Theor. Appl. Genet. 52, 201-207 (1978)

© by Springer-Verlag 1978

The Influence of Aberrant Values on the Statistics Related to a Selection Program

B. Thompson
Statistical Research Service, Research Branch Agriculture Canada, Ottawa, Ontario (Canada)

Summary Examples are presented to illustrate some of the effects aberrant values, in particular, measurement errors, may have on estimates of the genetic parameters related to selection studies. It is shown that aberrant values may cause observed response to selection pressure to differ considerably from predicted response. Possible dangers of indiscriminate screening are also discussed.

Key words: Outliers – Selection experiment – Analysis of variance – Heritability – Data screening

Introduction

The set of measurements obtained from almost any genetic experiment concerned with quantitative characters will include outlying observations. Some of these outliers will be 'true values' arising from the tails of the underlying distribution; others may be 'aberrant values' arising from various sources both biological (e.g., illness) and experimental (e.g., recording or measurement error). In general, it would be desirable to remove the aberrant values if, in the process, the true values were retained. Geneticists adopt various approaches in dealing with outliers: some reject them outright; some employ rules based on statistical theory in deciding which to reject; some choose to include them; etc. Whatever the approach, the justification is often intuitive, based largely on the experience of the geneticist. Rather surprisingly, little attention has been given to the theoretical aspects of the problems raised by aberrant values in genetic experiments.

One possible explanation of the lack of theoretical study is the difficulty in defining, in mathematical terms, the aberrant values which may appear in practical situations. However, in the analysis of two different sets of poultry data, the author has encountered a type of measurement error which can be readily represented in a mathematical model. It is the purpose of this paper to illustrate the influence of this type of aberrant value on the statistics associated with selection programs. In particular, it will be shown that predicted and observed responses to selection pressure may differ considerably if aberrant values are included.

This paper is also concerned with the effects on subsequent analyses of attempts to remove aberrant values. The detection of such values has been discussed at length in the statistical literature (see Grubbs (1950) and Anscombe (1960) for example). However, with notable exceptions such as Kruskal (1960) who commented on various practical aspects of the problem, most authors have devoted themselves to theoretical questions relating to the detection of aberrant values in random samples from the normal distribution. Recently, a wider range of topics has been considered (see, for example, Prescott (1975) on linear regression and Hawkins (1974) on multivariate data). Even so, few authors have discussed in any detail the effects on subsequent analyses of the various screening techniques which they propose. This aspect of data screening is of considerable importance to quantitative geneticists, who are often confronted with large data sets for which some method of screening seems imperative. An example will be presented which shows that an ill-conceived screening process can lead to more serious difficulties than those caused by the aberrant values it is intended to remove.

The discussion will relate to a selection program intended to improve production among progeny by selection of parents on the basis of individual performance. The populations involved are assumed to be sufficiently large to ensure that the various parameter estimates closely approximate their expected values. Although this assumption will limit the scope of the paper, the difficulties observed herein can scarcely be expected to be less in smaller populations.

0040-5752/78/0052/0201/$ 1.40

The Mathematical Model

The type of aberrant value to be considered was suggested by two large sets of data from control populations of laying pullets, each set consisting of more than 4000 individuals. The data were described in detail by Thompson (1974). The difficulties presented by these data have been alluded to by Finney (1974). Both data sets included body weight measurements taken at various times during the growth period of the pullets. As an example, the measurements from five of the birds in one data set are presented in Table 1. The measurements of the first bird represent a growth pattern typical of most birds in the data set, with relatively fast growth during the first few weeks followed by slower growth. Except for the underlined observations, the measurements from the other four pullets show a similar pattern. The discrepancies are approximately 500g for the first two birds and 1000g for the last two. Discrepancies of these magnitudes appeared sufficiently often in the data set to suggest that measurement errors of 500g and 1000g were not infrequent. In fact, the evidence in favour of such errors was considered sufficiently convincing in many cases (about 1% of the birds in the data set) that the aberrant values were adjusted by the appropriate amounts. In other cases where the evidence was less convincing, the observations were simply deleted.

A suggested explanation for the presence of these errors was the use of balance scales in the weighing operation. Such errors would be introduced by overlooking the removal (or addition) of a 500g or 1000g weight from (to) the balance pan. Similar errors with lighter weights would be sufficiently small to escape consistent detection.

If it is assumed that a measurement error of constant magnitude C occurs randomly, standard genetic models can be modified to include the aberrant values. Consider the hierarchical model

$$Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \epsilon_{ijk} \qquad \begin{aligned} i &= 1,\dots,s \\ j &= 1,\dots,d \\ k &= 1,\dots,n \end{aligned}$$

Table 1. Body weights (g) of five pullets at various stages during growth period

Age of bird	12 weeks	20 weeks	28 weeks	36 weeks	44 weeks
a)	1080	1540	1980	2110	2180
b)	680	1790[a]	1770	1820	1840
c)	950	1090	1910	2060	2060
d)	870	1480	1840	1840	2890
e)	1170	1720	2310	2360	1340

[a] Underlined weights appear to be aberrant

where Y_{ijk} is the response variable and the sire effect α_i, the dam effect β_{ij} and the random variation ϵ_{ijk} are independently distributed $N(0, \sigma_s^2)$, $N(0, \sigma_d^2)$ and $N(0, \sigma_e^2)$ respectively. An aberrant value with measurement error of C can be represented as

$$Y_{ijk} = \mu + C + \alpha_i + \beta_{ij} + \epsilon_{ijk}$$

It will be assumed that the measurement error appears as an 'upward' ($C > 0$) or 'downward' ($C < 0$) aberrant value with fixed frequencies q and r, respectively. The expected mean squares arising from this model are as follows:

Among sires	$\sigma_e^2 + pqC^2 + n\sigma_d^2 + nd\sigma_s^2$
Among dams/sire	$\sigma_e^2 + pqC^2 + n\sigma_d^2$
Within dam family	$\sigma_e^2 + pqC^2$

The effects of different magnitudes of error will be studied by varying C. The extension of the model to include the possibility of more than one magnitude of error in a data set is straightforward but this additional complication will not be considered.

In the examples to follow, it will be assumed that $\sigma_s^2 = \sigma_d^2 = 5000$ and $\sigma_e^2 = 30000$ so that $h^2 = .5$. These values are representative of the parameters of the 44-week body weights in the population from which the data in Table 1 were obtained.

Estimates of Heritability

Estimates of heritability are of fundamental importance in predicting response to selection pressure. The presence of aberrant values among the data may introduce considerable bias into these estimates. The magnitude of the bias will be considered in this section; its effect on prediction will be examined in the next section.

The expectation of the heritability estimate obtained from the hierarchical analysis of variance is approximated by

$$E(\hat{h}^2) \simeq \frac{4\,E(\hat{\sigma}_s^2)}{E(\hat{\sigma}_s^2) + E(\hat{\sigma}_d^2) + E(\hat{\sigma}_e^2)}$$

Strict equality cannot be assumed because the expectation of a ratio is not the ratio of the expectations. For sufficiently large populations, however, approximate equality will be achieved. In practice, this complication is usually disregarded.

It can be seen from the expected mean squares given in Section 2 that the only term in the formula above which is influenced by the type of aberrant value under study is $E(\hat{\sigma}_e^2)$, which becomes inflated as the magnitude and/or frequency of aberrant values increase. The effect on the expectation of the heritability is illustrated in Table 2.

Low magnitudes of measurement error introduce little bias for all frequencies of error considered. However, with larger errors the bias is considerable, even when the errors occur infrequently. Table 2 gives results only for upward aberrant values but downward aberrant values yield the same expectations and combinations of the two (where frequency = q + r) yield similar, though smaller, expectations of heritability.

It may seem unreasonable to assume that aberrant values occur as frequently as indicated in Table 2, but more than 2% of 44-week weights in one of the data sets motivating this study were judged aberrant by the author, even though the data set had already been screened. The additional aberrant values were detected by considering the growth patterns of each bird as a whole rather than by examining the measurements individually.

The practical consequences of aberrant values can be seen in the following example taken from the data set discussed in Section 2. Data screening suggested that 21 of 847 body weight measurements from one year were suspect. The necessary adjustments were carried out in three steps to illustrate the effect of the aberrant values; the results are shown in Table 3. It should be emphasized that these adjustments were dictated by the nature of the data and not by a desire to produce notable change in the estimates. Even in this relatively large data set, the inclusion of a few aberrant values has had a considerable influence on the heritability estimates. Less than 3% of the data has depressed the combined estimate by 1.5 standard deviations.

Table 2. Expectations of heritability estimates when upward or downward aberrant values occur

Magnitude (C) of measurement error	Frequency of aberrant value (q or r)				
	.01	.03	.05	.07	.10
100	.499	.496	.494	.492	.489
500	.471	.423	.386	.355	.320
1000	.401	.289	.229	.190	.154

Underlying parameters: $h^2 = .5$

$\sigma_e^2 = 30000$

Table 3. Estimates of heritability and σ_e^2 for 44-week body weights of 847 pullets

	Before screening	After screening[a]		
Estimates of heritability		(1)	(2)	(3)
From sire component	.52	.53	.60	.70
From dam component	.38	.46	.40	.47
Averaging the two estimates	.45	.50	.50	.58
Estimates of σ_e^2	33100	30800	29300	25200

Approximate standard errors of estimates: (based on analysis after screening)		
	sire	.19
	dam	.12
	pooled	.09

[a] Steps in adjusting the screened data:

(1) Five unexplained aberrant values removed
(2) Two observations apparently in error by 1000 g adjusted
(3) 14 observations apparently in error by 500 g adjusted

The results in Table 3 reflect another unfortunate effect of aberrant values: the estimates of heritability become more variable. If bias were the sole effect of aberrant values, the difference between the heritability estimates from the sire and dam components should remain consistent as the aberrant values are removed. This is clearly not the case in Table 3. In particular, it can be seen from the results in steps 1 and 2 that even two aberrant values in a large data set can have considerable influence on the relative magnitude of the two estimates. Hence, aberrant values could lead to differences in the estimates which might erroneously be considered evidence of genetic factors such as sex-linked genes or dominance.

As the results in Tables 2 and 3 show, the aberrant values tend to reduce the estimates of heritability. However, although the expectation of $\hat{h}^2$ must be smaller, the estimate itself may be inflated by the aberrent values in some situations. Consider, for example, a dam family consisting of a single individual from which an aberrant measurement is obtained. This observation will inflate the among-family sums of squares but will leave the within-family sums of squares unaffected. Thus, the increased variability will be attributed to genetic factors (or to maternal effects), and the heritability estimate will be inflated.

Aberrant Values in a Selection Experiment

Consider a hypothetical experiment involving a single generation of selection. A simple selection rule is assumed – any individual with a measured phenotype exceeding the population mean by more than one standard deviation (based on the estimate of σ_e^2) will be used as a parent. The parametric values $h^2 = .5$ and $\sigma_e^2 = 30000$ will be used throughout.

The standard prediction of response to selection, denoted below as 'predicted response', is the product of the estimated heritability and the apparent selection differential, the latter being determined as the difference between the mean of the parents and the mean of the population from which they were selected. However, this prediction is valid only when the usual assumptions of the genetic model hold. Generally, these assumptions cannot

Table 4. The effect of aberrant values on the predicted and observed response to selection

Frequency			Proportion in selected population			Selection differential				Response	
(1)[a]	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	Apparent	Predicted	Observed
.99	.00	.01	.940	–	.060	314	–	30	332	152	148 (143)[b]
.98	.00	.02	.877	–	.123	323	–	33	339	151	144 (134)
.95	.00	.05	.687	–	.313	348	–	42	384	148	126 (101)
.99	.01	.00	1.000^-	$.000^+$	–	306	751	–	311	146	153 (158)
.98	.02	.00	1.000^-	$.000^+$	–	307	752	–	317	141	153 (163)
.95	.05	.00	1.000^-	$.000^+$	–	307	753	–	332	128	154 (179)
.99	.005	.005	.970	$.000^+$	.030	310	756	29	317	149	151 (151)
.98	.01	.01	.939	$.000^+$	.061	315	762	31	328	146	149 (149)
.95	.025	.025	.840	$.000^+$	.160	329	778	35	362	138	141 (141)

[a] Headings:
(1) Measurement without error
(2) Measurement too small by 500 g
(3) Measurement too large by 500 g

[b] Values in brackets are the actual responses if no aberrant values are included among the progeny data

be justified when aberrant values are included among the data. In the present context, a valid prediction can be determined by first separating the selected population into two groups, the true and aberrant phenotypes. Because the necessary assumptions hold within each group, separate predictions of response follow directly. An overall prediction is obtained by averaging the separate predictions, weighted by the proportions included in the selected population. As this prediction should approximate the actual response to selection in a large population, it will be referred to as the 'observed response'. Note that in practice, unless the aberrant values are identified and the appropriate adjustments made, the sole prediction available to the breeder is the 'predicted response'; the 'observed response' will be reflected only in the difference between the means of the parent and progeny generations. It is the purpose of the present section to consider the effects of aberrant values on the predicted and observed responses, and in particular, on the discrepancy between these values. The results for various frequencies of aberrant values of magnitude 500 are presented in Table 4 and plotted in Figure 1.

It can be seen from Figure 1 that the observed response is almost unaffected by increasing numbers of

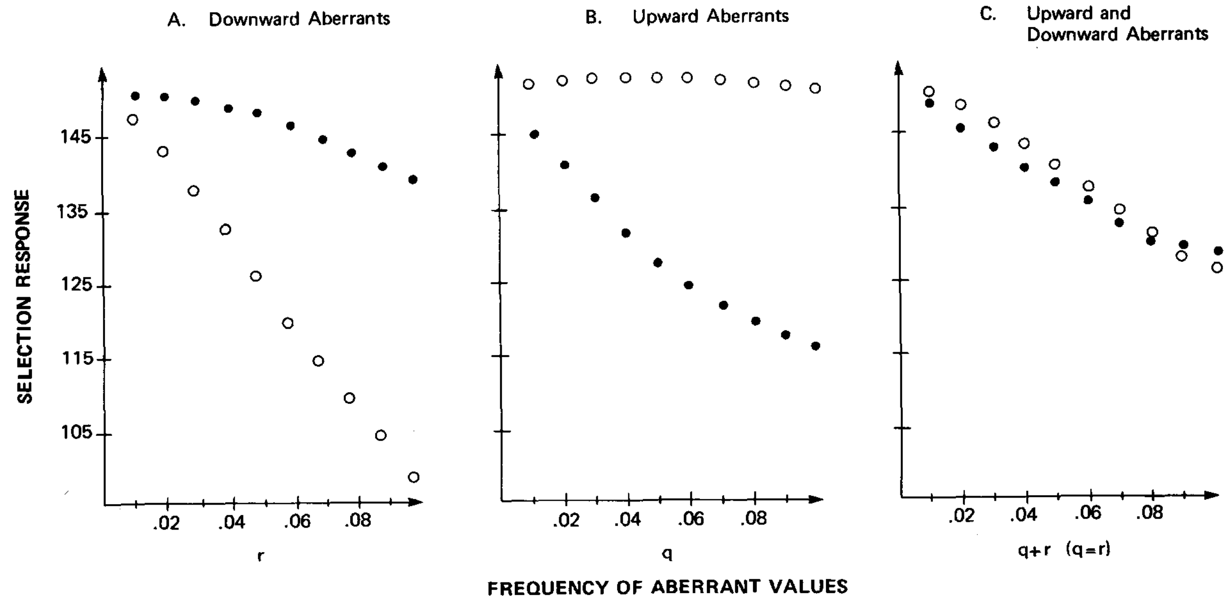

Fig. 1. The effect of aberrant values on the predicted and observed response to selection

downward aberrant values. This is to be expected because downward aberrant values simply mask superior phenotypes; in the large population assumed here, the superior but aberrant phenotypes lost to the selected population will be replaced by equivalent phenotypes. In smaller experiments where the superior phenotypes would be replaced by inferior phenotypes, the observed response would be reduced as the frequency of aberrant values increased. Figure 1 shows the predicted response to underestimate the observed response, a result due to the bias in the heritability estimate.

Upward aberrant values have a detrimental effect on observed response, as is apparent in Figure 1. Inferior, but aberrant, phenotypes will seem superior and hence will be included in the selected population. This problem will become severe as the frequency of aberrant values increases – when $q = .05$, 31% of the selected population has an effective selection differential of only 42. Although the heritability estimate is biased as in the case of downward aberrant values, the predicted response over-estimates the observed response because the apparent selection differential is inflated by the aberrant values in the selected population.

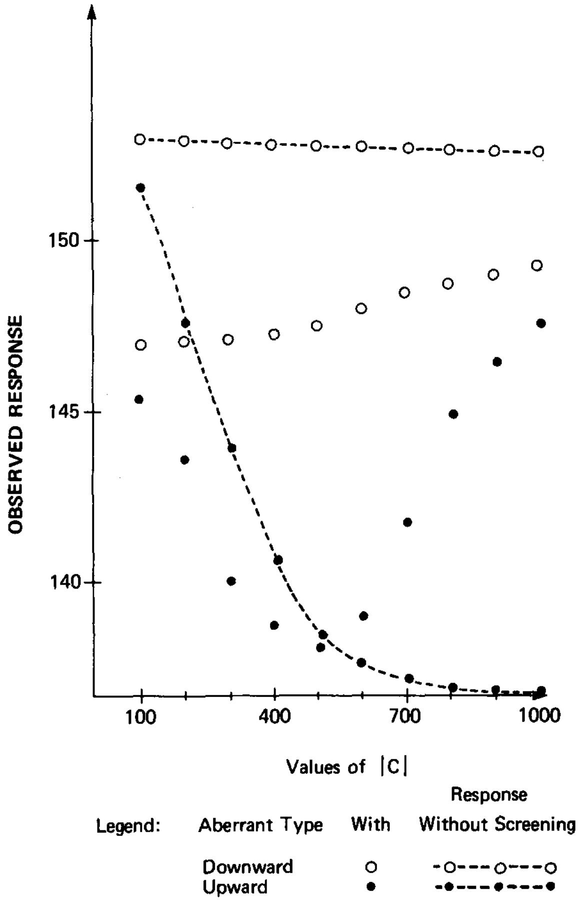

Fig. 2. The effect of screening on selection response

As Figure 1 shows, there is a marked difference between predicted and observed response, whether the aberrant values be downward or upward, when their frequency is not negligible. Regardless of which type of aberrant value appears, the difference would be increased even further if the cause of the aberrant values were removed in the progeny generation.

When both upward and downward aberrant values are included in equal proportions, the predicted response approximates the observed response fairly closely (Fig. 1). Again the observed response is reduced as more upward aberrant values are included in the selected population. The agreement between observed and predicted response is due to the predicted response being the product of an over-estimate of the selection differential and an under-estimate of the heritability.

The Detection of Aberrant Values

The examples of the last two sections have shown that the estimates associated with selection programs can be influenced considerably by aberrant values. It is natural then to consider methods of identifying these values so that remedial action can be taken. Although careful data screening will generally prove beneficial to subsequent statistical analyses, over-reaction to seemingly aberrant values may lead to more difficulties than those caused by the aberrant values themselves. It will be shown in this section that an arbitrary rejection rule may retard genetic progress in a selection program more than the aberrant values it is designed to remove.

Let us consider a rule by which all observations lying more than 2.5 standard deviations (based on the estimate of σ_e^2) from the observed mean are rejected as being aberrant. Assume that the underlying mean of the population

Table 5. Proportions of aberrant and true (non-aberrant) values from the original population included in the selected population

	True		Aberrant	
	Unscreened	Screened	Unscreened	Screened
a) Downward aberrant values				
C = −100	.159	.152	.067	.065
−400	.159	.153	.001	.001
−700	.159	.155	.000	.000
−1000	.159	.157	.000	.000
b) Upward aberrant values				
C = 100	.159	.153	.309	.287
400	.159	.154	.841	.579
700	.159	.156	.994	.253
1000	.159	.158	1.000	.032

free of aberrant values is 2000, that the frequency of aberrant values is 2% and that any individuals with phenotypes exceeding 2200 are selected as parents. The parametric values of Section 4 will again be assumed, that is, $h^2 = .5$ and $\sigma_e^2 = 30000$.

Figure 2 shows the response to selection, with and without screening, for varying magnitudes of measurement error. Although in each of the two situations represented a single type of aberrant value, either upward or downward, is considered, screening is assumed in both directions in both cases. Table 5 shows the proportions of the aberrant and true values which have been included in the selected populations.

Figure 2 shows that the screening strategem adopted here is counterproductive if the population includes only downward aberrant values. Owing to the rejection of superior phenotypes as being aberrant, the response to selection pressure from the screened data is considerably lower than from the unscreened data, even though relatively few true values have been rejected (Table 5). On the other hand, the rejection rule proves useful in processing data which include upward aberrant values with a measurement error in excess of 500. Although some true values are rejected, most of the aberrant values are removed from the selected population, thereby increasing the selection pressure. The screening process is less effective with a measurement error of less than 500 both because more true values representing superior phenotypes are rejected (see Table 5) and because the true phenotypes of the aberrant values which escape detection are inferior to those of the aberrant values which are detected.

This example may appear to paint too bleak a picture – no one would consider rejecting both extremely large and small observations if it were known that all measurement errors were in one direction only. However, seldom in practice will the geneticist be able to say with certainty that only one type of aberrant value is possible. The process used in the example would be reasonable if either upward or downward aberrant values might appear; the results could be expected to lie between the extremes presented here.

Whatever the approach, the geneticist will be confronted with the dilemma: is the extreme phenotypic value evidence of a superior genotype or is it simply an aberrant value? A simple ad hoc statistical rule cannot provide the answer. It is important that all information available, both statistical and biological, be used in reaching the best response. As the examples of the last two sections have shown, an incorrect decision on acceptance or rejection of an observation can lead to a considerable reduction in genetic improvement.

The real aim of the breeder is to achieve genetic improvement; to obtain reliable estimates of genetic parameters is of secondary interest. However, poor estimates may lead to faulty decisions concerning the selection program or to the construction of selection indices inappropriate to the needs of the breeders. As such, it may be advisable to attempt to remove aberrant values before estimating parameters even though the corresponding individuals clearly will not be included in the selected population. For example, in a poultry breeding program designed to improve egg production, hens which lay no eggs will certainly not be selected as prospective parents and yet their records, if not removed, may have an undesirable influence on estimates of heritability and selection differential.

The above is not meant to imply that once an observation has been rejected it can be ignored. On the contrary, unless the source of error is removed, the frequency and, wherever possible, the distribution of the aberrant values should be incorporated in the derivation of the prediction equations.

Conclusion

The examples presented in this paper are intended to emphasize the potential dangers of aberrant values and of the geneticists' response to them. It is the view of the author that careful data screening and prudent reactions to seemingly aberrant values will minimize these dangers. Aberrant values present problems specific to the data at hand and should be treated in context; undue reliance on ad hoc procedures can lead to serious deficiencies in subsequent analyses. Indeed, as the examples of this paper show, the large discrepancies often observed between predicted and actual response in selection programs may on occasion be attributable simply to the breeder's treatment of aberrant values.

Acknowledgement

This paper is based on work included in my Ph.D. thesis. As such, I would like to thank my two thesis supervisors, Profs. D.J. Finney and A. Robertson, both of the University of Edinburgh, for their support and guidance. I am also indebted to my colleagues in Statistical Research Service for their helpful comments concerning the manuscript.

Literature

Anscombe, F.J.: Rejection of outliers. Technometrics **2**, 123-147 (1960)

Finney, D.J.: Problems, data, and inference. J.Roy. Statist. Soc. A **137**, 1-23 (1974)

Grubbs, F.E.: Sample criteria for testing outlying observations. Ann. Math. Statist. **21**, 27-58 (1950)

Hawkins, D.M.: The detection of errors in multivariate data using principle components. J. Am. Statist. Ass. **69**, 340-344 (1974)
Kruskal, W.H.: Some remarks on wild observations. Technometrics **2**, 1-3 (1960)
Prescott, P.: An approximate test for outliers in linear regression. Technometrics **17**, 129-132 (1975)
Thompson, B.K.: Some statistical problems in the analysis of poultry data. Ph.D. Thesis. University of Edinburgh (1974)

Received December 5, 1977
Communicated by A. Robertson

Dr. Brian Thompson
Statistical Research Service
Research Branch Agriculture Canada,
Ottawa, Ontario (Canada)