TOWARDS ROBUST ANALYSIS OF VARIANCE

ALLAN H. SEHEULT & JOHN W. TUKEY
University of Durham Princeton University

SUMMARY

A distinctive feature of analysis of variance is the common occurrence of more than one error term. This feature calls attention to the two distinct potential roles of a single mean square. As a "numerator" it measures the variability visible at a given level in a design hierarchy, and as a "denominator" it measures how much variability has been "passed up" to higher levels, and may, if appropriate, serve as part of an error term. We propose a straightforward multiphase procedure that explicitly recognizes these two roles, and argue that, in general, such considerations preclude naive use of robust regression techniques for analysis of factorially designed experiments. Instead, an upsweeping-by-medians decomposition of the data is followed by a comparison-within-subtable analysis to flag *exotic* ("relatively large") entries in each of the subtables associated with the different sorts of variation. A classical analysis by means, after replacing each identified exotic entry by an algorithmically specified value, yields a decomposition of the data that is used to construct an analysis of variance table in which for each sort of variation there is both a list of any exotic entries and an *inner* ("denominator") mean square that 'excludes' those exotic entries. The analysis can then be completed by downsweeping the inner subtables that are insufficiently prominent, and providing (formally) appropriate error terms for analyzing table entries that remain. The results are displayed as a decomposition of the data into exotic values and those inner subtables, both simple and composite, that survive downsweeping. The exploratory nature of the approach is emphasized, and the method is applied to an example of a factorial experiment in which all factors have three or more versions.

KEY WORDS: ANOVA; Downsweeping; Exotic values; Factorial experiments; Fibian; Half-Winsorizing; Inner mean squares; Mean polish; Median polish: Middle-Median; Midmedian; Robust ANOVA; Upsweeping; Winsorizing.

1 INTRODUCTION

Today's challenge to a "robust" analysis of variance procedure is not to optimize the process of analysis, but rather to be organized in meeting a variety of complications, both individually or in as many combinations as we can handle. We describe below a multistep process that does this better than any other proposed process that we have seen.

As with other procedures for analyzing substantial bodies of data, we have to expect a frequent need to combine a formalized scheme of analysis with judgment-based adaptations to the combination of a particular dataset and a particular purpose of analysis. Part of the expected evolution of "current best procedure" will be the incorporation, into the formalized scheme of analysis, of techniques for dealing with an increasing variety of "unusual" behaviors.

In 1950, the formalized schemes dealing with analysis of variance were relatively simple, so that the demands on the analyst were diverse and frequently important. In 2000, the sort of formalized scheme illustrated below will handle a much greater diversity of complications—but not all—without reference to the judgment of the analyst.

The operation of the formalized scheme of analysis is illustrated by the application of its steps to a frequently analyzed $3 \times 5 \times 8$ dataset. In this dataset, values for 19 of 120 cells, 3 of 79 (one-dimensional) fibers, and 3 of 15 (2-dimensional) sheets of values appear exotic. (Of the 7 subtables of different kinds of effects and interactions, 5 contain instances of corresponding exoticity.) Clearly, this dataset calls for an analysis that is prepared to deal with a diverse collection of kinds of exoticity.

While the example does not show all the idiosyncrasies that deserve attention, it does show a reasonable diversity of misbehavior.

While our process is designed to restrict the impact of exotic valued fibers, sheets, etc., on the fit to non-exotic values, when, as in this dental gold experiment, we know the purpose for which the data was gathered, we have an obligation to ask whether there is any way to "clean" the analysis still further by omitting some of the initial data. In the present example we can do this by omitting one or more versions of two of the three factors, a process not yet formalized, but a definite obligation.

In this paper we offer a recipe for robust/resistant analysis of variance of data from factorial experiments in which all factors have three or more versions. The details presented are subject to further evolutionary changes, as we learn more. However, we believe the overall structure, concepts and attitudes to be important, useful, and likely to continue. Special modifications will be required for certain, other types of factorially designed experiment, including those in which some factors have only two versions, a situation we plan to treat in a separate account: however, see Seheult and Tukey (1982) and Seheult (1997) for robust treatments of (pure) 2^n experiments.

Analysis of variance continues to be one of the most widely used statistical methods. Not only the form of an analysis of variance table with its lines of mean squares and degrees-of-freedom associated with each of several sorts of variation, but the entire analysis, including confidence statements, is classically supposed to be determined by the design—the hierarchical structure, conduct, and the intent of the experiment—alone. The behavior of the data itself is, classically, not supposed to influence how its description is formatted. Hardly an exploratory attitude, and one for which the book *Fundamentals of Exploratory Analysis of Variance* (*FEAV*) has provided a meaningful alternative within the classical framework, where summarization by means is a basic. As Mallows and Tukey (1982) remark: "In our experience most analysis of variance is a matter of exploratory data analysis, rather than one of critical confirmatory analysis". In this account, rather than using a data-free structure to define our procedures, we provide a further stage of responding to the data's behavior, one where summarization is based on a robust alternative to the mean.

Given a format, to which we have been led by the data, we may well want to return to the data and present a careful analysis, at least in balanced cases. For classical analyses, such final analysis can be conducted by regression methods. Thus, it is tempting to suppose that at least the final analysis—and even, some optimists would dream, the process of feeling around for the right format—could be conducted by naive use of a conventional robust-regression technique, which allows for a certain number of individual points to misbehave individually. Such a tempting view—of only simple misbehavior—is far from being usually applicable. Properly designed robust techniques should (i) bring to our attention unusual behavior; (ii) prevent such behavior from disturbing our understanding and appreciation of the general, usual behavior; and (iii) prevent us from disturbing our summaries and conclusions by responding conventionally to unusual behavior not confined only to individual data points. As examples of designed experiments so frequently illustrate, unusual behavior is often exhibited by larger structured subsets of data points.

Such issues are illustrated in the example of a two-factor experiment with equal replication in each cell (factor combination), for which the analysis of variance can be derived from the results of sweeping out row and column means from the two-way table of replicate means. It follows that we need to be robust at least at two levels or lines of the analysis: at the replicate level, to obtain 'clean' estimates of cell effects and replicate error, and at the interaction level, where substantial interaction is often concentrated in one or a few cells, or in one or a few rows or columns. Analysis by means 'smears' such concentrated interactions both over the other cells and into the main effects. In some instances such interactions may be of genuine scientific interest, in others they may be dismissible as "bias" or "model failure", but in either case their identification will be important; see Brown (1975) and Daniel (1976, 1978) for careful classical treatments of examples of factorial data exhibiting these and other types of nonstandard behavior.

As with least squares, a standard robust regression analysis of this problem, which treats main effects and interactions as an undifferentiated whole, to be fitted all at once, will not usually isolate and sterilize such concentrated interactions. However, a simple procedure such as median polish (Tukey, 1977) of the replicate medians will usually do what is needed at the interaction level and for main effects and, if there are at least three replications per cell, at the replication level as well; see, also, Besag and Seheult (1988). A robust regression procedure applied to cell medians would then have some chance of providing a useful analysis at the upper level, provided that both the number of rows and the number of columns is not too small. Oehlert (1994) suggests a scheme in which prominent one-cell interactions are identified as 'outliers' from a previously identified simple model, and Besag (1981) gives a resistant decomposition of data from a latin square experiment as a preliminary to a final classical analysis.

An important related issue arises when, for example, one of the two factors is treated as random. In a classical analysis the interaction mean square is the appropriate error term for making comparisons between summaries for versions of the *other* factor. It is often important that such concentrated or *exotic* interactions do not contribute to these summaries and hence should not contribute to the corresponding error term.

In experiments involving more factors, where we will regard the data as a many-way array, recognition of "cells" (containing single or duplicate observations), or of "fibers" (lines of cells), or of "sheets" (two-way arrays of cells) or of "blocks" (many-way arrays of cells) as exotic raises the question of how our observations should be summarized. Should we include exotic values in our basic (overall) summaries? On balance, we believe the answer is much more often "no" than "yes", especially when exoticity is so profound as to make the non-robust summary silly or when the exotic values are "clearly something else" than the bulk of the observations. In those infrequent instances

where classical summaries are reasonable and relevant, we would use the classical analysis of variance table. Otherwise, we 'exclude' exotic values from basic summaries and, once we do this, we also exclude them from the error terms used to assess these basic summaries. However, we need to take these exotic values seriously, and would want to assess their variability separately.

These considerations emphasize that in classical analysis a single mean square can have two roles: as a "numerator" it indicates how much variability is there at the given level, and as a "denominator", or error term, it indicates how much variability has been "passed up" to "higher lines" in the design hierarchy. In a robust analysis, we choose to emphasize this distinction explicitly, having at each line of an analysis of variance table (i) a list of any exotic entries; and (ii) a denominator or *inner mean* square that 'excludes' the contributions of exotic entries and indicates how much variability has been passed up to non-exotic aspects of higher levels by whatever robust summaries have been used. Our recipe, for the present at least, is as follows:

Pre-decomposition. In Section 2, we apply a simple robust summary, a type of median, through the same pattern of sweep operations used in the corresponding classical analysis by means in order to determine an additive pre-decomposition or breakdown of data into subtables of effects focused on different sorts of variation. The primary purpose of this "median polish" breakdown is to aid in identification of "exotic" values.

Identification. Each subtable of the pre-decomposition is examined separately to flag exotic entries; that is, entries which are large relative to the overall variability of the entries in the subtable. A simple procedure for doing this is described in Section 3.

Re-decomposition. At the next stage, each identified exotic entry in each subtable is first replaced by a value related in size to the non-exotic entries in its subtable. The resulting 'decomposition' is then re-polished by means to produce an *inner decomposition* undisturbed by identified exotic values. Different possibilities for replacement values are discussed in Section 4.

Robust analysis of variance. For any line whose subtable includes exotic entries, there will now be (i) an *inner* mean square proportional to the sum of squares of the entries in the corresponding subtable of the inner decomposition, (ii) a degrees-of-freedom and (iii) this list of exotic entries. Details of proportionality and calculation are discussed in Section 5.

Downsweeping. Downsweeping, a process which involves pooling of mean squares, is applied to the inner mean squares following the pattern for classical analysis of variance described in *FEAV*, Chapter 11. There, when a subtable (or equivalently an "overlay") is downswept onto a candidate subtable, namely one involving "one more factor", corresponding entries from the equivalent overlays are simply added together, and the identity of the downswept subtable is lost. The identity of a subtable is preserved if it is not downswept. Overall, the downsweeping process is intended to concentrate the data for further analyses, not to choose a prescribed framework of final analysis. Details of when and where to downsweep and subsequent pooling of mean squares are treated in Section 6.

Evaluation. Finally, error mean squares are calculated for each subtable, whether simple or complex, that survives the downsweeping process, The details and how these error mean squares might be converted into standard errors and confidence statements are discussed briefly in Section 7.

This recipe may seem unduly complicated, but the steps parallel those used in the non-robust case, as described in *FEAV*. Basford and Tukey (1995) apply some of the above ingredients to a randomized block experiment. Paying even more attention to details of a decomposition can be worthwhile and is addressed by Johnson and Tukey (1987) for the non-robust case and by Johnson (1988) and also Tukey (1993).

2 PRE-DECOMPOSITION

Before considering details of decomposition, we introduce an example to illustrate the ideas in this and other sections of the paper.

Dental Gold Data

Brown (1975) analyzed a three-factor experiment described in Xhonga (1971). The response variable was a measure of hardness of gold alloy fillings obtained by five dentists (D), each using three methods of condensation (C), with each of eight types of gold alloy (G). The data, reproduced from Brown's paper in Table 1, will be analyzed as a three-factor experiment for which the standard analysis of variance is given in Table 2. A full description of the background to these data can be found in Xhonga (1971) and Chapter 11 of *FEAV*. They will be used here to illustrate both the classical decomposition of factorial data into main effects and interactions using means and the analogous robust decomposition using medians.

Mean Polish

Table 3 displays the mean or least squares decomposition of the data in Table 1 into eight ($= 2^3$) subtables associated with the corresponding lines in the analysis of variance in Table 2: an overall effect (1), three main effects D, C and G, three two-factor interactions DC, DG and CG, and a three-factor interaction DCG. All the entries here and hereafter have been rounded to the nearest integer.

Each datum in Table 1 is the appropriate sum of eight entries, one from each of the eight subtables; for example, the first data entry 792 is decomposed as

$$792 = 737 + 48 + 49 + (-9) + (-16) + 6 + (-54) + 30$$

which, in an obvious notation, corresponds to

$$y_{111} = (1) + d_1 + c_1 + g_1 + (dc)_{11} + (dg)_{11} + (cg)_{11} + (dcg)_{111}$$

In classical analysis of variance, the mean square for an interaction or main effect can be obtained as the sum of squares of the entries in its corresponding subtable, times the number of data values contributing to each entry, and divided by the associated degrees of freedom. Thus, for example, the mean square for DC is (conventionally) computed as

$$\frac{8 \times [(-16)^2 + \cdots + (-88)^2]}{(5-1) \times (3-1)} = 32930$$

A convenient way of obtaining the subtables in Table 3 is as follows. First, expand the $5 \times 3 \times 8$ data array into a $6 \times 4 \times 9$ array with *zero* entries in each of the three additional last planes. Next, sweep out in the D direction the 4×9 array of means of the *first* five planes and add them to the sixth remaining 4×9 plane of the $6 \times 4 \times 9$ array; and repeat a corresponding operation in each of the C and G directions. The decomposition given in Table 3 can be obtained from the resulting $6 \times 4 \times 9$ array as follows: the original data array has been replaced by the three-way table for the three-factor interaction, and attached to it are seven bordering tables—a two-way table for each of the three two-factor interactions, a one-way table for each of the three main effects and a zero-way table for the overall effect. This process, which we will refer to as *mean polish*, can be applied

to any many-way complete factorial data array, and can be extended to data from most balanced experimental designs, with or without pure replication. Note that, for each proper subtable the *mean* of any *fiber*—of any complete line of cell entries, holding all but one of the subtable's factors fixed—is zero.

Median Polish

To obtain a *robust/resistant* decomposition, we choose to use a *median* instead of a mean in the above cycle and iterate until 'convergence'. A median is the "middle" value for an odd number of values, otherwise, it can be any value between the two central values. The lower of the two central values is called the *lomedian*, the higher is called the *himedian*, and their average is called the *midmedian*. At every stage leading towards our median based decomposition, we choose to summarise a fiber by its *fibian* which is one these four medians. The fibian is the (unique) median for an odd number of values in a fiber. For an even number, it is either the lomedian or himedian, whichever, when swept out from the fiber, minimizes the size of the resulting swept-into value; and hence could be the midmedian if these sizes are the same. It is usually convenient to label it as the *roof* of the midmedian when the midmedian is a half-integer.

The fibian decomposition of the gold fillings data, using the "long-fibers-first" sweep order GDC, is displayed in Table 4.

What can we learn from the fibian decomposition in Table 4 that perhaps is not so readily apparent from the mean decomposition in Table 3 ? We notice, for example, that the two emboldened entries **-146** at $(4, 3, \cdot)$ and **-208** at $(5, 3, \cdot)$ in the DC subtable are large relative to the other entries there, and are designated as exotic according to the criterion set forth in the next section. Brown (1975) discovered this appearance (that dentists 4 and 5 had difficulty with condensation method 3) by first noting the *significance* of the DC mean square relative to the DCG mean square and then examining the DC subtable from the mean decomposition to ascertain the cause of this significant interaction.

Examination of the CG subtable reveals the distinctive nature of the entry **-172** at $(\cdot, 3, 8)$. Similarly, the fifth dentist and the hand condensation method ($C3$) produce distinctively low levels of hardness, whereas the sixth gold appears to produce distinctively high levels of hardness. The 19 exotic entries in the DCG table are mostly associated with dentists 4 and 5 and hand condensation, suggesting perhaps that different levels of variability are associated with the different levels of hardness in different parts of the experiment. It might have been anticipated that hand condensation would produce more variable results, but not necessarily overall low levels of hardness.

At this point readers may have the impression that they are being asked to examine carefully $6 \times 4 \times 9 = 216$ numbers in the place of the $5 \times 3 \times 8 = 120$ data values! While this is likely to be worthwhile (Johnson and Tukey, 1987 and Johnson, 1988), this paper proposes only less striking modifications. In any analysis of variance, robust or otherwise, careful examination of the details of any subtable that appears to exhibit unusual behavior will be worthwhile.

Remarks on fibian decompositions. Clearly, the choice of fibian is somewhat arbitrary but it has certain advantages over other simple robust/resistant measures of location. Like the median, its computation is simple and does not require an internal assessment of scale. Moreover, if the original data are integers, which in practice we can nearly always arrange to be the case, then all the entries of the fibian decomposition will also be integers. If so, convergence of fibian polish in a finite number of iterations is assured. Note that, in a final fibian decomposition at least one of the lomedian, midmedian, roof of midmedian, or the himedian of every fiber in each proper subtable

will be zero. Thus, there is either at least one zero in a fiber, or the two smallest values, one of each sign, sum to 0 or -1. Unlike mean polish, but like all other median-based polishes, the result of a fibian polish will usually depend on the order of sweep directions. We choose to polish in declining order of number of versions of the factors.

These somewhat arbitrary aspects of fibian polish do not appear to be important in the present context, where the main aim is to isolate genuine substantial entries in each subtable. In our experience with real data, this aim is mostly achieved without real difficulty. These remarks also apply to the lomedian and himedian: the fibian just chooses between them according to a sensible criterion.

3 IDENTIFYING EXOTIC ENTRIES

Inevitably, any procedure for flagging exotic entries in a subtable will be somewhat arbitrary. Thus, when Brown (1975) analyzed the dental gold data and noted the significant DC interaction mean square, he redid the analysis setting the largest sized entry in the DC subtable as "missing." And, on finding that the new DC mean square was still significant, he continued the procedure until nonsignificance: in this instance, a total of two passes were required. By contrast, in our procedure each subtable is inspected, and a decision to flag an entry in any particular subtable is determined by the entries in *that* subtable, rather than by reference to that table's "error term." This is appropriate, since we are seeking to identify *exoticity* not *significance*. A very well-behaved interaction with no trace of exoticity can quite easily be highly significant.

A simple flagging procedure, the one we use, runs as follows. Let ν denote the conventional degrees-of-freedom assigned to a particular subtable of a decomposition, and denote by $0 \le z_\nu \le z_{\nu-1} \le \cdots \le z_1$ the ordered ν largest sizes of entries in the subtable. If none of the entries in the subtable are exotic the $z's$ will tend to resemble an ordered sample of size ν from a half-Gaussian distribution with unknown standard deviation σ. Let c_i denote the median of the sampling distribution of the i-th order statistic from a unit half-Gaussian distribution and put $s_i = z_i/c_i$ for $i = 1, \ldots, \nu$. Then, s, the *middle-median*[1] of $s_1, \ldots, s_\nu$, is a simple robust/resistant estimate of σ, a measure of the overall variability in the subtable that is resistant to large ratios for small subtable entries as well as those for large entries. Next, for each $i = 1, \ldots, \nu$, and for some positive "cutoff" value K, we flag the entry corresponding to z_i if $s_j > Ks$ for *all* $j \le i$. Thus, flagged entries in a subtable must form an uninterrupted top-down sequence. This procedure is similar to the use of half-normal plots for informally assessing controlling factors and interactions in 2^n factorial experiments; see Daniel (1959) and Johnson and Tukey (1987).

We now indicate how to calculate the c_i and choose a value for K. An adequate (very good) approximation to c_i, which we sometimes refer to as the i-th *working value*, is found as the solution to

$$2\Phi(c_i) - 1 = \frac{\nu - i + 1}{\nu + \frac{2}{3}}$$

where $\Phi(\cdot)$ denotes the unit Gaussian cumulative distribution function. This working value formula makes $c_{\nu+1} = 0$.

Unpublished simulation studies for two-way tables by Edward Fowlkes (Bell Communications Research, Inc.), Jean McRae (AT&T Bell Laboratories) and the present authors indicated that a value of about 1.5 for K is a reasonable general choice which comes close to optimizing the lesser of two efficiencies—one for Gaussian perturbations, the other for the very stretched-tail perturbations

[1]The middle-median of a set of s's is the median of $s_{q+1}, \ldots, s_{\nu-q}$, where q is the floor of $(\nu+1)/4$.

generated from the slash distribution—the ratio of a unit Gaussian random variable to an independent unit uniform random variable. Here, we will use a single cutoff value of 1.5 for any subtable, regardless of its dimensions.

We need to protect ourselves from problems with excessive numbers of zeros, especially for small subtables. This is the case with the *DG* subtable from the fibian decomposition in Table 4, where there are 25 non-zero entries, three less than the nominal degrees-of-freedom. Our present recommendation is to replace ν by the lesser of ν and "one plus the number of nonzero entries"; for example, $\nu = 28$ for the DG subtable is replaced by 26, the lesser of 28 and (25+1). Notice that the modification implies that if all-but-one of $z_1,\ z_2\ , \cdots,\ z_\nu$ are zero, the single nonzero value is treated as exotic.

Sometimes there can be more than ν nonzero entries; for example, DCG has 68 rather than 56. In such cases, our present recommendation is to subtract $z_{\nu+1}$ from each of $z_1\ , \ldots,\ z_\nu$, and then apply the exotic identification procedure to the ν differences $z_1 - z_{\nu+1}\ , \ldots,\ z_\nu - z_{\nu+1}$.

Table 12 shows the appropriate calculations for the *DC* subtable given in Table 4. Here, there are exactly $\nu = 8$ non-zero entries, and, according to the criterion above, the entries in *DC* cells $(5, 3, \cdot)$ and $(4, 3, \cdot)$ are identified as exotic. Table 4 displays all such exotic entries in a large bold font.

4 EXOTIC REPLACEMENT AND RE-DECOMPOSITION

Having isolated any exotic entries in each subtable, we now consider how they should be treated and analyzed in conjunction with non-exotic subtable entries. The key idea is to first replace an exotic entry by an exotic replacement value 'similar in size' to the non-exotic entries in its subtable, and then mean polish the resulting decomposition, adjusting the exotic values to retain an overall additive decomposition. This type of analysis lets us take advantage of the flexibility and power of a mean analysis while ensuring that exotic values have no chance to take advantage of the general failure of mean analysis to be robust.

The simplest replacement strategy is to replace each exotic value by *zero*. An alternative is to Winsorize each exotic value by replacing it with the nearest non-exotic value (sign considered) in the same subtable. We have investigated these and other replacement strategies, but on balance, at least in the present example and for the remainder of this paper, we prefer the compromise "half-Winsorization" strategy whereby each exotic entry is replaced by an amount equal to one-half of its Winsorized value.

Note that the "zero" strategy is appropriate when an exotic value is 'something other than the rest', whereas "Winsorization" treats each exotic value as if it were generated from a stretched-tail distribution, as when a particular contribution to error is usually zero but sometimes large. "Half-Winsorization", the presently preferred compromise between "zeroing" and "Winsorization", is something to use when we do not understand the nature of the exotics in the data before us, and seems a reasonable default if we can indulge ourselves with the luxury of a single answer. However, it will often be desirable to compare the results with those from "zeroing" and from "Winsorization".

Table 5 shows the (rounded) half-Winsorized exotic replacements for the exotic values in the fibian decomposition in Table 4 for the dental gold data. Table 6 gives the *inner decomposition* (rounded to the nearest integer), the result of mean polishing the replacement decomposition in Table 5: The emboldened values highlight the cells which were flagged exotic in the fibian decomposition.

The values of the *exotic supplements* , the difference between the exotic values and their replace-

ments, will be known before redecomposition; for example, the exotic value -57 in the D subtable becomes the exotic supplement $-52 = -57 - (-5)$, since -5 is half of the Winsorized value -10, the non-exotic subtable entry, sign considered, nearest to -57.

Finally, Table 7 shows a full *additive* decomposition of the data, formed by taking the sum of the inner decomposition in Table 6 and the exotic supplements. Thus, each emboldened entry in Table 7 is the sum of its inner value and its exotic supplement, and other entries are just inner values. For example, the emboldened entry -42 for the fifth dentist in the D subtable is obtained as $-42 = 10 + (-52)$, whereas the effect for the first dentist is 16, its inner value from Table 6.

Observe that in this additive decomposition, previously identified exotics still appear to stand out. Additionally, the entry 63 for the seventh gold also appears to stand out. Thus, the last three golds, which correspond to the same gold alloy but sintered at three different temperatures, appear to offer increased hardness in comparison with the other five golds.

5 A ROBUST ANALYSIS OF VARIANCE TABLE

In this section, we introduce a new style of analysis of variance table in which for each line of the analysis, we give the conventional degrees-of-freedom, an *inner mean square*, a list of the signed identities of any exotics, and include the classical mean square for comparison. Calculation of inner mean squares proceeds exactly as in a classical analysis. Thus, for example, the inner mean square for DC is derived from its inner subtable (Table 6), and is simply

$$\frac{[11^2 + (-18)^2 + \cdots + 23^2 + (-20)^2] \times 8}{(5-1) \times (3-1)} = 4218$$

Table 8 shows the robust analysis of variance table for the dental gold data.

What do we learn from Table 8? First, we discover the identity and signs of any exotic values, except where there are too many, when the number of each sign is just given, as in the DCG line. Secondly, when we compare standard mean squares with inner mean squares for each line in Table 8, we notice reductions of (i) 87% for dentists, indicating the very low level of hardness for the gold fillings produced by the fifth dentist; (ii) 99% for condensation methods, indicating an extremely low level of hardness produced when using the third condensation method (hand malleting); (iii) 87% for the dentists $\times$ condensation methods interaction, indicating the very low level of hardness achieved when the fourth and fifth dentist use hand malleting; and (iv) 85% for the condensation methods $\times$ gold alloys interaction, indicating the low level of hardness attained when hand malleting is used with the eighth gold alloy. Thus, the fifth dentist and/or hand malleting tend to produce very soft fillings, even when hand malleting is applied to the eighth gold alloy which itself appears to produce reasonably hard fillings. Thirdly, we see that the 13 positive and 6 negative three-factor exotic combinations are not exhibiting equally extreme behaviour, as judged by the smaller (77% reduction) in the DCG mean square. Lastly, considered alone, the standard analysis of variance table tells us none of this!

6 DOWNSWEEPING

The process of upsweeping separates the data into as many identifiable subtables as possible. The subsequent process of downsweeping preserves (by not downsweeping) a subtable if its mean square is at least twice the mean square of what it could be swept down into, namely, a subtable involving one more factor—"Paull's rule-of-two"; see Paull (1950) and Chapter 11 of *FEAV* for

details on downsweeping in the conventional setting.

In the conventional non-robust setting, this downsweeping process can be suggested by starting from an analysis treating all factors as "random", but it is *not* intended to depend on either the appropriateness or inappropriateness of such a choice. It is intended to provide a standard reduction of complexity likely to be useful in any of a variety of situations. The results of downsweeping can be incorporated in further analyses that treat some factors as "fixed", some as "random". It is intended to concentrate the decomposed data, not to choose a detailed framework of analysis.

Downsweeping Conventional Subtables

Table 9, based on the conventional, and less than satisfactory, analysis of the dental gold data shows a detailed application of downsweeping. The resulting downswept analysis is given in Table 10. Notice that the (pooled) mean square for a surviving composite subtable is computed as the weighted average of the mean squares of its component subtables, with weights proportional to the degrees-of-freedom of the corresponding components. Of the eight subtables from the complete upsweeping, five have preserved their identity and three have been downswept.

Downsweeping Inner Subtables

For the inner subtables, there seems to be no reason to alter the techniques used for classical subtables. Degrees-of-freedom are as in a conventional analysis. The resulting downswept analysis is given in Table 11. This time, the downsweeping leaves only the common term and the compound two-way and three-way interactions DG^* and DCG^{**}.

The possibility of downsweeping exotic supplements is considered briefly in Section 10.

7 STANDARD ERRORS

Now that we have completed robust upsweeping, identification, redecomposition, the robust analysis of variance table, and downsweeping, and are left with survival lists of inner subtables and of exotics that appear to deserve individual attention, we now consider the final part of our recipe, *evaluation*; that is, "standard errors" and "confidence statements" for subtable entries. These evaluations are intended for guidance only, as presently there is no distribution theory or simulation results available for our robust procedure.

We plan to associate a standard error with the label of each surviving inner subtable. Here we are concerned with subtables as "upper lines", and we want to assess stability and confidence for subtable entries. Thus, if there are exotic values associated with a surviving inner subtable, we should recombine the inner subtable and the exotic supplements into a single table, within which to make comparisons or assessments. The standard errors will be applied to both exotic and inner entries.

If we have a full subtable containing m entries, and would be prepared to look at any simple comparison among them, there is little doubt that $q(m, \nu, 5\%)$, the upper five-percent point of the studentized-range of comparisons among m values, based upon ν degrees of freedom, is the natural critical value to use; see *FEAV*, Table A-4, p414. For comparison with *zero*, we can use the Boole-Bonferroni Student's t based allowance $t(\nu, 2.5\%|m)$ for a 95% confidence interval; see *FEAV*, Table A-1, p405. If we wish to have a more informative analysis of two-factor interactions, we are led to work with simple bicomparisons—double differences involving the 4 entries from 2 rows and 2 columns; see *FEAV*, p355.

If we were making a standard analysis, the formulas for standard errors are simple and uniform, namely

$$\text{SE} = \left[\frac{\text{Mean Square providing error term}}{\text{number of observations per entry}}\right]^{\frac{1}{2}}$$

However, as we are making a robust analysis based upon inner mean squares, there are two processes that should be accounted for in a more careful analysis when dealing with any inner subtable other than the all-factor-interaction subtable. First, because of flagging in a subtable contributing to the standard error, fewer observations may contribute to some of the values we are about to compare. Secondly, because of the possibility of flagging, whether or not realized in any particular situation, and thus removing from the inner mean square large entries in subtables contributing to the standard error, the inner mean square will be slightly biased downward. While we cannot provide a definitive procedure to deal with both of these effects, a plausible interim measure, would be to expand the naive standard error of each surviving subtable by the factor

$$1.05 \times \text{maximum}\left\{\frac{1}{\text{contraction}} \mid \text{contributing subtables}\right\}$$

where the "contraction" of a subtable contributing to the standard error is simply the proportion of unflagged entries in that subtable. Here, the 1.05 is a rough allowance for shrinking because of possible flagging, while the maximum over contributing subtables is a conservative allowance for the effects of absence of flagged values from the computation.

However, note that this interim procedure does not take account of the consequences of unbalance of flagged values across the versions that we are comparing. Still greater precision and extreme care would lead us to use different standard errors for different determinations, comparisons or bicomparisons of the entries in a single subtable. Only in quite extreme cases does this seem likely to be important.

We now illustrate some of these calculations for the dental gold example. The naive standard error for the "common" value based upon the inner mean square for $DG^\star$ in Table 11 is $\sqrt{8262/120} = \pm 8.3$ and $t(39, 2.5\%|1) = 2.023$, leading to a Boole-Bonferroni allowance of ± 16.8. However, the corresponding allowance based upon the $DCG^\star$ inner mean square is only ± 8.9. Only the larger of these two standard errors allows for unfixedness (possibly randomness) in the versions of C.

The naive standard error for $DG^\star$, which can only be based on inner $DCG^\star$, is $\sqrt{2398/3} = \pm 28.3$ and $t(80, 2.5\%|40) = 3.3107$, leading to a Boole-Bonferroni allowance of ± 93.6. On the other hand, the Studentized-range tabular value is $q(40, 80, 5\%) = 5.596$, leading to an allowance of ± 158.2 for simple comparisons of $DG^\star$ entries. In the example, the standard errors change only moderately when they are inflated to take account of the two possible effects of flagging described above.

8 REPLICATION AND NESTING

We now consider what sort of analysis we should do at the lowest level, provided this is a level of pure replication. If we have at least three replicates in each cell, we can pass upward, to the cell level, the fibian of the cell replicates. The resulting sets of residuals, one set for each cell, can be processed as before using the exotic identification method described in Section 3, except that here we have a choice between either applying it to each set of residuals separately, or to the pooled set of residuals from all cells, or to each of a few pooled sets of residuals, probably divided to reflect cell levels. The second of these is likely to be the choice in most balanced situations (at least until we learn more about the problem), as it accords with the usual assumption of constant variance in

the classical analysis.

Whichever procedure is adopted in any particular instance, we apply a half-Winsorized analysis at each cell; that is, for each cell, we replace any exotic entry by its half-Winsorized value and then sweep out the mean of the modified residuals and add it to the fibian of the original replicate values. These modified cell summaries then become the input to the analysis described in the previous sections.

The *inner* replicate mean squares will be calculated as before by analogy with the corresponding classical analysis. The inner mean square will estimate the assumed constant variance, while the existence of one or more exotic values will signal unusual behavior at the pure replicate level.

One possible reason for unusual behavior is non-constancy of variance, which is more likely than not to be associated with differences in level. Thus we should, if exotic replications arise, compare the fibians for cells with exotic replications with the fibians for the other cells.

All that we have said about pure replication applies equally well to the nesting of a factor (with three or more versions) within a factor or within the interaction of two or more factors.

9 DISCUSSION FOR DENTAL GOLD DATA

In this section we give more details about the dental gold experiment and discuss what we seem to have learned from our robust analysis.

What Do We Know About The Experiment?

Xhonga (1971) reported that the primary objective of the experiment was to find a dental gold filling with increased hardness, and it was known that condensation when carried out properly increases hardness. The three methods of condensation compared in the experiment were: hand condensation ($C3$) in which hand pressure is applied using a handpiece; hand malleting ($C2$) which is the same as hand condensation, except that blows are delivered by a hand mallet; and electromalleting ($C1$) in which blows are delivered mechanically at constant frequency. It was anticipated that hardness levels achieved with both hand condensation and hand malleting would be more variable and more dependent on the ability of the dentists than with electromalleting,

Eight golds were compared: the first comprised pure gold cylinders; the second comprised pellets of powdered gold in a gold foil envelope; and the remaining six were in a 2×3 factorial structure, comprising two types of "direct gold alloy" each sintered at three different temperatures (1500° F, 1600° F, 1700° F). In this paper we have not attempted to incorporate this underlying additional factorial structure; however, see *FEAV* for an exploratory treatment of the classical four-way analysis of the last six golds.

Five dentists participated in the experiment, each of them applying each of the three condensation methods to each of the eight golds, thus requiring twenty-four specimens in all. Morton Brown, in correspondence with Donald Preece, reported that dentist 5 had been physically tired prior to the experiment; Preece(1983).

Each reported hardness number in Table 1 is the sum of ten unavailable measurements taken on the corresponding specimen. Each such measurement is a measure of surface hardness called "Vickers diamond hardness" computed as $1.8544\,W/d^2$ where W in this experiment was taken to be 1 kg and d is surface indentation in mm. Thus, for example, hardness values of 245 and 1115 in Table 1 convert to indentations of 0.28 mm and 0.13 mm, respectively.

Preece (1983) noticed that equally spaced indentations transform to the hardness values in Table 1 which are 5 units apart for low values and 13 apart for high values, thus presenting enough

discreteness for some to question a Gaussian assumption for errors.

What Have We Learned From Our Analysis?

Roughly, we seem to have learnt the following:

- The common or "undisturbed" (by the factors investigated) hardness is about 78 units give or take about 0.8, a crude standard error calculated in Section 7 from the second line in Table 11.
- The sixth and seventh golds (Au Ca sintered at 1500° F and 1600° F) produce somewhat harder fillings. In fact the last three golds, which outperformed the other five, were of one type of direct gold alloy (Au Ca), which apparently all five dentists reported handled with greater ease.
- Hand condensation seems to be inferior for all dentists, particularly for the fourth and fifth,
- The combination of hand condensation and the eighth gold alloy (Au Ca 1700° F) seems to produce a further softening.
- Of the 19 DCG exotics, 12 were associated with the fourth and fifth dentists ($D4$ or $D5$), 11 with hand condensation ($C3$), and 16 were associated with either $C3$ or $D4$ or $D5$. Of the remaining lines of the analysis involving C or D, all exotics involve $C3$ or $D4$ or $D5$.
- The larger exotics appear to be associated with low hardness, whereas many more positive exotics seem to be associated with $C1$ and $C2$. The inhomogeneity of variability suggested by these observations has been noted previously in Section 2.
- The one surviving subtable, $DG^\star$, shows that there is substantial interaction between dentists and the golds, although this was the only subtable that showed no signs of exoticity.

What Might We Do Next?

The aim of our dental gold analysis is choices that make fillings harder. Thus we can consider setting aside $C3$ and $D4$ and $D5$, leaving a $3 \times 2 \times 8$ data pattern, that could be analysed by itself. If the versions of the two factors set aside are involved in distinctive interactions, removing obviously low values might lead to a cleaner analysis.

The analysis just described is appropriate if we are interested in performance by skilled dentists. A very different option arises if we are asking for reasonable performance for any of a wide variety of dentists. An extreme analysis of our dental gold data would involve finding the worst (i.e. lowest hardness) result for the five dentists. We can do this for each of the 3×8 combinations of condensations and golds, and we can analyse the resulting 3×8 table by fibian polish. This analysis may be more sensitive to outliers than we would like. A compromise would be the same sort of analysis of the next-to-lowest hardness achieved by the five dentists.

Table 13 reports the gold and condensation main effects (from the appropriate fibian polish) for four analyses, the analysis of earlier sections and the three just suggested. All analyses agree that $G6$ and condensation $C1$ or $C2$ is a desirable choice. Most golds and the satisfactory condensations show no strong trend. Gold $G8$ and condensation $C3$, both of which are to be avoided as result of each analysis, show rather strong trends from one analysis to another. Together, these last two statements imply interaction-like behavior, suggesting that, given the concerns to be addressed, analysing less than all the data is desirable, as does the absence of exotics in the DCG subtable of

the 3 dentist-2 condensation analysis (not shown here) and the appearance of only small exotics in the main effects and two-factor interactions.

10 DISCUSSION

In this paper, we have set forth a recipe for robust analysis of variance for factorial data (with at least three versions for each factor) which explicitly seeks to identify exotic behaviour associated with the different sorts of variation while insulating normal variation from the disturbing effects of such exotic behaviour.

The recipe comprises the following sequence. First, an upsweeping-by-medians decomposition of the data into subtables of main effects and interactions (associated with the different sorts of variation) is followed by a comparison-within-subtables to flag unusually large (exotic) entries in each of the subtables. Next, a classical analysis by means, after replacing exotic entries by values similar in magnitude to the non-exotic entries in their respective subtables, yields a decomposition of the data into inner subtables and associated adjusted exotic values (exotic supplements) which is used to construct an analysis of variance table in which, for each sort of variation, there is a list of exotic supplements and an inner (or denominator) mean square that excludes the exotic values. The analysis continues by downsweeping those inner subtables that are inadequately prominent, as judged by Paull's rule-of-two. The results are displayed as a decomposition of the data into exotic values and those inner subtables, both simple and composite, that survive downsweeping. Although the exploratory nature of the recipe is emphasized, an approximate confirmatory stage is suggested in which the inner mean squares corresponding to the surviving subtables can be used to construct appropriate error terms for analysing subtable entries that remain.

While we believe the recipe represents the right approach to robust analysis of variance and have kept the individual ingredients as simple as possible, we recognise that there may be scope for sharpening some of them. For example, (i) to some, it might seem desirable that the initial breakdown should be unique with the number of non-zero entries in each subtable being no more than the nominal degrees-of-freedom; (ii) flagging procedures which depend both on the degrees-of-freedom and on the level of the subtable in the design hierarchy may prove eventually to be better suited to identifying distinctive subtable entries; (iii) rather than simply regard subtable entries as either exotic or not, an alternative approach would be to weight each entry according to its size compared to the overall variability in the subtable, and then do a weighted mean analysis; (iv) we might consider downsweeping exotic values, and decide to downsweep an exotic supplement if an appropriate multiple of the square of its value is less than twice the mean square of the inner entries in a candidate fiber; (v) standard errors and associated confidence statements deserve careful study, as allowances will have to be made for the process of flagging exotic values and for subsequent processes in the recipe, such as downsweeping.

The procedures described in this paper can be readily adapted to other balanced designs such as Latin squares, balanced incomplete blocks, and those involving both crossed and nested factors, such as split-plot experiments. However, whenever factors with two versions are involved, different procedures are likely to be needed such as those developed by Seheult and Tukey (1982) and by Seheult (1997) for the special case where all the factors have just two versions. Unbalanced designs, including those with unequal replication, will also require special attention.

What has all this accomplished ? Mainly two kinds of things:

- recognition, and proper treatment, of those subtables, most of whose entries deserve to be downswept, although some are so large that they should not be downswept. Here, proper treat-

ment means partial downsweeping, accomplished by downsweeping "inner" subtables while leaving "exotic supplements" unmoved.

- arranging that unusually large entries (a) do not distort the summaries at higher levels and (b) do not contribute unduly to to that subtable's mean square, thus avoiding distortion of downsweeping or of variance estimates.

While we may have been led to consider robust analysis of variance as a way to make improvements, possibly minor, in the precision of our summaries, we now recognize its greatest virtue is its ability to cope much more adequately with data whose factorial structure may be almost absent, except perhaps for a sub-factorial structure, and to do this without compromise to its ability to cope adequately with data in which factorial structure is clearly evident. Its recognition has markedly increased the palette of formats through which we can describe real data.

ACKNOWLEDGMENT

This paper was prepared in part in connection with research at Princeton University, supported by the Army Research Office, Durham, DAAL03-91-G-0138.

REFERENCES

Basford, K. E. and Tukey, J. W. (1995). Intercomparing residuals to find outliers in randomized blocks. *Australian Journal of Agricultural Research*, **46**, 451-461.

Besag, J. E. (1981). On resistant techniques and statistical analysis. *Biometrika*, **68**, 463-469.

Besag, J. E. and Seheult, A. H. (1981). Resistance. In *Encyclopedia of Statistical Sciences, Volume 8* (eds S. Kotz, N.L. Johnson and C.B. Read), pp 98-101. Wiley. New York.

Brown, M. B. (1975). Exploring interaction effects in the ANOVA. *Applied Statistics*, **24**, 288-298.

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

Daniel, C. (1976). *Applications of Statistics in Industrial Experimentation.* Wiley. New York.

Hoaglin, D. C., Mosteller, F. and Tukey, J. W., eds (1991). *Fundamentals of Exploratory Analysis of Variance.* Wiley. New York.

Johnson, E. G. (1988). *Robust Analysis of Factorial Designs via Elemental Subsets and Outlier Sterilization.* PhD Thesis, Princeton University, New Jersey.

Johnson, E. G. and Tukey, J. W. (1987). Graphical exploratory analysis of variance illustrated on a splitting of the Johnson and Tsao data. In *Design, Data, and Analysis: By Some Friends of Cuthbert Daniel* (ed. C. L. Mallows), pp. 171-244. Wiley. New York.

Mallows, C. L. and Tukey, J. W. (1982). An overview of techniques of data analysis, emphasizing its exploratory aspects. In *Some Recent Advances in Statistics* (eds J. Tiago de Oliveira and B. Epstein), pp. 111-172. Academic Press. London.

Oehlert, G. W. (1994). Isolating one-cell interactions. *Technometrics*, **36**, 403-408.

Paull, A. E. (1950). On a preliminary test for pooling mean squares in the analysis of variance. *Annals of Mathematical Statistics*, **21**, 539-556.

Preece, D. A. (1983). *Personal Communication*.

Seheult, A. H. (1997). Bonus decompositions for robust analysis of 2^n factorial experiments. In *The Practice of Data Analysis: Essays in Honor of John W. Tukey* (eds D.R.Brillinger, L.T.Fernholtz and S.Morgenthaler), pp. 299-316. Princeton University Press. Princeton, New Jersey.

Seheult, A. H. and Tukey, J. W. (1982). Some resistant procedures for analysing 2^n factorial experiments. *Utilitas Mathematica*, **21B**, 57-98.

Tukey, J. W. (1977). *Exploratory Data Analysis.* Addison-Wesley. Reading, MA.

Tukey, J. W. (1993). Exploratory analysis of variance as providing examples of strategic choices. In *New Directions in Data Analysis and Robustness, Monte Verita* (eds S.Morgenthaler, E.Ronchetti and W.A.Stahel), pp. 205-258. Birkhöuser Verlag. Basel.

Xhonga, F. (1971). Direct gold alloys – Part II. *Journal of the American Academy of Gold Foil Operators*, **14**, 5-15.

D	C	G							
		1	2	3	4	5	6	7	8
	1	792	824	813	792	792	907	792	835
1	2	772	772	782	698	665	1115	835	870
	3	782	803	752	620	835	847	560	585
	1	803	803	715	803	813	858	907	882
2	2	752	772	772	782	743	933	792	824
	3	715	707	835	715	673	698	734	681
	1	715	724	743	627	752	858	762	724
3	2	792	715	813	743	613	824	847	782
	3	762	606	743	681	743	715	824	681
	1	673	946	792	743	762	894	792	649
4	2	657	743	690	882	772	813	870	858
	3	690	245	493	707	289	715	813	312
	1	634	715	707	698	715	772	1048	870
5	2	649	724	803	665	752	824	933	835
	3	724	627	421	483	405	536	405	312

Table 1: Dental gold data.

Source	df	Mean Square
(1)	1	65118387
D	4	54394
C	2	298808
G	7	31477
DC	8	32930
DG	28	7458
CG	14	14984
DCG	56	9969

Table 2: Standard analysis of variance table for the dental gold data.

(1)	737							
D	48	43	5	-37	-59			
C	49	50	-100					
G	-9	-22	-12	-27	-48	84	58	-23
	-16	-22	38					
	-6	-34	40					
DC	-53	-25	78					
	32	35	-67					
	43	45	-88					
	6	36	9	-54	27	87	-114	2
	-14	3	6	14	12	-34	-26	39
DG	24	-38	37	-30	10	-26	12	11
	-17	-34	-30	105	-44	23	67	-70
	1	33	-22	-35	-5	-51	60	18
	-54	38	-20	-26	29	-12	16	29
CG	-53	-20	-3	-6	-30	31	11	70
	107	-18	24	32	1	-19	-27	-99
	30	-47	18	81	-34	-70	13	9
	15	-36	-26	-28	-98	99	67	8
	-45	83	8	-53	132	-29	-80	-17
	57	-39	-82	19	-2	-3	36	14
	32	15	-15	4	13	56	-46	-58
	-89	24	97	-23	-11	-53	10	45
	15	8	0	-27	23	74	-62	-31
DCG	64	29	25	40	-85	-31	0	-42
	-79	-36	-25	-13	62	-44	62	73
	-28	182	73	-90	44	18	-131	-68
	-49	33	-51	25	108	-111	-51	96
	77	-215	-22	65	-152	93	182	-28
	-74	-104	-9	16	-31	-19	144	76
	-62	-40	67	-41	62	-13	31	-3
	136	144	-58	24	-31	32	-175	-73

Table 3: Separate subtables for mean decomposition of the dental gold data.

(1)	771							
D	20	1	0	-10	**-57**			
C	1	0	**-65**					
G	-9	0	1	-17	-18	**95**	38	43
	0	-19	0					
	30	-11	0					
DC	-48	0	9					
	0	0	**-146**					
	0	27	**-208**					
	9	32	25	-46	0	26	-38	0
	0	0	-4	17	0	-60	0	36
DG	0	-39	27	-42	28	-30	0	-43
	-80	-1	-58	149	0	17	33	-109
	-72	0	0	0	0	0	116	0
	0	0	-9	0	18	20	0	0
CG	0	-17	14	-11	0	-12	38	51
	56	0	0	25	-16	0	-11	**-172**
	0	0	4	63	0	-26	0	0
	0	-15	-30	0	-89	**234**	25	4
	0	45	0	-68	**143**	0	**-155**	-12
	9	0	-76	0	10	0	66	0
	0	28	0	32	0	**149**	-45	-67
	-39	0	**131**	-17	0	-44	0	67
	0	39	0	-38	0	49	0	0
DCG	30	0	0	42	**-168**	0	0	-40
	0	-70	0	0	34	-65	82	**138**
	0	**185**	96	**-151**	0	0	-41	-47
	-15	0	-28	0	29	-48	0	**112**
	173	**-304**	0	0	**-227**	53	**203**	0
	0	0	0	0	0	-58	**179**	**112**
	-11	0	47	-48	29	0	0	0
	308	**186**	-21	34	-2	0	**-179**	0

Table 4: Separate subtables for fibian decomposition of the dental gold data. Identified exotic entries are in bold.

(1)	771							
D	20	1	0	-10	**-5**			
C	1	0	**0**					
G	-9	0	1	-17	-18	**21**	38	43
	0	-19	0					
	30	-11	0					
DC	-48	0	9					
	0	0	**-24**					
	0	27	**-24**					
	9	32	-25	-46	0	26	-38	0
	0	0	-4	17	0	-60	0	36
DG	0	-39	27	-42	28	-30	0	-43
	-80	-1	-58	149	0	17	33	-109
	-72	0	0	0	0	0	116	0
	0	0	-9	0	18	20	0	0
CG	0	-17	14	-11	0	-12	38	51
	56	0	0	25	0	-16	-11	**-8**
	0	0	4	63	0	-26	0	0
	0	-15	-30	0	-89	**48**	25	4
	0	45	0	-68	**48**	0	**-44**	-12
	9	0	-76	0	10	0	66	0
	0	28	0	32	0	**48**	-45	-67
	-39	0	**48**	-17	0	-44	0	67
	0	39	0	-38	0	49	0	0
DCG	30	0	0	42	**-44**	0	0	-40
	0	-70	0	0	34	-65	82	**48**
	0	**48**	96	**-44**	0	0	-41	-47
	-15	0	-28	0	29	-48	0	**48**
	48	**-44**	0	0	**-44**	53	**48**	0
	0	0	0	0	0	-58	**48**	**48**
	-11	0	47	-48	29	0	0	0
	48	**48**	-21	34	-2	0	**-44**	0

Table 5: Separate subtables for the half-Winsorized fibian decomposition of the dental gold data. Replacement values are in bold.

(1)	781							
D	16	10	-19	-18	**10**			
C	1	2	**-3**					
G	-27	-14	-9	-12	-26	**0**	63	25
	11	-18	7					
	22	-19	-3					
DC	-34	9	25					
	5	4	**-9**					
	-3	23	**-20**					
	33	38	16	-60	-17	46	-73	17
	14	5	-14	9	-1	-47	-22	56
DG	43	-44	35	-45	-30	-14	8	-12
	-42	-1	-33	124	-6	34	10	-86
	-47	1	-3	-29	-6	-18	77	25
	-21	18	-8	-5	21	13	-1	-17
CG	-21	-13	9	-6	-12	-1	20	24
	43	-5	0	12	-9	-13	-20	**-7**
	-2	-27	5	60	5	-34	-7	1
	5	-21	-15	-5	-61	**30**	43	23
	-3	48	9	-56	**56**	5	**-36**	-23
	24	-20	-67	-2	4	4	52	5
	12	19	12	17	8	**32**	-44	-55
	-35	1	**55**	-14	-12	-36	-8	50
	-9	36	-4	-40	-2	57	-37	-1
DCG	25	14	6	33	**-27**	-6	-16	-28
	-16	-49	-2	7	29	-51	53	**29**
	-5	**38**	74	**-26**	4	5	-49	-41
	-22	2	-45	7	48	-62	9	**63**
	27	**-40**	-29	19	**-52**	56	**40**	-21
	-7	-26	-9	8	-11	-32	**41**	**37**
	-20	-14	42	-52	32	6	8	-3
	28	**40**	-34	45	-21	26	**-49**	-35

Table 6: Inner subtables from the mean re-decomposition of the half-Winsorized fibian decomposition of the dental gold data in Table 5.

(1)	781							
D	16	10	-19	-18	**-42**			
C	1	2	**-68**					
G	-27	-14	-9	-12	-26	**73**	63	25
	11	-18	7					
	22	-19	-3					
DC	-34	9	25					
	5	4	**-131**					
	-3	23	**-204**					
	33	38	16	-60	-17	46	-73	17
	14	5	-14	9	-1	-47	-22	56
DG	43	-44	35	-45	-30	-14	8	-12
	-42	-1	-33	124	-6	34	10	-86
	-47	1	-3	-29	-6	-18	77	25
	-21	18	-8	-5	21	13	-1	-17
CG	-21	-13	9	-6	-12	-1	20	24
	43	-5	0	12	-9	-13	-20	**-170**
	-2	-27	5	60	5	-34	-7	1
	5	-21	-15	-5	-61	**216**	43	23
	-3	48	9	-56	**151**	5	**-147**	-23
	24	-20	-67	-2	4	4	52	5
	12	19	12	17	8	**133**	-44	-55
	-35	1	**138**	-14	-12	-36	-8	50
	-9	36	-4	-40	-2	57	-37	-1
DCG	25	14	6	33	**-151**	-6	-16	-28
	-16	-49	-49	7	29	-51	53	**119**
	-5	**175**	74	**-132**	4	5	-49	-41
	-22	2	-45	7	48	-62	9	**127**
	152	**-299**	-29	19	**-234**	56	**195**	-21
	-7	-26	-9	8	-11	-32	**172**	**101**
	-20	-14	42	-52	32	6	8	-3
	288	**178**	-34	45	-21	26	**-183**	-35

Table 7: The additive decomposition of the dental gold data in which each emboldened entry is the sum of its inner value from Table 6 and its exotic supplement, and the other entries are inner values.

Label	Df	Standard MS	Inner MS	Signed exotic labels
D	4	54394	6978	$-D5$
C	2	298808	206	$-C3$
G	7	31477	13768	$+G6$
DC	8	32930	4218	$-D4C3$ $-D5C3$
DG	28	7458	7068	
CG	14	14984	2253	$-C3G8$
DCG	56	9969	2253	13^+ 6^-

Table 8: Robust analysis of variance table (including the standard analysis) of inner mean squares summarizing the inner subtables resulting from the half-Winsorized decomposition of the dental gold data displayed in Table 6. There are 13 positive and 6 negative DCG exotics.

Label	df	MS	Possible subtables to downsweep into (⇓)			Action
common	1	65118387	54394(D)	298808(C)	31477(G)	hold
D	4	54394	32930(DC)	7458(DG)		⇓ into DC
C	2	298808	32930(DC)	14984(CG)		hold
G	7	31477	7458(DG)	14984(CG)		hold
DC	12	40085	9969(DCG)			hold
DG	28	7458	9969(DCG)			⇓ into DCG
CG	14	14984	9969(DCG)			⇓ into DCG
DCG	56	9969				

Table 9: Application of Paull's rule-of-two to the conventional mean squares for the dental gold data of Table 1.

Label	df	MS
common	1	65118387
C	2	298808
G	7	31477
DC*	12	40085
DCG*	98	9968

Table 10: Final downswept conventional mean squares for the dental gold data following from Table 9, where DC* results from pooling DC and D, and DCG* results from pooling DCG, DG and CG.

Label	df	MS
common	1	73159398
DG*	39	8262
DCG*	80	2398

Table 11: Final downswept inner mean squares for the dental gold data, where DG* is the result of pooling D, G and DG, and DCG* is the result of pooling D, C, DC, CG and DCG.

Ordered sizes z_i	Working values c_i	Scales s_i	Ratios s_i/s
208	1.770	117.5	2.73*
146	1.304	111.9	2.60*
48	1.020	47.0	1.09
30	0.801	37.4	0.87
27	0.615	43.9	1.02
19	0.448	42.4	0.98
11	0.293	37.5	0.87
9	0.145	62.0	1.44
		$s = 43.1$**	

* These ratios exceed the cutoff value $K = 1.5$, so that the corresponding cell entries -208 and -146 are identified as exotic.
** 43.1 is the middle-median of the eight scales; that is, the median of $47.0, 37.4, 43.9, 42.4$.

Table 12: Calculations for identifying exotic entries in the DC subtable of the fibian decomposition of the dental gold data given in Table 4.

Golds	All 5 dentists	3 dentists and 2 compactions	Next to worst dentist	Worst dentist	Comments (range)
6	95	105	143	148	high in each (53)
7	38	28	77	80	moderately high (52)
3	1	0	0	25	neutral (25)
2	0	−10	9	33	neutral (43)
4	−17	−10	−17	0	possibly low (17)
5	−18	0	−50	−52	moderately low (52)
1	−9	−10	−42	−16	weak trend (33)
8	43	33	9	−33	strong trend (76)
Compactions					
1	1	20	0	17	small (20)
2	0	0	0	0	neutral (0)
3	−65		−160	−277	strong trend (212)

Table 13: Gold and compaction effects from fibian analyses of the $G \times C$ tables of four analyses.