.672	.189	.502	.762
Random Sample	269	51.7111	9.6217	.5911				

Table 6: Comparisons of Interval Estimation of Mean

Sample	N	M	SD	Std. Error of Mean	CI Lower	CI Upper
Small Sample	20	53.3331	9.2380	2.0657	49.0059	57.6566
S-SMART Sample	220	53.3136	10.1527	.6845	51.9645	54.6626
Random Sample	265	51.7111	9.6217	.5911	50.5474	52.8749

small sample is not representative of the population due to the sampling bias, the sample is not valid to be used for any statistical analysis or any other resampling methods; therefore, in the case of sampling bias, researchers' judgment must be relied upon to assess the quality of the data collected.

S-SMART is a Robust Procedure

From the simulation study, it was found that the S-SMART samples based on the original small samples from various distributions are robust to outliers. By using the middle 95% of percentiles instead of all the data points in the original sample, the S-SMART technique can reduce the influence of the extreme values or potential outliers in the original small sample. At the same time, some beneficial information carried by the extreme values can be retained through the estimation of the percentiles from the original sample.

The S-SMART Sample Statistics Are Stable

The results of the F-tests in ANOVA of the simulation study have shown that the means of the S-SMART samples are statistically equal across all different sizes of replenished samples for each of the different sizes of the original small samples. The stability of the means of the S-SMART samples makes the method reliable in representing the mean values of the original small samples at any times of amplification. The F-tests also show that the homogenous S-SMART sample means are not significantly different from that of the original small sample: this property of imitating the mean of the original small sample makes the S-SMART sample reflect the essential statistic of the original small sample well.

As evident in the Brown and Forsythe's Tests, with few reasonable exceptions, the S-SMART samples have equal variances across almost all different sizes of replenishing samples for each of the different sizes of the original small samples; and the stable S-SMART sample variances are not significantly different from that of the original small sample. The similarity between the homogeneous sample variances of the S-SMART samples and the original small sample makes the S-SMART sample closely mirror the original small sample for its variability. The variation of the S-SMART sample come from two sources: one is from the random errors and the other from the simulation procedure. However, even with the two sources of variation, the S-SMART procedure still produces the amplified samples with the similar variation to that of the original small sample. The stability of the amplified sample together with its robustness to the influence of outliers makes the S-SMART sample representative of the population or local data from which the original small sample is drawn.

The S-SMART Sampling Distribution is Normal

The sampling distribution of the mean of the S-SMART sample is also examined through a series of histograms. The sampling distributions for the S-SMART data from both symmetric and asymmetric distributions are normal as expected from the central limit theorem. The means of the S-SMART samples distribute normally and center around the small sample means. Even though the shape and dispersion depend on the original sample, the variation from sample to sample is not noticeable.

The S-SMART Samples Produce Accurate Confidence Intervals

The confidence intervals for both the mean and standard deviation of the amplified samples produced by the S-SMART method have been studied. The findings in the confidence interval study are promising in that the confidence intervals of the S-SMART samples aptly covered the means. The confidence intervals for the mean of the S-SMART sample are more accurate with the narrower range than the confidence intervals of the original samples for the symmetric and asymmetric distributions. The confidence intervals for the standard deviation of the S-SMART sample are better than those of the original small sample, with the exception of the skewed distribution. It requires more research on the adjustment of the skewness in the S-SMART procedure to make a better estimation for the confidence intervals.

Implications

The findings suggest that the S-SMART procedure has potential to lessen some limitations of the existing resampling methods. First, S-SMART can reduce the influence of outliers, a problem from which the bootstrap method has long suffered. It is known that outliers can severely influence statistics such as mean and variance. Reducing outlier influence can greatly improve the validity of statistical inferences, thus improving the quality of quantitative studies with small samples. Second, the S-SMART sample is the union of the sub-samples randomly generated from the Gaussian kernels centered at the quantiles with a random noise instead of repeatedly selecting resamples from the same data points in the original small sample; therefore, it has independent observations conditionally on the original small sample. Third, the S-SMART procedure produces amplified samples with larger statistical power than the original small sample. As is known, small samples suffer from problems of small statistical power in detecting significant effects of interventions. When only small samples are available, researchers can directly apply the amplified S-SMART sample to statistical analysis in their research to draw more accurate statistical inferences than using the original small sample.

Some researchers may have a concern that the S-SMART method would produce samples with too large power. It is true that researchers can amplify their small sample as large as they wish. However, samples size for any statistical analysis should be determined by a statistical power analysis (Cohen, 1988). The S-SMART is the right tool to help researchers amplify their small samples as large as required by the corresponding statistical power.

In sum, because the S-SMART samples are unbiased in imitating the original sample in terms of distributions and statistical behaviors with less influence of outliers through its robust procedures, they can better represent the population or local data from which the small samples are drawn. With this property the S-SMART samples have the potential to be used for any kind of statistical analysis in quantitative research with small samples.

Limitations

Like all other resampling methods, the S-SMART method relies on how well the small sample represents its population. Because S-SMART produces amplified samples based on the original small samples, if the original small sample is randomly selected and represents its population, S-SMART can produce the corresponding amplified samples representing the population; however, if the original small sample is not randomly selected, the S-SMART can only produce the corresponding amplified samples similar to the original small sample in terms of distribution and other statistical behaviors locally. In this case, the statistical analysis using the S-SMART samples cannot provide reliable statistical inferences to be generalized to the population even though the sample size is amplified. With this concern, researchers should judge the quality of their samples to see if their original small samples are randomly selected so that the statistical results can be generalized to the population; otherwise, statistical results from either the original small sample or the amplified S-SMART sample are only valid locally.

It is also worth noting that the S-SMART method has some restrictions for the sample sizes when estimating the confidence interval of the standard deviation of the data amplified from a skewed population distribution. This limitation requires further investigations. In addition, again like all other resampling methods, the present research of the S-SMART method lacks in-depth mathematical derivations. Adding to the numerical evidence from the simulation study and empirical example, mathematical investigations on the equalities of the means and variances between the S-SMART samples and the corresponding original small sample would make the study of S-SMART stronger.

Further Research

Simulation study on the new resampling method S-SMART has produced promising findings; however, it is desirable to have more mathematical investigations on sample distributions, sampling distribution, sample means, standard errors, and sampling bias. This

study focused on the basic concept, simulation procedures, and verification of the newly-developed S-SMART method; therefore, the simulation study was only conducted over the small samples with univariate data. Besides the current univariate small sample simulation study, investigations with the S-SMART method to amplify multivariate data is in progress.

Future studies could also involve more real-world data to examine how to solve real research questions with the S-SMART samples and thereafter to compare the data analysis results from the S-SMART samples with the results from the other resampling methods. In addition to the above considerations for future studies, it is also desirable to compile a stand-alone computer program package with a user-friendly interface.

References

Adcock, C. J. (1997). Sample size determination: A review. *The Statistician*, *46*(*2*), 261-283.

Bai, H., & Pan, W. (2008). Resampling methods revisited: Advancing the understanding and applications in educational research. *International Journal of Research and Methods in Education*, *31*(1), 45-62.

Brown, M. B., & Forsythe, A. B. (1974). Robust tests for equality of variances. *Journal of the American Statistical Association*, *69*, 364 - 367.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, (*2ⁿᵈ Ed.*). Hillsdale: Lawrence Erlbaum.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago: Rand McNally College Publishing Company.

Davison, D. C., & Hinkley, D. V. (1999). *Bootstrap methods and their application*. New York: Cambridge University Press.

de Angelis, D., & Young, G. A. (1992). Smoothing the bootstrap. *International Statistical Review*, *60*, 45-56.

Diaconis, P., & Efron, B. (1983). Computer-intensive methods in statistics. *Scientific American*, *248,* 96-108.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annual of Statistics*, *7*, 1-26.

Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans, *Number 38 in CBMS-NSF Regional Conference Series in Applied mathematics*. Philadelphia: SIAM.

Efron, B. (1994). Missing data, imputation, and the bootstrap (with discussion). *Journal of the American Statistical Association*, *89*, 463-470.

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, *37*(*1*), 36-48.

Efron, B., & Tibshirani, R. J. (1998). *An introduction to the bootstrap*. New York: Chapman & Hall, Inc.

Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyd.

Gilchrist, W. G. (2000). *Statistical modelling with quantile functions*. Boca Raton, FL: Chapman & Hall/CRC.

Gravetter, F. J., & Wallnau, L. B. (2004). *Statistics for the behavioral sciences* (*6ᵗʰ Ed.*). Belmont, CA: Wadsworth/Thomson Learning.

Hall, P., DiCiccio, T. J., & Romano, J. P. (1989). On smoothing and the bootstrap. *The Annals Statistics*, *17*, 692-704.

Hansen, M. H., Madow, W. G., & Tepping, B. J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, *78*(384), 776-793.

Hesterberg, T. C. (1999). *Smoothed bootstrap and jackboot sampling* (Research Rep. No. 87). Seattle, WA: MathSoft, Inc.

Hinkley, D. V. (1988). Bootstrap methods (with discussion). *Journal of Royal Statistical Society*, *Serial B*, *50*, 321-337.

Johnson, M. E. (1987). *Multivariate statistical simulation*. New York: John Wiley & Sons.

Kurtz, A. K. (1948). A research test of Rorschach test. *Personnel Psychology*, *1*, 41-51.

Lahiri, S. N. (2003). *Resampling methods for dependent data*. New York: Springer-Verlag.

Lindley, D. V. (1997). The choice of sample size. *The Statistician*, *46*(*2*), 129-138.

O'Brien, P. C. (1992). Robust procedures for testing equality of covariance matrices. *Biometrics*, *48*, 819–827.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, *33*(*3*), 1065-1076.

Quenouille, M. G. (1949). Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society, 11*, 68-84.

SAS Institute Inc. (2001). *SAS/STAT user's guide, Version 8.02*. Cary, NC: SAS Institute Inc.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.

Silverman, B. W., & Young, G. A. (1987). The bootstrap: to smooth or not smooth. *Biometrika, 74*, 649-479.

Simonoff, J. S. (1996). *Smoothing methods in statistics*. New York: Springer.

Stine, R. (1989). An introduction to bootstrap methods. *Sociological Methods and Research, 18*(2), 243-291.

Tabatabai, M. (1985). Robust procedures for comparing several means and testing for parallelism of several straight lines under heteroscedasticity and nonnormality. *Dissertation Abstracts International Part B: Science and Engineering, 45* (10), 95.

Troendle, J. F., Korn, E. L., & McShane, L. M. (2004). An example of show convergence of the bootstrap in high dimensions. *The American Statistician, 58,* 25-29.

Tukey, J. W. (1958). Bias and confidence in not quite large samples (Abstract). *The Annals of Mathematical Statistics*, *29*, 614.

Wu, C. F. J. (1986). Jackknife bootstrap and other resampling plans in regression analysis (with discussion). *The Annals of Statistics, 14*, 1261-1350.