HE Plots for Multivariate Linear Models

Michael FRIENDLY

Multivariate analysis of variance (MANOVA) extends the ideas and methods of univariate ANOVA in simple and straightforward ways. But the familiar graphical methods typically used for univariate ANOVA are inadequate for showing how measures in a multivariate response vary with each other, and how their means vary with explanatory factors. Similarly, the graphical methods commonly used in multiple regression are not widely available or used in multivariate multiple regression (MMRA). We describe a variety of graphical methods for multiple-response (MANOVA and MMRA) data aimed at understanding what is being tested in a multivariate test, and how factor/predictor effects are expressed across multiple response measures.

In particular, we describe and illustrate: (a) Data ellipses and biplots for multivariate data; (b) HE plots, showing the hypothesis and error covariance matrices for a given pair of responses, and a given effect; (c) HE plot matrices, showing all pairwise HE plots; and (d) reduced-rank analogs of HE plots, showing all observations, group means, and their relations to the response variables. All of these methods are implemented in a collection of easily used SAS macro programs.

Key Words: Biplot; Canonical discriminant plot; Data ellipse; HE plot matrix; MANOVA; Multivariate multiple regression; MMRA; Scatterplot matrix.

1. INTRODUCTION

Among all the inventions in the history of statistics, the classical univariate general linear model (LM), $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, has arguably been the most generally useful, in terms of its range and frequency of use in applied research settings. It has also been the most prolific, in terms of its progeny and extensions.

As is well known, the LM includes as special cases all forms of regression, ANOVA, ANCOVA, and response surface models. Extensions of this basic model include *generalized* linear models, $g(\boldsymbol{y}) = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, such as Poisson regression, logistic regression, and loglinear models, all with non-Gaussian, heteroscedastic error structures, and versions that substitute robust estimation for standard least squares.

Equally remarkable is the range of diagnostic and graphical methods associated with the LM family, both for assessing departures of the data from model assumptions, and for

Michael Friendly is Professor, Psychology Department, York University, Toronto, ON, M3J 1P3 Canada (E-mail: *friendly@yorku.ca*).

© 2007 *American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of North America*

Journal of Computational and Graphical Statistics, Volume 16, Number 2, Pages 421-444
DOI: 10.1198/106186007X208407

assisting the viewer in understanding and communicating the nature of effects, many of which are now widely used in applied research.

As if this is not enough, the classical LM extends quite naturally to the multivariate response setting, at least for a multivariate normal collection of p responses, $(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_p) \equiv \boldsymbol{Y}$. The multivariate linear model (MLM) then becomes $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{U}$. Thus, multivariate analysis of variance (MANOVA) extends the ideas and methods of univariate analysis of variance (ANOVA) in simple and straightforward ways, just as multivariate multiple regression (MMRA) extends univariate multiple regression (MRA) to the multiple response setting.

It is therefore surprising that corresponding graphical methods for multivariate responses are not widely developed, or at least, are not widely known and used. This article describes a variety of graphical methods for multivariate data in the context of the multivariate LM aimed at understanding how variation is reflected in multivariate tests and showing how factor/predictor effects are expressed across multiple response measures.

The outline of this article is as follows: Section 2 describes bivariate plots based on the use of a data ellipse as a sufficient visual summary and extensions of this technique to multivariate displays. Section 3 focuses on HE plots that provide a direct visual comparison of the hypothesis and error covariance matrices involved in any general linear test for a multivariate response. Finally, Section 4 describes a low-dimensional view for MANOVA data based on canonical discriminant analysis. The Appendix describes software for SAS and R that implement these displays.

2. BIVARIATE PLOTS FOR MULTIVARIATE DATA

The principal new ideas we introduce concern the use of ellipses to visualize covariation against multivariate null hypotheses ($\boldsymbol{H}$) relative to error covariation ($\boldsymbol{E}$), and their use in low-rank approximations. To set the stage, we briefly describe the data ellipse and biplot that provide simple conceptual bases for these ideas.

2.1 DATA ELLIPSE

The data ellipse (Dempster 1969; Monette 1990) provides a remarkably simple and effective display for viewing and understanding bivariate marginal relationships in multivariate data. It is typically used to add a visual summary to a scatterplot, indicating the means, standard deviations, correlation, and slope of the regression line for two variables.

It is historically appropriate to illustrate the data ellipse and describe its properties using Galton's (1886, Table I) data (reproduced in Stigler (1986, Table 8.2, p. 286)) on the average heights of parents and their children, shown in Figure 1, where the frequency at each point is shown by a sunflower symbol. We also overlay the 40%, 68%, and 95% data ellipses, as described below.

Galton smoothed the numbers by averaging each set of four adjacent cells, and noticed that: (a) contours of equal frequency formed concentric ellipses, and (b) the locus of vertical and horizontal tangent lines to these ellipses corresponded closely to the traces of the

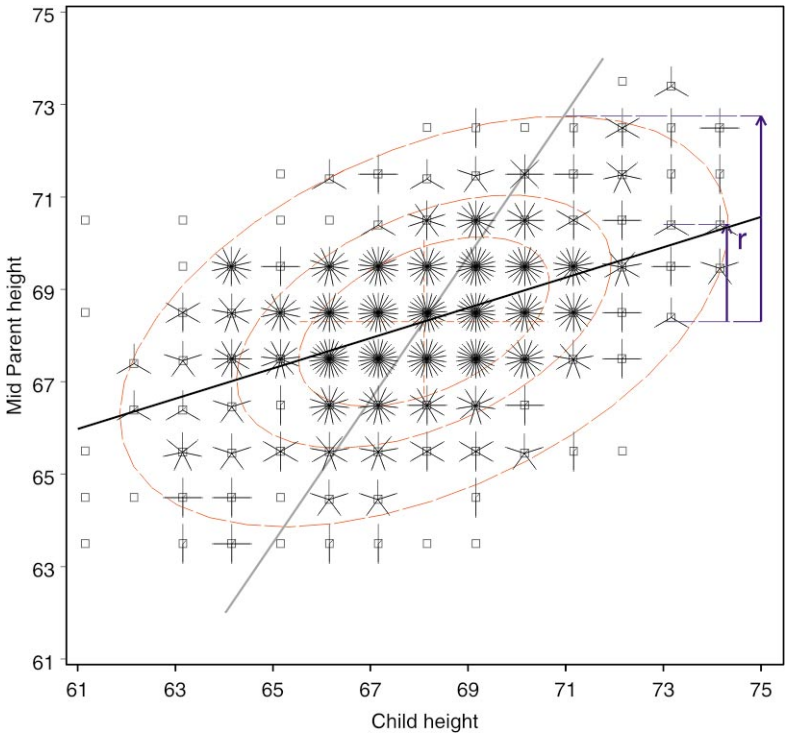

Figure 1. Sunflower plot of Galton's data on heights of parents and their children (in.), with 40%, 68%, and 95% data ellipses and the regression lines of y on x (black) and x on y (gray). The ratio of the vertical to the regression line (labeled "r") to the vertical to the top of the ellipse gives a visual estimate of the correlation ($r = 0.46$, here). Plotting children's height on the abscissa follows Galton.

conditional means, $\bar{y} \mid x$ and $\bar{x} \mid y$ that he called the regression lines of y on x and of x on y, respectively. Galton's graphical insight led directly to the development of correlation and regression and the theory of the bivariate normal distribution (where the ellipses are exact contours of equal density) and later to principal components (Pearson 1901), the major and minor axes of these ellipses, and ultimately to much of the theory of multivariate correlation and regression.

In Figure 1 the ellipses have the mean vector $(\bar{x}, \bar{y})$ as their center; the lengths of arms of the central cross show the standard deviation of each variable, which may be seen to correspond to the shadows of the 40% ellipse. In addition, the correlation coefficient may be visually estimated as the fraction of a vertical tangent line from $\bar{y}$ to the top of the ellipse that is below the regression line $\hat{y}|x$, shown by the arrow labeled "r." See Monette (1990, Fig. 5.2) and Friendly (1991, p. 183) for illustrations and further discussion of the properties of the data ellipse.

Formally, for two variables, Y_1, Y_2, the sample data ellipse $\mathcal{E}_c$ of size c is defined as the set of points $\boldsymbol{y} = (y_1, y_2)'$ whose squared Mahalanobis distance, $D^2(\boldsymbol{y}) = (\boldsymbol{y}-\bar{\boldsymbol{y}})' \boldsymbol{S}^{-1} (\boldsymbol{y}-\bar{\boldsymbol{y}})$, from the means, $\bar{\boldsymbol{y}}$, is less than or equal to c^2,

$$\mathcal{E}_c(\boldsymbol{y}; \boldsymbol{S}, \bar{\boldsymbol{y}}) \equiv \{\boldsymbol{y} : (\boldsymbol{y} - \bar{\boldsymbol{y}})' \boldsymbol{S}^{-1} (\boldsymbol{y} - \bar{\boldsymbol{y}}) \leq c^2\} \ , \tag{2.1}$$

where $\boldsymbol{S}$ is the sample variance-covariance matrix, $\boldsymbol{S} = (n-1)^{-1} \sum_{i=1}^{n} (\boldsymbol{y}_i - \bar{\boldsymbol{y}})'(\boldsymbol{y}_i - \bar{\boldsymbol{y}})$.

When $\boldsymbol{y}$ is (at least approximately) bivariate normal, $D^2(\boldsymbol{y})$ has a large-sample χ^2_2 distribution (χ^2 with 2 df), so taking $c^2 = \chi^2_2(0.68) = 2.28$ gives a "1 standard deviation bivariate ellipse," an analog of the standard interval $\bar{y} \pm 1s$, while $c^2 = \chi^2_2(0.95) = 5.99 \approx 6$ gives a data ellipse of 95% coverage. A bivariate ellipse of $\approx 40\%$ coverage has the property that its shadow on the y_1 or y_2 axes (or any linear combination of y_1 and y_2) corresponds to a univariate $\bar{y} \pm 1s$ interval. In small samples, the distribution of $D^2(\boldsymbol{y})$ can be approximated more closely by $[2(n-1)/(n-2)]F_{2,n-2} \approx 2F_{2,n-2}$; except in tiny samples ($n < 10$), the difference is usually too small to be noticed in a graphical display.

The boundary of the data ellipse, $\mathcal{E}_c$ (where equality holds in Equation (2.1)) may easily be computed as a transformation of a unit circle, $\mathcal{U} = (\sin\boldsymbol{\theta}, \cos\boldsymbol{\theta})$ for $\boldsymbol{\theta} = 0$ to 2π in radians. Let $\boldsymbol{A} = \boldsymbol{S}^{1/2}$ be the Choleski square root of $\boldsymbol{S}$ in the sense that $\boldsymbol{S} = \boldsymbol{A}\boldsymbol{A}'$, whose columns form an orthonormal basis for the inner product $(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}\boldsymbol{S}^{-1}\boldsymbol{v}$. Then $\mathcal{E}_c = \bar{\boldsymbol{y}} + c\boldsymbol{A}\mathcal{U}$ is an ellipse centered at the means, $\bar{\boldsymbol{y}} = (\bar{y}_1, \bar{y}_2)$, whose size reflects the standard deviations of y_1 and y_2 and whose shape (eccentricity) reflects their correlation. For bivariate normal data, the data ellipse is a level curve through the bivariate density.

All of the above results extend immediately to p variables, $y_1, y_2, \ldots, y_p$, giving a p-dimensional $(1-\alpha)$ data ellipsoid $\mathcal{E}_c$ with $c^2 = \chi^2_p(1-\alpha)$ or $c^2 = [p(n-1)/(n-p)]F_{p,n-p}(1-\alpha)$ in small samples.

As useful as the data ellipse might be for a single, unstructured sample, its value as a visual summary increases with the complexity of the data. For example, Figure 2 shows a scatterplot matrix of all pairwise plots of the variables from Anderson's (1935) classic data on three species of iris flowers found in the Gaspé Peninsula, later used by Fisher (1936) in his development of discriminant analysis. The data ellipses show clearly that the means, variances, correlations, and regression slopes differ systematically across the three iris species in all pairwise plots. Exactly how they differ can be shown in several forms, as explained below.

We recognize that a normal-theory summary (first and second moments), shown visually or numerically, can be distorted by multivariate outliers, particularly in smaller samples. Such effects can be countered by using robust covariance estimates such as multivariate trimming (Gnanadesikan and Kettenring 1972) or the high-breakdown bound minimum volume ellipsoid (MVE) and minimum covariance determinant (MCD) methods developed by Rousseeuw and others (Rousseeuw and Leroy 1987; Rousseeuw and Van Driessen 1999). In what follows, it should be noted that robust covariance estimates could, in principle, be substituted for the classical, normal-theory estimates in all cases. To save space, we do not explore these possibilities further here.

2.2 BIPLOTS

The data ellipse, used within a scatterplot matrix as in Figure 2, may show quite effectively the relations among a reasonably large number of variables. Yet there are even better methods for display of complex, high-dimensional datasets.

Among these, the biplot (Gabriel 1971, 1981, 1995) must rank among the most gen-

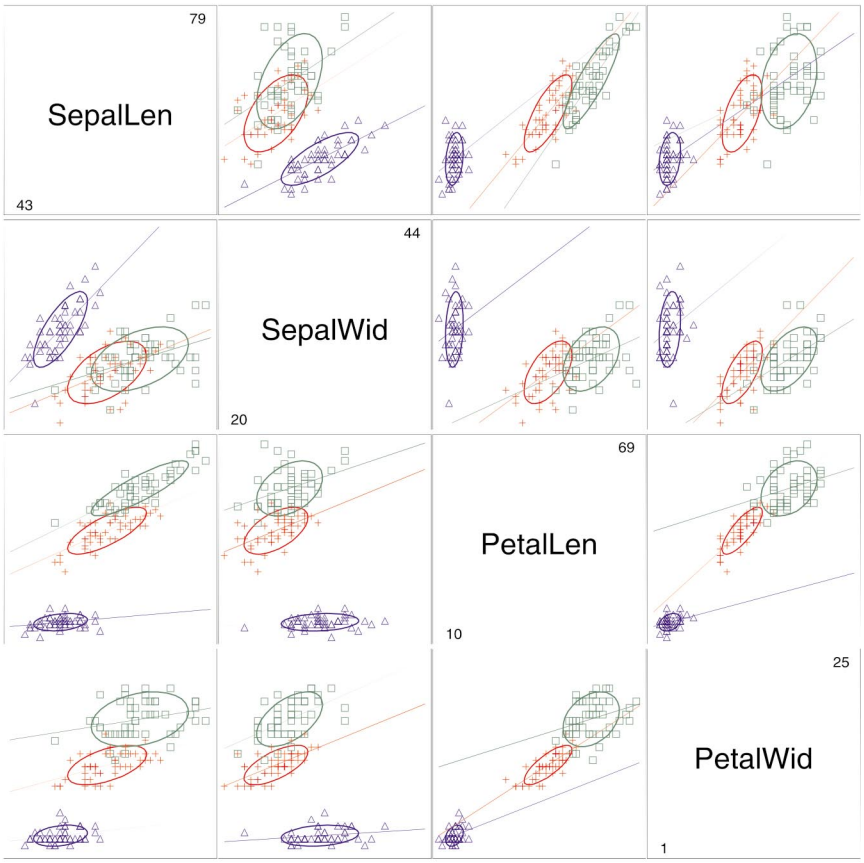

Figure 2. Scatterplot matrix of Anderson's iris data, showing separate 68% data ellipses and regression lines for each species. Key: *Iris setosa*: blue, △; *Iris versicolor*: red, +; *Iris virginica*: green, □.

erally useful. The name "biplot" comes from the fact that this technique displays all the observations and variables in the *same plot*, in a way that depicts their *joint* relationships. For the present purposes, it is useful to explain how the ideas behind the biplot provide a low-dimensional view of a dataset that contributes to an understanding of multivariate LMs, particularly for MANOVA designs.

Biplots can be regarded as the multivariate analog of scatterplots (Gower and Hand 1996), obtained by projecting a multivariate sample into a low-dimensional space (typically of two or three dimensions) accounting for the greatest variance in the data. The (symmetric) scaling of the biplot described here is equivalent to a plot of principal component scores for the observations (shown as points), together with principal component coefficients for the variables (shown as vectors) in the same 2D (or 3D) space (see Figure 3). When there are classification variables dividing the observations into groups, we may also overlay data ellipses for the scores to provide a low-dimensional visual summary of differences among groups in means and covariance matrices.

For brevity, we omit formal description of the biplot (see Gower and Hand 1996) and just note that the row scores and variable vectors in a d-dimensional least-squares approximation

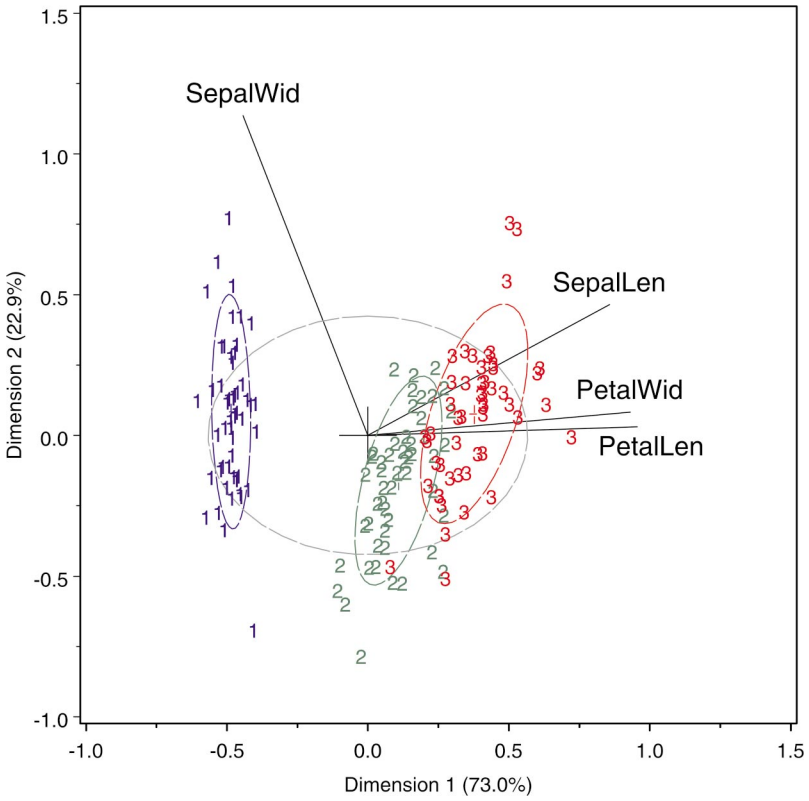

Figure 3. Biplot for iris data, showing observations (points) and variables (vectors), together with the 68% data ellipses (calculated in the reduced space) for each species (*setosa*: blue (1); *versicolor*: green (2); *virginica*: red (3)) and for all species (in gray).

are easily obtained from a singular value decomposition of the mean-centered data matrix, $\boldsymbol{Y}^\star$, of the form

$$\boldsymbol{Y}^\star \approx \sum_{k=1}^{d} \lambda_k \boldsymbol{u}_k \boldsymbol{v}_k' \ , \tag{2.2}$$

The squared singular values, λ_i^2 measure the variance accounted for by each biplot dimension, so the goodness of fit of a d-dimensional solution is $\sum^d \lambda_k^2 / \sum^K \lambda_k^2$ where $K = \min(n, p)$.

As a simple example, Figure 3 shows the biplot for the iris data. In the following discussion it will be instructive for the reader to compare this figure with the bivariate, marginal view of all pairwise plots shown in Figure 2. The 2D projection of the 4D dataset accounts for 95% of the total variation, of which most (73%) is accounted for by the first dimension.

In this plot: (a) the variable vectors have their origin at the mean on each variable, and point in the direction of positive deviations from the mean on each variable. Negative deviations from the mean may be imagined as vectors in the opposite direction for each variable; (b) the angles between variable vectors portray the correlations between them, in

the sense that the cosine of the angle between any two variable vectors approximates the correlation between those variables (in the reduced space). Thus, vectors at right angles reflect correlations of zero, while vectors in the same direction reflect perfect correlations; (c) the relative length of each variable vector indicates the proportion of variance for that variable represented in the low-rank approximation; (d) the orthogonal projections of the observation points on the variable vectors show approximately the value of each observation on each variable; (e) by construction, the observations, shown as principal component scores are uncorrelated, as may be seen from the total sample ellipse (gray ellipse in Figure 3); (f) within-sample correlations, means, and variances in the reduced space are shown by the separate data ellipses, in relation to the grand mean $Y_{\cdot\cdot}$ at the origin, and in relation to the variable vectors.

The interpretation of Figure 3 is as follows: In the total sample, petal width and petal length are nearly perfectly correlated, and these are both highly correlated with sepal length; the two sepal variables are nearly uncorrelated. As well, the three iris species differ primarily along the first dimension, and so are ordered by increasing means on both petal variables (see Figure 2, panel 3:4 in row 3, column 4), but the variances and covariances differ as well.

In addition, several observations stand out as unusual for their species, in relation to the total sample covariation: one "runt" *setosa* (1) with very small sepal size, one *virginica* (3) that looks more like a *versicolor* (2), and perhaps several other *virginica* observations. Robust estimation methods for the data ellipses would more effectively highlight potential outliers. Canonical discriminant plots (Section 4) provide another low-dimensional view for MANOVA data (see Figure 13, p. 21).

3. HE PLOTS

For the multivariate linear model, *any* hypothesis test may be calculated from an analog of the univariate F, where $p \times p$ matrices, $\boldsymbol{H}$ and $\boldsymbol{E}$ play the roles of univariate sums of squares, SS_H and SS_E. But, in the multivariate case, the variation against the null hypothesis ($\boldsymbol{H}$) may be large in one or more dimensions relative to the error variation ($\boldsymbol{E}$).

The HE plot provides a two-dimensional visualization of the size of the $\boldsymbol{H}$ matrix relative to the size of the $\boldsymbol{E}$ matrix for any multivariate test. It does this, in several forms explained below, in terms of the covariance ellipses for two response variables from the $\boldsymbol{H}$ and $\boldsymbol{E}$ matrices. This is useful for understanding the differences among the various multivariate test statistics, for visualizing the sizes of various model effects, and how predictive power of a non-null model is distributed across the various responses.

In the MANOVA context the HE plot shows directly how the group means differ on the two variables jointly, in relation to the within-group variation. The canonical discriminant plot (Section 4) shows an alternative view for *all* variables, in a two-dimensional space that maximizes the differences among groups. These methods extend directly to other cases of the MLM.

3.1 BRIEF REVIEW OF THE MULTIVARIATE LM

To establish notation and context for HE plots, we begin with a capsule summary of the multivariate LM and the *general linear test* for any hypothesis. For details, see, for example, Timm (1975) or Muller, LaVange, Ramey, and Ramey (1992).

When there are p responses, $(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_p) = \boldsymbol{Y}$, the multivariate LM

$$\underset{(n\times p)}{\boldsymbol{Y}} = \underset{(n\times q)}{\boldsymbol{X}}\,\underset{(q\times p)}{\boldsymbol{B}} + \underset{(n\times p)}{\boldsymbol{U}} \quad , \tag{3.1}$$

with $\text{vec}(\boldsymbol{U}) \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$, where $\otimes$ is the Kronecker product, is a natural extension of the univariate version. Except for the fact that hypotheses are tested using multivariate tests, model Equation (3.1) is equivalent to the set of p models for each separate response, $\boldsymbol{y}_i = \boldsymbol{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$ for $i = 1, 2, \ldots p$, where the columns of $\boldsymbol{B} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots \boldsymbol{\beta}_p)$ in Equation (3.1) are the model coefficients for the separate responses. In all cases, the least squares estimates of the coefficients, $\boldsymbol{B}$ can be calculated as $\widehat{\boldsymbol{B}} = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{Y}$, where $\boldsymbol{A}^{-}$ denotes a generalized inverse. For simplicity, we do not distinguish here among various parameterizations for factor variables (e.g., sum-to-zero constraints, first/last parameter = 0, contrasts, etc.) that provide different unique solutions for parameter estimates, but which all yield identical overall tests for model effects.

Then, just as all linear models can be cast in the form of the LM all tests of hypotheses (for null effects) can be represented in the form of a general linear test,

$$H_0 : \underset{(h\times q)}{\boldsymbol{C}}\,\underset{(q\times p)}{\boldsymbol{B}} = \underset{(h\times p)}{\boldsymbol{0}} \quad , \tag{3.2}$$

where $\boldsymbol{C}$ is a matrix of constants whose rows specify h linear combinations or contrasts of the parameters to be tested simultaneously by a multivariate test. (For repeated measures designs, an extended form of the general linear test, $\boldsymbol{C}\boldsymbol{B}\boldsymbol{A} = \boldsymbol{0}$, where $\boldsymbol{A}$ is a $p \times k$ matrix of constants, provides analogous contrasts or linear combinations of the responses to be tested. We do not pursue this straightforward extension here.)

For *any* such hypothesis of the form Equation (3.2), the analogs of the univariate sums of squares for hypothesis (SS_H) and error (SS_E) are the $p \times p$ sum of squares and crossproducts (SSCP) matrices (Timm 1975, chap. 3, 5)

$$\boldsymbol{H} = (\boldsymbol{C}\widehat{\boldsymbol{B}})' \, [\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}']^{-1} \, (\boldsymbol{C}\widehat{\boldsymbol{B}}) \quad , \tag{3.3}$$

and

$$\boldsymbol{E} = \boldsymbol{Y}'\boldsymbol{Y} - \widehat{\boldsymbol{B}}'(\boldsymbol{X}'\boldsymbol{X})\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{U}}'\widehat{\boldsymbol{U}} \quad . \tag{3.4}$$

For example, in a one-way MANOVA, using the cell-means model for the vector of responses of subject j in group i, $\boldsymbol{y}_{ij} = \boldsymbol{\mu}_i + \boldsymbol{e}_{ij}$, the $\boldsymbol{H}$ and $\boldsymbol{E}$ SSCP matrices for testing $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots \boldsymbol{\mu}_g$ are

$$\boldsymbol{H} = \sum_{i=1}^{g} n_i \, (\bar{\boldsymbol{y}}_{i.} - \bar{\boldsymbol{y}}_{..}) \, (\bar{\boldsymbol{y}}_{i.} - \bar{\boldsymbol{y}}_{..})' \quad , \tag{3.5}$$

and

$$\boldsymbol{E} = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_{i\cdot})\,(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_{i\cdot})' \ . \tag{3.6}$$

Then, the multivariate analog of the univariate F statistic,

$$F = \frac{\mathrm{SS}_H/\mathrm{df}_h}{\mathrm{SS}_E/\mathrm{df}_e} = \frac{\mathrm{MS}_H}{\mathrm{MS}_E}, \text{ which implies } (\mathrm{MS}_H - F\,\mathrm{MS}_E) = 0$$

is

$$\det(\boldsymbol{H} - \lambda \boldsymbol{E}) = 0 \ , \tag{3.7}$$

where λ represents the $s = \min(p, \mathrm{df}_h)$ nonzero latent roots of the $\boldsymbol{H}$ matrix relative to ("in the metric of") the $\boldsymbol{E}$ matrix, or equivalently, the ordinary latent roots of the matrix $\boldsymbol{H}\boldsymbol{E}^{-1}$

$$\det(\boldsymbol{H}\boldsymbol{E}^{-1} - \lambda \boldsymbol{I}) = 0 \ . \tag{3.8}$$

The ordered latent roots, $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_s$ measure the "size" of $\boldsymbol{H}$ relative to the "size" of $\boldsymbol{E}$ in up to s orthogonal directions, and are minimal sufficient statistics for all multivariate tests. These tests can also be expressed in terms of the eigenvalues $\theta_i = \lambda_i/(1 + \lambda_i)$ of $\boldsymbol{H}\boldsymbol{T}^{-1}$, where $\boldsymbol{T} = \boldsymbol{H} + \boldsymbol{E}$, and $\theta_i = \rho_i^2$ are the generalized squared canonical correlations. The various multivariate test statistics (Wilks' Λ, Pillai's trace criterion, Hotelling-Lawley trace criterion, Roy's maximum root criterion) reflect different ways of combining this information across the dimensions, ranging from functions of their product (Wilks' Λ), to functions of their sum (Pilai, Hotelling-Lawley), to their maximum (Roy).

Thus, in univariate problems ($p = 1$), or for 1 degree-of-freedom hypotheses ($\mathrm{df}_h = 1$), there is $s = 1$ nonzero latent root, that has an exact relation to a univariate F. When $s > 1$, the number of "large" dimensions indicate how many different aspects of the responses are being explained by the hypothesized effect. These ideas can be illustrated conceptually, and understood for data through plots of the $\boldsymbol{H}$ and $\boldsymbol{E}$ ellipses, as described below.

3.2 DIDACTIC DISPLAYS FOR MANOVA TESTS

From the description above, it is relatively easy to provide a visual explanation of the essential ideas behind all multivariate tests, particularly in the MANOVA context, as shown in Figure 4.

Figure 4(a) shows the individual-group data ellipses for two hypothetical variables, Y_1, Y_2. The variation due to differences in the group means is captured by the $\boldsymbol{H}$ matrix, while the pooled within-sample variation is captured by the $\boldsymbol{E}$ matrix, as illustrated in panel (b). The answer to the question, "How big is $\boldsymbol{H}$ relative to $\boldsymbol{E}$" is shown geometrically in the last two panels.

The transformation from Equation (3.7) to Equation (3.8) can be represented (panel (c)) as a rotation of the variable space to the principal components of $\boldsymbol{E}$, giving the matrix $\boldsymbol{E}^\star$. The same transformation applied to the $\boldsymbol{H}$ matrix gives $\boldsymbol{H}^\star$. The axes in panels (c)

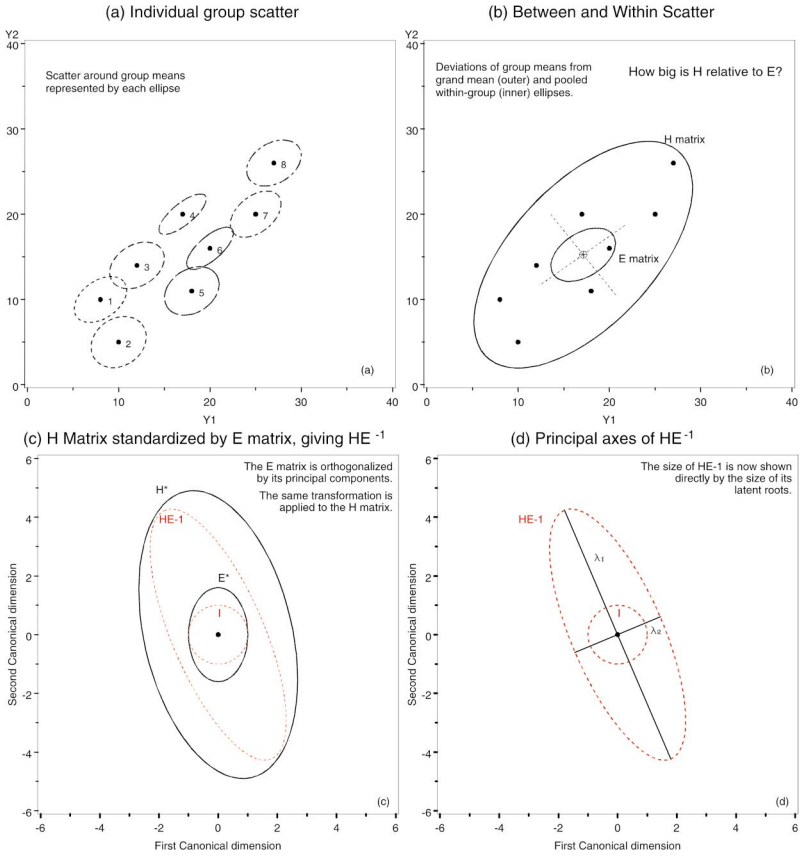

Figure 4. Conceptual plots showing the essential ideas behind multivariate tests, in terms of the hypothesis ($\boldsymbol{H}$) and error ($\boldsymbol{E}$) matrices, for a 1-way MANOVA design with two response variables (Y_1, Y_2): (a) Bivariate means (points) and within-group variance-covariance matrices (ellipses); (b) The hypothesis ($\boldsymbol{H}$) matrix reflects the variation of bivariate group means around the grand mean. The error ($\boldsymbol{E}$) reflects the pooled within-group dispersion and covariation. (c) Standardizing: The $\boldsymbol{E}$ matrix can be standardized, first to its principal components ($\boldsymbol{E}^\star$) and then normalized. The same transformations are applied to the $\boldsymbol{H}$ matrix, giving $\boldsymbol{H}\boldsymbol{E}^{-1}$. (d) The ellipsoid of $\boldsymbol{H}\boldsymbol{E}^{-1}$ then shows the size and dimensionality of the variation in the group means in relation to a spherical error ellipsoid.

and (d) turn out to be the canonical discriminant dimensions discussed in the following section (Section 4). In this space, the errors (residuals) are all uncorrelated, that is, $\boldsymbol{E}^\star$ is diagonal, but with possibly different variances. Standardizing then transforms $\boldsymbol{E}^\star \mapsto \boldsymbol{I}$ and $\boldsymbol{H}^\star \mapsto \boldsymbol{H}\boldsymbol{E}^{-1}$.

Because the transformed errors are now uncorrelated and standardized to unit variance, we can focus only on the ellipse for $\boldsymbol{H}\boldsymbol{E}^{-1}$ as shown in panel (d), where the latent roots, λ_1, λ_2 are the half-lengths of the major and minor axes.

3.3 VARIETIES OF HE PLOTS

From Figure 4 and the preceding discussion it may be seen that there are several different ways to display the $\boldsymbol{H}$ and $\boldsymbol{E}$ matrices for a given effect in a multivariate test, as illustrated

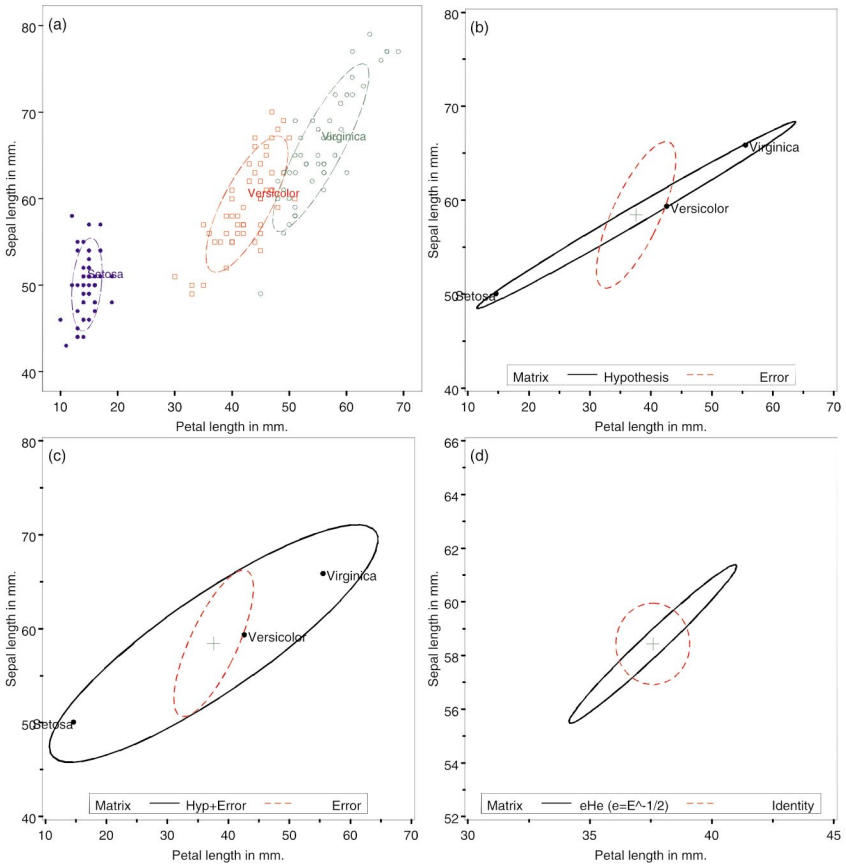

Figure 5. Data and HE plots for iris data, showing the relation between sepal length and petal length in the iris data. (a) data ellipses; (b) $\boldsymbol{H}$ and $\boldsymbol{E}$ matrices; (c) $\boldsymbol{H}+\boldsymbol{E}$ and $\boldsymbol{E}$ matrices; (d) $\boldsymbol{H}\boldsymbol{E}^{-1}$ and $\boldsymbol{I}$ matrices.

in Figure 5.

Panel (a) shows the observations and the data ellipses for sepal length and petal length, as in panel (1:3) in Figure 2. The $\boldsymbol{H}$ and $\boldsymbol{E}$ matrices are shown in panel (b). The shape of the $\boldsymbol{H}$ ellipse shows that the variation in the group means is essentially 1D, a dimension of overall (petal + sepal) length.

Alternatively, it is sometimes useful to plot the ellipses for $\boldsymbol{H}+\boldsymbol{E}$ and $\boldsymbol{E}$ as shown in panel (c). This form is particularly useful for multiway designs, so that each effect (e.g., $\boldsymbol{H}_A$, $\boldsymbol{H}_B$, $\boldsymbol{H}_{AB}$) can be seen in relation to error ($\boldsymbol{E}$) variation (see Figure 9, p. 15). When the variation due to a given hypothesis is small relative to error—leading to acceptance of H_0—the $\boldsymbol{H}$ and $\boldsymbol{E}$ ellipses will nearly coincide. The lengths of the major/minor axes of $\boldsymbol{H}+\boldsymbol{E}$ are $1+\lambda_i$, and Wilks' $\Lambda = \prod_{i=1}^{s}(1+\lambda_i)^{-1}$ is inversely proportional to the area (volume when $s > 2$) of the $\boldsymbol{H}+\boldsymbol{E}$ ellipse.

In these plots, $\boldsymbol{E}$ in Equation (3.4) is scaled to a covariance matrix (giving $\boldsymbol{S}_{\text{pooled}} = \boldsymbol{E}/\text{df}_e$ for a MANOVA), so that it is on the same scale as the data ellipses, and the same scaling is applied to $\boldsymbol{H}$ (or $\boldsymbol{H}+\boldsymbol{E}$), so we plot $\mathcal{E}_c(\boldsymbol{y}; \boldsymbol{H}/\text{df}_e, \bar{\boldsymbol{y}})$ and $\mathcal{E}_c(\boldsymbol{y}; \boldsymbol{E}/\text{df}_e, \bar{\boldsymbol{y}})$. This also allows us to show the group means on the plot as an aid to interpretation, and the H matrix

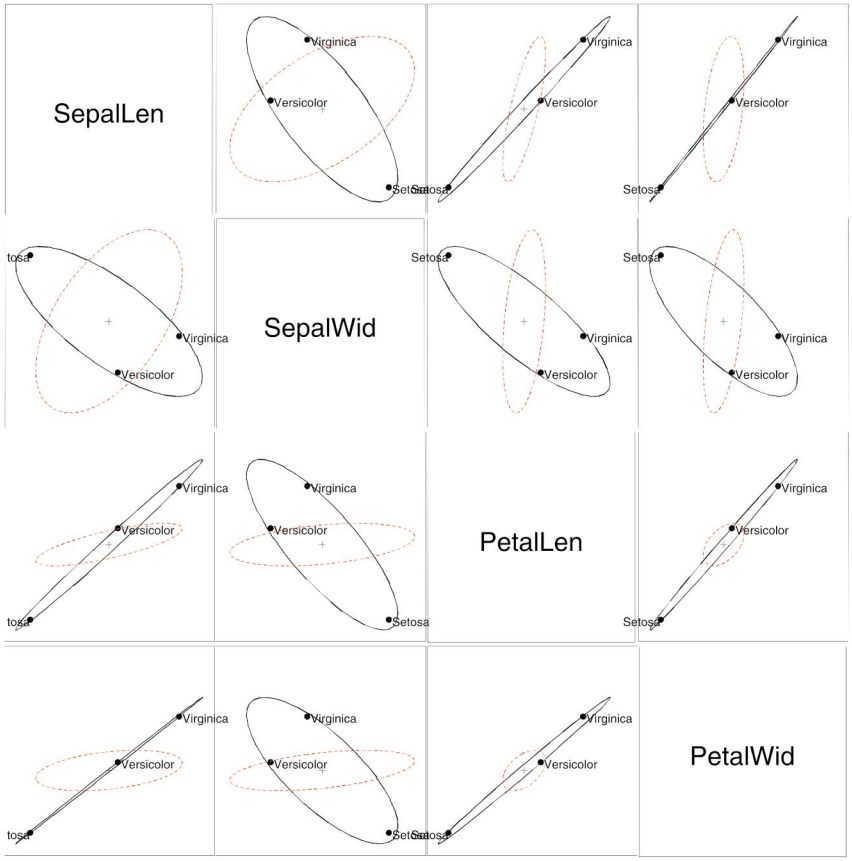

Figure 6. HE plot matrix for iris data. Each panel displays the $\boldsymbol{H}$ (solid, black) and $\boldsymbol{E}$ (dashed, red) bivariate ellipsoids.

then reflects the effect size (similar to the square of Cohen's (1977) $d = (\bar{x}_1 - \bar{x}_2)/s_{\text{pooled}}$) as well as its orientation and shape.

Finally, one may plot the ellipse for $\boldsymbol{H}\boldsymbol{E}^{-1}$ (or the equivalent, symmetric matrix, $\boldsymbol{E}^{-1/2}\boldsymbol{H}\boldsymbol{E}^{-1/2}$) in relation to the identity matrix, $\boldsymbol{I}$, representing uncorrelated errors of unit variance, as shown in panel (d). The Hotelling-Lawley trace statistic, HLT = $\text{tr}(\boldsymbol{H}\boldsymbol{E}^{-1}) = \sum \lambda_i$, reflects the sum of lengths of the major and minor axes; the length of the major axis reflects Roy's criterion, $\theta_1 = \lambda_1/(1+\lambda_1)$. This does not allow the group means to be shown on the same plot, however.

We resist the temptation (suggested by a reviewer) to provide a scaling or visual guides to whether a multivariate hypothesis test would lead to "rejection," because different multivariate tests focus on different aspects of the "size" of the matrices displayed, and this cannot be done in a general and unambiguous way. Rather, we regard the HE plot as showing the size and shape of predictive variation for a hypothesized effect relative to a reduced model in which that effect is null, with the representation of $\boldsymbol{E}$ providing the metric.

To simplify discussion, we illustrate HE plots below using $\boldsymbol{H}$ and $\boldsymbol{E}$, and standard 1 s.d. (68%) coverage, unless otherwise noted.

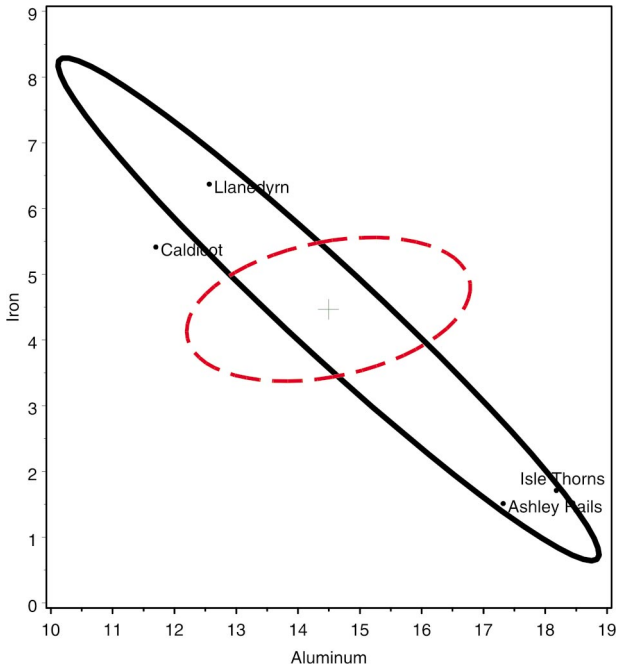

Figure 7. HE plot for pottery data, showing iron and aluminum. H matrix: (solid); E matrix: (dashed).

3.4 HE PLOT MATRICES

The HE plots described above show the variation due to hypothesis relative to error for two response variables. But it is not hard to show the pairwise relations for *all* response variables, using the framework of a scatterplot matrix, plotting all pairs of response variables, in an HE plot matrix.

Figure 6 shows a simple example, again for the iris data, and displaying the separate $\boldsymbol{H}$ and $\boldsymbol{E}$ bivariate ellipses in each panel. Comparing this with the full scatterplot matrix (Figure 2, p. 5) one can regard the HE plot matrix as a "visual thinning" of the data, here focused on the predictive variation due to group mean differences relative to within-group variation. As well, the negative relations of species means on sepal width again stand out, compared with the strong positive relations for all other variables (see Figure 3, p. 6).

3.5 MANOVA EXAMPLES

Tubb, Parker, and Nickless (1980) used atomic absorption spectrophotometry to analyze data on the element composition of 26 samples of Romano-British pottery found at four different kiln sites in Britain with a view to determining if the chemical content of aluminium, iron, magnesium, calcium, and sodium could differentiate those sites; see also Hand et al. (1994, p. 252). If so, the chemical content of pottery of unknown origin might be used for classification purposes. The data thus comprise a one-way MANOVA design with four groups and five variates.

Figure 7 shows the plot of the $\boldsymbol{H}$ and $\boldsymbol{E}$ matrices for two of the variables, iron and

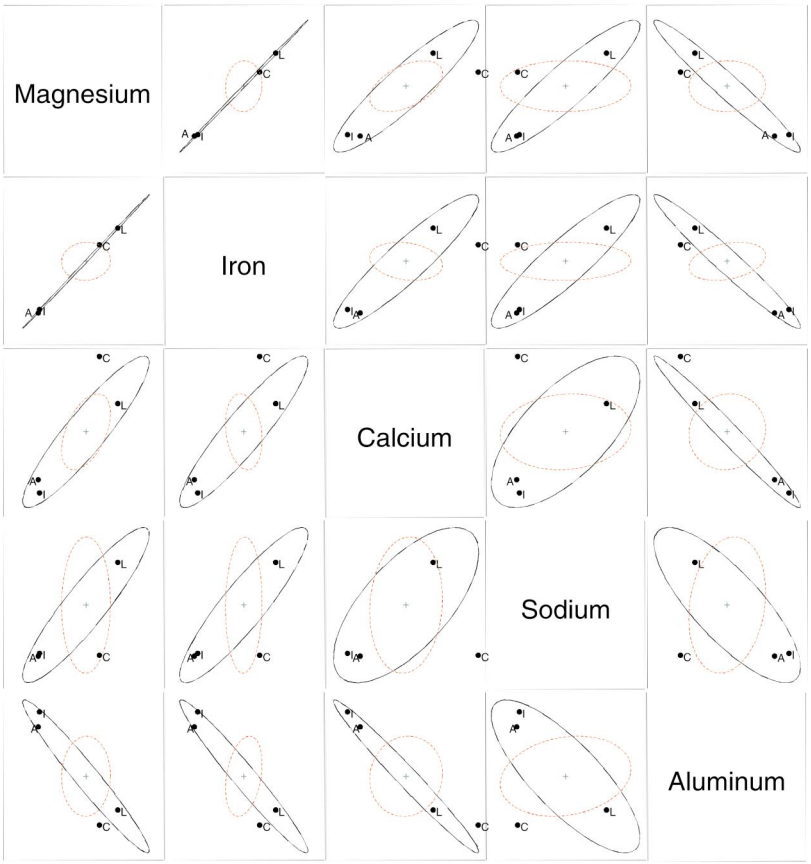

Figure 8. HE plot matrix for pottery data. Group means for the sites are labeled by their initial letters: **A**shley Rails, **C**aldicot, **I**sle Thorns, **L**lanedryn.

aluminum, and the means for the four sites. It is clear that the sites differ primarily in terms of a contrast between Caldicot and Llanedryn versus Ashley Rails and Isle Thorns, and that the means on these two chemical components are negatively related, while the pooled within group scatter shows a weak positive relation.

The variation across sites on *all* chemical components may be seen in the HE plot matrix in Figure 8. Quite a lot may be read directly from this plot. For example: the site means for magnesium and iron are nearly perfectly correlated, and have the same pattern with all other variables, while all mean differences for aluminium are in the opposite direction; the relations for calcium and sodium also differ somewhat from those for magnesium and iron in that Caldicot samples are quite high on calcium, while Llanedryn is high on sodium.

For two-way and higher-order MANOVA designs, HE plots provide a compact, visual summary of the multivariate tests for various main effects and interactions. To illustrate, Figure 9 uses a textbook example (Morrison 1990, p. 217, Table 5.5) dealing with possible toxic effects of three drugs (A, B, C) on rats, also classified by sex (M, F), where the responses are weight losses on two consecutive weeks (Week1, Week2), treated here as a two-way MANOVA design.

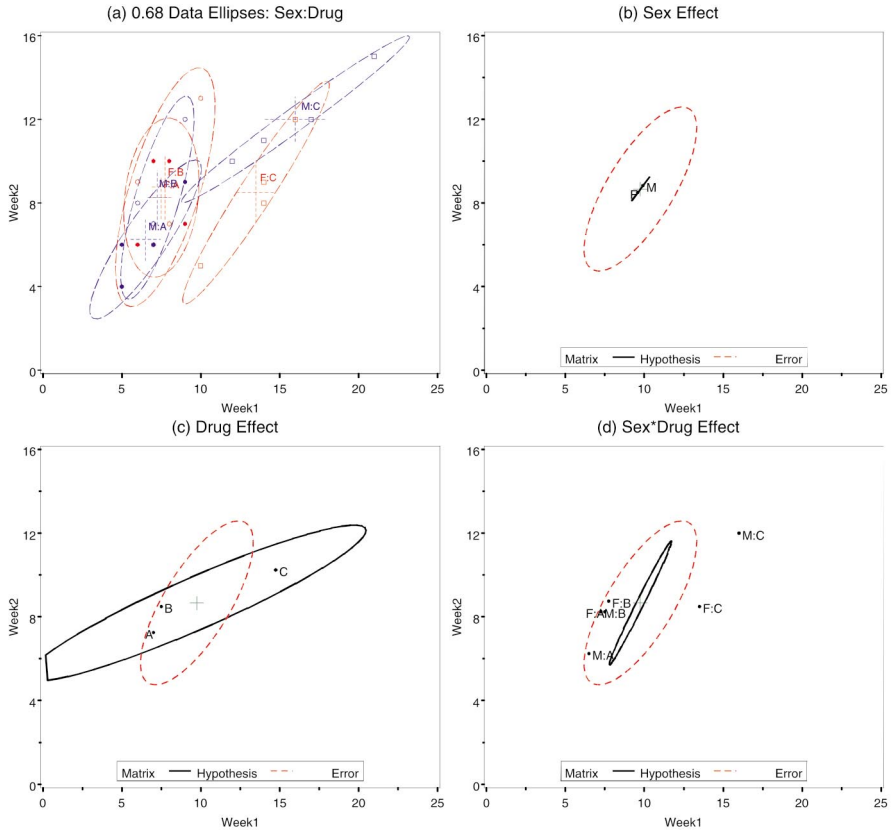

Figure 9. Data ellipses and HE plots for two-way design: Sex (M, F) × Drug (A, B, C). (a) Data ellipses; (b) Sex effect; (c) Drug effect; (d) Sex * Drug interaction.

From the data ellipses (Figure 9 (a)) it is apparent that groups given drug C differ substantially from the remaining groups, which do not appear to differ among themselves, with the possible exception of group M:A. These are very small samples ($n = 4$ per cell); with larger samples, the assumption of equal within-group covariance matrices might be questioned. The HE plots (Figure 9 (b)–(d)) show that differences among drugs are quite large; the main effect of sex is inconsequential, and any hint of a sex*drug interaction is confined to the larger and opposite sex difference with drug C than the other two. Note that for a one degree-of-freedom test ($s = 1$), such as sex in this example, the H ellipse degenerates to a line, a result we exploit below.

3.6 CONTRASTS

Just as in univariate ANOVA designs, important overall effects ($\text{df}_h > 1$)) in MANOVA may be usefully explored and interpreted by the use of contrasts among the levels of the factors involved. In the general linear test Equation (3.2), contrasts are easily specified as one or more ($h_i \times q$) $\boldsymbol{C}$ matrices, $\boldsymbol{C}_1, \boldsymbol{C}_2, \ldots$, each of whose rows sum to zero.

As an important special case, for an overall effect with df_h degrees of freedom (and

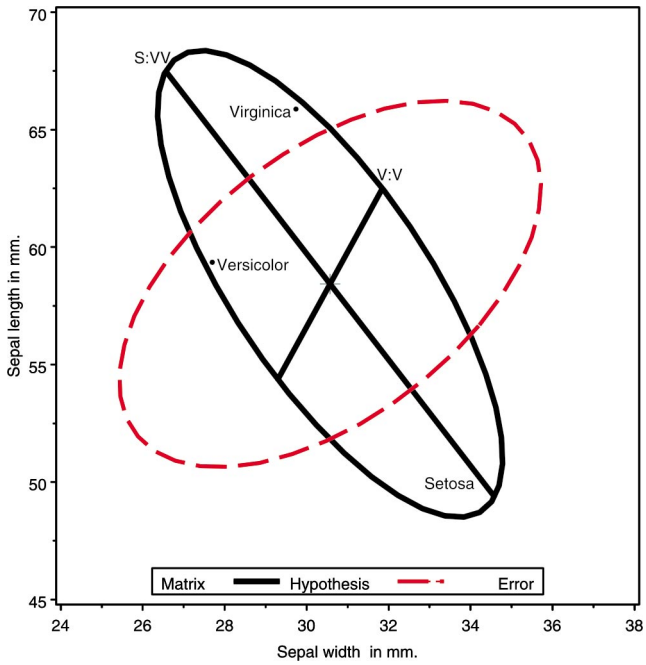

Figure 10. $\boldsymbol{H}$ and $\boldsymbol{E}$ matrices for sepal width and sepal length in the iris data, together with $\boldsymbol{H}$ matrices for testing two orthogonal contrasts in the species effect.

balanced sample sizes), a set of df_h pairwise orthogonal $(1 \times q)$ $\boldsymbol{C}$ matrices ($\boldsymbol{C}_i' \boldsymbol{C}_j = 0$) gives rise to a set of df_h rank 1 $\boldsymbol{H}_i$ matrices that additively decompose the overall hypothesis SSCP matrix,

$$\boldsymbol{H} = \boldsymbol{H}_1 + \boldsymbol{H}_2 + \cdots + \boldsymbol{H}_{\text{df}_h} \ ,$$

exactly as the univariate SS_H may be decomposed in an ANOVA. Each of these rank 1 $\boldsymbol{H}_i$ matrices will plot as a vector in an HE plot, and their collection provides a visual summary of the overall test, as partitioned by these orthogonal contrasts.

To illustrate, we show in Figure 10 an HE plot for the sepal width and sepal length variables in the iris data, corresponding to panel (1:2) in Figure 2 (p. 5). Overlayed on this plot are the 1 df $\boldsymbol{H}$ matrices obtained from testing two orthogonal contrasts among the iris species: *setosa* versus the average of *versicolor* and *virginica* (labeled "S:VV"), and *versicolor* versus *virginica* ("V:V"), for which the contrast matrices are

$$\begin{aligned} \boldsymbol{C}_1 &= \begin{pmatrix} -2 & 1 & 1 \end{pmatrix} \\ \boldsymbol{C}_2 &= \begin{pmatrix} 0 & 1 & -1 \end{pmatrix} \end{aligned}$$

where the species (columns) are taken in alphabetical order.

This HE plot shows that, for the two sepal variables, the greatest between-species variation is accounted for by the contrast between *setosa* and the others, for which the effect is very large in relation to error (co-)variation. The second contrast, between the *versicolor*

and *virginica* species is relatively smaller, but still explains some variation of the sepal variables among the species.

The general method described above applies more widely than we have illustrated. Multiple-df tests may be expressed in terms of $\boldsymbol{C}$ matrices with $h_i > 1$ rows. In a bivariate HE plot, their $\boldsymbol{H}$ matrices will appear as ellipses for these contrasts contained within the H ellipse for the overall test.

3.7 HE PLOTS FOR MMRA

The ideas behind HE plots extend naturally to multivariate multiple regression, with the only difference that the $\boldsymbol{X}$ matrix in the MLM model (Equation (3.1)) consists only of quantitative regressors (possibly transformed, or with polynomial terms). In this context, several types of hypotheses are of interest, all of which can be espressed in the form of a multivariate general linear test Equation (3.1), and shown visually in an HE plot.

In MMRA, it is typical to carry out an initial overall multivariate test of the hypothesis $\boldsymbol{B} = \boldsymbol{0}$ that all regression coefficients for all responses are zero, analogous to the use of an overall test in a univariate analysis to control for multiple tests. But as well, the multivariate test takes the correlations of responses into account, so it is possible that an overall multivariate test can show significance, while some or even none of the univariate overall or individual MRA tests do so. The converse is also sometimes true: an overall multivariate test may fail significance, while tests for individual predictors may prevail. HE plots for these tests can help to make these results understandable.

For MMRA, the overall test, $\boldsymbol{B} = \boldsymbol{0}$, of no linear relation between the X and Y variables collectively corresponds to $\boldsymbol{C} = \boldsymbol{I}$ in Equation (3.3) and the $(p \times p)$ H matrix becomes

$$\boldsymbol{H} = \widehat{\boldsymbol{B}}' (\boldsymbol{X}'\boldsymbol{X}) \widehat{\boldsymbol{B}} = \widehat{\boldsymbol{Y}}'\widehat{\boldsymbol{Y}} \ ,$$

where $\boldsymbol{H}$ is of rank $s = \min(p, q)$ and therefore has s nonzero latent roots. (For simplicity, we assume that all response variables are expressed in terms of deviations from their means, so all intercepts are zero and can be ignored here.)

For any two responses, the overall test can be shown as an HE plot as we have shown before. The $\boldsymbol{H}$ ellipse is simply the data ellipse of the fitted values $\widehat{\boldsymbol{Y}}$; the $\boldsymbol{E}$ ellipse is the data ellipse of the residuals, $\boldsymbol{U} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}$ (shifted to the centroid). For an individual regressor, the test of $H_0 : \boldsymbol{\beta}_i = \boldsymbol{0}$ for the ith row of $\boldsymbol{B}$ also gives rise to a $(p \times p)$ H matrix, obtained using the $1 \times q$ matrix $\boldsymbol{C} = (0, 0, \ldots 1, 0, \ldots 0)$, with a 1 in the ith position. In this case $\boldsymbol{H}_i = \hat{\boldsymbol{\beta}}_i'(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}}_i$, is a matrix of rank 1, with one non-zero latent root, so the ellipse for the H matrix degenerates to a line.

Thus, in an HE plot for multivariate multiple regression, the overall H ellipse shows how the predictors relate collectively to the two chosen response variables. The H ellipse for an individual predictor appears as a vector, whose orientation shows how that variable is related and whose relative length shows the strength of the relations to the response variables displayed.

The collection of these individual vectors helps to show how the various predictors contribute to the overall test. For example, when the predictors are related in the same or

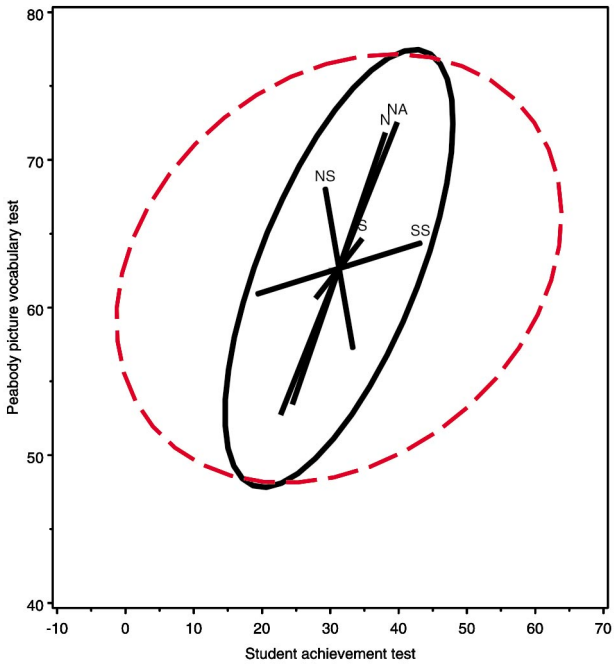

Figure 11. HE plot for MMRA, showing the H ellipse for an overall test, $\boldsymbol{B} = \mathbf{0}$, and the H ellipses for individual predictors.

similar way to the responses, only the first λ_i will be large, the overall H ellipse will be highly eccentric, and the individual predictor ellipse vectors will be aligned. In this case, there is nearly one dimension (linear combination) of the responses being predicted (and Roy's maximum root test will be most powerful). Alternatively, when two or more λ_i are large, the H ellipsoid will have several large dimensions, reflecting the fact that two or more *different* aspects of the responses are being predicted (and the other multivariate tests will be more powerful).

To illustrate, we use data from a study by Rohwer (Timm 1975, Table 4.7.1) on $n = 37$ children designed to determine how well a set of paired-associate (PA) tasks predicted performance on the Peabody Picture Vocabulary test (PPVT), a student achievement test (SAT), and the Raven Progressive matrices test. The PA tasks varied in how the stimuli were presented, and are called *named* (N), *still* (S), *named still* (NS), *named action* (NA), and *sentence still* (SS).

For this data, univariate multiple regressions carried out for each response variable show that the SAT and RAVEN fail significance on an overall test for the $q = 5$ predictors (SAT: $F(5, 31) = 1.63$, $R^2 = 0.201$; RAVEN: $F(5, 31) = 1.77$, $R^2 = 0.222$). For the PPVT, the overall univariate test is significant ($F(5, 31) = 6.47$, $R^2 = 0.510$), but among the partial tests for individual predictors, only one (NA) attains significance.

From these results, one might conclude that PA tasks are at best marginally related to the intellectual and achievement tests. However, the overall multivariate test, $\boldsymbol{B} = \mathbf{0}$, is highly significant. The HE plot for SAT and PPVT in Figure 11 helps to understand these results.

It may be seen that although the error covariance for these variables is nearly circular, the H matrix structure is more highly concentrated, with generally positive correlations among the predictors, particularly for the N and NA tasks. This allows the multivariate tests to "pool strength" across predictors, resulting in greater power for overall test.

4. REDUCED RANK HE PLOTS

The HE plot shows the relation of variation in the group means on two variables relative to error variance. To visualize these relations on more than two variables, we have used the framework of the scatterplot matrix.

However, just as with the biplot, we can see all variables together in a single plot by projecting the data and variable vectors into a two-dimensional subspace that captures most of the variation due to hypothesis relative to error. Canonical discriminant analysis (CDA) is a dimension-reduction technique intimately related to MANOVA, canonical correlation and linear discriminant analysis that provides the basis for such a view.

CDA may be characterized as a method for finding linear combinations of the variables which maximally discriminate among the group mean vectors, in the sense of giving the largest possible univariate F statistics. It is also equivalent to a canonical correlation analysis between the set of response variables and a set of dummy variables representing group membership.

To illustrate this idea, consider the HE plot matrix of the pottery data shown in Figure 8 (p. 14). Each pairwise plot shows a marginal 2D projection of the 5D $\boldsymbol{H}$ and $\boldsymbol{E}$ matrices and corresponding group means. But in fact the $\boldsymbol{H}$ and $\boldsymbol{E}$ matrices are of rank $s = \min(p, \mathrm{df}_h) = 3$, and so can be embedded perfectly in a 3D space. The 2D canonical discriminant projection that accounts for the greatest variation between group means relative to error is shown in Figure 12, an HE plot of the first two canonical variates. In this case, it turns out that two dimensions account for 99.92% of the between-group variation, so the plot may be regarded as a much simpler portrayal of the differences among groups on all variables with no loss of information.

From this plot, it can be seen that most (96.4%) of the between-group variance may be accounted for by the difference between two pairs of sites: (Caldicot, Llanedyrn) versus (Ashley Rails, Isle Thorns), where the latter is higher on aluminum (Al) content and the former is higher on all others, but principally on magnesium (Mg) and iron (Fe) content. There is little difference between Ashley Rails and Isle Thorns on any of these variables, but Caldicot and Llanedyrn differ somewhat on the second canonical dimension (3.5%), with Caldicot higher on calcium (Ca). This type of display provides an optimally compact low-dimensional summary of the MLM in the MANOVA setting. We briefly describe some of the theory below; see Seber (1984) for more statistical details and Friendly (1991) for computational methods.

Formally, for a one-way design with g groups and p-variate observations $\boldsymbol{y}_{ij}$, CDA finds a set of $s = \min(p, g-1)$ linear combinations, $z_1 = \boldsymbol{c}_1'\boldsymbol{y}$, $z_2 = \boldsymbol{c}_2'\boldsymbol{y}$, $\ldots$, $z_s = \boldsymbol{c}_s'\boldsymbol{y}$, so that: (a) all z_k are mutually uncorrelated; (b) the vector of weights $\boldsymbol{c}_1$ maximizes the univariate F-statistic for the linear combination z_1; (c) each successive vector of weights,

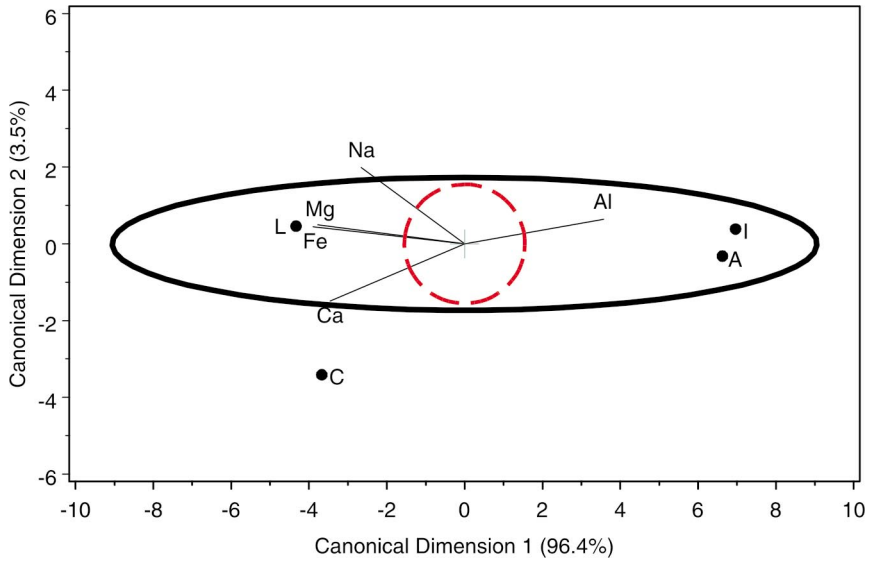

Figure 12. HE plot in 2D canonical discriminant space for the pottery data. Variable vectors indicate the correlations with the linear combinations that best discriminate among groups.

$\boldsymbol{c}_k, k = 2, \ldots, s$ maximizes the univariate F-statistic for z_k, subject to being uncorrelated with all other linear combinations.

The canonical weights, $\boldsymbol{c}_i$ are simply the eigenvectors of $\boldsymbol{H}\,\boldsymbol{E}^{-1}$ associated with the ordered eigenvalues, $\lambda_i, i = 1, \ldots, s$, and a MANOVA of all s linear combinations is statistically equivalent to that of the raw data. The λ_i are proportional to the fractions of between-group variation expressed by these linear combinations. Hence, to the extent that the first one or two eigenvalues are relatively large, a two-dimensional display will capture the bulk of between group differences. The canonical discriminant HE plot is then simply an HE plot of the scores $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ on the first two canonical dimensions.

Because the $\boldsymbol{z}$ scores are all uncorrelated, the $\boldsymbol{H}$ and $\boldsymbol{E}$ matrices will always have their axes aligned with the canonical dimensions; when, as here, the $\boldsymbol{z}$ scores are standardized, the $\boldsymbol{E}$ ellipse will be circular, assuming that the axes are equated so that a unit data length has the same physical length on both axes, as in Figure 12.

In this plot, the order and separation of the group means on each canonical variable indicates how that linear combination of the responses discriminates among the groups. As an additional aid to interpretation we also draw vectors on the plot indicating the correlation of each of the observed variables with the canonical dimensions. For a given response, a vector is drawn from the origin (representing the grand mean on the canonical variates) to a point proportional to the correlation (canonical structure coefficients) of that variable with each canonical variate, $(r_{y_i z_1}, r_{y_i z_2})$. With axes equated, the relative length of each variable vector will be proportional to its contribution to discriminating among the groups. As well, the angles between the variable vectors approximately indicate the correlations among group mean differences, based on the standardized $\boldsymbol{H}$ matrix projected into the space of the canonical dimensions.

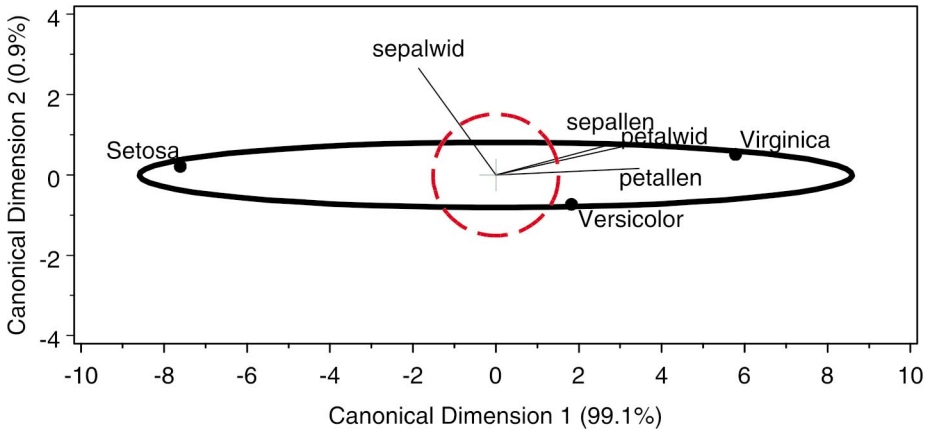

Figure 13. Canonical HE plot for the iris data.

Figure 13 presents the canonical discriminant HE plot for the iris data. With $g = 3$ groups, the two-dimensional plot is exact, but it is apparent that the structure of group means is essentially one-dimensional, with 99.1% of between-species variation accounted for by the large differences among the species on all variables except sepal width. These same effects may be observed in the scatterplot matrix (Figure 2, p. 5), but the canonical discriminant HE plot is a more compact summary. This plot is also similar to the biplot (Figure 3, p. 6), however the biplot displays the dimensions that account for *total* variation, whereas the dimensions in the canonical discriminant plot account for between-group variation.

This method can be extended to two-way (and higher) designs in two different ways. First, one may simply code the combinations of all factors *interactively*, so that $\boldsymbol{H}$ expresses all group differences, for example, $\boldsymbol{H} = \boldsymbol{H}_A + \boldsymbol{H}_B + \boldsymbol{H}_{AB}$ in a two-way design. Second, the method may be applied to *adjusted* response variate vectors, which are essentially the residuals (deviations from the means) of the model effects adjusted for. In the two-way setting, for example, the reduced-rank HE plot for the AB effect, $\boldsymbol{H}_{AB}$, is equivalent to the analysis of $\boldsymbol{y} \mid (A, B)$, that is, $\boldsymbol{y}_{ijk} - \bar{\boldsymbol{y}}_{A(i)} - \bar{\boldsymbol{y}}_{B(j)}$.

5. DISCUSSION

Formally, the MLM is a direct extension of the univariate LM so that all univariate response models (ANOVA, ANCOVA, MRA) and test statistics have straightforward generalizations to the multivariate response case. Paradoxically, the extensive collection of plots and graphical methods for univariate models do not, at first glance, translate easily to multivariate responses.

In particular, the methods we have presented are largely distinct from the kinds of plots one thinks of for univariate response models: plots of means ($\pm$ standard errors) for n-way ANOVA designs; response-surface plots for bivariate regression models; the generalization of these in "effect plots" (Fox 1987) for the generalized linear model. Diagnostic plots in the

univariate response setting include: (a) QQ plots of residuals; (b) spread-level plots to diagnose nonconstant variance; (c) influence-leverage plots to detect influential observations; (d) partial regression and added-variable plots to show conditional relations (controlling for other predictors) and diagnose nonlinear relations, for both the classical and generalized linear models; see Fox (1991, 1997); Friendly (1991) for descriptions and examples of these.

In the multivariate response setting, the graphical methods listed above can, of course, be usefully applied to each response individually, but information about the correlations among responses is thereby lost. The more general extension of these ideas to the MLM therefore provides interesting, and as yet unsolved, topics for future research.

In this article we have described and illustrated a variety of graphical displays for multivariate GLMs, designed to focus on the relationships between two sets of variables: predictors (regressors) and responses. Some of these methods are new (HE plots), some are old (biplots), and some have been extended here to a wider context (data ellipse). There are several general themes, statistical ideas, and graphical notions that connect the cases we have described here.

First, the data ellipse, as used here, provides a visual summary of bivariate relations, depicting means, variances, and covariances (or correlations), for either the classical, normal-theory estimators, or any robust estimator. These provide useful exploratory and confirmatory displays in a variety of multivariate contexts, can be used to show multiple-group MANOVA data, and can be embedded in a scatterplot matrix form to show all pairwise, bivariate relations.

Second, the idea of HE plots provides ways to visualize and understand the results of multivariate tests in both the MANOVA and MMRA contexts. Group means (for MANOVA) or 1-df H matrix vectors (for MMRA) can be overlayed on these plots to aid interpretation, and the pairwise relations for *all* responses can be seen in the HE plot matrix.

Third, we have used several dimension-reduction techiques (biplot, canonical discriminant analysis) to display two-dimensional summaries of the salient characteristics of multivariate data related to various aspects of the MLM. Overlaying variable vectors, data ellipses, and reduced-rank scores for observations, helps to make these plots more interpretable in relation to both the original data and the low-dimensional summary.

APPENDIX: SOFTWARE

All of the plots shown here were produced using macro programs for SAS software, many of which originated in Friendly (1991, 2000). These are described in a companion article (Friendly 2006) and are also available with documentation and examples at *http://www.math.yorku.ca/SCS/sasmac/*. The principal programs used here are:

`biplot`	Generalized biplot display of variables and observations
`canplot`	Canonical discriminant structure plots
`ellipses`	Plot bivariate data ellipses
`heplot`	Plot H and E matrices for a bivariate MANOVA effect

`hemat`	HE plots for all pairs of response variables
`hemreg`	Extract H and E matrices for multivariate regression
`scatmat`	Scatterplot matrices

For R users, the `heplots` package, available at *http://www.r-project.org/*, provides equivalent functionality and more for for `mlm` objects via the `Anova` and `linear.hypothesis` functions in the `car` package (Fox 2006).

ACKNOWLEDGMENTS

This work is supported by Grant 8150 from the Natural Sciences and Engineering Research Council of Canada. I am grateful to John Fox, Ernest Kwan, and Georges Monette for comments on earlier drafts. Several reviewers and especially the editor helped to strengthen and focus the article.

[Received February 2004. Revised October 2006.]

REFERENCES

Anderson, E. (1935), "The Irises of the Gaspé Peninsula," *Bulletin of the American Iris Society*, 35, 2–5.

Cohen, J. (1977), *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.) New York: Academic Press.

Dempster, A. P. (1969), *Elements of Continuous Multivariate Analysis*, Reading, MA: Addison-Wesley.

Fisher, R. A. (1936), "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, 8, 379–388.

Fox, J. (1987), "Effect Displays for Generalized Linear Models," in *Sociological Methodology, 1987*, ed. C. C. Clogg, San Francisco: Jossey-Bass, pp. 347–361.

——— (1991), *Regression Diagnostics: An Introduction*, Beverly Hills, CA: Sage Publications.

——— (1997), *Applied Regression Analysis, Linear Models, and Related Methods*, Thousand Oaks, CA: Sage Publications.

——— (2006), *car: Companion to Applied Regression*, R package version 1.2-0.

Friendly, M. (1991), *SAS System for Statistical Graphics* (1st ed.) Cary, NC: SAS Institute.

——— (2000), *Visualizing Categorical Data*, Cary, NC: SAS Institute.

——— (2006), "Data Ellipses, HE Plots and Reduced-Rank Displays for Multivariate Linear Models: SAS Software and Examples," *Journal of Statistical Software*, 17, 1–42.

Gabriel, K. R. (1971), "The Biplot Graphic Display of Matrices with Application to Principal Components Analysis," *Biometrics*, 58, 453–467.

——— (1981), "Biplot Display of Multivariate Matrices for Inspection of Data and Diagnosis," in *Interpreting Multivariate Data*, ed. V. Barnett, London: Wiley, chap. 8, pp. 147–173.

——— (1995), "MANOVA Biplots for Two-Way Contingency Tables," in *Recent Advances in Descriptive Multivariate Analysis*, ed. W. J. Kraznowski, Oxford, UK: Oxford University Press, chap. 10, pp. 227–268.

Galton, F. (1886), "Regression Towards Mediocrity in Hereditary Stature," *Journal of the Anthropological Institute*, 15, 246–263.

Gnanadesikan, R., and Kettenring, J. R. (1972), "Robust Estimates, Residuals, and Outlier Detection With Multiresponse Data," *Biometrics*, 28, 81–124.

Gower, J. C., and Hand, D. J. (1996), *Biplots*, London: Chapman & Hall.

Hand, D., Daly, F., Lunn, A. D., McConway, K. J., and Ostrowski, E. (1994), *A Handbook of Small Data Sets*, London: Chapman & Hall.

Monette, G. (1990), "Geometry of Multiple Regression and Interactive 3-D Graphics," in *Modern Methods of Data Analysis*, eds. J. Fox and S. Long, Beverly Hills, CA: Sage Publications, chap. 5, pp. 209–256.

Morrison, D. F. (1990), *Multivariate Statistical Methods* (3rd ed.), New York: McGraw-Hill.

Muller, K. E., LaVange, L. M., Ramey, S. L., and Ramey, C. T. (1992), "Power Calculations for General Linear Multivariate Models Including Repeated Measures Applications," *Journal of the American Statistical Association*, 87, 1209–1226.

Pearson, K. (1901), "On Lines and Planes of Closest Fit to Systems of Points in Space," *Philosophical Magazine*, 6, 559–572.

Rousseeuw, P., and Leroy, A. (1987), *Robust Regression and Outlier Detection*, New York: Wiley.

Rousseeuw, P., and Van Driessen, K. (1999), "A Fast Algorithm for the Minimum Covariance Determinant Estimator," *Technometrics*, 41, 212–223.

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

Stigler, S. M. (1986), *The History of Statistics: The Measurement of Uncertainty before 1900*, Cambridge, MA: Harvard University Press.

Timm, N. H. (1975), *Multivariate Analysis with Applications in Education and Psychology*, Belmont, CA: Wadsworth (Brooks/Cole).

Tubb, A., Parker, A., and Nickless, G. (1980), "The Analysis of Romano-British Pottery by Atomic Absorption Spectrophotometry," *Archaeometry*, 22, 153–171.