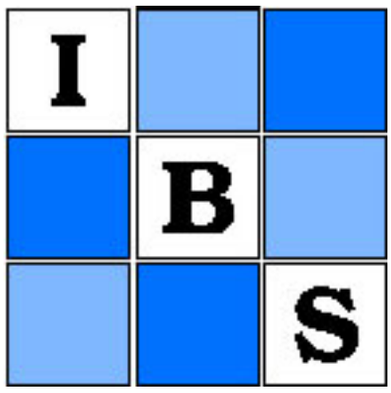

PROCESSING DATA FOR OUTLIERS[1]

W. J. Dixon

University of Oregon

1. *Introduction*

Every experimenter has at some time or other faced the problem of whether certain of his observations properly belong in his presentation of measurements obtained. He must decide whether these observations are valid. If they are not valid the experimenter will wish to discard them or at least treat his data in a manner which will minimize their effect on his conclusions. Frequently interest in this topic arises only in the final stages of data processing. It is the author's view that a consideration of this sort is more properly made at the recording stage or perhaps at the stage of preliminary processing.

This problem will be discussed in terms of the following general models. We assume that observations are independently drawn from a particular distribution or alternatively, we assume that an observation is occasionally obtained from some other population and that there is nothing in the experimental situation to indicate that this has happened except what may be inferred from the observational reading itself.[2]

We assume that if no extraneous observations occur, the observations (or some transformation of them, such as logs) follow a normal distribution. We shall also assume that the occasional extraneous observations are either from a population with a shifted mean or from a population with the same mean and a larger variance. These assumptions may not be completely realistic but procedures developed for these alternatives should be helpful.

If one is taking observations where either of these models apply there remain two distinct problems.

First, one may attempt to pick out the particular observation or observations which are from the different populations. One may be interested in this selection either to decide that something has gone wrong with the experimental procedure resulting in this observation (in which case he will not wish to include the result) or that this observation gives an indication of some unusual occurrence which the investigator may wish to explore further.

[1]This research sponsored by the Office of Naval Research.

[2]There is no attempt here to discuss the problem of rejecting observations statistically when there are known experimental conditions which make the observation suspect. For example, the dirty test tube or the rat that died of the wrong disease.

The second problem is not concerned with tagging the particular observation which is from a different population, but to obtain a procedure of analysis not appreciably affected by the presence of such observations. This second problem is of importance whenever one wishes to estimate the mean or variance of the basic distribution in a situation where unavoidable contamination occasionally occurs.

The first problem—tagging the particular observation—is of importance in looking for "gross errors" or mavericks, or the best or largest of several different products. Frequently the analysis of variance test for difference in means is used in the latter case. This is not really a very good procedure since many types of inequality of means have the same chance of being discovered. It should be noted that the power of the analysis of variance test decreases as more products are considered when testing in a situation of one product different from others which are all alike.

The problem of testing *particular* observations as outliers was discussed in reference (1). The power of numerous criteria was investigated and recommendations were made there for various circumstances.

This paper will concern itself primarily with the problem of contamination occurring according to the following model:

Outliers occur with a certain probability each time an observation is made. Let $N(\mu, \sigma^2)$ represent a normal population with mean, μ, and variance, σ^2. An observation from $N(\mu + \lambda\sigma, \sigma^2)$ introduced into a sample from $N(\mu, \sigma^2)$ is termed a location error. An observation from $N(\mu, \lambda^2\sigma^2)$ introduced into a sample from $N(\mu, \sigma^2)$ is a scalar error. It will be convenient to use the notation $C_+(N, \gamma, \lambda)$ or $C_\times(N, \gamma, \lambda)$ to represent samples of size N drawn from a population $N(\mu, \sigma^2)$ contaminated γ proportion from $N(\mu + \lambda\sigma, \sigma^2)$ or from $N(\mu, \lambda^2\sigma^2)$, respectively.

Section 2 will discuss the estimation of μ by use of the mean and median. Section 3 discusses the estimation of σ and σ^2 by the sample variance and the range. Section 4 gives recommended rules for processing data under various conditions of contamination.

2. *Effects of Contamination on the Mean and Median.*

The median has often been proposed as an estimator for μ under certain conditions of contamination. The ability of the mean and median to estimate μ can be compared by computing the mean square error (MSE) of the estimates for various types of contamination. The biases will be listed in several cases. The bias of the arithmetic mean is defined as $E(\bar{x} - \mu)/\sigma$ and the MSE is defined as $E(\bar{x} - \mu)^2/\sigma^2$. The criteria of *better* estimate of mean to be used here is *smaller* MSE.

TABLE I. SAMPLES NOT TREATED FOR CONTAMINATION

	$C_+(5, .10, \lambda)$				$C_+(5, .01, \lambda)$				
	Mean		Median			Mean		Median	
λ	Bias	MSE	Bias	MSE	λ	Bias	MSE	Bias	MSE
0	0	.200	0	.287	0	0	.200	0	.287
2	.2	.313	.15	.41	2	.02	.208	.02	.30
3	.3	.455	.18	.48	3	.03	.219	.02	.30
5	.5	.908	.20	.61	5	.05	.252	.02	.30
7	.7	1.588	.22	.80	7	.07	.302	.02	.30

From Table I, it can be concluded that the median is superior to the mean of untreated data for 10% contamination in samples of size 5, only

Contours indicating equality of MSE of median of untreated data and MSE of $\bar{x}$ of treated data.

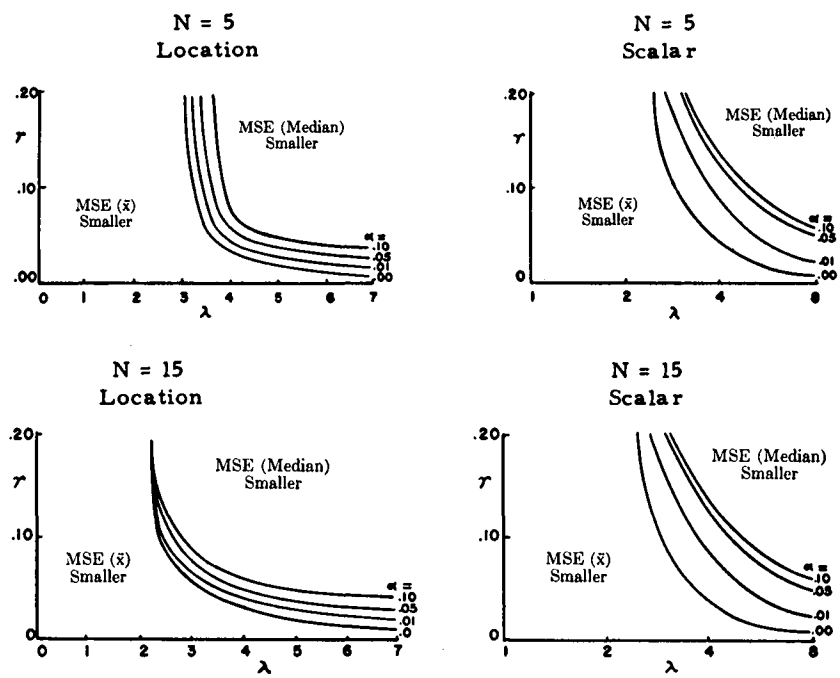

FIGURE 1.

if contamination is centered about 3.3σ or further from the mean. For 1% contamination the untreated mean is superior to the median for λ as large as 7. The reader is referred to Section 5 for the accuracy and method of obtaining the values in the above and succeeding tables.

The curves labelled $\alpha = 0$ in Figure 1 show the frequency and extent of contamination which can be tolerated before the MSE of the mean exceeds the MSE of the median. Curves are given for samples of size 5 and 15 and for location and scalar contamination. For example, for samples of size 5 which are 5% contaminated, the MSE, for $\bar{x}$ is smaller when the contaminating distribution is shifted 3σ but not when it is shifted 4σ.

Let us now consider changes in the above results when some of the contamination has been removed by the use of one of the r criteria of reference 2. A selection of critical values for these criteria is given in the Appendix.

Investigation was made using the 1, 5, and 10% levels of significance. The sample was tested until no further observations could be removed i.e. if a rejection was obtained at a certain level of significance, the reduced sample was again tested for outlier using the same level for α. This means, of course, that α should no longer be called a level of significance.

The additional curves in Figure 1 indicate the larger regions in which the MSE of $\bar{x}$ for treated samples is smaller than the MSE of the median for untreated samples when larger values of α are used. For extreme contamination an $\alpha = .20$ or .30 would further reduce the MSE for $\bar{x}$. This was not investigated in detail but it is known that this would not materially increase the size of λ and α which can be tolerated before the median should be used in preference to the mean.

In samples of size 5 use of the mean for treated samples results in most cases in a MSE *considerably smaller* than the MSE for the median. In cases of extreme contamination where the MSE for the median is smaller it is only slightly smaller. However, the MSE for the mean or the median is very large for heavy contamination. The use of the best treatment procedure still does not give us all that might be hoped for since the MSE is still large. The ratio of MSE for mean of treated data to MSE of mean for data with no contamination is an index of the extent of the contamination. We can also see from this index how much better the better estimate is. These ratios are given in Table II for samples of size 5 and 15. For samples of size 15 the picture is changed since the MSE for the mean becomes very large in the region where the MSE for the median is less than the MSE for the mean. Treatment is at the level $\alpha = .00, .01, .05$, or .10 which gives minimum MSE of $\bar{x}$.

TABLE II

$$\frac{\text{MSE}[\bar{x}, \text{ treated data}, C_+(5, \gamma, \lambda)]}{\text{MSE}[\bar{x}, N(\mu, \sigma^2)]}$$

γ \ λ	0	2	3	5	7
.01	1.00	1.04	1.08	1.2	1.3
.05	1.00	1.3	1.5	1.8	2.2
.10	1.00	1.6	2.2	3.4	5.1
.20	1.00	2.4	4.0	8.5	15.

$$\frac{\text{MSE [median}, C_+(5, \gamma, \lambda)]}{\text{MSE}[\bar{x}, N(\mu, \sigma^2)]}$$

γ \ λ	0	2	3	5	7
.01	1.43	1.5	1.5	1.5	1.5
.05	1.43	1.8	1.7	1.9	2.0
.10	1.43	2.1	2.4	3.1	4.0
.20	1.43	3.0	4.4	8.2	14.1

$$\frac{\text{MSE}[\bar{x}, \text{ treated data}, C_+(15, \gamma, \lambda)]}{\text{MSE}[\bar{x}, N(\mu, \sigma^2)]}$$

γ \ λ	0	2	3	5	7
.01	1.0	1.1	1.1	1.1	1.1
.05	1.0	1.3	1.7	1.9	2.3
.10	1.0	1.9	2.9	5.1	8.2
.20	1.0	4.0	7.6	20.	35.

$$\frac{\text{MSE [median}, C_+(15, \gamma, \lambda)]}{\text{MSE}[\bar{x}, N(\mu, \sigma^2)]}$$

γ \ λ	0	2	3	5	7
.01	1.6	1.6	1.6	1.6	1.6
.05	1.6	1.8	1.8	1.8	1.8
.10	1.6	2.3	2.3	2.3	2.3
.20	1.6	4.3	5.0	6.	7.

From these ratios we can see that in samples of size 5 even with extreme contamination the median is not a satisfactory substitute for $\bar{x}$ if the samples are treated for contamination. However, in samples of size 15 one should use the median if contamination beyond the $\alpha = .10$ curve (Figure 1) is expected.

The increase in MSE for $\bar{x}$ caused by removing occasional values which are not contaminators is very small in comparison to the reduction of MSE of $\bar{x}$ obtained by the removal of extreme contaminators.

Use of a large value of α will discover more contaminators but, of course, will increase the MSE of $\bar{x}$ if samples do not contain outliers. This effect is, however, small. Use of $\alpha = .10$ in samples of size 5 containing no contamination will increase MSE of $\bar{x}$ from .200 to approximately .216. Therefore, unless the contamination is believed to be slight a fairly large α should be used. In order to obtain minimum MSE for the estimate of μ, we can consider using one of the following procedures:

a) use of $\bar{x}$ after treating for rejection with $\alpha = .01$.
b) use of $\bar{x}$ after treating for rejection with $\alpha = .05$.
c) use of $\bar{x}$ after treating for rejection with $\alpha = .10$.
d) use of the median.

Table III gives the procedure resulting in the least MSE among the four procedures considered for the various types of contamination and sample sizes. The numbers in parentheses are the MSE resulting.

The MSE figures in Table III would not be increased by more than 5% by the use of $\alpha = .10$ in place of $\alpha = .01$ or .05 (or by 10% over no treatment). This is a small effect compared to the 50% to 200% increase in MSE resulting if certain extreme contamination is not removed. This fact will be taken into account in laying down general rules.

3. *Bias of s^2 and the range.*

The effect of contamination on the estimate of variance can be assessed by computing the amount of bias resulting. Let us define $B = E(s^2)/\sigma^2$. Since removal of outliers will reduce the variance, it is possible to make $B = 1$ for any γ and λ by choosing α sufficiently large. Since investigation was carried out only for $\alpha = .01, .05, .10$ (and a few results for $\alpha = .20$) it will not be possible to state the appropriate α for heavy or extreme contamination. Table IV lists the values obtained. It is believed that the consideration of bias alone is sufficient for the estimate of variance since in general the mean and MSE of the variance are closely related.

TABLE III. MINIMUM MSE FOR FOUR TREATMENTS

$N = 5$, location contamination

γ \ λ	0	2	3	5	7
.01		$a(.21)$	$a(.22)$	$c(.23)$	$c(.23)$
.02	use $\bar{x}$	$a(.22)$	$b(.24)$	$c(.26)$	$c(.27)$
.05	(.200)	$b(.25)$	$b(.30)$	$c(.37)$	$c(.44)$
.10		$b(.32)$	$c(.43)$	$d(.62)$	$d(.80)$
.20		$b(.49)$	$c(.81)$	$d(1.63)$	$d(2.82)$

$N = 15$, location contamination

γ \ λ	0	2	3	5	7
.01		$a(.07)$	$a(.07)$	$b(.07)$	$a(.07)$
.02		$a(.07)$	$a(.07)$	$c(.08)$	$b(.07)$
.05	use $\bar{x}$	$b(.08)$	$b(.11)$	$c(.13)$	$d(.12)$
.10	(.067)	$b(.13)$	$d(.16)$	$d(.16)$	$d(.16)$
.20		$b(.27)$	$d(.33)$	$d(.4)$	$d(.4)$

$N = 5$, scalar contamination

γ \ λ	1	2	4	8
.01		no treat. (.21)	$a(.22)$	$b(.24)$
.02	use $\bar{x}$	$a(.21)$	$b(.23)$	$b(.26)$
.05	(.200)	$b(.23)$	$c(.26)$	$c(.36)$
.10		$b(.26)$	$c(.34)$	$d(.50)$
.20		$b(.31)$	$d(.49)$	$d(1.0)$

$N = 15$, scalar contamination

γ \ λ	1	2	4	8
.01		$a(.07)$	$a(.07)$	$a(.07)$
.02		$a(.07)$	$a(.07)$	$a(.07)$
.05	use $\bar{x}$	$b(.07)$	$b(.08)$	$b(.10)$
.10	(.067)	$b(.08)$	$c(.11)$	$c(.24)$
.20		$c(.08)$	$c(.23)$	$d(.70)$

TABLE IV. APPROPRIATE α TO REMOVE BIAS IN s^2

$N = 5$, Location errors

γ \ λ	2	3	5	7
.01	.03	.05	.07	.05
.02	.06	.08	.10	.12
.05	.12			
.10				
.20				

$N = 15$, Location errors

γ \ λ	2	3	5	7
.01	.12	.10	.08	.01
.02	.14	.12	.10	.02
.05	.20	.25		
.10				
.20				

$N = 5$, Scalar errors

γ \ λ	2	4	8
.01	.02	.04	.05
.02	.03	.07	.10
.05	.07	.12	
.10	.12		
.20			

$N = 15$, Scalar errors

γ \ λ	2	4	8
.01	.10	.10	.05
.02	.20	.20	.10
.05			
.10			
.20			

TABLE V

Bias in s^2 for $C_+(5, \gamma, 5)$.

α \ γ	0	.10	.20
.00	1.00	3.52	6.0
.01	.99	1.90	5.3
.05	.95	1.52	4.4
.10	.91	1.26	4.0

The notation $\alpha = .00$ indicates results for untreated data.

Here again it is much more serious to allow contamination to remain than to remove non-contaminators incorrectly so that in general one should lean toward a large α. Table V illustrates this effect. For example, if contaminatoin is not present and we use $\alpha = .10$, we underestimate

TABLE VI

Appropriate α to remove bias of range estimate of σ.

Location errors

γ \ λ	2	3	5	7
.01	.01	.03	.04	.04
.02	.02	.05	.06	.05
.05	.06	.09	.10	.12
.10	.10			
.20				

Scalar errors

γ \ λ	2	4	8
.01	<.01	.01	.05
.02	.01	.04	.09
.05	.03	.05	
.10	.09		
.20			

σ^2 by 9%; but if 10% contamination at 5σ is present the use of the same rejection criterion will give us an overestimate of only 26% in place of 250% in samples of size 5.

In very small samples the *range* is often used to estimate the population standard deviation. Contamination will, of course, seriously affect the sample range, but the rejection criteria can effectively remove the bias in the range estimate of the population standard deviation for samples of size 5. Table VI shows the α which will result in an unbiased range estimate of σ in samples of size 5.

Table VII shows the bias of the range estimate of σ for one type of contamination. As before, it is much more serious to leave contamination in than to remove a few observations from samples which contain no contamination.

TABLE VII

Bias of the range estimate of σ for $C_+(5, \gamma, 5)$.

α \ γ	0	.10	.20
.00	1.00	1.66	2.11
.01	.99	1.50	1.91
.05	.97	1.27	1.63
.10	.93	1.13	1.48

The above results indicate that the range estimate of σ is less affected by contamination than s^2 even if no treatment is applied. Table VIII has been constructed to compare:

(1) s^2, the estimate of σ^2.
(2) ks, the estimate of σ where $E(ks) = \sigma$.
(3) the range estimate of σ.

TABLE VIII

Appropriate α to remove bias in $C_+(5, \gamma, 2)$.

γ	s^2	range estimate	ks
.01	.02	.01	.02
.02	.03	.02	.04
.05	.07	.06	.10
.10	.12	.10	>.10

The comparison is given in terms of the level α necessary to remove the bias in each of the estimates. The bias may be removed from the range estimate with a smaller value of α.

4. *Recommended Rules for Processing Data for Outliers.*

The problem of test of significance for tagging an individual as extraneous, extreme or as a "gross error" is pretty straightforward. We choose a level of significance, using the standard considerations and make a test on the set of observations we are processing. If a significant ratio is obtained we declare the extreme value to be from a population differing from that of the remaining observations. Depending on the practical situation we then declare the extreme individual a "gross erorr" or an exceptional individual. The best* statistic for this test if σ is known is the range over σ for outliers in either direction or the ratio $(x_n - \bar{x})/\sigma$ for a one-sided test. x_n represents the largest observation. For a one-sided test in the other direction we substitute $\bar{x} - x_1$ for $x_n - \bar{x}$. Here x_1 represents the smallest observation. The power of these tests is discussed in reference [1]. Critical values for range over σ are given in reference [4] and for $(x_n - \bar{x})/\sigma$ in reference [3].

If an independent estimate of σ is available, the best tests for outliers are the same as above with s replacing σ. Critical values for these tests are in references [3] and [4]. If no external estimate of σ is available the best statistics are the r-ratios of reference [2]. Critical values for these ratios are given in the Appendix.

Now, suppose that in place of tagging an individual observation from some different distribution, we wish to estimate the parameters of the basic distribution free from these contaminating effects. How might we process the data to come closer to the mean and variance of this basic distribution?

If very little is known about the contamination to be expected, about the best one can do is to "tag" observations as above and remove them from estimates of μ and σ.

If even a moderate amount of information about the type of contamination to be expected is available, a process can be prescribed which will minimize the effects of contamination on the estimates of mean and dispersion in small samples. The following rules result from the investigation of sections 2 and 3 for samples of size 5 and 15. Rules will be presented for these two sample sizes with the expectation that rules for samples of approximately these sizes will be approximately the same.

An attempt has been made to present simple rules for the estimation

*Best is used here in the sense of power greater than or equal to all other tests investigated in reference [1].

of both μ and σ. As a consequence the minimum MSE will not always be obtained. In most cases, however, the suggested procedure will yield a MSE which is not more than 5% larger than the minimum MSE. The rules will yield bias B between .90 and 1.10 for the indicated estimate of dispersion except in cases noted specifically in the rules.

Rules:

Process data using rejection criteria from Appendix I unless use of median is indicated. The appropriate α will be indicated in the rules below. Repeat application of criteria until no further observations are rejected. Use level of α as indicated in following statements.

$N = 5$, *Location contamination*

1. if $\gamma\lambda < .10$, use $\alpha = \gamma\lambda$, $\bar{x}$ for average, either s^2 or range for dispersion.
2. if $.10 < \gamma\lambda < .45$, use $\alpha = .10$, $\bar{x}$ for average, range for dispersion.
3. if $\gamma\lambda > .45$ use median for average, use $\alpha = .10$ and range to estimate dispersion. The estimate of dispersion will be biased, giving an overestimation of σ of more than 10%. ($B \simeq 1.1$ for $\gamma\lambda = .45$, $B \simeq 1.5$ for $\gamma\lambda = 1.00$).

$N = 5$, *Scalar contamination*

4. if $\gamma\lambda < .45$, use $\alpha = \frac{1}{2}\gamma\lambda$, $\bar{x}$ for average, s^2 for dispersion (for $\gamma\lambda > .30$ use range for dispersion. Bias for both range and s^2 over 10%).
5. if $\gamma\lambda > .45$, use median for average, range for dispersion. The estimate of dispersion will be biased, giving an overestimate of σ of more than 40%.

$N = 15$, *Location contamination*

6. if $\gamma\lambda < .30$, use $\alpha = .10$, $\bar{x}$ for average, s^2 for dispersion. For $\gamma > .02$ the estimate of dispersion will have considerable bias. ($B \simeq 1.2$ for $\gamma\lambda = .2$; $B \simeq 1.4$ for $\gamma\lambda = .3$).
7. if $\gamma\lambda > .30$ use the median for average, use $\alpha = .10$ and s^2 for dispersion. s^2 is considerably biased ($B \simeq 2$ for $\gamma\lambda = .50$).

$N = 15$, *Scalar contamination*

8. if $\gamma\lambda < 1.00$, use $\alpha = .10$, $\bar{x}$ for average, s^2 for dispersion. For $\gamma > .02$ (unless $\gamma\lambda < .15$) the estimate of dispersion will have considerable bias. ($B \simeq 1.1$ for $\gamma\lambda = .2$; $B \simeq 1.5$ for $\gamma\lambda = .4$).

9. if $\gamma\lambda > 1.00$ use median for average and use $\alpha = .10$ and s^2 for dispersion. s^2 is considerably biased ($B \simeq 10$ for $\gamma\lambda = 1.6$).

An application of the above rules is given as Example 1.

Example 1. Suppose samples of size 5 are taken from each lot. It is expected that about 10% of the observations will be location errors of 3 to 4 standard deviations. Here $\gamma = .10$ and $\lambda = 3$ to 4. Then $\gamma\lambda = .30$ to .40 and we use rule 2. Observations are recorded in order of size of measurement for a sample of five and treatment process indicated.

$x_1 = 23.2$ For $N = 5$ and $\alpha = .10$, the critical value of $r_{10} = .557$

$x_2 = 23.4$ By inspection $x_1 = 23.2$ is acceptable

$x_3 = 23.5$ The test for $x_5 = 25.5$ is

$x_4 = 24.1$ $$r_{10} = \frac{25.5 - 24.1}{25.5 - 23.2} = \frac{1.4}{2.3} = .609$$

$x_5 = 25.5$ x_5 is rejected

For $N = 4$ and $\alpha = .10$, the critical value of $r_{10} = .679$

The test for x_4 is

$$r_{10} = \frac{24.1 - 23.5}{24.1 - 23.2} = \frac{.6}{.9} = .667$$

x_4 is accepted.

The average is $(23.2 + 23.4 + 23.5 + 24.1)/4 = 23.55$.

The range is $24.4 - 23.2 = 0.9$.

The estimate of standard deviation is $(.486)(0.9) = .44$.

It may not be possible to decide which type of contamination might be expected in a particular sampling situation. Also the type of contamination might not be of the comparatively simple sort discussed here. However, if a large number of observations (say 50 to 100) are collected we can estimate the amount and type of contamination present. If we are willing to assume that the observations can be considered to be drawn from a population composed of two normal populations in different proportions and with possibly different means and variances there is a method for estimating these components. The estimation may be done by trial and error or graphically as described in References [5] and [6] and then one of the simpler models discussed here may be selected as representing approximately the actual conditions of contamination. A

trial and error method will be preferable if the population is not *very closely* represented by two normal populations.

Example 2. A series of chemical determinations are made on known chemical solutions giving a distribution as follows:

Error	Frequency		Fitted curve is:
	Observed	Fitted	
>1.25	2	2	
.9	6	5	80 *percent* $\mu = -.175$
.3	59	60	$\sigma = .333$
−.3	142	142	
−.9	38	39	
−1.5	11	9	20 *percent* $\mu = -.55$
−2.1	0	4	$\sigma = 1.0$
−2.7	2	2	
<−2.95	3	0	
	263		

There is a shift in mean of slightly more than one standard deviation unit. The important factor of contamination here is the large standard deviation of the second distribution. Considering only the scalar contamination, we have $\gamma = .20$ and $\lambda = 3$ or $\gamma\lambda = .6$. In samples of size 5 from the above population, the rules suggest use of the median and range.

The Appendix gives critical values for criteria for processing contaminated data and a table of multipliers for estimating the standard deviation from the range.

5. *Accuracy of Tabular Values.*

Although some known results are included and some were determined analytically, most results were obtained by sampling methods. Most of the sampling results for $N = 5$ are based on 100 samples and those for $N = 15$ on 66 samples. However, since the results quoted are weighted sums of several determinations each based on 100 (for $N = 5$) the effective sample size is greater than 100. Furthermore, sampling results were obtained for several values of a parameter (e.g. λ) so that unfortunately large sampling deviations could in some cases be discovered and rectified by an increased amount of sampling. It is difficult to state the accuracy

to be associated with each figure, but the accuracy should be adequate for determining the comparatively large differences on which the recommended analysis is based. After the tables had been assembled several of the reported results were checked by additional sampling. No MSE or bias differed from the tabulated results by more than 15%. Errors of 15 or 20% would not change the recommended procedures appreciably.

The values known to be correct are all results reported for $\lambda = 0$, the quantities for the mean in Table I, the results for no contamination in Table II and III and the first line of Table V.

REFERENCES

1. W. J. Dixon, "Analysis of Extreme Values," *Annals of Math. Stat.*, Vol. 21 (1950) pp. 488–506.
2. W. J. Dixon, "Ratios Involving Extreme Values," *Annals of Math. Stat.*, Vol. 22 (1951) pp. 68–78.
3. K. R. Nair, "Tables of Percentage Points of the 'Studentized' Extreme Deviate From the Sample Mean," *Biometrika*, Vol. 39 (1952) pp. 189–191.
4. Joyce M. May, "Extended and Corrected Tables of the Upper Percentage Points of the 'Studentized' Range," *Biometrika*, Vol. 39 (1952) pp. 192–193.
5. Carl Burrau, "The Half-Invariants of the Sum of Two Typical Laws of Errors, with an Application to the Problem of Dissecting a Frequency Curve into Components," *Skandinavisk Aktuarietidskrift*, Vol. 17 (1934), pp. 1–6.
6. Bengt Stromgren, "Tables and Diagrams for Dissecting a Frequency Curve into Components by the Half-invariant Method," *Skandinavisk Aktuarietidskrift*, Vol. 17 (1934), pp. 7–54.

APPENDIX

CRITICAL VALUES AND CRITERIA FOR TESTING FOR EXTREME VALUES

N \ α	.30	.20	.10	.05	.02	.01	.005	Criterion
3	.684	.781	.886	.941	.976	.988	.994	
4	.471	.560	.679	.765	.846	.889	.926	
5	.373	.451	.557	.642	.729	.780	.821	$r_{10} = \frac{x_N - x_{N-1}}{x_N - x_1}$
6	.318	.386	.482	.560	.644	.698	.740	
7	.281	.344	.434	.507	.586	.637	.680	
8	.318	.385	.479	.554	.631	.683	.725	
9	.288	.352	.441	.512	.587	.635	.677	$r_{11} = \frac{x_N - x_{N-1}}{x_N - x_2}$
10	.265	.325	.409	.477	.551	.597	.639	
11	.391	.442	.517	.576	.638	.679	.713	
12	.370	.419	.490	.546	.605	.642	.675	$r_{21} = \frac{x_N - x_{N-2}}{x_N - x_2}$
13	.351	.399	.467	.521	.578	.615	.649	
14	.370	.421	.492	.546	.602	.641	.674	
15	.353	.402	.472	.525	.579	.616	.647	
16	.338	.386	.454	.507	.559	.595	.624	
17	.325	.373	.438	.490	.542	.577	.605	
18	.314	.361	.424	.475	.527	.561	.589	
19	.304	.350	.412	.462	.514	.547	.575	
20	.295	.340	.401	.450	.502	.535	.562	$r_{22} = \frac{x_N - x_{N-2}}{x_N - x_3}$
21	.287	.331	.391	.440	.491	.524	.551	
22	.280	.323	.382	.430	.481	.514	.541	
23	.274	.316	.374	.421	.472	.505	.532	
24	.268	.310	.367	.413	.464	.497	.524	
25	.262	.304	.360	.406	.457	.489	.516	

RANGE ESTIMATE OF STANDARD DEVIATION WHERE SAMPLE RANGE = w.

N	Estimate
2	.886w
3	.591w
4	.486w
5	.430w
6	.395w
7	.370w
8	.351w
9	.337w
10	.325w