This article was downloaded by: [National Chiao Tung University 國立交通大學]
On: 27 April 2014, At: 20:42
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Multivariate Behavioral Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hmbr20

# A Comparative Study of Power and Sample Size Calculations for Multivariate General Linear Models

Gwowen Shieh

Published online: 10 Jun 2010.

To cite this article: Gwowen Shieh (2003) A Comparative Study of Power and Sample Size Calculations for Multivariate General Linear Models, Multivariate Behavioral Research, 38:3, 285-307, DOI: 10.1207/S15327906MBR3803_01

To link to this article: http://dx.doi.org/10.1207/S15327906MBR3803_01

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

*Multivariate Behavioral Research*, *38* (3), 285-307
Copyright © 2003, Lawrence Erlbaum Associates, Inc.

# A Comparative Study of Power and Sample Size Calculations for Multivariate General Linear Models

Gwowen Shieh
Department of Management Science
National Chiao Tung University

Repeated measures and longitudinal studies arise often in social and behavioral science research. During the planning stage of such studies, the calculations of sample size are of particular interest to the investigators and should be an integral part of the research projects. In this article, we consider the power and sample size calculations for normal outcomes within the framework of multivariate general linear models that represent the most fundamental method for the analysis of repeated measures and longitudinal data. Direct extensions of the existing generalized estimating equation and likelihood-based approaches are presented. The major feature of the proposed modification is the accommodation of both fixed and random models. A child development example is provided to illustrate the usefulness of the methods. The adequacies of the sample size formulas are evaluated through Monte Carlo simulation study.

## *Introduction*

Repeated measures and longitudinal studies arise often in social and behavioral analyses, in which repeated observations of a response variable and a set of independent variables are recorded on subjects across occasions. Because repeated observations are recorded on the same subject, the response variables are usually correlated. Methods for analyzing correlated data have recently received considerable attention in the literature, see Keselman, Algina and Kowalchuk (2001) for a comprehensive review and their references for related discussions. The generalized estimating equation (GEE) approach proposed by Liang and Zeger (1986), in particular, is widely used by researchers in a number of fields for the analysis of longitudinal data. In recent years, a number of articles are devoted to exemplifying the use of the GEE method. For application of the GEE method relevant to behavioral studies, see Duncan et al. (1995). Two of the attractive properties of the GEE approach are that it

---

This research was partially supported by the National Science Council. I wish to convey my appreciation to the reviewers, whose suggestions extended and strengthened the article's content immensely.

accommodates both discrete and continuous data, and it also allows for the flexibility in the correlation structure under a single framework. Furthermore, it only requires specification of the forms of the first two moments. The full joint distribution of correlated responses is not required.

Sample size calculations and power analyses are often critical for researchers to address specific scientific hypotheses and confirm credible treatment effects. Accordingly, it is of practical importance to be able to perform these tasks in a GEE setting. Liu and Liang (1997) proposed a sample size and power formula derived from the score statistic. Shih (1997) and Rochon (1998) proposed similar Wald test procedures to compute sample sizes and statistical powers in the context of GEE. All these GEE-based approaches represent a unified tool for both normal and non-normal responses of repeated measures and longitudinal studies. These procedures have been illustrated in these articles with simulated and real data sets for binary, count and normal outcomes. Due to the complex nature, however, there is little information on the discrepancy of these methods. It would be helpful to make direct comparisons of these approaches in terms of both calculated sample sizes and precision of estimated power with a familiar content, such as the commonly used analysis of variance models in the repeated measures studies of continuous outcome variables.

In this article, we examine the adequacy of the sample size formula for Gaussian outcomes of multivariate general linear models. Since the GEE approach is closely related to quasi-likelihood methods, a question of interest is how the GEE approach compares with the methods obtained from a likelihood-based viewpoint when the correlated responses have a joint multivariate normal distribution. Therefore, the existing likelihood-based approaches of O'Brien and Shieh (1992) are also discussed. However, it is important to note that these approaches are applicable to fixed (conditional) models that assumed all the levels of the independent variables to be predetermined before data collection. The results would be specific to the particular values of the independent variables that are observed or preset by the researcher. However, it is quite common in behavioral research to have studies in which the levels of the independent variables for each experimental unit cannot be controlled and are available only after the observations are made. These models are referred to as random (unconditional) models. A natural generalization to incorporate both fixed and random independent variables should be essential to these approaches for performing power and sample size calculations in practice. According to theoretical arguments, the GEE and likelihood-based approaches are modified to accommodate the two types of models.

Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

To test the multivariate general linear hypothesis, the test statistics and corresponding modification of sample size formulas of both the GEE and likelihood-based approaches are presented. These methods are illustrated with the child development example from Muller, LaVange, Ramey, & Ramey (1992). Since all the approaches considered here use large sample approximations, simulation studies are conducted to assess their adequacy for finite sample and robustness for noncentrality structure under various model configurations.

*The Fixed Model*

Consider the standard multivariate general linear model with all the levels of independent variables fixed a priori

$$\mathbf{Y} = \mathbf{XB} + \boldsymbol{\varepsilon}, \quad (1)$$

where $\mathbf{Y} = (\mathbf{y}_1, ..., \mathbf{y}_N)^T$ is a $N \times p$ matrix with $\mathbf{y}_i$ as the $p \times 1$ vector of observed sequence of measurements for the $i^{th}$ subject; $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_N)^T$ is a $N \times r$ design matrix with full column rank $r < N$, where $\mathbf{x}_i$ is the $r \times 1$ vector of independent variables associated with the $i^{th}$ subject; $\mathbf{B}$ is the $r \times p$ matrix of unknown regression coefficients; and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, ..., \boldsymbol{\varepsilon}_N)^T$ is a $N \times p$ matrix with $\boldsymbol{\varepsilon}_i$ as the $p \times 1$ vector of random errors associated the $i^{th}$ subject, for $i = 1, ..., N$. The errors $\boldsymbol{\varepsilon}_i$ are assumed to have independent and identical normal distribution $N_p(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a $p \times p$ positive-definite covariance matrix. We are concerned with the general linear hypothesis $H_0$: $\mathbf{CBA} = \boldsymbol{\Psi}_0$, where $\mathbf{C}$ is the $c \times r$ matrix of between-subject contrasts with full row rank $c \leq r$, and $\mathbf{A}$ is the $p \times a$ matrix of within-subject contrasts with full column rank $a \leq p$. The maximum likelihood estimators for $\mathbf{B}$ and $\boldsymbol{\Sigma}$ are $\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ and $\hat{\boldsymbol{\Sigma}} = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^T(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})/N$, respectively. The common statistics for $H_0$ are obtained from the eigenvalues of $\mathbf{E}^{-1}\mathbf{H}$, where $\mathbf{E} = (\mathbf{YA} - \mathbf{X}\hat{\mathbf{B}}\mathbf{A})^T(\mathbf{YA} - \mathbf{X}\hat{\mathbf{B}}\mathbf{A})$ and $\mathbf{H} = (\mathbf{C}\hat{\mathbf{B}}\mathbf{A} - \boldsymbol{\Psi}_0)^T[\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T]^{-1}(\mathbf{C}\hat{\mathbf{B}}\mathbf{A} - \boldsymbol{\Psi}_0)$. It follows from standard results that $\mathbf{E}$ has the Wishart distribution $W_a(N - r, \mathbf{A}^T\boldsymbol{\Sigma}\mathbf{A})$, $\mathbf{H}$ has the Wishart distribution $W_a(c, \mathbf{A}^T\boldsymbol{\Sigma}\mathbf{A}, \boldsymbol{\delta})$, and $\mathbf{E}$ and $\mathbf{H}$ are independent, where $\boldsymbol{\delta} = (\mathbf{A}^T\boldsymbol{\Sigma}\mathbf{A})^{-1}(\mathbf{CBA} - \boldsymbol{\Psi}_0)^T[\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T]^{-1}(\mathbf{CBA} - \boldsymbol{\Psi}_0)$ is the noncentrality parameter matrix. See Timm (2002) and Morrison (1990) for further discussion of the Wilks likelihood ratio $|\mathbf{E}(\mathbf{E} + \mathbf{H})^{-1}|$, Pillai trace $tr[\mathbf{H}(\mathbf{E} + \mathbf{H})^{-1}]$ and Hotelling-Lawley trace $tr(\mathbf{E}^{-1}\mathbf{H})$ test statistics, where $tr(\bullet)$ is the trace of a matrix.

Note that the analysis of repeated observations can be viewed as a regression model with correlated errors. This is exactly the formulation of the GEE approach. A brief review of the GEE methods is provided in the



Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

Appendix. Next, we present the important details of the test statistics of the GEE and likelihood-based approaches.

*The GEE Approach*

With the subject-wise orientation, the following results are derived through the GEE setting of Rochon (1998) and Liu and Liang (1997) for the multivariate general linear model shown by Equation 1. The Wald test statistic or generalized $T_0^2$ statistic proposed in Rochon is

$$(2) \qquad Q_W = N \bullet tr(\mathbf{E}^{-1}\mathbf{H})$$

where $tr(\mathbf{E}^{-1}\mathbf{H})$ is the Hotelling-Lawley trace mentioned above. Furthermore, it can be shown that the quasi-score statistic or Rao's score statistic in Liu and Liang (1997) is

$$(3) \qquad Q_S = N \bullet tr[\mathbf{H}(\mathbf{E} + \mathbf{H})^{-1}]$$

where $tr[\mathbf{H}(\mathbf{E} + \mathbf{H})^{-1}]$ is the aforementioned Pillai trace. The actual test is performed by referring the test statistics $Q_W$ and $Q_S$ to their asymptotic distribution under $H_0$, which is a central chi-square distribution with *ca* degrees of freedom.

In order to evaluate the power under alternative hypothesis, the distributions of both statistics shown by Equations 2 and 3 are approximated by a noncentral chi-square distribution with *ca* degrees of freedom. It follows from Equation 6 of Rochon (1998) and Equation 4 of Liu and Liang (1997) that the noncentrality parameter $\nu = tr(\boldsymbol{\delta})$, where $\boldsymbol{\delta}$ is defined earlier. In this case, the asymptotic property of Rochon's $Q_W$ and Liu and Liang's $Q_S$ agree with that of the Hotelling-Lawley trace and Pillai trace, respectively (see Anderson, 1984, p. 330 and Seber, 1984, p. 415 for details). Consequently, the Wald statistic $Q_W$ and Rao's statistic $Q_S$ have identical asymptotic distributions under both null and alternative hypotheses.

*The Likelihood-based Approach*

To test a general linear hypothesis under multivariate general linear model shown by Equation 1, the common statistics are computed from $|\mathbf{E}(\mathbf{E} + \mathbf{H})^{-1}|$, $tr[\mathbf{H}(\mathbf{E} + \mathbf{H})^{-1}]$ and $tr(\mathbf{E}^{-1}\mathbf{H})$, and their critical values have been widely tabled and charted (for example, see Seber, 1984, pp. 562-564). However, in practice, more tractable ones have been proposed by transforming them to *F*-type statistics. In this article, we focus on *F*



Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

approximations of the Hotelling-Lawley trace statistic $tr(\mathbf{E}^{-1}\mathbf{H})$ for two principle reasons. First, it resembles very much the aforementioned statistic (Equation 2). Second, it has more desired properties than the other competing tests. To be specific, two transformations of $tr(\mathbf{E}^{-1}\mathbf{H})$ are considered here. Pillai and Samson (1959) proposed

$$(4) \qquad F_{T1} = df2_{T1} tr(\mathbf{E}^{-1}\mathbf{H})/sca$$

where $df2_{T1} = s(N - r - a - 1) + 2$ and $s = \text{minimum}(c, a)$. Moreover, McKeon (1974) presented

$$(5) \qquad F_{T2} = df2_{T2} tr(\mathbf{E}^{-1}\mathbf{H})/hca,$$

where $df2_{T2} = (ca + 2)g + 4$, $g = [(N - r)^2 - (N - r)(2a + 3) + a(a + 3)]/[(N - r)(c + a + 1) - (c + 2a + a^2 - 1)]$, and $h = (df2_{T2} - 2)/(N - r - a - 1)$. Under the null hypothesis, both $F_{T1}$ and $F_{T2}$ are compared to an $F$ distribution with numerator degrees of freedom $ca$, and denominator degrees of freedom $df2_{T1}$ and $df2_{T2}$, respectively.

To approximate the distributions of $F_{T1}$ and $F_{T2}$ under the alternative hypothesis, O'Brien and Shieh (1992) described a direct extension of the respective $F$ distributions to their noncentral counterpart with the noncentrality parameter $\nu$ as mentioned earlier. These procedures are motivated and partially justified by the knowledge of the following facts. First, a noncentral $F$ distribution with numerator and denominator degrees of freedom ($df1$, $df2$) and noncentrality parameter $\Delta$ converges to a noncentral chi-square distribution with degrees of freedom $df1$ and noncentrality parameter $\Delta$ when $df2$ tends to infinity. In this case, both $F_{T1}$ and $F_{T2}$ converge to the same asymptotic distribution of $N \bullet tr(\mathbf{E}^{-1}\mathbf{H})$, which is a noncentral chi-square distribution with degrees of freedom $ca$ and noncentrality parameter $\nu$. Second, when $s = 1$, the proposed noncentral $F$ distribution is exact for both $F$-type statistics. Although similar transformation has been presented by Muller and Peterson (1984) and Muller et al. (1992), their method provides smaller noncentrality values for $s > 1$ as discussed in O'Brien and Shieh (1992).

### *The Random Model*

Traditionally, we treat the model described in Equation 1 as a conditional random model because the values of the independent variables are fixed and known. It is important to recognize and account for the extra variability stemming from the fact that, in another replication of the same study, different settings for the independent variables will be obtained. Thus, the



Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

random model is more appropriate for the studies in which the values of the independent variables cannot be predetermined. We now define the formal random model associated with the multivariate general linear model.

Assume that the independent variables ($\mathbf{x} = \mathbf{x}_i$, $i = 1, ..., N$) follow a distribution $f(\mathbf{x})$ with finite moments. The form of $f(\mathbf{x})$ is assumed to be depended on none of the unknown parameters $\mathbf{B}$ and $\mathbf{\Sigma}$. It follows from the standard asymptotic result that

$$\mathbf{X}^T\mathbf{X} / N = \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^T / N$$

converges in probability to $\mathbf{K}$ where $\mathbf{K} = E_{\mathbf{x}}(\mathbf{x}\mathbf{x}^T)$ and $E_{\mathbf{x}}(\bullet)$ denotes the expectation taken with respect to the distribution of $\mathbf{x}$. With this additional assumption and the application of Slutsky's Theorem, it can be demonstrated that $N^{1/2}(\hat{\mathbf{B}} - \mathbf{B}) = N^{1/2}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}$ has a limiting matrix normal distribution $N_{r \times p}(\mathbf{0}, \mathbf{K}^{-1}, \mathbf{\Sigma})$. Similarly, $[\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T]^{-1/2}(\mathbf{C}\hat{\mathbf{B}}\mathbf{A} - \mathbf{CBA})$ has a limiting matrix normal distribution $N_{c \times a}(\mathbf{0}, \mathbf{I}_c, \mathbf{A}^T\mathbf{\Sigma}\mathbf{A})$, where $\mathbf{I}_c$ is the $c \times c$ identity matrix. Hence, under the null hypothesis $H_0$: $\mathbf{CBA} = \mathbf{\Psi}_0$, $\mathbf{H}$ converges in distribution to the Wishart distribution $W_a(c, \mathbf{A}^T\mathbf{\Sigma}\mathbf{A}, \mathbf{0})$. It can be shown that the distribution of $\mathbf{E}$ is the Wishart distribution $W_a(N - r, \mathbf{A}^T\mathbf{\Sigma}\mathbf{A})$ under both hypotheses for the fixed and random models. Consequently, under the null hypothesis, the proposed central chi-square and $F$ approximations for the four statistics $Q_W$, $Q_S$, $F_{T1}$ and $F_{T2}$ described above are extended to the case of random independent variables.

For the statistics under the alternative hypothesis, we note that $[\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T]^{-1/2}(\mathbf{C}\hat{\mathbf{B}}\mathbf{A} - \mathbf{\Psi}_0)$ and $N^{1/2}(\mathbf{C}\mathbf{K}^{-1}\mathbf{C}^T)^{-1/2}(\mathbf{C}\hat{\mathbf{B}}\mathbf{A} - \mathbf{\Psi}_0)$ are equivalent in asymptotic distribution. For the purpose of relating asymptotic power function calculations to the local alternatives to $H_0$, an asymptotically equivalent distribution of $[\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T]^{-1/2}(\mathbf{C}\hat{\mathbf{B}}\mathbf{A} - \mathbf{\Psi}_0)$ can be defined in the form of

$$N_{c \times a}[N^{1/2}(\mathbf{C}\mathbf{K}^{-1}\mathbf{C}^T)^{-1/2}(\mathbf{CBA} - \mathbf{\Psi}_0),\ \mathbf{I}_c,\ \mathbf{A}^T\mathbf{\Sigma}\mathbf{A}],$$

in which case, we propose to consider the distribution of $\mathbf{H}$ with the operational and asymptotically equivalent Wishart distribution $W_a(c, \mathbf{A}^T\mathbf{\Sigma}\mathbf{A}, N\bar{\boldsymbol{\delta}})$ and $\bar{\boldsymbol{\delta}} = (\mathbf{A}^T\mathbf{\Sigma}\mathbf{A})^{-1}(\mathbf{CBA} - \mathbf{\Psi}_0)^T(\mathbf{C}\mathbf{K}^{-1}\mathbf{C}^T)^{-1}(\mathbf{CBA} - \mathbf{\Psi}_0)$.

We proceed to approximate the distributions of the GEE-based statistics $Q_W$ and $Q_S$ by a noncentral chi-square distribution with $ca$ degrees of freedom. The noncentrality parameter used in the approximation is defined as

$$\nu_N = N\bar{\nu}, \tag{6}$$



Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

where $\bar{\nu} = tr(\bar{\boldsymbol{\delta}})$. Similarly, we approximate the distributions of the likelihood-based statistics $F_{T1}$ and $F_{T2}$ by noncentral $F$ distributions with noncentral parameter $\nu_N$, common numerator degrees of freedom $ca$ and denominator degrees of freedom $df2_{T1}$ and $df2_{T2}$ defined in Equations 3 and 4, respectively. Essentially, the noncentral parameter $\nu_N$ is the counterpart of $\nu$ with substitution of $\mathbf{X}^T\mathbf{X}$ with $N\mathbf{K}$.

Note that the probability distribution $f(\mathbf{x})$ of $\mathbf{x}$ can be discrete, continuous and both. In general, there is no simple and tractable expression for the distribution $\hat{\mathbf{B}}$ except in some special cases. For the common additional assumption that $\mathbf{x}$ has a multivariate normal distribution as in Sampson (1974), it can be shown that $\hat{\mathbf{B}}$ follows a matrix $t$ distribution, see Dickey (1967) for detailed discussion. As described in Dickey (1967), the limiting distribution of a matrix $t$ agrees with the matrix normal distribution. Therefore, all the asymptotic properties claimed previously for $\hat{\mathbf{B}}$ and related statistics still apply under the normality assumption of $\mathbf{x}$.

## *Sample Size Calculations*

### *The GEE Approach*

The actual implementation of sample size calculations of the GEE approaches using the statistics $Q_S$ and $Q_W$ is as follows. With specified parameter values $\mathbf{B}$ and $\boldsymbol{\Sigma}$, and chosen probability distribution $f(\mathbf{x})$, the sample size needed to test hypothesis $H_0$: $\mathbf{CBA} = \boldsymbol{\Psi}_0$ with specified significance level $\alpha$ and power $1 - \beta$ against the alternative $H_1$: $\mathbf{CBA} \neq \boldsymbol{\Psi}_0$ is determined by the following two steps. First, find the noncentrality $\nu_N$ of a noncentral chi-square distribution with $ca$ degrees of freedom such that its $100{\bullet}\beta^{th}$ percentile is equivalent to the $100(1 - \alpha)^{th}$ percentile of a central chi-square distribution with $ca$ degrees of freedom. Second, the sample size estimate is computed as $\nu_N / \bar{\nu}$, where $\bar{\nu}$ is defined in Equation 6. Note that both methods result in the same estimated sample size.

Note that although the notion of fixed values for independent variables was introduced in Liu and Liang (1997), their presentation and simulation studies did not emphasize the special feature of random independent variable. Moreover, our proposed formulation naturally accommodates the extension of unequal allocation in Rochon (1998). Therefore, the approach described here differs from that of Liu and Liang (1997) and Rochon (1998) with respect to the independent variables distribution where they proposed to consider only the fixed model. This is naturally extended here to the random case and the distribution $f(\mathbf{x})$ could be either discrete or continuous with a finite number of levels such as Bernoulli or multinomial distributions or an infinite number of



Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

values, for example, Poisson and normal distributions. If $\mathbf{x}$ is discrete with $m$ distinct values $\mathbf{x}_{uj}$ and $f(\mathbf{x}_{uj}) = \pi_j$, $j = 1, \ldots, m$, where

$$\sum_{j=1}^{m} \pi_j = 1,$$

then

$$\mathbf{K} = E_{\mathbf{x}}(\mathbf{x}\mathbf{x}^T) = \sum_{j=1}^{m} \pi_j \mathbf{x}_{uj} \mathbf{x}_{uj}^T.$$

Accordingly, this representation is also valid for the case of $\mathbf{x}_{uj}$ being considered as fixed levels rather than random component as in Liu and Liang (1997) and Rochon (1998).

*The Likelihood-based Approach*

In a manner analogous to the GEE approach described above, we keep the same assumption that the independent variables $\mathbf{x}$ has a joint probability function $f(\mathbf{x})$, which can be either discrete or continuous, and propose the factorization defined in Equation 6 for the noncentral parameter $\nu_N$ of the noncentral $F$ approximations of $F_{T1}$ and $F_{T2}$ under the alternative hypothesis. Therefore, this procedure becomes a direct generalization of O'Brien and Shieh (1992) in which their results are limited to those applications where all the levels of the independent variables in the model are fixed in advance. Hence, given parameter values $\mathbf{B}$ and $\boldsymbol{\Sigma}$, chosen probability distribution $f(\mathbf{x})$, and sample size $N$, the statistical power achieved for testing hypothesis $H_0$: $\mathbf{CBA} = \boldsymbol{\Psi}_0$ with specified significance level $\alpha$ against the alternative $H_1$: $\mathbf{CBA} \neq \boldsymbol{\Psi}_0$ is the probability

$$P[F(ca,\ df2,\ N\bar{\nu}) > F_{ca,\ df2,\ \alpha}]$$

with $df2 = df2_{T1}$ and $df2_{T2}$ for the two $F$-type test statistics $F_{T1}$ and $F_{T2}$ defined in Equations 4 and 5, respectively, where $F(df1, df2, \Delta)$ denotes a noncentral $F$ random variable with $(df1, df2)$ degrees of freedom and noncentrality parameter $\Delta$, and $F_{df1, df2, \alpha}$ denotes the $100(1-\alpha)^{\text{th}}$ percentile of a central $F$ random variable with $(df1, df2)$ degrees of freedom. Note that this procedure can be reversed to calculate the sample size needed in order to attain the specified power. However, it usually involves an iterative process to find the solution because both $F(df1, df2, \Delta)$ and $F_{df1, df2, \alpha}$ depend on the sample size $N$.

Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014



*An Example*

We illustrate in this section the power and sample size calculation procedures in a child development example that has motivated our work. Muller et al. (1992) presented a power analysis for a longitudinal study of a child's intellectual performance as a function of mother's estimated verbal intelligence. With child IQ measurements at 12, 24, and 36 months ($p$ = 3), and with intercept, linear, quadratic, and cubic trends in mother's standardized IQ (MSIQ) as independent variables ($r$ = 4), this yields

$$\mathbf{x} = \begin{bmatrix} 1 \\ MSIQ \\ MSIQ^2 \\ MSIQ^3 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} \beta_{I.12} & \beta_{I.24} & \beta_{I.36} \\ \beta_{L.12} & \beta_{L.24} & \beta_{L.36} \\ \beta_{Q.12} & \beta_{Q.24} & \beta_{Q.36} \\ \beta_{C.12} & \beta_{C.24} & \beta_{C.36} \end{bmatrix},$$

where $\beta_{I.t}$ is the intercept, while $\beta_{L.t}$, $\beta_{Q.t}$, and $\beta_{C.t}$ are the corresponding coefficients of linear, quadratic, and cubic values of MSIQ for time $t$ = 12, 24, and 36, respectively. Here the hypothesized relationship between mother and child competence of interest corresponds to a test of the time × mother's IQ interaction. Hence, the between-subject and within-subject contrast matrices ($c$ = 3 and $a$ = 2) are

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{A} = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{6} \\ 0 & -2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix},$$

respectively. According to the previous study of the Infant Health and Development Program (IHDP, see Ramey, Bryant, Wasik, Sparling, Fendt, & LaVange, 1992), the model parameter estimates are

$$\hat{\mathbf{B}} = \begin{bmatrix} 114.46 & 104.66 & 98.83 \\ 2.88 & 8.77 & 10.67 \\ -0.71 & -0.90 & -1.30 \\ -0.21 & -0.54 & -0.72 \end{bmatrix} \text{ and } \hat{\mathbf{\Sigma}} = \begin{bmatrix} 218.48 & 83.66 & 72.19 \\ 83.66 & 251.92 & 158.60 \\ 72.19 & 158.60 & 244.58 \end{bmatrix}.$$

As for the values of the independent variables, one of the three schemes assumed that through screening the sample would be evenly distributed across the four groups of mother's IQ: namely retarded (IQ < 70), borderline

(IQ 70-85), normal (IQ 85-100) and high (IQ > 100). Furthermore, it was also assumed that the spread of mother's IQ scores within each group would follow the IHDP pattern, and the actual mother's IQ scores were treated as continuous values in the power analysis. At first sight, this may produce reasonable approximation to the true mother's IQ score distribution of the current study. However, the authors considered it impractical and expected an increase in power associated with over-sampling of the extreme values of mother's IQ. The primary reason of inducing such scheme and other approximations is due to the lack of a proper procedure that accounts for the nature of continuous distribution. Therefore, the absence of consensus in determining the discretization of continuous independent variables distribution and the failure to embed the method in a general setup are obvious limitations of the existing approaches in Muller et al. (1992), O'Brien and Shieh (1992), Liu and Liang (1997) and Rochon (1998).

For illustrative purpose, we assume that the mother's standardized IQ (MSIQ) has a standard normal distribution. Note that the mother's IQ score has a population mean 100 and standard deviation 15. Thus, it follows that

$$\mathbf{K} = E_X(\mathbf{x}\mathbf{x}^T) = \begin{bmatrix} 1 & m_1 & m_2 & m_3 \\ m_1 & m_2 & m_3 & m_4 \\ m_2 & m_3 & m_4 & m_5 \\ m_3 & m_4 & m_5 & m_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 3 \\ 1 & 0 & 3 & 0 \\ 0 & 3 & 0 & 15 \end{bmatrix},$$

where $m_i$ is the $i^{\text{th}}$ moment of a standard normal distribution. We are interested in how many subjects are needed to detect the time × mother's IQ interaction $H_0$: $\mathbf{CBA} = 0$ in terms of the matrices $\mathbf{C}$ and $\mathbf{A}$ just stated. With these specifications, it follows from Equation 6 that $\bar{\nu} = 0.1328$. Assuming the significance level $\alpha = 0.05$, the sample size estimates of the Wald test (Equation 2) and Rao's test (Equation 3) and $F$ transforms (Equations 4 and 5) are 132, 132, 135 and 137, respectively, for power $1 - \beta = 0.90$, while the corresponding sample size estimates are 158, 158, 161 and 162, respectively, for power $1 - \beta = 0.95$. The achieved power levels of the four tests are 0.9858, 0.9858, 0.9843 and 0.9836 for a given sample size $N = 200$. These numbers are comparable with the results obtained from Muller et al. (1992) shown in Table 6.

With these sample size and power calculations, the most powerful approach can be easily identified. However, the proposed approaches use large sample approximations, there is no guarantee that the one that gives higher power will always be more accurate in achieving the nominal power. Hence, we continue to compare the accuracy of these formulas in terms of



Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

the discrepancy between estimated actual power and nominal power, where they all use the same sample size. This is demonstrated in the following simulation studies.

*Simulation Studies*

In order to evaluate the accuracy of the proposed approaches, Monte Carlo simulation studies are performed for three different multivariate general linear models with Gaussian responses. To reinforce the concept of fixed or random models and to emphasize the flexibility of the proposed methods, the three models are chosen as follows.

*Fixed MANOVA Model*

For the fixed MANOVA model, we consider two designs for the standard MANOVA model with equal group sizes. The first design has $r = 4$, whereas the second design is set to have $r = 3$. For the first design, the vector of independent variable contains only indicator variables, taking the values of zero or one, $\mathbf{x} = [1\ 0\ 0\ 0]^T$, $[0\ 1\ 0\ 0]^T$, $[0\ 0\ 1\ 0]^T$ or $[0\ 0\ 0\ 1]^T$, and each of these vectors is replicated with the same number of times in the design matrix $\mathbf{X}$. In this case, it is trivial that $\mathbf{X}^T\mathbf{X} = (N/4)\mathbf{I}_4$. Hence, even for the fixed independent variable structure, the noncentrality parameter can still be written in the form $\nu_N$ in Equation 6 with $\mathbf{K} = (1/4)\mathbf{I}_4$. We are interested in a test of no group effects ($c = 3$) with

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}.$$

Accordingly, the independent variable vector in the second design is $\mathbf{x} = [1\ 0\ 0]^T$, $[0\ 1\ 0]^T$ or $[0\ 0\ 1]^T$. Furthermore, $\mathbf{X}^T\mathbf{X} = (N/3)\mathbf{I}_3$ and

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \text{with } c = 2.$$

*Random MANOVA Model*

In contrast to the fixed MANOVA model, we generalize the configurations of independent variables into a random setting of discrete uniform distribution



Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

with $\pi_j = 1/r$, $j = 1, ..., r$. Hence, the design matrix $\mathbf{X}$ is now composed of vectors, which follow a discrete uniform distribution. It is interesting to note that $\mathbf{K} = E_x(\mathbf{xx}^T) = (1/r)\mathbf{I}_r$ for $r = 3$ and 4. Hence, for any one of the four proposed approaches, the sample size formulas are identical for the first two models.

*Development Model*

The third model is patterned after the prescribed child development data set with $r = 4$ repeated measures. We consider the vector of independent variables composed of powers of a standard normal variable, specifically $\mathbf{x} = [1\ z\ z^2\ z^3]^T$, where $z$ has a standard normal distribution. Hence, the matrix $\mathbf{K}$ has the same form as defined in the previous section for the child development example. Here, the concern is whether there is a linear, quadratic and cubic trend relationship between response and the between-subject contrast matrix is defined as

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{with } c = 3.$$

As in the previous two models, two different designs are studied as well.

Note that all four sample size formulas depend on the identical noncentrality parameter $\nu_N$ as defined in Equation 6, which in turn relies on $tr(\bar{\boldsymbol{\delta}})$ or the sum of eigenvalues of $\bar{\boldsymbol{\delta}}$, where $\bar{\boldsymbol{\delta}} = (\mathbf{A}^T\boldsymbol{\Sigma}\mathbf{A})^{-1}(\mathbf{CBA} - \boldsymbol{\Psi}_0)^T(\mathbf{CK}^{-1}\mathbf{C}^T)^{-1}(\mathbf{CBA} - \boldsymbol{\Psi}_0)$. Therefore, without loss of generality, we assume $\mathbf{A}^T\boldsymbol{\Sigma}\mathbf{A} = \mathbf{I}_a$, and $\boldsymbol{\Psi}_0$ is a $c \times c$ null matrix throughout the simulation study. Furthermore, we let $\bar{\boldsymbol{\delta}} = diag(\bar{\delta}_1, ..., \bar{\delta}_a)$, a diagonal matrix be with diagonal elements $(\bar{\delta}_1, ..., \bar{\delta}_a)$, where $(\bar{\delta}_1, ..., \bar{\delta}_a)$ are the eigenvalues of $\bar{\boldsymbol{\delta}}$ and

$$\text{tr}(\bar{\boldsymbol{\delta}}) = \sum_{l=1}^{a} \bar{\delta}_l.$$

For each of the three models described above, the values of $a$ are set as 3 and 2 for the two designs, respectively. In order to study the robustness properties of the sample size formulas with respect to a wide range of designs and sample sizes, the parameter matrix $\mathbf{BA} = (\mathbf{B}_1^T, \mathbf{B}_2^T)^T$ is defined such that $(\bar{\delta}_1, ..., \bar{\delta}_a)$ has the following four different structures:



Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

$$\text{Equal: } \bar{\delta}_l = \bar{\nu}^*/a,\ l = 1, ..., a;$$

$$\text{Linear: } \bar{\delta}_l = l \cdot \bar{\nu}^*/(\sum_{l=1}^{a} l),\ l = 1, ..., a;$$

$$\text{Geometric: } \bar{\delta}_l = 2^{l-1} \cdot \bar{\nu}^*/(\sum_{l=1}^{a} 2^{l-1}),\ l = 1, ..., a;$$

$$\text{Extreme: } \bar{\delta}_1 = \bar{\nu}^*, \text{ and } \bar{\delta}_l = 0,\ l = 2, ..., a.$$

The value of $\bar{\nu}^*$ is predetermined by $P[F(ca, df2_{T1}, N\bar{\nu}^*) > F_{ca, df2_{T1}, 0.05}] = 0.70$ for each design. In fact, for simplicity, $\mathbf{B}_1$ is set as a $1 \times a$ null vector in all cases, and $\mathbf{B}_2$ can thus be decided, since it is the only unknown element in $\bar{\boldsymbol{\delta}}$ for each of the four different eigenvalue structures. It is important to note that the equal and extreme eigenvalue structures cover the two opposite cases of ($\bar{\delta}_1$, ..., $\bar{\delta}_a$) for all possible combinations of $\mathbf{A}$, $\boldsymbol{\Sigma}$, $\mathbf{B}$, and $\boldsymbol{\Psi}_0$ under the specifications of $\mathbf{C}$ and $\mathbf{K}$. Obviously, the other two structures stand for some of the intermediate situations. Given the sample sizes $N$ and eigenvalues structure $\bar{\nu}^*$, the nominal powers of the four approaches can be computed according to the procedures described earlier. Note that the nominal power of the $F_{T1}$ statistic is always 0.70. In general, the nominal power of $F_{T2}$ statistic is slightly lower, whereas the nominal power of the Wald and Rao's statistics tends to be higher than 0.70.

Estimates of actual Type I error rate and power associated with the given sample sizes and model configurations are then computed through Monte Carlo simulation using 5000 replicate data sets. The adequacy of the sample size formula is determined by the difference between the estimated and nominal values of Type I error rate and power. All calculations are performed using programs written with SAS/IML (SAS Institute, 1999). The results of the simulation studies are presented in Tables 1–6 for the three models and each with two different designs. We also conducted the simulation study for two different levels of power, namely $P[F(ca, df2_{T1}, N\bar{\nu}^*) > F_{ca, df2_{T1}, 0.05}] = 0.50$ and 0.95. The results are similar to those of power 0.70, and are thus not shown here.

As would be expected, the results in Tables 1–6 suggest that the accuracy of the competing approaches increases with the sample size, and varies with the structure of the eigenvalues. For the errors between the estimated Type I error rate and the nominal level 0.05, it is clear that the method $Q_W$ gives the largest errors while the other three methods have much better performance of achieving the nominal level in all six tables. With respect to the accuracy of power calculations, the Rao's statistic $Q_S$ of the GEE approach is the most sensitive one among the four approaches with respect to four different eigenvalue structures. This situation is more

Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

Table 1
The Errors Between Actual and Nominal Values of Type I Error Rate and Power for Fixed MANOVA Model with Equal Group Size ($c = 3$, $r = 4$, $a = 3$, $\alpha = 0.05$)

| Sample size | Method | Error for α | Nominal power | Error for power Eigenvalues structure of $\bar{\boldsymbol{\delta}}$ | | | |
|---|---|---|---|---|---|---|---|
| | | | | Equal | Linear | Geometric | Extreme |
| $N = 20$ | $Q_W$ | 0.2772 | 0.8075 | 0.1367 | 0.1361 | 0.1359 | 0.1281 |
| | $Q_S$ | -0.0160 | 0.8075 | -0.1509 | -0.1743 | -0.1855 | -0.4079 |
| | $F_{T1}$ | 0.0148 | 0.7000 | -0.0336 | -0.0322 | -0.0322 | -0.0556 |
| | $F_{T2}$ | 0.0018 | 0.6008 | 0.0156 | 0.0150 | 0.0170 | -0.0076 |
| $N = 40$ | $Q_W$ | 0.1034 | 0.7446 | 0.1086 | 0.1040 | 0.1054 | 0.0920 |
| | $Q_S$ | -0.0098 | 0.7446 | -0.0712 | -0.0774 | -0.0806 | -0.1562 |
| | $F_{T1}$ | 0.0024 | 0.7000 | -0.0162 | -0.0160 | -0.0174 | -0.0208 |
| | $F_{T2}$ | -0.0026 | 0.6591 | 0.0013 | 0.0029 | 0.0029 | -0.0019 |
| $N = 60$ | $Q_W$ | 0.0586 | 0.7281 | 0.0807 | 0.0753 | 0.0745 | 0.0597 |
| | $Q_S$ | -0.0036 | 0.7281 | -0.0431 | -0.0475 | -0.0471 | -0.1027 |
| | $F_{T1}$ | 0.0048 | 0.7000 | -0.0100 | -0.0076 | -0.0076 | -0.0220 |
| | $F_{T2}$ | 0.0002 | 0.6743 | 0.0045 | 0.0051 | 0.0021 | -0.0097 |
| $N = 80$ | $Q_W$ | 0.0394 | 0.7205 | 0.0453 | 0.0463 | 0.0443 | 0.0451 |
| | $Q_S$ | -0.0072 | 0.7205 | -0.0421 | -0.0389 | -0.0423 | -0.0751 |
| | $F_{T1}$ | -0.0004 | 0.7000 | -0.0146 | -0.0122 | -0.0152 | -0.0204 |
| | $F_{T2}$ | -0.0024 | 0.6813 | -0.0045 | -0.0049 | -0.0063 | -0.0099 |
| $N = 100$ | $Q_W$ | 0.0250 | 0.7161 | 0.0515 | 0.0499 | 0.0499 | 0.0409 |
| | $Q_S$ | -0.0050 | 0.7161 | -0.0251 | -0.0281 | -0.0287 | -0.0541 |
| | $F_{T1}$ | -0.0004 | 0.7000 | -0.0048 | -0.0044 | -0.0052 | -0.0110 |
| | $F_{T2}$ | -0.0024 | 0.6853 | 0.0029 | 0.0021 | 0.0013 | -0.0061 |

prominent when the sample size is small in all models. On the contrary, the two likelihood-based approaches are much more robust with respect to the level of variation among eigenvalues. Furthermore, the performance of the two $F$-transform approaches appears to be excellent over the whole range of conditions that we have considered. Although the differences between the two approaches are small, it shows a clear pattern that the $F_{T2}$ approach



Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

Table 2
The Errors Between Actual and Nominal Values of Type I Error Rate and Power for Fixed MANOVA Model with Equal Group Size ($c = 2$, $r = 3$, $a = 2$, $\alpha = 0.05$)

| Sample size | Method | Error for $\alpha$ | Nominal power | Error for power Eigenvalues structure of $\bar{\boldsymbol{\delta}}$ | | |
|---|---|---|---|---|---|---|
| | | | | Equal | Linear/Geometric | Extreme |
| $N = 15$ | $Q_W$ | 0.1602 | 0.8106 | 0.0964 | 0.0912 | 0.0802 |
| | $Q_S$ | -0.0126 | 0.8106 | -0.1218 | -0.1384 | -0.2852 |
| | $F_{T1}$ | 0.0080 | 0.7000 | -0.0172 | -0.0170 | -0.0446 |
| | $F_{T2}$ | 0.0006 | 0.6293 | 0.0237 | 0.0215 | -0.0031 |
| $N = 30$ | $Q_W$ | 0.0708 | 0.7462 | 0.0744 | 0.0696 | 0.0588 |
| | $Q_S$ | 0.0000 | 0.7462 | -0.0546 | -0.0590 | -0.1072 |
| | $F_{T1}$ | 0.0076 | 0.7000 | -0.0078 | -0.0064 | -0.0228 |
| | $F_{T2}$ | 0.0034 | 0.6696 | 0.0088 | 0.0112 | -0.0024 |
| $N = 45$ | $Q_W$ | 0.0432 | 0.7291 | 0.0395 | 0.0441 | 0.0407 |
| | $Q_S$ | -0.0014 | 0.7291 | -0.0443 | -0.0457 | -0.0713 |
| | $F_{T1}$ | 0.0034 | 0.7000 | -0.0128 | -0.0140 | -0.0174 |
| | $F_{T2}$ | 0.0016 | 0.6807 | -0.0019 | -0.0023 | -0.0071 |
| $N = 60$ | $Q_W$ | 0.0312 | 0.7213 | 0.0409 | 0.0423 | 0.0313 |
| | $Q_S$ | -0.0010 | 0.7213 | -0.0285 | -0.0321 | -0.0495 |
| | $F_{T1}$ | 0.0034 | 0.7000 | -0.0036 | -0.0044 | -0.0128 |
| | $F_{T2}$ | 0.0008 | 0.6859 | 0.0041 | 0.0045 | -0.0051 |
| $N = 75$ | $Q_W$ | 0.0214 | 0.7167 | 0.0367 | 0.0331 | 0.0371 |
| | $Q_S$ | 0.0010 | 0.7167 | -0.0159 | -0.0193 | -0.0313 |
| | $F_{T1}$ | 0.0044 | 0.7000 | 0.0022 | -0.0018 | -0.0022 |
| | $F_{T2}$ | 0.0036 | 0.6889 | 0.0093 | 0.0055 | 0.0045 |

is more accurate with the estimated actual power than the other $F_{T1}$ approach. The performance for the method $Q_W$ is very good throughout the six tables regardless of the eigenvalues structure and sample size. However, this is counteracted by the inflated estimates of Type I error rate $\alpha$ mentioned earlier. Therefore, the practical use of $Q_W$ is questionable at least for the multivariate general linear models.



Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

Table 3
The Errors Between Actual and Nominal Values of Type I Error Rate and Power for Random MANOVA Model with Uniformly Distributed Group Size ($c = 3$, $r = 4$, $a = 3$, $\alpha = 0.05$)

| Sample size | Method | Error for $\alpha$ | Nominal power | Error for power<br>Eigenvalues structure of $\bar{\boldsymbol{\delta}}$<br>Equal | <br><br>Linear | <br><br>Geometric | <br><br>Extreme |
|---|---|---|---|---|---|---|---|
| $N = 20$ | $Q_W$ | 0.2804 | 0.8075 | 0.1239 | 0.1229 | 0.1233 | 0.1101 |
| | $Q_S$ | -0.0192 | 0.8075 | -0.1793 | -0.2049 | -0.2169 | -0.4195 |
| | $F_{T1}$ | 0.0102 | 0.7000 | -0.0490 | -0.0574 | -0.0580 | -0.0830 |
| | $F_{T2}$ | -0.0030 | 0.6008 | -0.0032 | -0.0072 | -0.0102 | -0.0376 |
| $N = 40$ | $Q_W$ | 0.1072 | 0.7446 | 0.0954 | 0.0868 | 0.0830 | 0.0754 |
| | $Q_S$ | -0.0094 | 0.7446 | -0.0864 | -0.0974 | -0.0998 | -0.1784 |
| | $F_{T1}$ | 0.0048 | 0.7000 | -0.0316 | -0.0368 | -0.0346 | -0.0482 |
| | $F_{T2}$ | -0.0022 | 0.6591 | -0.0145 | -0.0167 | -0.0157 | -0.0271 |
| $N = 60$ | $Q_W$ | 0.0644 | 0.7281 | 0.0663 | 0.0677 | 0.0667 | 0.0519 |
| | $Q_S$ | -0.0016 | 0.7281 | -0.0547 | -0.0561 | -0.0643 | -0.1069 |
| | $F_{T1}$ | 0.0082 | 0.7000 | -0.0200 | -0.0160 | -0.0194 | -0.0340 |
| | $F_{T2}$ | 0.0032 | 0.6743 | -0.0065 | -0.0023 | -0.0065 | -0.0211 |
| $N = 80$ | $Q_W$ | 0.0506 | 0.7205 | 0.0525 | 0.0487 | 0.0507 | 0.0427 |
| | $Q_S$ | 0.0018 | 0.7205 | -0.0433 | -0.0405 | -0.0439 | -0.0709 |
| | $F_{T1}$ | 0.0098 | 0.7000 | -0.0198 | -0.0136 | -0.0114 | -0.0172 |
| | $F_{T2}$ | 0.0062 | 0.6813 | -0.0099 | -0.0041 | -0.0021 | -0.0065 |
| $N = 100$ | $Q_W$ | 0.0290 | 0.7161 | 0.0443 | 0.0475 | 0.0469 | 0.0371 |
| | $Q_S$ | -0.0060 | 0.7161 | -0.0329 | -0.0293 | -0.0309 | -0.0589 |
| | $F_{T1}$ | -0.0010 | 0.7000 | -0.0110 | -0.0074 | -0.0062 | -0.0116 |
| | $F_{T2}$ | -0.0032 | 0.6853 | -0.0071 | -0.0013 | -0.0005 | -0.0053 |

As mentioned on the previous page, the sample size formulas in the first two models are identical. Hence, the discrepancy between the results in Tables 1–2 and Tables 3–4 are the direct consequence of using fixed or random independent variables. In general, the absolute errors in Tables 3–4 are slightly larger than those in Tables 1–2. Such phenomena shall continue

Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

Table 4
The Errors Between Actual and Nominal Values of Type I Error Rate and Power for Random MANOVA Model with Uniformly Distributed Group Size ($c = 2$, $r = 3$, $a = 2$, $\alpha = 0.05$)

| Sample size | Method | Error for $\alpha$ | Nominal power | Error for power Eigenvalues structure of $\bar{\boldsymbol{\delta}}$ | | |
|---|---|---|---|---|---|---|
| | | | | Equal | Linear/Geometric | Extreme |
| $N = 15$ | $Q_W$ | 0.1614 | 0.8106 | 0.0788 | 0.0756 | 0.0536 |
| | $Q_S$ | -0.0166 | 0.8106 | -0.1634 | -0.1796 | -0.3102 |
| | $F_{T1}$ | 0.0024 | 0.7000 | -0.0476 | -0.0532 | -0.0754 |
| | $F_{T2}$ | -0.0054 | 0.6293 | -0.0019 | -0.0075 | -0.0315 |
| $N = 30$ | $Q_W$ | 0.0624 | 0.7462 | 0.0508 | 0.0522 | 0.0476 |
| | $Q_S$ | -0.0022 | 0.7462 | -0.0782 | -0.0816 | -0.1248 |
| | $F_{T1}$ | 0.0042 | 0.7000 | -0.0286 | -0.0314 | -0.0368 |
| | $F_{T2}$ | 0.0004 | 0.6696 | -0.0112 | -0.0148 | -0.0208 |
| $N = 45$ | $Q_W$ | 0.0388 | 0.7291 | 0.0299 | 0.0283 | 0.0249 |
| | $Q_S$ | -0.0008 | 0.7291 | -0.0581 | -0.0573 | -0.0857 |
| | $F_{T1}$ | 0.0058 | 0.7000 | -0.0278 | -0.0226 | -0.0346 |
| | $F_{T2}$ | 0.0034 | 0.6807 | -0.0157 | -0.0101 | -0.0211 |
| $N = 60$ | $Q_W$ | 0.0342 | 0.7213 | 0.0315 | 0.0301 | 0.0227 |
| | $Q_S$ | 0.0058 | 0.7213 | -0.0373 | -0.0369 | -0.0583 |
| | $F_{T1}$ | 0.0100 | 0.7000 | -0.0110 | -0.0106 | -0.0198 |
| | $F_{T2}$ | 0.0074 | 0.6859 | -0.0021 | -0.0019 | -0.0139 |
| $N = 75$ | $Q_W$ | 0.0212 | 0.7167 | 0.0267 | 0.0249 | 0.0261 |
| | $Q_S$ | -0.0048 | 0.7167 | -0.0239 | -0.0279 | -0.0427 |
| | $F_{T1}$ | 0.0000 | 0.7000 | -0.0050 | -0.0100 | -0.0086 |
| | $F_{T2}$ | -0.0016 | 0.6889 | -0.0003 | -0.0021 | -0.0037 |

to exist between random and fixed models in other settings. For the results associated with the development model in Tables 5–6, comparatively much larger sample sizes are required to achieve the same accuracy than the previous two cases in Tables 1–4. Due to the polynomial function of the standard normal covariates, the magnitude of the errors between the

Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

Table 5
The Errors Between Actual and Nominal Values of Type I Error Rate and Power for Development Model with Standard Normal Independent Variable ($c = 3$, $r = 4$, $a = 3$, $\alpha = 0.05$)

| Sample size | Method | Error for $\alpha$ | Nominal power | Error for power Eigenvalues structure of $\bar{\boldsymbol{\delta}}$ Equal | Linear | Geometric | Extreme |
|---|---|---|---|---|---|---|---|
| $N = 100$ | $Q_W$ | 0.0294 | 0.7161 | -0.0019 | 0.0057 | 0.0079 | 0.0181 |
| | $Q_S$ | -0.0042 | 0.7161 | -0.0823 | -0.0751 | -0.0727 | -0.0885 |
| | $F_{T1}$ | -0.0002 | 0.7000 | -0.0554 | -0.0514 | -0.0454 | -0.0432 |
| | $F_{T2}$ | -0.0024 | 0.6853 | -0.0481 | -0.0425 | -0.0393 | -0.0375 |
| $N = 150$ | $Q_W$ | 0.0206 | 0.7105 | 0.0069 | 0.0141 | 0.0135 | 0.0085 |
| | $Q_S$ | 0.0018 | 0.7105 | -0.0415 | -0.0381 | -0.0391 | -0.0569 |
| | $F_{T1}$ | 0.0048 | 0.7000 | -0.0280 | -0.0212 | -0.0212 | -0.0304 |
| | $F_{T2}$ | 0.0032 | 0.6904 | -0.0230 | -0.0166 | -0.0170 | -0.0258 |
| $N = 200$ | $Q_W$ | 0.0194 | 0.7078 | -0.0044 | -0.0024 | 0.0034 | 0.0144 |
| | $Q_S$ | -0.0004 | 0.7078 | -0.0432 | -0.0390 | -0.0386 | -0.0334 |
| | $F_{T1}$ | 0.0026 | 0.7000 | -0.0336 | -0.0284 | -0.0264 | -0.0090 |
| | $F_{T2}$ | 0.0012 | 0.6929 | -0.0301 | -0.0245 | -0.0225 | -0.0047 |
| $N = 250$ | $Q_W$ | 0.0162 | 0.7062 | 0.0008 | 0.0080 | 0.0098 | 0.0024 |
| | $Q_S$ | 0.0028 | 0.7062 | -0.0278 | -0.0226 | -0.0230 | -0.0392 |
| | $F_{T1}$ | 0.0060 | 0.7000 | -0.0192 | -0.0130 | -0.0144 | -0.0196 |
| | $F_{T2}$ | 0.0056 | 0.6943 | -0.0159 | -0.0103 | -0.0127 | -0.0177 |
| $N = 300$ | $Q_W$ | 0.0094 | 0.7052 | -0.0032 | 0.0034 | 0.0056 | 0.0104 |
| | $Q_S$ | 0.0006 | 0.7052 | -0.0298 | -0.0242 | -0.0226 | -0.0226 |
| | $F_{T1}$ | 0.0024 | 0.7000 | -0.0234 | -0.0154 | -0.0146 | -0.0084 |
| | $F_{T2}$ | 0.0010 | 0.6953 | -0.0203 | -0.0133 | -0.0125 | -0.0067 |

estimated power and the nominal power tends to decrease when the eigenvalue structure changes from "equal" to "extreme". The results in the two MANOVA models show the opposite pattern that the "equal" eigenvalue structure produces the smallest errors among the four different structures, while the "extreme" eigenvalue structure gives the largest errors. This should be the case that is commonly encountered in the standard MANOVA model.



Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

Table 6
The Errors Between Actual and Nominal Values of Type I Error Rate and Power for Development Model with Standard Normal Independent Variable ($c = 3$, $r = 4$, $a = 2$, $\alpha = 0.05$)

| Sample size | Method | Error for $\alpha$ | Nominal power | Error for power<br>Eigenvalues structure of $\bar{\boldsymbol{\delta}}$<br>Equal | <br><br>Linear/Geometric | <br><br>Extreme |
|---|---|---|---|---|---|---|
| $N = 40$ | $Q_W$ | 0.0726 | 0.7464 | 0.0174 | 0.0202 | 0.0274 |
| | $Q_S$ | -0.0020 | 0.7464 | -0.1296 | -0.1288 | -0.1614 |
| | $F_{T1}$ | 0.0048 | 0.7000 | -0.0704 | -0.0626 | -0.0672 |
| | $F_{T2}$ | 0.0020 | 0.6764 | -0.0590 | -0.0498 | -0.0560 |
| $N = 80$ | $Q_W$ | 0.0310 | 0.7217 | -0.0003 | 0.0009 | -0.0001 |
| | $Q_S$ | -0.0016 | 0.7217 | -0.0761 | -0.0745 | -0.0869 |
| | $F_{T1}$ | 0.0028 | 0.7000 | -0.0480 | -0.0424 | -0.0458 |
| | $F_{T2}$ | 0.0018 | 0.6891 | -0.0417 | -0.0377 | -0.0385 |
| $N = 120$ | $Q_W$ | 0.0230 | 0.7142 | -0.0098 | -0.0094 | -0.0080 |
| | $Q_S$ | 0.0020 | 0.7142 | -0.0542 | -0.0560 | -0.0662 |
| | $F_{T1}$ | 0.0046 | 0.7000 | -0.0362 | -0.0370 | -0.0380 |
| | $F_{T2}$ | 0.0032 | 0.6929 | -0.0319 | -0.0333 | -0.0357 |
| $N = 160$ | $Q_W$ | 0.0138 | 0.7105 | -0.0003 | 0.0015 | -0.0051 |
| | $Q_S$ | 0.0006 | 0.7105 | -0.0381 | -0.0357 | -0.0437 |
| | $F_{T1}$ | 0.0032 | 0.7000 | -0.0252 | -0.0212 | -0.0246 |
| | $F_{T2}$ | 0.0026 | 0.6947 | -0.0229 | -0.0173 | -0.0217 |
| $N = 200$ | $Q_W$ | 0.0108 | 0.7083 | 0.0019 | -0.0001 | -0.0039 |
| | $Q_S$ | 0.0000 | 0.7083 | -0.0293 | -0.0323 | -0.0409 |
| | $F_{T1}$ | 0.0014 | 0.7000 | -0.0184 | -0.0200 | -0.0246 |
| | $F_{T2}$ | 0.0008 | 0.6958 | -0.0164 | -0.0186 | -0.0226 |

## *Conclusions*

The primary aim of the present article is to provide guidance in the choice of approach for sample size and power calculations within the framework of multivariate general linear models with Gaussian responses. The proposed approaches are the direct extension of the work by Liu and Liang (1997) Rochon (1998) and O'Brien and Shieh (1992) to accommodate both fixed and random independent variables. Their methods have been restricted to the simplifying assumption that all levels of the independent variables in the model are completely fixed nonrandom. According to the explicit sample size formulas obtained in this article, the general setup of the proposed approaches provides feasible solutions for studies involving both fixed and random independent variables.

The GEE approach described in Liu and Liang (1997) and Rochon (1998) represents a unified tool for sample size and power calculations for both normal and non-normal responses of repeated measures and longitudinal studies. Unfortunately, it is found in this article that their approaches have potential problem associated with the control of Type I error rate or the sensitivity to the unbalanced design for the most common research paradigm of multivariate general linear models with normal responses. More importantly, it is outperformed by the likelihood-based approaches of two *F* transforms of Hotelling-Lawley trace statistic. Such phenomenon continues to exist in the proposed modification of their approaches over all the conditions that we have considered. The simulation results suggest the proposed $F_{T2}$ approach developed from the approximation proposed by McKeon (1974) performs extremely well and shall be of practical use.

## References

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis* (2nd Ed.), New York: John Wiley.

Dickey, J. M. (1967). Matricvariate generalizations of the multivariate *t* distribution and the inverted multivariate *t* distribution. *Annals of the Mathematical Statistics*, *38*, 511-518.

Duncan, T. E., Duncan, S. C., Hops, H., & Stoolmiller, M. (1995). An analysis of the relationship between parents and adolescent marijuana use via generalized estimating equation methodology. *Multivariate Behavioral Research*, *30*, 317-339.

Dunlop, D. D. (1994). Regression for longitudinal data: A bridge from least squares regression. *The American Statistician*, *48*, 299-303.

Horton, N. J. & Lipsitz, S. R. (1999). Review of software to fit generalized estimating equation regression models. *The American Statistician*, *53*, 160-169.

Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2001). The analysis of repeated measures designs: a review. *British Journal of Mathematical and Statistical Psychology*, *54*, 1-20.

Liang, K. Y. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13-22.

Liu, G. & Liang, K. Y. (1997). Sample size calculations for studies with correlated observations. *Biometrics*, *53*, 937-947.

McKeon, J. J. (1974). *F* approximations to the distribution of Hotelling's $T_0^2$. *Biometrika*, *61*, 381-383.

Morrison, D. F. (1990). *Multivariate statistical methods*. McGraw Hill: New York.

Muller, K. E., LaVange, L. M., Ramey, S. L., & Ramey, C. T. (1992). Power calculations for general linear multivariate models including repeated measures applications. *Journal of the American Statistical Association*, *87*, 1209-1226.



Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

Muller, K. E. & Peterson, B. L. (1984). Practical methods for computing power in testing the multivariate general linear hypothesis. *Computational Statistics and Data Analysis*, *2*, 143-158.

O'Brien, R. G. & Shieh, G. (1992, August). Pragmatic, unifying algorithm gives power probabilities for common *F* tests of the multivariate general linear hypothesis. Paper presented at the Annual Joint Statistical Meetings of the American Statistical Association, Boston, Massachusetts.

Pillai, K. C. S. & Samson, P., Jr. (1959). On Hotelling's generalization of $T^2$. *Biometrika*, *46*, 160-168.

Ramey, C. T., Bryant, D. M., Wasik, B. H., Sparling, J. J., Fendt, K. H., & LaVange, L. M. (1992). The infant health and development program for low birth weight, premature infants: program elements, family participation. *Pediatrics*, *89*, 454-465.

Rochon, J. (1998). Application of GEE procedures for sample size calculations in repeated measures experiments. *Statistics in Medicine*, *17*, 1643-1658.

Sampson, A. R. (1974). A tale of two regressions. *Journal of the American Statistician Association*, *69*, 682-689.

SAS Institute. (1999). *SAS/IML software: Usage and reference, Version 8*. Carey, NC: Author.

Seber, G. A. F. (1984). *Multivariate observations*. New York: John Wiley.

Shih, W. J. (1997). Sample size and power calculations for periodontal and other studies with clustered samples using the method of generalized estimating equations. *Biomedical Journal*, *39*, 899-908.

Timm, N. E. (2002). *Applied multivariate analysis*. New York: Springer-Verlag.

Ziegler, A. & Gromping, U. (1998). The generalised estimating equations: a comparison of procedures available in commercial statistical software packages. *Biometrical Journal*, *40*, 245-260.

*Accepted February, 2003.*

## *Appendix*

### *Brief Review of Generalized Estimating Equations (GEE)*

Consider the repeated measures design or longitudinal study, and let $y_{ij}^*$ be the response and $\mathbf{x}_{it}^*$ be the $K \times 1$ vector of independent variables or covariates for the $i^{\text{th}}$ subject and the $t^{\text{th}}$ condition or time point, where $i = 1, ..., N$ and $t = 1, ..., T_i$. Here for simplicity we consider the situation that all $T_i = T$; the case of varying $T_i$ can be handled in a similar way. Let $\mathbf{y}_i^*$ be the $T \times 1$ vector $(y_{i1}^*, ..., y_{iT}^*)^T$ and $\mathbf{X}_i^*$ be the $T \times K$ matrix $(\mathbf{x}_{i1}^*, ..., \mathbf{x}_{iT}^*)^T$. It is assumed throughout that $\mathbf{y}_i^*$ and $\mathbf{y}_j^*$ are independent for any $i \neq j$. The GEE approach fits a "marginal model" in which a mean function and a covariance structure are specified, but a full likelihood is not required. The mean function specifies the relationship between the marginal mean $E(y_{it}^* \mid \mathbf{x}_{it}^*) = \mu_{it}$ and the vector $\mathbf{x}_{it}^*$ through a generalized linear model $g(\mu_{it}) = \mathbf{x}_{it}^{*T}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a $K \times 1$ vector of parameters, and $g$ is a known link function. Common choices for the link function might be



Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

identity link $g(\mu) = \mu$ for Gaussian data, log link $g(\mu) = \log(\mu)$ for count data, and logit link $g(\mu) = \log[\mu/(1-\mu)]$ for binary data. The variance function $V(y_{it}^* \mid \mathbf{x}_{it}^*)$ $= \nu_{it}$ is expressed as $\nu_{it} = \nu(\mu_{it})/\phi$, where $\nu$ is a known function and $\phi$ is a scale parameter which may be estimated. Under the GEE model, the key step is to write the covariance matrix of $\mathbf{y}_i^*$ given $\mathbf{X}_i^*$ as $\mathbf{V}_i = \mathbf{A}_i^{1/2}\mathbf{R}_i(\boldsymbol{\alpha})\mathbf{A}_i^{1/2}$, where $\mathbf{A}_i$ is a $T \times T$ diagonal matrix with $\nu_{it}$ as the $t^{\text{th}}$ diagonal element, and $\mathbf{R}_i(\boldsymbol{\alpha})$ is the working correlation matrix which is fully specified by the $q \times 1$ vector of unknown parameters $\boldsymbol{\alpha}$. The vector $\boldsymbol{\alpha}$ typically contains parameters that characterize the correlation as a function of time lag or distance separation.

The major feature of GEE is the relaxation of the full likelihood function. For the classical univariate and multivariate linear models with normal responses, the mean and variance functions (the first two moments) fully determine the likelihood function, but this assumption is violated for many types of data such as binary or count outcomes. To specify the entire likelihood, additional assumptions about higher order moments are also necessary. Even if additional assumptions are made, the likelihood is often intractable and involves many nuisance parameters in addition to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ that must be estimated. GEE alleviate these restrictions.

A link can be built from familiar linear regression and generalized linear models to GEE methodology through the form of the estimating equations. Estimating equations represent a set of equations the solution of which give parameter estimates. For linear regression, it is well known that the least squares estimates are obtained from solving the normal equations. In similar fashion, the parameter estimates of generalized linear model are solutions to the estimating equations obtained by maximizing the likelihood function of the exponential family. The estimating equations of generalized linear models were extended by Liang and Zeger (1986) to account for correlated measurements from longitudinal data. Specifically, the GEE approach estimates $\boldsymbol{\beta}$ by solving the following generalized estimating equations:

$$\sum_{i=1}^{N} \mathbf{D}_i^T \mathbf{V}_i^{-1}\left(\mathbf{y}_i^* - \boldsymbol{\mu}_i\right) = \mathbf{0},$$

where $\mathbf{D}_i = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}^T$ is the $T \times K$ gradient matrix and $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{iT})^T$. The extra term "generalized" distinguishes these as estimating equations in generalized linear models that accommodate the correlation structure $\mathbf{R}_i(\boldsymbol{\alpha})$. Liang and Zeger (1986) showed that the solution to the generalized estimating equations gives a consistent estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ that is asymptotically multivariate normal with the covariance matrix given by

Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}) = \left( \sum_{i=1}^{N} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \right)^{-1} \left[ \sum_{i=1}^{N} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathrm{cov}\left(\mathbf{y}_i^*\right) \mathbf{V}_i^{-1} \mathbf{D}_i \right] \left( \sum_{i=1}^{N} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \right)^{-1}.$$

Liang and Zeger (1986) proposed to estimate cov($\hat{\boldsymbol{\beta}}$) by replacing $\boldsymbol{\beta}$, $\phi$ and $\boldsymbol{\alpha}$ with their estimators and replacing cov($\mathbf{y}_i^*$) by $(\mathbf{y}_i^* - \hat{\boldsymbol{\mu}}_i)(\mathbf{y}_i^* - \hat{\boldsymbol{\mu}}_i)^T$, where $\hat{\boldsymbol{\mu}}_i = (\hat{\mu}_{i1}, ..., \hat{\mu}_{iT})$ and $\hat{\boldsymbol{\mu}}_{iT} = g^{-1}(\mathbf{x}_{it}^{*T}\hat{\boldsymbol{\beta}})$, $t = 1, ..., T$. Another useful feature of the GEE methodology is that $\boldsymbol{\beta}$ and cov($\hat{\boldsymbol{\beta}}$) are consistently estimated even if the correlation structure is misspecified. A good discussion of the connection between the GEE approach and the well-known least squares regression methodology was given by Dunlop (1994). Review of commercial software packages to fit GEE models can be found in Ziegler and Gromping (1998) and Horton and Lipsitz (1999).

Downloaded by [National Chiao Tung University ] at 20:42 27 April 2014