TECHNOMETRICS ©, VOL. 21, NO. 2, MAY 1979

On the Use of AIC for the Detection of Outliers

Genshiro Kitagawa
The Institute of Statistical Mathematics
4-6-7 Minami-Azabu Minato-ku
Tokyo, JAPAN

An objective procedure for the detection of outliers is given by using Akaike's information criterion. Numerical illustrations are given, using data from Grubbs [8] and Tietjen and Moore [17].

KEY WORDS
Outliers
AIC
Entropy
Information
Hypothesis test
Modeling

1. INTRODUCTION

We consider the problem of detecting outliers. In its simplest form it may be stated as follows. 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.

Much work has been done within the framework of statistical hypothesis testing. Various statistics have been proposed for the test of the null hypothesis that all the observations are drawn from the same normal population. Frequently used are those by Grubbs [8], Dixon [6], David, Peason and Hartley [5], Ferguson [7], Shapiro and Wilk [15], Paulson [13], Murphy [10] and Kudo [9]. However, as pointed out by Tietjen and Moore [17], each of the tests is geared to a particular kind of outliers and the application of these procedures to a given set of data at a given level of significance often leads to conflicting conclusions with respect to the null hypothesis. Tietjen and Moore emphasized the importance of the determination of the number of outliers previous to the testing.

Akaike [2] has proposed an information criterion

$$\begin{aligned} \text{AIC} = &-2 \log (\text{maximum likelihood}) \\ &+ 2 (\text{number of independently} \\ &\qquad \text{adjusted parameters}), \end{aligned} \tag{1}$$

for the identification of an optimal model from a class of competing models. The information criterion has been used for modeling in various fields of statistics, engineerings and numerical analysis. See, for example, Akaike [1], [3], Otomo et al. [11], Otsu et al. [12], Sakamoto and Akaike [14] and Tanabe [16].

In this paper, two classes of models are proposed for the detection of outliers. The best approximating model is obtained by minimizing AIC. The problem of outlier detection is solved within the scope of estimation theory and we no longer need a two-step procedure of determination of type and number of outliers and hypothesis testing. By numerical examples it is shown that the proposed procedure will be effective for every type of outliers.

2. AIC AND MINIMUM AIC PROCEDURE

For the convenience of the readers, we present here a brief description of AIC and minimum AIC procedure and refer to Akaike [2], [4] for further details.

In statistical inference situations, Akaike [4] proposed consistent use of the entropy

$$B(f; g) = \int \log \left\{ \frac{g(z; x)}{f(z)} \right\} f(z)\, dz, \tag{2}$$

where x is the vector of observations, $f(z)$ and $g(z; x)$ are the probability density functions of the true and fitted models, respectively. According to the entropy maximization principle [4], the object of statistical inference is to estimate $f(z)$ from the data x and try to find $g(z; x)$ which will maximize the expected entropy

Received October 1977; revised June 1978

$$E_x B(f; g) = \int B(f; g) f(x)\, dx.$$

For the sake of simplicity, we will consider the simple situation where $f(x) = g(x|\theta^*)$ with $\theta^* \in R^L$. It should be noted that in [2] a more general situation is treated. Now, by expanding $B(\theta^*; \theta) \equiv B(g(x|\theta^*); g(x|\theta))$ in a Taylor series around θ^*, we have an approximation

$$B(\theta^*; \theta) = -\tfrac{1}{2} \|\theta - \theta^*\|_J^2, \tag{3}$$

where $\|\theta - \theta^*\|_J^2 = (\theta - \theta^*)' J(\theta - \theta^*)$ and J is the Fisher information matrix. Suppose ${}_k\theta^*$ is the projection of θ^* to a lower dimensional subspace R^k and ${}_k\hat{\theta}$ is the maximum likelihood estimate of ${}_k\theta^*$ obtained from N independent observations. Then it is well known that approximately

$$-2\, N\, B(\theta^*; {}_k\hat{\theta}) = N\, \|{}_k\theta^* - \theta^*\|_J^2 + N\, \|{}_k\hat{\theta} - {}_k\theta^*\|_J^2 \tag{4}$$

and the second term on the right-hand side of equation (4) is asymptotically a chi-square random variable with the degrees of freedom equal to k. It follows that approximately

$$-2N\, E_x B(\theta^*; {}_k\hat{\theta}) = N\, \|{}_k\theta^* - \theta^*\|_J^2 + k, \tag{5}$$

with E_x being expectation operator. On the other hand, it may be shown that

$${}_k\eta_L = -2 \sum_{n=1}^{N} \log \left\{ \frac{g(x_n \mid {}_k\hat{\theta})}{g(x_n \mid {}_L\hat{\theta})} \right\} \tag{6}$$

is asymptotically a non-central chi-square random variable with the noncentrality parameter $N\, \|{}_k\theta^* - \theta^*\|_J^2$ and the degree of freedom equal to $(L - k)$. It follows that ${}_k\eta_L + 2k - L$ is a suitable estimate of $-2NE_x B(f; g)$.

Thus Akaike [2] derived, by ignoring the constant terms, a criterion called Akaike's Information Criterion (AIC),

$$\begin{aligned} \text{AIC} = -2 \log\,&(\text{maximum likelihood}) \\ &+ 2(\text{number of independently} \\ &\qquad\qquad \text{adjusted parameters}), \end{aligned} \tag{7}$$

as an estimate of the quantity $-2NE_x B(f; g)$.

The AIC has a clear interpretation in model fitting. The first term of (7) indicates the badness of fit and the second the increased unreliability due to the increased number of parameters. The best approximating model is the one which achieves the most satisfactory compromise.

The model with the minimum of AIC is termed the Minimum AIC Estimate (MAICE) and the procedure having an intention of obtaining MAICE of the model is called the minimum AIC procedure. By the minimum AIC procedure, it is hoped that the entropy of the model will, at least approximately, be maximized.

3. TWO MODELS FOR DETECTING OUTLIERS

Let the sample of n observations be denoted in order of increasing magnitude by $x_1 \leqq x_2 \leqq \cdots \leqq x_n$. We assume that x_i is a realization of the random variable specified by a probability density function $f_i(x)$. Hereafter, we will use the notation $\phi(x; \mu, \sigma^2)$ for the normal distribution with mean μ and variance σ^2 and $\phi_{i,n}(x; \mu, \sigma^2)$ for the probability density function of the ordered variate $X_{i,n}$ $(i = 1, \cdots, n)$ from the normal population, i.e.,

$$\phi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{- \frac{(x - \mu)^2}{2\sigma^2}\right\} \tag{8}$$

and

$$\begin{aligned} \phi_{i,n}(x; \mu, \sigma^2) = C(i, n - i + 1)^{-1} \Phi(x)^{i-1} \\ \times \{1 - \Phi(x)\}^{n-i} \phi(x; \mu, \sigma^2), \end{aligned} \tag{9}$$

where

$$C(i, j) = \frac{\Gamma(i)\Gamma(j)}{\Gamma(i + j)} = \frac{(i - 1)!\,(j - 1)!}{(i + j - 1)!}$$

and

$$\Phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-(t-\mu)^2/2\sigma^2}\, dt.$$

Now, we consider the following two models:

MODEL 1

$$f_i(x) = \begin{cases} \phi(x; \mu_1, \sigma^2) \\ \qquad (i = 1, \cdots, n_1) \\ \phi_{i-n_1, n-n_1-n_2}(x; \mu, \sigma^2) \\ \qquad (i = n_1 + 1, \cdots, n - n_2) \\ \phi(x; \mu_2, \sigma^2) \\ \qquad (i = n - n_2 + 1, \cdots, n) \end{cases} \tag{10}$$

MODEL 2

$$f_i(x) = \begin{cases} \phi(x; \mu, \tau^2) \\ \qquad (i = 1, \cdots, n_1) \\ \phi_{i-n_1, n-n_1-n_2}(x; \mu, \sigma^2) \\ \qquad (i = n_1 + 1, \cdots, n - n_2) \\ \phi(x; \mu, \tau^2) \\ \qquad (i = n - n_2 + 1, \cdots, n) \end{cases} \tag{11}$$

Model 1 means that n_1 observations $x_1, \cdots, x_{n_1}$, $n - n_1 - n_2$ observations $x_{n_1+1}, \cdots, x_{n-n_2}$ and n_2 observations $x_{n-n_2+1}, \cdots, x_n$ each are realizations of normally distributed variables with the common variance σ^2 and mean μ_1, μ and μ_2, respectively. Model 2 means that $n - n_1 - n_2$ observations $x_{n_1+1}, \cdots, x_{n-n_2}$ and $n_1 + n_2$ observations $x_1, \cdots, x_{n_1}, x_{n-n_2+1}, \cdots, x_n$ are realizations of normally distributed variables with the common mean μ and variance σ^2 and τ^2, respectively.

In each model, we will consider the observations $x_1, \cdots, x_{n_1}$ and $x_{n-n_2+1}, \cdots, x_n$ as outliers. Thus, in Model 1, outliers are assumed to be distributed with the same variance σ^2 as the main group but different mean μ_1 or μ_2 ($\mu_1 < \mu < \mu_2$). And in Model 2 they are assumed to be normally distributed with the same mean μ as the main group but with the different variance $\tau^2 (> \sigma^2)$.

The likelihood function of Model 1 is given by

$$L(x; n_1, n_2, \mu, \mu_1, \mu_2, \sigma^2) = \prod_{i=1}^{n_1} \phi(x_i; \mu_1, \sigma^2) \prod_{i=n_1+1}^{n-n_2} \phi_{i-n_1, n-n_1-n_2}(x_i; \mu, \sigma^2) \times \prod_{i=n-n_2+1}^{n} \phi(x_i; \mu_2, \sigma^2). \tag{12}$$

Thus, the log likelihood $l_1(x; n_1, n_2, \mu, \mu_1, \mu_2, \sigma^2)$ of Model 1 is obtained by

$$-\frac{1}{2}\left\{n \log 2\pi + n \log \sigma^2 + \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu^i)^2\right\} - \sum_{i=n_1+1}^{n-n_2}[\log C(j, k-j+1) - (j-1)\log\{\Phi(x_i)\} - (k-j)\log\{1-\phi(x_i)\}] \tag{13}$$

where $j = i - n_1$, $k = n - n_1 - n_2$ and

$$\mu^i = \begin{cases} \mu_1 & 1 \leqq i \leqq n_1 \\ \mu & n_1 < i \leqq n - n_2 \\ \mu_2 & n - n_2 < i \leqq n. \end{cases}$$

Hence, it is now clear that the MAICE of the best approximating numbers of the outliers in the low side and high side are those values of n_1 and n_2 which minimize $\mathrm{AIC}_1(i, j)$ given by

$$\mathrm{AIC}_1(i,j) = \begin{cases} -2l_1(x; i, j, \hat{\mu}, \cdot, \cdot, \hat{\sigma}^2) + 2 \times 2 & (i = j = 0) \\ -2l_1(x; i, j, \hat{\mu}, \hat{\mu}_1, \cdot, \hat{\sigma}^2) + 2 \times 3 & (i \neq 0, j = 0) \\ -2l_1(x; i, j, \hat{\mu}, \cdot, \hat{\mu}_2, \hat{\sigma}^2) + 2 \times 3 & (i = 0, j \neq 0) \\ -2l_1(x; i, j, \hat{\mu}, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2) + 2 \times 4 & (i \neq 0, j \neq 0) \end{cases} \tag{14}$$

with the maximum likelihood estimates $\hat{\mu}$, $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\sigma}^2$.

Likewise the log likelihood $l_2(x; n_1, n_2, \mu, \sigma^2, \tau^2)$ of Model 2 is obtained by

$$-\frac{1}{2}\left\{n \log 2\pi + \sum_{i=1}^{n}\log\sigma^2(i) + \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma(i)^2}\right\} - \sum_{i=n_1+1}^{n-n_2}[\log C(j, k-j+1) - (j-1)\log\{\Phi(x_i)\} - (k-j)\log\{1-\phi(x_i)\}] \tag{15}$$

where

$$\sigma(i)^2 = \begin{cases} \sigma^2 & n_1 + 1 \leqq i \leqq n - n_2 \\ \tau^2 & 1 \leqq i \leqq n_1 \text{ or } n - n_2 < i \leqq n. \end{cases}$$

Now, $\mathrm{AIC}_2(i, j)$ of our Model 2 is given by

$$\mathrm{AIC}_2(i,j) = \begin{cases} -2l_2(x; i, j, \hat{\mu}, \hat{\sigma}^2, \cdot) + 2 \times 2 & (i = j = 0) \\ -2l_2(x; i, j, \hat{\mu}, \hat{\sigma}^2, \hat{\tau}^2) + 2 \times 3 & (\text{otherwise}) \end{cases} \tag{16}$$

with the maximum likelihood estimates $\hat{\mu}$, $\hat{\sigma}^2$ and $\hat{\tau}^2$ of Model 2.

TABLE 1—*The AIC's of Model 1 fitted to the data of Example 1 (* indicates minimum).*

Outliers (low side)	Outliers (high side): none	4.13	4.13, 4.11	4.13, 4.11, 4.05
none	1.61	4.27	5.09	5.82
2.02	0.01	2.74	3.56	4.23
2.02, 2.22	-6.79*	-4.30	-4.10	-4.63
2.02, 2.22, 3.04	-3.14	-0.23	0.89	2.42

4. NUMERICAL ILLUSTRATIONS

In this section, we will use the following classical test statistics for comparison with our procedure [8], [17],

(1) Test for single outlier:

(i) $T_1 = \dfrac{\bar{x} - x_1}{s}$, $\quad T_n = \dfrac{x_n - \bar{x}}{s}$.

(ii) 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 $n \leqq 7$, $i = j = 1$ for $n = 8, 9, 10$, $i = 2, j = 1$ for $n = 11, 12, 13$ and $i = j = 2$ for $n \geqq 14$.

(iii) Grubb's test

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

TABLE 2—*The AIC's of Model 2 fitted to the data of Example 1 (* indicates minimum).*

Outliers (low side)	Outliers (high side): none	4.13	4.13, 4.11	4.13, 4.11, 4.05
none	1.61	5.41	7.97	11.08
2.02	3.19	5.83	8.43	11.24
2.02, 2.22	1.01*	4.30	7.30	10.02
2.02, 2.22, 3.04	2.47	6.13	9.45	12.34

where

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2, \quad S_1^2 = \sum_{i=2}^{n} (x_i - \bar{x}_1)^2,$$

$$S_n^2 = \sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2,$$

$$\bar{x}_1 = \frac{1}{n-1} \sum_{i=2}^{n} x_i \quad \text{and} \quad \bar{x}_n = \frac{1}{n-1} \sum_{i=1}^{n-1} x_i.$$

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

$$\text{(iv)} \qquad L_k = \frac{S_k^2}{S^2}, \quad L_{n-k} = \frac{S_{n-k}^2}{S^2},$$

where

$$S_k^2 = \sum_{i=k+1}^{n} (x_i - \bar{x}_k)^2, \quad S_{n-k}^2 = \sum_{i=1}^{n-k} (x_i - \bar{x}_{n-k})^2,$$

$$\bar{x}_k = \frac{1}{n-k} \sum_{i=k+1}^{n} x_i \text{ and } \bar{x}_{n-k} = \frac{1}{n-k} \sum_{i=1}^{n-k} x_i.$$

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

$$\text{(v)} \qquad E_k = \frac{\sum_{i=1}^{n-k} (z_i - \bar{z}_k)^2}{\sum_{i=1}^{n} (z_i - \bar{x})^2}$$

where z_i is the value of x_i with the i-th smallest distance from the mean $\bar{x}$ and

$$\bar{z}_k = \frac{1}{n-k} \sum_{i=1}^{n-k} z_i.$$

The critical values for the statistics for some levels of significance are given in [8] or [17]. Throughout the whole examples, we will use the 5% significance level.

Example 1 The first example is the data given in Grubbs [8]:

2.02 2.22 3.04 3.23 3.59 3.73 3.94 4.05 4.11 4.13.

In this example two lowest observations are highly suspect. We find

$$T_1 = \frac{3.4060 - 2.02}{0.7711} = 1.7975$$

$$r_{11}^{1} = \frac{2.22 - 2.02}{4.11 - 2.02} = 0.0957$$

$$L_1 = \frac{3.217}{5.351} = 0.6011.$$

Thus, none of these tests can reject the smallest value, 2.02, since T_1 and r_{11} do not exceed their critical values 2.18 and 0.477, respectively, and L_1 is not less than its critical value 0.418.

On the other hand, since

$$L_2 = \frac{1.197}{5.351} = 0.224$$

is less then the 5% significance level 0.2305, both 2.02 and 2.22 are rejected simultaneously as outliers. This example shows that the sequential one-at-a-time procedures are not always useful because of the masking effect. On the other hand, it should be noted that since

$$L_3 = \frac{0.6564}{5.3510} = 0.1227$$

and

$$L_4 = \frac{0.2424}{5.3510} = 0.0453$$

TABLE 3—*The AIC's of Model 1 fitted to the data of Example 2 (* indicates minimum).*

Outliers (low side) \ Outliers (high side)	none	1.01	1.01, 0.63	1.01, 0.63, 0.48	1.01, 0.63, 0.48, 0.39
none	-18.52	-16.44	-14.13	-11.39	-8.17
-1.40	-23.34	-25.76	-26.29*	-25.85	-24.30
-1.40, -0.44	-17.68	-18.85	-17.63	-14.57	-9.01
-1.40, -0.44, -0.30	-13.78	-13.74	-10.74	-4.85	5.26
-1.40, -0.44, -0.30, -0.24	-10.55	-9.18	-3.98	5.94	23.48

TABLE 4—*The AIC's of Model 2 fitted to the data of Example 2 (* indicates minimum).*

Outliers (low side)	Outliers (high side): none	1.01	1.01 0.63	1.01 0.63 0.48	1.01 0.63 0.48 0.39
none	-18.52	-13.52	-9.07	-4.32	0.94
-1.40	-19.46	-20.05*	-17.89	-15.22	-11.96
-1.40 -0.44	-14.78	-15.90	-14 08	-11.72	-8.75
-1.40 -0.44 -0.30	-10.74	-12.09	-10.42	-8.17	-5.28
-1.40 -0.44 -0.30 -0.24	-6.86	-8.46	-6.97	-4.90	-2.17

are less than the critical values 0.129 and 0.070, respectively, the smallest four observations are rejected as outliers.

The values of the AIC's for Models 1 and 2 are given in Tables 1 and 2, respectively. We consider the model with the minimum AIC as the best approximating model. From both tables, it is concluded that 2.02 and 2.22 are outliers.

Example 2 The second example is the data given in [8] and [17]:

−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.

Grubbs [8] shows that by the successive application of tests for a single outlier, we cannot reject the largest value 1.01 as an outlier. Tietjen and Moore [17] show that since

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

is smaller than the 5% critical value of 0.317, −1.40 and 1.01 were rejected simultaneously. It should be noted that if a 10% significance level were to be taken, 0.63 and even 0.48 would also be rejected as outliers, since

$$E_3 = \frac{0.8774}{4.2496} = 0.2065$$

is almost equal to the 5% critical value of 0.205 and

$$E_4 = \frac{0.6287}{4.2496} = 0.1479$$

is less than the 10% critical value of 0.160.

The values of the AIC's are given in Tables 3 and 4. From these tables it is observed that by Model 1, − 1.40, 1.01 and 0.63 are rejected as outliers, and by Model 2, −1.40 and 1.01 are rejected as outliers.

It should be remembered that the AIC is an estimate of minus twice the entropy of the true probability density function with respect to the fitted model. Thus, we can also compare the goodness of fit of Model 1 and Model 2. The AIC's of the MAICE's (the minimum AIC estimates) of Model 1 and Model 2 are −26.29 and −20.05, respectively. Therefore, the MAICE of Model 1 is the overall MAICE and −1.40, 1.01 and 0.63 are considered to be outliers.

Example 3 Grubbs [8] presents an example of interlaboratory testing. Three test observations each were obtained from 12 laboratories, as shown in Display 1. The tests for outliers had been applied to isolate the particular laboratories whose results gave rise to the significant variation. The basic assumption in Grubbs' analysis was that the variances of the readings within all laboratories were equal. From the analysis of variance, he showed that the observations

DISPLAY 1—*Observations from 12 laboratories.*

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

TABLE 5—*The AIC's of Model 1 fitted to the data of Example 3.*
(indicates minimum).*

		Outliers (high side)		
		none	Lab.12	Lab.12 Lab. 8
Outliers (low side)	none	35.49	63.97	63.89
	Lab.10	-125.89	-154.68*	-129.92
	Lab.10 Lab. 3	94.61	61.59	112.26

from Laboratory 10 were highly significant. Then he applied his T' statistic to the mean values of the observations from each laboratory and concluded that the test method of Laboratory 10 should be investigated. Applying the two procedures successively, he concluded that the test methods of Laboratories 12 and 10 should be investigated, and that all the laboratories except 10 and 12 exhibited the same capability in testing procedure.

Our models are still useful in this situation. The AIC's of the models are given in Tables 5 and 6. From the tables it is shown that our procedure detects the observations from Laboratories 10 and 12 as outliers.

5. DISCUSSION

One may consider a criterion of fit obtained by replacing the second term of AIC (7) by $\alpha k (\alpha \neq 2)$ or by some other increasing function of k and N, where k and N are the number of independently adjusted parameters and the data length, respectively. Nevertheless the present definition of AIC seems to be the most reasonable as it gives an unbiased estimate of our cost function, the expected entropy $E_x B(f; g)$.

There might be some generalizations of our models. An example of such a generalization is:

Model 1'

$$f_i(x) = \begin{cases} \phi(x; \mu_1{}^j, \sigma^2) \\ \qquad (i = n_1{}^{j-1} + 1, \cdots, n_1{}^j; j = 1, \cdots, k_1) \\ \phi_{i-n_1, n-n_1-n_2}(x; \mu, \sigma^2) \\ \qquad (i = n_1 + 1, \cdots, n - n_2) \\ \phi(x; \mu_2{}^j, \sigma^2) \\ \qquad (i = n_2{}^{j-1} + 1, \cdots, n_2{}^j; j = 1, \cdots, k_2) \end{cases}$$

where $n_1{}^0 = 0$, $n_2{}^0 = n - n_2$,

$$\sum_{j=1}^{k_1} n_1{}^j = n_1 \quad \text{and} \quad \sum_{j=1}^{k_2} n_2{}^j = n_2.$$

In this generalized model, the outliers are grouped into several clusters having different mean values. The AIC of this model is obtained by

$$\text{AIC} = -2 \log (\text{maximum likelihood}) + 2(2 + k_1 + k_2).$$

TABLE 6—*The AIC's of Model 2 fitted to the data of Example 3.*
(indicates minimum).*

		Outliers (high side)		
		none	Lab. 12	Lab.12 Lab. 8
Outliers (low side)	none	35.49	68.89	69.68
	Lab.10	-109.58	-132.65*	-115.66
	Lab.10 Lab. 3	-91.80	-114.20	-94.78

Table 7 shows the values of AIC's when this generalized model is fitted to the data given in Example 3. From the table, it is observed that the same observations are detected as outliers but the value of AIC is considerably different when there are two groups of outliers in one side. The generalized model will be useful in such cases. However, the fitting of a model with too many parameters sometimes brings disastrous results. Thus, this generalized model is only recommended in the case where the variance σ^2 is known or the sample size is very large compared with the number of parameters of the model.

One of the referees pointed out the possibility of applying our procedure to the detection of outliers in linear regression and/or in two-way tables. This problem is the subject of future study.

6. CONCLUSION

There are two difficulties in applying classical procedures for detecting several outliers [17]: (i) By the successive use of a test for detecting a single outlier, we sometimes fail to detect any outlier because of the masking effect of other outliers. On the other hand, (ii) by testing for several outliers, we sometimes reject an observation which is not really an outlier because of the effect of an outlier.

Tietjen and Moore [17] emphasize that the proper choice of k, the number of outliers, is important, and present a heuristic procedure for the determination of k.

Another difficulty in applying classical procedures, which is inherent in statistical hypothesis testing, is that we do not know which significance level, e.g.,

TABLE 7—*The AIC's of Model 1' fitted to the data of Example 3.*
(indicates minimum).*

		Outliers (high side)		
		none	Lab.12	Lab.12 Lab. 8
Outliers (low side)	none	35.49	63.97	64.69
	Lab.10	-125.89	-154.68*	-143.43
	Lab.10 Lab. 3	-118.41	-142.62	-132.71

10%, 5% or 1%, should we take. Thus the procedures cannot be objective ones.

In this paper, we proposed a procedure for the detection of outliers. Our approach is quite different from the former ones. We used an explicit modeling of outliers and an objective decision procedure using an information theoretic criterion AIC.

The most significant merits of our procedure are that (i) the determination of k and the "test" can be done simultaneously by comparing the values of AIC's; (ii) various situations (e.g., single outlier, several lowest or highest outliers, two-sided case, and the grouped case such as Example 3) can be treated equally by our two models; and (iii) the procedure is free from the choice of the significance level, and one can make quite an objective decision.

7. ACKNOWLEDGEMENTS

The author would like to express his appreciation to Dr. H. Akaike of the Institute of Statistical Mathematics for his helpful suggestions and to Miss F. Tada for her assistance in computation. The author is also grateful to Mr. K. Hirano for his helpful discussions. Thanks are also due to the editor and the referees for their valuable suggestions. This research was partially supported by a grant from the Japanese Ministry of Education, Science and Culture.

REFERENCES

[1] AKAIKE, H. (1970). Statistical predictor identification. *Ann. Inst. Statis. Math., 22,* 203–217.

[2] AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory,* eds. B. N. Petrov and F. Csaki, pp. 267–281. Budapest: Akademiai Kiado.

[3] AKAIKE, H. (1976). Canonical correlation analysis of time series and the use of an information criterion. *System identification,* eds. R. K. Mehra and D. G. Lainiotis, pp. 27–96. New York: Academic Press.

[4] AKAIKE, H. (1977). On entropy maximization principle. *Proc. Symposium on Applications of Statistics,* ed. P. R. Krishnaiah, pp. 27–47. Amsterdam: North-Holland.

[5] DAVID, H. A., HARTLEY, H. O. and PEARSON, E. S. (1954). The distribution of the ratio, in a single normal sample, of range to standard deviation. *Biometrika, 41,* 482–493.

[6] DIXON, W. J. (1953). Processing data for outliers. *Biometrics, 22,* 74–89.

[7] FERGUSON, T. S. (1961). On the rejection of outliers. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability,* ed. J. Neyman, pp. 253–287.

[8] GRUBBS, F. E. (1956). Procedures for detecting outlying observations in samples. *Technometrics, 11,* 1–21.

[9] KUDO, A. (1956). On the testing of outlying observations. *Sankhya, 37,* 67–76.

[10] MURPHY, R. B. (1951). On tests for outlying observations. Unpublished Ph.D. thesis, Princeton University.

[11] OTOMO, T., NAKAGAWA, T. and AKAIKE, H. (1972). Statistical approach to computer control of cement rotary kiln. *Automatica, 8,* 35–48.

[12] OTSU, K., HORIGOME, M. and KITAGAWA, G. (1976). On the prediction and stochastic control of ship's motion. *Proceedings of the 2nd IFAC/IFIP Symposium,* eds. M. Pitkin, J. J. Roche and T. J. Williams, pp. 69–76. Washington, D. C.

[13] PAULSON, E. (1952). An optimum solution to the k-sample slippage problem for the normal distribution. *Ann. Math. Statist., 23,* 610–616.

[14] SAKAMOTO, Y. and AKAIKE, H. (1977). Analysis of cross classified data by AIC. *Ann. Inst. Statist. Math., B,* 30–1.

[15] SHAPIRO, S. S. and WILK, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52,* 591–611.

[16] TANABE, K. (1974). Statistical regularization of a noisy ill-conditioned system of linear equations by Akaike's information criterion. Research Memo. No. 60, The Institute of Statistical Mathematics, Tokyo, Japan.

[17] TIETJEN, G. L. and MOORE, R. H. (1972). Some Grubbs-type statistics for the detection of several outliers. *Technometrics, 14,* 583–597.