The Study of Outliers: Purpose and Model

Author(s): Vic Barnett

Source: *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 1978, Vol. 27, No. 3 (1978), pp. 242-250

Published by: Wiley for the Royal Statistical Society

Stable URL: https://www.jstor.org/stable/2347159

Appl. Statist. (1978),
27, *No.* 3, *pp.* 242–250

The Study of Outliers: Purpose and Model

By Vic Barnett

University of Sheffield, Britain

[Received January 1978. Revised April 1978]

Summary

Outliers may influence the analysis of a set of data in various different ways. Some practical examples are used to motivate a categorization of the different aims in handling outliers and of the different models which might be employed to reflect the presence of outliers.

Keywords: OUTLIERS; OUTLIER-GENERATING MODELS; TEST OF DISCORDANCY

1. Introduction

The legal case of *Hadlum* v. *Hadlum*, held in 1949, is interesting! Mr Hadlum was appealing against the failure of his earlier petition for divorce on the grounds of Mrs Hadlum's claimed adultery. The sole evidence of adultery consisted of the birth of a child to Mrs Hadlum on 12 August 1945: 349 days after Mr Hadlum had left for military service abroad. The claim was essentially that the *outlier* of 349 days (compared with an average of 280 days) was *discordant*: statistically unreasonable in relation to the distribution of human gestation periods. The appeal judges agreed that the limit of credibility had to be drawn somewhere, but that on medical evidence 349, whilst improbable, was scientifically possible. The appeal failed. (Later, in 1951, in *Preston-Jones* v. *Preston-Jones*, the House of Lords drew the limit at 360 days: in *M.-T.* v. *M.-T.*, 1949, 340 days had been ruled "impossible in the light of modern gynaecological evidence"). Let us look at the distribution of gestation times (which is more than was done in any of the court cases). Fig. 1 (based on Chamberlain, 1975) shows the distribution of lengths of gestation for a sample of 13,634 births, and the outliers described above. Fig 1

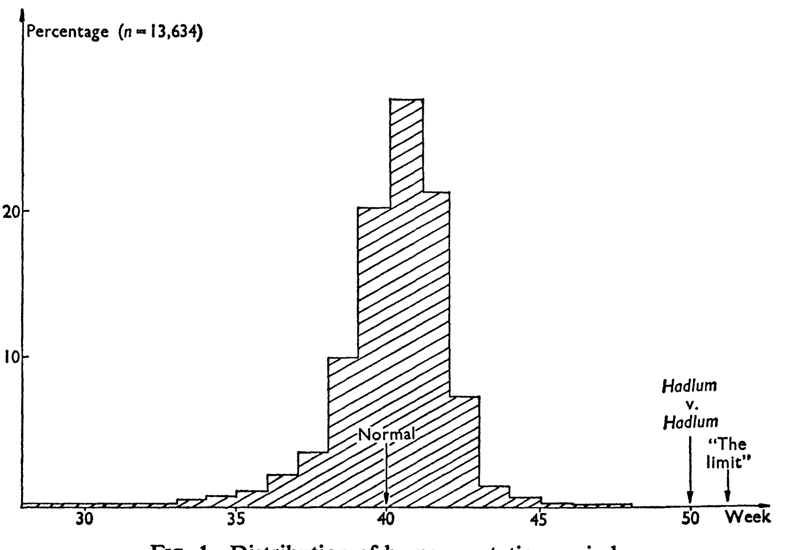

Fig. 1. Distribution of human gestation periods.

Anyone who was earlier sceptical of the importance of studying outliers must, in the light of this example, be duly chastened! Mr Hadlum's interest was clearly in *identifying* a discordant outlier as important in its own right; he surely had no concern for its *rejection* as a precursor to valid scientific study of the distribution of lengths of gestation period in the population at large. There is a message here. Statisticians have long been preoccupied with whether or not it is reasonable to reject outliers. We cannot deny that outliers exist, but their *rejection* is only one prospect among many. I shall try to sketch the alternative prospects and how they relate to the ways in which outliers can arise in data sets.

From the earliest efforts to harness and employ the information implicit in collected data there has been a concern for "unrepresentative", "rogue", "spurious", "maverick" or "outlying" observations in a data set. What should we do about the "outliers" in a sample: those observations which appear "surprisingly far away from the main group"? Should we automatically reject them, as alien contaminants, thus restoring the integrity of the data set or take no notice of them unless we have overt practical evidence that they are unrepresentative?

These two extreme viewpoints were much in evidence long ago. Bessel and Baeuer (1838) claimed never to have rejected an observation merely because of its extreme nature ("large residual"). Legendre (1805) relied on subjective judgement of experienced computors in recommending rejection of deviations "adjudged too large to be admissible". Astronomers of the nineteenth century veered to the Legendre, Boscovitch view and readily threw out dubious observations perhaps supporting their action in terms of *ad hoc* (and often dubious) rejection rules such as those of Pierce (1852), Chauvenet (1863) or Wright (1884).

Even in our present-day enlightened and sophisticated statistical climate we detect a polarization of attitude to outliers: with latter-day Bessels maintaining the objective stance of defending the integrity of the whole data mass ("who are we to cast a stone") and pragmatic neo-Legendres freely dispensing with awkward interlopers. At least the pragmatists have a vast armoury of techniques to choose from in support of their actions. There are more than 400 relevant references, and we are currently witnessing a relative avalanche of new developments in outlier techniques. Such work formalizes the basis for rejecting outliers, admits courses of action intermediate between rejection and retention by applying differential weights to the observations, implicitly distinguishes different aims in the processing of outliers and different models to explain their presence and extends technique from single univariate samples to multivariate or structured data sets.

I shall make no effort to review the mass of detailed proposals. What seems more useful is to place them in perspective by categorizing the study of outliers in response to a few fundamental interrelated questions:

(i) what are the possible causes of outliers in statistical data?
(ii) in what way do outliers influence data analysis, OR what are our different possible aims in processing outliers?
(iii) what probabilistic models might be employed to explain the presence of outliers?

In the last resort the propriety of outlier methods rests on the extent to which they suit a relevant combination of aim and outlier-generating model for a particular problem under examination. Some, but all too few, methods have been examined from this standpoint, but much more remains to be done.

2. Some Examples

That there are a variety of different possible *causes of outliers*, *ways of dealing with outliers*, and *models for outlier-generation* is readily established by considering a few simple examples.

2.1. *Barnett and Lewis (1967)*

In connection with a study of low-temperature probabilities throughout the U.K. over the winter months masses of hourly temperature readings were examined. On New Year's Eve, 1961, in Wick the following readings were noted:

43 43 41 41 41 42 43 **58** **58** 41 41
↑
midnight

The outliers **58, 58** were disturbing!

2.2. *Finney* (*1974*)

Growth of poultry: for one bird the weights (in kg) on successive weighings at regular intervals were

1·20 1·60 1·90 **1·55** 2·20 2·25.

Here, whilst not an extreme, **1·55** is an outlier in its marked pattern-disrupting form.

2.3. *A student exercise*

Ten dice were thrown ten times; one record of the numbers of sixes had the form

2, 0, 3, **12**, 2, 0, 1, 1, 3

with an outlier **12**, and apparently a missing value.

All these examples are linked by the non-statistical (*deterministic*) origin of the outlier as a gross *measurement error* or *execution error*. (In Subsection 2.1 the measurement scale changed at midnight from °F to 0·1 °C—in °F the readings become

43 43 41 41 41 42 43 42 42 39 39.

In Subsection 2.2 there was a precise 0·5 kg recording error—1·55 should be 2·05. In Subsection 2.3 a comma has been omitted!) Such an error is more or less readily detected and remedied either by rejection, replacement or correction. No statistical criteria are involved, just native wit! For examination of the outlier *per se*, no probabilistic outlier-generating model is required. Processing of such outliers is non-controversial, but sadly such tangibly promoted outliers are rare and we must consider what to do (beyond the "nothing" which some might recommend) when the origin of the outlier is more problematic, and perhaps probabilistic.

2.4. "*A normal sample*"

We are presented with a small sample as follows:

1·74 1·46 −0·28 −0·02 −0·40 0·02 **3·89** 1·35 −1·10 0·71

The value **3·89** attracts our attention as an outlier, but there is no obvious deterministic explanation. In what sense do we regard it as an outlier? As well as being extreme (and all samples must have extremes) it appears extremely extreme! If the data came from **N**(0, 1) not only is **3·89** subjectively surprising, it is also statistically unreasonable on the basis of any sensible "test of discordancy" (e.g. with test statistic $[x_{(10)} - \bar{x}]/s$, the outlier is significant beyond the 0·1 per cent level). So what do we conclude? It depends on having an alternative model to adopt—one possibility is that all but one observation came from **N**(0, 1) and one observation comes from **N**(a, 1), $a > 1$ (a *slippage-type alternative model*, see below). But the outlier may not imply anything specific about its *own* generating source, instead it may reflect on the origin of the whole sample. Although our doubts were raised by the value **3·89**—these doubts arose in relation to an assumed normal distribution as the homogeneous source for the data—it could be that in relation to some other *homogeneous* source, the sample (and in particular, the value **3·89**) does not appear strange. This is indeed the case here; the sample arose at random from a standardized Cauchy distribution.

Note the distinguishing features here:

(i) the outlier is a random manifestation, rather than a deterministic feature;

(ii) the outlier can only engender surprise relative to some (at least subjective or implicit) initial (working) model for the whole sample;

(iii) assessment of discordancy is relative to the initial model: we need an alternative model in whose favour we reject the initial model;

(iv) the alternative model may be an *inherent* one: under which the data now appear as a homogeneous sample (Cauchy rather than Normal).

2.5. *Fisher, Corbet and Williams (1943)*

A familiar example is the nocturnal Macrolepidoptera data. A random sample of numbers of individuals in different species caught in a light-trap over 4 years was:

11, 54, 5, 7, 4, 15, **560**, 18, 120, 24, 3, 51, 3, 12, 84.

The outlier **560** is here an *inherent* feature of the natural data pattern; extensive sampling shows that it is in no way anomalous (except, say, in relation to an erroneous Poisson model).

Returning to the earlier sample

1·74 1·46 −0·28 −0·02 −0·40 0·02 3·89 1·35 −1·10 0·71

we might now ask what we want to do with the data and how the outlier affects our aim. If the purpose is non-model-specific location estimation, preliminary *rejection* of the outlier following a test of discordancy is *irrelevant*. We need an estimator which *accommodates* outliers; that is, it is robust against their presence. An example is the median, or some other appropriate trimmed mean.

2.6. *Karl Pearson (1931)*

The capacities (in ml) of Moriori skulls for a sample of 17 skulls was as follows:

1230	1318	1380	1240	**1630**	1378
1348	1380	1470	1445	1360	1410
1540	1260	1364	1410	1545	

What of the outlier **1630**? It proves highly discordant on various tests. It might be prudent simply to *reject* it as unrepresentative prior to any analysis of the residual sample; or we could again accommodate the outlier using robust inference. Alternatively it could merely reflect the propriety of a positively skew underlying distribution (by no means unprecedented) as a homogeneous model to explain the whole sample (including the outlier). Then there is yet another prospect. The observation **1630** may be the very manifestation the archaeologist hoped to find, indicative of a mixing of cultures—a small number of some other race of individuals is mixed with the "pure-bred Moriori". The outlier *identifies* this prospect for us and encourages further examination of such mixing. (In this respect *rejection* or *accommodation* of the outlier becomes irrelevant).

2.7. *Daniel* (*1959*)

2^5 *experiment; 31 contrasts* in order of absolute value

0·000	0·028	−0·056	−0·084	−0·098	0·126	0·168
0·196	0·225	−0·253	0·295	−0·309	0·393	0·407
0·421	0·435	0·463	−0·477	0·547	0·660	0·744
−0·744	−0·758	−0·814	−0·814	−0·898	1·080	−1·305
2·147	**−2·666**	**−3·143**				

The three outliers are highly discordant on the null normal model. But this is what we were seeking. We *identify* the outliers as indications of features of practical importance, rather than as tedious reflections of possible inadequacies in the model or in our measuring technique.

Finally we consider a consumer purchasing example.

2.8. *Chatfield (private communication 1974)*

Purchases of packets of cereals for 13 weeks by 2000 customers yielded the frequency distribution shown in Table 1.

We have outliers at 39 and 52. How are we to react to these?

A negative binomial distribution has often been successfully employed to describe consumer purchasing behaviour. But suppose that, in relation to such a basic model, F, the outliers prove discordant. How could we account for their presence? A clue appears in the fact that we are recording the number of purchases over 13 weeks: the outliers 39 and 52 are precise multiples of 13. Is this pure coincidence? We notice relative excess frequencies at 26 and 13 also (frequencies 3 and 33). It is plausible that the data reflect two types of purchasing behaviour: casual purchases and regular purchases of 1, 2, 3 or 4 packets each week, representing a mixing of two types of purchaser. Thus the outliers suggest a more sophisticated model of the mixture type

$$(1-\lambda)F+\lambda G,$$

where λ is the small proportion of regular purchasers whose purchases are described by a distribution G.

TABLE 1

Frequency distribution for numbers of packets of cereal purchased over 13 weeks by 2000 customers

No. of packets	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Frequency	1149	119	129	87	71	43	49	46	44	24	45	22	23	33	8	2	7	2	3
No. of packets	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	
Frequency	1	2	0	0	1	0	1	3	2	0	1	1	0	0	0	0	0	0	
No. of packets	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	⩾ 54	
Frequency	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	

3. CAUSES OF OUTLIERS, METHODS OF HANDLING THEM

The examples outlined above provide illustrative responses to all three areas of enquiry:

causes of outliers; ways of handling outliers; models for outlier-generation.

Let us try to draw together some of the threads, initially under the first two headings.

Def

Following Grubbs (1969) I shall regard an *outlier* as an observation which "appears to deviate markedly from other members of the sample in which it appears". As such it is a subjective post-data manifestation. In broad terms the purpose of a statistical study of outliers is to assess whether our subjective declaration of outliers in a sample has important objective implications for an analysis of the sample. The answer must depend on all three factors previously defined: cause, manner of handling and modelling of the outliers within a sample.

3.1. *Causes of Outliers*

The examples reinforce a categorization introduced by others (e.g. Anscombe, 1960; Grubbs, 1969). Outliers may reflect either

(i) measurement errors
(ii) execution faults
or (iii) intrinsic variability.

In certain cases (i) and (ii) are readily dealt with. Gross errors of measurement or clear inclusion of obviously unrepresentative sample members can yield outliers in a sample. Their origin is deterministic—we need no statistical method to identify or interpret them. No controversy arises in weeding them out and possibly replacing or amending them. In the rare cases of such obvious contamination, rejection (or replacement) is the only answer.

More often, however, there is no simple objective origin for an outlier. It may be caused by unrecognized measurement or execution error or may merely reflect intrinsic variability in the data source. The outlier is now a random (or at least inexplicable) phenomenon, and we cannot avoid studying it *relative to some initial model for the data generation.* Even the subjective declaration of an outlier involves some primitive notion of model, otherwise how does an observation "deviate markedly" from the rest of the sample: how is it a "surprising value"? If we specify some initial model F we can proceed to examine our outlier relative to F. An outlier is an "extreme value". But we can test (using a *test of discordancy*) whether it is in probabilistic terms *too extreme* to have arisen from F and in this way exhibit it possibly as a *discordant outlier* (N.B. some authors reserve the term "outlier" for a "discordant outlier" using "suspect value" or some such term for an outlier in the present sense). If the outlier is discordant we must abandon F as a reasonable homogeneous model for the whole data set. What we do about adopting an alternative model is confounded with how we intend to react to outliers.

3.1. *Ways of Handling Outliers*

We can distinguish four basic ways of handling outliers.

(i) *Accommodation*

Here we are set on drawing inferences (perhaps about location or dispersion) irrespective of the presence of outliers. We accommodate, or protect against them, by employing robust methods: e.g. *for location*, we might employ Winsorization or trimming (that "old French custom": Huber, 1972), or sophisticated refinements such as the Huber or Bickel estimators (see Andrews *et al.*, 1972); *for dispersion*, we might use Hampel's (1974) *median deviation*

$$S_m = \text{median}\{|x_j - \tilde{x}|\}$$

or an associated quantity (where $\tilde{x}$ is the sample median).

Here there need be no specification of even an initial model F—no test of discordancy of outliers. But note that all the premium-protection study of outliers (stemming from Anscombe, 1960) is concerned with accommodation and is highly model-specific.

(ii) *Incorporation*

A discordant outlier may lead us to replace F with another *homogeneous* model G for the *whole sample* (incorporating the outlier) in relation to which no observations appear discordant (cf. Macrolepidoptera data above).

(iii) *Identification*

The discordant outlier may serve to identify *per se* an important feature of the population; (cf. purchasing data above). This may lead us to set up an alternative model, possibly of a mixture form, to explain the presence of the outliers within the larger sample. Or it may focus attention specifically on the new feature and lead to new (associated) experimentation based on an appropriate model for the population yielding the outliers.

(iv) *Rejection*

At last we come to what many seem to regard as the only prospect for discordant outliers: *viz. rejection.* In fact its role is limited to cases where the initial model is inviolable, when there may be no alternative but to discard the "contaminant" and treat the residual sample appropriately [*sic*] as coming from the model F.

4. Models for Outlier-generation

Any test of discordancy is a test of an initial (basic, null) hypothesis

$$H:F$$

which declares that the sample arises from a distribution F. A significant result implies rejecting H in favour of an alternative hypothesis $\bar{H}$ which explains the presence of a discordant outlier (for illustration we limit interest to *upper* outliers and to at most one such). *Ad hoc* test statistics such as $[x_{(n)}-x_{(n-1)}]/[x_{(n)}-x_{(1)}]$ (Dixon, 1950) can be set up with obvious intuitive appeal, and their null distribution (under H) is employed as a reference for a test of discordancy. $\bar{H}$ need not be specified—and until relatively recently little concern was given to the alternative model. But the identification aspect of outlier study requires some $\bar{H}$ to be accepted in place of H. More generally, we can only hope to compare rival tests of discordancy, and measure their power, in terms of a postulated alternative $\bar{H}$. It is vital, therefore, to consider what possibilities exist for the alternative (outlier-generating) model. Suppose we consider the case of a single *upper* outlier.

(i) *Deterministic alternative*

Suppose that x_i is *known* to be spurious. We have:

$$H\colon x_j \in F \quad (j=1,2,\ldots,n), \quad \bar{H}\colon x_j \in F \quad (j \neq i).$$

(ii) *Inherent alternative*

$$H\colon x_j \in F \quad (j=1,2,\ldots,n), \quad \bar{H}\colon x_j \in G \not\equiv F \quad (j=1,2,\ldots,n).$$

The outlier triggers rejection of the model F for the *whole* sample in favour of G for the *whole* sample (e.g. log-normal not normal). Whilst the outlier may be the stimulus for contrasting F and G, the test of discordancy will not, of course, be the only basis for comparison (Shapiro *et al.*, 1968; Shapiro and Wilk, 1972).

Between the extremes of an obvious specified contaminant, and a replacement homogeneous model, are intermediate possibilities more directed to the concept of outlying values.

(iii) *Mixture alternative*

Here we contemplate the possibility (under $\bar{H}$) that the sample contains a *small* proportion λ of observations from a distribution G. Then

$$H\colon x_j \in F \quad (j=1,2,\ldots,n), \quad \bar{H}\colon x_j \in (1-\lambda)F+\lambda G \quad (j=1,2,\ldots,n).$$

To make sense of $\bar{H}$ being reflected by outliers, G needs to have appropriate form: e.g. greater dispersion than F (with λ small enough we might encounter just one outlier) (Dixon, 1953; Tukey, 1960; Box and Tiao, 1968, using Bayesian methods).

(iv) *Slippage alternative*

This is by far the most common type of outlier model employed in the literature (Dixon, 1950; Grubbs, 1950; Anscombe, 1960; Ferguson, 1961; McMillan and David, 1971; Guttman, 1973).

In its general form we have

$$H\colon x_j \in F \quad (j=1,2,\ldots,n), \quad \bar{H}\colon \begin{cases} x_j \in F & (j \neq i), \\ x_i \in G, \end{cases} \tag{*}$$

where, again, G is such that $\bar{H}$ is likely to be reflected by *outliers*.

Examples involving slippage of location, or of dispersion, respectively, arise where F has mean μ and variance σ^2 and G has the same form as F but mean $\mu+a\ (a>0)$ or variance $b\sigma^2\ (b>1)$. Thus we have

$$F \sim (\mu, \sigma^2)$$

and G takes the form

$$G_1 \sim (\mu+a, \sigma^2) \quad (a>0) \quad \text{or} \quad G_2 \sim (\mu, b\sigma^2) \quad (b>1),$$

respectively.

An interesting dilemma arises here. $\bar{H}$ declares merely that *some* observation x_i comes from G. Yet we distinguish an *upper* outlier as the extreme $x_{(n)}$—a *particular* observation. Should we not, then, take

$$H\colon x_j \in F \quad (j=1,2,\ldots,n), \quad \bar{H}\colon \begin{cases} x_{(j)} \in F & (j=1,2,\ldots,n-1), \\ x_{(n)} \in G\ ? \end{cases} \tag{†}$$

We could term this a *labelled* slippage alternative model. In fact the slippage model for outliers is always used in unlabelled form as (*). In two cases at least (mean slippage normal, scale slippage gamma) the use of (†) and (*) lead to identical *maximum likelihood ratio tests* of discordancy, but this correspondence may not always hold. See Barnett and Lewis (1978).

(v) *Exchangeable alternative*

Work by Kale, Sinha, Veale and others (see, for example, Kale and Sinha, 1971) extends the unlabelled slippage-type model (*) by further assuming that the index i of the discordant value is equally likely to be $(1,2,\ldots,n)$.

$$H\colon x_j \in F \quad (j=1,2,\ldots,n) \quad \bar{H}\colon \begin{cases} x_j \in F & (j \neq i), \\ x_i \in G, \\ p(i=j) = n^{-1} & (j=1,2,\ldots,n) \end{cases}$$

$X_1 \ldots X_n$ are no longer independent, but they are exchangeable.

This is half-way towards the alternative models adopted in some Bayesian approaches (e.g. Guttman, 1973).

Related to model formulation are the ideas of Neyman and Scott (1971) and of Green (1974) on *outlier proneness*. Defining $P(\kappa, n \mid F) = p[x_{(n)} - x_{(n-1)} > \kappa(x_{(n-1)} - x_{(1)}) \mid F]$ where $F \in \mathscr{F}$ (some family of distributions), $\mathscr{F}$ is *(κ, n)-outlier-resistant* if $\sup_{\mathscr{F}} P(\kappa, n \mid F) < 1$. Otherwise $\mathscr{F}$ is *(κ, n)-outlier-prone*. In a rather nebulous way this distinguishes families of distributions in terms of the extent to which outliers will be manifest, but no statistical applications have yet appeared.

5. Conclusion

The study of methods of handling outliers continues to interest statisticians. It seems crucial that we clearly distinguish and categorize the different ways of handling and modelling outliers if such study is to lead to fruitful further developments.

References

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J. Rogers, W. H. and Tukey, J. W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton, N.J.: Princeton University Press.

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

Barnett, V. D. and Lewis, T. (1967). A study of low-temperature probabilities in the context of an industrial problem (with Discussion). *J. R. Statist. Soc.* A, **130**, 177–206.

—— (1978). *Outliers in Statistical Data*. Chichester: Wiley.

Bessel, F. W. and Baeuer, J. J. (1838). *Gradmessung in Ostpreussen und ihre Verbindung mit Preussischen und Russischen Dreiecksketten*. Berlin. Reprinted in *Abhendlungen von F. W. Bessel* (R. Englemen, ed.), Leipzig, 1876.

Box, G. E. P. and Tiao, G. C. (1968). A Bayesian approach to some outlier problems. *Biometrika*, **55**, 119–129.

Chamberlain, R. (1975). In *British Births 1970. Vol. 1: The First Week of Life*. Heinemann, London.

Chauvenet, W. (1863). Method of least squares. Appendix to *Manual of Spherical and Practical Astronomy*, **2**, Lippincott, Philadelphia, pp. 469–566.

Daniel, C. (1959). Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics*, **1**, 311–341.

Dixon, W. J. (1950). Analysis of extreme values. *Ann. Math. Statist.*, **21**, 488–506.

—— (1953). Processing data for outliers. *Biometrics*, **9**, 74–89.

Ferguson, T. S. (1961). On the rejection of outliers. *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, **1**, 253–287.

Finney, D. J. (1974). Problems, data and inference: The Address of the President (with Proceedings). *J. R. Statist. Soc.* A, **137**, 1–23.

Fisher, R. A., Corbet, A. S. and Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Animal Ecol.*, **12**, 42–57.

Green, R. F. (1974). A note on outlier-prone families of distributions. *Ann. Statist.*, **2**, 1293–1295.

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

—— (1969). Procedures for detecting outlying observations in samples. *Technometrics*, **11**, 1–21.

Guttman, I. (1973). Care and handling of univariate or multivariate outliers in detecting spuriosity—a Bayesian approach. *Technometrics*, **15**, 723–738.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Ass.*, **69**, 383–393.

Huber, P. J. (1972). Robust statistics: a review (The 1972 Wald Lecture). *Ann. Math. Statist.*, **43**, 1041–1067.

Kale, B. K. and Sinha, S. K. (1971). Estimation of expected life in the presence of an outlier observation. *Technometrics*, **13**, 755–759.

Legendre, A. M. (1805). *Nouvelles Méthodes pour la Determination des Orbites des Comètes*. Courcier, Paris (especially "Appendice sur la méthode des moindres quarrés", pp. 72–80).

McMillan, R. G. and David, H. A. (1971). Tests for one of two outliers in normal samples with known variance. *Technometrics*, **13**, 75–85.

Neyman, J. and Scott, E. L. (1971). Outlier proneness of phenomena and of related distributions. In *Optimising Methods in Statistics* (J. Rustagi, ed.). New York: Academic Press.

Pearson, K. (1931). *Tables for Statisticians and Biometricians*. Biometric Lab., University College, London.

Pierce, B. (1852). Criterion for the rejection of doubtful observations. *Astr. J.*, **2**, 161–163.

Shapiro, S. S. and Wilk, M. B. (1972). An analysis of variance test for the exponential distribution (complete samples). *Technometrics*, **14**, 355–370.

Shapiro, S. S., Wilk, M. B. and Chen, M. J. (1968). A comparative study of variance tests for normality. *J. Amer. Statist. Ass.*, **63**, 1343–1372.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics* (I. Olkin, ed.). Stanford, Calif.: University Press.

Wright, T. W. (1884). *A Treatise on the Adjustment of Observations by the Method of Least Squares*. New York: Van Nostrand.

Citations from the official law reports on the decisions of the courts of England and Wales.

Hadlum v. *Hadlum*	1949	*Probate*	197
Preston-Jones v. *Preston-Jones*	1951	*Appeals*	391–420
M.-T. v. *M.-T.* and the Official Solicitor	1949	*Probate*	331–335