Journal of Modern Applied Statistical Methods

Volume 1 | Issue 1 Article 8

5-1-2002

A Measure Of Relative Efficiency For Location Of A Single Sample

Shlomo S. Sawilowsky
Wayne State University, shlomo@wayne.edu

Recommended Citation
Sawilowsky, Shlomo S. (2002) "A Measure Of Relative Efficiency For Location Of A Single Sample," *Journal of Modern Applied Statistical Methods*: Vol. 1 : Iss. 1 , Article 8.
DOI: 10.22237/jmasm/1020254940

Journal Of Modern Applied Statistical Methods
Winter 2002, Vol. 1, No. 1, 52-60

1538 - 9472/02/$30.00

A Measure Of Relative Efficiency For Location Of A Single Sample

Shlomo S. Sawilowsky
Evaluation & Research
College of Education
Wayne State University

The question of how much to trim or which weighting constant to use are practical considerations in applying robust methods such as trimmed means (L-estimators) and Huber statistics (M-estimators). An index of location relative efficiency (LRE), which is a ratio of the narrowness of resulting confidence intervals, was applied to various trimmed means and Huber M-estimators calculated on seven representative data sets from applied education and psychology research. On the basis of LREs, lightly trimmed means were found to be more efficient than heavily trimmed means, but Huber M-estimators systematically produced narrower confidence intervals. The weighting constant of $\psi = 1.28$ was found to be superior to various competitors suggested in the literature for $n < 50$.

Keywords: Huber M-estimator, ψ, Trimmed mean, Point estimator, Relative efficiency

Introduction

The development of robust methods in the past thirty-five years has produced a plethora of modern techniques for succinctly describing the most salient characteristics of a data set. With regard to measures of central tendency (location), the familiar mean, median, mode, and geometric mean have been augmented with L-estimators (*L*inear combination of order statistics), M-estimators (generalized *M*aximum likelihood), and R-estimators (inverted tests on *R*anks of absolute values).These modern methods are examples of the "outright rejection" and the "accommodation" approaches to handling outliers (Barnett & Lewis, 1994, p. 29). To understand the concepts of these robust methods, the mathematically inclined reader is referred to Hampel, et al. (1986), Hoaglin, Mosteller, and Tukey (1983), Huber (1981), or Staudte & Sheather (1990). An excellent "first course" textbook introducing L- and M-estimators is Wilcox (1996).

Consider trimmed means, which are typical of L-estimators. The trimmed mean is calculated by sorting the data set, trimming a certain percentage of observations from the top and bottom of the scores, and calculating the average of the remaining scores. For example, the arithmetic mean of a data set containing the scores 75, 71, 70, 76, 72, 73, 73, 70, 30, and 74 is 68.4. A 2x10% trim (symmetric trim of 10% of the smallest and largest observations) is calculated as follows:

1. sort observations: 30, 70, 70, 71, 72, 73, 73, 74, 75, 76
2. trim 10% × 10 scores = 1 score from both ends: 70, 70, 71, 72, 73, 73, 74, 75
3. calculate average:

$$\frac{70+70+71+72+73+73+74+75}{8} = 72.25$$

In this example, the 2x10% trimmed mean is shown to be resistant to the extreme score (30), resulting in a value of 72.25 which is more indicative of bulk of the scores. The arithmetic mean, in contradistinction, chased after the extreme low score, resulting in a measure of location that was lower than ninety percent of scores in the data set. Thus, the arithmetic mean is said to have a low breakdown point, because it is strongly influenced by even a single value, such as an outlier.

The data analyst might wonder if a different amount of trim would improve the estimate of location. The literature on this question is divided into two camps: the "heavily trim" (e.g., a 2x25% trim was recommended by Rosenberger & Gasko, 1983, p. 332-333; a 2x20% trim was adopted by Wilcox, 1996, p. 16; 1998) and the "lightly trim" (either a 2x10% or 2x5% trim, considered by Hill & Dixon, 1982; Huber, 1977, p. 1090; Stigler, 1977, p. 1063; Staudte & Sheather, 1990, p. 133-134). Simulation evidence on a contaminated normal distribution indicated that the variances of trimmed means were minimized (and are thus one measure of the optimum trim or "optrim") for sample size n = 10 when the trim was 16.1%; for samples of size n = 20 it was almost half as much, as the optrim was 8.7% (Rosenberger & Gasko, 1983, p. 319). The variance of estimators was minimized for data sampled from the double exponential and Cauchy distributions for samples of size n = 10 with optrims of 34% and 40% (p. 330), respectively, and was 37% and 39% (p. 331), respectively, for samples of size n = 20. Wilcox (1996) noted, "Currently there is no way of being certain how much trimming

Shlomo S. Sawilowsky, Wayne Statue University Distinguished Faculty Fellow, is Professor and Chair, Educational Evaluation and Research, 351 College of Education, Wayne State University, Detroit, MI, 48202, e-mail: shlomo@edstat.coe.wayne.edu. His interests are in non-parametric, robust, exact, permutation, and other computer-intensive methods, especially via Fortran.

should be done in a given situation, but the important point is that some trimming often gives substantially better results, compared to no trimming" (p. 16).

The problem of selecting parameters in the robust measures literature, such as how much to trim, is not restricted to L-estimators. As an example with M-estimators, there are many choices pertaining to the weighting constant ψ (also referred to as the bending constant or the tuning factor) used in the one-step Huber M-estimator. The M-estimator has a high breakdown point, determines empirically how much and from which side to trim, and has other desirable properties (Wilcox, 1996, p. 146, 204).

The formula for the one-step Huber M-estimator, with a weighting constant of ψ = 1.28, is

$\text{Huber}_{\psi 1.28}$ M-Estimator =

$$\frac{1.8977(MAD)(i_2 - i_1) + \sum_{i=i_1+1}^{n-i_2} x_i}{n - i_1 - i_2} \qquad (1)$$

It is calculated on the ten scores as follows:

1. sort observations:30, 70, 70, 71, 72, 73, 73, 74, 75, 76
2. calculate median: 72.5
3. calculate MAD (Median Absolute Difference):
   a. calculate $|x_i$ - median$|$: |30-72.5|=42.5, |70-72.5|=2.5, |70-72.5|=2.5, |71-72.5|=1.5, |72-72.5|=.5, |73-72.5|=.5, |73-72.5|=.5, |74-72.|=1.5, |75-72.5|=2.5, |76-72.5|=3.5
   b. sort results: .5, .5, .5, 1.5, 1.5, 2.5, 2.5, 2.5, 3.5, 42.5
   c. MAD = median from step b: 2
4. calculate $\frac{.6745(x_i - median)}{MAD}$

$$\frac{.6745(30-72.5)}{2} = -14.333$$

$$\frac{.6745(70-72.5)}{2} = -.843$$

$$\frac{.6745(70-72.5)}{2} = -.843$$

$$\frac{.6745(71-72.5)}{2} = -.506$$

$$\frac{.6745(72-72.5)}{2} = -.169$$

$$\frac{.6745(73-72.5)}{2} = .169$$

$$\frac{.6745(73-72.5)}{2} = .169$$

$$\frac{.6745(74-72.5)}{2} = .506$$

$$\frac{.6745(75-72.5)}{2} = .843$$

$$\frac{.6745(76-72.5)}{2} = 1.180$$

5. count i_1 (number of values in step 4 < -1.28): 1 (observation a)
6. count i_2 (number of values in step 4 > 1.28): 0
7. calculate $\text{Huber}_{\psi 1.28}$:

$$= \frac{1.8977(2)(0-1)+(70+70+71+72+73+73+74+75+76)}{10-1-0}$$

= 72.24

The one-step M-estimator is nearly the same as the median in this illustration. Wilcox noted that typically the M-estimator is between the median and the mean (Wilcox, 1996, p. 147). (The constants .6745 and 1.8977 appearing in the calculation of the one-step Huber $_{\psi 1.28}$ M-estimator refer to the inverse cumulative distribution function for $x = 0.75$ for the standard normal curve ($\mu = 0$, $\sigma = 1$), and $\frac{\psi}{.6745}$, respectively. This formula, and its constants, are given by Wilcox, 1996, p. 147. It should be pointed out that the second constant, 1.8977, pertains only to ψ = 1.28. For example, use 2.2239 in calculating the one-step $\text{Huber}_{\psi 1.5}$ M-estimator.)

Some commonly used weights for the one-step Huber M-estimator include the following: (a) 1.28, the value used in the illustration, which corresponds to the .9 quantile of the standard normal distribution (Staudte & Sheather, 1990, p. 117), (b) 1.339, used by a popular statistics software package (*SPSS*, 1996), (c) 1.345, because it represents the "0.95 asymptotic efficiency on the standard normal distribution" (Rey, 1980, p. 108-109), (d) 1.4088 (Hampel, et al., 1983, p. 167), and (e) 1.5, (Andrews, et al., 1972, p. 13), which Stigler (1977) stated "does particularly well" (p. 1064). Other values are cited less frequently in the literature (e.g., Huber, 1964, p. 84-85, examined the upper bounds of the asymptotic variance for ψ = 0 - 3.0 (.1)), but in the absence of current usage they are not considered further.

The question remains as to which weighting constant should be used in practice. As noted by Lee (1995), the efficiency of M-estimation will "depend on the choice of ψ" (p. 521). The casual approach in choosing ψ was summarized by Hogg (1977), who stated that for optimum performance, "let the k *[ψ]* of the Huber statistic decrease as measures of skewness and kurtosis... increase; that is, for illustration, use Huber's P20 *[ψ = 2.0]*, P15 *[ψ = 1.5]*, and P12 *[ψ = 1.2]*, respectively, as those measures increase. Users of the M-estimators actually seem to do this in practice anyway... to suit the problems at hand" (p. 1089, bracketed material added for clarification). The casual approach is obviously not satisfactory.

Location Relative Efficiency (LRE)

The issues raised above in terms of how much to trim and which weighting constant to use are important because they relate to the primary question of which procedure and which parameters of a procedure to use in estimating location in the single sample problem. Stigler's (1977) approach toward discovery of the best estimator was "to measure the absolute magnitude of an estimator's error relative to the size of errors achieved by other estimators" (p. 1062) for some famous physical science data sets. However, this technique is only applicable in the improbable situation where θ, the population parameter for location, is exactly known. This is necessary in order to measure the variability of error, ε, which is taken to be $(\theta - \hat{\theta})$, the difference between the actual location parameter and the estimated location. (For other limitations, see Andrews, 1977, p. 1079; Hoaglin, 1977, p. 1087; Huber, 1977, p. 1091; and Pratt, 1977, p. 1092.)

Another method is to quantify the comparative efficiency of robust measures with their competitors. An example is the Cramér-Rao efficiency, which is the ratio of the lower bound of a competitor with that of the best estimator. Another method, proposed by Gastwirth and Cohen (1970), is the ratio of the variance of the best estimator divided by the variance of the competitor. Their statistic is called the relative efficiency. (See Rosenberger & Gasko, 1983, p. 327, for further discussion, and Staudte, 1980, p. 15, for a typical application.)

A problem with these two techniques is the need to know the best estimator, which if known, of course, would obviate the initial question. The best estimator is dependent on a variety of factors, including the distribution and sample size. In practice, unfortunately, the evidence cited in favor of the best estimator is typically just an assertion. Among other difficulties, the lower bound of the Cramér-Rao efficiency is frequently impossible to obtain. The Gastwirth and Cohen relative efficiency depends on the asymptotic variance, as opposed to the actual variance. (See Hampel, et al., 1986, p. 398, for a polemic favoring asymptotic variance over the actual variance, when taken together with asymptotic normality.)

A statistic was recently proposed (Sawilowsky, 1998) for determining the comparative efficiency of a location estimator for a single sample that precludes the necessity of knowing the best estimator. It avoids problems associated with assuming asymptotic normality in order to generalize asymptotic variances to actual variances. It is not hampered by limitations associated with the arithmetic mean and its impact on actual variances. Finally, it is simple to compute. This index is called the Location Relative Efficiency (LRE).

The impetus for the LRE was from Dixon and Tukey (1968). They calculated the 95% bracketed interval for the mean and for one through five units of Winsorized means. Then, they compared the length of the intervals of the Winsorized means with the length of the confidence interval associated with Student's t, characterizing this ratio as the "apparent efficiency" (p. 86). This is a Type III performance measure of a bracketed interval in the classification scheme discussed by Barnett and Lewis (1994; see their discussion for Types I and II), where they noted that, "a natural measure of its efficiency is the ratio of the lengths of the intervals" (p. 75). Indeed, Huber (1972) noted that, "While for years one had been concerned mostly with what was latter called 'robustness of validity' (that the actual confidence levels should be close to, or at least on the safe side of the nominal levels), one realized that 'robustness of performance' (stability of power, *or the length of confidence intervals*) was at least as important" (p. 1045-1046, emphasis added).

The LRE for the 95% bracketed interval is defined as:

$$LRE = \frac{Huber_{\psi 1.28(U95\%C.I.)} - Huber_{\psi 1.28(L95\%C.I.)}}{X_{(U95\%C.I.)} - X_{(L95\%C.I.)}} \quad (2)$$

where the LRE is the range (U = upper bound, L = lower bound) for the 95% bracketed interval for the one-step Huber M-estimator divided by the range for the 95% bracketed interval of the competitor. The $Huber_{\psi 1.28}$ is not asserted to be the best estimator. Rather, the resulting ratio may be greater than or less than one. LREs greater than one indicate the competitor yields confidence intervals that are narrower and thus more efficient than the $Huber_{\psi 1.28}$. LREs less than one indicate the $Huber_{\psi 1.28}$ is more efficient in that it produces narrower bracketed intervals.

The LRE was used to evaluate the performance of bracketed intervals produced by a variety of procedures (Sawilowsky, 1998). Four data sets given in Staudte and Sheather (1990, p. 133-137) were analyzed and the results were as follows: Huber $_{\psi 1.28}$ (1.000), 2x10% trim (.945), Sign (.895), Wilcoxon (.869), 2x5% trim (.862), and Student's t (.625). However, note that the location parameter θ is not necessarily the same for these varied procedures (although Pratt argued that "It doesn't matter what an estimator estimates, as long as it is a location parameter", 1977, p. 1092); and these results were based on four small data sets. The data sets pertained to empirical measures of the velocity of light (n = 66), the percentage of seafood in a product to determine if it complied with proposed labeling guidelines (n = 18), the proportion of DDT in kale (n = 15), and Darwin's data on plant height (n = 15).

The physical science data sets explored by Stigler (1977) leave the same question "of whether these data adequately reflect anticipated applications, for example in the *social sciences*. I have little to add to my previous comments on this, other than to reiterate that I would welcome evidence on this point" (Stigler, 1977, p. 1098, emphasis added). The same is true of most published simulation work in the robustness literature, which was conducted on theoretical distributions such as the deMoivreian (Gaussian), logistic, one-out (one score has scale three times the rest), one-wild (one score has scale ten times the rest), double exponential, and Cauchy (e.g., Rosenberger & Gasko, 1983, p. 326-330). (The reliance on theoretical distributions has led to ridiculous statements, such as Hampel et al., 1986, citing Huber to say the following about the t distribution with three degrees of freedom: "t_3 is a suitable example for what high-quality data can look like", p. 23!)

Micceri (1986, 1989) canvassed the education and psychology literature and highlighted representative distributions as the most prevalent in social and behavioral science research. (These data sets were previously investigated, in terms of their impact on the t-test, by Sawilowsky & Blair, 1992.) Descriptive statistics on the data sets, in the order that they were presented by Micceri (1986), are compiled in Table 1. The ordering does not, however,

Table 1. Descriptive Statistics For Seven Real Data Sets from Micceri (1986).

Distribution	N	Median	μ	SE	σ	γ_1	γ_2	$Z_{U95\%CI}$	$Z_{L95\%CI}$
Achievement Smooth Symmetric	5,375	13	13.186	.013	4.907	.005	-.340	13.211	13.160
Achievement Discrete Mass At Zero	2,429	13	12.919	.019	4.415	-.034	.312	12.956	12.881
Achievement Extreme Asymmetry	2,768	27	24.497	.018	5.788	-1.330	1.106	24.553	24.462
Psychometric Extreme Asymmetry	2,047	11	13.667	.021	5.754	1.638	1.522	13.709	13.626
Achievement Digit Preference	3,063	535	536.900	.680	37.644	-.065	-.240	536.981	536.914
Psychometric Extreme Bimodality	665	4	2.971	.037	1.687	-.078	-1.696	3.044	2.899
Achievement Multimodality Lumpy	467	18	21.148	.044	11.917	.194	-1.199	21.234	21.062

Notes: μ = population mean, σ = population standard deviation, SE = standard error of the mean, γ_1 = skew, γ_2 = kurtosis, $Z_{U95\%CI}$, $Z_{L95\%CI}$ = Upper and Lower 95% Confidence Interval based on Z and the SE. Parameters are reported here in accordance with Micceri (1986), who took the position that the data sets were considered of sufficient size to proxy the population.

reflect prevalence of occurrence. Clearly, the physical science data sets and convenient theoretical distributions mentioned above have little relevance for the applied social and behavioral science researcher.

Purpose of the Study

In the current study, the data sets provided by Micceri (1986, 1989) are used to assess the LRE of some robust methods of estimating location of a single sample. The first question to be considered is whether the "heavily trim" or the "lightly trim" approach is more efficient in estimating location in real education and psychology data sets. Only symmetric trimming is considered because the applied researcher will most likely not have à priori knowledge of whether or not the parent population from which the data were sampled is asymmetric, and therefore will not know which side to trim. The second question is whether Huber's M-estimator is more efficient in estimating

Table 2. Median Location Relative Efficiency For Various Robust Measures Of Location and Sample Sizes For The Real Data Sets From Micceri (1986); Huber$_{\psi 1.28}$ = 1.000; 1,000 Repetitions.

	n									
Statistic	10	20	30	40	50	60	70	80	90	100
Smooth Symmetric (Achievement)										
Huber$_{\psi 1.339}$	.983	.983	.985	.986	.991	.994	.994	.993	.997	1.000
Huber$_{\psi 1.345}$	.979	.982	.986	.992	.996	.996	.997	.997	1.000	1.000
Huber$_{\psi 1.4088}$	.971	.984	.996	.997	1.001	1.004	1.009	1.006	1.010	1.013
Huber$_{\psi 1.5}$	1.000	1.000	1.003	1.006	1.009	1.008	1.012	1.013	1.013	1.014
Trim$_{2x25\%}$	.731	.828	.891	.895	.908	.915	.920	.930	.937	.938
Trim$_{2x20\%}$	.731	.854	.891	.916	.918	.931	.938	.942	.942	.951
Trim$_{2x10\%}$	.793	.895	.933	.947	.959	.967	.970	.975	.975	.982
Trim$_{2x5\%}$	*	.922	.960	.968	.986	.988	.994	.996	1.002	1.000
Discrete Mass At Zero (Achievement)										
Huber$_{\psi 1.339}$	.979	.982	.984	.985	.986	.987	.986	.986	.986	.987
Huber$_{\psi 1.345}$	.978	.981	.982	.984	.987	.990	.991	.992	.990	.989
Huber$_{\psi 1.4088}$	.968	.982	.981	.984	.994	.999	.996	1.004	1.005	1.000
Huber$_{\psi 1.5}$	.994	.994	.997	.998	1.003	1.004	1.008	1.007	1.008	1.009
Trim$_{2x25\%}$	.716	.830	.888	.886	.915	.920	.920	.922	.932	.932
Trim$_{2x20\%}$	.716	.859	.895	.908	.928	.931	.934	.938	.948	.947
Trim$_{2x10\%}$	.794	.899	.929	.949	.960	.967	.971	.968	.975	.980
Trim$_{2x5\%}$	*	.910	.948	.967	.978	.981	.986	.984	.993	.997
Extreme Asymmetry (Achievement)										
Huber$_{\psi 1.339}$	.977	.977	.975	.978	.978	.978	.978	.978	.977	.978
Huber$_{\psi 1.345}$	.975	.975	.972	.975	.976	.976	.976	.976	.976	.976
Huber$_{\psi 1.4088}$	.959	.958	.953	.961	.962	.965	.964	.966	.963	.966
Huber$_{\psi 1.5}$	.955	.949	.946	.953	.950	.952	.950	.954	.951	.952
Trim$_{2x25\%}$	.647	.694	.734	.748	.757	.754	.748	.754	.762	.753
Trim$_{2x20\%}$	.647	.713	.741	.755	.761	.758	.764	.767	.773	.770
Trim$_{2x10\%}$	.685	.755	.744	.750	.750	.758	.746	.748	.749	.747
Trim$_{2x5\%}$	*	.740	.733	.748	.747	.755	.746	.754	.749	.748

Statistic	n: 10	20	30	40	50	60	70	80	90	100
					Extreme Asymmetry (Psychometric)					
Huber$_{\psi 1.339}$	.967	.966	.965	.966	.965	.966	.966	.966	.966	.966
Huber$_{\psi 1.345}$	.964	.962	.962	.962	.962	.963	.962	.962	.962	.962
Huber$_{\psi 1.4088}$	.931	.928	.929	.927	.927	.928	.928	.928	.927	.928
Huber$_{\psi 1.5}$	.904	.901	.897	.892	.892	.891	.889	.887	.887	.887
Trim$_{2x25\%}$	.510	.503	.449	.457	.443	.444	.424	.443	.418	.432
Trim$_{2x20\%}$	.510	.473	.417	.421	.404	.408	.394	.401	.386	.389
Trim$_{2x10\%}$	.431	.408	.375	.375	.365	.354	.355	.354	.359	.356
Trim$_{2x5\%}$	*	.382	.361	.361	.352	.341	.342	.343	.346	.341
					Digit Preference (Achievement)					
Huber$_{\psi 1.339}$	.980	.984	.986	.996	1.000	.999	.997	.996	.998	1.002
Huber$_{\psi 1.345}$	.978	.983	.988	1.003	1.000	.998	.997	1.002	1.001	1.002
Huber$_{\psi 1.4088}$	.969	1.000	.994	.999	1.003	1.003	1.004	1.004	1.003	1.006
Huber$_{\psi 1.5}$	1.000	1.000	1.000	1.007	1.009	1.010	1.008	1.011	1.009	1.010
Trim$_{2x25\%}$	.727	.831	.882	.893	.909	.923	.923	.934	.941	.935
Trim$_{2x20\%}$	.727	.841	.892	.904	.917	.929	.932	.933	.945	.936
Trim$_{2x10\%}$	.786	.895	.932	.950	.962	.968	.973	.974	.981	.979
Trim$_{2x5\%}$	*	.913	.952	.966	.983	.984	.990	.991	.999	.994
					Extreme Bimodality (Psychometric)					
Huber$_{\psi 1.339}$	.976	.979	.980	.981	.981	.982	.982	.983	1.000	1.000
Huber$_{\psi 1.345}$	.975	.977	.978	.980	.979	.980	.981	.982	1.000	1.000
Huber$_{\psi 1.4088}$	.955	.961	.970	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Huber$_{\psi 1.5}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Trim$_{2x25\%}$	.545	.523	.594	.588	.603	.598	.620	.607	.619	.617
Trim$_{2x20\%}$	.545	.609	.651	.675	.678	.687	.702	.704	.707	.720
Trim$_{2x10\%}$	.629	.763	.826	.849	.884	.897	.910	.937	.946	.925
Trim$_{2x5\%}$	*	.851	.964	.959	1.021	1.013	1.043	1.056	1.079	1.043
					Multimodality & Lumpy (Achievement)					
Huber$_{\psi 1.339}$	.987	1.000	1.000	1.004	1.006	1.009	1.013	1.012	1.014	1.015
Huber$_{\psi 1.345}$	.986	1.000	1.000	1.006	1.008	1.013	1.014	1.013	1.016	1.017
Huber$_{\psi 1.4088}$	1.000	1.000	1.013	1.019	1.019	1.026	1.029	1.029	1.032	1.035
Huber$_{\psi 1.5}$	1.000	1.002	1.027	1.036	1.035	1.042	1.049	1.048	1.053	1.057
Trim$_{2x25\%}$	.657	.662	.714	.708	.734	.726	.741	.731	.745	.740
Trim$_{2x20\%}$	.657	.732	.757	.775	.788	.798	.805	.805	.811	.815
Trim$_{2x10\%}$	.726	.835	.875	.903	.903	.924	.929	.941	.941	.944
Trim$_{2x5\%}$	*	.897	.951	.963	.963	.983	.995	.995	1.003	1.002

Note: *Insufficient sample size to trim.

location in real education and psychology data sets with a 1.28, 1.339, 1.345, 1.4088, or 1.5 weighting constant. Both of these questions will be answered on the basis of LREs. That is, evidence in support of a procedure will be in the form of a more efficient or narrower confidence interval.

Methodology

A Monte Carlo program was written for Minitab (1996) Release 11.1 using the data sets from Micceri (1986) and Minitab "macros" D.2.1 (one-sample trimmed mean, p. 318-319), and D.2.3-D.2.5 (one step Huber M-estimate, p. 321-323) from Staudte and Sheather (1990). Each data set was randomly sampled to produce sample sizes of n = 10 (10) 100. (See Stigler, 1977, for a contrary view on "subsampling from large data sets", p. 1057.) These sample sizes were noted to be of interest by Goodall (1983, p. 395).

For the 2β-trimmed mean (i.e, two-sided trim of β percent), degrees of freedom (df) = n - [2βn] - 1. The df are due to Tukey and McLaughlin (1963). The standard error for the one-step Huber M-estimator is the square root of the estimated asymptotic variance (Staudte & Sheather, 1990, p. 132, formula 4.6.2).

As noted by Rosenberger and Gasko (1983), "Sometimes, in order to obtain a specified amount of trimming exactly, we need to trim a fraction of an observation; for example, a 5%-trimmed mean from a sample of size 10 requires trimming half of each of the largest and smallest observations" (p. 309). They accomplished this feat by "giving fractional weight" (p. 310) to the remaining fraction. They noted that "Some authors trim only an integer number of observations from each extreme" (p. 310). The debate on this issue is amazingly involved; the reader is referred to their discussion on the matter (Rosenberger & Gasko, 1983, p. 310-311). In the current study, trimming is rounded down to the whole number, which was the approach taken by Staudte and Sheather (1990, p. 134) and Wilcox (1998, in press). Therefore, the 2x5% trim cannot be conducted for sample size n=10. Another anomaly is that the 2x25% and 2x20% trims yield identical samples (and therefore results) when n = 10.

Results

The study proceeded as follows. The LRE was calculated for each statistic. This process was repeated 1,000 times, sampling with replacement from the data set. Then, the median LRE for each statistic was computed from the 1,000 samples. The results are compiled in Table 2.

The first question pertained to the amount of trimming that would yield the most efficient estimator. With the exception of the extreme asymmetric psychometric data set, lightly trimmed means produced narrower confidence intervals. In the best case (i.e., extreme bimodality psychometric with n = 20), the 2x5% trim produced confidence intervals about 63% narrower than the 2x25% trimmed mean. In general, as the sample size increased, the confidence intervals produced by the various levels of trimming converged, although in half of the data sets the results were less than satisfactory even for n = 100 in the sense that the LREs were substantially less than 1.0.

Indeed, the first question appears to be rather moot. On the basis of LREs, trimmed means systematically performed worse than the various Hubers. The trimmed means were only competitive for n ≥90 with the smooth symmetric achievement and the digit preference psychometric data sets. The latter data set is essentially a smooth symmetric data set with certain scores enjoying a propensity to protrude. In the worst case (i.e., extreme asymmetry psychometric with n = 60), trimmed means produced confidence intervals as much as 293% wider than the $\text{Huber}_{\psi 1.28}$. Jackson (1986) noted that "A disadvantage of both trimming and Winsorizing is that they down-weight the highest and lowest order statistics whether or not all observations are sound. Thus, a proportion of the data values are always either omitted altogether or have their values changed towards the centre of the distribution" (p. 27). Perhaps this is the reason for the poor performance.

Again, some researchers express disdain in comparing statistics which estimate different quantities, and therefore, would not compare trimmed means directly with Huber statistics. Support for those researchers who find the comparisons useful is available from Pratt (1977), who argued that "It doesn't matter what an estimator estimates, as long as it is a location parameter" (p. 1092).

The second question pertained to the choice of ψ in the one-step Huber M-estimator. The results in Table 2 suggest that ψ = 1.28 is the best choice regardless of the nature of the data set for n < 50. For situations where 50 ≤ n ≤ 100, ψ = 1.28 remains an excellent choice, although ψ = 1.5 produced narrower confidence intervals more frequently.

Conclusion

The selection of robust methods requires more consumer input than clicking on a pull-down menu in a statistical package. This is because many robust procedures require making choices, such as the amount to trim or the value of a tuning parameter. Although there are many opinions to be found in the literature on which values to use, there has not been a systematic study of the impact of these choices. For example, this article considered the bracketed interval around the location for a single sample.

The results in this article pertain to the 95% bracketed interval. It was chosen because it was the level used by Dixon and Tukey (1968) (who provided the impetus for the creation of the LRE). Another reason is, "The 95% confidence level appears to be used more frequently in practice than any other level" (Hahn & Meeker, 1991, p. 38). Obviously, the results in this study should not be generalized to other levels (e.g., 90% or 99%).

There were some assumptions made in this paper. First, it was assumed that bracketed intervals should have a fixed length. There have been attempts to improve on fixed-length bracketed intervals with those that "adjust so that their expected length depends ... on the data" (Low, 1997, p. 2548). Second, it was assumed that the bracketed confidenc intervals should be symmetric with respect to $\hat{\theta}$.

To restate the interpretation of the LRE, values less than one indicate the length of the bracketed interval is wider for a competitor than for the $\text{Huber}_{\psi 1.28}$. It is desirable that the choice of ψ for constructing the numerator of the LRE predominately result in a ratio less than one, and only occasionally should a competitor stand out in terms of its comparative performance. This study showed that the value of $\psi = 1.28$ met this requirement. Specifically, for samples of size $n \leq 30$, the LREs were greater than 1.0 for only 4 out of 184 (2.2%) outcomes, and for only 10 out of 240 (4.2%) outcomes for $n < 50$.

The results were also generally less than 1.0 for all estimators for $50 \leq n \leq 100$, save the $\text{Huber}_{\psi 1.5}$. An advantage in maintaining $\psi = 1.28$ is that its bracketed intervals were never more than 6% wider than those for $\psi = 1.5$ for $50 \leq n \leq 100$, whereas results for $\psi = 1.5$ were as much as 13% wider than $\psi = 1.28$ for certain distributions and sample sizes. (Although it would complicate a simple statistic, and therefore is not recommended, the Monte Carlo results indicate that it would be beneficial to compute the LREs with $\psi=1.28$ for $10 \leq n < 50$, and $\psi = 1.5$ for $50 \leq n \leq 100$.)

Huber (1981) defined robustness as "*insensitivity to small deviations from the assumptions*" (p. 1). Indeed, many previously conducted studies concentrated on robustness against contamination in the form of small deviations (e.g., one-out or one-wild). Barnett and Lewis (1994) noted that "Many such published procedures are robust against the possibility that the entire sample comes from some other distribution, possibly gamma or Cauchy, not too dissimilar to the normal but perhaps somewhat skew or fatter-tailed" (p. 56). The purpose of this paper was to examine the LREs of some robust measures where the sample comes from applied social and behavioral science data sets where the shape is quite dissimilar to the normal curve. The results indicate narrower 95% bracketed intervals for the one step Huber M-estimator when $\psi = 1.28$ (as opposed to $\psi = 1.339$, 1.345, 1.4088, and 1.5) for samples less than fifty. The results also indicate that, although lightly trimmed means of 2x5% yield narrower 95% bracketed intervals than heavily trimmed means of 2x25%, trimmed means almost always result in significantly wider bracketed intervals than M-estimators for the real education and psychology data sets and sample sizes studied.

References

Andrews, D. F. (1977). Discussion. *The Annals of Statistics*, *5*, 1078-1079.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., & Tukey, J. W. (1972). *Robust estimates of location: survey and advances*. Princeton, NJ: Princeton University Press.

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data*. (3rd ed.) Chichester, UK: Wiley.

Dixon, W. J., & Tukey, J. W. (1968). Approximate behavior of the distribution of Winsorized *t* (Trimming/Winsorization 2). *Technometrics*, *10*, 83-98.

Gastwirth, J. L, & Cohen, M. L. (1970). Small sample behavior of some robust linear estimators of location. *Journal of the American Statistical Association*, *65*, 946-973.

Goodall, C. (1983). M-estimators of location: An outline of the theory. In D. C. Hoaglin, F. Mosteller, and J. W. Tukey (Eds.), *Understanding robust and exploratory data analysis*. Chapter11. NY: Wiley.

Hahn, G. J., & Meeker, W. Q. (1991). *Statistical intervals: A guide for practitioners*. NY: Wiley.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. NY: Wiley.

Hill, M., & Dixon, W. J. (1992). Robustness in real life: A study of clinical laboratory data. *Biometrics*, *38*, 377-396.

Hoaglin, D. C. (1977). Discussion. *The Annals of Statistics*, *5*, 1087-1088.

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1983). *Understanding robust and exploratory data analysis*. NY: Wiley.

Hogg (1977). Discussion. *The Annals of Statistics*, *5*, 1088-1090.

Huber, P. J. (1981). *Robust statistics*. NY: Wiley.

Huber, P. J. (1977). Discussion. *The Annals of Statistics*, *5*, 1090-1091.

Huber, P. J. (1972). The 1992 Wald Lecture. Robust statistics: A review. *The Annals of Mathematical Statistics*, *43*, 1041-1067.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, *35*, 73-101.

Jackson, P. J. (1986). Robust methods in statistics. In (A. D. Lovie, Ed.), *New developments in statistics for psychology and the social sciences*, Chapter 2. London: The British Psychological Society and Methuen.

Lee, H. (1995). Outliers in business surveys. In B. G. Cox, D. A. Binder, B. N. Chinnappa, M. J. Colledge, & P. S. Kott (Eds.), *Business survey methods*. Chapter 26. NY: Wiley

Low, M. G. (1997). On nonparametric confidence intervals. *The Annals of Statistics, 25*, 2547-2554.

Micceri, T. (1986, November). A futile search for that statistical chimera of normality. Paper presented at the 31st Annual Convention of the Florida Educational Research Association, Tampa, Fl.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156-166.

Minitab. (1996). *Minitab*, Release 11.1. State College, PA: Minitab, Inc.

Pratt, J. W. (1977). Discussion. *The Annals of Statistics, 5*, 1092-1094.

Rey, W. J. J. (1980). *Introduction to robust and quasi-robust statistical methods*. Berlin: Springer-Verlag.

Rosenberger, J. L., & Gasko, M. (1983). Comparing location estimators: trimmed means, medians, and trimean. In D. C. Hoaglin, F. Mosteller, and J. W. Tukey (Eds.) *Understanding robust and exploratory data analysis*, Chapter 10. NY: Wiley.

Sawilowsky, S. S. (1998). Comment: Using robust statistics in social and behavioral science. *British Journal of Mathematical and Statistical Psychology, 51*, 49-52.

Sawilowsky, S. S., & Blair, R. C. (1992). A more realistic look at the robustness and Type II error properties of the t test to departures from population normality. *Psychological Bulletin, 111*, 352-360.

Sawilowsky, S. S., & Hillman, S. B. (1992). Power of the independent samples t test under a prevalent psychometric measure distribution. *Journal of Consulting and Clinical Psychology, 60*, 240-243.

SPSS (1996). *SPSS* Base 7.0 Applications Guide. Chicago: SPSS.

Staudte, Jr., R. G. (1980). Robust estimation: Queen's papers in pure and applied mathematics No. 53. Kingston, Ontario, Canada: Queen's University.

Staudte, Jr., R. G., & Sheather, S. J. (1990). *Robust estimation and testing*. NY: Wiley.

Stigler, S. M. (1977). Do robust estimators work with real data? *The Annals of Statistics, 5*, 1055-1098.

Tukey, J. W., & McLaughlin, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: trimming/Winsorization 1, *Sankhya, 25*, 331-352.

Wilcox, R. R. (1998). What are the goals and strategies of modern robust methods, why do they greatly increase chances of detecting differences between groups and associations among variables, and how do they enhance our understanding of data? *British Journal of Mathematical and Statistical Psychology, 51*, 1-40.

Wilcox, R. R. (1997). Some practical reasons for reconsidering the Kolmogorov-Smirnov test. *British Journal of Mathematical and Statistical Psychology, 50*, 9-20.

Wilcox, R. R. (1996). *Statistics for the social sciences*. San Diego, CA: Academic Press.