Research Note

ANOVA PROCEDURES IMPROVED[1]

The variance of the data collected in an experiment may be analyzed by the method of fitting constants. This method entails the formulation of an estimation equation, the preparation of observation and normal equations, the solution of the latter by inversion of their matrix of coefficients, the evaluation of the statistics representing the effects of the independent factors and the estimation of the variances of the statistics and/or of their differences.

The ANOVA procedures in current use eliminated the need to invert the above-mentioned matrices, since at the time of their development there were no high-speed computers available to invert the matrices in a reasonably short time. However, when the balance or partial balance of an experimental design is altered by missing values, in order to use these procedures there is need to estimate such missing values before proceeding with the analysis of the data. As a matter of fact, these procedures are obsolete because of availability of high-speed computers.

Furthermore, since the matrix of coefficients of a set of normal equations and their inverts are definite for any given design, irrespective of the values of the dependent factor, it is possible and convenient to calculate in advance the values of the elements of the inverts (Gauss multipliers) of the experimental designs most commonly used. This would allow starting the analysis at that advanced stage of the process. The time and effort required to perform the analysis starting at this stage are considerably less than those required by the procedures in common use.

On calculating the Gauss multipliers of various experimental designs, the author found that the multipliers of a design may be expressed as simple rational fractions with a common denominator. Thus, to analyze the data, the required operations may be performed by considering only the numerators of said fractional multipliers, making the corresponding divisions later on in the analysis. Furthermore, the values of the multipliers of the different sizes of a given design may be summarized into relatively simple formulas. The formulas may be used to calculate the values of the multipliers which correspond to any given size of the design.

With these formulas or the tables of multipliers prepared from them, it is thus possible, even advantageous, to analyze experimental data by the method of fitting constants, without preparing the normal equations, and of inverting their matrix of coefficients, that is, by calculating directly

[1] Manuscript submitted to Editorial Board February 18, 1981.

the effects of the various levels of the independent factors from the multipliers and the corrected sums of products with the values of the dependent factor. The sum of squares explained by the independent factors, whether individually or jointly, may then be calculated from these effects and the corrected sums of products and, by subtraction from the total sum of squares, the error sum of squares is obtained as usual.

In balanced or partially-balanced designs, the corrected sums of products of the qualitative independent factors, with the values of the dependent factor, may be calculated by simple substractions, a fact which still further simplifies the analysis.

The analysis of experimental data by the proposed modification of the method of fitting constants is based on the assumption of constant effects of the levels of the independent factors. When the estimation equation expresses these effects as deviations from the mean, for each qualitative factor the estimation equation includes one less constant than the number of levels of the factor.

TABLE 1.—*Calculation of the corrected sums of products and the total sum of squares*

	T_1	T_2	T_3	Total	Sby	Squares		
B_1	5	7	4	16	3	25	49	16
B_2	3	6	4	13	0	9	36	16
Total	8	13	8	29		34	85	32
Sty	0	5	0					

Grand total = 151

Correction factor = $(29)^2/6$ = 140 1/6

Syy = 10 5/6

First example (Complete blocks)

In the case of the 3-treatment 2-block complete block design, the estimation equation is

$$y'_{tb} = x_{t1}T1 + x_{t2}T2 + x_{b1}B1,$$

where y'_{tb} is the estimated deviation of the observation on the dependent factor from the general mean; *T1, T2* and *B1* represent both the corresponding treatments or blocks and their corresponding effects as deviations from the mean; and x_{t1}, x_{t2} and x_{b1} are equal to 1 when the observation corresponds to them, or zero if not.

The 6 observation equations in this case are

$$\begin{aligned} T_1 + B1 &= y'_{11}, \\ T_1 - B1 &= y'_{12}, \\ T_2 + B1 &= y'_{21}, \\ T_2 - B1 &= y'_{22}, \\ -T_1 - T_2 + B1 &= y'_{31}, \\ \text{and} \quad -T_1 - T_2 - B1 &= y'_{32}. \end{aligned}$$

From these, the 3 normal equations are

$$(4T1 + 2T2 + 0B1 = Sx_{t1}y = (y_{11} + y_{12}) - (y_{31} + y_{32}) = (Y_{11} + Y_{12}) - (Y_{31} + Y_{32}),$$

$$(2T1 + 4T2 + 0B1 = Sx_{t2}y = (y_{21} + y_{22}) - (y_{31} + y_{32}) = (Y_{21} + Y_{22}) - (Y_{31} + Y_{32}),$$

and

$$0T1 + 0T_2 + 6B1 = Sx_{b1}y = (y_{11} + y_{21} + y_{31}) - (y_{12} + y_{22} + y_{32}) = (Y_{11} + Y_{21} + Y_{31}) - (Y_{12} + Y_{22} + Y_{32}).$$

The invert of the matrix of coefficients of this set of normal equations is

	$T1$	$T2$	$B1$
$T1$	2/6	−1/6	0
$T2$	−1/6	2/6	0
$B1$	0	0	1/6

The values of these elements are constant regardless of the values of the dependent factor. Therefore, these values need not be recalculated

TABLE 2.—*Calculation of effects and their contribution to Sy'y'*

	T_1	T_2	B_1	Sxy	Effects	Contribution to Sy'y'
	Common denominator = 6					
T_1	2	−1	0	0	$T_1 = -5/6$	0
T_2	−1	2	0	5	$T_2 = 10/6$	50/6
B_1	0	0	1	3	$B_1 = 3/6$	9/6
					Total Sy'y' = 59/6 = 9 5/6	

TABLE 3.—*ANOVA*

Source of variation	df	SS
Blocks	1	1.5000
Treatments	2	8.3333
Error	2	1.0000
Total	5	10.8333
	Ve = 1/2 = 0.5000	

for any 3-treatment 2-block complete block design. Adding a column with the corrected sums of products to the tabulation of the invert presented above facilitates the calculation of the treatment and block effects, and of their contribution to the sum of squares explained by said effects.

Thus, if $Y_{11} = 5$, $Y_{12} = 3$, $Y_{21} = 7$, $Y_{22} = 6$, $Y_{31} = 4$ *and* $Y_{32} = 4$, the analysis would be conveniently performed as shown in tables 1 through 3:

By definition, $T_3 = -(-5/6 + 10/6) = -5/6$.

The variance of the difference between any 2 treatment effects is

$$(C_{ii} - 2Cij + Cjj)Ve = ((2 - 2\ (-1) + 2)/6)0.5000 = 0.5000.$$

Second example (Partially-balanced regular divisible design R 27).

TABLE 4.—*Per plot yield data and calculation of the corrected sums of products of the treatment and block levels*

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}	T_{13}	T_{14}	T_{15}	Total	Sx_{b_y}
B_1	26								25				20		24	95	−17
B_2	27				27		28	24								106	−6
B_3	28	22								26				24		100	−12
B_4		21	23								31				34	109	−3
B_5				33		41	29								33	136	24
B_6	30		28	32								34				124	12
B_7								26				32		25	24	107	−5
B_8			23		27	23								24		97	−15
B_9		26		28	28								25			107	−5
B_{10}						26				25		27	28			106	−6
B_{11}			24				26		26	23						99	−13
B_{12}		25				27		27	26							105	−7
B_{13}					30				36		32	32				130	18
B_{14}							30				25		28	24		107	−5
B_{15}				25				32		24	31					112	0
Total	111	94	98	118	112	117	113	109		98	119	125	101	97	115	1640	
	−4	−21	−17	3	−3	2	−2	−6	113	−17	4	10	−14	−18	0		
Sx_{t_y}									−2								

Table 4 presents the yield data (multiplied by 10) and the calculation of the corrected sums of products with the values of the dependent factor of a 15-treatment 15-block experiment presented by Bose et al.[2]

For this type and size of design, the common denominator of the Gauss multipliers is 900. Table 5 shows also the values of the numerators of the multipliers, the corrected sums of products of the treatments and blocks with the values of the dependent factor, the values of the effects of the treatment and block levels, and the calculation of the sum of squares jointly explained by the treatments and blocks.

These calculations follow exactly the same general procedure used in the analysis of the variance of the complete block experiment used as first example. Again, the values of the multipliers are constant for this type and size of design, and need not be recalculated.

The total sum of squares calculated in the usual way, is 917.3333. Thus the ANOVA table is

Source of variation	*df*	*S S*
Treatments + blocks	*28*	*648.7444*
Error	*31*	*268.5889*
Total	*59*	*917.3333*

$$Ve = 268.5889/31 = 8.6641$$

In this case the value of Cij needed for the evaluation of the variance of the difference between 2 treatment effects depends on whether the treatments to be compared happen to occur together in some block, in which case $Cij = -16/900$; or whether they do not appear together in any block, in which case $Cij = -36/900$. Whether or not any 2 treatments appear in the same block may be seen in table 4, and the Cij values appear in table 5.

Again $T_{15} = -(T_1 + T_2 + \cdots + T_{14})$.

To date, the values of the Gauss multipliers of many of the more commonly used experimental designs have been evaluated and are available in the form of tables and/or formulas[3]. Among these are the following:

Unrestricted designs, complete blocks, complete blocks either with one or two missing values, Latin squares, Latin squares with one missing value, lattices, rectangular lattices, incomplete block designs of the $T = Z^2 - Z + 1$ type, and 2-plot block designs.

[2] Bose, R. C., Clatworthy, W. H., and Shrikhande, S. S., 1954. Tables of partially balanced designs with two associate classes, N. C. Agric. Exp. Stn. Tech. Bul. 107.

[3] Capó, B. G., 1979. Statistical analysis of research data (2nd ed), B. G. Capó, Ed, Rio Piedras, P. R.

Table 3.- Calculation of effects of treatments and blocks and of the sum of squares explained by them. (Common denominator of Gauss multipliers = 900)

	Numerators of Gauss multipliers																														
	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}	T_{13}	T_{14}	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}	B_{11}	B_{12}	B_{13}	B_{14}	B_{xy}	Effects X 900	Contribution to $Sy'y'$ X 900
T_1	264	-16	-16	-16	-16	-36	-16	-16	-16	-16	-36	-16	-16	-16	-54	-54	-54	21	21	-54	16	21	16	21	16	21	21	21	-4	T_1 = 1120	-4,480
T_2	-16	264	-16	-16	-16	-16	-36	-16	-16	-16	-16	-36	-16	-16	16	21	-54	-54	21	21	21	16	-54	21	21	-54	21	21	-21	T_2 = -3335	70,035
T_3	-16	-16	264	-16	-16	-16	-16	-36	-16	-16	-16	-16	-36	-16	21	21	16	-54	16	-54	21	-54	21	21	-54	21	16	21	-17	T_3 = -2565	43,605
T_4	-36	-16	-16	264	-16	-16	-16	-16	-36	-16	-16	-16	-16	-36	21	16	21	16	-54	-54	21	21	-54	16	21	21	21	21	3	T_4 = - 490	- 1,470
T_5	-16	-16	-16	-16	264	-16	-16	-16	-16	-36	-16	-16	-16	-16	21	-54	21	21	21	16	21	-54	-54	21	21	16	-54	16	- 3	T_5 = 610	- 1,830
T_6	-36	-16	-16	-16	-16	264	-16	-16	-16	-16	-36	-16	-16	-16	21	21	21	21	-54	21	16	-54	16	-54	16	-54	21	21	2	T_6 = 1495	2,990
T_7	-16	-36	-16	-16	-16	-16	264	-16	-16	-16	-16	-36	-16	-16	16	-54	21	21	-54	21	21	16	21	21	-54	21	21	-54	- 2	T_7 = 340	- 680
T_8	-16	-16	-36	-16	-16	-16	-16	264	-16	-16	-16	-16	-36	-16	21	-54	16	21	16	21	-54	21	21	21	21	-54	16	21	- 6	T_8 = 660	- 3,960
T_9	-16	-16	-16	-36	-16	-16	-16	-16	264	-16	-16	-16	-16	-36	-54	16	21	16	21	21	21	21	21	16	-54	-54	-54	21	- 2	T_9 = 1760	- 3,520
T_{10}	-16	-16	-16	-16	-36	-16	-16	-16	-16	264	-16	-16	-36	-16	21	21	-54	21	21	16	21	21	21	-54	-54	16	21	16	-17	T_{10} = - 1835	31,195
T_{11}	-36	-16	-16	-16	-16	-36	-16	-16	-16	-16	264	-16	-16	-16	21	21	21	-54	21	21	16	21	16	21	16	21	-54	-54	4	T_{11} = 1045	4,180
T_{12}	-16	-36	-16	-16	-16	-16	-36	-16	-16	-16	-16	264	-16	-16	16	21	21	21	21	-54	-54	16	21	-54	21	21	-54	21	10	T_{12} = 2515	25,150
T_{13}	-16	-16	-36	-16	-16	-16	-16	-36	-16	-16	-16	-16	264	-16	-54	21	16	21	16	21	21	21	-54	-54	21	21	16	-54	-14	T_{13} = - 615	8,610
T_{14}	-16	-16	-16	-36	-16	-16	-16	-16	-36	-16	-16	-16	-16	264	21	16	-54	16	21	21	-54	-54	21	16	21	21	21	-54	-18	T_{14} = -1690	30,420
B_1	-54	16	21	21	21	21	16	21	-54	21	21	16	-54	21	264	-16	-16	-16	-16	-16	-16	-36	-16	-16	-16	-16	-36	-16	-17	B_1 = -4040	68,680
B_2	-54	21	21	16	-54	21	-54	-54	16	21	21	21	21	16	-16	264	-16	-36	-16	-16	-16	-16	-16	-36	-16	-16	-16	-16	- 6	B_2 = -1435	8,610
B_3	-54	-54	16	21	21	21	21	16	21	-54	21	21	16	-54	-16	-16	264	-16	-36	-16	-16	-16	-16	-16	-16	-16	-36	-16	-12	B_3 = - 660	7,920
B_4	21	-54	-54	16	21	21	21	21	16	21	-54	21	21	16	-16	-36	-16	264	-16	-16	-16	-16	-16	-36	-16	-16	-16	-16	- 3	B_4 = 890	- 2,670
B_5	21	21	16	-54	21	-54	-54	16	21	21	21	21	16	21	-16	-16	-36	-16	264	-16	-16	-16	-16	-16	-16	-16	-36	-16	24	B_5 = 5415	129,960
B_6	-54	21	-54	-54	16	21	21	21	21	16	21	-54	21	21	-16	-16	-16	-16	-16	264	-16	-16	-16	-16	-16	-36	-16	-36	12	B_6 = 3155	37,860
B_7	16	21	21	21	21	16	21	-54	21	21	16	-54	21	-54	-16	-16	-16	-16	-16	-16	264	-16	-36	-16	-36	-16	-16	-16	- 5	B_7 = -1145	5,725
B_8	21	16	-54	21	-54	-54	16	21	21	21	21	16	21	-54	-36	-16	-16	-16	-16	-16	-16	264	-16	-16	-16	-16	-16	-16	-15	B_8 = 2240	33,600
B_9	16	-54	21	-54	-54	16	21	21	21	21	16	21	-54	21	-16	-16	-16	-16	-16	-16	-36	-16	264	-16	-36	-16	-16	-16	- 5	B_9 = 430	- 2,150
B_{10}	21	21	21	16	21	-54	21	21	16	-54	21	-54	-54	16	-16	-36	-16	-36	-16	-16	-16	-16	-16	264	-16	-16	-16	-16	- 6	B_{10} = -1135	6,810
B_{11}	16	21	-54	21	21	16	-54	21	-54	-54	16	21	21	21	-16	-16	-16	-16	-16	-16	-36	-16	-36	-16	264	-16	-36	-16	-13	B_{11} = -1745	22,685
B_{12}	21	-54	21	21	16	-54	21	-54	-54	16	21	21	21	21	-16	-16	-16	-16	-16	-36	-16	-16	-16	-16	-16	264	-16	-36	- 7	B_{12} = -1120	7,840
B_{13}	21	21	16	21	-54	21	21	16	-54	21	-54	-54	16	21	-36	-16	-36	-16	-36	-16	-16	-16	-16	-16	-16	-16	264	-16	18	B_{13} = 3165	56,970
B_{14}	21	21	21	21	16	21	-54	21	21	16	-54	21	-54	-54	-16	-16	-16	-16	-16	-36	-16	-16	-16	-16	-16	-36	-16	264	- 5	B_{14} = - 295	1,475

$Sy'y'$ = 583,020/900
= 648.3444.

Considerable progress has also been attained in the evaluation of the multipliers of other partially-balanced designs such as the R-27 design used as the second example in this paper.

Now, for those who have excess computer availability in comparison with their punching facilities, thus able to lose the time needed for the computer to invert the corresponding matrix, there is a computer program available for the analysis of data with this method[4]. This program has a capacity of up to 100 levels of the qualitative independent factors and up to 8 quantitative (co-variance) independent factors.

Bernardo G. Capó
Statistics Section

[4] Antoni, M. and Capó, B. G., 1978. A computer program for the analysis of experimental data, J. Agric. Univ. P. R. 62: 431–47.