한국통계학회논문집 제2권 2호
1995년 12월 pp. 184-197

Detecting Multiple Outliers Using the Gaps of Order Statistics

Kim, Hyun Chul[1)]

Abstract

An objective and one-step detection procedure of multiple outliers is suggested by using the gaps of the order statistics. The detection procedure can be used as a routine outlier detection method of a statistical analysis computer program. The procedure is applied to some examples including the data selected by Kitagawa.

KEY WORDS: Outliers, Gaps, Consecutive tests, AIC, Order statistics, Grubbs-type statistic

1. Introduction

We consider the problem of detecting outliers. As some authors have pointed out, most definitions of outliers have some vague and subjective form(Prescott 1978, Beckman and Cook 1983). For example, Grubbs(1969, recited from Prescott 1978) says,

> "An outlying observation, or outlier, is one that appears to deviate markedly from the other members of the sample in which it occurs."

The following definition of detecting outliers is still ambiguous; however, when stated in the simplest form, it can be described as(Kitagawa 1979):

> "In estimating the mean value of normal population, it may happen that one or two values are surprisingly far away from the main group. We are tempted to consider that the suspicious observations are taken from a different population or that the sampling technique is at fault. However, we know that even from the normal population, there is a positive, although very small, probability that such observations will be drawn. The problem is to introduce an objective procedure for the detection and rejection of such outlying observations."

Most of the test procedures proposed for outlier detection consider the problem as one of

1) Department of Computer Science and Statistics, Kunsan National University, Kunsan, 573-701, KOREA.

hypothesis testing. To be brief, the null hypothesis is that the sample comes from a normal distribution with unspecified mean and variance, while the alternative hypothesis is that one or more of the observations come from a different distribution. Percentage points of a test statistic may be determined under the null conditions and compared with computed values of the test statistic when applied to specific samples.

The following criteria are frequently used for testing outliers in the single sample situation. The first three criteria follow conventional hypothesis testing and the last one has some other form. These tests will form the basic test procedures to be compared below.

Grubbs' criteria

Ratios of the sum of squares for a reduced sample excluding suspicious points to the sum of squares for the whole sample.

Dixon's criteria

Ratios of differences between an extreme value and its nearest or next nearest neighbor to the range, or some other measure of spread of the sample(possibly omitting the extreme observation).

Tietjen and Moore's criteria

Extend Grubbs statistic for the detection of multiple outliers.

Minimum AIC criteria

By defining a series of models where the smallest r_1 observations belong to one outlying population and the largest r_2 to another, the model with the minimum AIC(maximum entropy) is determined, and the corresponding r_1 smallest and r_2 largest observations declared outliers.

Most test procedures have some problems. For Dixon's and Grubbs' criteria are designed for testing single outlier, these methods should be modified or extended to be applicable to multiple outliers testing. Barnett and Lewis(1993 p.126) dubbed those modified methods as a consecutive test which apply single outlier tests repeatedly for the rejection of multiple outliers. However, numerous authors, realizing that the consecutive application of single outlier tests leads to some problems, proposed tests for the simultaneous rejection of outliers.

Tietjen and Moore(1972, 1979) proposed two Grubbs-type statistics for the detection of multiple outliers. They emphasized there the importance of the determination of the number of outliers before testing. They also suggested that the number of outliers should be estimated by use of the largest gap in order statistics. Rosner(1975) developed test statistics that are not prone to masking and require only knowledge of the maximum number of possible outliers. However, these two tests have a similar problem. If a number is selected a priori as the number of outliers and is smaller than the number of contaminants present, we run the risk of not detecting any contaminants because of masking. On the other hand, if the number is too large, we may discriminate valid observations as outliers because of swamping. Historically, the most annoying issue has been how to determine the number of outliers or an estimate thereof(Beckman and Cook 1983).

The fact that the number of outliers should be estimated a priori implies these test procedures are subjective at some degree. Most detection procedures are subjective from the fact that one should decide whether the data contain suspicious values or not. Gentleman and Wilk(1975) advocated automatic screening of data for outliers as part of data analysis. They stated that development of routine outlier detection methods is motivated from several considerations including the fact that in these days computers handle and analyze data, therefore data are invisible and that data are often huge so that it is hard to spot peculiar values by inspection.

The methods fitting distributions to the entire sample, such as Minimum AIC method, are advocated as one of the best multiple outliers test procedures by Beckman and Cook(1983). Moreover, minimum AIC no longer needs a two-step procedure of determination of number of outliers and hypothesis testing. However, in reality it is a time consuming and tiresome method.

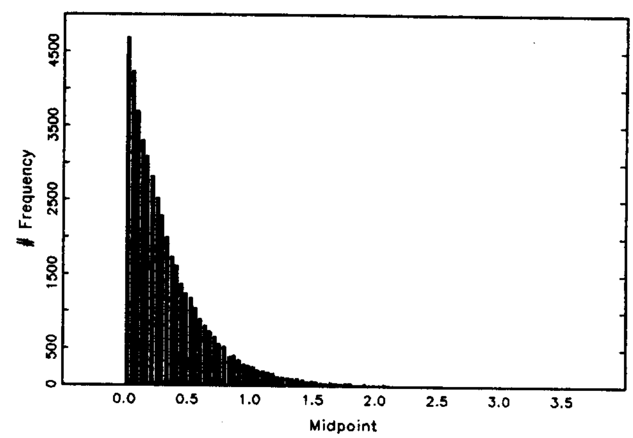
In this paper, the gaps test is proposed for the detection of outliers. The gaps test uses gaps in order statistics which were called blocks by Wilks(1962). The largest gap is used by Tietjen and Moore as a tool for estimating the number of outliers. If the gaps test is used, it needs not be modified or extended for multiple outliers test because single outlier testing and multiple outliers testing are no longer different. Furthermore, we need not estimate the number of outliers a priori, which has been the most annoying issue. It is not only objective but appropriate for routine outlier detection.

2. The gaps and multiple outliers testing

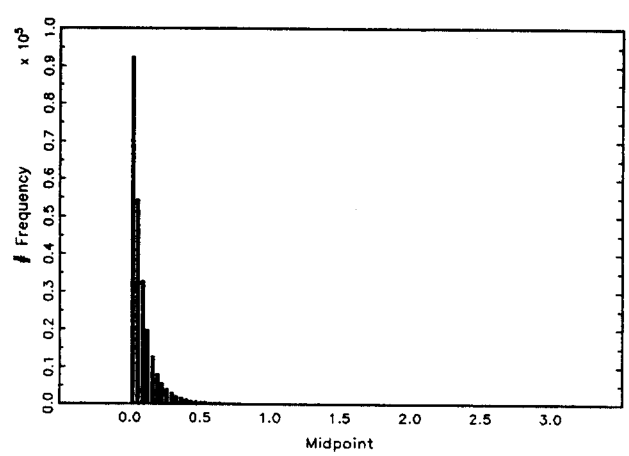
Suppose $x_1, \cdots, x_n$ is a sample from a population having a continuous c.d.f. $F(x)$. Let $x_1, \cdots, x_n$ be rearranged in ascending order and let the ordered values be $x_{(1)}, \cdots, x_{(n)}$. The intervals $(-\infty, x_{(1)}]$, $(x_{(1)}, x_{(2)}]$, $\cdots$, $(x_{(n)}, \infty)$ are called sample blocks $B_1, \cdots, B_n$ respectively, and the functions $F(x_{(1)})$, $F(x_{(2)})-F(x_{(1)})$, $\cdots$, $1-F(x_{(n)})$ of these blocks are called coverages $u_1, \cdots, u_{n+1}$ respectively. Since the sum of the u's is 1, we shall omit u_{n+1} as usual. Wilks(1962) showed that the coverages are random variables having the n-variate Dirichlet distribution. Since any $k(\leq n)$ of the coverages have the k-variate Dirichlet distribution, sample range may be used to transform coverages to blocks. However, the resulting distribution is still multivariate distribution which is difficult to be used for outlier testing.

Let $x_{(n)}$ be a suspicious point. If we use Dixon-type statistic, $(x_{(n)} - x_{(n-1)})/(x_{(n)} - x_{(1)})$ is used in this case. But if $x_{(n)}$ and $x_{(n-1)}$ or more points are suspicious, we must extend and modify the statistic over and over. The gaps test uses the gaps between the order statistics. If any gap is too large to be a gap from a normal sample, we conclude that the points outside the gap are outliers. In this procedure, we can disregard the number of points outside the gap(which is the number of outliers). This aspect makes the gaps test be applicable to both single outlier testing and multiple outliers testing. And we no longer need a two-step procedure of determination of number of outliers and hypothesis testing. Furthermore, since we need not information a priori, we can apply the test to automatic screening of data for outliers. If we can calculate the tail probability of the distribution of gaps, we will use the gaps test with ease.

However, in practice we can not still use the distribution of gaps, so we simulate the distribution by Monte Carlo method. Let's assume without loss of generality that the population follows a standard normal distribution. We generate pseudo normal random numbers of size n as a sample from $N(0,1)$, where n is a sample size(n=10(5)50(10)100). Gaps are obtained by sorting and differencing of the sample. This procedure is iterated 5,000 times and these 5,000(n-1) values mimic the distribution of blocks for each sample size. We get an empirical distribution from the values, and the distribution is regarded as a distribution of gaps in each sample size. All computations are done with GAUSS386 VM in a Pentium PC.



(a) n=10



(b) n=50

<Figure 1> Empirical distributions of gaps when sample size is (a) 10, (b) 50

<Table 1> Critical values at each significance levels and sample sizes

sample size \ significance level	α = 0.01	α = 0.05	α = 0.1
10	1.55271132	1.00682876	0.77887127
15	1.30756814	0.76804316	0.57114254
20	1.10818230	0.63302389	0.45399849
25	0.99924390	0.53233635	0.37905606
30	0.90494360	0.46129193	0.32426890
35	0.82531062	0.40964810	0.28215136
40	0.76367885	0.37174177	0.25307804
45	0.72271847	0.33516637	0.22786549
50	0.67026354	0.30728640	0.20636251
60	0.60882522	0.26588398	0.17441541
70	0.55695711	0.23377068	0.15254523
80	0.50492129	0.20545287	0.13376256
90	0.47340948	0.18783534	0.12017719
100	0.44344056	0.17022222	0.10870477

As expected, the shape of distribution has a peak at the neighbourhood of zero and a long right tail. The distribution is more short-tailed as the sample size grows up(Figure 1). We can read the value whose right tail area is 1%, 5%, 10% of the empirical distribution for each sample size(Table 1). These are the critical values of our gaps test.

$$
\begin{aligned}
\frac{1}{P_{0.01}} &= \underset{(0.01520)}{0.428955} + \underset{(0.000665)}{0.024039\,n} - \underset{(5.989E-6)}{0.000059\,n^2} \qquad \overline{R}^2 = 0.9991 \\
\frac{1}{P_{0.05}} &= \underset{(0.02216)}{0.437616} + \underset{(0.000969)}{0.058147\,n} - \underset{(8.734E-6)}{0.000040\,n^2} \qquad \overline{R}^2 = 0.9998 \\
\frac{1}{P_{0.1}} &= \underset{(0.01694)}{0.409425} + \underset{(0.000741)}{0.089478\,n} - \underset{(6.677E-6)}{0.000017\,n^2} \qquad \overline{R}^2 = 1.0
\end{aligned}
\tag{1}
$$

We can easily find, at a glance, that the critical values are fitted well by the reciprocal function of sample size n. Fitting results are equation (1), where $P_{0.01}$, $P_{0.05}$, and $P_{0.1}$ are critical values at significance level 0.01, 0.05, 0.1, respectively and the values in the parentheses are standard error of estimates. Equation (1) can be used to interpolate the table 1.

Testing procedure

With a data set, testing procedure is as follows.

Step 1. We must standardize the data, because we had simulated the distribution of gaps with standard normal variates.

Step 2. Get gaps by sorting and differencing the results.

Step 3. Compare some large values with the values of table 1(or calculated from the equation (1)). If a gap is greater than the critical value with a significance level α, the points outside the gap are regarded as outliers under the significance level α.

For example, we test the results of a 2^5 factorial experiment(Daniel 1959, recited from Barnett and Lewis 1993 p.40). Barnett and Lewis regarded the three observations(2.1470, -3.1430, -2.6660) as discordant outliers on the normal model.

-3.1430	-2.6660	-1.3050	-0.8980	-0.8138	-0.8138	-0.7577	-0.7437	-0.4771
-0.3087	-0.2526	-0.0982	-0.0842	-0.0561	0.0000	0.0281	0.1263	0.1684
0.1964	0.2245	0.2947	0.3929	0.4069	0.4209	0.4350	0.4630	0.5472
0.6595	0.7437	1.0800	2.1470					

In this example, the largest standardized gap is 1.3609777(between -2.666 and -1.3050), the second is 1.0669825(between 2.1470 and 1.0800) and the third is 0.447699220(between -2.6660 and -3.1430). But the third is ignored because the gap is outside the largest gap. Critical value of sample size 31 must be calculated by equation (1), because it is not tabulated. The value is 0.454187 with α=0.05. Now we can conclude that the three points, 2.1470, -3.1430, -2.6660 are multiple outliers.

3. Some theoretical backgrounds and modified testing procedure

We proposed gaps test for detecting multiple outliers in the previous section. For this purpose we did Monte Carlo simulation for standard normal data to get critical values. And we detect outliers from standardized data. However, in order for the gaps test to be accepted, the invariant property of the distributions of gaps from $N(\mu,\sigma^2)$ and $N(0,1)$ must be proved. In this section, we show this property and other problems.

<Lemma>

Let $X_1,\cdots,X_n$ be a random sample of size n from $N(\mu,\sigma^2)$, where μ and σ^2 are known. And $X_{(1)},\cdots,X_{(n)}$ be the order statistics of X_is. Then the distribution of gaps of standardized $X_{(i)}$s $\left(\text{i.e. } \frac{X_{(i)}-\mu}{\sigma}\right)$, where $i=1,\cdots,n$, is equivalent to the distribution of gaps of order statistics from $N(0,1)$, in the sense of gaps test.

<Proof>

Let $D_i^* = X_{(i)}^* - X_{(i-1)}^*$, where $i=1,\cdots,n$, and $X_{(i)}^*$s are the order statistics from $N(0,1)$, and suppose D_i^* has some density function f.

$D_i^* \sim f(\cdot)$

Now let $D_i = X_{(i)} - X_{(i-1)}$, where $i=1,\cdots,n$ and $X_{(i)}$s are the order statistics from $N(\mu,\sigma^2)$. Then

$$\frac{D_i}{\sigma} = \frac{X_{(i)}-\mu}{\sigma} - \frac{X_{(i-1)}-\mu}{\sigma}$$
$$= X_{(i)}^{**} - X_{(i-1)}^{**}$$

where $X_{(i)}^{**} = \frac{X_{(i)}-\mu}{\sigma}$. And the $X_{(i)}^{**}$s are order statistics from $N(0,1)$ too.

$$\therefore \frac{D_i}{\sigma} \sim f(\cdot)$$

So standardized order statistics from $N(\mu,\sigma^2)$ can be treated as order statistics from $N(0,1)$. ■

When μ and σ^2 are unknown, we can standardize using $\overline{X}$ and S, the estimates of μ and σ. If an independent estimate of μ and σ^2 is available, these could also be used. But for the estimates are affected by outliers, the power of test will be lessened in general. As a minor modification of other methods, we can replace $\overline{X}$ and S with estimates which might possibly be based on a restricted sample obtained by excluding outlying observations.

This idea can be applied to the gaps test to result in modified testing procedure. The new procedure lessen the sensitivity of the estimates and the masking effect as shown in the examples of following section.

Modified testing procedure

With a data set, modified new testing procedure are as follows.

Step 1. We must standardize the data, because we had simulated the distribution of gaps with standard normal variates.

Step 2. Get gaps by sorting and differencing the results.

Step 3. Compare the largest value with the values of table 1(or calculated from the equation (1)). If a gap is greater than the critical value with a significance level α, the points outside the gap are regarded as outliers under the significance level α. Otherwise we conclude there is no outliers, and stop.

Step 4. Recalculate the estimates of μ and σ^2, based on a restricted sample excluding the points that are regarded as outliers in Step 3.

Step 5. Apply the steps from 1 to 4 recursively.

4. Comparisons with other methods

In this section, we show some typical examples selected by Kitagawa(1979). First, we display the selected data set by a dot plot for the convenience sake of visualization of each example. Second, we detect outliers for the data set using both the listed test procedures and the gaps test, and compare the results.

We will use the following classical test statistics for comparison with our procedure.

(1) Test for single outlier:

(i) Dixon's test

$$r_{ij}^{1} = \frac{x_{i+1}-x_1}{x_{n-j}-x_1}, \qquad r_{ij}^{n} = \frac{x_n-x_{n-i}}{x_n-x_{j+1}}.$$

where i=1, j=0 for $n\le 7$, $i=j$=1 for n=8, 9, 10, i=2, j=1 for n=11, 12, 13 and $i=j$=2 for $n\ge 14$.

(ii) Grubbs' test

$$L_1 = \frac{S_1^2}{S^2}, \qquad L_n = \frac{S_n^2}{S^2}.$$

where

$$S^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2, \quad S_1^2 = \sum_{i=2}^{n}(x_i-\overline{x_1})^2, \quad S_n^2 = \sum_{i=1}^{n-1}(x_i-\overline{x_n})^2,$$

$$\overline{x_1} = \frac{1}{n-1}\sum_{i=2}^{n}x_i, \text{ and } \overline{x_n} = \frac{1}{n-1}\sum_{i=1}^{n-1}x_i .$$

(2) Test for several outliers(one-sided case):

(iii) Tietjen and Moore's test 1

$$L_k = \frac{S_k^2}{S^2}, \qquad L_{n-k} = \frac{S_{n-k}^2}{S^2}.$$

where

$$S_k^2 = \sum_{i=k+1}^{n}(x_i - \overline{x_k})^2, \quad S_{n-k}^2 = \sum_{i=1}^{n-k}(x_i - \overline{x_{n-k}})^2,$$

$$\overline{x_k} = \frac{1}{n-k}\sum_{i=k+1}^{n} x_i, \text{ and } \overline{x_{n-k}} = \frac{1}{n-k}\sum_{i=1}^{n-k} x_i.$$

(3) Test for several outliers(two-sided case):

(iv) Tietjen and Moore's test 2

$$E_k = \frac{\sum_{i=1}^{n-k}(z_i - \overline{z_k})^2}{\sum_{i=1}^{n}(z_i - \overline{x})^2}$$

where z_i is the value of x_i with the ith smallest distance from the mean $\overline{x}$ and

$$\overline{z_k} = \frac{1}{n-k}\sum_{i=1}^{n-k} z_i.$$

(4) Minimum AIC criteria:

(v) Model 1(mean shift model)

$$f_i(x) = \begin{cases} \phi(x;\mu_1, \sigma^2) & i = 1,\cdots,n_1 \\ \phi_{i-n_1,n-n_1-n_2}(x;\mu, \sigma^2) & i = n_1+1,\cdots,n-n_2 \\ \phi(x;\mu_2, \sigma^2) & i = n-n_2+1,\cdots,n \end{cases}$$

(vi) Model 2(variance inflation model)

$$f_i(x) = \begin{cases} \phi(x;\mu, \tau^2) & i = 1,\cdots,n_1 \\ \phi_{i-n_1,n-n_1-n_2}(x;\mu, \sigma^2) & i = n_1+1,\cdots,n-n_2 \\ \phi(x;\mu, \tau^2) & i = n-n_2+1,\cdots,n \end{cases}$$

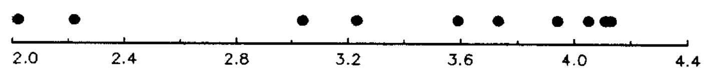
where, $\phi(x;\mu_1, \sigma^2)$ is the normal probability density function(p.d.f.) with mean μ and variance σ^2, and $\phi_{i,n}(x;\mu_1, \sigma^2)$ is the p.d.f. of the order statistic $X_{(i)}$ (i=1, $\cdots$, n) from the normal population.

Throughout the whole examples, unless mentioned otherwise, we will use the 5% significance level.

Example 1

The first example is the data given in Grubbs(1969, recited from Kitagawa 1979).

2.02 2.22 3.04 3.23 3.59 3.73 3.94 4.05 4.11 4.13



< Figure 2> A dot plot of the data of example 1

In this example, two lowest observations are highly suspicious points(Figure 2). As Kitagawa(1979) mentioned, all the consecutive tests are not appropriate to this example, because of the masking effect. On the other hand, because L_2 statistic of Tietjen and Moore(1972) is less than the 5% significance level, both 2.02 and 2.22 are rejected simultaneously as outliers. However, since L_3 and L_4 are less than the critical values too, the smallest four observations are rejected as outliers. By the minimum AIC criterion, 2.02 and 2.22 are rejected as outliers.

Since the largest gap(between 2.22 and 3.04), 1.0634474 is greater than the critical value, 1.00682876(n=10, α=0.05), we conclude the two lowest observations are outliers. The gaps test is less prone to masking effect than consecutive methods in this example. The next largest gap(between 3.23 and 3.59), 0.46687934 is even less than the critical value, 0.77887127(n=10, α =0.1), we accept 3.04 and 3.23 are ordinary points. And the results of modified gaps test are the same. This result shows the gaps test is superior to the test of Tietjen and Moore.

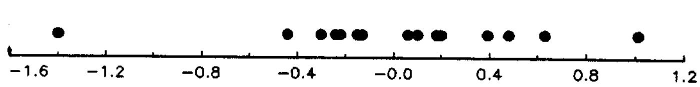
<Table 2> Testing results for example 1

Methods	Consecutive test		Tietjen & Moore's	Minimum AIC		Gaps test	Modified Gaps test
	Dixon's	Grubbs'		Model 1	Model 2		
Outliers	no outliers	no outliers	2.02, 2.22, 3.04, 3.23	2.02, 2.22	2.02, 2.22	2.02, 2.22	2.02, 2.22

Example 2

The second example is the data given in Tietjen and Moore(1972).

-1.40 -0.44 -0.30 -0.24 -0.22 -0.13 -0.15 0.06 0.10 0.18 0.20 0.39
0.48 0.63 1.01



< Figure 3> A dot plot of the data of example 2

The lowest point is highly suspected and the largest point is suspected slightly(Figure 3).

In this example, by the consecutive tests, we can not reject the largest point 1.01 as an outlier. Tietjen and Moore's test statistic is

$$E_2 = \frac{1.24089}{4.24964} = 0.292$$

The value is smaller than the 5% critical value of 0.317, -1.40 and 1.01 were rejected simultaneously. On the other hand, as Kitagawa pointed out, if a 10% significance level were to be taken, 0.63 and even 0.48 would also be rejected outliers. By model 1 of minimum AIC, -1.40, 1.01 and 0.63 are rejected as outliers, and by model 2, -1.40 and 1.01 are rejected as outliers. Kitagawa argued that since the value of the AIC of model 1 is the overall minimum, -1.40, 1.01 and 0.63 are rejected as outliers. But it seems to go a little bit too far that 0.63 is regarded as outlier(Figure 3).

Since the largest gap(between -1.40 and -0.44), 1.7377628 is greater than the critical value, 0.76804316(n=15, α=0.05), we conclude that the lowest observation is an outlier. Likewise the next largest gap(between 1.01 and 0.63), 0.68786443 is also greater than the critical value, 0.57114254(n=15, α=0.1), so we reject 1.01 as a outlier under the significance level 0.1. But modified gaps test isolates -1.40 and 1.01 simultaneously as outliers under α=0.05. The modified gaps test seems to be less affected by masking effect and less sensitive to outliers on estimating μ and σ.

<Table 3> Testing results for example 2

Methods	Consecutive test		Tietjen & Moore's	Minimum AIC		Gaps test	Modified Gaps test
	Dixon's	Grubbs'		Model 1**	Model 2		
Outliers	-1.40	-1.40	-1.40, 1.01, 0.63*, 0.48*	-1.40, 1.01, 0.63	-1.40, 1.01	-1.40, 1.01*	-1.40, 1.01

* under α=0.1
** the value of the AIC of model 1 is overall minimum

Example 3

Grubbs(1969, recited from Kitagawa 1979) presented an example of inter-laboratory testing. Each of three test observations was obtained from 12 laboratory. We show a dot plot of means of the each laboratory(Figure 4).

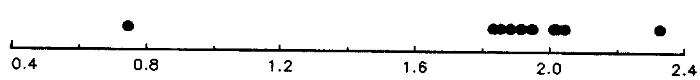
In this example, outliers tests had been applied to isolate the particular laboratories whose results varied significantly. From the ANOVA, Grubbs showed that the observations from Lab. 10 were highly significant. After doing ANOVA, we apply three multiple comparison methods to the data. By LSD and Tukey's methods, Lab. 10, 12 exhibit significantly different result from others. But only Lab. 10 is significantly different from others by the method of Scheffe. All other laboratories exhibited the same capability in testing procedure.

And our test procedure is useful in this situation too. The largest gap between Lab. 10 and

Lab. 3(2.8799882) is greater than the critical value which is calculated from equation (1). The second largest gap between Lab. 12 and Lab. 8 is greater than only the critical value of significance level 0.1. Especially modified gaps test could isolate Lab. 10 and Lab. 12 as outliers under α=0.05. Thus the modified gaps test turns out more improved for testing outliers than the original gaps test.

<Table 4> Data set of example 3

Lab.	Observations			Mean
1	1.893	1.972	1.876	1.914
2	2.046	1.851	1.949	1.949
3	1.874	1.792	1.829	1.832
4	1.861	1.998	1.983	1.947
5	1.992	1.881	1.850	1.884
6	2.082	1.958	2.029	2.023
7	1.992	1.980	2.066	2.013
8	2.050	2.181	1.903	2.045
9	1.831	1.883	1.855	1.856
10	0.735	0.722	0.777	0.745
11	2.064	1.794	1.891	1.916
12	2.475	2.403	2.102	2.327



< Figure 4> A dot plot of means of example 3

<Table 4> Testing results for example 3

Methods	ANOVA	Grubbs' conclusion	Minimum AIC		Gaps test	Modified Gaps test
			Model 1	Model 2		
Outliers	Lab. 10, 12**	Lab. 10, 12	Lab. 10, 12	Lab. 10, 12	Lab. 10, 12*	Lab. 10, 12

* under α=0.1

** Lab. 10, 12 are isolated by LSD and Tukey's multiple comparison methods, but by Scheffe's method only Lab. 10 is isolated.

5. Conclusion

As Tietjen and Moore pointed out, there are some difficulties in applying classical procedures for detecting several outliers. (i)By the successive use of a test for detecting a single outlier, we sometimes fail to detect any outlier when two or more outliers are located in the same direction, because of the masking effect of other outliers. On the other hand, (ii)by testing several outliers, we sometimes reject an observation which is not really an outlier because of the swamping effect of an outlier. So, some multiple outliers testing methods, such as Tietjen and Moore's or Rosner's, were suggested. But these methods are subjected to a priori information of k, the number of outliers. This fact makes them to be a two-step and subjective method.

Minimum AIC is an excellent criterion having the merits, such as (i)the determination of k and the test can be done simultaneously, (ii)various situations can be treated equally (Kitagawa 1979). However, since the values of AIC's must be calculated for the various combinations of possible outliers, the procedure becomes a tiresome and time consuming method.

In this paper, we proposed the gaps test for the detection of outliers. We used the gaps of order statistics directly, and the test may be further modified to a sequential method for lessening the masking effect and sensitivity on outliers in estimating μ and σ. The critical values are simulated by Monte Carlo method.

Some significant merits of new tests are similar to those of minimum AIC. (i)The method needs no a priori information of k, (ii)various situations can be treated equally as seen in the examples(e.g., single outlier, several lowest or highest outliers, two-sided case, and the grouped case such as example 3). (iii)The new tests are less prone to masking effect than other compared methods as shown in the examples. Moreover, our procedure is very simple to understand and apply. And the procedure can be applied to routine outliers test of statistical analysis computer package, because it is an objective method independent of a priori information of k and possible sets of outliers.

Acknowledgement

I would like to thank to the editor and referees for their comments in improving this paper.

References

[1] Barnett,V. and Lewis,T.(1993). *Outliers in Statistical Data 3rd ed.*, John Wiley, New York.

[2] Beckman,R.J. and Cook,R.D.(1983). Outlier..........s, *Technometrics*, Vol. 25, No. 2, 119-149.

[3] Dixon,W.J.(1953). Processing Data for Outliers, *Biometrics*, Vol. 9, 74-89.

[4] Gentleman,J.F. and Wilk,M.B.(1975). Detecting Outliers II. Supplementing the Direct Analysis of Residuals, *Biometrics*, Vol. 31, 387-410.

[5] Grubbs,F.E.(1969). Procedures for Detecting Outlying Observations in Samples, *Technometrics*, Vol. 11, 1-21.

[6] Kitagawa,G.(1979). On the Use of AIC for the Detection of Outliers, *Technometrics*, Vol. 21, 193-199.

[7] Prescott,P.(1978). Examination of the Behaviour of Tests for Outliers when More than One Outlier is present, *Applied Statistics*, Vol. 27, 10-25.

[8] Rosner,B.(1977). Percentage Points for the RST Many Outlier Procedure, *Technometrics*, Vol. 19, 307-312.

[9] Tietjen,G.L. and Moore,R.H.(1972). Some Grubbs-Type Statistics for the Detection of Several Outliers, *Technometrics*, Vol. 14, 583-597.

[10] Tietjen,G.L. and Moore,R.H.(1979). Corrigendum to Some Grubbs-Type Statistics for the Detection of Several Outliers, *Technometrics*, Vol. 21, 396.

[11] Wilks,S.S.(1962). *Mathematical Statistics*, John Wiley & Sons, New York.