

Power Estimation in Multivariate Analysis of Variance

Sylvain Chartier

University of Ottawa

Institut Philippe Pinel de Montréal

Jean-François Allaire

Institut Philippe Pinel de Montréal

Power is often overlooked in designing multivariate studies for the simple reason that it is believed to be too complicated. In this paper, it is shown that power estimation in multivariate analysis of variance (MANOVA) can be approximated using a F distribution for the three popular statistics (Hotelling-Lawley trace, Pillai-Bartlett trace, Wilk's likelihood ratio). Consequently, the same procedure, as in any statistical test, can be used: computation of the critical F value, computation of the noncentral parameter (as a function of the effect size) and finally estimation of power using a noncentral F distribution. Various numerical examples are provided which help to understand and to apply the method. Problems related to post hoc power estimation are discussed.

Because it is a central tool in the design of any study, statistical power is introduced in all statistics textbooks. Power reflects the probability of rejecting a false null hypothesis (Figure 1). However, power is often overlooked when designing a multivariate study for the simple reason that it is believed to be too difficult to estimate. Because power is related to sample size, the latter must be adjusted accordingly. If the sample size is too small, the study will not be sensitive enough and if the sample size is too big, the study will waste valuable resources. Consequently, estimating power is essential and the choice of power depends on the research question, measurement, design and analysis plan (Cohen, 1988; Muller, LaVange, Landesman Ramey, & Ramey, 1992). In this paper, it is shown that power estimation in multivariate analysis of variance (MANOVA) is a direct generalization of its estimation in univariate analysis of variance (ANOVA) through the general linear model (GLM).

Power can be estimated before the study or after the study. The first situation will be referred to as prospective (or a priori) power estimation and the second as

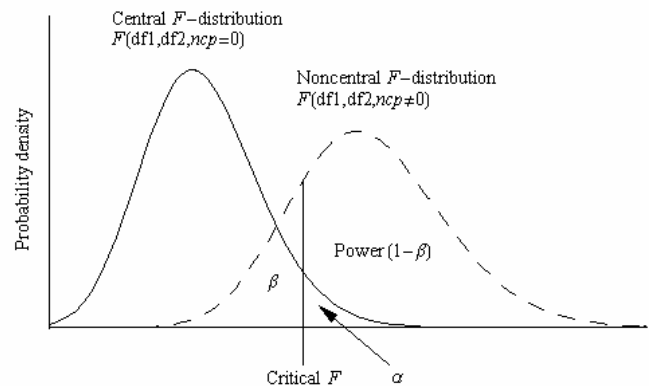


Figure 1. Power and errors illustrative concepts

retrospective (post hoc) power estimation. Power is important for design, not analysis (Lenth, 2001). Since studies can be statistically significant yet clinically insignificant; time should be spent on what is clinically significant in respect to the study hypothesis. Statistics are used to help in reaching that goal, not the other way around (Lenth, 2001). Post hoc power estimation may be useful for a pilot study where procedures are established and the obtained variance estimates are used for determining sample size or effect size for future experiments. The general procedure for power estimation is:

- 1- Compute the critical F value.

The authors are grateful to an anonymous reviewer for his help in preparing this article

Table 1 Univariate and multivariate representations of the GLM

#	DV	Form	IV	Form	Type of analysis
Univariate GLM	1	nominal	1	nominal	Phi coefficient / Chi-square
	1	continuous	1	nominal	<i>t</i> -test
	1	nominal	≥1	continuous and/or nominal	Logistic regression / Discriminant function
	1	continuous	1	continuous	Simple correlation / regression
	1	continuous	≥2	nominal	ANOVA (one-way, factorial, repeated measures)
	1	continuous	≥2	continuous and nominal	ANCOVA
	1	continuous	≥2	continuous	Multiple correlation / regression
Multivariate GLM	≥2	nominal	≥2	nominal	Correspondance
	≥2	nominal	≥2	continuous and/or nominal	Multivariate logistic regression / Discriminant functions
	≥2	continuous	≥2	nominal	MANOVA
	≥2	continuous	≥1	latent	Principal component / Factor
	≥2	continuous and/or nominal	≥1	latent	Multidimensional scaling
	≥2	continuous and/or latent	≥1	continuous and/or latent	Structural equation modeling
	≥2	continuous	≥2	continuous	Canonical correlation

- 2- Compute the noncentral parameter ncp (as a function of the effect size)
- 3- Compute the power using the noncentral F distribution

Whatever the design (regression, ANOVA, MANOVA, etc.), the same procedure is always applied. Using this procedure, the sample size required to reach a given power can be computed using the Mathematica function provided in Listing 2 (see the Numerical example III section).

Multivariate analysis of variance

When a research design contains two or more dependent variables there is two possible type of analysis: multiple univariate tests or one multivariate test. Performing multivariate analysis of variance (MAOVA) has many advantages (Baguley, 2004; Stevens, 1992). First, MANOVA does not have the problem of inflated overall type I error rate (α) that occurs when using multiple univariate tests. In addition, when multiple univariate tests are used, estimation of the type I error inflation is difficult since the multiple tests are not independents. Second, univariate tests ignore the correlations among the variables. On the other

hand, in multivariate tests, the correlation matrix (via the covariance) is directly incorporated into the analysis. Finally, since the groups are combined in the overall effect, multivariate tests are more powerful than multiple univariate tests; the same argument applies when choosing one ANOVA instead of multiple t tests.

In this paper, attention will be focus on MANOVA. MANOVA requires two or more categorical independent variables (predictors), just like in ANOVA. However, unlike ANOVA, it uses two or more dependent variables (criteria). MANOVA is thus employed to test the main and interaction effects on multiple dependent variables. Like any other linear statistics, Table 1 shows that MANOVA is part of the general linear model (GLM) and is in fact a special case of canonical correlation analysis (Thompson, 1984).

For concision, only a two-factor multivariate factorial analysis will be considered in this paper. The concepts presented here easily extended to situation involving more than 2 factors. In a factorial MANOVA, each score is given by a base mean for the h^{th} measurement (μ_{th}), some error (e_{ijhr}) and hopefully some effect due to one (α_{th}), and/or the other

Listing 1. Mathematica function for power estimation

```

power[df1_, df2_, fObs_, signif_ : 0.05] := Module[{tmp, fcrit, ncp, obsPow},
  fcrit = Quantile[FRatioDistribution[df1, df2], 1 - signif];
  ncp = fObs * df1;
  obsPow = 1 - CDF[NoncentralFRatioDistribution[df1, df2, ncp], fcrit];
  Return[{ncp, obsPow}];
];

```

factor (β_{ih}) and/or their interaction ($\alpha\beta_{ijh}$). This is described by the following

$$y_{ijhr} = \mu_h + \alpha_{ih} + \beta_{jh} + \alpha\beta_{ijh} + e_{ijhr} \quad (1)$$

$h = 1, 2, \dots, q; i = 1, 2, \dots, k; j = 1, 2, \dots, l; r = 1, 2, \dots, n.$

With an appropriate coding matrix, it is possible to represent this situation in general linear term.

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \quad (2)$$

where the matrix \mathbf{B} contains the coefficients used in Equation 1, such that,

$$\mathbf{B} = \begin{bmatrix} \mu_1, \alpha_{11}, \alpha_{21}, \dots, \alpha_{k-1,1}, \beta_{11}, \beta_{21}, \dots, \beta_{l-1,1}, (\alpha\beta)_{111}, (\alpha\beta)_{121}, \dots, (\alpha\beta)_{k-l+1,1} \\ \mu_2, \alpha_{21}, \alpha_{22}, \dots, \alpha_{k-1,2}, \beta_{21}, \beta_{22}, \dots, \beta_{l-1,2}, (\alpha\beta)_{112}, (\alpha\beta)_{122}, \dots, (\alpha\beta)_{k-l+1,2} \\ \vdots \\ \mu_l, \alpha_{l1}, \alpha_{l2}, \dots, \alpha_{k-1,l}, \beta_{l1}, \beta_{l2}, \dots, \beta_{l-1,l}, (\alpha\beta)_{1l1}, (\alpha\beta)_{1l2}, \dots, (\alpha\beta)_{k-l+1,l} \end{bmatrix}^T$$

\mathbf{X} represents the predictor matrix obtained from a coding matrix (see below), q is the number of dependant variables, k and l are the number of groups for treatment α and β respectively.

The coding matrix \mathbf{X} is not different from the one used in univariate analysis. This full matrix is constructed by building two coding matrices for each factor (α and β) and then multiplying each column of the first factor with each column of the second factor to give the interaction coding matrix ($\alpha\beta$). An example of effect coding is shown in Table 2. Although an effect coding matrix is used for full factorial analysis, any other type of coding (e.g. contrast, dummy) can be used as well as any kind of mixed coding (e.g. α : contrast; β : dummy). From the coding matrix in figure 2 the predictor matrix \mathbf{X} is constructed as follows.

$$\mathbf{X}_i = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (3)$$

where $i \in \{\text{full } (\alpha, \beta \text{ and } \alpha\beta); \alpha; \beta; \text{ or } \alpha\beta\}$, n is the total number of subject and p is the number of independent predictors. In the full factorial model p will be set to $(k \times l) - 1$. Finally, the criterion matrix \mathbf{Y} contains the observed data and is organized as follow:

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1q} \\ y_{21} & y_{22} & \dots & y_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \dots & y_{nq} \end{bmatrix} \quad (4)$$

where q the number of criterions (dependant variables). These two matrices can be concatenated into a single matrix

$$\mathbf{M}_i = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} & y_{11} & y_{12} & \dots & y_{1q} \\ x_{21} & x_{22} & \dots & x_{2p} & y_{21} & y_{22} & \dots & y_{2q} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} & y_{n1} & y_{n2} & \dots & y_{nq} \end{bmatrix} \quad (5)$$

where, x_{ij} , represents the i^{th} subject of j^{th} predictor, y_{ih} represents the i^{th} subject of the h^{th} criterion, n the number of subjects per groups, p the number of predictor and q the number of criterion. In the case of the full model (Table 2) the sum of square and cross product (SSCP) matrix will be needed. The SSCP matrix is obtained by

$$\text{SSCP} = \mathbf{M}_{Full}^T \mathbf{M}_{Full} - (\mathbf{1}^T \mathbf{M}_{Full})^T (\mathbf{1}^T \mathbf{M}_{Full}) / n \quad (6)$$

where $\mathbf{1}$ is defined as the vector in which all the elements are 1, and T denotes the matrix transposition operation. As a further optional processing, when the SSCP is divided by the corresponding degrees of freedom ($n - 1$) then the variance/covariance matrix is obtained. Therefore, the SSCP is a convenient way to represent all information about variability in one matrix.

By partitioning the SSCP matrix correctly, the canonical correlation matrix, \mathbf{R} , can be obtained. The SSCP must be divided in four sections named \mathbf{S}_{pp} , \mathbf{S}_{pc} , \mathbf{S}_{cp} , \mathbf{S}_{cc} which represents the sum of squares of the predictors alone, the sum of cross-product between the predictors and the criterion, the sum of cross-product between the criterion and the predictors (note that $\mathbf{S}_{cp} = \mathbf{S}_{pc}^T$) and the sum of squares of

Table 2 Effect coding matrix for two-level factorial ANOVA

	α		β		$\alpha\beta$							
	x_1	x_2	x_{m-1}	x_m	x_{m+1}	x_{m+n-2}	x_{m+n-1}	x_{m+n}	x_{m+n-1}			
$\alpha\beta$	1	0	...	0	1	0	...	0	1	0	...	0
$\alpha\beta$	0	1	...	0	1	0	...	0	0	0	...	0
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\ddots	\vdots
$\alpha\beta$	-1	-1	...	-1	1	0	...	0	-1	0	...	0
$\alpha\beta$	1	0	...	0	0	1	...	0	0	0	...	0
$\alpha\beta$	0	1	...	0	0	1	...	0	0	0	...	0
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\ddots	\vdots
$\alpha\beta_n$	-1	-1	...	-1	-1	-1	...	-1	1	1	...	1

the criterion alone respectively This is summarized using:

$$\mathbf{SSCP} = \begin{bmatrix} \mathbf{S}_{pp} & \mathbf{S}_{pc} \\ \mathbf{S}_{cp} & \mathbf{S}_{cc} \end{bmatrix} \quad (7)$$

The canonical correlation matrix can be obtained by the following matrix multiplication

$$\mathbf{R} = \mathbf{S}_{pc}^T \mathbf{S}_{pp}^{-1} \mathbf{S}_{pc} \mathbf{S}_{cc}^{-1} \quad (8)$$

From \mathbf{R} , the error matrix (\mathbf{E}) can be computed

$$\mathbf{E} = (\mathbf{I} - \mathbf{R}) \mathbf{S}_{cc} \quad (9)$$

The degree of freedom associated with the error is given by

$$df_{err} = n - k \times l \quad (10)$$

The total variation is the sum of the various hypothesis variation add to the error variation, i.e. $\mathbf{T} = \mathbf{E} + \mathbf{H}_\alpha + \mathbf{H}_\beta + \mathbf{H}_{\alpha\beta}$.

Each matrix \mathbf{H} is obtained by

$$\mathbf{H}_i = \mathbf{Y}^T \mathbf{M}_i (\mathbf{M}_i^T \mathbf{M}_i)^{-1} \mathbf{M}_i^T \mathbf{Y} \quad (11)$$

where $i \in \{\alpha, \beta, \alpha\beta\}$; the full model is omitted when performing hypothesis testing. In univariate, the statistics used is based on the F -ratio distribution.

$$F = \frac{R^2 df_2}{(1 - R^2) df_1} \quad (12)$$

However, in MANOVA there is no unique statistic. Four statistics are commonly used: Hotelling-Lawley trace (HL), Pillai-Bartlett trace (PB), Wilk's likelihood ratio (W) and Roy's largest root (RLR). Those statistics can be obtained from Equations 9 and 11, as seen here.

The HL statistic is defined as

$$HL_i = \text{tr}(\mathbf{H}_i \mathbf{E}^{-1}) = \sum_{k=1}^s \lambda_k \quad (13)$$

where $s = \min(df_i, q)$, i represents the tested effect ($i \in \{\alpha, \beta, \alpha\beta\}$), df_i is the degree of freedom associated with the hypothesis under investigation (α, β or $\alpha\beta$) and λ_k is k^{th} eigenvalue, i.e. the k^{th} nonzero root of the following characteristic equation

$$|\mathbf{H} - \lambda \mathbf{E}| = 0 \quad (14)$$

where $|\mathbf{H} - \lambda \mathbf{E}|$ is the determinant of $\mathbf{H} - \lambda \mathbf{E}$.

The PB statistic is defined as

$$PB_i = \text{tr}(\mathbf{H}_i (\mathbf{E} + \mathbf{H}_i)^{-1}) = \sum_{k=1}^s \frac{\lambda_k}{1 + \lambda_k} \quad (15)$$

The W statistic is defined as

$$W_i = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}_i|} = \prod_{k=1}^s \frac{1}{1 + \lambda_k} \quad (16)$$

Finally, the RLR statistic is defined as

$$RLR_i = \frac{\text{Max}(\lambda_k)}{1 + \text{Max}(\lambda_k)} \quad (17)$$

All the statistics are equivalent when $s = 1$. In general there is no exact formula for finding the associated p value except on

Table 3 Summary table of the m and i symbolic variables.

	HL_i	Type of analysis (i)	$\frac{\alpha}{\beta}$ $\alpha\beta$
Type of statistics (m)	PB_i	Type of analysis (i)	$\frac{\alpha}{\beta}$ $\alpha\beta$
	W_i	Type of analysis (i)	$\frac{\alpha}{\beta}$ $\alpha\beta$

rare situations¹. Nevertheless, a convenient and sufficient approximation exists for all but RLR (Muller et al., 1992). Since RLR is the least robust (Olson, 1974), attention will be focused on the first three statistics: HL , PB and W . These three statistics' distributions are approximated using an F distribution which has the advantage of being simple to understand. The multivariate F -ratio is function of their multivariate association coefficient.

$$F(m_i) = \frac{\eta_{m_i}^2 df_2(m_i)}{(1 - \eta_{m_i}^2) df_1} \quad (18)$$

where df_1 represents the numerator degree of freedom ($df_1 = q^* df_i$), $df_2(m)$ the denominator degree of freedom for each statistic m ($m \in \{HL_i, PB_i \text{ and } W_i\}$) and η_m^2 is the multivariate measure of association for each statistic m (see Table 3 for the detailed values of the symbolic indices used in Equation 18).

The multivariate measure of association for HL is given by

$$\eta_{HL_i}^2 = \frac{HL_i}{HL_i + s} \quad (19)$$

and its denominator df by

$$df_2(HL_i) = s(df_{err} - q - 1) + 2 \quad (20)$$

The multivariate measure of association for PB is given by

$$\eta_{PB_i}^2 = \frac{PB_i}{s} \quad (21)$$

and its denominator df by

$$df_2(PB_i) = s(df_{err} - q + s) \quad (22)$$

The multivariate measure of association for W is given by

$$\eta_{W_i}^2 = 1 - W_i^{\frac{1}{s}} \quad (23)$$

and its denominator df by

¹ For example, Tatsuoka (1988) gives the exact formula for the Wilk's likelihood ratio.

Listing 2. Mathematica function for sample size estimation

```

sampleSizePillBart[etaP, k, l, q, alpha, power] :=
  Module[{n, res, dlHyp, dlerr, s, df1, df2, fObs, fcrit, ncp},
    n = 2;
    done = False;
    While[!done,
      If[l == 1, dlHyp = (k - 1), dlHyp = (k - 1) (l - 1)];
      dlerr = k * l * (n - 1);
      s = Min[q, dlHyp];
      df1 = dlHyp * q;
      df2 = s * (dlerr - q + s);
      fObs = (etaP * df2) / ((1 - etaP) df1);
      fcrit = Quantile[FRatioDistribution[df1, df2], 1 - alpha];
      ncp = fObs * df1;
      res = 1 - CDF[NoncentralFRatioDistribution[df1, df2, ncp], fcrit];
      If[res < power, n = n + 1, done = True];
    ]];
Return[{n, res}]];
1;

```

$$df_2(W_i) = og + 1 - \frac{df_1}{2} \quad (24) \quad Power(m_i) \approx 1 - CDF\left(NoncentralFDistribution(df_1, df_2, ncp(m_i)), F_{crit}\right) \quad (27)$$

where

$$o = df_{err} - (q + 1 - df_i) / 2 \quad \text{and}$$

$$g = \left((q^2 df_i^2 - 4) / (q^2 + df_i^2 - 5) \right)^{1/2}.$$

Equation 23 (like Equations 19 and 21) are related with Cohen effect size f^2 (Cohen, 1988) using the following equivalence

$$f_i^2 = \frac{\eta_{W_i}^2}{1 - \eta_{W_i}^2} \quad (25)$$

The only missing information to obtain the power is the value of the noncentral parameter ncp . It is obtained by the following

$$ncp(m_i) = df_1 F(m_i) = \frac{\eta_{m_i}^2 df_2(m_i)}{(1 - \eta_{m_i}^2)} = f_i^2 df_2(m_i) \quad (26)$$

Thus power can be estimated if an F -value (a priori or observed) or a multivariate association coefficient $\eta_{m_i}^2$ or a Cohen's f_i^2 is given. Now that all the relevant information has been obtained following the general procedure previously described it is possible to find the estimated power

Where F_{crit} is given by the inverse of a central F -distribution

$$F_{crit} = CDF^{-1}(1 - \alpha, df_1, df_2) \quad (28)$$

The Listing 1 shows a *Mathematica* function that evaluates the power given the numerator and denominator df , the multivariate association coefficient ($\eta_{m_i}^2$) and the type-I error (α); the default value is set to $\alpha = 0.05$.

Numerical example I: Post hoc power

In this section, power estimations are computed for the weight losses data given by Table 5.5 of Morrison (1976) reproduced in Table 4. The two factors are the sex of the rats

Table 4 Weight loss and time to complete a maze as a function of the sex of the rat and the type of drug ingested

	Drug		
	A	B	C
Male	5, 6	7, 6	21, 15
	5, 4	7, 7	14, 11
	9, 9	9, 12	17, 12
	7, 6	6, 8	12, 10
Female	7, 10	10, 13	16, 12
	6, 6	8, 7	14, 9
	9, 7	7, 6	14, 8
	8, 10	6, 9	10, 5

and the type of drug administered (among three) resulting in 2x3 design ($\alpha = 2, \beta = 3$). There were four rats of each sex assigned to each drug. Finally, weight losses in grams and the time to run the maze in second were observed; Therefore, there is two dependent variables ($q = 2$). Using Table 2 we construct the following coding matrix.

	α	β		$\alpha\beta$	
$\alpha_1\beta_1$	1	1	0	1	0
$\alpha_1\beta_2$	1	0	1	0	1
$\alpha_1\beta_3$	1	-1	-1	-1	-1
$\alpha_2\beta_1$	-1	1	0	-1	0
$\alpha_2\beta_2$	-1	0	1	0	-1
$\alpha_2\beta_3$	-1	-1	-1	1	1

From this matrix, it is then possible to obtain the matrix **M** by assigning the corresponding row, and the corresponding criterion value, to each subject. For example, the first row [1, 1, 0, 1, 0] correspond to the “male” subjects that has received the “drug A”. There are four subjects that belong to this category. Therefore, this row will be repeated four times and each one of them will be aggregated with the corresponding subject scores on the two criterions (weight loss and time).

	X					Y	
	α	β	$\alpha\beta$				
s1	1	1	0	1	0	5	6
s2	1	1	0	1	0	5	4
s3	1	1	0	1	0	9	9
s4	1	1	0	1	0	7	6
s5	1	0	1	0	1	7	6
s6	1	0	1	0	1	7	7
s7	1	0	1	0	1	9	12
s8	1	0	1	0	1	6	8
s9	1	-1	-1	-1	-1	21	15
s10	1	-1	-1	-1	-1	14	11
s11	1	-1	-1	-1	-1	17	12
M = s12	1	-1	-1	-1	-1	12	10
s13	-1	1	0	-1	0	7	10
s14	-1	1	0	-1	0	6	6
s15	-1	1	0	-1	0	9	7
s16	-1	1	0	-1	0	8	10
s17	-1	0	1	0	-1	10	13
s18	-1	0	1	0	-1	8	7
s19	-1	0	1	0	-1	7	6
s20	-1	0	1	0	-1	6	9
s21	-1	-1	-1	1	1	16	12
s22	-1	-1	-1	1	1	14	9
s23	-1	-1	-1	1	1	14	8
s24	-1	-1	-1	1	1	10	5

From **M** the **SSCP** matrix (Equations 6 and 7) can be obtained

SSCP=	24	0	0	0	0	4	4
	0	16	8	0	0	-62	-24
	0	8	16	0	0	-58	-14
	0	0	0	16	8	-14	-22
	0	0	0	8	16	-12	-16
	4	-62	-58	-14	-12	410.5	196
	4	-24	-14	-22	-16	196	183.3

Using Equation 8 the canonical correlation matrix **R** can be obtained

$$R = \begin{bmatrix} 0.93673 & -0.349632 \\ 0.2258 & 0.136781 \end{bmatrix}$$

And the error matrix **E** (Equation 9) is then equal to

$$E = \begin{bmatrix} 94.5 & 76.5 \\ 76.5 & 114 \end{bmatrix}$$

Finally, the error *df* (Equation 10) is $n-k \times l = 24-2 \times 3=18$. With this last information, each statistics and their corresponding power can be computed. The first step is to test for interaction. Using the interaction part ($\alpha\beta$) of the full coding matrix defined previously, the hypothesis matrix **H_{αβ}** is obtained using Equation 11.

$$H_{\alpha\beta} = \begin{bmatrix} 14.33 & 21.33 \\ 21.33 & 32.33 \end{bmatrix}$$

Hotelling-Lawley trace

Using Equation 13, $HL_{\alpha\beta} = 0.2897$ and its corresponding effect size (Equation 19) is 0.1265 which gives an *F*-value (Equation 18) of 1.1588 (*sig* = 0.3473) with *df₁* equal to 4 and *df₂* (Equation 20) equal to 32. The critical *F* (Equation 28) is 2.66844 and the *ncp* (Equation (26) is 4.635. From those outputs, the corresponding power ($\alpha = 0.05$) is obtained using Equation 27 or the *Mathematica* algorithm (Listing 1)

$$POWER(HL_{\alpha\beta}) = 0.32106$$

or using the *Mathematica* function:

$$power[4, 32, 0.1265, 0.05] = 0.32106.$$

Pillai-Bartlett trace

Using Equation 15, $PB_{\alpha\beta} = 0.22695$ and its corresponding effect size (Equation 21) is 0.11348 which gives an *F*-value (Equation 18) of 1.152 (*sig* = 0.34808) with *df₁* equal to 4 and *df₂* (Equation 22) equal to 36. The critical *F* (Equation 28) is 2.6335 and the *ncp* (Equation (26) is 4.608. From those outputs the corresponding power ($\alpha = 0.05$) is obtained using Equation 27 or the *Mathematica* algorithm (Listing 1)

$$POWER(PB_{\alpha\beta}) = 0.32407$$

or using the *Mathematica* function:

$$power[4, 36, 0.11348, 0.05] = 0.32407.$$

Wilk's likelihood ratio

Using Equation 16, $W_{\alpha\beta} = 0.7744$ and its corresponding effect size (Equation 23) is 0.12 which gives an F -value (Equation 18) of 1.15933 ($sig = 0.3459$) with df_1 equal to 4 and df_2 (Equation 24) equal to 34. The critical F (Equation 28) is 2.6499 and the ncp (Equation (26) is 4.6373. From those outputs the corresponding power ($\alpha = 0.05$) is obtained using Equation 27 or the *Mathematica* function (Listing 1)

$$\text{POWER}(W_{\alpha\beta}) = 0.32375$$

or using the *Mathematica* function:

$$\text{power}[4, 34, 0.12, 0.05] = 0.32375.$$

In a similar fashion, these steps are repeated for the main effects α (sex) and β (drug). Table 5 summarized the results.

Numerical example II: A priori power

In this example, it is assumed that there are 3 groups where in each group there are 20 participants ($n=60$). In the present example, there is only one factor ($l = 1$). Each participant is measured with two variables ($q = 2$). Now suppose that the expected effect size is $\eta^2 = 0.15$, what will be the power? The complete details will be given for the Wilk's likelihood ratio. The required information is

$$\begin{aligned} n &= 60 & \eta_W^2 &= 0.15 \\ k &= 3 & \alpha &= 0.05 \\ q &= 2 & df_\alpha &= k - 1 = 2 \\ l &= 1 \end{aligned}$$

Using Equation 10, the error degree of freedom is $df_{err} = n - k \times l = 60 - 3 \times 1 = 57$ and $df_1 = df_\alpha \times q = 2 \times 2 = 4$. From Equation 24, df_2 is obtained

$$\begin{aligned} g &= \sqrt{\frac{q^2 df_\alpha - 4}{q^2 + df_\alpha - 5}} = \sqrt{\frac{2^2 \cdot 2 - 4}{2^2 + 2 - 5}} = 2 \\ o &= df_{err} - \frac{q + 1 - df_\alpha}{2} = 57 - \frac{2 + 1 - 2}{2} = 56.5 \\ df_2 &= o \times g + 1 - \frac{df_1}{2} = 56.5 \times 2 + 1 - \frac{4}{2} = 112 \end{aligned}$$

Then the critical F (Equation 28) is computed

$$F_{crit} = CDF^{-1}(1 - \alpha, df_1, df_2) = CDF^{-1}(1 - 0.05, 4, 112) = 2.453$$

Finally, the noncentral parameter is computed according to Equation (26).

$$ncp(W) = \frac{\eta_W^2 df_2(W)}{(1 - \eta_W^2)} = \frac{0.15 \times 112}{(1 - 0.15)} = 19.76$$

The power can then be estimated using Equation 27.

$$\begin{aligned} \text{Power}(W) &\approx 1 - CDF(\text{NoncentralFdistribution}(df_1, df_2, ncp(W)), F_{crit}) \\ \text{Power}(W) &\approx 1 - CDF(\text{NoncentralFdistribution}(4, 112, 19.76), 2.453) \approx 0.954 \end{aligned}$$

or using the *Mathematica* function (Listing 1)

Table 6 Power estimation as a function of sample size when the estimated effect size η^2 is 0.1, the number of group for the first factor is 3 and the number of group for the second factor is 1 (i.e. there is no second factor) and there are three variables measured

n	power
2	0.076
3	0.124
4	0.185
5	0.254
6	0.329
7	0.406
8	0.481
9	0.553
10	0.620
11	0.681
12	0.735
13	0.782
14	0.823

$$\text{power}[4, 112, 19.76/4, 0.05] = 0.954$$

A similar procedure can be repeated for the other two statistics (HL and PB). After the various computations, those results are nearly identical.

$$\text{Power}(HL) \approx 0.95$$

$$\text{Power}(PB) \approx 0.958$$

Numerical example III: Sample size estimation

For sample size estimation given a power probability, tables like the one found in statistical book could be used. However, *Mathematica* can be used to find the appropriate n . The easiest method is to estimated power starting with the smallest possible n ($=2$) and then, if the resulting power is less than the desired value (e.g. $1 - \beta = 0.8$), the n value is increase by 1 ($n = n + 1$) and the new power is computed. This is repeated until the power is equal or higher of the desired value (ex. $1 - \beta = 0.8$). The Listing 2 illustrated this algorithm using *Mathematica* for the PB statistic. This algorithm can be extended to the other two methods (W and HL).

For example, assumed that there are 4 groups where each participant is measured with three variables ($q = 3$). Now suppose that the expected effect size is $\eta^2 = 0.1$ and the type I error rate is set to 0.05, then what is the minimum sample size if a power of 0.8 is desired? Table 6 indicates the

various powers as a function of n . From the table, the required number of participant per group is about 14. Thus, since there are 4 groups, the total number of participant needed is $4 \times 14 = 56$. The algorithm given in Listing 2 will output only the last row; the other rows were added to show the power

Table 1: MANOVA Summary

Interaction								
Statistics	Value	η^2	df_1	df_2	F	Sig	ncp	power
HL	0.2897	0.127	4	32	1.1587	0.3473	4.6531	0.321
PB	0.22695	0.114	4	36	1.152	0.3481	4.608	0.324
W	0.7744	0.12	4	34	1.1593	0.3459	4.637	0.324

Effect α (sex)								
Statistics	Value	η^2	df_1	df_2	F	Sig	ncp	power
HL	0.0075	0.0075	2	17	0.6391	0.9383	0.1278	0.0582
PB	0.0075	0.0075	2	17	0.6391	0.9383	0.1278	0.0582
W	0.993	0.0075	2	17	0.6391	0.9383	0.1278	0.0582

Effect β (drug)								
Statistics	Value	η^2	df_1	df_2	F	Sig	ncp	power
HL	4.64	0.699	4	32	18.59	<0.0001	74.23	>0.9999
PB	0.88	0.44	4	36	7.077	0.0003	28.31	0.989
W	0.169	0.589	4	34	12.2	<0.0001	48.8	0.9999

progression as a function of n . Although the *Mathematica* function is for factorial MANOVA, the present example is simply a special case where there is only one factor ($l=1$).

Discussion and conclusion

Unfortunately there is no single test that is the most powerful if the MANOVA assumptions are not met. If there is a violation of homogeneity of the covariance matrices or the multivariate normality, then the PB statistic is the most robust while RLR is the least robust statistic. If the noncentrality is concentrated (when the population centroids are largely confined to a single dimension), *RLR* provides the most power test. If on the other hand, the noncentrality is diffuse (when the population centroids differ almost equally in all dimensions) then *PB*, *HT* or *W* will all give good power. However, in most cases, power differences among the four statistics are quite small (<0.06) (Muller et al., 1992; Stevens, 1992), thus it does not matter which statistics is used.

Power estimation is essential in a priori testing. However, for post hoc estimation, careful interpretation must be considered (Lenth, 2001). Since power and p value are inversely proportional (Hoenig & Heisey, 2001) a non significant p value will give a low power; this is

tautological and uninformative. Thus, it makes no sense to use retrospective power to “enhance” the interpretation of a significance test (Baguley, 2004). In fact, for post hoc analysis, likelihood and Bayesian methods (Gelman, Carlin, Stern, & Rubin, 2004) or equivalence testing (Berger & Hsu, 1996; Chervoneva, Hyslop, & Hauck, 2007; Schuirman, 1987) should be preferred. Therefore attention should be focused on the effect size. Consequently, a low observed power reflects the absence of an interesting effect rather than a low sample size (Baguley, 2004).

Moreover, standardized effect sizes, like effect sizes proposed by Cohen (1988), should be avoided (Baguley, 2004; Lenth, 2001). Their uses are particularly wrong when they do not provide any information about the practical importance of a given effect. Moreover standardized effect sizes are function of the measurement error. Thus, it is wrongly assumed that studies with similar standardized effect sizes have similar effect amplitude (Baguley, 2004; Lenth, 2001). In power analysis like any other type of statistical analysis, flawed design can never be corrected by a given analysis. Therefore much attention must be spent in the design phase before it is too late.

It has been shown that power estimation principles for MANOVA is not different from ANOVA. The same steps

are carried out as illustrated in Figure 1. Only independent groups were considered here, however the same principles can be applied for repeated measures (Muller et al., 1992).

References

- Baguley, T. (2004). Understanding Statistical Power in the Context of Applied Research. *Applied Ergonomics*, 35, 73-80.
- Berger, R. L., & Hsu, J. C. (1996). Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Sets. *Statistical Science*, 11(4), 283-319.
- Chervoneva, I., Hyslop, T., & Hauck, W. (2007). A Multivariate Test for Population Bioequivalence. *Statistics in Medicine*, 26, 1208-1223.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian Data Analysis* (2 nd ed.). Boca Raton, Florida: Chapman & Hall/CRC.
- Hoenig, J. M., & Heisey, S. M. (2001). The Abuse of Power: The Pervasive Fallacy of Power Calculations fo Data Analysis. *The American Statistician*, 55(1), 19-24.
- Lenth, R. V. (2001). Some Practical Guidelines for Effective Sample-Size Determination. *The American Statistician*, 55(3), 187-193.
- Morrison, D. F. (1976). *Multivariate statistical methods* (2nd ed.). New York: McGraw-Hill.
- Muller, K. E., LaVange, L. M., Landesman Ramey, S., & Ramey, C. T. (1992). Power Calculations for General Linear Multivariate Models Including Repeated Measures Applications. *Journal of the American Statistical Association*, 87(420), 1209-1226.
- Olson, C. L. (1974). Comparative Robustness of Six Tests in Multivariate Analysis of Variance. *Journal of American Statistical Association*, 69(348), 894-908.
- Schuirmann, D. J. (1987). A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, 15(2), 657-680.
- Stevens, J. (1992). *Applied multivariate statistics for the social sciences* (2nd ed.). Hillsdale, N.J.: L. Erlbaum Associates.
- Tatsuoka, M. M. (1988). *Multivariate analysis : techniques for educational and psychological research* (2nd ed.). New York: Macmillan.
- Thompson, B. (1984). *Canonical correlation analysis: Uses and interpretation*. Beverly Hills: Sage.

Manuscript received November 26th, 2006

Manuscript accepted October 1st, 2007