Graphical methods for multivariate linear models in psychological research: An ℝ tutorial

Michael Friendly[a, ✉] and Matthew Sigal[a]

[a]York University

Abstract ∎ This paper is designed as a tutorial to highlight some recent developments for visualizing the relationships among response and predictor variables in multivariate linear models (MLMs), and implemented in convenient packages for R. These models include multivariate multiple regression analysis (MMRA), multivariate analysis of variance (MANOVA) and multivariate analysis of covariance (MANCOVA). The methods we describe go well beyond what can be understood and explained from simple univariate graphical methods for the separate response variables. We describe extensions of these methods for the case of more than just a few response variables, where the important relationships can be readily seen in the low-dimensional (2D) space that accounts for most of the relevant information. As befits the tutorial nature of this paper, we analyze some sample psychological research studies utilizing these multivariate designs, showing examples in R. In the process, we also take up several practical problems related to the assumptions of MLMs, and how these can be dealt with using graphical methods. Finally, we provide guidelines to aid researchers in conducting multivariate research, pertaining to the analysis, visualization, and reporting of such designs. The graphical and statistical methods described here are all freely available and implemented in the R packages candisc, car, heplots, and mvinfluence.

Keywords ∎ Canonical discriminant plots, Hypothesis-Error (HE) plots, MANOVA, MANCOVA, MMRA, visual tests of significance. **Tools** ∎ R.

Introduction

Multivariate response designs are common in applied psychological research, and are utilized to analyze a wide range of phenomena within that domain. A particular research outcome (e.g., depression, neuro-cognitive functioning, academic achievement, self-concept, attention deficit hyperactivity disorders) might take on a multivariate form if it has several observed measurement scales or related aspects by which it is quantified, or if there are multiple theoretically distinct outcomes that should be assessed in conjunction with each other (e.g., using depression, generalized anxiety, and stress inventories to model overall happiness). In this situation, the primary concern of the researcher is to ascertain the impact of potential predictors on two or more response variables simultaneously.

For example, if academic achievement is measured for adolescents by their reading, mathematics, science, and history scores, do predictors such as parent encouragement, socioeconomic status and school environmental variables affect all of these outcomes? Do they affect them in the *same* or *different* ways? Similarly, if psychiatric patients in various diagnostic categories are measured on a battery of tests related to social skills and cognitive functioning, we might want to know which measures best discriminate among the diagnostic groups and which are most predictive of positive outcomes. Further, how are the *relationships* between the outcomes affected by the predictors? Such questions obviously concern more than just the separate univariate relations of each response to the pre-

---

[1]Structural equation modeling (SEM) offers another route to explore and analyze the relationships among multiple predictors and multiple re-

dictors. Equally, or perhaps more importantly, are questions of how the response variables are predicted *jointly*.[1]

Statistically, this is easy, because the classical univariate response model for ANOVA and regression, $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$, with $\boldsymbol{u} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$ generalizes directly to an analogous multivariate linear model (MLM), $\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{U}$ for multiple responses (as will be discussed in detail). Happily as well, hypothesis tests for the MLM are also straightforward generalizations of the familiar F and t-tests for univariate response models.

However, with two or more response variables, visualizations for multivariate models are not as simple as they are for their univariate counterparts for understanding the effects of predictors, model parameters, or model diagnostics. Consequently, the results of such studies are often explored and discussed solely in terms of coefficients and significance, and visualizations of the relationships are only provided for one response variable at a time, if at all. This tradition can mask important nuances, and lead researchers to draw erroneous conclusions.

The aim of this paper is to describe and illustrate some central methods that we have developed over the last ten years that aid in the understanding and communication of the results of multivariate linear models (Friendly, 2007; Friendly & Meyer, 2016; Friendly, Monette, & Fox, 2013). These methods rely on *data ellipsoids* as simple, minimally sufficient visualizations of variance that can be shown in 2D and 3D plots. As will be demonstrated, the Hypothesis-Error (HE) plot framework applies this idea to the results of multivariate tests of linear hypotheses.

Further, in the case where there are more than just a few outcome variables, the important nectar of their relationships to predictors can often be distilled in a multivariate juicer— a projection of the multivariate relationships to the predictors in the low-D space that captures most of the flavor. This idea can be applied using canonical correlation plots and with canonical discriminant HE plots.

Overall, the goal of this paper is to provide a substantive and computational tutorial on how these ideas can be implemented (in R software) and interpreted for popular designs: multivariate multiple regression analysis (MMRA), multivariate analysis of variance (MANOVA) and multivariate analysis of covariance (MANCOVA). Each are prevalent in psychological research applications, and we hope that this paper will aid researchers in properly interpreting and presenting their results.

We assume the reader to have at least a basic familiarity with R. While R fundamentals are outside the scope of the current paper, we believe that this language provides a rich set of resources, far beyond that offered by other statistical software packages, and is well worth learning. For those not familiar with R, we recommend Matloff (2011), Wickham (2014), and Cotton (2013) for introductions to programming in the language, and Fox and Weisberg (2011) and Teetor (2011) for learning about how to conduct basic statistical analyses.

Problems in understanding and communicating MLM results

In our consulting practice within the Statistical Consulting Service at York University, we see hundreds of clients each year ranging from advanced undergraduate thesis students, to graduate students and faculty from a variety of fields. Over the last two decades, and across each of these groups, we have noticed an increasing desire to utilize multivariate methods. As researchers are exposed to the utility and power of multivariate tests, they see them as an appealing alternative to running many univariate ANOVAs or multiple regressions for each response variable separately.

However, multivariate analyses are more complicated than such approaches, especially when it comes to understanding and communicating results. Output is typically voluminous, and researchers will often get lost in the numbers. While SPSS and SAS make tabular summary displays easy, these often obscure the findings that researchers are most interested in. The most common analytic oversights that we have observed are:

- **Atomistic data screening**: Researchers have mostly learned the assumptions (the Holy Trinity of normality, constant variance and independence) of univariate linear models, but then apply univariate tests (e.g., Shapiro-Wilk) and diagnostic plots (normal QQ plots) to every predictor and every response.

- **Bonferroni everywhere**: Faced with the task of reporting the results for multiple response measures and a collection of predictors for each, a common tendency is to run (and sometimes report) each of the separate univariate response models and then apply a correction for multiple testing.

- **Reverting to univariate visualizations**: To display results, SPSS and SAS make some visualization methods available through menu choices or syntax, but usually these are the wrong (or at least unhelpful) choices, in that they generate separate univariate graphs for the individual responses.

It is the purpose of this paper to discuss a few essential procedures for multivariate linear models, how their interpretation can be aided through the use of well-crafted

---

sponses. They have the advantage of being able to test potentially complex systems of linear equations in very flexible ways; however, these methods are often far removed from data analysis *per se* and except for path diagrams offer little in the way of visualization methods to aid in understanding and communicating the results. The graphical methods we describe here can also be useful in a SEM context.

(though novel) visualizations, and provide replicable sample code in R to showcase their use in applied psychological research. A closing section provides some practical guidelines for analyzing, visualizing and reporting such models to help avoid these and other problems.

Brief review of the MLM

The general multivariate linear model (MLM) can be understood as an extension of the univariate linear model, with the main difference being that there are multiple response variables instead of just one. In this context, there are multiple techniques that can be applied depending on the structure of the variables at hand. For instance, with one or more continuous predictors and multiple response variables, one could use multivariate regression to obtain estimates useful for prediction. Instead, if the predictors are categorical, multivariate analysis of variance (MANOVA) can be applied to test for differences between groups. Again, this is akin to multiple regression and ANOVA in the univariate context – the same underlying model is utilized by multiple statistical techniques.

In each of these cases, the underlying MLM is given most compactly using the matrix equation,

$$\underset{(n\times p)}{\boldsymbol{Y}} = \underset{(n\times q)}{\boldsymbol{X}}\ \underset{(q\times p)}{\boldsymbol{B}} + \underset{(n\times p)}{\boldsymbol{U}}, \tag{1}$$

where $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_p)$ is the matrix of n observations on p responses; $\boldsymbol{X}$ is the model matrix with columns for q regressors, which typically includes an initial column of 1s for the intercept; $\boldsymbol{B}$ is a matrix of regression coefficients, one column for each response variable; and $\boldsymbol{U}$ is a matrix of errors. The structure of the model matrix $\boldsymbol{X}$ is the same as the univariate linear model, and may contain, therefore, dummy regressors representing factors, polynomial or regression-spline terms, interaction regressors, and so on.[2]

The assumptions of the multivariate linear model entirely concern the behavior of the errors: Let $\boldsymbol{u}_i'$ represent the ith row of $\mathbf{U}$. Then it is assumed that $\boldsymbol{u}_i' \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a non-singular error-covariance matrix, constant across observations; $\boldsymbol{u}_i'$ and $\boldsymbol{u}_j'$ are independent for $i \neq j$; and $\mathbf{X}$ is fixed or independent of $\mathbf{U}$. These statements are simply the multivariate analogs of the assumptions of normality, constant variance and independence of the errors in univariate models.

In univariate response models, statistical hypothesis tests and model summaries (like R^2) are based on the familiar decomposition of the total sum of squares SS_T into regression or hypothesis (SS_H) and error (SS_E) sums of squares. In the multivariate linear model a similar decomposition is applied to the total *sum of squares and cross products* (*SSP*) matrix into regression and residual SSP matrices. For the MLM this is expressed as,

$$\underset{(p\times p)}{\mathbf{SSP}_T} = \boldsymbol{Y}'\boldsymbol{Y} - n\overline{\boldsymbol{y}}\,\overline{\boldsymbol{y}}' = \left(\widehat{\boldsymbol{Y}}'\widehat{\boldsymbol{Y}} - n\overline{\boldsymbol{y}}\,\overline{\boldsymbol{y}}'\right) + \widehat{\mathbf{U}}'\widehat{\mathbf{U}}$$

$$= \boldsymbol{SSP}_H + \boldsymbol{SSP}_E \equiv \boldsymbol{H} + \boldsymbol{E},$$

where $\overline{\boldsymbol{y}}$ is the $(p \times 1)$ vector of means for the response variables; $\widehat{\boldsymbol{Y}} = \boldsymbol{X}\widehat{\boldsymbol{B}}$ is the matrix of fitted values; and $\widehat{\mathbf{U}} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}$ is the matrix of residuals. This is the decomposition that we visualize below in HE plots, where the size and direction of $\boldsymbol{H}$ and $\boldsymbol{E}$ are represented as ellipsoids. An animated display of these ideas and the relations between data ellipses and HE plots can be seen at http://www.datavis.ca/gallery/animation/manova/.

The univariate F test statistic, $F = \frac{SS_H/df_h}{SS_E/df_e}$ assesses "how big" SS_H is relative to SS_E. This has a direct multivariate analog in terms of the $s = \min(p, df_h)$ non-zero latent roots (eigenvalues), λ_i, of $\boldsymbol{H}\boldsymbol{E}^{-1}$ ($\boldsymbol{H}$ "divided by" $\boldsymbol{E}$); that is, the values λ_i that solve

$$\det(\boldsymbol{H}\boldsymbol{E}^{-1} - \lambda\boldsymbol{I}) = 0 \implies \boldsymbol{H}\boldsymbol{E}^{-1}\lambda_i = \lambda_i\boldsymbol{v}_i. \tag{2}$$

The various multivariate test statistics such as Wilks' Λ, the Pillai and Hotelling trace criteria, and Roy's maximum root test are all functions of the λ_i. These reflect different geometric properties of the "size" of the $\boldsymbol{H}$ ellipsoid relative to the size of the $\boldsymbol{E}$ ellipsoid. The statistical and geometric details of this relationship are described in Friendly (2007) and Friendly et al. (2013).

The overall multivariate test for the model in Eqn. (1) is essentially a test of the hypothesis $\boldsymbol{B} = \boldsymbol{0}$ (excluding the intercept row), or equivalently a test based on the *incremental* $\boldsymbol{SSP}_H$ for the hypothesized terms in the model—that is, the difference between the $\boldsymbol{SSP}_H$ for the full model and the null (intercept only) model.

Testing contrasts and linear hypotheses

More generally, these tests apply to *every* linear hypothesis concerning the coefficients in $\boldsymbol{B}$. Suppose we want to test the hypothesis that a subset of rows (predictors) and/or columns (responses) simultaneously have null effects. This can be expressed in the general linear test, $H_0 : \boldsymbol{C}_{h\times q}\,\boldsymbol{B}_{q\times p} = \boldsymbol{0}_{h\times p}$, where $\boldsymbol{C}$ is a full rank $h \leq q$ hypothesis matrix of constants, that selects subsets or linear combinations (contrasts) of the coefficients in $\boldsymbol{B}$ to be tested in a h degree-of-freedom hypothesis. In this case, the SSP matrix for the hypothesis has the form

---

[2]For brevity, we assume that $\boldsymbol{X}$ is of full column-rank q; allowing for less than full rank cases would introduce additional notation, but not fundamentally change any of the results presented here.

$$H = (C\widehat{B})^{\mathsf{T}} [C(X^{\mathsf{T}}X)^{-1}C^{\mathsf{T}}]^{-1} (C\widehat{B}) \qquad (3)$$

where there are $s = \min(h, p)$ non-zero eigenvalues of HE^{-1}. In Eqn. (3) H measures the (Mahalanobis) squared distances (and cross products) among the linear combinations $C\widehat{B}$ from the origin under the null hypothesis.

Repeated Measure Designs

Finally, these ideas extend directly to *repeated measure designs*, where the multivariate responses in Y arise when a given variable is assessed for the same individual on different occasions or under different circumstances, or even a collection of different outcome variables (e.g, response speed, accuracy) measured at different occasions.[3]

In this case, any linear hypothesis to be tested can be expressed as $H_0 : C_{h \times q} B_{q \times p} M_{p \times v} = 0_{h \times v}$ where M is a suitable "within-subjects" response transformation matrix expressing contrasts among the repeated measures. See, e.g., O'Brien and Kaiser (1985), Timm (1980), Hand and Taylor (1987) for details on the multivariate approach to repeated measure designs; Friendly (2010) gives details and examples of the application of the HE plot framework for these models.

Example: Father perceived parenting data

We use a simple example of a three-group multivariate design in this section and the next to illustrate the basic ideas of fitting MLMs in R and the graphical methods we propose for understanding the results of such models.

The data come from an exercise (10B) in Meyers, Gamst, and Guarino (2006) and are probably contrived, but modeled on a real study in which fathers were assessed on three subscales of a *Perceived Parenting Competence Scale*: (a) caring, or the degree of caretaking responsibilities corresponding to the father; (b) emotion, the emotional support provided by the father to the child; and (c) play, the recreational time the father spent with the child. Each of the above variables are discrete, and range from 0 to 10. The dataset Parenting comprises 60 fathers selected from three groups: (a) fathers of a child with no disabilities; (b) fathers with a physically disabled child; (c) fathers with a mentally disabled child. The design can appropriately be modeled using MANOVA, with three levels for the group factor and three response variables. The main questions concern whether group means differ on these scales, and what the nature of these differences are.

First, we illustrate fitting a MANOVA model to this data. In R, a MLM is fit with the standard $\mathrm{lm}()$ function simply by using a matrix of variables as the response (left-hand) side of the model formula. In the call to $\mathrm{lm}()$, $\mathrm{cbind}()$ combines the response vectors into a matrix object. For testing model effects, we recommend use of the $\mathrm{Anova}()$ function from the car packge (Fox & Weisberg, 2011, 2015), which provides *partial* (Type II) tests of model terms for the null hypothesis that the three response means are all equal across the three groups (with only one term, the standard $\mathrm{anova}()$ function would return equivalent results).

```
data(Parenting, package="heplots")
library(car)
parenting.mod <- lm(
    cbind(caring, play, emotion) ~ group,
    data=Parenting
)
Anova(parenting.mod)
```

The results would be:

```
## Type II MANOVA Tests: Pillai test statistic
##       Df test stat approx F num Df den Df    Pr(>F)
## group 2   0.94836   16.833      6    112 8.994e-14 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
      0.1 ' ' 1
```

By default, $\mathrm{Anova}()$ reports the Pillai trace statistic; all test statistics and the H and E matrices are provided by $\mathrm{summary}(\mathrm{Anova}(\mathrm{parenting.mod}))$ and any tests of linear hypotheses or contrasts can be obtained using $\mathrm{linearHypothesis}(\mathrm{parenting.mod})$ from car.

The MANOVA test above tells us that the means of the groups differ significantly on the three scales collectively, but it doesn't indicate *how* they differ. In a univariate approach to visualization, we could make *separate* boxplots for each of the response variables (shown in the supplementary materials), but this would not show how they are *related*. The following section details how one can generate and interpret informative graphics for this model and will then be demonstrated using this example for the Parenting data.

Graphical Methods for MLMs

In this section, we describe three simple, yet fundamental ideas behind our approach to visualizing data in relation to MLMs. These methods are explained in more detail in Friendly and Sigal (2014) and Fox, Friendly, and Monette (2009).

(a) For any multivariate normal data, the graphical analog of the minimally sufficient statistics (μ, Σ) (mean vector and covariance matrix) is a *data ellipsoid* centered at μ whose size and shape is determined by Σ, and which can be viewed in 2D, 3D, and by other means.

(b) From the theory of tests for the MLM outlined above,

---

[3]The main practical limitation that using this MLM approach for repeated measures has is that it requires complete data. In all software environments, any missing data for a given subject causes that case to be deleted from the analysis. A mixed model approach avoids this.

we can show *any* hypothesis test by superposing an $\boldsymbol{H}$ ellipsoid representing variation against a null hypothesis on an $\boldsymbol{E}$, representing error variation. The HE-plot method we describe provides a *visual* test of significance by one common test statistic (Roy's maximum root), signaled by the $\boldsymbol{H}$ ellipsoid projecting outside that of $\boldsymbol{E}$ as illustrated below.

(c) When there are more than a few response variables, these same relationships can be more readily seen via projection onto a smaller space that gives a maximally efficient 2D (or 3D) representation. This often provides additional insight into the results of a complex multivariate analysis.

Detailed examples of the application of these ideas to a selection of psychological research topics are provided in the Section "Examples" and guidelines for their use is explored in the Discussion, but first more information is required to understand what is being plotted in order to properly interpret their properties.

Data ellipsoids: Minimally sufficient visual summaries

Under classical theory that a p-dimensional variate $\boldsymbol{y}$ is multivariate normally distributed, $\boldsymbol{y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the sample mean vector $\bar{\boldsymbol{y}}$ and covariance matrix $\boldsymbol{S}$ are *sufficient statistics*, in the sense that they provide all the information necessary for making inferences and predictions. Geometrically, contours of constant density correspond to concentric ellipsoids located at the centroid – this fact was first discovered by Francis Galton (1886), and exemplified in the case of his bivariate plot of heights of parents and their offspring (Friendly & Denis, 2005; Hanley, 2004).

Similarly, in applied data analysis, the *data ellipse* (or *concentration ellipse*, Dempster (1969), Monette (1990) provides a remarkably simple and effective display for viewing and understanding bivariate relationships in multivariate data and generalizes to ellipsoids in p dimensions. (In what follows, we use the term *ellipsoid* for the general case and *ellipse* for the bivariate case, but this should be clear from the context.)

For example, all properties of bivariate regression (means, standard deviations, correlation, and slopes of both regression lines, residual standard error, etc.) can be immediately seen in terms of properties of the data ellipse (see Figure 1). See Friendly et al. (2013) for a complete discussion of the role of ellipsoids in statistical data visualization. Moreover, because a data ellipsoid is an exact geometrical representation of multivariate location ($\boldsymbol{\mu}$) and scatter ($\boldsymbol{\Sigma}$), we call this a *sufficient visual summary*.

Formally, for a p-dimensional sample, $\boldsymbol{Y}_{n \times p}$, with mean vector $\bar{\boldsymbol{y}}$, the $p \times p$ covariance matrix $\boldsymbol{S}$ can be represented by the p-dimensional data ellipsoid, $\mathcal{E}_c$ of size ("radius") c translated to $\bar{\boldsymbol{y}}$. This is defined as the set of all points $\boldsymbol{y}$ satisfying

$$\mathcal{E}_c(\bar{\boldsymbol{y}}, \boldsymbol{S}) := \{\boldsymbol{y} : (\boldsymbol{y} - \bar{\boldsymbol{y}})^\mathsf{T} \, \boldsymbol{S}^{-1} \, (\boldsymbol{y} - \bar{\boldsymbol{y}}) \leq c^2\}, \quad (4)$$

It is readily seen that the quadratic form in Eqn. (4) corresponds to the set of points whose squared Mahalanobis distances $D_M^2(\boldsymbol{y}) = (\boldsymbol{y} - \bar{\boldsymbol{y}})^\mathsf{T} \, \boldsymbol{S}^{-1} \, (\boldsymbol{y} - \bar{\boldsymbol{y}})$, from the centroid of the sample, $\bar{\boldsymbol{y}} = (\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_p)^\mathsf{T}$, are less than or equal to c^2.

When the data variables are multivariate normal, the data ellipsoid approximates a contour of constant density in their joint distribution. In this case $D_M^2(\boldsymbol{y})$ has a large-sample χ_p^2 distribution, or, in finite samples, approximately $[p(n-1)/(n-p)]F_{p,n-p}$. Hence, in the bivariate case, taking $c^2 = \chi_2^2(0.95) = 5.99 \approx 6$ encloses approximately 95% of the data points under normal theory. A 68% coverage data ellipse with $c^2 = \chi_2^2(0.68) = 2.28$ gives a bivariate analog of the standard $\bar{x} \pm 1s_x$ and $\bar{y} \pm 1s_y$ intervals.[4]

Data ellipses are illustrated in Figure 1, using a bivariate normal sample of $n = 300$ observations with $(\mu_X, \mu_Y) = (0, 0)$, $(\sigma_X, \sigma_Y) = (10, 10)$ and $\rho = .50$. In this figure, the ellipses have the sample mean vector $(\bar{x}, \bar{y})$ as their center; the lengths of arms of the central cross (dashed) show the standard deviation of each variable, which may be seen to correspond to the shadows of the 40% ellipse. In what follows, we typically use a 68% ellipse to show a standard bivariate "unit" region.

### HE plots: Visualizing sources of significance in MLMs

With the mathematical properties of data ellipsoids described above, a 2D HE plot for two response variables (see: Figure 2) simply overlays two types of data ellipses for these variables; a 3D HE plot does the same for three response variables. The first is a standard 68% data ellipse showing residual (error) variation reflected in the $\boldsymbol{E}$ matrix of the residuals in $\boldsymbol{U}$. This is scaled as $\boldsymbol{E}/df_e$ so that its size is analogous to a mean square error in univariate tests. This 68% scaling allows researchers to "read" the residual standard deviation as the half-length of the shadow of the E ellipse on any axis. Translating the E ellipse to the overall (grand) means of the variables plotted also allows us to show the means for factor levels on the same scale, facilitating interpretation.

The second is an ellipse representing the variation in the means of a factor (or any other term reflected in

---

[4]Of course no real data is ever exactly multivariate normal, but when the data differs from this theory by only moderate skewness and kurtosis and/or the possible presence of even many outliers, the classical mean and product-moment estimate, $\widehat{\boldsymbol{S}} = \sum (\boldsymbol{y} - \bar{\boldsymbol{y}})(\boldsymbol{y} - \bar{\boldsymbol{y}})^\mathsf{T}/(n-1)$ can be readily replaced by robust estimates (Rousseeuw & Leroy, 1987) of location and scatter. Popular choices are provided for example by the minimum covariance determinant (MCD) algorithm (Rousseeuw & Van Driessen, 1999) or minimum volume ellipsoid (MVE) method.

Figure 1 ■ Plot of artificial data showing properties of data ellipses. The ellipses have theoretical coverage of 40% (blue), 68% (red) and 95% (gray) of the data points. The regression line of y on x (black) goes through the points where the ellipses have vertical tangent lines. The ratio of the vertical distance from the mean of y to the regression line (labeled 'r') to the vertical to the top of the ellipse gives a visual estimate of the sample correlation ($r = 0.51$, here).
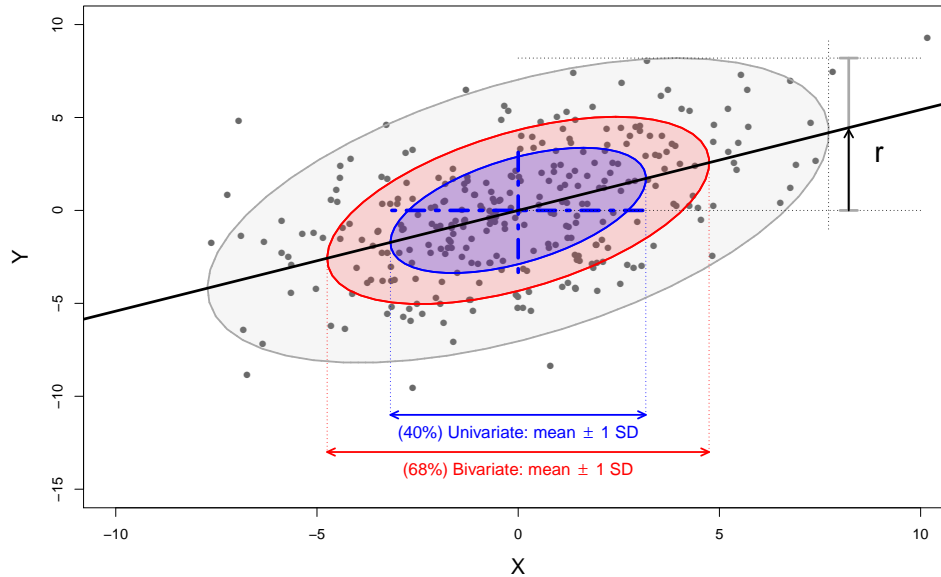
Eqn. (3) in the $\boldsymbol{H}$ matrix) which is simply the data ellipse of the *fitted values* for that term. Scaling the $\boldsymbol{H}$ matrix as $\boldsymbol{H}/df_e$ puts this on the same scale as the $\boldsymbol{E}$ ellipse. We refer to this as *effect size scaling*, because it is similar to an effect size index used in univariate models, e.g., $ES = (\bar{x}_2 - \bar{x}_2)/s$ in a two-group, univariate design.

For the Parenting data, and the two responses caring and play, effect size scaling is shown in the left panel of Figure 2, indicating that the variation of means of the groups is about the same size as within-group variance. Means for the groups indicate that fathers of normal children have higher scores on play, while fathers of physically disabled children are highest on caring.

Significance scaling

The geometry of ellipsoids and multivariate tests allow us to go further, to provide *visual tests of significance* for terms in a MLM. This is achieved simply by dividing $\boldsymbol{H}/df_e$ further by the α-critical value of the corresponding test statistic. Among the various multivariate test statistics, Roy's maximum root test gives $\boldsymbol{H}/(\lambda_\alpha df_e)$ has the attractive visual property that the scaled $\boldsymbol{H}$ ellipsoid will protrude *somewhere* outside the standard $\boldsymbol{E}$ ellipsoid if and only if Roy's test is significant at significance level α. For these data, the HE plot using significance scaling is shown

in the right panel of Figure 2. Note that although we use a conventional $\alpha = 0.05$ criterion in these plots, the emphasis is *not* on simplistic p-values for significance testing, but rather on visual assessment of the strength of evidence against a null hypothesis and how this relates to the response variables shown.

Example: Father perceived parenting data

In our R implementation, the HE plots shown in this figure are produced directly from the "mlm" object, parenting.mod generated earlier. This uses heplot() from the heplots package (Fox & Friendly, 2014), which can be installed and loaded as follows:

```
install.packages("heplots", dependencies
    =TRUE)
library("heplots")
```

When heplot() is given the "mlm" object, it produces a bivariate HE plot for any pair of response variables in the model (the first two, by default). At their core, the two plots in Figure 2 are produced by these calls to heplot():

```
# effect scaling
heplot(parenting.mod, size="effect")
# significance scaling is the default,
# so no size argument is required
```
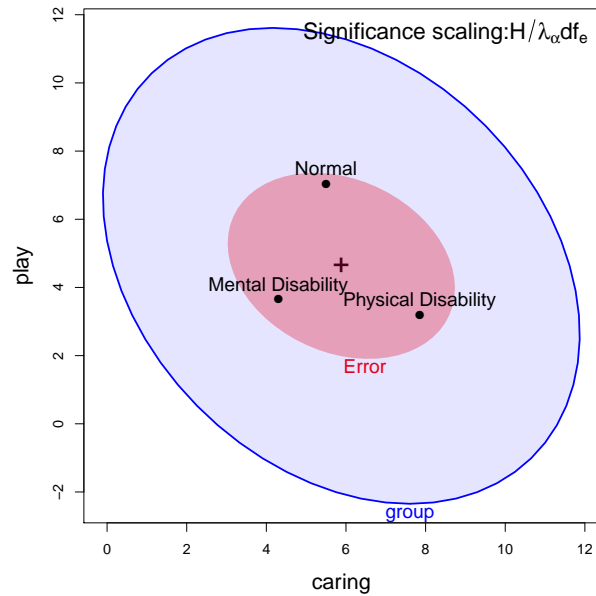
Figure 2 ∎ HE plots for the variables caring and play in the Parenting data set. (a) Effect scaling of the $\boldsymbol{H}$ matrix. (b) Significance scaling of the $\boldsymbol{H}$ matrix, $\alpha = 0.05$.

(a)
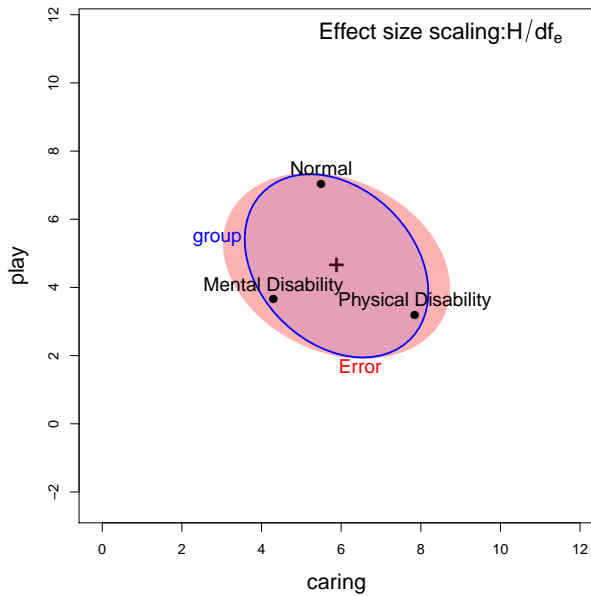
(b)


```
heplot(parenting.mod)
```

We can interpret these plots as follows: for caring and play, the within-group residuals ($\boldsymbol{E}$) are slightly negatively correlated: within a given group, fathers who score higher on caring score slightly lower on play. The variation among the group means ($\boldsymbol{H}$) is in the same direction: The group (Physical Disability) scoring highest on caring scores lowest on play. In general, for any HE plot, the direction of the $\boldsymbol{H}$ ellipsoid relative to that of $\boldsymbol{E}$ reflects the linear combinations of the response variables shown with the greatest effect.

In this example, the Parenting data has $p = 3$ dimensions, and the residual $\boldsymbol{E}$ ellipsoid is 3D. However, with three groups, $df_h = 2$, so the $\boldsymbol{H}$ ellipsoid has only $s = \min(p, df_h) = 2$ non-zero dimensions. To see these relations for all variables together, it is easy to use a scatterplot matrix format, using the pairs() function on a "mlm" object, or in 3D using heplot3d(). The 3D plots are interactive, in that they can be rotated and zoomed by mouse control, and dynamic, in that they can be made to spin and saved as movies. To save space, these plots are not shown here.

```
# all pairwise 2D projections
pairs(parenting.mod)
# interactive 3D plot
heplot3d(parenting.mod)
```

Visualizing contrasts and linear hypotheses

In a MANOVA, it is often desired to follow a significant overall test for a factor by tests of contrasts (e.g., $\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2 = 0$) or other linear hypotheses ($\boldsymbol{\beta}_1 = 0$) As noted earlier, such tests are just examples of the general linear test, $H_0 : \boldsymbol{C}\,\boldsymbol{B} = \boldsymbol{0}$ for a specified matrix of constants $\boldsymbol{C}$ giving linear combinations of the parameters to be tested.

As an important special case, for an overall effect with df_h degrees of freedom (and balanced sample sizes), a set of df_h pairwise orthogonal $(1 \times q)$ $\boldsymbol{C}$ matrices ($\boldsymbol{C}_i^\mathsf{T}\boldsymbol{C}_j = 0$) gives rise to a set of df_h rank 1 $\boldsymbol{H}_i$ matrices that additively decompose the overall hypothesis SSCP matrix,

$$\boldsymbol{H} = \boldsymbol{H}_1 + \boldsymbol{H}_2 + \cdots + \boldsymbol{H}_{df_h},$$

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 degenerate ellipse— a *line* in an HE plot. Their collection provides a visual summary of the overall test, as partitioned by these orthogonal contrasts, but more importantly, an "explanation" for the overall test in terms of answers to df_h independent questions.

To illustrate, we can set up and test two orthogonal contrasts for the Parenting data to answer the following questions: (a) Do fathers of children with no disabilities differ from the average of those with either physical or mental disabilities? (b) Is there a difference between fathers of

physically disabled children and those with a mental disability? In R, this is done by specifying two contrast vectors as the columns of a matrix assigned to group with the $\mathrm{contrasts}()$ function.

```
contrasts(Parenting$group) <- matrix(
  c(2, -1, -1,
    0,  1, -1), nrow=3, ncol=2)
contrasts(Parenting$group)
```

```
##                     [,1] [,2]
## Normal                2    0
## Physical Disability  -1    1
## Mental Disability    -1   -1
```

Then, after fitting the model $\mathrm{parenting.mod}$ with $\mathrm{lm}()$, the $\mathrm{linearHypothesis}()$ function from the car package can be used to test them individually, where the set of contrasts are labelled sequentially, e.g., $\mathrm{group1}$, $\mathrm{group2}$:

```
print(linearHypothesis(parenting.mod,
      "group1"), SSP=FALSE)
```

```
## Multivariate Tests:
##      Df test stat approx F numDf denDf    Pr(>F)
## Pillai 1 0.5210364 19.94376    3    55 7.1051e-09 ***
## Wilks  1 0.4789636 19.94376    3    55 7.1051e-09 ***
## H-L    1 1.0878413 19.94376    3    55 7.1051e-09 ***
## Roy    1 1.0878413 19.94376    3    55 7.1051e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
    0.1 ' ' 1
```

in which H-L denotes the Hotteling-Lawley test. For a one degree-of-freedom hypotheses like this, all test statistics will produce the same F values and p-values.

Similarly, $\mathrm{heplot}()$ allows any such set of linear hypotheses to be visualized via the $\mathrm{hypotheses}$ argument. The following call (simplified here) gives Figure 3, in which the two linear hypotheses appear as lines overlaid on the hypothesis and error ellipsoids:

```
hyp <- list("N:MP" = "group1",
            "M:P"  = "group2")
heplot(parenting.mod, hypotheses = hyp)
```

This HE plot shows that, for these two variables, the greatest between-group difference is the contrast of the Normal group versus the others (labeled "N:MP"), and that this contrast largely reflects differences on the play scale. The second contrast ("M:P") appears mainly in terms of caring: fathers of physically disabled engage in more care-taking than those with mentally disability. In general, the directions of these hypotheses in a given plot show *how* the group means differ in terms of a given contrast.[5]

---

[5]That the $\boldsymbol{H}$ ellipses for the contrasts subtend that for the overall test of group is no accident. In fact this is true in p-dimensional space for any linear hypothesis, and orthogonal contrasts have the additional geometric property that they form conjugate axes for the overall $\boldsymbol{H}$ ellipsoid relative to the $\boldsymbol{E}$ ellipsoid (Friendly et al., 2013).

Low-D views: Generalized canonical discriminant HE plots

The HE plot framework so far provides views of all the effects in a MLM in *variable* space, that we can see in 2D (3D) for selected pairs (triples) of response variables, or for all pairwise views in scatterplot matrix format.

However in a one-way MANOVA design with more than three response variables, it is difficult to visualize how the groups vary on all responses together, and how the different variables contribute to discrimination among groups. In this situation, canonical discriminant analysis (CDA) is often used, to provide a low-D visualization of between-group variation, analogous to the biplot technique for purely quantitative variables.

The key to this is the eigenvalue decomposition, $\boldsymbol{H}\boldsymbol{E}^{-1}\lambda_i = \lambda_i \boldsymbol{v}_i$ (Eqn. (2)) of $\boldsymbol{H}$ relative to $\boldsymbol{E}$. The eigenvalues, λ_i, give the "size" of each s orthogonal dimensions on which the multivariate tests are based, but also, the corresponding eigenvectors, $\boldsymbol{v}_i$, give the weights for the response variables in s linear combinations that maximally discriminate among the groups.

Thus, CDA amounts to a transformation of the p responses, $\boldsymbol{Y}_{n \times p}$ into the canonical space, $\boldsymbol{Z}_{n \times s} = \boldsymbol{Y}\boldsymbol{E}^{-1/2}\boldsymbol{V}$, where $\boldsymbol{V}$ contains the eigenvectors of $\boldsymbol{H}\boldsymbol{E}^{-1}$ and $s = \min(p, df_h)$. It is well-known (e.g., Gittins, 1985) that *canonical discriminant plots* of the first two (or three, in 3D) columns of $\boldsymbol{Z}$ corresponding to the largest canonical correlations provide an optimal low-D display of the variation between groups relative to variation within groups.

For a one-way design, the *canonical HE plot* is simply the HE plot of the canonical scores in the analogous MLM model that substitutes $\boldsymbol{Z}$ for $\boldsymbol{Y}$. This is shown in Figure 4 for the Parenting data. The interpretation of this plot is the same as before: if the hypothesis ellipse extends beyond the error ellipse, then that dimension is significant. Vectors for each predictor are then superimposed and show the relation between each and the two canonical dimensions; these are the structure coefficients (correlations) of the responses in $\boldsymbol{Y}$ with the canonical scores in $\boldsymbol{Z}$, scaled to conveniently fill the plot window. The relative lengths of the variable vectors is proportional to their contributions to group mean differences.

The interpretation of this plot is simple: in canonical space, variation of the means for the groups is entirely two-dimensional. Dimension 1 (60% of the group effect) is largely attributable to the play scale. The remaining 40% is attributable to the differences on caring, which separates the two disability groups. The $\mathrm{emotion}$ scale contributes a little, but in the opposite direction. In this example, there

Figure 3 ■ HE plot for two variables in the Parenting data set, showing the (degenerate) H ellipses corresponding to two orthogonal contrasts among the groups.
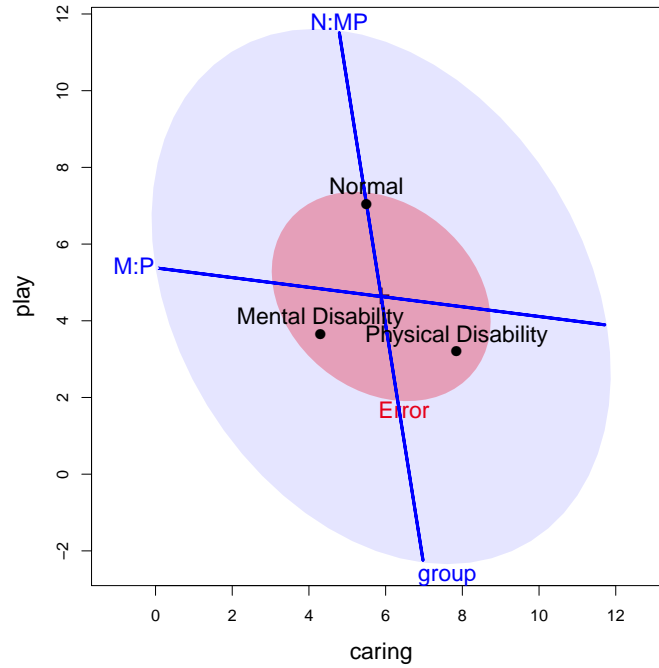
is little gained from viewing the model in canonical space. The advantage over views in variable space comes when there are more response variables, as we shall see in the "Examples" section.

This and related methods for canonical correlation are implemented in the R package candisc (Friendly & Fox, 2013). candisc() calculates the canonical analysis for one term in a MLM, returning a "candisc" object or for multiple terms giving a "candiscList". The heplot.candisc() and heplot3d.candisc() methods in that package give 2D and 3D plots in canonical space. The essence of Figure 4 is produced by these calls:

```
library(candisc)
parenting.can <- candisc(parenting.mod)
heplot(parenting.can)
```

Examples

This section provides a set of examples applying these graphical methods to a prototypical selection of research problems employing MMRA, MANOVA and MANCOVA designs. The main goal is to illustrate how these methods can be useful in helping the researcher to understand the substantive interpretation provided beyond what can be seen in tables of statistical tests and model coefficients.

All examples have been run directly with R software (R version 3.2.5) using the knitr (Xie, 2015) package. To save space, we often omit most of the R details of code used to generate analyses and figures, sometimes replacing arguments that just make graphs prettier by "...". Some sample annotated code is collected in the Appendix and the full scripts pertaining to all analyses and figures are available in online supplements to this article (https://mattsigal.github.io/graph-MLM/).

Multivariate multiple regression

Rohwer data

Imagine the following scenario: You are a methodology specialist in your department, and a colleague comes to you with the following problem: She has done a study assessing the relationship between a set of "ability" measures (a scholastic aptitude test (SAT), the Peabody Picture Vocabulary Test (PPVT), and the Raven progressive matrices test) in kindergarden children and a collection of memory tasks (measures of paired associate (PA) learning, under different conditions) she had them perform. She ran a set of separate multiple regressions for each of the responses, and was disappointed to find only very weak evidence: some of the ability measures had only small R^2, and in these,

Figure 4 ■ A canonical HE plot, visualizing the two canonical dimensions of the Parenting dataset, with α = .05. The contributions of the response variables to differences in group means is shown by the direction and relative length of the variable vectors.
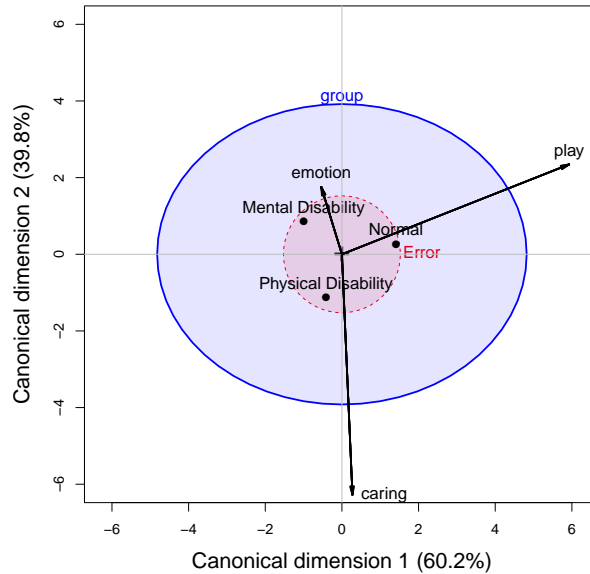
only a few of the PA measures showed individual significance of their regression coefficients. Can you help rescue her?

This scenario is all too common in our consulting work, and the context we described comes from a classical teaching example of a MMRA research study by William Rohwer given in Timm (1975), examples 4.3, 4.7, and 4.23). The complete data for this study pertains to 69 children in two groups (schools) stratified by socioecomomic status (SES) and contained in the data set Rohwer in the heplots package. In this example we focus on the analysis of the relationships between the ability and PA measures in the high SES group ($n = 32$), corresponding to Timm's Example 4.7. We return to this data set in the example on MANCOVA, where the SES variable is also analyzed.

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), but the details of these conditions are unknown. The scores are the total number of items correct on two presentations of 20 items, and thus have a potential range of 0–40.

The separate univariate multiple regressions can be run as follows:

```
data("Rohwer", package="heplots")
Rohwer2 <- subset(Rohwer,
  subset = group==2 )
```

```
# subset for high SES
rownames(Rohwer2)<- 1:nrow(Rohwer2)

rohwer.mod1 <- lm ( SAT   ~ n + s + ns +
    na + ss, data = Rohwer2)
rohwer.mod2 <- lm ( PPVT  ~ n + s + ns +
    na + ss, data = Rohwer2)
rohwer.mod3 <- lm ( Raven ~ n + s + ns +
    na + ss, data = Rohwer2)
```

The results of these models are conveniently summarized in a publication-quality table (Table 1) using the stargazer package:

```
stargazer(rohwer.mod1, rohwer.mod2,
    rohwer.mod3)
```

However, they are indeed disappointing. Only the model for SAT is strongly significant overall and only two predictors, ns and na, have coefficients significantly different from 0. However, the multivariate regression model tells a quite different, and more encouraging story. Here, three of the five predictors are significant by the multivariate test on all three response.

```
rohwer.mlm <- lm(cbind(SAT, PPVT, Raven)
    ~ n + s + ns + na + ss, data=Rohwer2
    )
```

Table 1 ■ Univariate regression models for Rohwer data.

	Dependent variable:		
	SAT	PPVT	Raven
n	3.26*	0.07	0.06
	(1.30)	(0.62)	(0.13)
s	3.00	0.37	0.49**
	(1.50)	(0.72)	(0.15)
ns	−5.86***	−0.37	−0.16
	(1.54)	(0.74)	(0.16)
na	5.67***	1.52*	0.12
	(1.34)	(0.64)	(0.14)
ss	−0.62	0.41	−0.12
	(1.14)	(0.54)	(0.12)
Observations	32	32	32
R^2	0.56	0.35	0.31
Adjusted R^2	0.47	0.23	0.18
Residual Std. Error (df = 26)	25.67	12.25	2.61
F Statistic (df = 5; 26)	6.54***	2.85*	2.32

Note. *p<0.05; **p<0.01; ***p<0.001

```
Anova(rohwer.mlm)
```

```
## Type II MANOVA Tests: Pillai test statistic
##    Df test stat approx F num Df den Df  Pr(>F)
## n   1   0.20177   2.0222      3     24 0.13760
## s   1   0.30970   3.5891      3     24 0.02836 *
## ns  1   0.35795   4.4601      3     24 0.01260 *
## na  1   0.46516   6.9576      3     24 0.00157 **
## ss  1   0.08873   0.7790      3     24 0.51725
## ---
## Signif.codes:'***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
    ' 1
```

Moreover, we can also test the overall null hypothesis for the entire MLM, $H_0 : \boldsymbol{B} = \boldsymbol{0}$ and this test very strongly rejects H_0 by all test criteria.

```
linearHypothesis(rohwer.mlm,
         c("n", "s", "ns", "na", "ss"))
```

```
## Multivariate Tests:
##        Df test stat approx F num Df den Df   Pr(>F)
## Pillai  5   1.0386     2.753     15  78.00 0.001912 **
## Wilks   5   0.2431     2.974     15  66.65 0.001154 **
## H-L     5   2.0615     3.115     15  68.00 0.000697 ***
## Roy     5   1.4654     7.620      5  26.00 0.000160 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
    0.1 ' ' 1
```

So, how do we help the client understand these results? HE plots to the rescue! Figure 5 shows pairwise views of all of the multivariate tests carried out above, including the overall test of $H_0 : \boldsymbol{B} = \boldsymbol{0}$ labeled "Regr". $\boldsymbol{E}$ has rank 3, so the error ellipsoid is 3D. For the multivariate tests of each predictor, each $\boldsymbol{H}_i$ has only one non-zero latent root, so these plot as degenerate lines. With the multivariate regression model rohwer.mlm, this plot is produced by our

pairs() function for "mlm" objects as follows:

```
hyp <- list("Regr" = c("n", "s", "ns", "
    na", "ss"))
pairs(rohwer.mlm, hypotheses = hyp, ...)
```

To interpret Figure 5, recall that: (a) Any hypothesis ellipsoid that protrudes *anywhere* outside the $\boldsymbol{E}$ ellipsoid is a significant effect (at $\alpha = 0.05$) by Roy's test. (b) The *length* of each predictor line indicates the strength of its relationship to the two responses jointly. (c) The *orientation* of each predictor line shows its relationship to the two response variables in a given panel, and contributes to making the $\boldsymbol{H}$ ellipsoid for the overall test "fat" in its direction.

Thus, in the plot for (PPVT, SAT) in the second row, first column (2, 1) panel, only na and ns appear individually significant. However, the s and n tests are also positively associated with SAT, and contribute something to the Regr test. The significant effect of the s test is only seen in the panels involving the Raven, shown in the (3, 3) panel. Overall, the predicted values for all three responses are positively correlated and because of this the separate predictors each contribute power to the multivariate test.

To complete this example, we note that for multivariate multiple regression, canonical correlation analysis provides similar dimension reduction strategy to that described earlier for canonical discriminant HE plots in MANOVA designs. With $p = 3$ responses and $q = 5$ predictors for this problem, the canonical correlation analysis of $[\boldsymbol{Y}, \boldsymbol{X}]$ results in $s = \min(p, q) = 3$ non-zero eigenvalues (canonical correlations), and the associated eigenvec-

Figure 5 ■ HE scatterplot matrix for the Rohwer data, high SES group. Each panel shows one bivariate view of the 3D H and E matrices. The multivariate tests for the predictor variables each have 1 df, and so appear as lines. The numbers in the diagonal panels give the range of each variable.
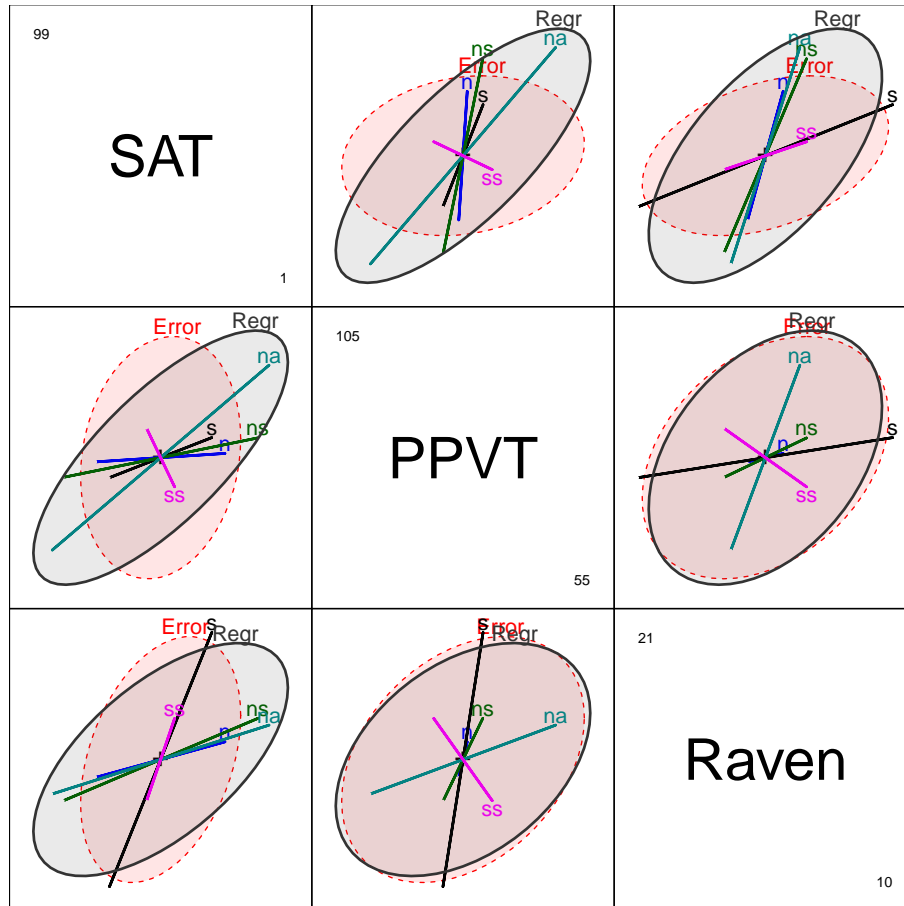
tors give the linear combinations of the Y variables most predictable from linear combinations of the X variables (and vice-versa). The candisc package contains a $\mathrm{cancor}()$ function for the analysis and $\mathrm{plot}()$ and $\mathrm{heplot}()$ methods for visualizing these results in the HE plot framework.

MANOVA designs

Multivariate analysis of variance is of course just a special case of the MLM (Eqn. (1)), where the X matrix is composed entirely of columns (dummy variables) that represent *discrete factors* and their possible interactions. However, they differ from regression designs in terms of useful and applicable data visualization methods, particularly for data exploration. In R, this is mostly masked by object-oriented methods; for example, $\mathrm{plot}(\mathrm{y} \sim \mathrm{x})$ gives a scatterplot for a quantitative x, but a boxplot when x is a factor. HE plots don't care about this distinction, because every

term in a model is just an ellipsoid for the corresponding H matrix.

The following example takes up a larger MANOVA design with several sets of analyses, and also serves as a vehicle for illustrating the kinds of model checking that should be part of any statistical analysis of an MLM. Further, it utilizes real world data from the psychological literature.

Neuro- and Social-Cognitive Measures in Psychiatric Groups

This example is concerned with the following substantive questions:

- To what extent can patients diagnosed as schizophrenic or with schizoaffective disorder (using the DSM-IV Structured Clinical Interview) be distinguished from a normal control sample using a well-validated, comprehensive neurocognitive battery specifically designed for individuals with psychosis (Heinrichs et al., 2015)?

- If they differ, do any of the cognitive domains (processing speed, attention, working memory, verbal learning, visual learning, problem solving) show larger or smaller differences among these groups? Do they discriminate in the same or different ways?
- Moreover, are the schizophrenic patients distinguishable from those with schizoaffective disorder on such measures (Heinrichs, Ammari, McDermid Vaz, & Miles, 2008)?
- For the above questions, how do the results differ when assessed by measures of *social* cognition instead of the neurocognitive measures?

These are the questions addressed in a recent PhD dissertation by (Hartman, 2016) at York University. Her study involved a total clinical sample comprising 116 patients who had a diagnosis of schizophrenia (n = 70) or schizoaffective disorder (n = 46), screened for absence of learning disability, neurological disorder and substance use disorder. A demographically matched sample of non-psychiatric control participants consisted of n = 146 subjects, screened for absence of medical and psychiatric illness and history of substance abuse.

Neurocognitive function in all subjects was assessed using the MATRICS Consensus Cognitive Battery (MCCB; Nuechterlein et al., 2008), developed by the NIMH Measurement and Treatment to Improve Cognition in Schizophrenia (MATRICS) initiative. A subset of 139 of the subjects from all groups ($n = 43, 30, 66$ in the three groups) also completed a battery of tests designed to measure aspects of social cognition (called here: managing emotions, theory of mind, externalizing bias, personalizing bias).

The data presented in these analyses are subsets that are comprised only of observations with complete data. By the kind permission of Leah Hartman, these data are included (and described further) in the data sets NeuroCog and SocialCog in the heplots package. See the help pages (via ? NeuroCog and ? SocialCog) for further details on the structure of the dataset and pertinent references.

In the brief descriptions of the results of analyses of these data, we focus on those that are important for answering the above substantive questions and visualizations that support their interpretation.

Neurocognitive measures The main factor here is diagnostic group, called Dx in the dataset. To facilitate answering questions regarding group differences, the following contrasts were applied: the first column compares the control group to the average of the diagnosed groups, the second compares the schizophrenia group against the schizoaffective group.

```
data(NeuroCog, package="heplots")
```

```
contrasts(NeuroCog$Dx)
```

```
##                   [,1] [,2]
## Schizophrenia    -0.5    1
## Schizoaffective  -0.5   -1
## Control           1.0    0
```

The MLM for the six neurocognitive response measures can be fit and tested as follows:

```
NC.mlm <- lm(cbind(Speed, Attention,
    Memory, Verbal, Visual, ProbSolv) ~
    Dx,  data=NeuroCog)
Anova(NC.mlm)
```

```
## Type II MANOVA Tests: Pillai test statistic
##    Df test stat approx F num Df den Df    Pr(>F)
## Dx  2    0.2992   6.8902     12    470 1.562e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
##   0.1 ' ' 1
```

We also test the linear hypotheses for the contrasts, but for brevity do not display the results here. The contrast Dx1 for control vs. diagnosed is highly significant, but the second contrast Dx2 comparing the schizophrenic and schizoaffective group is not.

```
print(linearHypothesis(NC.mlm, "Dx1"),
    SSP=FALSE)
print(linearHypothesis(NC.mlm, "Dx2"),
    SSP=FALSE)
```

So the question becomes: how to understand these results. HE plots tell a very simple story (Figure 6): The control group performs higher on all measures than the diagnosed groups, which do not differ between themselves. This pattern is consistent across all of the response variables, as we see from a plot of pairs(NC.mlm) (not shown to conserve space).

```
heplot(NC.mlm, variables=1:2, ...)
```

We can gain further insight, and a simplified plot showing all the response variables by projecting the MANOVA into the canonical space, which is entirely 2-dimensional (because $df_n = 2$). However, the output from candisc() shows that 98.5% of the mean differences among groups can be accounted for in one canonical dimension.

```
NC.can <- candisc(NC.mlm)
NC.can
```

```
## Canonical Discriminant Analysis for Dx:
##
##      CanRsq Eigenvalue Difference Percent Cumulative
## 1 0.2929499  0.4143270    0.40803  98.504     98.504
## 2 0.0062532  0.0062926    0.40803   1.496    100.000
##
## Test of H0: The canonical correlations in the
## current row and all that follow are zero
```
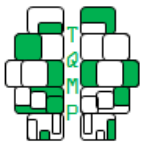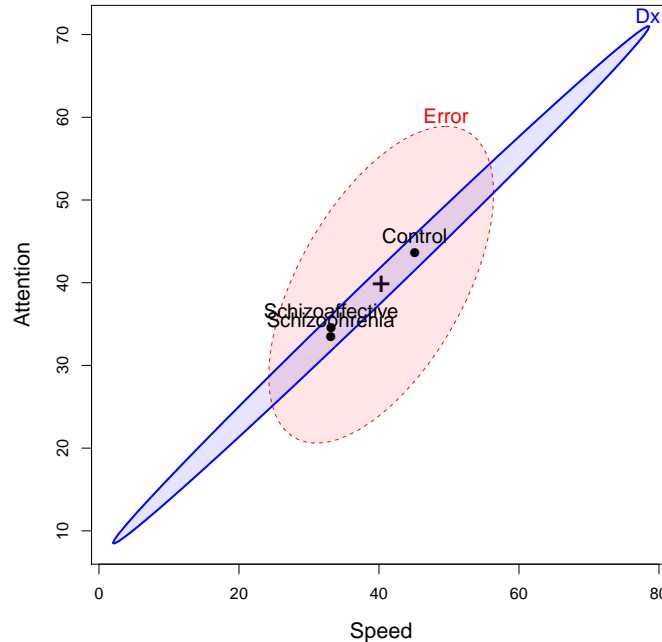
Figure 6 ■ HE plot for the 'NeuroCog' data MANOVA showing Attention vs. Speed of processing. All other pairwise plots follow the same pattern.


```
##
##     LR test stat approx F num Df den Df Pr(> F)
## 1      0.70263  22.9659      4    476  <2e-16 ***
## 2      0.99375   1.5039      1    239  0.2213
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
##    0.1 ' ' 1
```

Figure 7 is the result of the $\mathrm{plot}()$ method for "candisc" objects, that is, the result of calling $\mathrm{plot}(\mathrm{NC.can}, ...)$. It plots the two canonical scores, $\boldsymbol{Z}_{n\times 2}$ for the subjects, together with data ellipses for each of the three groups.[6] As noted earlier, the projections of the variable vectors in this plot on the coordinate axes are proportional to the correlations of the responses with the canonical scores.

The interpretation of Figure 7 is again fairly straightforward. The separation of the groups is essentially one-dimensional, with the control group higher on all measures. Moreover, the variables processing speed and visual memory are the purest measures of this dimension, but all variables contribute positively. The second canonical dimension accounts for only 1.5% of group mean differences and is non-significant (by a likelihood ratio test). Yet, if we were to interpret it, we would note that the schizophrenia group is slightly higher on this dimension, scoring better in problem solving and slightly worse on working memory,

attention, and verbal learning tasks.

Social cognitive measures The social cognitive measures were designed to tap various aspects of the perception and cognitive processing of emotions of others. Emotion perception was assessed using a Managing Emotions score from the MCCB. A "theory of mind" (ToM) score assessed ability to read the emotions of others from photographs of the eye region of male and female faces. Two other measures, externalizing bias ($\mathrm{ExtBias}$) and personalizing bias ($\mathrm{PersBias}$) were calculated from a scale measuring the degree to which individuals attribute internal, personal or situational causal attributions to positive and negative social events.

The analysis of the $\mathrm{SocialCog}$ data proceeds in a similar way: first we fit the MANOVA model, then test the overall differences among groups using $\mathrm{Anova}()$, and finally test the same two contrasts using $\mathrm{linearHypothesis}()$. We find that the overall multivariate test is again significant, but now *both* contrasts are significant (Dx1: $F(4, 133) = 5.21, p < 0.001$; Dx2: $F(4, 133) = 2.49, p = 0.0461$), the test for $\mathrm{Dx2}$ just barely.

```
data(SocialCog, package="heplots")
SC.mlm <-  lm(cbind(MgeEmotions,ToM,
```

---

[6]A canonical HE plot, such as shown in Figure 4 would result from calling $\mathrm{plot}(\mathrm{NC.can}, ...)$.

Figure 7 ■ Canonical discriminant plot for the 'NeuroCog' data MANOVA. Scores on the two canonical dimensions are plotted, together with 68% data ellipses for each group.
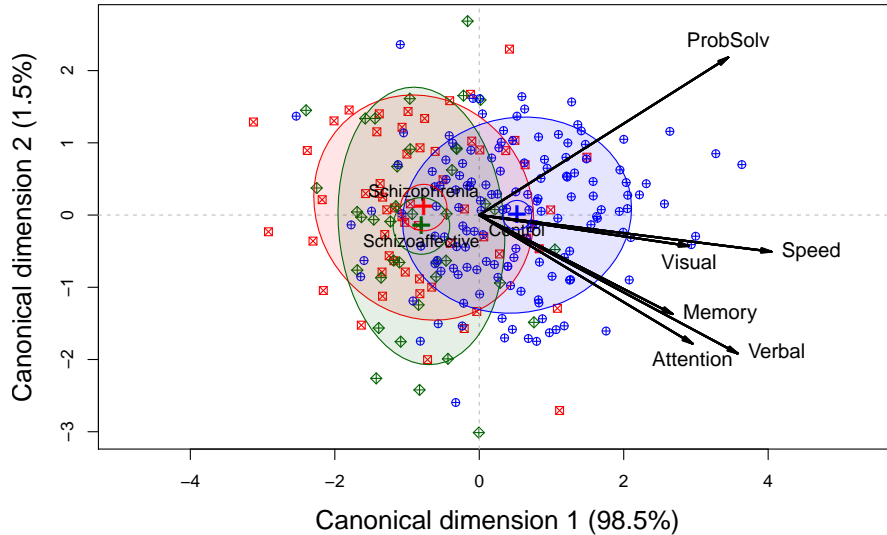

```
        ExtBias, PersBias) ~ Dx,
                    data=SocialCog)
Anova(SC.mlm)
linearHypothesis(SC.mlm, "Dx1")
linearHypothesis(SC.mlm, "Dx2")
```

These results are important, because, if they are reliable and make sense substantively, they imply that patients with schizophrenia and schizoaffective diagnoses *can* be distinguished by their performance on tasks assessing social perception and cognition.

At this point, it is useful to visualize the nature of these differences among groups with HE plots. Corresponding to Figure 6 for the neurocognitive responses, Figure 8 shows the HE plot for the first two response variables, together with the (degenerate) ellipses representing the tests of the Dx1 and Dx2 contrasts.

```
heplot(SC.mlm, variables=1:2, hypotheses
    =list("Dx1"="Dx1", "Dx2"="Dx2"), ...)
```

It can be seen that the three group means are approximately equally spaced on the ToM measure, whereas for MgeEmotions, the control and schizoaffective groups are quite similar, and both are higher than the schizophrenic group. This ordering of the three groups was somewhat

similar for the other responses, as we could see in a pairs(SC.mlm) plot.

Model checking

Normally, we would continue this analysis, and consider other HE and canonical discriminant plots to further interpret the results, in particular the relations of the cognitive measures to group differences, or perhaps an analysis of the relationships between the neuro- and social-cognitive measures. We don't pursue this here for reasons of length, and this example actually has a more important lesson to demonstrate.

Before beginning the MANOVA analyses, extensive data screening was done by the client using SPSS, in which all the response *and* predictor variables were checked for univariate normality and multivariate normality (MVN) for both sets. This traditional approach yielded a huge amount of tabular output and no graphs, and did not indicate any major violation of assumptions.[7]

A simple visual test of MVN and the possible presence of multivariate outliers is related to the theory of the data ellipse: Under MVN, the squared Mahalanobis distances $D_M^2(\boldsymbol{y}) = (\boldsymbol{y} - \bar{\boldsymbol{y}})^\mathsf{T} \boldsymbol{S}^{-1} (\boldsymbol{y} - \bar{\boldsymbol{y}})$ should follow a χ_p^2 distribution. Thus, a quantile-quantile plot of the ordered D_M^2

---

[7]Actually, multivariate normality of the predictors in $\boldsymbol{X}$ is not required in the MLM. This assumption applies only to the contidional values $\boldsymbol{Y} \mid \boldsymbol{X}$, i.e., that the errors $\boldsymbol{u}_i' \sim \mathcal{N}_P(\boldsymbol{0}, \boldsymbol{\Sigma})$ with constant covariance matrix. Moreover, the widely used MVN test statistics, such as Mardia's (1970) test based on multivariate skewness and kurtosis are known to be quite sensitive to mild departures in kurtosis (Mardia, 1974) which do not threaten the validity of the multivariate tests.

Figure 8 ■ HE plot for the 'SocialCog' data MANOVA showing theory of mind vs. managing emotions. The embedded lines show the H ellipses for the contrasts 'Dx1' and 'Dx2'.
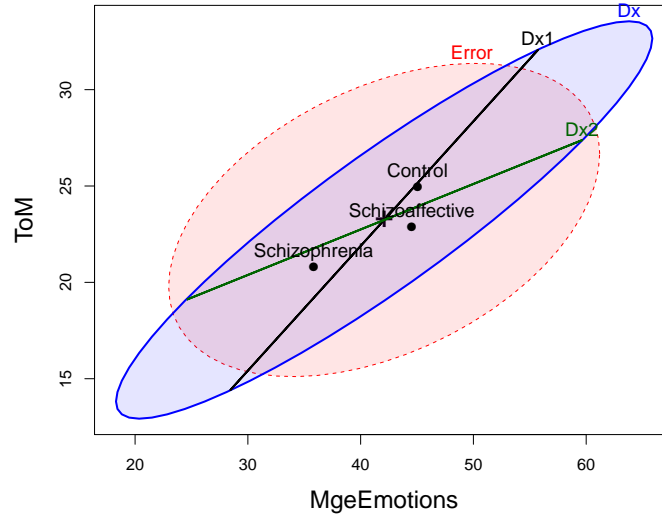
values vs. corresponding quantiles of the χ^2 distribution should approximate a straight line (Cox, 1968; Healy, 1968). Note that this should be applied to the *residuals* from the model – residuals(SC.mlm) – and not to the response variables directly.

The heplots package implements a cqplot() method for "mlm" objects that does this. Calling this function for the model SC.mlm produces Figure 9. It is immediately apparent that there is one extreme multivariate outlier; the remaining observations are within the 95% confidence envelope (using the classical product-moment estimate of S).

```
cqplot(SC.mlm, id.n=0, ...)
```

Further checking revealed that this was a data entry error where one case (15) in the schizophrenia group had a score of -33 recorded on the ExtBias measure, whose valid range was (-10, +10). In R, it is very easy to re-fit a model to a subset of observations (rather than modifying the dataset itself) using update(). The result of the overall Anova() and the test of Dx1 were unchanged; however, the multivariate test for the most interesting contrast Dx2 comparing the schizophrenia and schizoaffective groups became non-significant at the $\alpha = 0.05$ level ($F(4, 133) = 2.18, p = 0.0742$).

```
SC.mlm1 <- update(SC.mlm, subset=
    rownames(SocialCog)!="15")
Anova(SC.mlm1)
print(linearHypothesis(SC.mlm1, "Dx1"),
```

```
    SSP=FALSE)
print(linearHypothesis(SC.mlm1, "Dx2"),
    SSP=FALSE)
```

This outcome creates a bit of a quandry for further analysis (univariate follow-up tests? try a robust model?) and reporting (what to claim about the Dx2 contrast?) that we don't explore here. Rather, we proceed to attempt to interpret the MLM with the aid of canonical analysis and a canonical HE plot. The canonical analysis of the model SC.mlm1 now shows that both canonical dimensions are significant, and account for 83.9% and 16.1% of between group mean differences respectively.

```
SC.can1 <- candisc(SC.mlm1)
SC.can1
```

```
## Canonical Discriminant Analysis for Dx:
##
##     CanRsq Eigenvalue Difference Percent Cumulative
## 1 0.164496   0.196882    0.15909  83.897     83.897
## 2 0.036412   0.037788    0.15909  16.103    100.000
##
## Test of H0: The canonical correlations in the
## current row and all that follow are zero
##
##   LR test stat approx F num Df den Df   Pr(> F)
## 1      0.80508   7.6715      4    268 7.254e-06 ***
## 2      0.96359   5.1014      1    135   0.02551 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
##    0.1 ' ' 1
```

The HE plot version of this canonical plot is shown in Figure 10. Because the heplot() method for a "candisc"

Figure 9 ∎ χ^2 quantile-quantile plot for residuals from the model 'SC.mlm'. The confidence band gives a point-wise 95% envelope, providing information about uncertainty. One extreme multivariate outlier is highlighted.
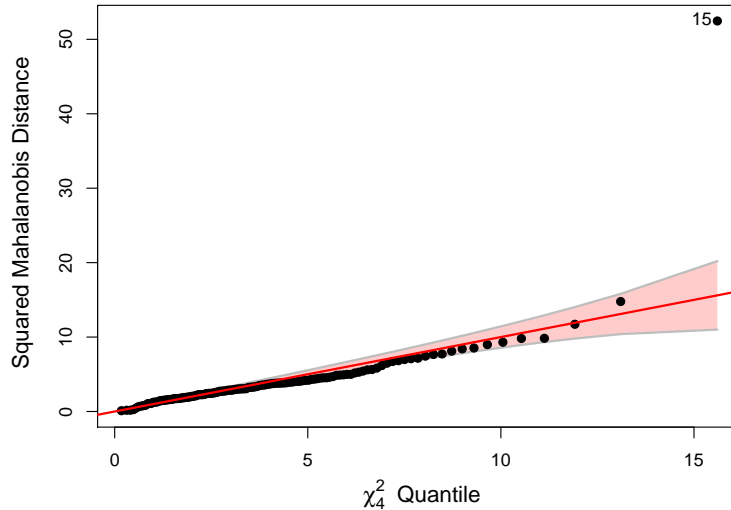
object refits the original model to the Z canonical scores, it is easy to also project other linear hypotheses into this space. Note that in this view, both the Dx1 and Dx2 contrasts project outside E ellipse[8].

```
heplot(SC.can1, fill=TRUE,
        hypotheses=list("Dx1"="Dx1", "Dx2
   "="Dx2"), ...)
```

This canonical HE plot has a very simple description: (a) Dimension 1 orders the groups from control to schizoaffective to schizophrenia, while dimension 2 separates the schizoaffective group from the others; (b) Externalizing bias and theory of mind contributes most to the first dimension, while personal bias and managing emotions are more aligned with the second; and, (c) the relations of the two contrasts to group differences and to the response variables can be easily read from this plot.

MANCOVA and homogeneity of regression

As an extension of the MANOVA model, MANCOVA incorporates one or more continuous predictors (covariates) whose effect is "averaged out" of the model for the factors, thus controlling for their effects. For the same data structure (a mixture of discrete factors and continuous predictors) and a focus on regression relations in MMRA, but with an additional discrete factor, the questions of interest relate to *homogeneity of regression*.

Rohwer data

The Rohwer data used in the MMRA example has a discrete two-level variable pertaining to the socioeconomic status (SES) of the children, and in the previous analysis only the high SES group were analyzed. If we were to model the complete dataset, the intercepts for the low and high SES student groups would be constrained to be equal if the SES variable is not included in the model. The MANCOVA model allows the intercepts to be freely estimated and provides a test of group differences in means on the responses, controlling for any included covariates.

To extend our previous model, this can be incorporated by simply adding the relevant factor variable name (SES) to the call to lm():

```
rohwer.mod <- lm(cbind(SAT, PPVT, Raven)
    ~ SES + n + s + ns + na + ss,
   data=Rohwer)
```

The summary output from this model indicates that SES does have a significant effect— indicating that the intercepts cannot be considered equal. This relationship can be visualized pairwise, as shown in Figure 11. The (degenerate) H ellipsoid for the SES effect has a positive slope in all panels, showing that the high SES group has greater conditional means for all the ability response variables. The significant regression effects of na and ns can also be

---

[8]The direct application of significance tests to canonical scores probably requires some adjustment because these are computed to have the optimal between-group discrimination.

Figure 10 ■ Canonical HE plot for the corrected 'SocialCog' MANOVA. The variable vectors show the correlations of the responses with the canonical variables. The embedded lines show the projections of H ellipses for the contrasts 'Dx1' and 'Dx2' in canonical space.
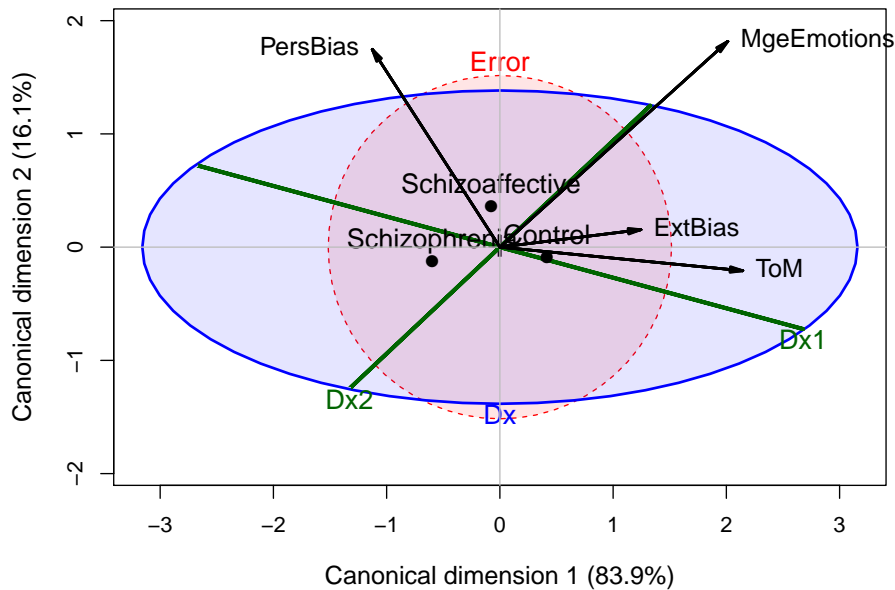
seen in most panels, as can the combined effect ("Regr") for all the paired association tests. A useful exercise for the reader of this tutorial is to compare the pattern of relationships for the MANCOVA model in Figure 11 with those in the MMRA model for the high SES group shown in Figure 5.

```
pairs(rohwer.mod,
      hypotheses=list("Regr" =
      c("n", "s", "ns", "na", "ss")),
      fill=TRUE, fill.alpha=0.1)
```

However, the MANCOVA test of the SES effect assumes that the *slopes* for the low and high SES groups are equivalent, an assumption that seems dubious at best. A convenient feature of $\mathrm{lm}()$ in R is that it allows for incorporating interactions between the predictors and the covariate. This is available in model formulas by using "*" and parentheses, which indicate that all two-way interactions with SES should be added to the model (along with the main effects that are marginal to the interactions).

```
rohwer.mod1 <- lm(cbind(SAT, PPVT, Raven
    ) ~ SES * (n + s + ns + na + ss),
                  data=Rohwer)
Anova(rohwer.mod1)
```

```
## Type II MANOVA Tests: Pillai test statistic
##         tDf test stat approx F numDf den Df     Pr(>F)
## SES      1   0.39123  11.7822     3     55  4.55e-06 ***
## n        1   0.07901   1.5727     3     55 0.2063751
## s        1   0.12524   2.6248     3     55 0.0595192 .
## ns       1   0.25412   6.2461     3     55 0.0009995 ***
## na       1   0.30663   8.1077     3     55 0.0001459 ***
## ss       1   0.06017   1.1738     3     55 0.3281285
## SES:n    1   0.07231   1.4290     3     55 0.2441738
## SES:s    1   0.09942   2.0240     3     55 0.1211729
## SES:ns   1   0.11757   2.4425     3     55 0.0738258 .
## SES:na   1   0.14801   3.1850     3     55 0.0308108 *
## SES:ss   1   0.05733   1.1150     3     55 0.3509357
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
    0.1 ' ' 1
```

The above table reveals only 1 significant interaction with SES by the separate multivariate tests. However, the MANCOVA model assumes that *all* slopes are equal, and this can be tested collectively using $\mathrm{linearHypothesis}()$:

```
# store coefficient names in a vector
coefs <- rownames(coef(rohwer.mod1))
# only test for interaction effects
print(linearHypothesis(rohwer.mod1,
    coefs[grep(":", coefs)]), SSP=FALSE)
```

```
## Multivariate Tests:
##         Df test stat approx F numDf   denDf     Pr(>F)
## Pillai   5 0.4179376 1.845226    15 171.000 0.0320861 *
## Wilks    5 0.6235824 1.893613    15 152.232 0.0276949 *
## H-L      5 0.5386513 1.927175    15 161.000 0.0239619 *
```
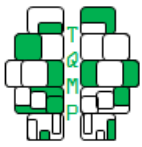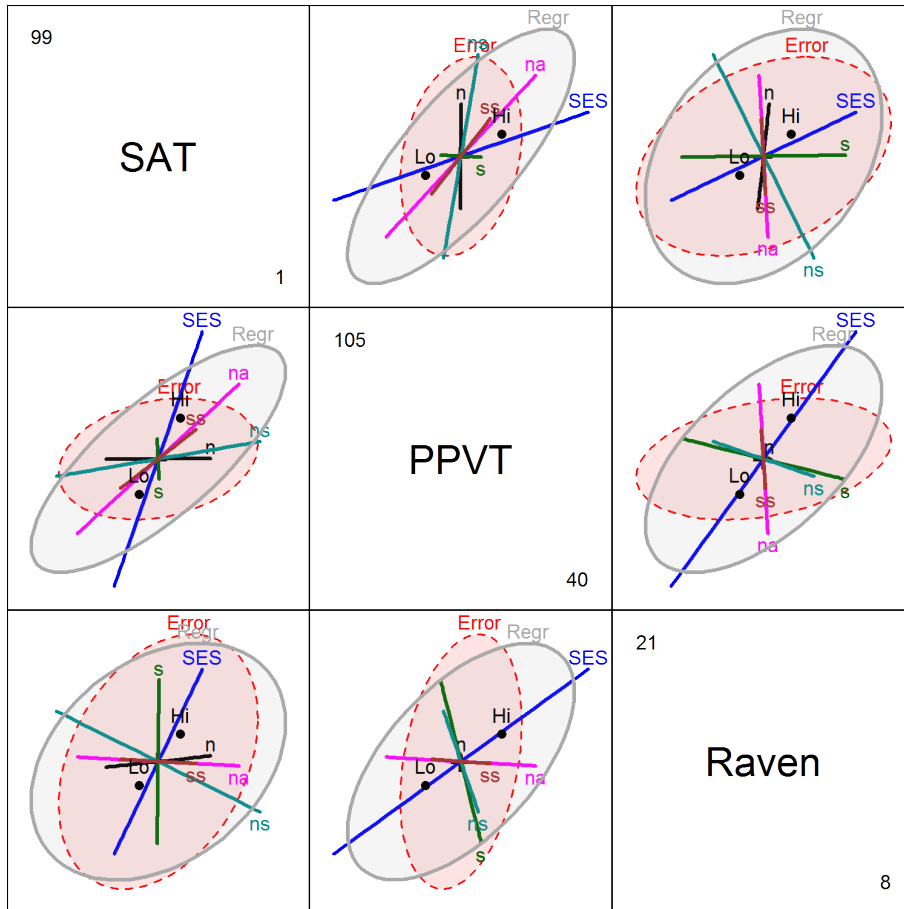
Figure 11 ■ Pairwise HE plot for MANCOVA model for the Rohwer data. The effect of SES, with 1 df, plots as a line determined by the conditional means of the high and low SES groups. The numbers in the diagonal panels give the range of each variable.


```
## Roy     5 0.3846488 4.384997     5  57.000 0.0019053 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
   0.1 ' ' 1
```

The above shows that collectively the interactions are significant, no matter which criteria is chosen.

Another way to view the question of homogeneity of regression is to fit *separate* multivariate regression models for the two groups, using the subset argument to lm(). This has the disadvantage of being less efficient than the homogeneity of regression test from the model rohwer.mod1 above, but the advantage of allowing us to visualize possible differences in within-group error covariance.

```
rohwer.ses1 <- lm(cbind(SAT, PPVT, Raven
    ) ~ n + s + ns + na + ss, data =
    Rohwer, subset = SES == "Hi")
rohwer.ses2 <- lm(cbind(SAT, PPVT, Raven
```

```
    ) ~ n + s + ns + na + ss, data =
    Rohwer, subset = SES == "Lo")
```

Here we show the complete code to produce a visualization of these models allowing them to be visualized *together* and compared for all features: intercepts, slopes and error covariances. It illustrates the high level of control of graphic features available in R and in the heplots and other packages for producing publication-quality graphs. Figure 12 is generated by two calls to heplot() using add=TRUE to overlay them.

```
# Low SES students:
heplot(rohwer.ses2,
    col = c("red", rep("black",5),"blue"),
    hypotheses = list("B=0, Low SES" = c("
        n", "s", "ns", "na", "ss")),
    level = 0.5, cex = 1.25,
```

```
    fill = c(TRUE, FALSE), fill.alpha =
    0.05,
    xlim = c(-15, 110),
    ylim = c(40, 110),
    xlab = "Student Achievement Test",
    ylab = "Peabody Picture Vocabulary
    Test",
    label.pos = c(1, rep(NULL, 5), 1))

# High SES students:
heplot(rohwer.ses1, col = c("red", rep("
    black", 5), "blue"),
  hypotheses = list("B=0, High SES" = c(
    "n", "s", "ns", "na", "ss")),
    level = 0.5, cex = 1.25,
    # place both plots on same graphic
    add = TRUE,
    # error ellipse not drawn by default
    error = TRUE,
    fill = c(TRUE, FALSE), fill.alpha =
    0.05,
    xlim = c(-15, 110),
    ylim = c(40, 110))
```

In Figure 12 the greater mean scores for the high SES group on both measures is readily apparent. It can also be seen that the overall regression ellipsoid testing $H_0 : \boldsymbol{B} = \boldsymbol{0}$ is more aligned with the SAT axis for the high SES group than for the low group, reflecting better prediction of SAT than PPVT for the high group. Among the individual predictors, na and ns are also more important in the high SES group.

Guidelines for analyzing, visualizing and reporting MLMs

From the above description and examples, and from our broader experience in consulting with researchers in psychology, we offer the following guidelines for analysis, visualization and reporting of research using multivariate linear models.

Analysis

- **Reproducible analysis**: Errors and confusion can creep in when an analysis is conducted from many discrete, manual steps, especially if done via unreplicable GUI menu choices. It is highly recommended to always keep and maintain a script of syntax that can be re-run to recreate your research data set from its source. Document special features and modifications in comments.

- **Data screening vs. model checking**: Data screening is important, but if done in a rote univariate manner (test everything for normality) and without data displays, can generate too much output to comprehend, and rely

on tests that are neither necessary nor sufficient. Instead, we recommend the utilization of visualizations that allow direct inspection of these assumptions. For example, the assumption of homogeneity of variance in MANOVA is often tested by univariate Levene-type tests (Brown & Forsythe, 1974; Levene, 1960). The heplots package implements several graphical methods for this question (e.g., plot methods for Box's M test; Box, 1949).

- **Residuals** from a multivariate linear model contain *all* the information about what doesn't fit the data, and careful analysis of residuals is more important than rote data screening. MLMs are more complex, because there are residuals for each response. Mahalanobis squared distances of the residuals provide a good summary of outlyingness and a χ^2 QQ plot such as Figure 9 is an excellent screening device, providing information relevant to the assumption of multivariate normality and the presence of outliers.

- **Sensitivity tests**: Many questions regarding the details of choices made in an analysis (include or exclude outliers?) or the impact of potential violation of assumptions can be answered by conducting a sensitivity test: Run the the analysis in a different way (e.g., use a robust method) and ask whether it makes a difference in your conclusions. If not, great (and report this); if it does, investigate why the results differ.

Visualization

- **Tweaking graphs**: Graphs for analysis purposes only need to be good enough for you to see the important features. Often the defaults are sufficient, but plot annotations, such as adding smoothed ("loess") curves (Cleveland, Grosse, & Shu, 1992) and data ellipses, can be helpful; see Fox (2008) for these methods.

- **The 80-20 rule**: For publication graphs, remember the 80-20 rule (Friendly & Meyer, 2016, §1.4.7): You can usually get 80% of a useful graph with 20% of total effort, but the remaining 80% of effort may be required to get the remaining 20%. Among other things, this means paying careful attention to font sizes, axis ranges, titles and tick mark values, grouping attributes (color, style) for points and lines, area fill to highlight regions (like data ellipses), and so forth.

- **Raw data**: For multivariate data, scatterplot matrices, supplemented by data ellipsoids, regression lines and/or smoothed curves (the scatterplotMatrix() in the car package is the most useful) are very helpful. Organize the variables as $(\boldsymbol{Y}, \boldsymbol{X})$ to examine the relations within and between sets. Make points small if n is large, or remove them entirely to the see the summaries alone. The function covEllipses() in heplots can

Figure 12 ■ HE plot for the models fitting separate slopes and intercepts to the high and low SES groups.
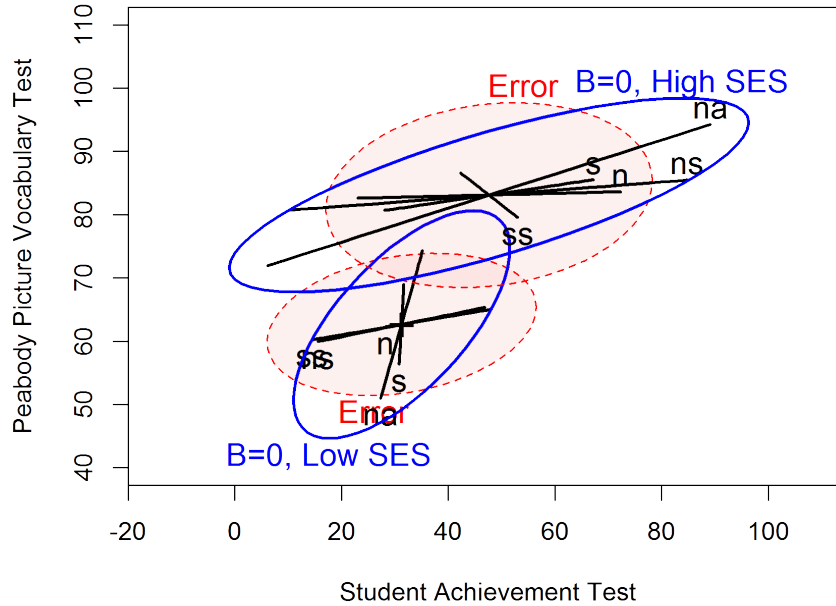
produce just the raw data ellipses.

- **Transformations**: Linear models assume that the residuals are normally distributed, but makes no such assumptions about the predictors. In regression designs, the question of linearity of regression relations is more important. Thus, don't worry so much about univariate measures of skewness and kurtosis. Rather, use non-parametric smoothed (e.g., loess) curves in scatterplots to see the trend in the data. If this is decidedly non-linear, consider transformations such as $\sqrt{x}$ or $\log x$. The car package contains many useful tools for data transformations (e.g., $\mathrm{boxCox}()$, $\mathrm{powerTransform}()$).

- **Low-D** views: For more than a few variables, biplots (Greenacre, 2010) can show a 2D (or 3D) summary view of the data points and variable relations (as vectors), using results of a principal component analysis. The nicest implementation in R is the ggbiplot (Vu, 2011) package.

- **Model-appropriate visualizations**: Always try to rely on visualizations that are model appropriate: if the model of interest is multivariate, remember that univariate and bivariate displays only show limited views

of the data, and can obscure the higher level relationships. When the right-hand side of the model is complex (many terms or interactions), *effect plots* (Fox, 2003) are among the most effective visual summaries for interpretation.

- **HE plots** and their scatterplot-matrix version can be used to visualize the effects of model terms in relation to the response variables as we demonstrated above. In a MANOVA setting, dimension reduction techniques via *canonical* HE plots should be used to show the dimensions that account for maximal discrimination among groups. For MMRA designs, related canonical correlation plots are implemented in the candisc package.

Reporting

- **Reproducible reporting**: Try to avoid the temptation to just cut and paste analysis results into a paper or presentation. The facilities available in R— knitr, rmarkdown, and the user interface provided by RStudio (https://www.rstudio.com/) provide an integrated environment for both doing analysis and writing it up. There is a bit of a learning curve, but well worth the effort. The important benefit is that if the data or de-

tails of the analysis change, the paper or slides can be *automatically* updated.

- **Tabular displays**: Multivariate models provide a plethora of potential results to present, and so require greater selectivity in *what* to present and *how* to present it. Tables like Table 1 can be easily generated (by stargazer) if you want to record the model coefficients, but often it is just the significance stars your readers will see. Gelman, Pasarica, and Dodhia (2002) provide some useful advice for turning tables into graphs, and ideas for visualizing tables using shaded backgrounds or other cues (e.g., tableplots; Kwan & Friendly, 2012) to show importance or patterns are becoming increasingly available.

- **Graphs** are more common in some journals than others, but you have to choose wisely and make them most effective for what you want to show and say. If the reviewer or reader cannot see or understand what is important and why, you may need to make the essential features more salient, and/or describe clearly in the plot title or figure caption what should be understood from the graph.

- **Novel graphical methods** like HE plots and their canonical cousins may be unfamiliar to most readers, but if you've understood the examples in this paper, you should be able to explain them sufficiently, or else just use them as analysis graphs that lead to your understanding of the results. Remember that Playfair (1801) took five pages to explain what a barchart represented, but barcharts are now (perhaps too) commonplace.

Discussion

Graphical methods for univariate linear models are now widely developed and widely used, from exploratory plots to tools for model diagnosis (influence plots of leverage vs. residuals) and interpretation (effect plots). Although the theory of multivariate linear models is a straightforward generalization of their univariate counterparts, the corresponding extension of visualization techniques has only been relatively recent and remains incomplete in some respects. The goal of this paper was to illustrate, with tutorial examples in R, a collection of methods we have developed over the last decade to fill this gap, and which we believe are sufficiently advanced to be helpful to psychological researchers.

So, what is still missing to complete this extension of univariate graphical methods to the multivariate domain?

- In R (and most other software) it is now relatively easy to obtain a reasonable collection of standard diagnostic plots for *univariate* response models with a simple command, menu option or mouse click. For example, the generic $\mathrm{plot}()$ in R produces the "regression quartet" (residuals vs. fitted values, normal QQ plots, spread-level plots for checking homogeneity of variance, and leverage-influence plots) when applied to a univariate "lm" or "glm" (generalized linear model) object.

- Yet, this idea has not yet been extended to MLMs: The call $\mathrm{plot}(\text{my.model})$ for a "mlm" object gives the disheartening message Error: 'plot.mlm' is not implemented yet. Another package we have developed, mvinfluence implements multivariate analogs of leverage and influence (Cook's D) and a set of $\mathrm{influencePlot}()$ methods to make some steps in this direction.

- Robust methods for *univariate* linear (and generalized linear) models have come a long way in recent years, and their implementation in software is growing and maturing (e.g., $\mathrm{lmRob}()$ in the robust package; PROC ROBUSTREG in SAS); however their widespread use by researchers remains to catch up to theory. These are important because they provide a means to protect researchers from some of the effects of messy data. Analogous robust methods for MLMs are not yet fully developed, but our implementation of $\mathrm{robmlm}()$ in the heplots package goes a long way toward filling this gap.

- Other visualizations for MLMs: Aside from the questions of multivariate normality of residuals and absence of multivariate outliers, the other main assumption of MANOVA is that of equality of covariance matrices. We recently implemented several graphical methods related to Box's M test (Box, 1949) and a multivariate generalization of Levene's test (Levene, 1960) in the heplots package.

The HE plot framework we have described serves as a means to help researchers visualize the sources of significance for model effects and the relationships between the predictors and the various response outcomes. Their geometry provides a visual means of judging both effect size and significance. The canonical, dimension-reduction techniques we described makes it easier to understand these relations for relatively many response variables. As we have illustrated, this often leads to surprisingly simple interpretations of MLMs that could not readily be achieved otherwise.

References

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, *36*(3-4), 317–346. doi:10.1093/biomet/36.3-4.317

Brown, M. B. & Forsythe, A. B. (1974). Robust tests for equality of variances. *Journal of the American Statistical Association*, *69*(346), 364–367. doi:10.1080/01621459.1974.10482955

Cleveland, W. S., Grosse, E., & Shu, W. M. (1992). Local regression models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical models in S* (pp. 309–372). Pacific Grove, CA: Wadsworth.

Cotton, R. (2013). *Learning R*. Sebastopol, CA: O'Reilly Media.

Cox, D. (1968). Notes on some aspects of regression analysis. *Journal of the Royal Statistical Society Series A*, *131*, 265–279.

Dempster, A. P. (1969). *Elements of continuous multivariate analysis*. Reading, MA: Addison-Wesley.

Fox, J. (2003). Effect displays in R for generalized linear models. *Journal of Statistical Software*, *8*(15), 1–27.

Fox, J. (2008). *Applied regression analysis and generalized linear models (2nd ed.)* Thousand Oaks, CA: Sage Publications.

Fox, J. & Friendly, M. (2014). Heplots: Visualizing hypothesis tests in multivariate linear models (Version 0.1). Retrieved from http://CRAN.R-project.org/package=heplots

Fox, J., Friendly, M., & Monette, G. (2009). Visualizing hypothesis tests in multivariate linear models: The heplots package for R. *Computational Statistics*, *24*(2), 233–246. doi:10.1007/s00180-008-0120-1

Fox, J. & Weisberg, S. (2011). *An R companion to applied regression (2nd ed.)* Thousand Oaks, CA: Sage.

Fox, J. & Weisberg, S. (2015). car: companion to applied regression (Version 1.0). Retrieved from http://R-Forge.R-project.org/projects/car/

Friendly, M. (2007). HE plots for multivariate general linear models. *Journal of Computational and Graphical Statistics*, *16*(2), 421–444. doi:10.1198/106186007X208407

Friendly, M. (2010). HE plots for repeated measures designs. *Journal of Statistical Software*, *37*(4), 1–37. Retrieved from http://www.jstatsoft.org/v37/i04

Friendly, M. & Denis, D. (2005). The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences*, *41*(2), 103–130. doi:10.1002/jhbs.20078

Friendly, M. & Fox, J. (2013). Candisc: Visualizing generalized canonical discriminant and canonical correlation analysis (Version 1.0). Retrieved from http://CRAN.R-project.org/package=candisc

Friendly, M. & Meyer, D. (2016). *Discrete data analysis with R: Visualization and modeling techniques for categorical and count data*. Boca Raton, FL: Chapman & Hall/CRC.

Friendly, M., Monette, G., & Fox, J. (2013). Elliptical insights: Understanding statistical methods through elliptical geometry. *Statistical Science*, *28*(1), 1–39. doi:10.1214/12-STS402

Friendly, M. & Sigal, M. (2014). Recent advances in visualizing multivariate linear models. *Revista Colombiana de Estadistica*, *37*(2), 261–283. doi:10.15446/rce.v37n2spe.47934

Gelman, A., Pasarica, C., & Dodhia, R. (2002). Let's practice what we teach: Turning tables into graphs. *The American Statistician*, *56*(2), 121–130.

Gittins, R. (1985). *Canonical analysis: A review with applications in ecology*. Berlin: Springer-Verlag.

Greenacre, M. (2010). *Biplots in practice*. Fundación BBVA. Retrieved from https://books.google.ca/books?id=dv4LrFP7U/_EC

Hand, D. J. & Taylor, C. C. (1987). *Multivariate analysis of variance and repeated measures: A practical approach for behavioural scientists*. London: Chapman & Hall.

Hanley, J. A. (2004). Transmuting women into men: Galton's family data on human stature. *The American Statistician*, *58*(3), 237–243. doi:10.1198/000313004X1558

Hartman, L. I. (2016). *Schizophrenia and schizoaffective disorder: One condition or two?* (Doctoral dissertation, York University).

Healy, M. J. R. (1968). Multivariate normal plotting. *Journal of the Royal Statistical Society Series C*, *17*(2), 157–161.

Heinrichs, R. W., Ammari, N., McDermid Vaz, S., & Miles, A. A. (2008). Are schizophrenia and schizoaffective disorder neuropsychologically distinguishable? *Schizophrenia Research*, *99*(1), 149–154. doi:10.1016/j.schres.2007.10.007

Heinrichs, R. W., Pinnock, F., Muharib, E., Hartman, L., Goldberg, J., & McDermid Vaz, S. (2015). Neurocognitive normality in schizophrenia revisited. *Schizophrenia Research: Cognition*, *2*(4), 227–232. doi:10.1016/j.scog.2015.09.001

Kwan, E. & Friendly, M. (2012). Tableplot: Represents tables as semi-graphic displays (Version 1.0). Retrieved from http://CRAN.R-project.org/package=tableplot

Levene, H. (1960). Robust tests for equality of variances. In S. G. G. I. Olkin, W. Hoeffding, W. G. Madow, & H. B. Mann (Eds.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 278–292). Stanford: Stanford University Press.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*(3), 519–530. doi:10.2307/2334770

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya: The Indian Journal of Statistics, Series B*, *36*(2), 115–128.

Matloff, N. (2011). *The art of R programming: A tour of statistical software design*. San Francisco, CA: No Starch Press.

Meyers, L., Gamst, G., & Guarino, A. (2006). *Applied multivariate research: Design and interpretation*. Thousand Oaks, CA: SAGE Publications.

Monette, G. (1990). Geometry of multiple regression and interactive 3-d graphics. In J. Fox & S. Long (Eds.), *Modern methods of data analysis* (pp. 209–256). Beverly Hills, CA: SAGE Publications.

Nuechterlein, K., Green, M., Kern, R., Baade, L., Barch, D., Cohen, J., … Marder, S. (2008). The matrics consensus cognitive battery, part 1: Test selection, reliability, and validity. *American Journal of Psychiatry*, *165*(2), 203–213. doi:10.1176/appi.ajp.2007.07010042

O'Brien, R. G. & Kaiser, M. K. (1985). Manova method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin*, *97*, 316–333.

Rousseeuw, P. & Leroy, A. (1987). *Robust regression and outlier detection*. New York: John Wiley; Sons.

Rousseeuw, P. & Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, *41*, 212–223.

Teetor, P. (2011). *R cookbook*. Sebastopol, CA: O'Reilly Media.

Timm, N. H. (1975). *Multivariate analysis with applications in education and psychology*. Belmont, CA: Wadsworth (Brooks/Cole.

Timm, N. H. (1980). Multivariate analysis of variance of repeated measurements. In P. R. Krishnaiah (Ed.), *Handbook of statistics (vol. 1)* (pp. 41–87). Amsterdam: North-Holland.

Vu, V. Q. (2011). Ggbiplot: A ggplot2 based biplot (Version 1.0). Retrieved from http://github.com/vqv/ggbiplot

Wickham, H. (2014). *Advanced R*. Boca Raton, FL: Chapman & Hall/CRC.

Xie, Y. (2015). Knitr: A general-purpose package for dynamic report generation in R (Version 1.0). Retrieved from http://CRAN.R-project.org/package=knitr

Appendix A: Sample Code for Multivariate Analyses

The following block contains annotated R code that produces the core visualizations and statistical analyses featured in this paper. These excerpts provide an excellent starting point for working with such data. The full script, including code that enhances the graphical output for publication, is available online.

```r
# Core Packages
library("car")     # for Anova()
library("heplots") # for HE plots framework and datasets
library("candisc") # for Canonical Discrimination analysis

## MULTIVARIATE MULTIPLE REGRESSION #####
data("Rohwer", package="heplots")
Rohwer2 <- subset(Rohwer, subset = group==2) # subset the high SES children

rohwer.mlm <- lm(cbind(SAT, PPVT, Raven) ~ n + s + ns + na + ss, data = Rohwer2)
Anova(rohwer.mlm)  #  Model MANOVA tests

# Evaluate and print multivariate test
print(}linearHypothesis(rohwer.mlm, c("n", "s", "ns", "na", "ss")), SSP=FALSE)

# HE scatterplot matrix
hyp <- list("Regr" = c("n", "s", "ns", "na", "ss"))
pairs(rohwer.mlm, hypotheses = hyp)

## MANOVA #####
data(NeuroCog, package="heplots")
NC.mlm <- lm(cbind(Speed, Attention, Memory, Verbal, Visual, ProbSolv) ~ Dx,
      data = NeuroCog)
Anova(NC.mlm) # Model MANOVA tests
cqplot(NC.mlm) # Model checking
```
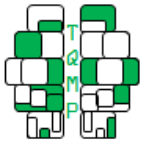
```r
# Evaluate linear contrasts:
print(linearHypothesis(NC.mlm, "Dx1"), SSP=FALSE)
print(linearHypothesis(NC.mlm, "Dx2"), SSP=FALSE)

heplot(NC.mlm, variables=1:2)

## CANONICAL DISCRIMINATION #####

NC.can <- candisc(NC.mlm)
NC.can # View canonical model

plot(NC.can) # Basic CD plot
heplot(NC.can,  # HE plot of canonical analysis
       fill=TRUE, fill.alpha=.1)

## MANCOVA #####

rohwer.mod <- lm(cbind(SAT, PPVT, Raven) ~ SES + n + s + ns + na + ss,
       data = Rohwer) # Main effects model

pairs(rohwer.mod, # Basic pairwise HE plot
       hypotheses = list("Regr" = c("n", "s", "ns", "na", "ss")),
       fill = TRUE, fill.alpha = 0.1)

rohwer.mod1 <- lm(cbind(SAT, PPVT, Raven) ~ SES * (n + s + ns + na + ss),
       data = Rohwer) # Interaction model
Anova(rohwer.mod1) # View model coefficients

coefs <- rownames(coef(rohwer.mod1)) # store coefficient names in a vector
print(linearHypothesis(rohwer.mod1,    # only test for interaction effects
       coefs[grep(":", coefs)]), SSP = FALSE)

# Fit separate models by SES level:
rohwer.ses1 <- lm(cbind(SAT, PPVT, Raven) ~ n + s + ns + na + ss,
       data = Rohwer, subset = SES == "Hi")

rohwer.ses2 <- lm(cbind(SAT, PPVT, Raven) ~ n + s + ns + na + ss,
       data = Rohwer, subset = SES == "Lo")

# Low SES students:
heplot(rohwer.ses2, col = c("red", rep("black",5), "blue"),
        hypotheses = list("B=0, Low SES" = c("n", "s", "ns", "na", "ss")),
        level = 0.5, cex = 1.25,
        fill = c(TRUE, FALSE), fill.alpha = 0.05,
        xlim = c(-15, 110), ylim = c(40, 110),
        xlab = "Student Achievement Test",
        ylab = "Peabody Picture Vocabulary Test",
        label.pos = c(1, rep(NULL, 5), 1))
```
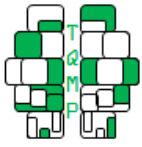
```
# High SES students:
heplot(rohwer.ses1, col = c("red", rep("black", 5), "blue"),
       hypotheses = list("B=0, High SES" = c("n", "s", "ns", "na", "ss")),
       level = 0.5, cex = 1.25,
       add = TRUE,   # place both plots on same graphic
       error = TRUE, # error ellipse is not drawn by default with add = TRUE
       fill = c(TRUE, FALSE), fill.alpha = 0.05,
       xlim = c(-15, 110), ylim = c(40, 110))
```

Open practices

⬣ The *Open Material* badge was earned because supplementary material(s) are available on the journal's web site.

Citation

Friendly, M. & Sigal, M. (2017). Graphical methods for multivariate linear models in psychological research: An R tutorial. *The Quantitative Methods for Psychology*, *13*(1), 20–45. doi:10.20982/tqmp.13.1.p020