

ONE DEGREE OF FREEDOM FOR NON-ADDITIVITY*†

JOHN W. TUKEY

Princeton University

INTRODUCTION

IN DISCUSSING the possible shortcomings of the analysis of variance, much attention has been paid to non-constancy and non-normality of the "error" contribution. (The recent papers in *Biometrics* by Eisenhart [4], Cochran [3] and Bartlett [1] discuss these matters and give references.) The present writer is usually much more concerned with and worried about non-additivity, and until recently has suffered from the lack of a systematic way to seek it out, and then to measure it. (Conversations with Frederick F. Stephan have contributed greatly to this development and presentation.)

The purpose of the present paper is to indicate such a way, when the data is in the form of a row-by-column table. (The professional practitioner of the analysis of variance will have no difficulty in extending the process to more complex designs.) We shall show how to isolate one degree of freedom from the "residue", "error", "interaction" or "discrepancy", call it what you will. There are two known situations to which this single degree of freedom is expected to react by swelling:

- (1) when one or more observations are unusually discrepant;
- (2) when the analysis has been conducted in terms where the effects of rows and columns are not additive.

The first situation is quite familiar and requires little explanation. The second occurs often enough, but may not be noticed. An example may help to fix the ideas.

Let us construct an artificial example with 3 rows and 4 columns, with each entry contributed to overall, by rows, by columns, and by cells. Suppose that these contributions are as follows:

*Prepared in connection with research sponsored by the Office of Naval Research.

†Presented to the Biometrics Section and the Biometric Society at Cleveland, December 29, 1948.

Construction of the example (don't worry about this)

Starting point

Add a row effect

Add a col effect

Add a 'cell' effect

ONE DEGREE OF FREEDOM FOR NON-ADDITIVITY

233

in general

by rows

by columns

by cells

1	1	1	1
1	1	1	1
1	1	1	1

4	4	4	4
-3	-3	-3	-3
-3	-3	-3	-3

6	1	-4	0
6	1	-4	0
6	1	-4	0

1	-2	1	0
0	-1	2	-3
0	-2	-1	0

Then the tables and corresponding analyses for the sum of all contributions are:

TABLE 1
ILLUSTRATIVE EXAMPLE IN ORIGINAL TERMS

		Values and Means				Analysis of Variance					
		12	4	2	5	23	5.8		DF	SS	MS
		4	-2	-4	-5	-7	-1.8				
		4	-3	-7	-2	-8	-2.0				
Sums		20	-1	-9	-2	8		Rows	2	140	70
Means		6.7	-0.3	-3.0	-0.7		0.7	Columns	3	157	52
								R × C	6	26	4

Now let us square the entries and divide by 10, rounding to integers. The resulting tables and analyses are:

The final example (also shown in Table3)

TABLE 2
ILLUSTRATIVE EXAMPLE IN TERMS OF SQUARES

		Values and Means				Analysis of Variance					
		14	2	1	2	19	4.8		DF	SS	MS
		2	0	2	2	6	1.5				
		2	1	5	0	8	2.0				
Sums		18	3	8	4	33		Rows	2	24.5	12.2
Means		6.0	1.0	2.7	1.3		2.8	Columns	3	46.9	15.6
								R × C	6	84.8	14.1

Notice that all semblance of row or column effects have now vanished, although Table 1 showed large and significant effects. The use of the squared scale has concealed the real effects. (It may be argued that squaring numbers which range from plus to minus is unrealistic. The answer is that this *is* an extreme example, but one that can be slowly and smoothly changed into a very mild one. There probably is a differ-

ence in degree between this example and what happens in practice, but there is no difference in kind.)

PROCEDURE

How then do we isolate the single degree of freedom? The process is simple, and runs as follows:

- (A) To the row-by-column table, already bordered with sums and means, add a new border of deviations of means from the grand mean (decimal places may be reduced, but the sums of deviations, by rows and by columns *must* be forced to vanish).
- (B) Add an extra column (or row) and enter in each cell the sum of products of the deviations by columns and the entries in its row (or column).
- (C) Accumulate the sum of products between the deviations of row (or column) means and the new entries of (B).
- (D) Calculate the sum of squares of deviations by columns and by rows.
- (E) Divide the square of the number from (C) by the product of the numbers from (D). This is the mean square (and also the sum of squares) for the single degree of freedom.

The process is illustrated on the same example below:

TABLE 3
SAMPLE CALCULATION

					Sums	Means	Deviations	Sums of x-products
	14	2	1	2	19	4.75	2.0	38.4
	2	0	2	2	6	1.50	-1.2	3.6
	2	1	5	0	8	2.00	-0.8	4.6
Sums	18	3	8	4	33		0.0	68.8
Means	6.00	1.00	2.67	1.33		2.75	6.08	
Deviations	3.2	-1.8	0.0	-1.4	0.0	15.44	50.9	

$$(B): 14(3.2) + 2(-1.8) + 1(0.0) + 2(-1.4) = 38.4$$

$$2(3.2) + 0(-1.8) + 2(0.0) + 2(-1.4) = 3.6$$

$$2(3.2) + 1(-1.8) + 5(0.0) + 0(-1.4) = 4.6$$

$$(C): 38.4(2.0) + 3.6(-1.2) + 4.6(-0.8) = 68.8$$

$$(D): (3.2)^2 + (-1.8)^2 + (-0.0)^2 + (1.4)^2 = 15.44$$

$$(2.0)^2 + (-1.2)^2 + (-0.8)^2 = 6.08$$

$$\frac{(68.8)^2}{(15.44)(6.08)} = 50.9.$$

Assigning the mean square 50.9 to the degree of freedom for non-additivity, which is subtracted from "R × C", the analysis of variance of Table 2 becomes:

Rows	2	24.5	12.2
Columns	3	46.9	15.6
Non-additivity	1	50.9	50.9
Balance	5	33.9	6.8

Thus the obvious thing about the illustrative example was its non-additivity. The corresponding *F* value of 7.3 on 1 and 5 degrees of freedom is significant at the 5% level.

EXPLANATION

We have explained what we are looking for—non-additivity—and how to look—last section—but we have not explained what we are really doing. This we shall now try to do. Those experienced with single degrees of freedom may have already recognized the computation as a short-cut method of eliminating the single degree of freedom labeled by

$$\begin{bmatrix} 6.40 & -3.60 & 0.00 & -2.80 \\ -3.84 & 2.16 & 0.00 & 1.68 \\ -2.56 & 1.44 & 0.00 & 1.12 \end{bmatrix} = \begin{bmatrix} 2.0 \\ -1.2 \\ -0.8 \end{bmatrix} \cdot \begin{bmatrix} 3.2 & -1.8 & 0.0 & -1.4 \end{bmatrix}$$

where 6.40 = (2.0)(3.2), -3.60 = (-1.8)(2.0), 2.16 = (-1.8)(-1.2) and so on. We have used the products of the deviations of the row means and the deviations of the column means to label this single degree of freedom. Since the sum of each column and of each row is zero, this degree of freedom is orthogonal to rows and to columns. It must be a part of "R × C". This is what we did, but why?

Let us take a special case, where there are row contributions, and column contributions, *and nothing else*. We start with perfect additivity. If x_i is the column contribution (where i goes from 1 to c , the number of columns), and if y_j is the row contribution (where j goes from 1 to r , the number of rows), then the ij entry in the table is

$$a_{ij} = x_i + y_j .$$

Now let us start to analyze a slightly nonlinear function of the a_{ij} . Instead of a_{ij} , consider

$$f_\lambda(a_{ij}) = a_{ij} + \lambda(a_{ij} - a)^2$$

where λ is a small constant, and a is, for convenience, the average $\bar{x} + \bar{y}$ of all the a_{ij} . We find that we can write

$$f(a_{ij}) = [x_i + \lambda(x_i - \bar{x})^2] + [y_j + \lambda(y_j - \bar{y})^2] + \lambda(x_i - \bar{x})(y_j - \bar{y}).$$

The first two terms depend, respectively, on the column alone and on the row alone, so the last one contains all the *non-additive* effect due to analysis in terms of $f(a)$ instead of in terms of a . Notice that this non-additive effect is a multiple of

$$(x_i - \bar{x})(y_j - \bar{y}).$$

This means that it occurs in a single degree of freedom, which is identified in terms of $x_i - \bar{x}$ and $y_j - \bar{y}$.

We assumed no error of measurement, or the like, and we wrote $a_{ij} = x_i + y_j$ without an additional term. This means that the difference between the i -th column mean and the grand mean is

$$(x_i - \bar{x}) + \lambda\{(x_i - \bar{x})^2 - \overline{(x_i - \bar{x})^2}\}$$

which is nearly $x_i - \bar{x}$ when λ is small. Thus a satisfactory approximation to the single degree of freedom we want is that indicated by the coefficients

$$(\text{column mean} - \text{grand mean})(\text{row mean} - \text{grand mean}).$$

This is exact for the combination of no error and a very slight change from a to $f(a)$, that is for no error and λ small. This fact plus empirical tests seems enough to warrant recommending general use of this single degree of freedom as a test of non-additivity.

WHAT OF SIGNIFICANCE?

Suppose that the test shows statistically significant evidence of non-linearity—what then? The simplest and laziest thing to do would

be to forget the degree of freedom for non-additivity and go on and use the mean square for the balance in considering for example, the significance of the row effects. *This is not recommended*, for the following reasons:

- (1) In general, results expressed in terms in which effects are additive apply in a broader region and are practically more useful.
- (2) If the "error" or fluctuating contribution is not normally distributed, then it is not known whether or not the use of the balance mean square unduly inflates the apparent significance of other mean squares (for the case of a normally distributed fluctuating contribution there is no distortion of significance.)

For these reasons, the occurrence of a large non-additivity mean square should lead to consideration of a transformation followed by a new analysis of the transformed variable.

This consideration should include two steps:

- (a) inquiry whether the non-additivity was due to analysis in the wrong form or to one or more unusually discrepant values;
- (b) in case no unusually discrepant values are found or indicated, inquiry into how much of a transformation is needed to restore additivity.

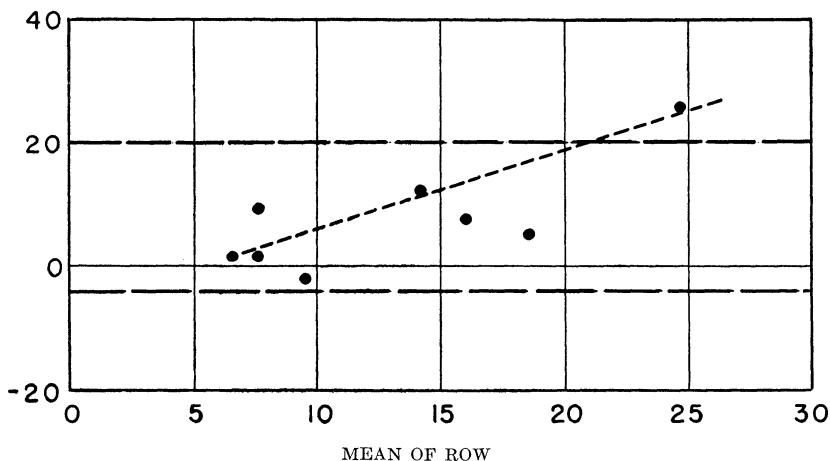
The decision under (a) will depend on an examination of the data and all the background information available in the field—in particular the result of similar inspections of other experiments for non-additivity. What seems to be the best way of inspecting the results of a single experiment so far proposed is to plot the entries in the new column (of sums of cross-products) against the corresponding row means. A single unusually discrepant observation will tend to be reflected by one point high or low and the others distributed around a nearly horizontal regression line. An analysis in the wrong terms will tend to be reflected by a slanting regression line.

The figure shows such a plot, including $2s$ limits, for

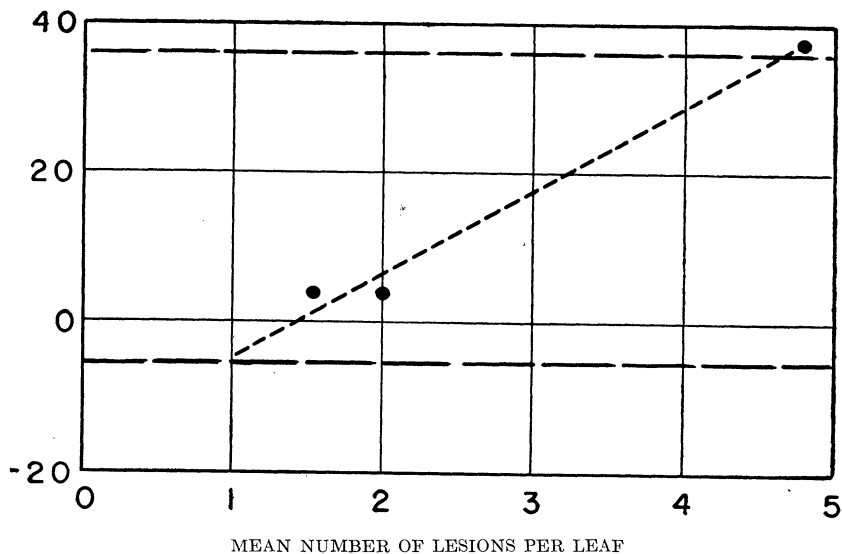
- (A) the illustrative example worked above,
- (B) Youden and Beale's data [6] as simplified by Snedecor [5, p. 44],
- (C) Beall's experiment VI [2] on insect infestation, with plots

GRAPHICAL ANALYSIS OF NONADDITIVITY
 (Ordinates are Sums of Cross Products, Dashed Lines are $2S$ Limits)

A—ILLUSTRATIVE

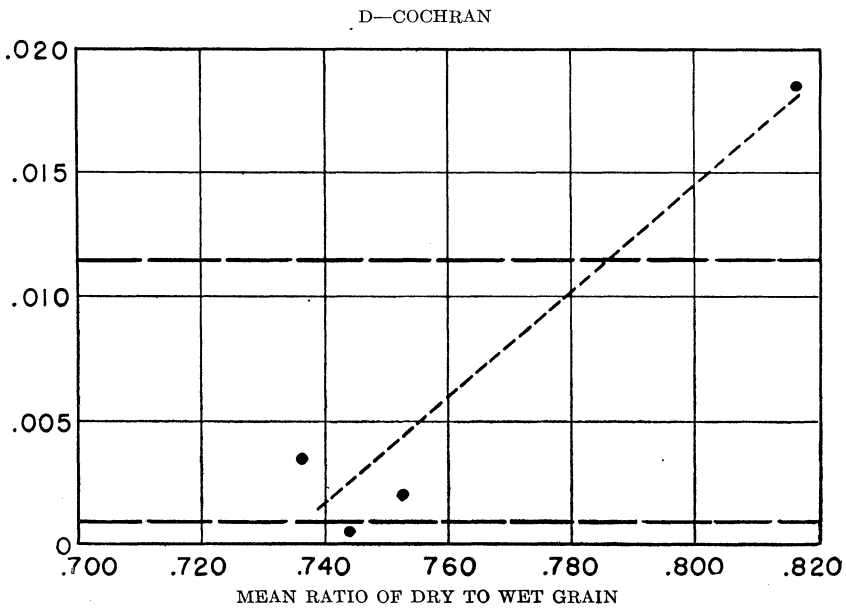
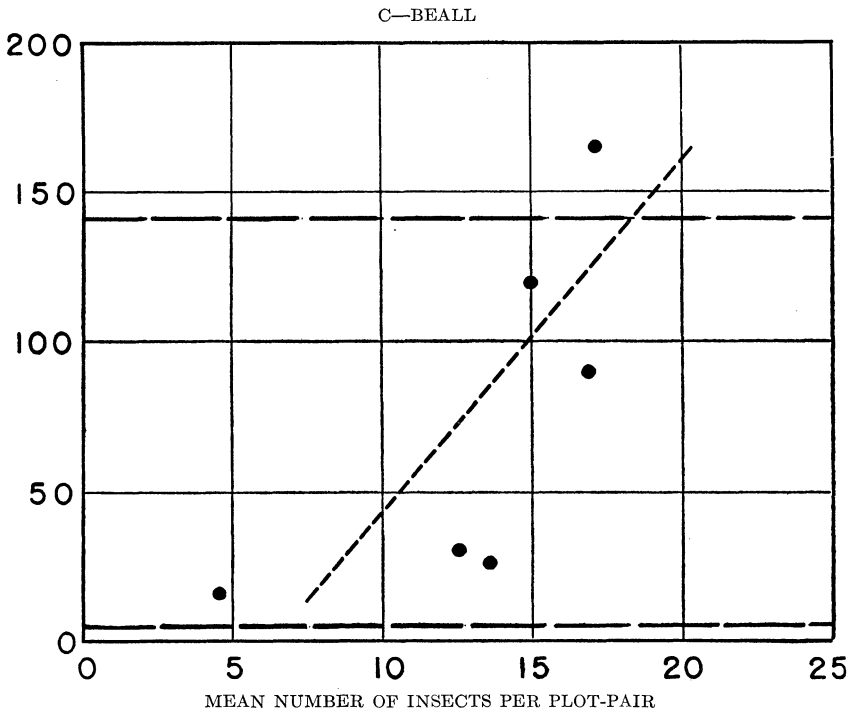


B—YOUDEN & BEALE



treated alike combined (analyzed in terms of numbers of insects).

(D) Cochran's example [3] of an obviously discrepant value.



The limits are set by the formula

$$\left(\begin{array}{c} \text{average} \\ \text{cross product} \end{array} \right) \pm 2 \left(\begin{array}{c} \text{sum of squares of} \\ \text{deviations of column means} \end{array} \right)^{\frac{1}{2}} \left(\begin{array}{c} \text{mean square} \\ \text{for balance} \end{array} \right)^{\frac{1}{2}}$$

For the illustrative example (Case A), this becomes

$$15.53 \pm 2 (15.44)^{\frac{1}{2}}(6.8)^{\frac{1}{2}} = 15.5 \pm 20.5 = -5.0 \text{ and } +36.0.$$

In every one of the four cases, the plotted points could be accounted for by non-additivity due to analysis in incorrect terms. Cases A and D can also be accounted for by a discrepant point. This suggests that it will be hard to make this distinction for single experiments on this scale. When several small experiments are available for analysis, agreement in signs of the slopes of the graphs or equivalently, the signs of the sums obtained in Step C may show up analysis in incorrect terms.

Why does the graph fail to decide about Cases A and D? The reason is simple—either explanation is plausible. If in Case A we alter the upper left-hand entry from 14 to 2, the analysis of variance becomes:

	DF	SS	MS
Rows	2	0.5	0.2
Columns	3	4.9	1.6
Non-additivity	1	0.2	0.2
Balance	5	12.6	2.5

Thus we see that our illustrative table of 3×4 entries *could* have perfectly well come from an additive situation where exactly one entry has been seriously disturbed.

Similarly in Case D, taken from Cochran's paper, if a nonlinear function is chosen so that

$$g(y) = \begin{cases} y, & .704 \leq y \leq .792, \\ .800, & y = 1.035, \end{cases}$$

then his table is converted into one where the F -ratio for non-additivity against balance is 0.8 instead of 27.6. We know that this table arose from an error in computation, but it *could* equally well have come from an additive table analyzed in the wrong terms.

In each case, the graphical solution has gone as far as it reasonably

could in assigning responsibility for the non-additivity. While the graphical analysis is not certain to settle Step (a), it may be expected to be a big help.

AID IN CHOOSING A TRANSFORMATION

If it has been decided that the wrong terms had been used, then the actual size of the mean square for non-additivity must be useful for choosing an appropriate transformation. We lack experience with the more delicate use of such information, so that it seems appropriate to stop here with the following table which shows the connection between the *sign* of the final sum of products (which was +68.8 in the illustrative example) and the type of transformation which may then be appropriate.

TABLE 4
SIGN OF FINAL SUM OF PRODUCTS WHEN CERTAIN TRANSFORMATIONS ARE APPROPRIATE (VALUES OF x OR $x + a$ NON-NEGATIVE)

Transformed values which are additive*	Conditions needed	Sign when x is analyzed	Important special cases
x^p or $(x + a)^p$	$0 \leq p < 1$	+	$\sqrt{x}, \sqrt{x + 1}$
	$p = 1$	0	(x)
	$1 < p$	-	x^2, x^3
$\log(x + a)$	(none)	-	$\log x, \log(1 + x)$

*Multiplication by a fixed constant and addition or subtraction of a fixed constant freely possible

While the removal of non-additivity by transformation usually tends to stabilize the variance, there may be cases where the variance is notably non-constant after transformation. In such cases, analysis of the transformed data using weights seems appropriate.

APPENDIX

VALIDITY OF THE ANALYSIS

This section is prepared for those who may feel that the method of obtaining the "single degree of freedom" may not produce quantities with the usual distribution.

The basic fact is this: If $u_1, u_2, \dots, u_k; v_1, v_2, \dots, v_m$ have some

joint distribution, and if, for fixed u_1, u_2, \dots, u_k , the distribution of v_1, v_2, \dots, v_m exists and is *always the same*, then the marginal distribution of v_1, v_2, \dots, v_m exists and, indeed, is the same, and, furthermore, u_1, u_2, \dots, u_k and v_1, v_2, \dots, v_m are independent. This can be established either by general considerations or by analytical detail.

To apply this in our case, let u_1, u_2, \dots, u_k be the row and column means, and let v_1 and v_2 be the sums of squares for non-additivity and for the balance. If the situation is additive, and the cell effects are normally distributed, and u_1, u_2, \dots, u_k are fixed, then v_1 and v_2 are independently distributed like σ^2 times chi-squares on 1 and $rc - r - c$ degrees of freedom. Hence v_1 and v_2 have these distributions, and are independent of all functions of row and column means. Thus the F -tests of rows, columns, or non-additivity against balance are valid.

In the presence of non-additivity and/or non-normality, the usual arguments indicate that the F -test is, if anything, conservative.

REFERENCES

- [1] Bartlett, Maurice S. The Use of Transformations. *Biometrics* 3, 39-57, 1947.
- [2] Beall, Geoffrey. The Transformation of Data from Entomological Field Experiments so that the Analysis of Variance becomes Applicable. *Biometrika* 32, 243-262, 1942.
- [3] Cochran, W. G. Some Consequences when the Assumptions for the Analysis of Variance are not Satisfied. *Biometrics* 3, 22-38, 1947.
- [4] Eisenhart, Churchill. The Assumptions underlying the Analysis of Variance. *Biometrics* 3, 1-21, 1947.
- [5] Snedecor, George. *Statistical Methods*. The Collegiate Press, Ames, Iowa; 4th edition, 1947.
- [6] Youden, W. J. and Beale, Helen Purdy. A Statistical Study of the Local Lesion Method for Estimating Tobacco Mosaic Virus. *Contributions from the Boyce Thompson Institute* 6, 437-454, 1934.