Statistical Methodology 6 (2009) 604–621

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

A comparison of two boxplot methods for detecting univariate outliers which adjust for sample size and asymmetry

Nancy J. Carter *, Neil C. Schwertman, Terry L. Kiser

Department of Mathematics & Statistics, California State University, Chico, United States

ARTICLE INFO

Article history:
Received 20 February 2009
Received in revised form
1 June 2009
Accepted 10 July 2009

Keywords:
Box-plot
Masking
Monte Carlo
Simulation
Swamping

ABSTRACT

It is important to identify outliers since inclusion, especially when using parametric methods, can cause distortion in the analysis and lead to erroneous conclusions. One of the easiest and most useful methods is based on the boxplot. This method is particularly appealing since it does not use any outliers in computing spread. Two methods, one by Carling and another by Schwertman and de Silva, adjust the boxplot method for sample size and skewness. In this paper, the two procedures are compared both theoretically and by Monte Carlo simulations. Simulations using both a symmetric distribution and an asymmetric distribution were performed on data sets with none, one, and several outliers. Based on the simulations, the Carling approach is superior in avoiding masking outliers, that is, the Carling method is less likely to overlook an outlier while the Schwertman and de Silva procedure is much better at reducing swamping, that is, misclassifying an observation as an outlier. Carling's method is to the Schwertman and de Silva procedure as comparisonwise versus experimentwise error rate is for multiple comparisons. The two methods, rather than being competitors, appear to complement each other. Used in tandem they provide the data analyst a more complete prospective for identifying possible outliers.

* Corresponding address: Department of Mathematics & Statistics, California State University, Chico, 400 West First Street, Chico, CA 95929-0525, United States. Tel.: +1 530 898 6562 (o); fax: +1 530 898 3097.

E-mail address: ncarter@csuchico.edu (N.J. Carter).

1572-3127/$ – see front matter
doi:10.1016/j.stamet.2009.07.001

1. Introduction

Observations that are inconsistent with the majority of the data often are classified as outliers. Such extreme observations can have a profound influence on the usual parametric data analyses and, as a consequence, can lead to erroneous conclusions. Therefore, it is not surprising there is considerable literature on the identification or adjustment for these unusual observations. See for example [1–22]. In this paper, simulations are used to compare two graphical methods for labeling outliers with respect to swamping, that is, mistaking clean observations as outliers and masking, failing to identify outliers.

In 1977, Tukey proposed a method for identifying outliers based on construction of a box-plot and the use of "inner" and "outer" fences. Tukey's method has become very popular and is often included in statistical texts, even basic introductory texts such as Milton [17]. The Tukey procedure uses the estimated interquartile range, often called the H-spread, which is the difference between the "hinges". That is, the H-spread is $q_3 - q_1$, where q_1 and q_3 are the first and third sample quartiles, respectively. The inner fences f_1 and f_3 and the outer fences F_1 and F_3 are then defined as $f_1 = q_1 - 1.5\,(q_3 - q_1)$, $f_3 = q_3 + 1.5\,(q_3 - q_1)$, $F_1 = q_1 - 3\,(q_3 - q_1)$ and $F_3 = q_3 + 3\,(q_3 - q_1)$. Tukey defined observations beyond the inner fences (f_1 and f_3) as mild outliers and those beyond the outer fences (F_1 and F_3) as extreme outliers.

While there are several very effective methods for labeling outliers that require more computation, the "fences" method is especially appealing because it is easy to use. Even more important than the simplicity, the "fences" do not incorporate the extreme potential outliers which can inflate the computation of a measure of spread and decrease the sensitivity for labeling of outliers. The usual "fences" method does not, however, take into account the size of the data set or the asymmetry of many possible distributions. One exception is the method proposed by Carling [10] which uses the generalized lambda distribution to make modest adjustments for both sample size and asymmetry. The modification of Tukey's fences suggested by Carling [10] and Schwertman and de Silva [21] require some computations but significantly less than most alternative procedures.

In practice a few "extreme" or unusual observations are to be expected to occur in "large" data sets. Extreme in this context could be those observations three or more standard deviations from the mean. For example, in a normally distributed data set of size 1000, you would expect to observe 2.7 observations more than three standard deviations from the mean, whereas in a sample of size 50, you would only expect on the average 0.135 observations three or more standard deviations from the mean.

Clearly, the size of the data set has a major influence on the occurrence of unusual observations. It seems logical that identification of outliers should incorporate sample size in determining the difference between extreme values that occur naturally and those which are perhaps contaminated by a recording error or measuring instrument malfunction.

Hoaglin and Iglewicz [15] proposed an adjustment for sample size in the standard Tukey box-plot method. Basically, their method provided coefficients for the interquartile range which are adjusted for sample size. Their method, commonly called "the box-plot rule", and accompanying tables are restricted to outside rates of 0.10 and 0.05 and are applicable only to distributions which are at least approximately normal. The outside rate as used in their paper is the probability an uncontaminated observation is outside the fences. Simulations in this paper and in [21] indicate that to avoid indicating many uncontaminated observations as outliers, called swamping, it is necessary to establish the outside rate at a very low level such as 0.01. While such a small outside rate may fail to find some mild outliers, the extreme outliers are still likely to be detected. The grossly contaminated observations can cause significant distortion of the parametric analysis while a few mild outliers are likely to have only a minimal influence.

The sequential fences method in [21], the procedure we are comparing to Carling's technique, is based on the Poisson distribution since in most data sets outliers are only a very small proportion of the observations. The Schwertman and de Silva simulations showed that the "box-plot rule" for sample size 20, was only slightly superior to their sequential fences procedure but the larger sample size 100 was much less satisfactory and less effective especially with multiple outliers. Overall, the sequential fences seemed to be superior to the "box-plot rule" and therefore we decided to focus on comparing Carling's method with the Schwertman/de Silva method of adjusting for sample size.

The purposes of this study are first to determine if Carling's method can be improved for asymmetric distributions by modifying it, that is by replacing the interquartile range (IQR) with the appropriate semi-interquartile range (SIQR), and, second to compare the modified Carling method to the procedures proposed by Schwertman and de Silva [21].

In Section 2 the proposed modification to Carling's method is described and then compared to the standard Carling procedure graphically. In Section 3 an examination of the effects of sample size over a very large spectrum of sample sizes from 10 to 400 is provided by a theoretical comparison of the Carling and Schwertman/de Silva methods. In addition to the theoretical comparison, in Section 4, data set simulations of two sample sizes and two different distributions were used to compare the methods. The simulation study provides a comparison of the two methods as they would actually be applied in data analysis. Section 5 contains two examples which compare the two procedures when the number of outliers is known and when they are not known. Finally, Section 6 states our conclusions and recommendations from this investigation.

2. Carling's fences for asymmetric distributions

The usual fences procedure makes no adjustment for data from asymmetrical distributions. To construct fences for such data, Kimber [16] and Schwertman, Owens, and Adnan [20] advocate using the semi-interquartile ranges to determine the appropriate fences. Specifically, the lower semi-interquartile range is $\text{SIQRL} = q_2 - q_1$ where q_2 is the sample second quartile and is used to compute a lower fence. Similarly, the upper semi-interquartile range is $\text{SIQRU} = q_3 - q_2$ and is used to compute the upper fence. To ascertain the effect of using the SIQR's compared to the IQR, computer studies were done using Mathematica which are discussed in Sections 3 and 4.

The usual fences suggested by Carling [10] are

$$q_2 \pm k_2\,(q_3 - q_1) = q_2 \pm k_2(\text{IQR}) \tag{2.1}$$

where k_2 is obtained from

$$\begin{aligned} 100\, r_n = {} & -8.07 + \frac{3.71}{n} + \frac{17.63}{k_2} - \frac{23.64}{n\,k_2} + 0.83\alpha_3 + 0.48\alpha_3^2 + 0.48\,(\alpha_4 - 3) \\ & - 0.04\,(\alpha_4 - 3)^2 \end{aligned} \tag{2.2}$$

where r_n is the outside rate (probability of misclassifying an observation as an outlier), n is the sample size, and α_3 and α_4 are the third and fourth moments respectively. Eq. (2.2) is applicable to both symmetric and asymmetric distributions with $\alpha_3 = 0$ for the symmetric case. It should be noted that Carling [10] actually considered only the upper tail. Eq. (2.1) is derived by using the analogous approach for the lower tail. To adjust for using the SIQR, Eq. (2.1) becomes

$$q_2 - 2k_2(\text{SIQRL}) = q_2 - 2k_2\,(q_2 - q_1) \quad \text{and} \quad q_2 + 2k_2\,(\text{SIQRU}) = q_2 + 2k_2(q_3 - q_2). \tag{2.3}$$

Various asymmetrical distributions from the gamma family were examined in a computer study by changing the gamma parameter α. It can be shown that the outside rate is unchanged for different values of the gamma scale parameter θ and hence only $\theta = 1$ was used. The third moment or skewness for the gamma distribution is $\alpha_3 = 2/\sqrt{\alpha}$ and the fourth moment, kurtosis, is $\alpha_4 = 3 + 6/\alpha$ where α is the first parameter of the gamma distribution. When α is small the distribution is very skewed but the skewness decreases as α increases. Mathematica was used to measure the theoretical outside rates for data as the skewness changed. Figs. 1–3 display the expected outside rates as a function of α using both the Carling IQR (2.1) and Carling SIQR (2.3) methods for sample size 20 and nominal outside rates (probability an observation falls outside the fence) 0.10, 0.05 and 0.01. Figures for sample sizes 50 and 100 are not included since they were nearly identical to those for sample size 20.

These figures are used to graphically compare Carling's method using IQR and the modified Carling method using the SIQR. Each figure consists of two graphs. The top graph compares the two methods over α ranging from 0 to 10. Since the most dramatic changes occur for small α values the second graph is an enlargement of the top graph but only over the α values ranging from 0 to 1. For the asymmetric distributions used for Figs. 1–3, Carling's SIQR method provides a better balance between the expected

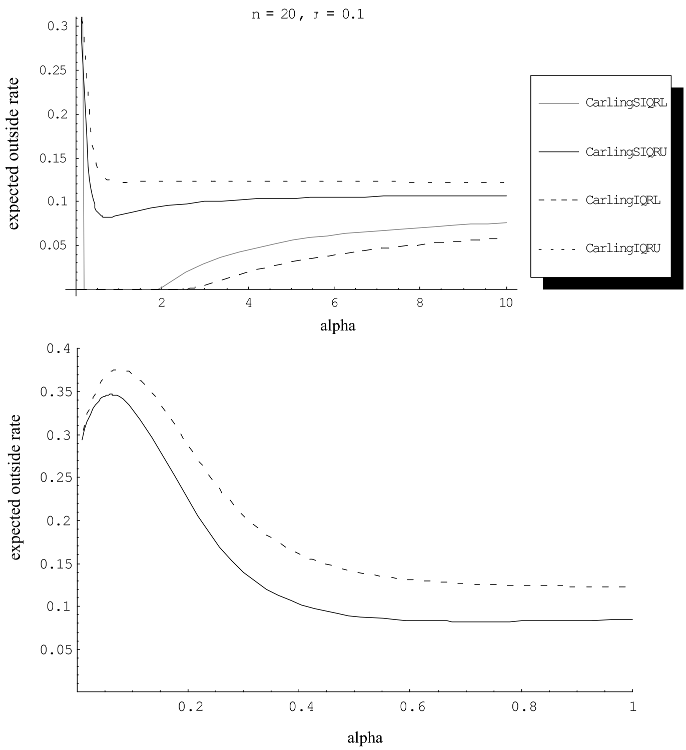


Fig. 1. Expected outside rate for Carling's method using IQR and SIQR for sample size 20 and nominal outside rate 0.1.

outside rates in the upper and lower tails than the IQR method. That is, using the SIQR approach, the expected lower tail outside rate increases while the upper tail outside rate decreases relative to the IQR outside rates. As would be expected this effect becomes less pronounced as the skewness decreases. Furthermore, the patterns are independent of sample size for a specified nominal outside rate.

3. A theoretical comparison of Carling and Schwertman/de Silva methods

The Carling [10] method for identifying outliers is given in Eq. (2.1). It uses a constant k_2 which is determined by Eq. (2.2) for some specified outside error rate, i.e. the nominal proportion of observations incorrectly identified as outliers. Eq. (2.2) incorporates sample size, n, skewness, α_3, and kurtosis, α_4, in the computation of k_2 and hence makes some adjustment for sample size, skewness and kurtosis while specifying the outside rate.

For data from distributions which are approximately normal, Schwertman, Owens and Adnan [20] suggested another method for constructing fences for a specified outside rate. Their procedure is similar to Carling's in that they too construct the fences from q_2. The computation of the coefficient to the IQR, however, is quite different.

Specifically, the Schwertman et al. [20] fences are

$$q_2 \pm \frac{q_3 - q_1}{k_n} Z_{\alpha/2} = q_2 \pm \frac{Z_{\alpha/2}}{k_n} \text{ (IQR)} \tag{3.1}$$

where $P\left(Z > Z_{\alpha/2}\right) = \alpha/2$, q_1, q_2 and q_3 are the first, second and third sample quartiles respectively and k_n is a constant, dependent on n, which relates the interquartile range to the standard deviation of the normal. The table for the k_n values is reproduced in Appendix A.

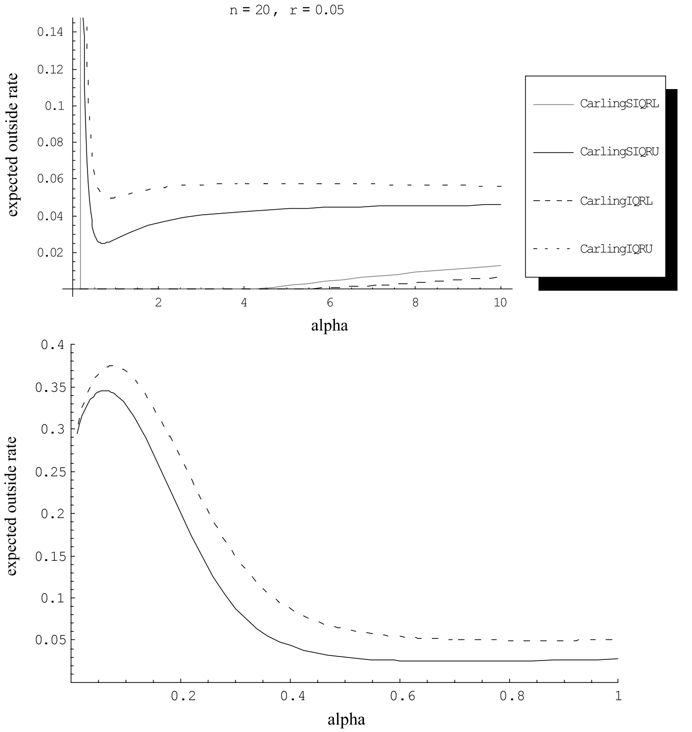


Fig. 2. Expected outside rate for Carling's method using IQR and SIQR for sample size 20 and nominal outside rate 0.05.

Schwertman/de Silva [21] suggested a modification of the Schwertman et al. [20] method to incorporate the sample size based on the Poisson distribution. Specifically, since an outlier should be an unusual event, the Poisson model seems appropriate with the probability of outliers equal to r, the specified outside rate. Then

$$e^{-n\alpha_n} = 1 - r \quad \text{and} \quad \alpha_n = \frac{-\ln(1-r)}{n}.$$

This α_n replaces $\alpha/2$ in Eq. (3.1). This method of determining α_n accomplishes the same effect as the Bonferroni method of splitting the overall significance level but the Poisson procedure, while slightly more difficult computationally, is more accurate in maintaining the overall significance level.

The Schwertman/de Silva fences, in practice, use the t distribution but for the theoretical study, since the estimate of variability is not calculated from the data, the Z distribution is more appropriate. The Schwertman/de Silva fences then become

$$q_2 \pm \frac{q_3 - q_1}{k_n} Z_{\alpha_n} = q_2 \pm \frac{Z_{\alpha_n}}{k_n} \, (\text{IQR}) . \tag{3.2}$$

Both the Carling and Schwertman/de Silva methods adjust for sample size but the Schwertman/de Silva is a much larger adjustment. The choice of adjustments appears to be analogous to choosing pairwise versus experimentwise error rates in multiple comparison procedures. The outside rate using Carling is much closer to the outside rate for each individual observation, but has an extremely high probability of misclassifying at least one uncontaminated observation as an outlier in larger samples. The Schwertman/de Silva adjustment for sample size is greater and results in outside rates closer to the specified value for not just each individual observation but the entire sample in aggregate.

To compare the procedures for both a symmetric and an asymmetric distribution, the standard normal distribution was used to represent the symmetric case and a chi-square distribution with

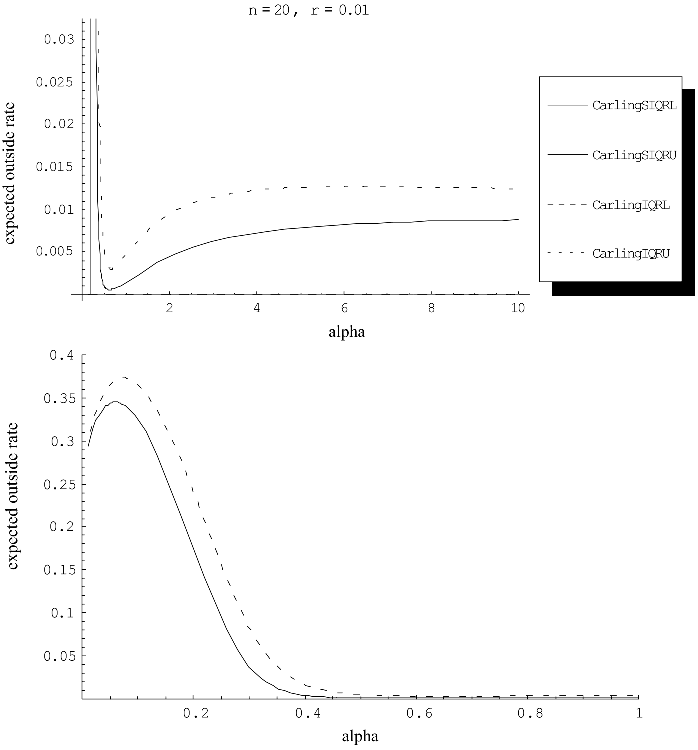


Fig. 3. Expected outside rate for Carling's method using IQR and SIQR for sample size 20 and nominal outside rate 0.01.

Table 1
Theoretical experimentwise error rates for various sample sizes with $r = 0.01$ when sampling from a $N(0, 1)$.

	n						
	10	20	30	50	100	200	400
Carling IQR	0.0861	0.1180	0.1538	0.2234	0.3753	0.5966	0.8319
Schwertman/de Silva IQR	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100

eight degrees of freedom was used for the asymmetric case. The Carling and Schwertman/de Silva methods using both the IQR approach and the adjusted procedure using the SIQR were compared for both types of distributions. That is, for the Carling approach, Eqs. (2.1) and (2.3) were used while the Schwertman/de Silva approach was investigated using Eqs. (3.2) and

$$q_2 - \frac{2Z_{\alpha_n}}{k_n}\ (\text{SIQRL}) \quad \text{and} \quad q_2 + \frac{2Z_{\alpha_n}}{k_n}\ (\text{SIQRU})\,. \tag{3.3}$$

These comparisons were made for nominal outside rates of 0.10, 0.05, and 0.01 for sample sizes 10, 20, 30, 50, 100, 200 and 400. The computation of the probabilities of incorrectly declaring at least one noncontaminated observation an outlier in the data set is given in Tables 1–6. The computations in the tables are based on the exact length of the IQR and SIQR, skewness (α_3), and kurtosis (α_4), characteristics of the standard normal and eight degree of freedom Chi-square distributions as well as the constants k_n and α_n from the tables in [21].

In Tables 1–3 only the IQR methods are included since the distribution is the symmetric standard normal and therefore the SIQR and IQR procedures are identical. As Tables 1–3 indicate for the Schwertman/de Silva method, the probability of incorrectly declaring at least one uncontaminated observation as an outlier is exactly the nominal outside rate since this is the basis of the theoretical

Table 2
Theoretical experimentwise error rates for various sample sizes with $r = 0.05$ when sampling from a $N(0, 1)$.

	n						
	10	20	30	50	100	200	400
Carling IQR	0.4164	0.5807	0.7028	0.8158	0.9742	0.9992	1
Schwertman/de Silva IQR	0.0501	0.0501	0.0500	0.0500	0.0500	0.0500	0.0500

Table 3
Theoretical experimentwise error rates for various sample sizes with $r = 0.10$ when sampling from a $N(0, 1)$.

	n						
	10	20	30	50	100	200	400
Carling IQR	0.7287	0.8969	0.9613	0.9946	1	1	1
Schwertman/de Silva IQR	0.1005	0.1003	0.1002	0.1000	0.1000	0.1000	0.1000

Table 4
Theoretical experimentwise error rates for various sample sizes with $r = 0.01$ when sampling from a $X^2(8)$.

	n						
	10	20	30	50	100	200	400
Carling IQR	0.1404	0.2182	0.2921	0.4210	0.6504	0.8727	0.9832
Schwertman/de Silva IQR	0.1188	0.1863	0.2331	0.3093	0.4404	0.5982	0.7630
Carling SIQR	0.0879	0.1354	0.1832	0.2721	0.4554	0.6954	0.9040
Schwertman/de Silva SIQR	0.0723	0.1121	0.1392	0.1844	0.2659	0.3746	0.5119

Table 5
Theoretical experimentwise error rates for various sample sizes with $r = 0.05$ when sampling from a $X^2(8)$.

	n						
	10	20	30	50	100	200	400
Carling IQR	0.4966	0.6946	0.8166	0.9342	0.9950	1	1
Schwertman/de Silva IQR	0.2321	0.3359	0.4046	0.5057	0.6592	0.8087	0.9228
Carling SIQR	0.4091	0.5826	0.7102	0.8613	0.9782	0.9995	1
Schwertman/de Silva SIQR	0.1593	0.2287	0.2748	0.3458	0.4645	0.6035	0.7489

Table 6
Theoretical experimentwise error rates for various sample sizes with $r = 0.10$ when sampling from a $X^2(8)$.

	n						
	10	20	30	50	100	200	400
Carling IQR	0.8429	0.9557	0.9880	0.9991	1	1	1
Schwertman/de Silva IQR	0.3058	0.4341	0.5091	0.6159	0.7633	0.8878	0.9650
Carling SIQR	0.8570	0.9599	0.9892	0.9992	1	1	1
Schwertman/de Silva SIQR	0.2295	0.3148	0.3691	0.4520	0.5813	0.7197	0.8471

analysis, while Carling's probability of making such an error is substantially higher. For Table 1 (nominal outside rate 0.01), while the theoretical Schwertman/de Silva's method maintains a probability of misclassifying at least one observation as an outlier at 0.01, the same probability for Carling's IQR method steadily increased from 0.0861 with $n = 10$ to 0.8319 when $n = 400$. While the Schwertman/de Silva procedure always maintains the nominal outside rates, for outside rates of 0.05 (Table 2) and 0.10 (Table 3), Carling's probabilities of misclassifying at least one observation as an outlier approaches one for $n \geq 200$ for outside rate 0.05 and for outside rate 0.10, it is 0.7287 for samples as small as 10 and approaches 1.0 for $n = 50$ or more.

In Tables 4–6, both the IQR and SIQR measures were used in both the Carling and Schwertman/de Silva procedures since the chi-square distribution is asymmetric and hence the SIQR method is likely to be more appropriate. For this asymmetrical distribution it appears that establishing the fences using the SIQR is preferable to using the IQR because the probability of misclassifying at least one observation as an outlier is less for SIQR using either the Carling or the Schwertman/de Silva approaches. The decrease (when averaged over the decreases for both methods combined) is typically about 0.11 for nominal outside rates of 0.01 and 0.05. The Carling method with a nominal outside rate 0.10 had virtually identical probabilities of misclassifying at least one observation as an outlier for both the SIQR and IQR approaches. As seen in these tables, the Schwertman/de Silva method always gives smaller probabilities of misclassifying at least one observation as an outlier than the corresponding Carling method. The typical decrease in probability of this error is 0.40, 0.36 and 0.10 for nominal outside rates of 0.10, 0.05 and 0.01 respectively. The probability of misclassifying at least one observation as an outlier increases with sample size n and nominal outside rate.

4. Simulations for two types of distributions

In Section 3, the Carling and Schwertman/de Silva procedures were compared theoretically, that is, the expected IQR and SIQR were used to compute the theoretical outside rates. In practice, the quartiles, IQR and SIQR, would be computed from the sample data. Therefore, as used in the Schwertman/de Silva paper [21], the Z_{α_n} in Eqs. (3.2) and (3.3) is replaced by the corresponding appropriate $t_{df,\alpha_{nm}}$. A Monte Carlo study was performed using Mathematica to further compare the methods from a more practical perspective. Ten thousand data sets each for sample sizes of 20, and 100 from two distributions, the standard normal and a chi-square with eight degrees of freedom were simulated. The two distributions provide symmetric and moderately asymmetric data sets. For both sample sizes and distributions data was generated containing no outliers, one outlier, or multiple outliers (three outliers when $n = 20$ and five outliers when $n = 100$). All outliers were in the upper tail only. The single outlier for the normal distribution simulations was 3.5 standard deviations above the mean while the corresponding outlier for the chi-square simulation was 30. Both outliers represented points with approximately the same tail probabilities. When simulations with three outliers for the normal with $n = 20$ were done, the outliers were 3.45, 3.5 and 3.55 deviations above the mean while the corresponding outliers for the chi-square simulations with $n = 20$ were 29.3, 29.8, and 30.2. These outliers represented points with approximately the same tail probabilities. When simulations with five outliers for the normal with $n = 100$ were done, the outliers were 3.40, 3.45, 3.50, 3,55, and 3.60 while the corresponding outliers for the chi-square simulations with $n = 100$ were 28.8, 29.3, 29.8, and 30.2 and 30.7. These outliers represented points with approximately the same tail probabilities as those used in the normal simulations. The proportions of the 10 000 simulations where outliers were identified are compared for the Carling and Schwertman/de Silva procedures and are tabulated in Tables 7–12.

For the normal distribution, Table 7 shows that the Schwertman and de Silva method is much better at not misidentifying observations as outliers when simulating no outliers for both $n = 20$ and $n = 100$. As expected, this pattern was consistent for both the upper and lower tails. Specifically, for sample size 20 and no outliers the Carling procedure misidentified outliers averaging over both tails in 0.8861, 0.5816 and 0.1948 of the simulations while the corresponding average rates for Schwertman and de Silva are 0.0557, 0.0322, and 0.0107 for nominal rates 0.10, 0.05, and 0.01 respectively. Similarly for sample size 100 and no outliers the Carling procedure misidentified outliers on average in 0.9999, 0.9542 and 0.4069 of simulations compared to Schwertman and de Silva rates of 0.0405, 0.0207 and 0.0043 for nominal rates 0.10, 0.05 and 0.01 respectively. For the larger sample size the Carling method misidentified more observations as outliers whereas for the Schwertman and de Silva method the converse was true, that is Schwertman and de Silva consistently had a lower proportion misidentified as outliers as sample size increased.

When there is one outlier, Carling's procedure was substantially better at identifying the observation as an outlier compared to the Schwertman and de Silva method. However, this increased likelihood came at the expense of misidentifying a second non-contaminated observation as an outlier at a significantly higher rate than the Schwertman and de Silva procedure. Specifically, for sample size 20 and one outlier the Carling method misidentified a lower tail outlier in 0.8568, 0.5368 and 0.1665 of the simulations compared to the Schwertman and de Silva rates of 0.0458, 0.0253, and 0.0060 for nominal outside rates of 0.10, 0.05, and 0.01 respectively. In the upper tail the Carling method

Table 7
Proportion classified as outliers for 10 000 simulations from a standard normal distribution.

	No outliers $n = 20$								
	Lower tail			Upper tail					
	$r = 0.10$	$r = 0.05$	$r = 0.01$	$r = 0.10$	$r = 0.05$	$r = 0.01$			
C	0.8847	0.5782	0.1943	0.8875	0.5849	0.1952			
SD	0.0538	0.0323	0.0108	0.0576	0.0320	0.0106			
	No outliers $n = 100$								
	$r = 0.10$	$r = 0.05$	$r = 0.01$	$r = 0.10$	$r = 0.05$	$r = 0.01$			
C	0.9999	0.9543	0.4077	0.9998	0.9540	0.4060			
SD	0.0405	0.0198	0.0041	0.0405	0.0215	0.0045			
	One outlier (upper tail) $n = 20$								
	Lower tail (no outliers) Incorrectly classified			Upper tail Correctly classified			Upper tail Incorrectly classified		
	$r = 0.10$	$r = 0.05$	$r = 0.01$	$r = 0.10$	$r = 0.05$	$r = 0.01$	$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.857	0.537	0.167	0.999	0.998	0.876	0.807	0.449	0.117
SD	0.046	0.025	0.006	0.529	0.387	0.162	0.165	0.117	0.059
	One outlier (upper tail) $n = 100$								
	Lower tail Incorrectly classified			Upper tail Correctly classified			Upper tail Incorrectly classified		
	$r = 0.10$	$r = 0.05$	$r = 0.01$	$r = 0.10$	$r = 0.05$	$r = 0.01$	$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.999	0.950	0.390	1.000	1.000	1.000	0.999	0.945	0.370
SD	0.036	0.019	0.003	0.810	0.619	0.216	0.211	0.148	0.067

C = Carling method.
SD = Schwertman and de Silva method.

Table 8
Proportion classified as outliers for 10 000 simulations from a Chi-square (degrees of freedom 8).

	No outliers $n = 20$								
	Lower tail			Upper tail					
	$r = 0.10$	$r = 0.05$	$r = 0.01$	$r = 0.10$	$r = 0.05$	$r = 0.01$			
C	0.5838	0.1902	0.0266	0.8863	0.6009	0.2334			
SD	0.0257	0.0167	0.0069	0.2228	0.1664	0.0922			
	No outliers $n = 100$								
	Lower tail			Upper tail					
	$r = 0.10$	$r = 0.05$	$r = 0.01$	$r = 0.10$	$r = 0.05$	$r = 0.01$			
C	0.8879	0.1286	0.0003	0.9999	0.9592	0.4837			
SD	0.0000	0.0000	0.0000	0.3762	0.2893	0.1558			
	One outlier (upper tail) $n = 20$								
	Lower tail Incorrectly classified			Upper tail Correctly classified			Upper tail Incorrectly classified		
	$r = 0.10$	$r = 0.05$	$r = 0.01$	$r = 0.10$	$r = 0.05$	$r = 0.01$	$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.563	0.174	0.025	1.000	0.996	0.879	0.770	0.435	0.136
SD	0.023	0.013	0.005	0.872	0.801	0.617	0.322	0.266	0.176
	One outlier (upper tail) $n = 100$								
	Lower tail Incorrectly classified			Upper tail Correctly classified			Upper tail Incorrectly classified		
	$r = 0.10$	$r = 0.05$	$r = 0.01$	$r = 0.10$	$r = 0.05$	$r = 0.01$	$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.884	0.117	0.000	1.000	1.000	0.997	1.000	0.945	0.438
SD	0.0000	0.0000	0.0000	0.989	0.970	0.857	0.654	0.580	0.432

C = Carling method.
SD = Schwertman and de Silva method.

correctly identified the outlier in 0.9999, 0.9976 and 0.8762 compared to the Schwertman and de Silva rates of 0.5291, 0.3870, and 0.1619 for nominal outside rates of 0.10, 0.05, and 0.01. However, the Carling procedure misidentified a second observation as an outlier in 0.8065, 0.4490 and 0.1171

Table 9
10 000 simulations with three outliers from a standard normal distribution with $n = 20$.

	Lower tail — Proportion Misclassified as outliers for $X_{(1)}$		
	$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.7512	0.3887	0.0905
SD	0.0195	0.0097	0.0028
	Lower tail — Proportion Misclassified as outliers for $X_{(2)}$		
	$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.4590	0.1277	0.0126
SD	0.0219	0.0125	0.0037

	Upper tail — Proportion Correctly classified as outliers for $X_{(20)}$		
	$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.9995	0.9671	0.6025
SD	0.2368	0.1426	0.0438
	Upper tail — Proportion Correctly classified as outliers for $X_{(19)}$		
	$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.9995	0.9617	0.5805
SD	0.6977	0.5799	0.3483
	Upper tail — Proportion Correctly classified as outliers for $X_{(18)}$		
	$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.9992	0.9557	0.5545
SD	0.9137	0.8452	0.6602
	Upper tail — Proportion Misclassified as outliers for $X_{(17)}$		
	$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.4499	0.1377	0.0193
SD	0.1971	0.1425	0.0725

$X_{(j)} = j$th order statistic.
C = Carling method.
SD = Schwertman and de Silva method.

of the simulations compared to the Schwertman and de Silva rates 0.1648, 0.1171, and 0.0589 for nominal outside rates 0.10, 0.05 and 0.01 respectively.

For the asymmetric chi-square distribution the same general pattern of classifying observations occurs, as was observed for the normal distribution. That is, Schwertman and de Silva was much better at not misclassifying on observation as an outlier whereas Carling was better at correctly identifying a true outlier but more likely to misclassify a second uncontaminated observation as an outlier compared to Schwertman and de Silva. There was one exception when $n = 20$ and $r = 0.01$ when Carling was better than Schwertman and de Silva. To illustrate the pattern for a different outside rate, when the sample size is 20 with no outliers at nominal outside rate $r = 0.05$ the Carling procedure misidentified observations as an outlier in the lower tail in 0.1902 of the simulations whereas the Schwertman and de Silva method misclassified in 0.0167 of the simulations. Similarly for the upper tail, the Carling procedure incorrectly identified an observation as an outlier in 0.6009 of the simulations while the Schwertman and de Silva rate was 0.1664. The corresponding simulation proportions for sample size 100 and nominal outside rate 0.05 are, for the lower tail using Carling's method, 0.1286 compared to 0.0000 for Schwertman and de Silva. For the upper tail the rate for Carling's method is 0.9592 while for Schwertman and de Silva the rate is 0.2893.

When there is one outlier with $n = 20$ in the upper tail using the nominal outside rate $r = 0.05$, Carling's method misidentified an outlier in the lower tail in 0.1738 of the simulations while the Schwertman and de Silva proportion was 0.0131, however Carling correctly detected the outlier in 0.9963 of the simulations compared to 0.8010 for Schwertman and de Silva. Carling however misidentified a second upper tail observation as an outlier in 0.4353 of the simulations compared to 0.2664 for Schwertman and de Silva. The same pattern occurred when n was 100.

From the simulations with multiple outliers, the same general previously observed trend occurred. Specifically, the Schwertman and de Silva method was superior almost always, and in several cases substantially superior, to the Carling procedure in not misclassifying an observation as an outlier while Carling's method was almost always superior in correctly identifying outliers. This was particularly true for identification of the largest outlier where Carling's method was substantially superior, especially with the large sample size. However, the superiority of the Carling method diminished with identification of the second, third, fourth, and fifth outliers and, in a few cases with small nominal

Table 10
10 000 simulations with five outliers from a standard normal distribution with $n = 100$.

	Lower tail — Proportion Misclassified as outliers for $X_{(1)}$				Upper tail — Proportion Correctly classified as outliers for $X_{(100)}$		
	$r = 0.10$	$r = 0.05$	$r = 0.01$		$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.9999	0.9307	0.3204		1.0000	1.0000	0.9847
SD	0.0223	0.0106	0.0019		0.2835	0.1298	0.0153
	Lower tail — Proportion Misclassified as outliers for $X_{(2)}$				Upper tail — Proportion Correctly classified as outliers for $X_{(99)}$		
	$r = 0.10$	$r = 0.05$	$r = 0.01$		$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.9975	0.7737	0.0880		1.0000	1.0000	0.9793
SD	0.0308	0.0163	0.0039		0.8889	0.7782	0.4621
					Upper tail — Proportion Correctly classified as outliers for $X_{(98)}$		
					$r = 0.10$	$r = 0.05$	$r = 0.01$
				C	1.0000	1.0000	0.9714
				SD	0.9845	0.9622	0.8490
					Upper tail — Proportion Correctly classified as outliers for $X_{(97)}$		
					$r = 0.10$	$r = 0.05$	$r = 0.01$
				C	1.0000	1.0000	0.9622
				SD	0.9981	0.9922	0.9606
					Upper tail — Proportion Correctly classified as outliers for $X_{(96)}$		
					$r = 0.10$	$r = 0.05$	$r = 0.01$
				C	1.0000	1.0000	0.9513
				SD	0.9997	0.9987	0.9884
					Upper tail — Proportion Misclassified as outliers for $X_{(95)}$		
					$r = 0.10$	$r = 0.05$	$r = 0.01$
				C	0.9984	0.8681	0.2346
				SD	0.7069	0.6333	0.4893

$X_{(j)}$ = jth order statistic.
C = Carling method.
SD = Schwertman and de Silva method.

outside rates, the Schwertman and de Silva procedure actually correctly identified outliers more frequently than the Carling method.

5. Examples

Example 1. To compare the two methods for detecting outliers, consider the following wood specific gravity data provided by Draper and Smith [11] but as contaminated by Rousseeuw and Leroy [19]. This same data set was also used as an example in the paper by Schwertman and de Silva [21].

Obs. number	1	2	3	4	5	6	7	8	9	10
Observation	0.534	0.535	0.570	0.450	0.548	0.431	0.481	0.423	0.475	0.48
Obs. number	11	12	13	14	15	16	17	18	19	20
Observation	0.554	0.519	0.492	0.517	0.502	0.508	0.520	0.506	0.401	0.568

The contaminated observations are 4, 6, 8, and 19. The sample quartiles are $q_1 = 0.478$, $q_2 = 0.507$, and $q_3 = 0.5345$. The IQR $= 0.0565$. As used in [21], a 75% confidence level was needed to improve the probability of identifying the four mild outliers. The computed Carling fences are 0.47874 and 0.53526 using Eq. (2.1) with a computed $k = 0.500175$ and with $\alpha_3 = 0$ and $\alpha_4 = 3$ for the normal distribution. The normal distribution was the one used by Rousseeuw and Leroy [19] and Schwertman and de Silva [21] and was therefore used in this example. These fences labeled observations 3, 4, 5, 6, 8, 9, 11, 19, and 20 as outliers. Thus the Carling procedure correctly classified all four outliers but

Table 11
10 000 simulations with three outliers from a chi square (8 df) with $n = 20$.

	Lower tail — Proportion Misclassified as outliers for $X_{(1)}$				Upper tail — Proportion Correctly classified as outliers for $X_{(20)}$		
	$r = 0.10$	$r = 0.05$	$r = 0.01$		$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.5195	0.1414	0.0183		0.9956	0.9175	0.5571
SD	0.0172	0.0096	0.0040		0.5424	0.4431	0.2733
	Lower tail — Proportion Misclassified as outliers for $X_{(2)}$				Upper tail — Proportion Correctly classified as outliers for $X_{(19)}$		
	$r = 0.10$	$r = 0.05$	$r = 0.01$		$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.2793	0.0510	0.0056		0.9950	0.9095	0.5375
SD	0.0276	0.0187	0.0089		0.8317	0.7705	0.6287
					Upper tail — Proportion Correctly classified as outliers for $X_{(18)}$		
					$r = 0.10$	$r = 0.05$	$r = 0.01$
				C	0.9943	0.8952	0.5730
				SD	0.9354	0.8973	0.8064
					Upper tail — Proportion Misclassified as outliers for $X_{(17)}$		
					$r = 0.10$	$r = 0.05$	$r = 0.01$
				C	0.3054	0.0937	0.0164
				SD	0.2138	0.1670	0.1044

$X_{(j)} = j$th order statistic.
C = Carling method.
SD = Schwertman and de Silva method.

also mislabeled five nonoutliers as outliers. The Schwertman and de Silva lower fences, as computed in [21], are: 0.4060, 0.4325, 0.4468, 0.4570, 0.4652, 0.4772 and the corresponding upper fences are: 0.6080, 0.5815, 0.5672, 0.5570, 0.5488, 0.5418 for $m = 1, 2, 3, 4, 5, 6$, respectively. This resulted in observations 4, 6, 8, and 19 being correctly labeled as outliers and no observations mislabeled as outliers.

Example 2. To illustrate the use of these methods on real data, the variable "Total Fat g" from the Candy Bar data set from JMP statistical software (Copyright (c) 2009, SAS Institute Inc., Cary, NC, USA, All Rights Reserved. Used with permission.) was used. In this example, rather than classifying an observation or observations as outliers, the techniques are used to identify candies that in a statistical sense differ substantially from the majority of other candies, similar in purpose to outlier identification. The data set is attached as Appendix B. Appendix C contains an analysis of the data using JMP. This data set has $q_1 = 8, q_2 = 12$, and $q_3 = 14$. The sample size is $n = 75$ with $\alpha_3 = 0.5020467 =$ skewness and $\alpha_4 = 3.9608338 =$ kurtosis when adjusted for the expected value for the standard normal which is 3.00. Carling's procedure produced $k_2 = 1.435895$ when the outside rate was 0.05. The Carling limits were $12-2(1.435895)(4) = 0.51284$ and $12+2(1.435895)(4) = 17.74358$. These limits identify nine data values as outliers. Specifically, the values labeled as outliers are: 0, 19, 20, 21, 22, 24, 25, 27, and 29. These values are labeled with a "c" in Appendix B.

The standard Tukey limits were $8 - 1.5(6) = -1$ and $14 + 1.5(6) = 23$. These limits identified four data values as mild outliers. Specifically, the values labeled as outliers are: 24, 25, 27, and 29. These values are labeled with an "a" in Appendix B.

In order to compute the Schwertman/de Silva limits, with $n = 75$, $\alpha_n = \frac{-\ln(1-0.05)}{75} = 0.000684$ which results in $Z_{\alpha_n} = 3.2013$. The Schwertman/de Silva limits were $12 + 6(3.2013)/1.36557 = -2.06577$ and $12+6(3.2013)/1.36557 = 26.065774$. These limits labeled two data values as outliers. These values are 27 and 29 and are labeled with a "b" in Appendix B.

The variable "Total Fat g" was chosen for illustration because the distribution of the values appeared to be approximately normal with the possible exception of those labeled as outliers. Subsequent testing for normality showed that the distribution including the outliers was statistically significant (that is, not normal). While we do not necessarily advocate the removal of the extreme

Table 12
10 000 simulations with five outliers from a chi square distribution (8 df) with $n = 100$.

	Lower tail — Proportion Misclassified as outliers for $X_{(1)}$			Upper tail — Proportion Correctly classified as outliers for $X_{(100)}$		
	$r = 0.10$	$r = 0.05$	$r = 0.01$	$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.9995	0.9291	0.3241	1.0000	1.0000	0.9849
SD	0.0237	0.0123	0.0020	0.2885	0.1332	0.0181
	Lower tail — Proportion Misclassified as outliers for $X_{(2)}$			Upper tail — Proportion Correctly classified as outliers for $X_{(99)}$		
	$r = 0.10$	$r = 0.05$	$r = 0.01$	$r = 0.10$	$r = 0.05$	$r = 0.01$
C	0.9964	0.7698	0.0886	1.0000	1.0000	0.9786
SD	0.0323	0.0164	0.0039	0.8897	0.7826	0.4649
				Upper tail — Proportion Correctly classified as outliers for $X_{(98)}$		
				$r = 0.10$	$r = 0.05$	$r = 0.01$
C				1.0000	1.0000	0.9705
SD				0.98451	0.9605	0.8466
				Upper tail — Proportion Correctly classified as outliers for $X_{(97)}$		
				$r = 0.10$	$r = 0.05$	$r = 0.01$
C				1.0000	1.0000	0.9607
SD				0.9982	0.9936	0.9694
				Upper tail — Proportion Correctly classified as outliers for $X_{(96)}$		
				$r = 0.10$	$r = 0.05$	$r = 0.01$
C				1.0000	1.0000	0.9485
SD				0.9997	0.9990	0.9896
				Upper tail — Proportion Misclassified as outliers for $X_{(95)}$		
				$r = 0.10$	$r = 0.05$	$r = 0.01$
C				0.9985	0.8696	0.2375
SD				0.7061	0.6372	0.4840

$X_{(j)}$ = jth order statistic.
C = Carling method.
SD = Schwertman and de Silva method.

observations identified by the various methods, it is of interest how their removal would effect the normality assumption for the remaining data. When the two outliers identified by the Schwertman/de Silva method were removed, the data set tested as normally distributed with a p-value of 0.0651 using the Shapiro–Wilks test. When Tukey's four outliers were removed, the data set was still statistically normal but the p-value was reduced to 0.0564. This indicates that removing the two additional values actually resulted in the remaining data having a distribution that is statistically less likely to be normal (small p-value). Deletion of the nine outliers identified by Carling's method resulted in the data set testing as extremely non-normal with p-value of 0.0009.

This example demonstrates that the removal of too many data values as in Carling's approach can reduce the likelihood of the remaining data having a normal distribution. Tukey's method maintained normality but since it does not adjust for sample size, it will tend to mislabel more data values as outliers as the sample size increases. The Schwertman/ de Silva method, while excluding fewer values, actually achieved a slightly better degree of normality than Tukey's method and clearly much better than Carling's method.

6. Conclusions and recommendations

The Carling approach is less subject to masking than the Schwertman and de Silva approach, that is, the Carling method is less likely to overlook an outlier. The Schwertman and de Silva method however is substantially less vulnerable to swamping, that is misclassifying an observation as an outlier.

For example, comparing the masking effect of the two methods, for normal data that included one outlier and outside rate $r = 0.05$, Carling's method correctly identified the outlier in nearly 100% of the simulations of both sample sizes compared to 39% and 62% for the Schwertman and de Silva method using sample size 20 and 100 respectively. Similarly for the chi-square study with one outlier the corresponding proportions for Carling remained at nearly 100% but the proportions for the Schwertman and de Silva method increased to 80% and 97% for sample sizes 20 and 100 respectively. Clearly the Carling procedure is superior with respect to masking.

To compare the swamping effect of the two procedures, consider for example, the same four data sets used to compare masking (i.e. normal and Chi square, $n = 20$ and 100, $r = 0.05$ and one outlier). The Carling approach misidentified a second observation as an outlier in the normal distribution in 45% and 95% of the simulations for sample sizes 20 and 100 respectively. The corresponding proportions for Schwertman and de Silva are 12% and 15% which clearly demonstrates the superiority of Schwertman and de Silva with respect to swamping.

Similarly, for the asymmetric Chi-square simulations with $r = 0.05$, Carling's method misclassified a second observation as an outlier in 44% and 95% of the simulations for samples sizes 20 and 100 respectively. The corresponding proportions for the Schwertman/de Silva procedure are 27% and 58% again demonstrating that the Schwertman/de Silva is better with respect to swamping.

Hence the Carling approach, while superior to masking effect, is sensitive to swamping and will almost always discard good data as the sample size increases. If this is a major concern then the alternative is to use the Schwertman and de Silva procedure which comes at the expense of overlooking some mild outliers but with a high likelihood of retaining noncontaminated data.

Our philosophy is that the decision to eliminate data as outliers is a serious matter and should not be taken lightly. Deletion of data subjects the experimenter to possible criticism. Unusual observations occasionally occur naturally in uncontaminated data and these observations still provide valuable information about the population. Therefore, we believe in a very conservative approach to establish the criteria for designating an observation as an outlier.

Our conservative criterion comes at the expense of failing to correctly identify some mild outliers. However, it is the extreme outliers which are grossly contaminated that cause the most difficulty in parametric data analysis. The process of detecting all outliers comes at the expense of eliminating many unusual but uncontaminated observations.

Based on this investigation of the Carling and Schwertman/de Silva modifications to the box-plot method of identifying outliers, the following patterns have emerged:

1. For symmetric or approximately symmetric data sets, use the IQR approach.
2. For asymmetric data sets use the SIQR approach.
3. As sample size n increases the Carling probability of misidentifying an observation as an outlier increases while the Schwertman and de Silva probability decreases for normal and for the thin tail of the asymmetric distribution. But in a skewed right distribution, the upper tail probability for both Carling and Schwertman and de Silva of misidentifying an observation as an outlier increases.
4. When compared to the Carling procedure, the Schwertman/de Silva method for both the symmetric and asymmetric distribution simulations has smaller proportions of misclassifying observations as outliers in both the upper and lower tails with the exception when $n = 20$ and $r = 0.01$ for the asymmetric distribution. For the asymmetric distribution both methods had a much larger probability of misidentifying an observation as an outlier than the nominal value. While the Schwertman/de Silva method had a substantially lower error rate than Carling's method, with nominal outside rate 0.01 the Schwertman/de Silva procedure still misidentifies an outlier in about 9% and 16% of the simulations of sample size 20 and 100 respectively.
5. The Carling method usually had a higher proportion of simulations which correctly identified the outlier for both the symmetric and asymmetric distributions. However, the proportions of correctly identified outliers for the two methods were considerably closer for the asymmetric simulations. The proportions for the Carling method were nearly identical for both the normal and the Chi-square distribution.

6. When the objective is to minimize the misclassification of uncontaminated observations while still eliminating the most extreme outliers, the Schwertman/de Silva method is well suited for this, particularly for larger data sets.
7. In the simulation of multiple outliers, the Schwertman and de Silva method generally outperformed the Carling procedure in not misclassifying observations as outliers. The Carling method was usually better at correctly labeling outliers, particularly in classifying the largest outlier where it was substantially superior especially for the larger sample. The Schwertman and de Silva method vastly improved in classifying the additional outliers, reducing the Carling method superiority compared to the Schwertman and de Silva method. In a few cases with small nominal outside rates, the Schwertman and de Silva procedure was superior in correctly labeling outliers.

Schwertman and de Silva [21] recommend stopping the sequential search when no additional observations are labeled by the new sequential fences. The simulations in this study suggest however that even if no outlier is identified with the initial fence, the search should continue by constructing the second fence before following the Schwertman/de Silva recommendation. In the multiple outlier simulations, stopping when the first fence failed to label an outlier obviously resulted in failing to recognize the multiple outliers when they existed. It is prudent to always search for at least two outliers using sequential fences.

The two methods, rather than being competitors, appear to complement each other. Carling's method is better at protecting against masking while the Schwertman/de Silva procedure is better at protecting against swamping. Used in tandem they provide the data analyst a more complete prospective for identifying outliers.

The Carling and Schwertman/de Silva procedures seem to be analogs to comparisonwise and experimentwise error rates in multiple comparisons. Just as comparisonwise and experimentwise error rates represent two different approaches to controlling errors in multiple comparisons, analogously, the Carling and Schwertman/de Silva methods represent two different approaches for avoiding mistakes in the identification of outliers. Specifically, the Schwertman/de Silva procedure is more conservative than those developed by Carling [10] and tends to maintain the overall "outside" experimentwise rate (the rate of misidentification of an observation as an outlier) at a more reasonable level for larger samples.

Appendix A

Conversion coefficients for IQR to σ (IQR $= k_n\sigma$)

n	**k**$_n$	**n**	**k**$_n$	**n**	**k**$_n$	**n**	**k**$_n$	**n**	**k**$_n$	**n**	**k**$_n$
5	1.65798	22	1.33333	39	1.38071	56	1.34361	73	1.36635	90	1.34535
6	1.28351	23	1.4023	40	1.34165	57	1.3713	74	1.34454	91	1.36267
7	1.51475	24	1.33753	41	1.38021	58	1.34329	75	1.36557	92	1.34562
8	1.32505	25	1.40096	42	1.34104	59	1.37004	76	1.34495	93	1.36258
9	1.50427	26	1.33587	43	1.37779	60	1.34394	77	1.36543	94	1.3455
10	1.31212	27	1.39455	44	1.34226	61	1.36981	78	1.34478	95	1.3621
11	1.45768	28	1.33894	45	1.37737	62	1.34366	79	1.36474	96	1.34576
12	1.32968	29	1.39355	46	1.34175	63	1.36871	80	1.34514	97	1.36201
13	1.45268	30	1.3377	47	1.37536	64	1.34424	81	1.36461	98	1.34565
14	1.32353	31	1.38876	48	1.34278	65	1.36851	82	1.34499	99	1.36157
15	1.42975	32	1.34004	49	1.37501	66	1.34399	83	1.36398	100	1.34588
16	1.33318	33	1.38799	50	1.34235	67	1.36754	84	1.34532	200	1.34740
17	1.42684	34	1.33909	51	1.37331	68	1.3445	85	1.36387	300	1.34792
18	1.32959	35	1.38428	52	1.34322	69	1.36737	86	1.34517	400	1.34818
19	1.41322	36	1.34092	53	1.37301	70	1.34429	87	1.3633	∞	1.34898
20	1.33568	37	1.38367	54	1.34285	71	1.3665	88	1.34548		
21	1.41132	38	1.34017	55	1.37156	72	1.34474	89	1.36319		

Appendix B

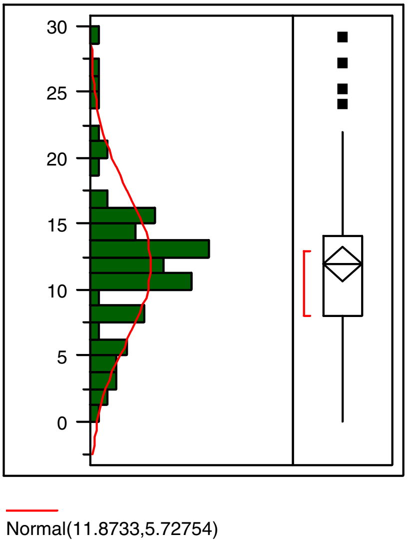
Brand	Name	Total fat g
M&M/Mars	Snickers Peanut Butter[c]	20
Hershey	Cookies 'n' Mint	12
Hershey	Cadbury Dairy Milk	12
M&M/Mars	Snickers	8
Charms	Sugar Daddy	2.5
M&M/Mars	Twix Peanut Butter	16
Hershey	Twizzler	1.5
Tobler	Toblerone	11
Nestle	Crunch	12
Hershey	Almond Joy	13
Sherwood	Elana Mint	10
Hershey	Krackel[c]	21
M&M/Mars	M&Ms Peanut	13
Bit-O-Honey	Bit-O-Honey	4
Nestle	100 Grand	8
Hershey	Skor	13
Hershey	Twix Caramel	14
M&M/Mars	Milky Way Lite	5
M&M/Mars	Mars	13
Pearson	Peanut Nut Roll	16
Nestle	Raisinet	8
Sherwood	Elana Mocca	13
Hershey	Bar None	15
Brown & Haley	Almond Roca[c]	19
Leaf	Payday	12
Just Born	Super Hot Tamales[c]	0
Hershey	Rolo	12
Nestle	Butterfinger	11
Myerson	Big Cherry	10
Hershey	Mr. Goodbar[a,b,c]	27
Hershey	Golden Collection[a,b,c]	29
Annabelle	U-No (Green)	17
Hershey	Reese's Peanut Butter Cup	14
M&M/Mars	Skittles	2.5
Hershey	Reese's Peanut Butter Cup Crunchy	16
Adams & Brooks	Cup O Gold	8
Nestle	Baby Ruth	12
Weider	Tiger Milk	6
M&M/Mars	Dove	13
Hershey	York Peppermint Patty	4
M&M/Mars	3 Musketeers	8
Annabelle	U-No (Blue)	17
Tootsie	Jr Mints	4
Hershey	Symphony (Blue)	15
Hershey	KitKat[c]	22
Leaf	Whoppers	10
Hershey	5th Avenue	12
Tootsie	Charleston Chew	7
Hershey	Kisses Almond	13
Hershey	Mound	13
Weider	Tiger Sport	2
Standard	Peanut Butter GooGoo	16
Hershey	Cadbury Roast Almond	13
M&M/Mars	Milky Way Dark	8
M&M/Mars	M&Ms Almond	11
Nabisco	Planters Original Peanut Bar	14
M&M/Mars	Snickers Munch	15
Hershey	Milk Chocolate	13
M&M/Mars	Milky Way	11
Annabelle	Look!	6
M&M/Mars	M&Ms Plain	10
Hershey	Cadbury Fruit & Nut	11

Brand	Name	Total fat g
Annabelle	Abba-Zabba	5
Hershey	Kisses	12
M&M/Mars	Whatchamacallit	13
Hershey	Symphony (Red)	14
Hershey	Special Dark[a,c]	24
M&M/Mars	M&Ms Peanut Butter	13
Hershey	Reese's Pieces	10
Nestle	Chunky	11
Hershey	Cadbury Caramello	9
Hershey	Milk Chocolate Almond[a,c]	25
Hershey	Reese's Nutrageous	14
Leaf	Heath	13
Annabelle	Big Hunk	3

[a] Tukey outlier.
[b] Schwertman/de Silva outlier.
[c] Carling outlier

Appendix C

Total fat g



Quantiles

100.0%	Maximum	29.000
99.5%		29.000
97.5%		27.200
90.0%		19.400
75.0%	Quartile	14.000
50.0%	Median	12.000
25.0%	Quartile	8.000
10.0%		4.000
2.5%		1.350
0.5%		0.000
0.0%	Minimum	0.000

Moments

Mean	11.873333
Std Dev	5.7275374
Std Err Mean	0.6613591
Upper 95% Mean	13.19112
Lower 95% Mean	10.555547
N	75
Sum Wgt	75
Sum	890.5
Variance	32.804685
Skewness	0.5020467
Kurtosis	0.9608338
CV	48.238664
N Missing	0

References

[1] D.F. Andrews, A robust method for multiple linear regression, Technometrics 16 (1974) 523–531.
[2] D.F. Andrews, D. Pregibon, Finding the outliers that matter, Journal of the Royal Statistical Society, Series B 44 (1978) 1–36.
[3] A.C. Atkinson, Fast very robust methods for the detection of multiple outliers, Journal of the American Statistical Association 89 (1994) 1329–1339.
[4] J. Bacon-Shone, W.K. Fung, A new graphical method for detecting single and multiple outliers in univariate and multivariate data, Applied Statistics 36 (1987) 153–162.
[5] V. Barnett, The study of outliers: Purpose and model, Applied Statistics 27 (1978) 242–250.
[6] V. Barnett, T. Lewis, Outliers in Statistical Data, 3rd ed., Wiley, Chicester, England, 1995.
[7] R.J. Beckman, R.D. Cook, Outlier.......... s, Technometrics 25 (1983) 119–163.
[8] M. Bhandary, Detection of the numbers of outliers present in a dataset using an information theoretic criterion, Communications in Statistics-Theory and Methods 21 (1992) 3263–3274.

[9] R. Brant, Comparing classical and resistant outlier rules, Journal of the American Statistical Association 85 (1990) 1083–1090.
[10] K. Carling, Resistant outlier rules and the non-Gaussian case, Computational Statistics & Data Analysis 33 (3) (2000) 249–258.
[11] N.R. Draper, H. Smith, Applied Regression Analysis, Wiley, New York, 1966.
[12] R. Gnanadesikan, J.R. Kettenring, Robust estimates, residuals, and outlier detection with multiresponse data, Biometrics 28 (1972) 81–124.
[13] F.R. Hampel, The breakdown points of the mean combined with some rejection rules, Technometrics 27 (1985) 95–107.
[14] D.M. Hawkins, Identification of Outliers, Chapman and Hall, London, 1980.
[15] D.C. Hoaglin, B. Iglewicz, Fine tuning, some resistant rules for outlier labeling, Journal of the American Statistical Association 82 (1987) 1147–1159.
[16] A.C. Kimber, Exploratory data analysis for possibly censored data from skewed distribution, Applied Statistics 39 (1990) 21–30.
[17] J.S. Milton, Statistical Methods in the Biological and Health Sciences, WCB/McGraw-Hill, New York, 1999.
[18] K.I. Penny, Appropriate critical values when testing for a single multivarate outlier by using the Mahalanobis distance, Applied Statistics 45 (1996) 73–81.
[19] P.J. Rousseeuw, A.M. Leroy, Robust Regression and Outlier Detection, Wiley, New York, 1987.
[20] N.C. Schwertman, M.A. Owens, R. Adnan, A simple more general box-plot method for identifying outliers, Computational Statistics and Data Analysis 47 (2004) 165–174.
[21] N.C. Schwertman, R. de Silva, Identifying outliers with sequential fences, Computational Statistics and Data Analysis 51 (2006) 3800–3810.
[22] J.W. Tukey, Exploratory Data Analysis, Addison-Wesley, 1977.