Center for Quality and Productivity Improvement
University of Wisconsin-Madison
610 Walnut Street
Madison, WI 53705

Report No. 19

An Investigation of the Method of Accumulation Analysis

George Box and Stephen Jones

December 1986

This research was supported by National Science Foundation Grant DMS-8420968, by the United States Army under Contract No. DAAG29-80-C-0041, and by the Vilas Trust of the University of Wisconsin--Madison; and was aided by access to the research computer of the University of Wisconsin--Madison Statistics Department. Completed November 1986.

CENTER FOR QUALITY AND PRODUCTIVITY IMPROVEMENT
UNIVERSITY OF WISCONSIN-MADISON

Report No. 19

AN INVESTIGATION OF THE METHOD OF ACCUMULATION ANALYSIS

George Box and Stephen Jones

December 1986

PRACTICAL SIGNIFICANCE

Results are frequently recorded in ordered categories which are assessed subjectively. Thus a weld may be visually judged as excellent, good, average, poor, or inferior. Various procedures have been proposed for analyzing such data. In particular, Taguchi (1974) has developed a complex method called accumulation analysis. In this paper this method is examined in the context of a paper by V.N.Nair (1986) and is compared with simpler alternative procedures. It is concluded that accumulation analysis lacks efficiency and is also unnecessarily complicated. Simpler, more intuitive methods are recommended which employ a simple scoring system with calculations performed along traditional lines. The ideas are illustrated with examples.

Key words: Accumulation analysis; ordered categorical data; location effects; dispersion effects.

This research was supported by National Science Foundation Grant DMS-8420968, the United States Army under contract No.DAAG29-80-C-0041, and by the Vilas Trust of the University of Wisconsin--Madison; and was aided by access to the research computer of the University of Wisconsin--Madison Statistics Department. Completed November 1986.

Investigation of the Method of Accumulation Analysis

George Box & Stephen Jones

Summary

Accumulation Analysis is a method introduced by Taguchi (1974) for the analysis of ordered categorical data. We believe that the researches carried out by Nair (1986) , Hamada and Wu (1986), and by us show that accumulation analysis is unnecessarily complicated, as well as inefficient, and consequently should not be taught or recommended.

1. Introduction

Industrial and other experimental results often appear as ordered categorical data. This occurs particularly when the measurements are subjective. For example, Fisher (1963), in his book "Statistical Methods for Research Workers," which first appeared in 1925, discussed the analysis of an experiment in which twelve samples of human blood tested with twelve different sera gave reactions represented in subjectively graded classes denoted by –, ?, w, (+), and +. Data of this kind are common. The food industry routinely uses such scales; for example, how good a cake tastes is frequently assessed by a subjective hedonic scale usually containing seven ordered classes. Many other examples come to mind from consulting experience, including assessment of how unpleasant was a particular odor, how soft to the touch was a particular fabric, and so on.

The Center for Quality and Productivity Improvement cares about your reactions to Reports. Please send your comments (general or specific) to: Report Feedback, Center for Quality and Productivity Improvement, 610 Walnut Street, Madison, WI 53705. All replies will be forwarded to the authors.

Often, as an aid to assessment, the levels are defined by using standard specimens. Thus, a set of standard fabrics may be chosen that are agreed to define, in an appropriately graded manner, the characteristic softness that is sought. Typically a panel will choose k such graded standards in a way that to them represents a scale having approximately equal intervals. It thus makes sense to designate these k classes by equally spaced scores such as $1, 2, ..., k$.

Although more elaborate methods, some of which are mentioned in Nair (1986), are available, we have usually analyzed such scores with standard procedures, such as those described in Davies (1958). Thus, for the analysis of fractional factorial designs we might use such scores to calculate effects and to analyze the results with Daniel plots on normal probability paper (Daniel (1959)). More recently we would add the Bayes plots proposed by Box and Meyer (1986). For more extensive data, we might employ the analysis of variance and derived t statistics. If preferred, nonparametric methods might be substituted.

It may be argued that such analysis is inappropriate because (1) the scale is subjective, (2) it is discrete rather than continuous, and (3) the necessary distributional assumptions are not satisfied. We feel that the first objection does not carry much weight, provided that the scale is carefully and thoughtfully chosen by a consensus of knowledgeable people and bearing in mind that the final results will be expressed on the same scale. On the second and third points, it is known that standard statistical procedures are remarkably robust to drastic rounding and to departures from distributional assumptions. In any case, we would argue that accumulation analysis avoids none of these difficulties.

2. Chi-square versus Accumulation Analysis

Accumulation analysis is an alternative approach for analyzing ordered categorical data. In a basic reference (Taguchi (1974)), to which later papers often refer, the arguments in favor of the method are illustrated using a particular set of hypothetical data. It seems to us appropriate therefore to reconsider Taguchi's method of analysis in terms of the same set of data that he used to introduce it, set out in Table 1.

It was supposed that for each of two drugs, A and B, a control and treated group each of eighty patients were tested and were allocated to the categories not effective (–), slightly effective (+), effective (++), and markedly effective (+++). We have added the estimated marginal probabilities $\hat{p}_i$ and the corresponding treatment-control differences y_i.

Drug A					Drug B				
	–	+	++	+++		–	+	++	+++
Control	40	24	10	6	Control	40	24	10	6
Treatment	24	40	10	6	Treatment	24	29	16	11
$\hat{p}_i$	$\frac{64}{160}$	$\frac{64}{160}$	$\frac{20}{160}$	$\frac{12}{160}$	$\hat{p}_i$	$\frac{64}{160}$	$\frac{53}{160}$	$\frac{26}{160}$	$\frac{17}{160}$
y_i	$\frac{-16}{80}$	$\frac{16}{80}$	$\frac{0}{80}$	$\frac{0}{80}$	y_i	$\frac{-16}{80}$	$\frac{5}{80}$	$\frac{6}{80}$	$\frac{5}{80}$
$\chi_3^2 = 8.00$ $\alpha = 4.7\%$					$\chi_3^2 = 7.33$ $\alpha = 6.0\%$				

Table 1. Hypothetical Data for Treatment and Control Patients with Two Therapeutic Drugs A and B, given by Taguchi (1974).

After noting the values of the Pearson χ^2 statistics (given below Table 1), Taguchi gives the following motivation for his alternative analysis " . . . the efficacy of the drug A is significant but that of drug B is not despite the fact that it is clear the efficacy of B is much greater than that of A. This example shows that the use of the χ^2 test is unsuitable in cases in which the . . . data are arranged in order. . . " (Taguchi (1974)).

Karl Pearson's test is for general discrepancies from expected proportions and is not, of course, intended to be sensitive to a particular discrepancy such as the ordering of categories of which it knows nothing. A way to look for a tendency toward an increase in the number of treated patients in the higher effect categories would have been to isolate the linear component in the expected proportions of the overall χ^2 statistic (see, for example, Snedecor and Cochran (1980)). If this is done we obtain the results shown in Table 2.

The significance levels associated with the linear components are then 16.0% and 1.1%. This clearly shows, as desired, that, after taking account of the ordering of the categories, drug B is superior to drug A in its therapeutic effect.

	df	Drug A	Drug B
Linear Component	1	1.98	6.49
Remainder	2	6.02	0.84
Total		8.00	7.33

Table 2. The Linear Components for the Drug Data

Instead of employing a simple standard decomposition of this kind, accumulation analysis was introduced. This used the cumulated frequencies shown in Table 3. In this table the frequencies in class *I* are the (–) frequencies, those in class *II* are the sum of the frequencies in the first two classes (–) and (+), and so on.

Analysis of variance tables were then constructed from these cumulated frequencies by calculating separate sums of squares for categories *I*, *II*, and *III*, scoring each patient as a 1 if they appear in the given category and a zero otherwise. The individual sums of squares for categories *I*, *II*, and *III* were then weighted with the reciprocal of the binomial variance appropriate for that category. As an example, Table 4 shows such an analysis of variance table for drug B taken from Taguchi (1974). Also shown is the general structure of the degrees of freedom and sums of squares for k categories ($k = 4$ in this example) and for n patients exposed to a treatment and n patients exposed to a control ($n = 80$ in this example).

It is not our purpose to rationalize this analysis of variance table. In particular, the very large number of degrees of freedom seems impossible to justify, since the total number of patients in this part of the trial was only $2n = 160$. For the purpose of subsequent discussion,

Drug A					Drug B				
	I	*II*	*III*	*IV*		*I*	*II*	*III*	*IV*
Control	40	64	74	80	Control	40	64	74	80
Treatment	24	64	74	80	Treatment	24	53	69	80
$\hat{q}_i$	$\frac{64}{160}$	$\frac{128}{160}$	$\frac{148}{160}$	1	$\hat{q}_i$	$\frac{64}{160}$	$\frac{117}{160}$	$\frac{143}{160}$	1

Table 3. Cumulated Frequencies for Data of Table 1

Source	df	Sum of Squares	Mean Square	
Treatments	$k-1=3$	$T_B=12.17$	4.06	$F=4.11$
Error	$2(n-1)(k-1)=474$	$2n(k-1)-T_B=467.83$	0.99	
Total	$(2n-1)(k-1)=477$	$2n(k-1)=480.00$		

Table 4. Taguchi's Accumulated Analysis of Variance Table (Drug B)

however, note that:

(i) The entries in the table only involve the frequency data via the quantity T, which we will call the *accumulation statistic*. Using subscripts A and B to denote its value for the two drugs we find $T_A = 6.67$ and $T_B = 12.17$

(ii) The "mean square for error" $\frac{2n(k-1)-T}{2(n-1)(k-1)}$ will usually be close to unity.

(iii) The stated number of error degrees of freedom $2(n-1)(k-1)$ will usually be fairly large, and consequently referral of the mean square ratio to an F distribution will be closely equivalent to referring the accumulation statistic T to a χ^2 distribution with $k-1$ degrees of freedom.

3. Nature of the Accumulation Statistic T

A simple way to understand the nature of the accumulation statistic T is shown in Table 5.

From the accumulated frequencies in Table 3 we have constructed three new tables for each drug. In these tables $\bar{J}$, $(J = I, II, III)$, means the complement of category J and refers to the number of patients not in category J and hence the number that did better than those in category J. The accumulation statistics T_A and T_B may be obtained by summing the individual

	Drug A					
	I	$\bar{I}$	II	$\bar{II}$	III	$\bar{III}$
Control	40	40	64	16	74	6
Treatment	24	56	64	16	74	6
	$\chi^2 = 6.67$		$\chi^2 = 0$		$\chi^2 = 0$	
	$T_A = 6.67$					
	Drug B					
	I	$\bar{I}$	II	$\bar{II}$	III	$\bar{III}$
Control	40	40	64	16	74	6
Treatment	24	56	53	27	69	11
	$\chi^2 = 6.67$		$\chi^2 = 3.85$		$\chi^2 = 1.65$	
	$T_B = 12.17$					

Table 5. Individual χ^2 Values for Separate 2×2 Tables

χ^2 values for the separate 2×2 tables. This correspondence between the accumulation statistic and the individual Pearson χ^2 values was noted by Takeuchi and Hirotsu (1982). The larger value is now obtained for drug B because the method does take account of the ordering of the categories. These T statistics do not, however, follow χ_3^2 distributions, nor is the F distribution appropriate in the analysis of variance (Table 4 for example) because the components are not independent.

We also explored the moment approximation mentioned by Nair (1986) and we found that for this setup T is approximately distributed as $g\chi^2(k-1)/g$ where $(k-1)/g$ is the fractional

degrees of freedom for chi-square,

$$g = 1 + \frac{2}{k-1} \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} r_i / r_j \ ,$$

$$r_i = \hat{q}_i / (1 - \hat{q}_i) \ ,$$

and $\hat{q}_i$ are the empirical cumulated marginal probabilities as in Table 3.

Using an arrow to mean "is approximately distributed as" we then find for this particular set of data

$$T_A / 1.36 \rightarrow \chi^2(2.2) \ ,$$
$$T_B / 1.43 \rightarrow \chi^2(2.1)$$

and the approximate significance levels are then 10% and 1.7%, respectively.

4. Eigenvalue-Eigenvector Analysis

As in Table 1, denote by y_i the difference in the *ith* category of the proportion falling in the treated and control groups. The accumulation statistics are then functions of these differences. Employing the usual asymptotic theory in what follows χ_1, χ_2, χ_3 will be independent normal deviates and we find

$$T_A = 6.67 = 1.82\chi_1^2 + 0.78\chi_2^2 + 0.39\chi_3^2$$

with

$$\chi_1 = -9.9y_1 - 5.2y_2 + 2.4y_3 + 12.8y_4 = 0.94$$
$$\chi_2 = 4.2y_1 - 7.9y_2 - 5.5y_3 + 9.2y_4 = -2.42$$
$$\chi_3 = -0.5y_1 + 5.1y_2 - 13.8y_3 + 9.2y_4 = 1.12$$

and

$$T_B = 12.19 = 1.91\chi_1^2 + 0.72\chi_2^2 + 0.37\chi_3^2$$

with

$$\chi_1 = -9.0y_1 - 4.2y_2 + 2.4y_3 + 10.8y_4 = 2.41$$
$$\chi_2 = 4.3y_1 - 7.3y_2 - 6.2y_3 + 9.2y_4 = 1.22$$
$$\chi_3 = -1.3y_1 + 6.5y_2 - 11.5y_3 + 6.3y_4 = 0.20$$

If equal numbers (i.e. 40 patients) happened to fall in each of the four categories then the accumulation statistic would have been of the form

$$T = 2.00\chi_1^2 + 0.66\chi_2^2 + 0.33\chi_3^2$$

with

$$\chi_1 = -8.5y_1 - 2.8y_2 + 2.8y_3 + 8.5y_4$$
$$\chi_2 = 6.3y_1 - 6.3y_2 - 6.3y_3 + 6.3y_4$$
$$\chi_3 = -2.8y_1 + 8.5y_2 - 8.5y_3 + 2.8y_4$$

In this last case the coefficients in the linear aggregates are seen to be proportional to (–3,–1,1,3), (1,–1,–1,1) and (–1,3,–3,1); the linear, quadratic and cubic orthogonal polynomials. Note that the coefficients in T_A and T_B also approximate these same functions. In all of the above expressions, employing the usual asymptotic approximation the components χ_1^2, χ_2^2, χ_3^2 are, on the null hypothesis, independent chi-square variates having single degrees of freedom.

Now if we broke out the linear, quadratic and cubic components for Pearson's overall χ^2 statistic P, we would have a decomposition $P = \chi_1^2 + \chi_2^2 + \chi_3^2$. A pure linear tendency, captured by χ_1^2, say, would thus be contaminated by the remaining two components, which would constitute two sources of unnecessary noise.

Taguchi's accumulation statistic goes partway toward remedying this by putting greater emphasis on the first component. The remaining two components, however, still introduce

unnecessary additional noise and, in general, we would clearly be better off to simply separate out the linear component as we did earlier. In a similar way if, for some reason, we wished to consider any other components of the overall χ^2, these could be separated out in the usual way. In particular, as Nair (1986) has indicated, the additional quadratic component might be used to measure dispersion.

5. Linear Component of χ^2 and Some Related Procedures

As is well known, one way to compute the χ^2 statistic is to perform a one-way analysis of variance for the four therapeutic classes (see Cochran (1950)). Within each class, patients from the control and treatment groups are indicated by two levels of a suitable indicator variable. This method of analysis has been more recently referred to as "minute analysis" by Professor Taguchi. It is known to provide an approximately valid analysis for independent (and specifically non-accumulated) data. The linear component of χ^2, in particular, may be computed by breaking out an appropriate sum of squares for regression on an equally spaced class variable, and other components can be found similarly. If we employ -1 and $+1$ for the indicator variable, the analysis of variance tables obtained for the drug data are as shown in Table 6.

	df	Sum of Squares	Mean Square
Drug A			
Between Classes	3	8.00	
Linear	1	1.9768	1.9768
Remainder	2	6.0232	3.0158
Error	156	152.00	0.9744
Total	159	160.00	
Drug B			
Between Classes	3	7.327	
Linear	1	6.485	6.485
Remainder	2	0.842	0.4209
Error	156	152.673	0.9787
Total	159	160.00	

Table 6. Analysis of Variance on the Four Therapeutic Classes for the Drug Data

For such an analysis, the sum of squares between classes yields the overall chi-square statistic and the linear regression sum of squares on one degree of freedom is then equal to the linear chi-square component. If instead of carrying through the chi-square test we perform the analysis of variance test, then we have the approximation

$$F = \chi^2[\frac{160-4}{3(160-\chi^2)}].$$

It has been argued that approximations of this type can be slightly more accurate (see Mantel (1963)), but for the present data the difference is negligible.

Mantel's Statistic

A regression test for ordered categorical data based on a chi-square statistic having one degree of freedom was proposed by Mantel (1963). His justification is slightly different from that above. In the case that we are considering, however, if we use P_L for the linear component of Pearson's χ^2 his statistic M (say) is $M = P_L \times (\frac{N-1}{N})$ with $N = 160$. Thus, for examples of this kind, the practical difference from the P_L statistic is negligible. In discussing the statistic, Mantel remarked "that a linear regression is being tested does not mean that an assumption of linearity is being made. Rather it is that test of a linear component of regression provides power for detecting any progressive association which may exist. [This] chi square may have various interpretations, the basic one being that it provides a test with power for any progressive relation..." (Mantel (1963), p.698).

ANOVA and the t statistic Using Scores as Data

At the beginning of this discussion we mentioned the simple way in which we have usually analyzed ordered categorical data in the past. We have adopted a suitable score and, for factorial and fractional factorial designs, used standard plotting procedures. For more extensive data, we have used ordinary parametric or, if preferred, "nonparametric" procedures. For example, for the drug example we might simply score patients showing the different therapeutic effects as 0, 1, 2, 3 (or equivalently, −3, −1, 1, 3) and perform a standard analysis of variance for the difference between control and treatment or equivalently calculate the corresponding t−statistic. We then obtain values for $F = t^2$ of ${t_A}^2 = 1.977$, ${t_B}^2 = 6.675$, corresponding with the differences in mean score between treated and control patients. Numerically these F $(= t^2)$ values are very close to the linear chi-square components obtained in the earlier analysis of variance of Table 6. Even this slight discrepancy is easily explained. The $F = t^2$ values obtained above are *precisely* what

we would get from the analysis in Table 6 if we included the two degrees of freedom for "remainder" in the error sums of squares. We see, therefore, that numerically the approximate t statistic, the linear component for chi-square and the Mantel test are essentially all the same, although they have somewhat different justifications and at first sight the necessary assumptions for the χ^2 tests look much less demanding than those for the t^2. The point that we wish to make is that the linear component of overall χ^2 which extracts the component of interest more efficiently and simply than accumulation analysis is precisely what we get from a straightforward analysis of variance or t analysis made on the scores.

Choice of Scoring Procedure

Equispaced scoring of ordered categories will not always be appropriate. Although moderate changes in the choice of the scores are unlikely to produce large differences in the final analysis, subject-matter knowledge may sometimes suggest a different allocation. In doubtful cases, as when different scores are suggested by different people of apparently equal discernment, analyses may be conducted with alternative scoring systems. A discriminant method for automatically assigning scores was proposed by Fisher (1963) for the analysis of his agglutination data referred to earlier. He also proposed a test for whether the discriminant scores differed significantly from the equispaced score. If we apply his technique to the data for drug B, for example, we obtain the scores shown in Table 7.

	Drug B			
Class	–	+	++	+++
Score	0	.63	.88	1

Table 7. The Scores for Drug B from Fisher's Method

The results of using this scoring system are not very different from those using the equispaced scores, nor, using Fisher's test, are the discriminant scores significantly different from the equispaced scores. The logic of the discriminant method, however, seems somewhat dubious to us and we would prefer scores based on human judgment.

6. The Window Size Example

The investigation conducted by Phadke et al. (1983) on integrated circuit fabrication is most valuable and of general interest. However, the post-etch window size data for that study, which are given in Table 1 of Nair (1986), possess rather peculiar characteristics.

(1) This set of data has been made categorical artificially. The actual measurements of window size are available (see Phadke et al. (1983)) and should, we feel, have been the quantities used for analysis. Splitting into categories (presumably so that accumulation analysis can be used) inevitably obscures information. In particular, different results from accumulation analysis and other methods will be obtained from different classifications and for different numbers of categories.

(2) A priori considerations, as well as the data themselves, suggest that the issue of *whether a window is open at all* ought to be separately studied. This can be done by noting the proportion open in each experimental run and using the usual arc sine transformation for variance stabilization. The results of such an analysis are shown in Table 8(a).

(3) The stated problem is to obtain windows of size as close as possible to 3μm. It is not clear to us how accumulation analysis, or for that matter the isolation of location and dispersion effects or of linear and quadratic components, contributes to solving this problem. A more direct approach would be to analyze, for those windows that are open, the mean squared error deviation from 3μm. (or more appropriately its log). Also, since mean squared error can be reduced both by making the mean closer to target and by reducing the variance, we have analyzed the mean and variance separately. Such an analysis is shown in Table 8(b) carried through for the original uncategorized data. This should be regarded as an adjunct to the analysis given in Table 8(a).

(4) Supposing that we could agree on some specific characteristic and an appropriate criterion to measure it, should we care about attempting to distinguish signal from noise using standard errors, Daniel plots, significance tests or Bayes plots? An argument, repeated in Nair (1986), has been advanced that we should not. This question was considered by one of us some years ago (Box (1966); see also Box and Ramírez (1986)). The conclusion was that if the only result an experiment could have would be the empirical adjustment of levels in the directions indicated to achieve an "optimal" combination, then testing procedures would be of little importance. For this to be true, however, the assumption must be made that no contribution is to be expected from the engineer as a result of his specialist knowledge and his thinking about the results (that is, we expect to proceed entirely by *empirical* feedback without *scientific* feedback).

Our experience is that statistical results studied by an expert in the field can suggest ideas not accessible by statistical reasoning alone that are often the most fruitful for further development. Because they are unexpected, they are also most likely to lead to advances that outdistance the competition. Statistical methods should, we feel, be regarded as a catalyst not a substitute for engineering science.

But if the engineer is to reason in this way about the effects (usually for several different responses) in a multifactor experiment, he must be given some way of understanding what effects could well be due to chance and what effects are almost certainly "real". Otherwise he will waste his time attempting to explain the noise and so, like the King in *Alice in Wonderland* explaining the White Rabbit's poem, may lead the whole investigation astray.

(5) Thus, as was emphasized by R. A. Fisher some 60 years ago, we need procedures for comparing an apparent signal with a relevant measure of noise. Approximation is acceptable and indeed inevitable, but bias such as induced by pooling the smallest mean squares to obtain an error term in an analysis of variance must surely be avoided.

After first scaling effects so that they have the same standard deviation we analyzed all of the data in Phadke et al. (1983) using Daniel plots (and more recently Bayes plots). For most of the data in this article, however, an additional simple procedure is available. The tacit assumption was made in this experiment that second-order effects due to interactions were unimportant. It seems logical, therefore, to consider an analysis based on the hypothesis that the remaining second-order effects from quadratic components are unimportant also. Daniel plots, not only for window size but also for the other responses in this experiment, do indeed show all quadratic effects plotting as noise. In Table 8(a) therefore, we have pooled these additional quadratic effects to produce a valid standard error having 8 degrees of freedom from which the t values

have been calculated. For the analysis of log mean squared deviation from 3μm., only 12 runs were available for data with windows open. We have analyzed the data using regression analysis. The t values computed for the analysis are shown in Table 8(b).

		(a) Proportion of windows not open	(b) Analysis of window size log mean square deviation	Mean $\bar{y}$	log s
Mask Dimension	(A)	1.10	0.39	1.33	0.15
Viscosity	(B)	2.48	0.85	-1.26	-0.19
Spin Speed	(C)	-2.70	-0.22	1.62	-0.63
Bake Temperature	(D)	0.55	-0.75	0.16	-0.18
Bake Time	(E)	-0.13	-0.49	-0.10	-0.44
Aperture	(F)	-1.19	-0.18	0.20	0.47
Exposure Time	(G)	1.59	0.10	-0.99	0.13
Developing Time	(H)	-1.27	-2.55	1.93	-1.33
Plasma Etch Time	(I)	-0.77	-1.37	0.31	0.18

Table 8. Values of t for Post-etch Window Size Data (Nair, Table 1)

Our conclusions from this analysis, which, since this is a highly fractionated design would need to be tentative and subject to confirmation, would be as follows: Increasing the spin speed (C) reduces the proportion of closed windows. Low viscosity (B) reduces the proportion of closed windows. Use of the longest developing time (H) reduces the mean squared deviation from the target so that at the highest level the window size is close to target. Also shown in Figures 1 and 2 are the corresponding Daniel plots. They tend to confirm the results while underlining the rather tenuous nature of the conclusions about the proportion of closed windows and the desirability of confirmation experiments.

Figure 1

Standardized effects for proportion of windows not open

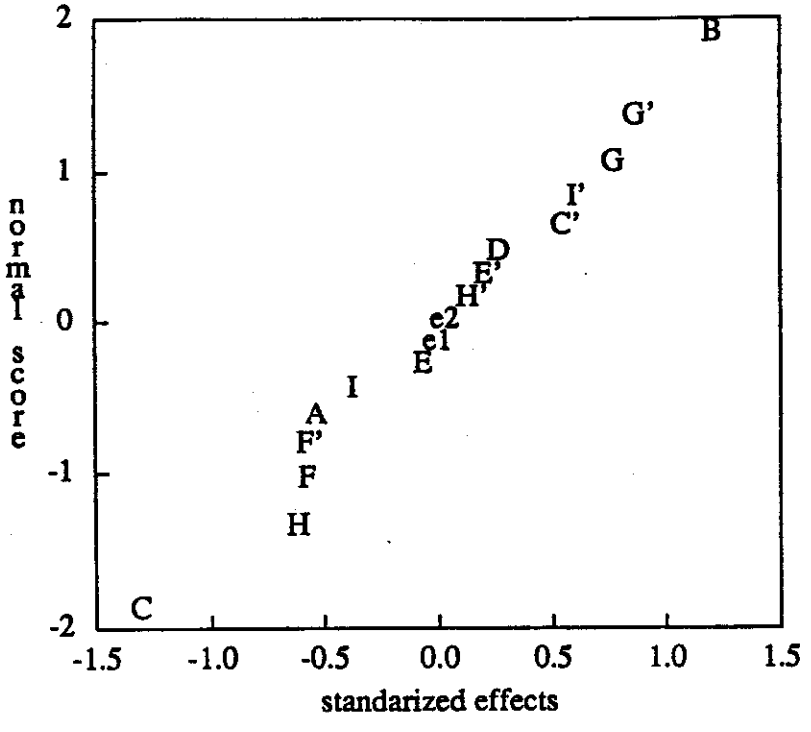
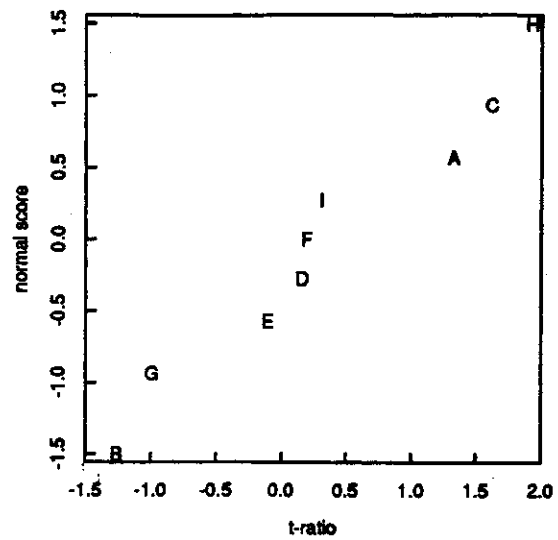
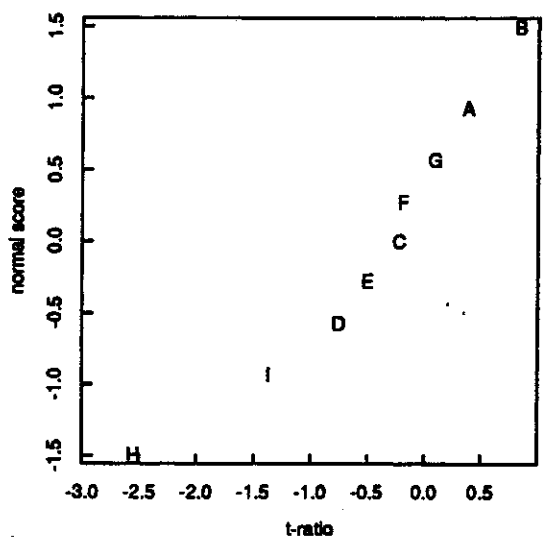


Figure 2 Normal Plots of Analysis of Window Size

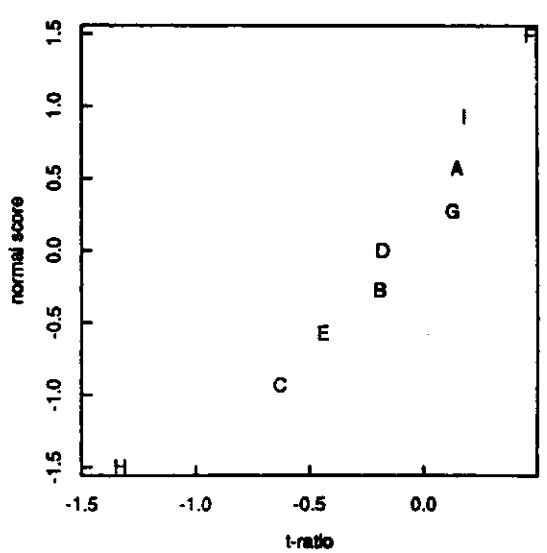
Mean $\bar{y}$



Log(mean squared deviation)



Log(s)



7. Conclusions

If we believe that our industry is presently at a serious competitive disadvantage in the quality of the products that it produces, then we must try to do better than our competition. To do this

(1) we must see that all useful techniques for quality improvement, and in particular those using statistical design and data analysis, are at last put to use in our industry.

(2) we must be ready to adopt all useful new ideas.

(3) we at the research institutions must, in cooperation with industry, continuously strive in our research to come up with even better concepts and ideas. If we always follow we will always be behind.

So far as the adoption of new ideas is concerned, Taguchi has made major contributions to quality engineering, particularly in his emphasis on (a) "closeness to target" versus "within specification," (b) using experimental design to minimize variance and (c) using experimental design to obtain products that are insensitive to environmental factors and to transmitted variation from components.

Nevertheless, we believe that his accumulation analysis is not a useful technique and that to attempt to teach it to engineers would be a serious mistake. Since the procedure is complicated, it could only be justified if it possessed clear advantages over alternatives. The fact that it does better against directional alternatives than Pearson's chi-square test is hardly surprising, since the latter is not a directional test. When compared with much simpler directional procedures, accumulation analysis can be expected to do poorly. So far as location effects are concerned, although it places moderate emphasis on something close to the linear component, it

dilutes that component with unnecessary noise. Furthermore, the weighting of this component seems to depend on matters that are irrelevant and can better be decided a priori. When we are interested in both location and dispersion effects, we would surely wish to avoid confounding the two, but this is what accumulation analysis does.

For ordered categorical data, we believe that the engineer should use his expertise to devise a simple scoring system and then analyze his results in the usual way. When the designs are fractional factorials or other orthogonal arrays an analysis with Daniel plots is usually adequate. Daniel plots are simpler than the analysis of variance and avoid the bias introduced by the pooling of mean squares which is recommended by Taguchi. If desired, Daniel plots may be supported by Bayes probability plots for active effects (see Box and Meyer (1986)).

We are doubtful whether accumulation analysis or the alternatives suggested by Nair (1986) are appropriate for the analysis of the window size data. Instead we have demonstrated a simpler approach that directly addresses the questions at issue.

We need to teach our engineers simple, reasonably efficient statistical techniques that can catalyze investigations in such a way that they are problem driven rather than technique driven. We should not be teaching them portmanteau methods that are complicated, lack efficiency and confound different aspects of the data.

Acknowledgement

Sponsored by the National Science Foundation Grant No. DMS-8420968, by the United States Army under Contract No. DAAG29-80-C-0041, and by the Vilas Trust of the University of Wisconsin-Madison; and was aided by access to the Statistics Department Research Computer at the University of Wisconsin-Madison.

References

Box, G.E.P. (1966), "A Simple System of Evolutionary Operation Subject to Empirical Feedback," *Technometrics*, 8, 19-26.

Box, G.E.P., and Ramírez, J.G. (1986) "Studies in Quality Improvement: Signal to Noise Ratios, Performance Criteria, and Statistical Analysis (Part II)." Report # 12, Center for Quality and Productivity Improvement, University of Wisconsin-Madison.

Box, G.E.P., and Meyer, R.D. (1986), "An Analysis for Unreplicated Fractional Factorials," *Technometrics*, 28, 11-18.

Cochran, W.G. (1950), "The Comparison of Percentages in Matched Samples," *Biometrika*, 37, 256-266.

Daniel, C. (1959), "Use of Half-Normal Plots in Interpreting Factorial Two Level Experiments," *Technometrics*, 1, 311-342.

Davies, O.L. (1958), *Statistical Methods in Research and Production*, New York: Hafner Publishing Co..

Fisher, R.A. (1963), *Statistical Methods for Research Workers*, Edinburgh: Oliver and Boyd. (13th edition)

Hamada, M., and Wu, C.F.J. (1986) "A Critical Look at Accumulation Analysis and Related Methods." Report # 20, Center for Quality and Productivity Improvement, University of Wisconsin-Madison.

Mantel, N. (1963), "Chi-Square Tests with One Degree of Freedom; Extensions of the Mantel-Haenszel Procedure," *Journal of the American Statistical Association*, 58, 690-700.

Nair, V. N. (1986), "Testing in Industrial Experiments with Ordered Categorical Data," *Technometrics*, 28, 283-291.

Phadke, M. S., Kackar, R. N., Speeney, D. V., and Grieco, M. J. (1983), "Off-Line Quality Control in Integrated Circuit Fabrication Using Experimental Design," *The Bell System Technical Journal*, 62, 1273-1310.

Snedecor, G. W., and Cochran, W. G. (1980), *Statistical Methods*, Ames, Iowa: The Iowa State University Press. (Seventh Edition)

Taguchi, G. (1974), "A New Statistical Analysis for Clinical Data, the Accumulating Analysis, in Contrast with the Chi-Square Test," *Saishin Igaku (The Newest Medicine)*, 29, 806-813.

Takeuchi, K., and Hirotsu, C. (1982), "The Cumulative Chi-Squares Method Against Ordered Alternatives in Two-Way Contingency Tables," *Rep. Stat. Appl. Res., JUSE*, 29, 1-13.