Exercises practising SAS/STAT^{R} GLM procedure - Analysis of response in MS Excel - #
SUPPLEMENT: Revision and Resume Added After (and In) Presentation (With the Software 'anore.xls').

Tadao SHIBAYAMA* (Retired: Nagoya Municipal Industrial Research Institute)
#Original Paper: (2006) SAS Forum User's Group (Japan) Academic Session 2006 Papers (Poster Session), p.321-328.
*Present address (home) : Sunshine Kanayama 502, 14-13 Furuwataricho, Naka-ku, Nagoya 460-0025 JAPAN
*Ph.-Fx.: +80-(0)52-331-8295 *E-mail: 64111256@people.or.jp

References: In this SUPPLEMENT the **References** [1]-[9] are that of the Original Paper # (above), p.328.
[0] Besides that, the Original Paper# is to be cited as the Ref. [0] whereas Ref. [10] below is added newly.
[10] T. Shibayama (2006): Analysis of response by software in Excel as proposed. (Japanese). Behaviormetrics Society of Japan 34th General Meeting (General Session: Mining Software 2006.09.13 0945-1130 Rm C), Papers, p.182-185.

A. Resume of presentation/revision (with aim of the study explicated).

The excellent documentations of SAS^{R} System help users ever [0] that written readably and well-organized with resources located precisely and almost thoroughly, still newcomers may often feel difficulties to be familiar with use of the important SAS/STAT^{R} GLM procedure that working on the estimable functions classified almost prototypically.

A cause of the situation may be that the procedure must cover very many cases of actual problems so extensively, details of identification of estimable functions, for example, cannot be exhausted so only summarized as to be seen in the typical cases, User's Guide of SAS/STAT^{R} 9.1 (2004), Ch.11, p.186 (**Table 11.10**) and 188 (**Table 11.12**), or almost equivalently, SAS/STAT^{R} Version 6, 4th edn, Vol. 1 (1990), Ch.9, p.122 (**Table 9.10**) and 123 (**Table 9.12**), or others.

Perhaps, the situation can be saved by practising the procedure in many cases utilizing the MODEL statement modified by the optional modifiers /E, /E1, /E2, /E3 or /E4 and, furthermore, the LSMEANS statement or so.

Anyway, the estimable functions or the estimable effect elements should be identified clearly. A newcomer may be helped by an analysis of response (ANORE) software for personal use coded in programs in FORTRAN77 [5] or Fortran90 etc., or in MS Excel 2003 VBA (Visual Basic Applications) [0] as demonstrated in this presentation.

In Excel system a matrix of 200 columns or so is displayed in its real image on an Excel worksheet (65536 Rows × 256 Columns, typical). An Excel book (workbook) usually comprises 3 such sheet. If a sheet is divided into 218 zones each comprising 300 adjacent rows × 256 columns, each zone may be used by a square or transposed matrix.

Each zone (of Sheet1) is identified arbitrarily by a mnemonic such as s1z0 (Row1-Row299), s1z1 (R300-R599), ⋯ , s1z9 (R2700-R2999), s1zA0 (R3000-R3299), s1zA1 (R3300-R3599), ⋯ , s1zA9 (R5700-R5999), s1zB0 (R6000-R6299), ⋯ , or on, up to s1zU8 (R65400-R65536, this zone is marginal so contains only 137 rows). Each mnemonic, furthermore, plays a role of the row number of the top row of a zone (in Sheet1) so that s1zA1 = 3300, for example.

A matrix of jh rows × ih columns may be placed in the zone s1zA1 (for example) from Cells(3311, 6), i.e., Cells(s1zA1 + 10 + 1, 5 + 1), of Sheet1, down and right, then, an element (jr, it) (jr = 1, ⋯ , jh; it = 1, ⋯ , ih) of the matrix is placed in or equal to Cells(s1zA1 + 10 + jr, 5 + it) (provided that s1zA1 = 3300) in MS Excel VBA notation.

With such the convention, the ANORE software coded in FORTRAN77 main programs [5] is transformed into an equivalent coded in MS Excel VBA procedures [0] much more powerful in its output, input or storage serviceabilities, with any matrix as the output, input or any intermediate displayed, written or stored in a zone in the real form.

Applicable to various cases, a few streams ([0] Tables p.325-326, but much revised now) are processed in parallel but independently to meet with various sets of TreatmentsList and EffectComponentsList or EffectElememtsList.

The lists may be modified in process with a run restrained (r) or missing (m), a factor omitted (f) or neglected (n), an effect component restrained (c) or excluded (e), an effect element ineffectuated (u) or annihilated (a), or an item (a run, a factor, an effect component, or an effect element) arbitrarily repressed (p) or discarded (q), etc. as processed so that each result of analysis in each modified stream may be compared to each other easily ([0], but much revised now).

By the proposal of analysis of response (ANORE) software in MS Excel VBA [0] or in FORTRAN77 [5], it is aimed that algorithms of analysis of response in various irregular cases clarified step by step precisely, accurately and exactly, and definitions of any effect components or elements, too, that are often mentioned only briefly or implicitly.

Particularly the software in MS Excel VBA [0] is useful because it can show changes of matrices successively and explicitly as these are processed. It may provide step-by-step explanation for newcomers to be familiar even with such the examples of estimable functions shown typically in SAS/STAT^{R} User's Guide that mentioned above.

B. Diagonal-elements solution of the canonically constrained normal equation: The convergence.

The method of solution was successful in ANORE (analysis of response) software in FORTRAN77 [5] to solve canonically constrained normal equation even of an irregularly abridged arrangement in a common way.

The principle of the method is briefed in AUXILIARY ILLUSTRATIONS (a handout of [5]: posted as a MS Word 2000 document file 'pa.doc' in the file 'anore' in ftp://ftp.sas.com/pub/webfiles/Japan/contrib/sugij01_anore), Figures 5 and 6, particularly, the process of recurrent iteration has experienced no problem of convergence so far, but it was successful only fortunately. Starting values and intermediate corrections should have been important.

Transcripted in MS Excel VBA and applied to various cases, the solution diverged sometimes, or more often, just immediately before the 'presentation' and after. The cause and any measure for convergence should be sought.

An important mistake in reasoning in Original Paper [0] has been detected and cured months after. Selection of starting values and intermediate corrections formulated. The convergence is achieved by procedures revised.

The mistake was as follows: A diagonally unitarized canonically constrained normal equation coefficients matrix as displayed on LCD, the head element of any row (i.e., any element of the 1st column) is equal to unity and the diagonal element of the row, too, so strikingly. This is effected by the situation such as follows.

Any row of the coefficients matrix of a canonically constrained or the original normal equation is built of rows of the design matrix such that pertaining to a particular jr-th (say) effect element. The jr-th element of such the row of the design matrix is unity (1) by the definition of the design matrix. Such the rows, n (say) in all, are summed to build a row (the jr-th) of the normal equation coefficients matrix. Consequently, its jr-th element is equal to n.

Whereas the head element of the row, too, because the head element of each of the original n rows is equal to unity (1) as any the head element pertains to the 1st effect element, i.e., the general mean.

Therefore, if elements of the jr-th row of the normal equation coefficients matrix built are normalized by the diagonal element (=n), the head element is to be equal to unity (1) as well as the diagonal element.

Forward elimination by the 1st row eliminates head elements of all the rows below [0]. That 'might' be equivalent to and 'could' dispense with canonical constrainment of each columns: A 'MISCONCEPTION' !

Checks of the normal equation and diagonal elements method of solution show that the columnwise canonical constrainment incorporates canonical constraints into the normal equation coefficients matrix really. It can be never dispensed with. Or it can be never replaced by the forward elimination that is of a much more trivial meaning.

The original software coded in FORTRAN77 [5] is completely correct and has no problem. Only the software translated in MS Excel Basic [0] was wrong as (mis-)'amended' by the misconception. It is cured now.

Starting values and intermediate corrections: Apart from the confusion and from the remedy, the recurrent iteration should have been affected by the starting values and intermediate corrections. Even if no problems been experienced so far, it owes perhaps so fortunate selection of the starting values and of the corrections.

A simple method of selection and of correction is formulated now and working successfully so far.

The diagonal elements solution of the normal equation gives the solution as the response effect-elements conversion matrix 'onrt'. The nt (:= nn-1)_th iteration gives an approximate 'onrt_nt' that is modified and employed as the starting value 'onrt_na' of the (next) nn_th iteration giving an approximate 'onrt_nn' recurrently.

The initial starting value 'onrt_0' is usually taken as equal to the refined response partial sums matrix 'onrs' corresponding to the refined diagonally unitarized canonically constrained coefficients matrix 'onrh'.

Any solution should be substituted into original normal equation, and into the constraints, to be checked ever.

The approximate 'onrt_nt' that is got of the nt_th iteration is modified by an equation (*) as follows

(*) onrt_na = onrt_nt . (1 - lambda) + (onrs - onrc . onrt_nt) . lambda

that gives the value 'onrt_na' to be employed as the starting value of the (next) nn_th iteration. The matrix 'onrc' is the off-diagonal matrix of the matrix 'onrh'. Degree of the modification 'lambda' is taken around 0.3.

The analysis of response (ANORE) software in MS Excel VBA applied easily to various problems, the solution diverged in some cases even after the 'wrong' algorithm by the 'misconception' revised, but it was cured by the starting values in the intermediate cycles of iteration modified by equation (*). The initial starting value 'onrt_0' affects the convergence much whereas the correction by equation (*) in the intermediate cycles proves more important.

C. Revisions of manuscripts of references.

1. In Original Paper [0] p.328, the product $L(X'.X)^{-1}L'$ just below eqn(3) should be read $[L(X'.X)^{-}L']^{-1}$.

2. In Ref. [4] */academic04_shibayama PRESENTATION AID: User's Guide 'Chapter 4' should be read 'Chapter 9'.

3. In Ref. [5] */sugij01_anore, Presn Handout (pa.doc) 'AUXHI…' / Readme2.txt 'Auxhi…' : 'H' / 'h' be deleted.

* | ftp://ftp.sas.com/pub/webfiles/Japan/contrib

D. Remarks: The software may be opened by any possible means on request (to the author) or otherwise.

POSTERS and HANDOUTS: Titles, Précis or Revision added after the presentation.

Posters (printed, in small size, Handout sheet 1-3). --

0-1 Typical Exercise 1 SAS/STAT[R] User's Guide* A 3x3 Factorial Design with Missing Diagonals.
*** 9.1(2004) Ch.11, p.186 Table 11.10.** Vn6, 4th edn Vol.1(1990) Ch.9, p.122 Table 9.10.

0-2 Typical Exercise 2 SAS/STAT[R] User's Guide* 3x3 Factorial Design with Four Missing Cells.
*** 9.1(2004) Ch.11, p.188 Table 11.12.** Vn6, 4th edn Vol.1(1990) Ch.9, p.123 Table 9.12.

1. Factorization of a design matrix X (Design Matrix) = J (Restoration Matrix) . Lu (Full Estimable Matrix) on the basis of that Lu (Full Estimable Matrix) = K (Contraction Matrix) . X (Design Matrix): An identity matrix I such that I (An Identity Matrix) = G0* (Inverse of a regular matrix) . G0 (A Regular Matrix) may be inserted in between J (Restoration Matrix) and Lu (Full Estimable Matrix) of the factorized X (Design Matrix), giving another set of J0* (Restoration Matrix) and Lu0 (Full Estimable Matrix) complicating the situation. [4]

2. The General Form & Linearly Independent Rows of the estimable or design matrix.are to be selected not uniquely. The details traced in MS Excel VBA etc. may help well to see the ground of the SAS/STAT[R] GLM procedure.

3. Treatment Matrix (Template 1 rewritten and pre-processed) - shown in poser partly (R288-315 x C1-13).
4. Effect Components List (Template 2 referred to: on the presn site) - shown in poster partly (R588-615 x C1-13).
5. Convergence of the Diagonal Elements Solution Should Be Checked - See partly (R65088-65115 x C1-13). If it diverges, should be called for 'Forward Elimination & Backward Substitution'. **Note:** At the time of the presentation, it diverged often, caused of a wrong algorithm. The necessary canonical constrainment operation columnwise was omitted wrongly by misconception (cf. **B.** of this SUPPLEMENT). Now revised completely, furthermore, the starting value and the intermediate correction to be selected appropriately, no problem having been experienced so far.
6. EffectElements vector to be got from the response vector by the conversion matrix - partly (R10488-10515 x C1-13) shown in poster. **Note:** Successfully even with the wrong program, owing the fortunate starting value perhaps.

Supplementary Handouts ---
(Handout sheet 4-6) **SUGI-J2003 PRESENTATION AID**, Shibayama, T. (2003):
ftp://ftp.sas.com/pub/webfiles/Japan/contrib/sugij03_shibayama
Indeterminacy of effect components of factorial experiments arising from the ways of expression.
(sheet 4) FORMULAS: PART 0, PART 1, PART 2 and PART 3.
(sheet 5) An example of hypothesis testing by SAS/STAT[R] GLM procedure.
A full estimable matrix, LuI, LuII, LuIII or Lu.
(sheet 6) Restoration matrix, JI, JII or JIII, built of the restoration matrix J of the general type
by the inverse sweep operator KKI, KKII or KKIII such that JI = J.KKI, JII = J.KKII or JIII = J.KKIII.
Furthermore, FORMULAS PART4 as continued from PART 3 (sheet 4).

(Handout sheet 7-11) **SUGI-J2001 PRESENTATION AID**, Shibayama, T. (2001):
ftp://ftp.sas.com/pub/webfiles/Japan/contrib/sugij01_anore (p.doc, pa.doc, pb.doc and pc.doc)
Simple input format to be employed in analysis of results of factorial experimentation.
(sheet 7) POSTER (p.doc)
AUXILIARY ILLUSTRATIONS (pa.doc)
Diagonal elements method solution of normal equation is explained briefly in Fig. 6.
(sheet 8) SUPPLEMENTARY NOTE (pb.doc)
Modularly classified contrasts interpreted in terms of functionally developed contrasts.
(sheet 9-11) Presentation Aid for SUGI-J '89, Shibayama, T. (1989) (pc.doc)
Introduction to the use of the FACTEX procedure of the SAS/QC[R] software.

Some barriers against newcomers to be familiar with SAS/STAT[R] GLM procedure may be removed by these materials.