REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/1266526?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

Some Remarks on Wild Observations*

WILLIAM H. KRUSKAL**

The University of Chicago

Editor's Note: At the 1959 meetings of the American Statistical Association held
in Washington D. C., Messrs. F. J. Anscombe and C. Daniel presented papers on the
detection and rejection of 'outliers', that is, observations thought to be maverick or
unusual. These papers and their discussion will appear in the next issue of *Techno-
metrics*. The following comments of Dr. Kruskal are another indication of the present
interest of statisticians in this important problem.

The purpose of these remarks is to set down some non-technical thoughts
on apparently wild or outlying observations. These thoughts are by no means
novel, but do not seem to have been gathered in one convenient place.

1. Whatever use is or is not made of apparently wild observations in a statisti-
cal analysis, it is very important to say something about such observations
in any but the most summary report. At least a statement of how many observa-
tions were excluded from the formal analysis, and why, should be given. It is
much better to state their values and to do alternative analyses using all or
some of them.

2. However, it is a dangerous oversimplification to discuss apparently wild
observations in terms of inclusion in, or exclusion from, a more or less conven-
tional formal analysis. An apparently wild (or otherwise anomalous) observation
is a signal that says: "Here is something from which we may learn a lesson,
perhaps of a kind not anticipated beforehand, and perhaps more important
than the main object of the study." Examples of such serendipity have been
frequently discussed—one of the most popular is Fleming's recognition of the
virtue of penicillium.

3. Suppose that an apparently wild observation is *really known* to have come
from an anomalous (and perhaps infrequent) causal pattern. Should we include
or exclude it in our formal statistics? Should we perhaps change the structure
of our formal statistics?

Much depends on what we are after and the nature of our material. For
example, suppose that the observations are five determinations of the percent
of chemical A in a mixture, and that one of the observations is badly out of

1

line. A check of equipment shows that the out of line observation stemmed
from an equipment miscalibration that was present only for the one observation.

   If the magnitude of the miscalibration is known, we can probably correct
for it; but suppose it is not known? If the goal of the experiment is only that
of estimating the per cent of A in the mixture, it would be very natural simply
to omit the wild observation. If the goal of the experiment is mainly, or even
partly, that of investigating the *method* of measuring the per cent of A (say in
anticipation of setting up a routine procedure to be based on one measurement
per batch), then it may be very important to keep the wild observation in.
Clearly, in this latter instance, the wild observation tells us something about
the frequency and magnitude of serious errors in the method. The kind of
lesson mentioned in 2 above often refers to methods of sampling, measurement,
and data reduction, instead of to the underlying physical phenomenon.

   The mode of formal analysis, with a known anomalous observation kept in,
should often be different from a traditional means-and-standard deviations
analysis, and it might well be divided into several parts. In the above very simple
example, we might come out with at least two summaries: (1) the mean of the
four good observations, perhaps with a $\pm$ attached, as an estimate of the per
cent of A in the particular batch of mixture at hand, and (2) a statement that
serious calibration shifts are not unlikely and should be investigated further.
In other situations, nonparametric methods might be useful. In still others,
analyses that suppose the observations come from a mixture of two populations
may be appropriate.

   The sort of distinction mentioned above has arisen in connection with military
equipment. Suppose that 50 bombs are dropped at a target, that a few go wildly
astray, that the fins of these wild bombs are observed to have come loose in
flight, and that their wildness is unquestionably the result of loose fins. If we
are concerned with the accuracy of the whole bombing system, we certainly
should not forget these wild bombs. But if our interest is in the accuracy of the
bombsight, the wild bombs are irrelevant.

   4. It may be useful to classify different degrees of knowledge about an ap-
parently wild observation in the following way:

   a. We may know, even *before* an observation, that it is likely to be wild,
or at any rate that it will be the consequence of a variant causal pattern. For
example, we may see the bomb's fins tear loose before it has fallen very far from
the plane. Or we may know that a delicate measuring instrument has been jarred
during its use.

   b. We may be able to know, *after* an observation is observed to be apparently
outlying, that it was the result of a variant causal pattern. For example, we
may check a laboratory notebook and see that some procedure was poorly
carried out, or we may ask the bombardier whether he remembers a particular
bomb's wobbling badly in flight. The great danger here, of course, is that it is
easy after the fact to bias one's memory or approach, knowing that the observa-
tion seemed wild. In complex measurement situations we may often find some-
thing a bit out of line for almost any observation.

   c. There may be *no evidence* of a variant causal pattern aside from the observa-

tions themselves. This is perhaps the most difficult case, and the one that has given rise to various rules of thumb for rejecting observations.

Like most empirical classifications, this one is not perfectly sharp. Some cases, for example, may lie between b and c. Nevertheless, I feel that it is a useful trichotomy.

5. In case c above, I know of no satisfactory approaches. The classical approach is to create a test statistic, chosen so as to be sensitive to the kind of wildness envisaged, to generate its distribution under some sort of hypothesis of non-wildness, and then to 'reject' (or treat differently) an observation if the test statistic for it comes out improbably large under the hypothesis of nonwildness. A more detailed approach that has sometimes been used is to suppose that wildness is a consequence of some definite kind of statistical structure—usually a mixture of normal distributions—and to try to find a mode of analysis well articulated with this structure.

My own practice in this sort of situation is to carry out an analysis both with and without the suspect observations. If the broad conclusions of the two analyses are quite different, I should view any conclusions from the experiment with very great caution.

6. The following references form a selected brief list that can, I hope, lead the interested reader to most of the relevant literature.

REFERENCES

1. C. I. BLISS, W. G. COCHRAN, AND J. W. TUKEY, "A rejection criterion based upon the range," *Biometrika*, 43 (1956), 418–22.
2. W. J. DIXON, "Analysis of extreme values," *Ann. Math. Stat.*, 21 (1950), 488–506.
3. W. J. DIXON, "Processing data for outliers," *Biometrics*, 9 (1953), 74–89.
4. FRANK E. GRUBBS, "Sample criteria for testing outlying observations," *Ann. Math. Stat.*, 21 (1950), 27–58.
5. E. P. KING, "On some procedures for the rejection of suspected data," *Jour. Amer. Stat. Assoc.*, 48 (1953), 531–3.
6. JULIUS LIEBLEIN, "Properties of certain statistics involving the closest pair in a sample of three observations," *Jour. of Research of the Nat. Bureau of Standards*, 48 (1952), 255–68.
7. E. S. PEARSON AND C. CHANDRA SEKAR, "The efficiency of statistical tools and a criterion for the rejection of outlying observations," *Biometrika*, 28 (1936), 308–320.
8. PAUL R. RIDER, "Criteria for rejection of observations," *Washington University Studies, New Series. Science and Technology*, 8 (1933).