ENCYCLOPEDIA OF STATISTICAL SCIENCES

VOLUME 3

Faà di Bruno's Formula — Hypothesis Testing

SAMUEL KOTZ & NORMAN L. JOHNSON
EDITORS-IN-CHIEF

CAMPBELL B. READ
ASSOCIATE EDITOR

where $r(\cdot)$ denotes the rank of a matrix. We note that the weights (w_{ij}) involve the $\{\theta_i\}$ and so an iterative method of solution is required (see Nelder [6] for one approach). However, (5) is in fact the set of likelihood equations for the $\{\theta_i\}$ based on error contrasts and an assumption of normality for $\mathbf{y}$, and their solution leads to *restricted maximum likelihood estimates*. In this case we may calculate the *information matrix* $I(\boldsymbol{\theta})$ of the $\{\hat{\theta}_i\}$, and $-2I(\boldsymbol{\theta})$ turns out to have diagonal elements

$$\theta_i^{-2}\left[r(\mathbf{C}_i) - \sum_{j\,:\,\lambda_{ij}>0} r(\mathbf{T}_j) + \sum_j (1-w_{ij})^2 r(\mathbf{T}_j)\right] \tag{6a}$$

and off-diagonal elements

$$\theta_i^{-1}\theta_{i'}^{-1}\sum_j w_{ij}w_{i'j}r(\mathbf{T}_j). \tag{6b}$$

A particular advantage of using this procedure with (GB) designs is the fact that a simple expression can be found for the right-hand side of (5) and Fisher's scoring method can be adopted using (6a) and (6b) without recalculating $\hat{\boldsymbol{\alpha}}$ explicitly (see Nelder [6]). In general this cannot be avoided.

We close with a remark about the calculation of BLUPs of random effects under models such as (2). This is really an aspect of the block structure defined by $\mathbf{C}_1, \ldots, \mathbf{C}_s$, so we do not give any details. It will suffice to observe that if $\mathbf{b}'\mathbf{U}\mathbf{y}$ is the best linear predictor of a contrast of random effects when $\boldsymbol{\alpha} = \mathbf{0}$, then $\mathbf{b}'\mathbf{U}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})$ is the BLUP of this same contrast for arbitrary $\boldsymbol{\alpha}$. The great advantage of having (GB) hold is the simplicity of the expression for $\hat{\boldsymbol{\alpha}}$ and the ease of estimation of the necessary dispersion parameters.

REMARKS

General balance was introduced by Nelder [5] with the intention of using the notions to permit the writing of very general computer programs for analyzing designed experiments. Many of the ideas described above have been implemented in a modified form in GENSTAT (see Alvey et al. [1, Chap. 6]). The example of a row-and-column design not satisfying (GB) comes from the thesis of Houtman [3], as does the combinability result given in the preceding section. For a recent review of the literature on variance components* and BLUPs which includes a discussion of (GB), see Thompson [9].

References

[1] Alvey, N. G. et al. (1977). *GENSTAT. A General Statistical Program*. Rothamsted Experimental Station, Harpenden, Herts., England.

[2] Brown, L. D. and Cohen, A. (1974). *Ann. Statist.*, **2**, 963–976.

[3] Houtman, A. (1980). *The Analysis of Designed Experiments*. Ph.D. dissertation, Princeton University Press, Princeton, NJ.

[4] Kempthorne, O. (1952). *The Design and Analysis of Experiments*. Wiley, New York.

[5] Nelder, J. A. (1965). *Proc. R. Soc. A*, **283**, 147–178.

[6] Nelder, J. A. (1968). *J. R. Statist. Soc. B*, **30**, 303–311.

[7] Pearce, S. C. (1963). *J. R. Statist. Soc. A*, **126**, 353–377.

[8] Preece, D. A. (1966). *Biometrics*, **22**, 1–25.

[9] Thompson, R. (1980). *Math. Operationsforsch. Statist. Ser. Statist.*, **11**.

[10] Yates, F. (1936). *Ann. Eugen.*, **7**, 121–140.

(BALANCE IN EXPERIMENTAL DESIGN
BLOCKS, BALANCED INCOMPLETE
DESIGN AND ANALYSIS OF EXPERIMENTS
ESTIMABILITY
GENERAL LINEAR MODEL
PARTIALLY BALANCED DESIGNS
SPLIT-PLOT DESIGNS
WEIGHTED LEAST SQUARES)

T. P. SPEED

GENERALIZED CANONICAL VARIABLES

A random vector $\mathbf{X}$ of interest poses both analytical and economical problems in many situations if it has too many compo-

nents. The large number of correlations associated with the components of $\mathbf{X}$ makes it difficult to comprehend overall or general relationships. A reduction in dimensionality together with some representative measure of relationships among the variables can thus be of practical importance. Generalized canonical variables (GCVs), with their associated correlations, termed the generalized canonical correlations (GCCs), attempt to serve this purpose.

It is assumed that a meaningful subgrouping of $\mathbf{X}: p \times 1$ into several disjoint subvectors $\mathbf{X}_1, \ldots, \mathbf{X}_k$, $\mathbf{X}_i: p_i \times 1$, $\sum_1^k p_i = p$ is already given. For $k = 2$, Hotelling [6] introduced the concept of canonical variables (CVs) and associated canonical correlations (CCs) (*see* CANONICAL ANALYSIS).

DEFINITION

The first GCV, $\mathbf{Y}^{(1)}$ with $\mathbf{Y}^{(1)\prime} = [Y_1^{(1)}, \ldots, Y_k^{(1)}] = \mathbf{f}^{(1)}(\mathbf{X})' = [f_1^{(1)}(\mathbf{X}_1), \ldots, f_k^{(1)}(\mathbf{X}_k)]$, where the $f_j^{(1)}$s are real-valued functions, is a k-dimensional random variable, the components of which are chosen so as to optimize a criterion based on some function of their correlation matrix. For each such function there will be a corresponding GCV as a generalization of CV. The higher-stage GCVs, $\mathbf{Y}^{(i)} = \mathbf{f}^{(i)}(\mathbf{X})$, $i = 1, 2, \ldots$, are also k-dimensional random variables, the components of which are chosen so as to optimize the same criterion with some additional constraints imposed at each stage regarding the relationships among the variables at a given stage with those in the preceding stages. [These constraints may be different for different methods. Also, the $f_j^{(i)}(\mathbf{X}_j)$s are usually linear functions of the $\mathbf{X}_j$s, $j = 1, 2, \ldots, k$, and have unit variances.] For a given GCV, the associated correlation can be loosely termed as GCC, although depending on the method of construction of the GCVs such GCCs may in fact be correlation matrices instead of a one-dimensional summary statistic. It may be reasonable to terminate the procedure at the sth stage if, depending on the method, at the $(s + 1)$th stage the scalar GCC is near zero or the off-diagonal elements of the GCC matrix are close to zeros.

There are two basic problems related to GCV analysis. One is concerned with the construction of GCVs when the population dispersion matrix is known. The other deals with statistical inference when only a sample from the population is available. Although the first problem had been attacked since 1951, the research on the second started only recently with Sen Gupta [12]. Also, a realistic interpretation of GCVs is of great practical importance.

CONSTRUCTION

For construction of GCVs, one faces at least three problems of optimization: (a) selection of the number k of subgroups and their corresponding elements, (b) determination of the compounding functions $f_j^{(i)}$s, and (c) deciding on the optimal stage of stopping for higher-order GCVs. As stated above, it is usually assumed for (a) that the number and the elements of the subgroups are completely specified. Otherwise, one may attempt to use cluster analysis techniques. For (b) several available methods will be discussed. With $k > 2$, for (c), the situation is "somewhat arbitrary" for some of these methods.

The algebraic derivations become simpler and essentially the same as in CC analysis if we can reduce the several groups effectively to two groups. Further, for multivariate normal populations, testing that the CCs are all zeros is equivalent to testing that the two sets are independent. These considerations motivated the construction of various conditional GCCs. A set or sets of variables are held fixed and effectively the two sets of residuals are analyzed. The GCCs can then be obtained through parallel test criteria of (conditional) independence as in the case of two-group CCs. Most recently this concept has led to g_1 and g_2 bipartial GCC analysis

(see Lee [8]). For each of the cases above, the GCVs are the normalized eigenvectors associated with the eigenvalues and GCCs corresponding to the determinantal equation

$$|{}_k\Sigma^*_{od} - (k-1)\rho_k {}_k\Sigma^*_d| = 0, \qquad {}_k\Sigma^* = {}_k\Sigma^*_{od} + {}_k\Sigma^*_d. \tag{1}$$

${}_k\Sigma^*_d$ is the diagonal supermatrix with elements Σ^*_{ii}, $i = 1, \ldots, k$, and ${}_k\Sigma^*$ is the modified covariance matrix, modified by the particular generalization under consideration. Although computationally quite convenient, these GCVs seem to be of limited practical utility because of their conditional nature.

McKeon [10] has suggested a (first) GCC obtained as a generalization of a modified intraclass correlation coefficient*. For the CVs ($k = 2$) (see, e.g., Anderson [1, Chap. 12, eq. (14)]) and the first GCV by the method of McKeon ($k = k$), ${}_k\Sigma^*$ of (1) is simply Σ, the dispersion matrix of $\mathbf{X}$. Carroll [2], Horst [5], and several other authors have also arrived at a similar solution, although from different viewpoints. Sen Gupta [12] derived new GCVs obtained by modifying with the equicorrelation constraint the criterion of minimizing the generalized variance* of $\mathbf{Y}$ (i.e., $|\Sigma_{\mathbf{Y}}|$; see Anderson [1, p. 305, Prob. 5]). These new GCVs are quite convenient for purposes of statistical inference and will be referred to later in that context.

Let the correlation (also covariance) matrix of $\mathbf{Y}^{(1)}$ be $\Sigma^{(1)} = (\rho^{(1)}_{ij})$. An important property of CC analysis is that the CVs are invariant under nonsingular transformations of either set. Exploiting this property, Horst [5] proposed maximizing $\sum\sum_{i<j=1}^{k} \rho^{(1)}_{ij}$ and also maximizing the largest eigenvalue of $\Sigma^{(1)}$. These are termed the SUMCOR and MAXVAR methods, respectively. Exploiting the same property, Steel [14] suggested the GENVAR method, where GCVs are obtained by minimizing the determinant $|\Sigma^{(1)}|$, the generalized variance of $\mathbf{Y}^{(1)}$. The methods of SSQCOR and MINVAR were proposed by Kettenring [7]. The former attempts to maximize $\sum\sum_{i<j=1}^{k} \rho^{(1)2}_{ij}$ or equivalently trace of $\Sigma^{(1)2}$, while the latter minimizes the smallest eigenvalue of $\Sigma^{(1)}$. Kettenring has given some interesting factor-analytic interpretations for the foregoing five procedures. However, for GENVAR, SSQCOR, and SUMCOR, the corresponding defining equations cannot be presented in simple form as in (1) and iterative procedures are required even for first GCCs and GCVs.

All of the methods above reduce to CC analysis for $k = 2$. Each method emphasizes some aspect of the correlation matrices of the GCVs and hence attempts to detect certain forms of linear relationships among the sets of variables. Thus, depending on the problem, it may be necessary to use several of these methods to better understand the underlying relationships among the sets of variables in $\mathbf{X}$.

For singular ${}_k\Sigma^*$, see Sen Gupta [11].

STATISTICAL INFERENCE

When only a sample is available from the population, the GCVs are usually estimated by the maximum likelihood* method, which effectively replaces the population parameters in (1) by their sample counterparts.

Exact distributions for GCVs are quite complicated and seem to be nearly intractable for those obtainable only by iterative procedures. A large-sample approximation by multivariate normality may be suggested.

Several problems of tests of hypotheses* were formulated in Sen Gupta [12]. Since GCVs obtained by different methods optimize different criteria, the test statistics and associated distributions need to be worked out separately for each method. We illustrate this below using the new GCVs obtained in Sen Gupta [12]. For simplicity, only tests based on the first new GCV, $\mathbf{Y}^{(1)} \equiv \mathbf{Y}$, will be considered. Let the corresponding correlation matrix of $\mathbf{Y}$ be $\Sigma_{\mathbf{Y}}$. Large samples are assumed.

1. It would be natural to explore to what extent the GCVs optimize the criteria, i.e., what value of the criterion is achieved by a particular GCV. For new GCVs, this leads to a test of $H_0: |\Sigma_{\mathbf{Y}}| = \sigma_0^{2k}$ (specified) against $H_1: |\Sigma_{\mathbf{Y}}| < \sigma_0^{2k}$. A test for $H_0: \rho^{(1)} = \rho_0^{(1)}$

(given) against $H_1: \rho^{(1)} \neq \rho_0^{(1)}$ can also be proposed, where $\rho^{(1)}$ is the new (first) GCC. These involve curved exponential families*. Likelihood ratio tests* and a locally most powerful test for $\rho^{(1)}$ are available. The latter test is globally one-sided unbiased and the exact distribution of the test statistic is available in terms of the (tabulated) Kummer's function (see Sen Gupta [13]).

2. Next, it would be of interest to determine whether with the same dimension k, reasonable alternative regroupings of $\mathbf{X}$ produce better results (see, e.g., Gnanadesikan [3, p. 77, para. 4]). If there are m reasonable regroupings, each of dimension k, then for the new GCVs, one would test $H_0: |\mathbf{\Sigma}_{jY}|$s all equal against $H_1: |\mathbf{\Sigma}_{jY}|$s are in a given order, where ${}_j\mathbf{Y}$ is the (first) GCV for the jth regrouping, $j = 1, \ldots, m$. If the same set of data is used to estimate the m GCVs, the tests will be quite involved, owing to the mutual dependence of the GCVs. Some judicious transformations are helpful here to yield simpler tests.

3. It may be possible to reasonably regroup $\mathbf{X}$ into various numbers k_i, $i = 1, \ldots, l$ of subsets, yielding GCVs of different dimensions. It is then worthwhile to explore whether a GCV of smaller dimension performs as good as, if not better than, one of a higher dimension. For such a case, the former GCV will naturally be preferred. For l subsets, with new GCVs, one would test $H_0: |\mathbf{\Sigma}_{jY}|^{1/k_j}$ all equal against $H_1: |\mathbf{\Sigma}_{iY}|^{1/k_i} < |\mathbf{\Sigma}_{jY}|^{1/k_j}$, $k_i < k_j$, $i \neq j$, $i, j = 1, \ldots, l$, where ${}_i\mathbf{Y}$ is the (first) new GCV of dimension k_i obtained by the ith mode of regrouping, $i = 1, \ldots, l$. Some solutions can be presented here through isotonic regression*. For more details, see Sen Gupta [12].

Example. From Thurstone and Thurstone [15] $k = 3$ sets of scores on three batteries of three tests each (i.e., $p_i = 3$, $i = 1, 2, 3$, $p = 9$) for several individuals are available. The vector variables $\mathbf{Z}_i: 3 \times 1$, $i = 1, 2, 3$, represent different measures of verbal, numerical, and spatial abilities of the subjects tested. A transformation was employed on $\mathbf{Z}' = (\mathbf{Z}_1', \mathbf{Z}_2', \mathbf{Z}_3')$ to give internally "sphericized" standardized variables $\mathbf{X}$. Horst [4], Kettenring, Gnanadesikan, and Sen Gupta have derived GCVs for $\mathbf{X}$ from its covariance (same as correlation) matrix $\mathbf{R}$ (see also McDonald [9]). Let $\mathbf{R} = (\mathbf{R}_{ij})$, the $\mathbf{R}_{ij}$ being 3×3 matrices. Then $\mathbf{R}_{ii} = \mathbf{I}$, $i = 1, 2, 3$, and

$$\mathbf{R}_{12} = \begin{bmatrix} 0.636 & 0.126 & 0.059 \\ -0.021 & 0.633 & 0.049 \\ 0.016 & 0.157 & 0.521 \end{bmatrix};$$

$$\mathbf{R}_{13} = \begin{bmatrix} 0.626 & 0.195 & 0.059 \\ 0.035 & 0.459 & 0.129 \\ 0.048 & 0.238 & 0.426 \end{bmatrix};$$

$$\mathbf{R}_{23} = \begin{bmatrix} 0.709 & 0.050 & -0.002 \\ 0.039 & 0.532 & 0.190 \\ 0.067 & 0.258 & 0.299 \end{bmatrix}.$$

Consider, including MAXVAR, the different approaches discussed above which yield the same result in the first stage as that of McKeon's method. For McKeon's method [e.g., in (1)] we have $\mathbf{\Sigma}^* = \mathbf{R}$, $\mathbf{\Sigma}_d^* = \mathbf{R}_d = \mathbf{I}$, and the first GCV is obtained from the eigenvector $\mathbf{v}^{(1)\prime} = (\mathbf{v}_1^{(1)\prime}, \ldots, \mathbf{v}_k^{(1)\prime})$ corresponding to the largest eigenvalue of the equation $|\mathbf{R} - \lambda\mathbf{I}| = 0$. The first GCC for McKeon's method is $\rho^{(1)} = 0.745$, which is larger than any pairwise correlation in $\mathbf{R}$. In the MAXVAR method, the usual requirement that the components $Y_i^{(1)} = \boldsymbol{\alpha}_i^{(1)\prime}\mathbf{X}_i$ of the first GCV $\mathbf{Y}^{(1)}$ have unit variance is achieved by letting $\boldsymbol{\alpha}_i^{(1)} = \mathbf{v}_i^{(1)}/\mathbf{v}_i^{(1)\prime}\mathbf{v}_i^{(1)}$, $i = 1, \ldots, k$. This yields

$$\begin{aligned}\mathbf{Y}^{(1)} = (&0.732X_{11} + 0.514X_{12} + 0.447X_{13}, \\ &0.659X_{21} + 0.625X_{22} + 0.420X_{23}, \\ &0.678X_{31} + 0.640X_{32} + 0.362X_{33}),\end{aligned}$$

where X_{ij} is the jth element of $\mathbf{X}_i$, $i, j = 1, 2, 3$.

It turns out that $\mathbf{Y}^{(1)}$ is "virtually identical" to the first GCVs obtained by iterative procedures for the SUMCOR method by Horst [4] and the SSQCOR and GENVAR methods by Kettenring. The same is true for the first GCC matrices for these methods, which are all almost identical to

$$\mathbf{\Sigma}^{(1)} = \begin{bmatrix} 1.000 & 0.735 & 0.756 \\ & 1.000 & 0.743 \\ & & 1.000 \end{bmatrix}.$$

For higher-stage GCVs further iterative computations are required (see Horst [4] and Kettenring [7] for details).

INTERPRETATIONS

Although all the original variables are needed to obtain the GCVs, the final result would indicate that only a few GCVs need to be retained. All future analysis can be limited to these retained GCVs. For example, these GCVs, instead of **X**, may constitute the variables in future regression analysis. A meaningful reduction of dimensionality is achieved.

The GCVs obtained above as linear combinations of the original variables have in most cases very little practical meaning by themselves. However, one can attempt to interpret the GCVs (see Timm [16] for the case of CCs) via the correlations of the linear compounds with the corresponding elements involved, i.e., correlations of $Y_i^{(t)}$ with X_{ij}s, $i = 1, \ldots, k$, $j = 1, \ldots, p_i$, $t = 1, \ldots, s$.

References

[1] Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York. (An excellent presentation of the theory of CVs.)

[2] Carroll, J. D. (1968). *Proc. Amer. Psychol. Ass.*, 227–228.

[3] Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. Wiley, New York.

[4] Horst, P. (1961). *Psychometrika*, **26**, 129–149.

[5] Horst, P. (1965). *Factor Analysis of Data Matrices*. Holt, Rinehart and Winston, New York.

[6] Hotelling, H. (1936). *Biometrika*, **28**, 321–377.

[7] Kettenring, J. R. (1971). *Biometrika*, **58**, 433–451. (Detailed study of construction of GCVs by five methods. Also gives interesting factor analytic interpretations.)

[8] Lee, S. Y. (1978). *Psychometrika*, **43**, 427–431. (Errata 1979, *Psychometrika*, **44**, 131.)

[9] McDonald, R. P. (1968). *Psychometrika*, **33**, 341–381.

[10] McKeon, J. J. (1966). *Psychometric Monogr.*, **13**. (Discusses applications of CVs.)

[11] Sen Gupta, A. (1980). Generalized Correlations in the Singular Case. *Tech. Rep. No. 46* (under Contract N00014-75-C-0442), Dept. of Statistics, Stanford University, Stanford, Calif.

[12] Sen Gupta, A. (1982). On the Problems of Construction and Statistical Inference Associated with a Generalization of Canonical Variables. *Tech. Rep. No. 168* (under NSF Grant MCS 78-07736), Dept. of Statistics, Stanford University, Stanford, Calif. (Introduces and formulates aspects of statistical inference related to GCVs and to new GCVs in particular.)

[13] Sen Gupta, A. (1982). On Tests for the Equicorrelation Coefficient and the Generalized Variance of Symmetric Standard Multivariate Normal Distribution. *Tech. Rep. No.* 55 (under Contract N00014-75-C-0442), Dept. of Statistics, Stanford University, Stanford, Calif.

[14] Steel, R. G. D. (1951). *Ann. Math. Statist.*, **22**, 456–460.

[15] Thurstone, L. L. and Thurstone, T. G. (1941). *Psychometric Monogr.*, **2**.

[16] Timm, N. H. (1975). *Multivariate Analysis with Applications in Education and Psychology*. Brooks/Cole, Monterey, Calif. (Discusses conditional GCCs.)

(CANONICAL ANALYSIS
GENERALIZED VARIANCE)

ASHIS SEN GUPTA

GENERALIZED HYPERGEOMETRIC DISTRIBUTIONS

Hypergeometric distributions* are defined by the probabilities

$$p(x) = \binom{M}{x}\binom{N-M}{n-x}\Big/\binom{N}{n} = \binom{n}{x}\binom{N-n}{M-x}\Big/\binom{N}{M} \qquad (1)$$

for $\max(0, n + M - N) \leqslant x \leqslant \min(n, M)$, where N, M, and n are positive integers. If $N \geqslant M + n$, this is written

$$p(x) = \frac{(N-M)!\,(N-n)!}{N!\,(N-M-n)!} \times \frac{(-M)^{[x]}(-n)^{[x]}}{(N-M-n+1)^{[x]}x!},$$

and the factor depending on x is a term of