Journal of Statistical Software

November 2010, Volume 37, Issue 4. *http://www.jstatsoft.org/*

HE Plots for Repeated Measures Designs

Michael Friendly
York University

Abstract

Hypothesis error (HE) plots, introduced in Friendly (2007), provide graphical methods to visualize hypothesis tests in multivariate linear models, by displaying hypothesis and error covariation as ellipsoids and providing visual representations of effect size and significance. These methods are implemented in the **heplots** for R (Fox, Friendly, and Monette 2009a) and SAS (Friendly 2006), and apply generally to designs with fixed-effect factors (MANOVA), quantitative regressors (multivariate multiple regression) and combined cases (MANCOVA).

This paper describes the extension of these methods to repeated measures designs in which the multivariate responses represent the outcomes on one or more "within-subject" factors. This extension is illustrated using the **heplots** for R. Examples describe one-sample profile analysis, designs with multiple between-S and within-S factors, and doubly-multivariate designs, with multivariate responses observed on multiple occasions.

Keywords: data ellipse, HE plot, HE plot matrix, profile analysis, repeated measures, MANOVA, doubly-multivariate designs, mixed models.

1. Introduction

Hypothesis error (HE) plots, introduced in Friendly (2007), provide graphical methods to visualize hypothesis tests in multivariate linear models, by displaying hypothesis and error covariation as ellipsoids and providing visual representations of effect size and significance. The **heplots** (Fox *et al.* 2009a) for R (R Development Core Team 2010) implements these methods for the general class of the multivariate linear model (MVLM) including fixed-effect factors (MANOVA), quantitative regressors (multivariate multiple regression, MMREG) and combined cases (MANCOVA). Here, we describe the extension of these methods to repeated measures designs in which the multivariate responses represent the outcomes on one or more "within-subject" factors.

1.1. Multivariate linear models: Notation

To set notation, we express the MVLM as

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

where, $\boldsymbol{Y} \equiv (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_p)$ is the matrix of responses for n subjects on p variables, $\boldsymbol{X}$ is the design matrix for q regressors, $\boldsymbol{B}$ is the $q \times p$ matrix of regression coefficients or model parameters and $\boldsymbol{U}$ is the $n \times p$ matrix of errors, with $\text{vec}(\boldsymbol{U}) \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$, where $\otimes$ is the Kronecker product.

A convenient feature of the MVLM for general multivariate responses is that *all* tests of linear hypotheses (for null effects) can be represented in the form of a general linear test,

$$H_0 : \underset{(h\times q)}{\boldsymbol{L}}\,\underset{(q\times p)}{\boldsymbol{B}} = \underset{(h\times p)}{\boldsymbol{0}} \quad , \tag{2}$$

where $\boldsymbol{L}$ is a matrix of constants whose rows specify h linear combinations or contrasts of the parameters to be tested simultaneously by a multivariate test. In R all such tests can be carried out using the functions `Anova()` and `linear.hypothesis()` in the **car**.[1]

For *any* such hypothesis of the form Equation 2, the analogs of the univariate sums of squares for hypothesis (SS_H) and error (SS_E) are the $p \times p$ sum of squares and crossproducts (SSP) matrices given by (Timm 1975, Chapters 3, 5):

$$\boldsymbol{H} \equiv \boldsymbol{SSP}_H = (\boldsymbol{L}\widehat{\boldsymbol{B}})^\top [\boldsymbol{L}(\boldsymbol{X}^\top\boldsymbol{X})^- \boldsymbol{L}^\top]^{-1} (\boldsymbol{L}\widehat{\boldsymbol{B}}) \ , \tag{3}$$

and

$$\boldsymbol{E} \equiv \boldsymbol{SSP}_E = \boldsymbol{Y}^\top\boldsymbol{Y} - \widehat{\boldsymbol{B}}^\top(\boldsymbol{X}^\top\boldsymbol{X})\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{U}}^\top\widehat{\boldsymbol{U}} \ , \tag{4}$$

where $\widehat{\boldsymbol{U}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}$ is the matrix of residuals. Multivariate test statistics (Wilks' Λ, Pillai trace, Hotelling-Lawley trace, Roy's maximum root) for testing Equation 2 are based on the $s = \min(p, h)$ non-zero latent roots of $\boldsymbol{HE}^{-1}$ and attempt to capture how "large" $\boldsymbol{H}$ is, relative to $\boldsymbol{E}$ in s dimensions. All of these statistics have transformations to F statistics giving either exact or approximate null hypothesis F distributions. The corresponding latent vectors give a set of s orthogonal linear combinations of the responses that produce maximal univariate F statistics for the hypothesis in Equation 2; we refer to these as the canonical discriminant dimensions.

In a univariate, fixed-effects linear model, it is common to provide F tests for each term in the model, summarized in an analysis-of-variance (ANOVA) table. The hypothesis sums of squares, SS_H, for these tests can be expressed as differences in the error sums of squares, SS_E, for nested models. For example, dropping each term in the model in turn and contrasting the resulting residual sum of squares with that for the full model produces so-called Type-III tests; adding terms to the model sequentially produces so-called Type-I tests; and testing each term after all terms in the model with the exception of those to which it is marginal produces so-called Type-II tests. Closely analogous MANOVA tables can be formed similarly by taking

[1] Both the **car** and the **heplots** are being actively developed. Except where noted, all results in this paper were produced using the old-stable versions on the Comprehensive R Archive Network (CRAN) at `http://CRAN.R-project.org/`, **car** 1.2-16 (2009-10-10) and **heplots** 0.8-11 (2009-12-08) running under R version 2.11.1 (2010-05-31).

differences in error sum of squares and products matrices ($\boldsymbol{E}$) for such nested models. Type I tests are sensible only in special circumstances; in balanced designs, Type II and Type III tests are equivalent. Regardless, the methods illustrated in this paper apply to any multivariate linear hypothesis.

1.2. Data ellipses and ellipsoids

In what follows, we make extensive use of ellipses (or ellipsoids in 3+D) to represent joint variation among two or more variables, so we define this here. The *data ellipse* (or covariance ellipse), described by Dempster (1969) and Monette (1990), is a device for visualizing the relationship between two variables, Y_1 and Y_2. Let $D^2_M(\boldsymbol{y}) = (\boldsymbol{y} - \overline{\boldsymbol{y}})^\top \boldsymbol{S}^{-1} (\boldsymbol{y} - \overline{\boldsymbol{y}})$ represent the squared Mahalanobis distance of the point $\boldsymbol{y} = (y_1, y_2)^\top$ from the centroid of the data $\overline{\boldsymbol{y}} = (\overline{Y}_1, \overline{Y}_2)^\top$. The data ellipse $\mathcal{E}_c$ of size c is the set of all points $\boldsymbol{y}$ with $D^2_M(\boldsymbol{y})$ less than or equal to c^2:

$$\mathcal{E}_c(\mathbf{y}; \mathbf{S}, \overline{\boldsymbol{y}}) \equiv \left\{ y\colon\ (\boldsymbol{y} - \overline{\boldsymbol{y}})^\top \boldsymbol{S}^{-1} (\boldsymbol{y} - \overline{\boldsymbol{y}}) \leq c^2 \right\} \tag{5}$$

Here, $\boldsymbol{S} = \sum_{i=1}^n (\boldsymbol{y} - \overline{\boldsymbol{y}})^\top (\boldsymbol{y} - \overline{\boldsymbol{y}}) / (n-1) = \widehat{\mathsf{Var}}(\boldsymbol{y})$ is the sample covariance matrix.

Many properties of the data ellipse hold regardless of the joint distribution of the variables, but if the variables are bivariate normal, then the data ellipse represents a contour of constant density in their joint distribution. In this case, $D^2_M(\boldsymbol{y})$ has a large-sample χ^2 distribution with 2 degrees of freedom, and so, for example, taking $c^2 = \chi^2_2(0.95) = 5.99 \approx 6$ encloses approximately 95 percent of the data. Taking $c^2 = \chi^2_2(0.68) = 2.28$ gives a bivariate analog of the univariate ± 1 standard deviation interval, enclosing approximately 68% of the data.

The generalization of the data ellipse to more than two variables is immediate: Applying Equation 5 to $\boldsymbol{y} = (y_1, y_2, y_3)^\top$, for example, produces a data ellipsoid in three dimensions. For p multivariate-normal variables, selecting $c^2 = \chi^2_p(1-\alpha)$ encloses approximately $100(1-\alpha)$ percent of the data.[2]

1.3. HE plots

The essential idea behind HE plots is that any multivariate hypothesis test Equation 2 can be represented visually by ellipses (or ellipsoids in 3D) which express the size of co-variation against a multivariate null hypothesis ($\boldsymbol{H}$) relative to error covariation ($\boldsymbol{E}$). The multivariate tests, based on the latent roots of $\boldsymbol{H}\boldsymbol{E}^{-1}$, are thus translated directly to the sizes of the $\boldsymbol{H}$ ellipses for various hypotheses, relative to the size of the $\boldsymbol{E}$ ellipse. Moreover, the shape and orientation of these ellipses show something more– the directions (linear combinations of the responses) that lead to various effect sizes and significance.

In these plots, the $\boldsymbol{E}$ matrix is first scaled to a covariance matrix ($\boldsymbol{E}/df_e = \widehat{\mathsf{Var}}(\boldsymbol{U}_i)$). The ellipse drawn (translated to the centroid $\overline{\boldsymbol{y}}$ of the variables) is thus the data ellipse of the residuals, reflecting the size and orientation of residual variation. In what follows (by default), we always show these as "standard" ellipses of 68% coverage. This scaling and translation also allows the means for levels of the factors to be displayed in the same space, facilitating interpretation.

[2] Robust versions of data ellipses (e.g., based on minimum volume ellipsoid, MVE, or minimum covariance determinant, MCD, estimators of $\boldsymbol{S}$) are also available, as are small-sample approximations to the enclosing c^2 radii, but these refinements are outside the scope of this paper.

The ellipses for $\boldsymbol{H}$ reflect the size and orientation of covariation against the null hypothesis. They always proportional to the data ellipse of the fitted effects (predicted values) for a given hypothesized term. In relation to the $\boldsymbol{E}$ ellipse, the $\boldsymbol{H}$ ellipses can be scaled to show either the *effect size* or strength of *evidence* against H_0 (significance).

For effect size scaling, each $\boldsymbol{H}$ is divided by df_e to conform to $\boldsymbol{E}$. The resulting ellipses are then exactly the data ellipses of the fitted values, and correspond visually to multivariate analogs of univariate effect size measures (e.g., $(\bar{y_1} - \bar{y_2})/s$ where $s =$ within group standard deviation). That is, the sizes of the $\boldsymbol{H}$ ellipses relative to that of the $\boldsymbol{E}$ reflect the (squared) differences and correlation of the factor means relative to error covariation.

For significance scaling, it turns out to be most visually convenient to use Roy's largest root test as the test criterion. In this case the $\boldsymbol{H}$ ellipse is scaled to $\boldsymbol{H}/(\lambda_\alpha df_e)$ where λ_α is the critical value of Roy's statistic. Using this gives a simple visual test of H_0: Roy's test rejects H_0 at a given α level if and only if the corresponding α-level $\boldsymbol{H}$ ellipse extends anywhere outside the $\boldsymbol{E}$ ellipse.[3] Consequently, when the rank of $\boldsymbol{H} = \min(p, h) \leq 2$, all significant effects can be observed directly in 2D HE plots; when $\text{rank}(\boldsymbol{H}) = 3$, some rotation of a 3D plot will reveal each significant effect as extending somewhere outside the $\boldsymbol{E}$ ellipsoid.

In our R implementation, the basic plotting functions in the **heplots** are `heplot()` and `heplot3d()` for `mlm` objects. These rely heavily on the `Anova()` and other functions from the **car** (Fox and Weisberg 2009) for computation. For more than three response variables, all pairwise HE plots can be shown using a `pairs()` function for `mlm` objects. Alternatively, the related **candisc** (Friendly and Fox 2010) produces HE plots in canonical discriminant space. This shows a low-rank 2D (or 3D) view of the effects for a given term in the space of maximum discrimination, based on the linear combinations of responses which produce maximally significant test statistics. See Friendly (2007); Fox, Friendly, and Monette (2009b) for details and examples for between-S MANOVA designs, MMREG and MANCOVA models.

2. Repeated measures designs

The framework for the MVLM described above pertains to the situation in which the response vectors (rows, $\boldsymbol{y}_i^\top$ of $\boldsymbol{Y}_{n\times p}$) are *iid* and the p responses are separate, not necessarily commensurate variables observed on individual i.

In principle, the MVLM extends quite elegantly to repeated-measure (or within-subject) designs, in which the p responses per individual can represent the factorial combination of one or more factors that structure the response variables in the same way that the between-individual design structures the observations. In the multivariate approach to repeated measure data, the same model Equation 1 applies, but hypotheses about between- and within-individual variation are tested by an extended form of the general linear test Equation 2, which becomes

$$H_0 : \underset{(h\times q)}{\boldsymbol{L}}\ \underset{(q\times p)}{\boldsymbol{B}}\ \underset{(p\times k)}{\boldsymbol{M}} = \underset{(h\times k)}{\boldsymbol{0}} \quad , \tag{6}$$

where $\boldsymbol{M}$ is a matrix of constants whose columns specify k linear combinations or contrasts among the responses, corresponding to a particular within-individual effect. In this case, the

[3] Other multivariate tests (Wilks' Λ, Hotelling-Lawley trace, Pillai trace) also have geometric interpretations in HE plots (e.g., Wilks' Λ is the ratio of areas – volumes – of the $\boldsymbol{H}$ and $\boldsymbol{E}$ ellipses – ellipsoids), but these statistics do not provide such simple visual comparisons. All HE plots shown in this paper use significance scaling, based on Roy's test.

$\boldsymbol{M}$ for within-S effects	Between-S effect tested			
	Intercept	$\boldsymbol{L} = \boldsymbol{L}_A$	$\boldsymbol{L} = \boldsymbol{L}_B$	$\boldsymbol{L} = \boldsymbol{L}_{AB}$
$\boldsymbol{M} = \boldsymbol{M}_1 = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}^\top$	$\boldsymbol{\mu}_{\cdot\cdot}$	A	B	A:B
$\boldsymbol{M} = \boldsymbol{M}_C = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}^\top$	C	A:C	B:C	A:B:C

Table 1: Three-way design: Tests for between- (A, B) and within-S (C) effects are constructed using various $\boldsymbol{L}$ and $\boldsymbol{M}$ matrices. Table entries give the term actually tested via the general linear test in Equation 6.

$\boldsymbol{H}$ and $\boldsymbol{E}$ matrices for testing Equation 6 become

$$\boldsymbol{H} = (\boldsymbol{L}\widehat{\boldsymbol{B}}\boldsymbol{M})^\top [\boldsymbol{L}(\boldsymbol{X}^\top\boldsymbol{X})^-\boldsymbol{L}^\top]^{-1} (\boldsymbol{L}\widehat{\boldsymbol{B}}\boldsymbol{M}) \ , \tag{7}$$

and

$$\boldsymbol{E} = (\boldsymbol{Y}\boldsymbol{M})^\top [\boldsymbol{I} - (\boldsymbol{X}^\top\boldsymbol{X})^-\boldsymbol{X}^\top] (\boldsymbol{Y}\boldsymbol{M}) \ . \tag{8}$$

This may be easily seen to be just the ordinary MVLM applied to the transformed responses $\boldsymbol{Y}\boldsymbol{M}$ which form the basis for a given within-individual effect. The idea for this approach to repeated measures through a transformation of the responses was first suggested by Hsu (1938) and is discussed further by Rencher (1995) and Timm (1980). In what follows, we refer to hypotheses pertaining to between-individual effects (specified by $\boldsymbol{L}$) as "between-S" and hypotheses pertaining to within-individual effects ($\boldsymbol{M}$) as "within-S."

In the general case, various $\boldsymbol{L}$ matrices provide contrasts or select the particular coefficients tested for between-S effects, while various $\boldsymbol{M}$ matrices specify linear combinations of responses for the within-S effects. This is illustrated in Table 1 for a three-way design with two between-S factors (A, B) and one within-S factor (C).

The between-S terms themselves are tested using the unit vector $\boldsymbol{M} = (\mathbf{1}_p)$, giving a test based on the sums over the within-S effects. This simply reflects the principle of marginality, by which effects for any term in a linear model are tested by averaging over all factors not included in that term. Tests using a matrix $\boldsymbol{M}$ of contrasts for a within-S effect provide tests of the interactions of that effect with each of the between-S terms. That is, $\boldsymbol{L}\boldsymbol{B}\boldsymbol{M} = \mathbf{0}$ tests between-S *differences* among the responses transformed by $\boldsymbol{M}$.

For more than one within-S factor, the full $\boldsymbol{M}$ matrices for various within-S terms are generated as Kronecker products of the one-way $\boldsymbol{M}$ contrasts with the unit vector ($\mathbf{1}$) of appropriate size. For example, with c levels of factor C and d of factor D,

$$\begin{aligned} \boldsymbol{M}_{C\otimes D} &= (\mathbf{1}_c, \boldsymbol{M}_C) \otimes (\mathbf{1}_d, \boldsymbol{M}_D) \\ &= (\mathbf{1}_c \otimes \mathbf{1}_d, \mathbf{1}_c \otimes \boldsymbol{M}_D, \boldsymbol{M}_C \otimes \mathbf{1}_d, \boldsymbol{M}_C \otimes \boldsymbol{M}_D) \\ &= (\boldsymbol{M}_1, \boldsymbol{M}_D, \boldsymbol{M}_C, \boldsymbol{M}_{CD}) \ . \end{aligned} \tag{9}$$

Each of the within-S terms combine with any between-S terms in an obvious way to give an extended version of Table 1 with additional rows for $\boldsymbol{M}_D$ and $\boldsymbol{M}_{CD}$.

In passing, we note that *all* software (SAS, SPSS, R, etc.) that handles repeated measure designs through this extension of the MLM effectively works via the general linear test Equation 2, with either implicit or explicit specifications for the $\boldsymbol{L}$ and $\boldsymbol{M}$ matrices involved in

testing any hypothesis for between- or within-S effects. This mathematical elegance is not without cost, however. The MLM approach does not allow for missing data (a particular problem in longitudinal designs), and the multivariate test statistics (Wilks' Λ, etc.) assume the covariance matrix of $\boldsymbol{U}$ is unstructured. Alternative analysis based on mixed (or hierarchical) models (e.g., Pinheiro and Bates 2000; Verbeke and Molenberghs 2000) are more general in some ways, but to date visualization methods for this approach remain primitive and the mixed model analysis does not easily accommodate multivariate responses.

The remainder of the paper illustrates these MLM analyses, shows how they may be performed in R, and how HE plots can be used to provide visual displays of what is summarized in the multivariate test statistics. We freely admit that these displays are somewhat novel and take some getting used to, and so this paper takes a more tutorial tone. We exemplify these methods in the context of simple, one-sample profile analysis (Section 3), designs with multiple between- and within-S effects (Section 4), and doubly-multivariate designs (Section 5), where two or more separate responses (e.g., weight loss and self esteem) are *each* observed in a factorial structure over multiple within-S occasions. In Section 6 we describe a simplified interface for these plots in the development versions of the **heplots** and **car** packages. Finally (Section 7) we compare these methods with visualizations based on the mixed model.

3. One sample profile analysis

The simplest case of a repeated-measures design is illustrated by the data on vocabulary growth of school children from grade 8 to grade 11, in the data frame `VocabGrowth`, recording scores on the vocabulary section of the Cooperative Reading Test for a cohort of 64 students. (The scores are scaled to a common, but arbitrary origin and unit of measurement, so as to be comparable over the four grades.) Since these data cover an age range in which physical growth is beginning to decelerate, it is of interest whether a similar effect occurs in the acquisition of new vocabulary. Thus, attention here is arguably directed to polynomial trends in grade: average rate of change (slope, or linear trend) and shape of trajectories (quadratic and cubic components).

```
R> some(VocabGrowth, 5)

   grade8 grade9 grade10 grade11
11  -0.95   0.41    0.21    1.82
42   1.03   2.10    3.88    2.81
49   1.10   2.65    1.72    2.96
56  -2.19  -0.42    1.54    1.16
60  -0.29   2.62    1.60    1.86
```

A boxplot of these scores (Figure 1) gives an initial view of the data. To do this, we first reshape the data from wide to long format (i.e., each 4-variate row becomes four rows indexed by `grade`). We can see that vocabulary scores increase with age, but the trend of means appears non-linear.

```
R> voc <- reshape(VocabGrowth, direction = "long",
+    varying = list(grade = 1:4), timevar = "Grade", v.names = "Vocabulary")
```

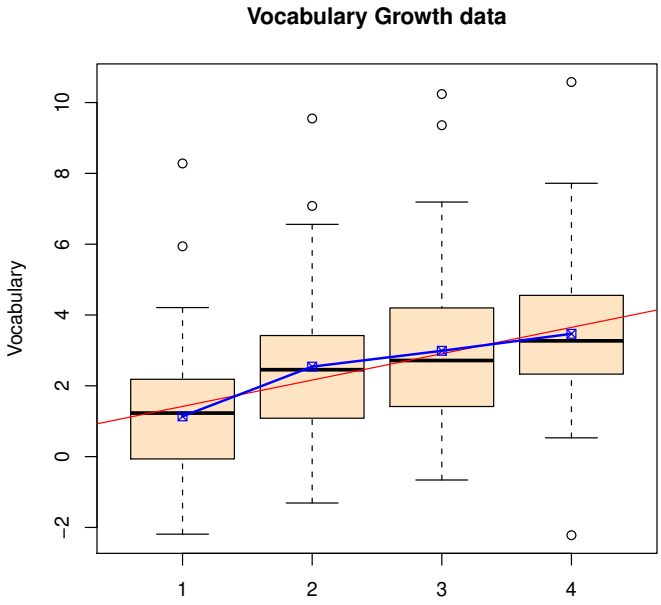

Figure 1: Boxplots of vocabulary score by grade, with linear regression line (red) and lines connecting grade means (blue).

```
R> boxplot(Vocabulary ~ Grade, data = voc, col = "bisque",
+    ylab = "Vocabulary", main = "Vocabulary Growth data")
R> abline(lm(Vocabulary ~ as.numeric(Grade), data = voc), col = "red")
R> means <- tapply(voc$Vocabulary, voc$Grade, mean)
R> points(1:4, means, pch = 7, col = "blue")
R> lines(1:4, means, col = "blue", lwd = 2)
```

The standard univariate and multivariate tests for the differences in vocabulary with grade can be carried out as follows. First, we fit the basic MVLM with an intercept only on the right-hand side of the model, since there are no between-S effects. The intercepts estimate the means at each grade level, $\mu_8, \ldots, \mu_{11}$.

```
R> (Vocab.mod <- lm(cbind(grade8, grade9, grade10, grade11) ~ 1,
+    data = VocabGrowth))

Call:
lm(formula = cbind(grade8, grade9, grade10, grade11) ~ 1, data = VocabGrowth)

Coefficients:
             grade8  grade9  grade10  grade11
(Intercept)  1.14    2.54    2.99     3.47
```

We could test the multivariate hypothesis that all means are simultaneously zero, $\mu_8 = \mu_9 =$

$\mu_{10} = \mu_{11} = 0$. This point hypothesis is the simplest case of a multivariate test under Equation 2, with $\boldsymbol{L} = \boldsymbol{I}$.

```
R> (Vocab.aov0 <- Anova(Vocab.mod, type = "III"))

Type III MANOVA Tests: Pillai test statistic
            Df test stat approx F num Df den Df Pr(>F)
(Intercept)  1    0.8577    90.38      4     60 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This hypothesis tests that the vocabulary means are all at the arbitrary origin for the scale. Often this test is not of direct interest, but it serves to illustrate the $\boldsymbol{H}$ and $\boldsymbol{E}$ matrices involved in any multivariate test, their representation by HE plots, and how we can extend these plots to the repeated measures case.

The $\boldsymbol{H}$ and $\boldsymbol{E}$ matrices can be printed with `summary(Vocab.aov0)`, or extracted from the `Anova.mlm` object. In this case, $\boldsymbol{H}$ is simply $n\overline{\boldsymbol{y}}\overline{\boldsymbol{y}}^\top$ and $\boldsymbol{E}$ is the sum of squares and crossproducts of deviations from the column means, $\sum_{i=1}^{n}(\boldsymbol{y}_i - \overline{\boldsymbol{y}})^\top(\boldsymbol{y}_i - \overline{\boldsymbol{y}})$.

```
R> Vocab.aov0$SSP

$`(Intercept)`
         grade8  grade9 grade10 grade11
grade8   82.810 185.037 217.547 252.525
grade9  185.037 413.461 486.104 564.262
grade10 217.547 486.104 571.509 663.398
grade11 252.525 564.262 663.398 770.062

R> Vocab.aov0$SSPE

         grade8  grade9 grade10 grade11
grade8  225.086 201.133 223.843 179.950
grade9  201.133 273.850 223.515 191.729
grade10 223.843 223.515 296.321 213.249
grade11 179.950 191.729 213.249 233.848
```

The HE plot for the `Vocab.mod` model shows the test for the `(Intercept)` term (all means = 0). To emphasize that the test is assessing the (squared) distance of $\bar{\boldsymbol{y}}$ from $\mathbf{0}$, in relation to the covariation of observations around the grand mean, we define a simple function to mark the point hypothesis $H_0 = (0, 0)$.

```
R> mark.H0 <- function(x = 0, y = 0, cex = 2, pch = 19, col = "green3",
+    lty = 2, pos = 2)
+  {
+    points(x, y, cex = cex, col = col, pch = pch)
+    text(x, y, expression(H[0]), col = col, pos = pos)
+    if (lty > 0) abline(h = y, col = col, lty = lty)
+    if (lty > 0) abline(v = x, col = col, lty = lty)
+  }
```

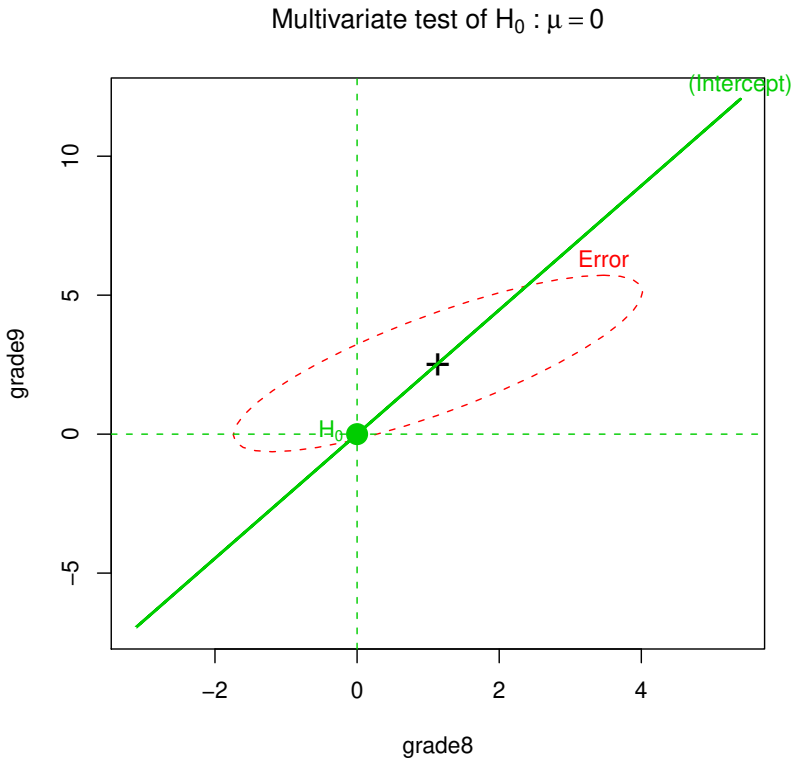

Figure 2: HE plot for vocabulary data, for the MANOVA test of $H_0 : \boldsymbol{\mu}_y = 0$. The size of the (degenerate) ellipse for the intercept term relative to that for error gives the strength of evidence for the difference between the sample means (marked by +) and the means under H_0 (marked by the cross-hairs and green dot). The projection of this $\boldsymbol{H}$ ellipse outside the $\boldsymbol{E}$ ellipse signals that this H_0 is clearly rejected.

Here we show the HE plot for the `grade8` and `grade9` variables in Figure 2. The $\boldsymbol{E}$ ellipse reflects the positive correlation of vocabulary scores across these two grades, but also shows that variability is greater in grade 8 than in grade 9. Its position relative to (0, 0) indicates that both means are positive, with a larger mean at grade 9 than grade 8.

```
R> heplot(Vocab.mod, terms = "(Intercept)", type = "III")
R> mark.H0(0, 0)
R> title(expression(paste("Multivariate test of ", H[0], " : ",
+    bold(mu) == 0)))
```

The $\boldsymbol{H}$ ellipse plots as a degenerate line because the $\boldsymbol{H}$ matrix has rank 1 (1 df for the MANOVA test of the intercept). The fact that the $\boldsymbol{H}$ ellipse extends outside the $\boldsymbol{E}$ ellipse (anywhere) signals that this H_0 is clearly rejected (for some linear combination of the response variables). Moreover, the projections of the $\boldsymbol{H}$ and $\boldsymbol{E}$ ellipses on the `grade8` and `grade9` axes, showing $\boldsymbol{H}$ widely outside $\boldsymbol{E}$, signals that the corresponding univariate hypotheses, $\mu_8 = 0$ and $\mu_9 = 0$ would also be rejected.

3.1. Testing within-S effects

For the `Anova()` function, the model for within-S effects— giving rise the $\boldsymbol{M}$ matrices in Equation 6— is specified through the arguments `idata` (a data frame giving the factor(s) used in the intra-subject model) and `idesign` (a model formula specifying the intra-subject design). That is, if $\boldsymbol{Z} = [\texttt{idata}]$, the $\boldsymbol{M}$ matrices for different intra-subject terms are generated from columns of $\boldsymbol{Z}$ indexed by the terms in `idesign`, with factors and interactions expanded expanded to contrasts in the same way that the design matrix $\boldsymbol{X}$ is generated from the between-S design formula.

Thus, to test the within-S effect of grade, we need to construct a `grade` variable for the levels of grade and use this as a model formula, `idesign = ~ grade` to specify the within-S design in the call to `Anova`.

```
R> idata <- data.frame(grade = ordered(8:11))
R> (Vocab.aov <- Anova(Vocab.mod, idata = idata, idesign = ~ grade,
+    type = "III"))

Type III Repeated Measures MANOVA Tests: Pillai test statistic
            Df test stat approx F num Df den Df   Pr(>F)
(Intercept)  1    0.6529   118.50      1     63 4.12e-16 ***
grade        1    0.8258    96.38      3     61  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As shown in Table 1, any such within-S test is effectively carried out using a transformation $\boldsymbol{Y}$ to $\boldsymbol{YM}$, where the columns of $\boldsymbol{M}$ provide contrasts among the grades. For the *overall* test of `grade`, *any* set of 3 linearly independent contrasts will give the same test statistics, though, of course the interpretation of the parameters will differ. Specifying `grade` as an ordered factor (`grade = ordered(8:11)`) will cause `Anova()` to use the polynomial contrasts shown in $\boldsymbol{M}_{\text{poly}}$ below.

$$\boldsymbol{M}_{\text{poly}} = \begin{pmatrix} -3 & 1 & -1 \\ -1 & -1 & 3 \\ 1 & -1 & -3 \\ 3 & 1 & 1 \end{pmatrix} \quad \boldsymbol{M}_{\text{first}} = \begin{pmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \boldsymbol{M}_{\text{last}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{pmatrix}$$

Alternatively, $\boldsymbol{M}_{\text{first}}$ would test the gains in vocabulary between grade 8 (baseline) and each of grades 9–11, while $\boldsymbol{M}_{\text{last}}$ would test the difference between each of grades 8–10 from grade 11. (In R, these contrasts are constructed with `M.first <- contr.sum(factor(11:8))[4:1, 3:1]`, and `M.last <- contr.sum(factor(8:11))` respectively.) In all cases, the hypothesis of no difference among the means across grade is transformed to an equivalent multivariate point hypothesis, $\boldsymbol{M}\boldsymbol{\mu}_y = \boldsymbol{0}$, such as we visualized in Figure 2.

Correspondingly, the HE plot for the effect of grade can be constructed as follows. For expository purposes we explicitly transform $\boldsymbol{Y}$ to $\boldsymbol{YM}$, where the columns of $\boldsymbol{M}$ provide contrasts among the grades reflecting linear, quadratic and cubic trends using $\boldsymbol{M}_{\text{poly}}$.

Using $\boldsymbol{M}_{\text{poly}}$, the MANOVA test for the `grade` effect is then testing $H_0 : \boldsymbol{M}\boldsymbol{\mu}_y = \boldsymbol{0} \leftrightarrow \mu_{\text{Lin}} = \mu_{\text{Quad}} = \mu_{\text{Cubic}} = 0$. That is, the means across grades 8–11 are equal if and only if their linear, quadratic and cubic trends are simultaneously zero.

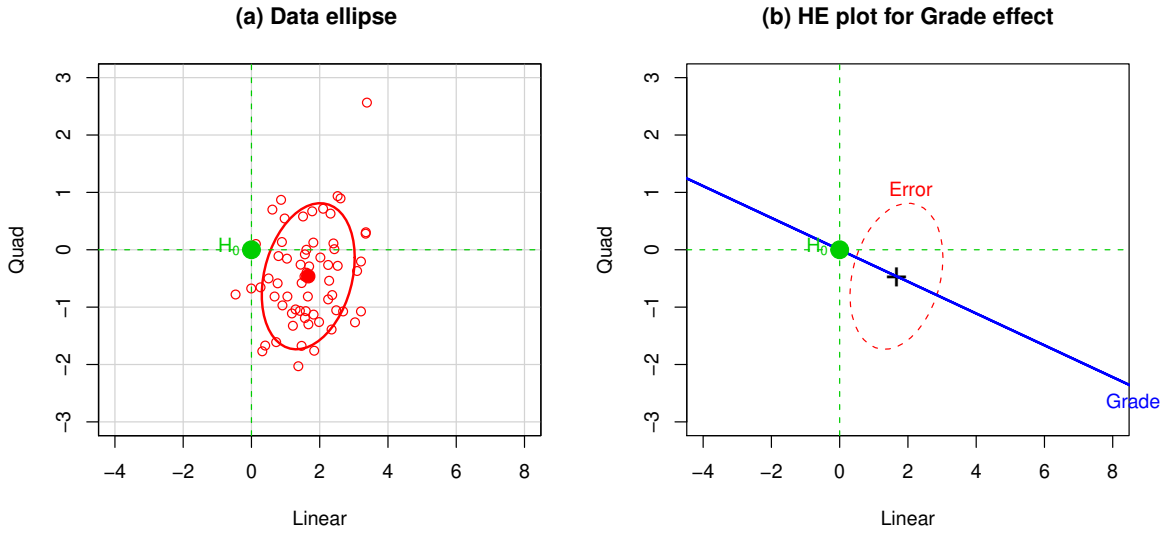

Figure 3: Plots of linear and quadratic trend scores for the vocabulary data. (a) Scatterplot with 68% data ellipse; (b) HE plot for the effect of grade. As in Figure 2, the size of the $\boldsymbol{H}$ ellipse for grade relative to the $\boldsymbol{E}$ ellipse shows the strength of evidence against H_0.

```
R> trends <- as.matrix(VocabGrowth) %*% poly(8:11, degree = 3)
R> colnames(trends) <- c("Linear", "Quad", "Cubic")
R> within.mod <- lm(trends ~ 1)
R> Anova(within.mod, type = "III")

Type III MANOVA Tests: Pillai test statistic
            Df test stat approx F num Df den Df Pr(>F)
(Intercept)  1    0.8258    96.38      3     61 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It is easily seen that the test of the `(Intercept)` term in `within.mod` is identical to the test of `grade` in `Vocab.mod` at the beginning of this subsection.

We can show this test visually as follows. Figure 3(a) shows a scatterplot of the transformed linear and quadratic trend scores, overlayed with a 68% data ellipse. Figure 3(b) is the corresponding HE plot for these two variables. Thus, we can see that the $\boldsymbol{E}$ ellipse is simply the data ellipse of the transformed vocabulary scores; its orientation indicates a slight tendency for those with greater slopes (gain in vocabulary) to have greater curvatures (leveling off earlier). Figure 3 is produced as follows:

```
R> op <- par(mfrow = c(1, 2))
R> data.ellipse(trends[, 1:2], xlim = c(-4, 8), ylim = c(-3, 3),
+    levels = 0.68, main = "(a) Data ellipse ")
R> mark.H0(0, 0)
R> heplot(within.mod, terms = "(Intercept)", col = c("red", "blue"),
+    type = "III", term.labels = "Grade", , xlim = c(-4, 8),
+    ylim = c(-3, 3), main = "(b) HE plot for Grade effect")
```

```
R> mark.H0(0, 0)
R> par(op)
```

We interpret Figure 3(b) as follows, bearing in mind that we are looking at the data in the transformed space ($\boldsymbol{YM}$) of the linear (slopes) and quadratic (curvatures) of the original data ($\boldsymbol{Y}$). The mean slope is positive while the mean quadratic trend is negative. That is, overall, vocabulary increases with grade, but at a decreasing rate. The $\boldsymbol{H}$ ellipse plots as a degenerate line because the $\boldsymbol{H}$ matrix has rank 1 (1 df for the MANOVA test of the intercept). Its projection outside the $\boldsymbol{E}$ ellipse shows a highly significant rejection of the hypothesis of equal means over grade.

In such simple cases, traditional plots (boxplots, or plots of means with error bars) are easier to interpret. HE plots gain advantages in more complex designs (two or more between- or within-S factors, multiple responses), where they provide visual summaries of the information used in the multivariate hypothesis tests.

4. Between- and within-S effects

When there are both within- and between-S effects, the multivariate and univariate hypotheses tests can all be obtained together using `Anova()` with the `idata` and `idesign` specifying the within-S levels and the within-S design, as shown above. `linear.hypothesis()` can be used to test arbitrary contrasts in the within- or between- effects.

However, to explain the visualization of these tests for within-S effects and their interactions using `heplot()` and related methods it is again convenient to explicitly transform $\boldsymbol{Y} \mapsto \boldsymbol{YM}$ for a given set of within-S contrasts, in the same way as done for the `VocabGrowth` data. See Section 6 for simplified code producing these HE plots directly, without the need for explicit transformation.

To illustrate, we use the data from O'Brien and Kaiser (1985) contained in the data frame `OBrienKaiser` in the **car**. The data are from an imaginary study in which 16 female and male subjects, who are divided into three treatments, are measured at a pretest, posttest, and a follow-up session; during each session, they are measured at five occasions at intervals of one hour. The design, therefore, has two between-subject and two within-subject factors.

For simplicity here, we initially collapse over the five occasions, and consider just the within-S effect of session, called `session` in the analysis below.

```
R> library("car")
R> OBK <- OBrienKaiser
R> OBK$pre <- rowMeans(OBK[, 3:7])
R> OBK$post <- rowMeans(OBK[, 8:12])
R> OBK$fup <- rowMeans(OBK[, 13:17])
R> OBK <- OBK[, -(3:17)]
```

Note that the between-S design is unbalanced (so tests based on Type II sum of squares and crossproducts are preferred, because they conform to the principle of marginality).

```
R> table(OBK$gender, OBK$treatment)
```

```
    control A B
  F       2 2 4
  M       3 2 3
```

The factors `gender` and `treatment` were specified with the following contrasts, $\boldsymbol{L}_{\text{gender}}$, and $\boldsymbol{L}_{\text{treatment}}$, shown below. The contrasts for `treatment` are nested (Helmert) contrasts testing a comparison of the control group with the average of treatments A and B (`treatment1`) and the difference between treatments A and B (`treatment2`).

```
R> contrasts(OBK$gender)

  [,1]
F    1
M   -1

R> contrasts(OBK$treatment)

        [,1] [,2]
control   -2    0
A          1   -1
B          1    1
```

We first fit the general MANOVA model for the three repeated measures in relation to the between-S factors. As before, this just establishes the model for the between-S effects.

```
R> mod.OBK <- lm(cbind(pre, post, fup) ~ treatment * gender, data = OBK)
R> mod.OBK

Call:
lm(formula = cbind(pre, post, fup) ~ treatment * gender, data = OBK)

Coefficients:
                    pre      post     fup
(Intercept)          4.4722   5.7361   6.2917
treatment1           0.1111   0.8264   0.9792
treatment2          -0.4167   0.0625   0.0208
gender1             -0.4722  -0.6528  -0.7083
treatment1:gender1  -0.3611  -0.5347  -0.1875
treatment2:gender1   0.6667   0.8125   0.8542
```

If we regarded the repeated measure effect of `session` as equally spaced, we could simply use polynomial contrasts to examine linear (slope) and quadratic (curvature) effects of time. Here, it makes more sense to use profile contrasts, testing (`post - pre`) and (`fup - post`).

```
R> session <- ordered(c("pretest", "posttest", "followup"),
+    levels = c("pretest", "posttest", "followup"))
R> contrasts(session) <- matrix(c(-1,  1, 0, 0, -1, 1), ncol = 2)
R> session
```

```
[1] pretest  posttest followup
attr(,"contrasts")
         [,1] [,2]
pretest    -1    0
posttest    1   -1
followup    0    1
Levels: pretest < posttest < followup
```

```
R> idata <- data.frame(session)
```

The multivariate tests for all between- and within- effects are then calculated as follows:

```
R> aov.OBK <- Anova(mod.OBK, idata = idata, idesign = ~ session, test = "Roy")
R> aov.OBK
```

```
Type II Repeated Measures MANOVA Tests: Roy test statistic
                         Df test stat approx F num Df den Df   Pr(>F)
(Intercept)               1     31.83    318.3      1     10 6.53e-09 ***
treatment                 2      0.93      4.6      2     10 0.037687 *
gender                    1      0.26      2.6      1     10 0.140974
treatment:gender          2      0.57      2.9      2     10 0.104469
session                   1      5.69     25.6      2      9 0.000193 ***
treatment:session         2      2.13     10.7      2     10 0.003309 **
gender:session            1      0.05      0.2      2      9 0.819997
treatment:gender:session  2      0.42      2.1      2     10 0.175303
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It is useful to point out here that the default `print` methods for `Anova.mlm` objects, as shown above, gives an optimally compact summary for all between- and within-S effects, using a given test statistic, yet all details and other test statistics are available using the `summary` method.[4] For example, using `summary(aov.OBK)` as shown below, we can display all the multivariate tests together with the $\boldsymbol{H}$ and $\boldsymbol{E}$ matrices, and/or all the univariate tests for the traditional univariate mixed model, under the assumption of sphericity and with Geiser-Greenhouse and Huhyn-Feldt corrected F tests. To conserve space in this article the results are not shown here.

```
R> summary(aov.OBK, univariate = FALSE)
R> summary(aov.OBK, multivariate = FALSE)
```

OK, now it is time to understand the *nature* of these effects. Ordinarily, from a data-analytic point of view, I would show traditional plots of means and other measures (as in Figure 1) or their generalizations in effect plots (Fox 1987, 2003; Fox and Hong 2009). But I am not going to do that here. Instead, I'd like for you to understand how HE plots provide a compact *visual* summary of an MLM, mirroring the tabular presentation from `Anova(mod.OBK)` above, but

[4] In contrast, SAS `proc glm` and SPSS General Linear Model provide only more complete, but often bewildering outputs that still recall the days of Fortran coding in spite of more modern look and feel.

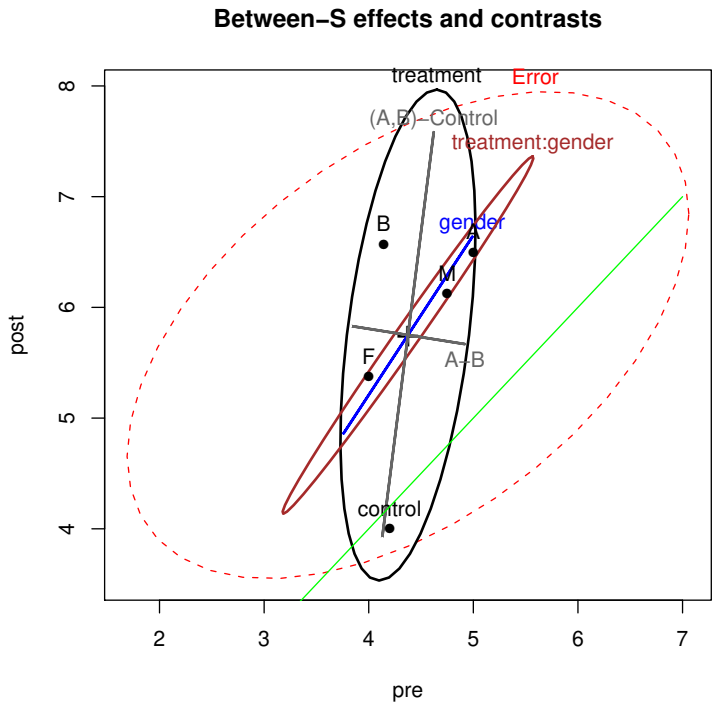

Figure 4: HE plot for the `mod.OBK` repeated measures model, showing between-S effects and contrasts in the space of the `pre` and `post` variables. Main effect means of the `treatment` (A, B, control) and `gender` (M, F) groups are marked by points. Contrasts among the treatment groups appear as lines labeled at one extreme. The green line of unit slope shows equality of `pre = post`.

which also reveals the nature of effects. Here, you have to bear in mind that between-S effects are displayed most naturally in the space of the response variables, while within-S effects are most easily seen in the contrast space of transformed responses ($\boldsymbol{YM}$).

HE plots for between-S effects can be displayed for any pair of responses with `heplot()`. Figure 4 shows this for `pre` and `post`. By default, $\boldsymbol{H}$ ellipses for all model terms (excluding the intercept) are displayed. Additional MLM tests can also be displayed using the `hypotheses` argument; here we specify the two contrasts for the `treatment` effect shown above as `contrasts(OBK$treatment)`.

```
R> heplot(mod.OBK, hypotheses = c("treatment1", "treatment2"),
+     col = c("red", "black", "blue", "brown", "gray40", "gray40"),
+     hyp.labels = c("(A,B)-Control", "A-B"),
+     main = "Between-S effects and contrasts")
R> lines(c(3, 7), c(3, 7), col = "green")
```

In Figure 4, we see that the treatment effect is significant, and the large vertical extent of this $\boldsymbol{H}$ ellipse shows this is largely attributable to the differences among groups in the post session. Moreover, the main component of the treatment effect is the contrast between the control group and groups A & B, which outperform the control group at post test. The effect

of gender is not significant, but the HE plot shows that that males are higher than females at both pre and post tests. Likewise, the treatment:gender interaction fails significance, but the orientation of the $\boldsymbol{H}$ ellipse for this effect can be interpreted as showing that the differences among the treatment groups are larger for males than for females. Finally, the line of unit slope shows that for all effects, scores are greater on post than pre.

Using heplot3d(mod.OBK, ...) gives an interactive 3D version of Figure 4 for pre, post, and fup, that can be rotated and zoomed, or played as an animation.

```
R> heplot3d(mod.OBK, hypotheses=c("treatment1", "treatment2"),
+    col = c("pink", "black", "blue", "brown", "gray40", "gray40"),
+    hyp.labels=c("(A,B)-Control", "A-B"))
R> play3d(rot8y <- spin3d(axis = c(0, 1, 0)), duration = 12)
```

This plot is not shown here, but an animated version can be generated from the code included in the supplementary materials.

Alternatively, all pairwise HE plots for the session means can be shown using pairs() for the mlm object mod.OBK, with the result shown in Figure 5.

```
R> pairs(mod.OBK, col = c("red", "black", "blue", "brown"))
```

Here we see that (a) the treatment effect is largest in the combination of post-test and follow up; (b) this 2 *df* test is essentially 1-dimensional in this view, i.e., treatment means at post-test and follow up are nearly perfectly correlated; (c) the superior performance of males relative to females, while not significant, holds up over all three sessions.

As before, for expository purposes, HE plots for within-S effects are constructed by transforming $\boldsymbol{Y} \mapsto \boldsymbol{Y}\boldsymbol{M}$, here using the (profile) contrasts for session.

```
R> OBK$session.1 <- OBK$post - OBK$pre
R> OBK$session.2 <- OBK$fup - OBK$post
R> mod1.OBK <- lm(cbind(session.1, session.2) ~ treatment * gender,
+    data = OBK)
R> Anova(mod1.OBK, test = "Roy", type = "III")

Type III MANOVA Tests: Roy test statistic
                Df test stat approx F num Df den Df   Pr(>F)
(Intercept)      1     4.366   19.645      2      9 0.000521 ***
treatment        2     2.186   10.932      2     10 0.003044 **
gender           1     0.071    0.319      2      9 0.734970
treatment:gender 2     0.417    2.083      2     10 0.175303
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the schematic summary in Table 1, with these (or any other) contrasts as $\boldsymbol{M}_{\text{session}}$, the tests of the effects contained in treatment * gender in mod1.OBK are identical to the interactions of these terms with session, as shown above for the full model in aov.OBK. The (Intercept) term in this model represents the session effect.

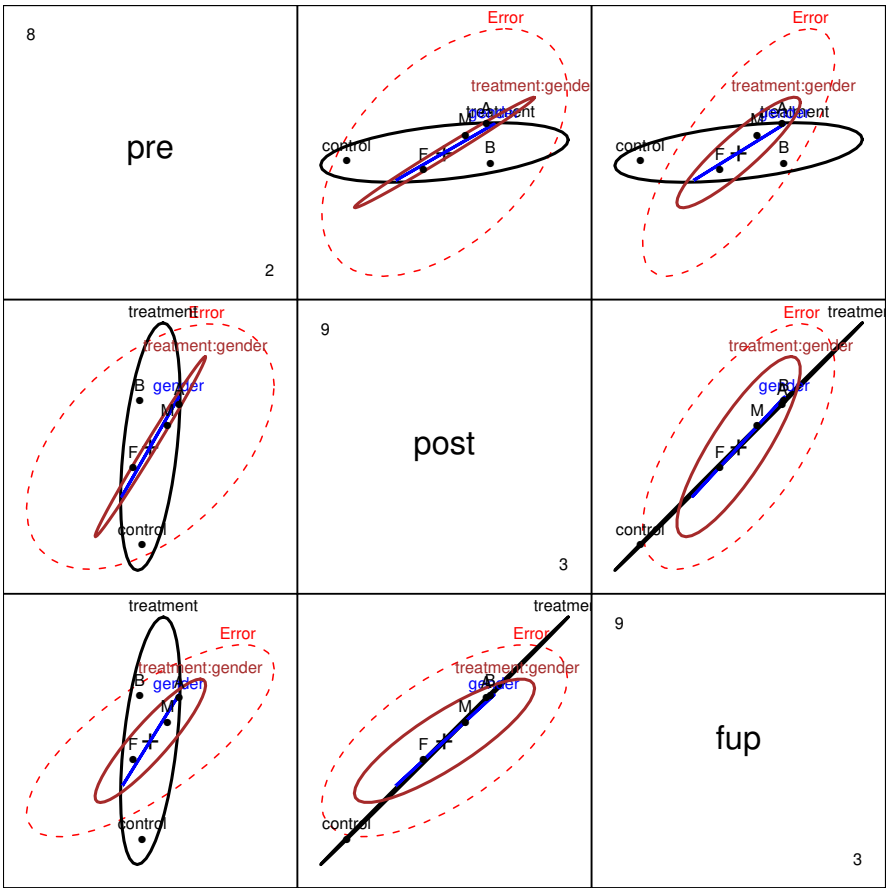

Figure 5: HE plot for the `mod.OBK` repeated measures model, showing between-S effects for all pairs of sessions. The panel in row 2, column 1 is identical to that shown separately in Figure 4.

The HE plot for within-S effects (Figure 6) is constructed from the `mod1.OBK` object as shown below. The main manipulation done here is to re-label the terms plotted to show each of them as involving `session`, as just described.

```
R> heplot(mod1.OBK, main = "Within-S effects: Session * (Treat*Gender)",
+     remove.intercept = FALSE, type = "III", xlab = "Post-Pre",
+     ylab = "Fup-Post", term.labels = c("session", "treatment:session",
+     "gender:session", "treatment:gender:session"),
+     col = c("red", "black", "blue", "brown"), xlim = c(-2, 4),
+     ylim = c(-2, 3))
R> mark.H0(0, 0)
```

Figure 6 provides an interpretation of the within-S effects shown in the MANOVA table shown above for `Anova(mod.OBK)`. We can see that the effects of `session` and `treatment:session` are significant. More importantly, for both of these, but the interaction in particular, the significance of the effect is more attributable to the `post-pre` difference than to `fup-post`.

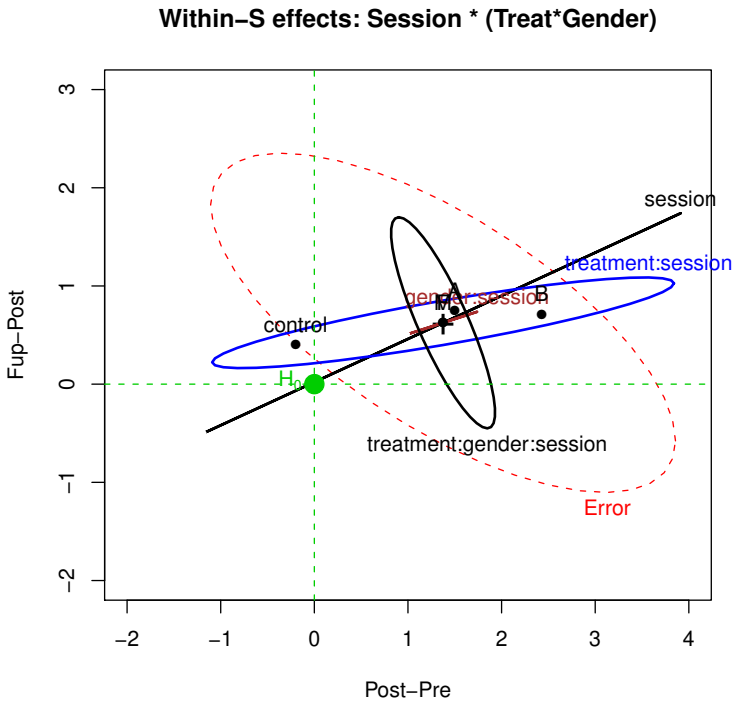

Figure 6: HE plot for the `mod1.OBK` model, showing within-S effects in the space of contrasts among sessions. The point labeled H_0 here marks the comparison point for no difference over session in contrast space.

4.1. Higher-order designs

The scheme described above and the obvious generalization of Table 1 easily accommodates designs with two or more within-S factors. Any number of between-S factors are handled automatically, by the model formula for between-S effects specified in the `lm()` call, e.g., `~ treatment * gender`.

For example, for the O'Brien-Kaiser data with `session` and `hour` as two within-S factors, first create a data frame, `within` specifying the values of these factors for the 3×5 combinations.

```
R> session <- factor(rep(c("pretest", "posttest", "followup"), c(5, 5, 5)),
+    levels = c("pretest", "posttest", "followup"))
R> contrasts(session) <- matrix(c(-1, 1, 0, 0, -1, 1), ncol = 2)
R> hour <- ordered(rep(1:5, 3))
R> within <- data.frame(session, hour)
```

The `within` data frame looks like this:

```
R> str(within)

'data.frame':        15 obs. of  2 variables:
 $ session: Factor w/ 3 levels "pretest","posttest",..: 1 1 1 1 1 2 2 2 2 2 ..
```

```
  ..- attr(*, "contrasts")= num [1:3, 1:2] -1 1 0 0 -1 1
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr  "pretest" "posttest" "followup"
  .. .. ..$ : NULL
 $ hour   : Ord.factor w/ 5 levels "1"<"2"<"3"<"4"<..: 1 2 3 4 5 1 2 3 4 5 ...
```

```
R> within
```

```
    session hour
1   pretest    1
2   pretest    2
3   pretest    3
4   pretest    4
5   pretest    5
6  posttest    1
7  posttest    2
8  posttest    3
9  posttest    4
10 posttest    5
11 followup    1
12 followup    2
13 followup    3
14 followup    4
15 followup    5
```

The repeated measures MANOVA analysis can then be carried out as follows:

```
R> mod.OBK2 <- lm(cbind(pre.1, pre.2, pre.3, pre.4, pre.5, post.1,
+    post.2, post.3, post.4, post.5, fup.1, fup.2, fup.3, fup.4,
+    fup.5) ~ treatment * gender, data = OBrienKaiser)
R> (aov.OBK2 <- Anova(mod.OBK2, idata = within, idesign = ~session * hour,
+    test = "Roy"))
```

```
Type II Repeated Measures MANOVA Tests: Roy test statistic
                            Df test stat approx F num Df den Df   Pr(>F)
(Intercept)                  1     31.83    318.3      1     10 6.53e-09 ***
treatment                    2      0.93      4.6      2     10 0.037687 *
gender                       1      0.26      2.6      1     10 0.140974
treatment:gender             2      0.57      2.9      2     10 0.104469
session                      1      5.69     25.6      2      9 0.000193 ***
treatment:session            2      2.13     10.7      2     10 0.003309 **
gender:session               1      0.05      0.2      2      9 0.819997
treatment:gender:session     2      0.42      2.1      2     10 0.175303
hour                         1     14.31     25.0      4      7 0.000304 ***
treatment:hour               2      0.23      0.5      4      8 0.758592
gender:hour                  1      0.41      0.7      4      7 0.602374
treatment:gender:hour        2      0.71      1.4      4      8 0.308786
```

```
session:hour                    1      1.22      0.5      8      3 0.832452
treatment:session:hour          2      0.58      0.3      8      4 0.936351
gender:session:hour             1      2.28      0.9      8      3 0.620208
treatment:gender:session:hour   2      0.80      0.4      8      4 0.875598
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note that the test statistics for `treatment`, `gender`, `session` and all interactions among them are identical to what was found in the simplified analysis above. Among the effects including `hour`, only the main effect is significant here.

The following $\boldsymbol{M}$ matrices, corresponding to profile contrasts for `session` and polynomial contrasts for `hour` are used internally in `Anova()` in calculating these effects (shown here as integers, rather than in normalized form).

$$\boldsymbol{M}_{\text{session}} = \begin{pmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} \qquad \boldsymbol{M}_{\text{hour}} = \begin{pmatrix} -2 & 2 & -1 & 1 \\ -1 & -1 & 2 & -4 \\ 0 & -2 & 0 & 6 \\ 1 & -1 & -2 & -4 \\ 2 & 2 & 1 & 1 \end{pmatrix}$$

Tests involving the interaction of `session:hour` use the Kronecker product, $\boldsymbol{M}_{\text{session}} \otimes \boldsymbol{M}_{\text{hour}}$.

For HE plots, it is necessary to explicitly carry out the transformation of $\boldsymbol{Y} \mapsto \boldsymbol{Y}\boldsymbol{M}_w$, where $\boldsymbol{M}_w$ conforms to $\boldsymbol{Y}$ and represents the contrasts for the within-S effect. In the present example, this means that $\boldsymbol{M}_{\text{session}}$ and $\boldsymbol{M}_{\text{hour}}$ are both expanded as Kronecker products with the unit vector,

$$\begin{aligned} \boldsymbol{M}_s &= \boldsymbol{M}_{\text{session}} \otimes \mathbf{1}_5 \ , \\ \boldsymbol{M}_h &= \mathbf{1}_3 \otimes \boldsymbol{M}_{\text{hour}} \ . \end{aligned}$$

These calculations in R are shown below:

```
R> M.session <- matrix(c(-1,  1, 0, 0, -1, 1), ncol = 2)
R> rownames(M.session) <-c("pre", "post", "fup")
R> colnames(M.session) <-paste("s", 1:2, sep = "")
R> M.hour <- matrix(c(-2, -1, 0, 1, 2, 2, -1, -2, -1, 1, -1, 2, 0, -2, 1,
+    1, -4, 6, -4, 1), ncol = 4)
R> rownames(M.hour) <- paste("hour", 1:5, sep = "")
R> colnames(M.hour) <- c("lin", "quad", "cubic", "4th")
R> M.hour

      lin quad cubic 4th
hour1  -2    2    -1   1
hour2  -1   -1     2  -4
hour3   0   -2     0   6
hour4   1   -1    -2  -4
hour5   2    1     1   1
```

```
R> unit <- function(n, prefix = "") {
+    J <-matrix(rep(1, n), ncol=1)
+    rownames(J) <- paste(prefix, 1:n, sep = "")
+    J
+  }
R> M.s <- kronecker(M.session, unit(5, "h"), make.dimnames = TRUE)
R> (M.h <- kronecker( unit(3, "s"), M.hour, make.dimnames = TRUE))
```

```
         :lin :quad :cubic :4th
s1:hour1   -2     2     -1    1
s1:hour2   -1    -1      2   -4
s1:hour3    0    -2      0    6
s1:hour4    1    -1     -2   -4
s1:hour5    2     1      1    1
s2:hour1   -2     2     -1    1
s2:hour2   -1    -1      2   -4
s2:hour3    0    -2      0    6
s2:hour4    1    -1     -2   -4
s2:hour5    2     1      1    1
s3:hour1   -2     2     -1    1
s3:hour2   -1    -1      2   -4
s3:hour3    0    -2      0    6
s3:hour4    1    -1     -2   -4
s3:hour5    2     1      1    1
```

```
R> M.sh <- kronecker(M.session, M.hour, make.dimnames = TRUE)
```

Using `M.h`, we can construct the within-model for all terms involving the `hour` effect,

```
R> Y.hour <- as.matrix(OBrienKaiser[, 3:17]) %*% M.h
R> mod.OBK2.hour <- lm(Y.hour ~ treatment * gender, data = OBrienKaiser)
```

We can plot these effects for the linear and quadratic contrasts of `hour`, representing within-session slope and curvature. Figure 7 is produced as shown below. As shown by the `Anova(mod.OBK2, ...)` above, all interactions with `hour` are small, and so these appear wholly contained within the $\boldsymbol{E}$ ellipse. In particular, there are no differences among groups (treatment × gender) in the slopes or curvatures over `hour`. For the main effect of `hour`, the linear effect is almost exactly zero, while the quadratic effect is huge.

```
R> labels <- c("hour", paste(c("treatment", "gender", "treatment:gender"),
+    ":hour", sep = ""))
R> colors <- c("red", "black", "blue", "brown", "purple")
R> heplot(mod.OBK2.hour, type = "III", remove.intercept = FALSE,
+    term.labels = labels, col = colors)
R> mark.H0()
```

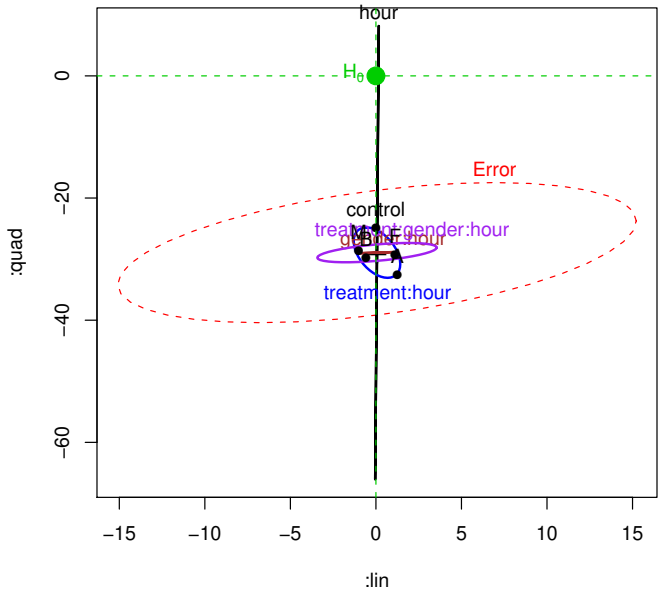

Figure 7: HE plot for the `mod.OBK2` repeated measures model, showing within-S effects for linear and quadratic contrasts in `hour`. As in Figure 6, we are viewing hypothesis and error variation in the transformed space of the repeated measures contrasts, here given by $\boldsymbol{M}_h$.

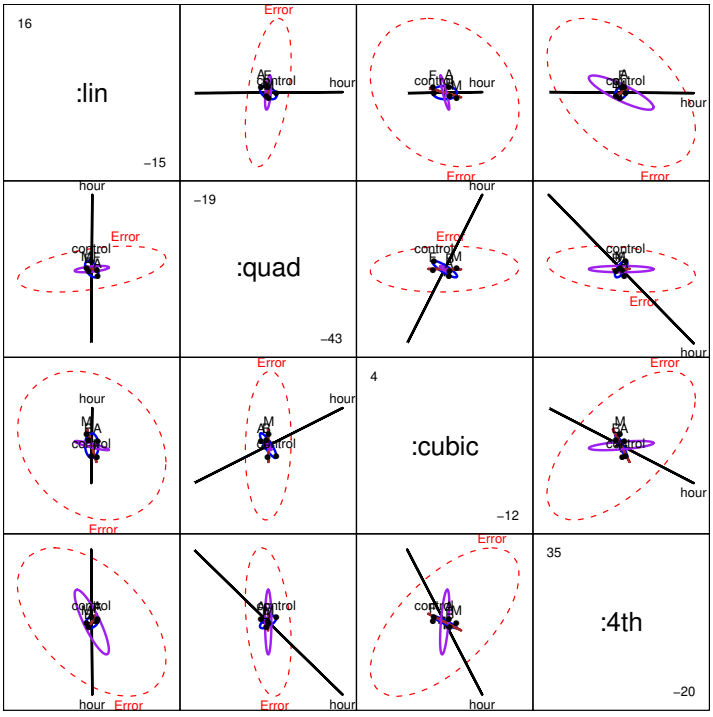

Figure 8: HE plot for the `mod.OBK2` repeated measures model, showing within-S effects for all pairs of contrasts in `hour`.

The `pairs()` function shows these effects (Figure 8) for all contrasts in `hour`. To reduce clutter, we only label the hour effect, since all interactions with hour are small and non-significant. The main additional message here is that the effects of `hour` are more complex than just the large quadratic component we saw in Figure 7.

```
R> pairs(mod.OBK2.hour, type = "III", remove.intercept = FALSE,
+    term.labels = "hour", col = colors)
```

5. Doubly-multivariate designs

In the designs discussed above the same measure is observed on all occasions. Sometimes, there are two or more *different* measures, $Y_1, Y_2, \ldots$, observed at each occasion, for example response speed and accuracy. In these cases, researchers often carry out separate repeated measures analyses for each measure. However the tests of between-S effects and each within-S effect can also be carried out as multivariate tests of $Y_1, Y_2, \ldots$ simultaneously, and these tests are often more powerful, particularly when the effects for the different measures are weak, but correlated.

In the present context, such doubly-multivariate designs can be easily handled in principle by treating the multiple measures as an additional within-S factor, but using an identity matrix as the $\boldsymbol{M}$ matrix in forming the linear hypotheses to be tested via Equation 6. For example, with two measures, Y_1, Y_2 observed on three repeated sessions, the full $\boldsymbol{M}$ matrix for the design is generated as in Equation 9 as

$$\boldsymbol{M}_{\text{CM}} = (\mathbf{1}, \boldsymbol{M}_{\text{session}}) \otimes \boldsymbol{I}_2 = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{10}$$

In R, we can express this as follows, using `M.measure` to represent Y_1, Y_2.

```
R> M.measure <- diag(2)
R> rownames(M.measure) <- c("Y1", "Y2")
R> colnames(M.measure) <- c("Y1", "Y2")
R> kronecker(cbind(1, M.session), M.measure, make.dimnames = TRUE)

        :Y1 :Y2 s1:Y1 s1:Y2 s2:Y1 s2:Y2
pre:Y1    1   0    -1     0     0     0
pre:Y2    0   1     0    -1     0     0
post:Y1   1   0     1     0    -1     0
post:Y2   0   1     0     1     0    -1
fup:Y1    1   0     0     0     1     0
fup:Y2    0   1     0     0     0     1
```

In the result, the first two columns correspond to the within-S Intercept term, and are used to test all between-S terms for Y_1, Y_2 simultaneously. The remaining columns correspond to the session effect for both variables and all interactions with session. In practice, this analysis

must be performed in stages because `Anova()` does not (yet)[5] allow such a doubly-multivariate design to be specified directly.

5.1. Example: Weight loss and self esteem

To illustrate, we consider the data frame `WeightLoss` originally from Andrew Ainsworth (http://www.csun.edu/~ata20315/psy524/main.htm), giving (contrived) data on weight loss and measures of self esteem after each of three months for 34 individuals, who were observed in one of three groups: Control, diet group, diet + exercise group. The within-S factors are thus measure (`wl, se`) and month (`1:3`).

```
R> table(WeightLoss$group)

Control    Diet  DietEx
     12      12      10

R> some(WeightLoss)

     group wl1 wl2 wl3 se1 se2 se3
6  Control   6   5   4  17  18  18
11 Control   4   2   2  16  16  11
12 Control   5   2   1  15  13  16
19    Diet   4   3   1  12  11  14
20    Diet   4   2   1  12  11  11
21    Diet   6   5   3  17  16  19
24    Diet   7   4   3  16  14  18
28  DietEx   3   4   1  16  13  17
29  DietEx   3   5   1  13  13  16
30  DietEx   6   5   2  15  12  18
```

Because this design is complex, and to facilitate interpretation of the effects we will see in HE plots, it is helpful to view traditional plots of means with standard errors for both variables. These plots, shown in Figure 9,[6] show that, for all three groups, the amount of weight lost each month declines, but only the diet + exercise maintains substantial weight loss through month 2. For self esteem, all three groups have a U-shaped pattern over months, and by month 3, the groups are ordered control < diet < diet + exercise.

Research interest in the differences among groups would likely be focused on the questions: (a) Do the two diet groups differ from the control group? (b) Is there an additional effect of exercise, given diet? These questions may be tested with the (Helmert) contrasts used below for `group`, which are labeled `group1` and `group1` respectively.

```
R> contrasts(WeightLoss$group) <- matrix(c(-2, 1, 1, 0, -1, 1), ncol = 2)
R> (wl.mod <- lm(cbind(wl1, wl2, wl3, se1, se2, se3) ~ group,
+    data = WeightLoss))
```

[5] The new version of the **car** (2.0-0) on CRAN now includes enhanced `Anova()` and `linearHypothesis()` which perform these analyses.

[6] These plots were drawn using `plotmeans()` in the **gplots** (Warnes 2010). The code is not shown, but is available in the R example code for this paper.

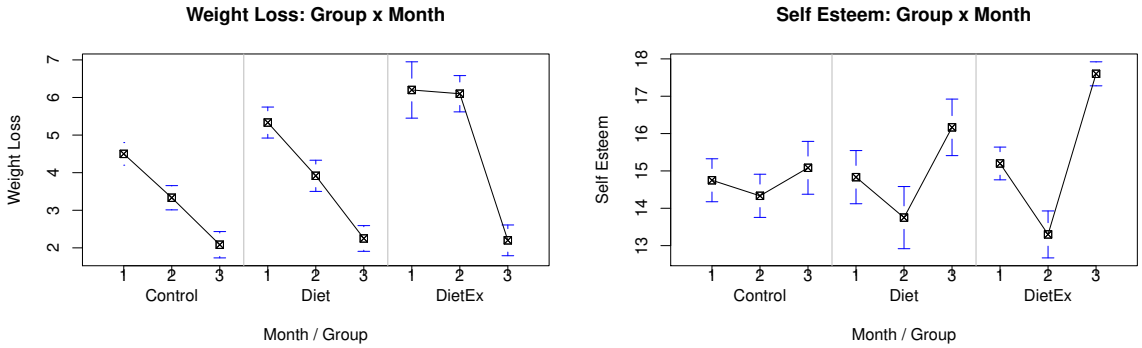

Figure 9: Means for weight loss and self esteem by group and month. Error bars show ±1 standard error for each mean.

```
Call:
lm(formula = cbind(wl1, wl2, wl3, se1, se2, se3) ~ group, data = WeightLoss)

Coefficients:
             wl1      wl2      wl3      se1      se2      se3
(Intercept)   5.3444   4.4500   2.1778  14.9278  13.7944  16.2833
group1        0.4222   0.5583   0.0472   0.0889  -0.2694   0.6000
group2        0.4333   1.0917  -0.0250   0.1833  -0.2250   0.7167
```

A standard between-S MANOVA, ignoring the within-S structure shows a highly significant `group` effect.

```
R> Anova(wl.mod)

Type II MANOVA Tests: Pillai test statistic
      Df test stat approx F num Df den Df  Pr(>F)
group  2    0.7255    2.562     12     54 0.00924 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As before, it is often useful to examine HE plots for pairs of variables in this analysis before proceeding to the within-S analysis. For example, Figure 10 shows the test of `group` and the two contrasts for weight loss and for self esteem at months 1 and 2.

```
R> op <- par(mfrow = c(1, 2))
R> heplot(wl.mod, hypotheses = c("group1", "group2"),
+    xlab = "Weight Loss, month 1", ylab = "Weight Loss, month 2")
R> heplot(wl.mod, hypotheses=c("group1", "group2"), variables = 4:5,
+    xlab = "Self Esteem, month 1", ylab = "Self Esteem, month 2")
R> par(op)
```

This is helpful, but doesn't illuminate the *overall* `group` effect for weight loss and self esteem for all three months, and, of course cannot shed light on any interactions of group with measure

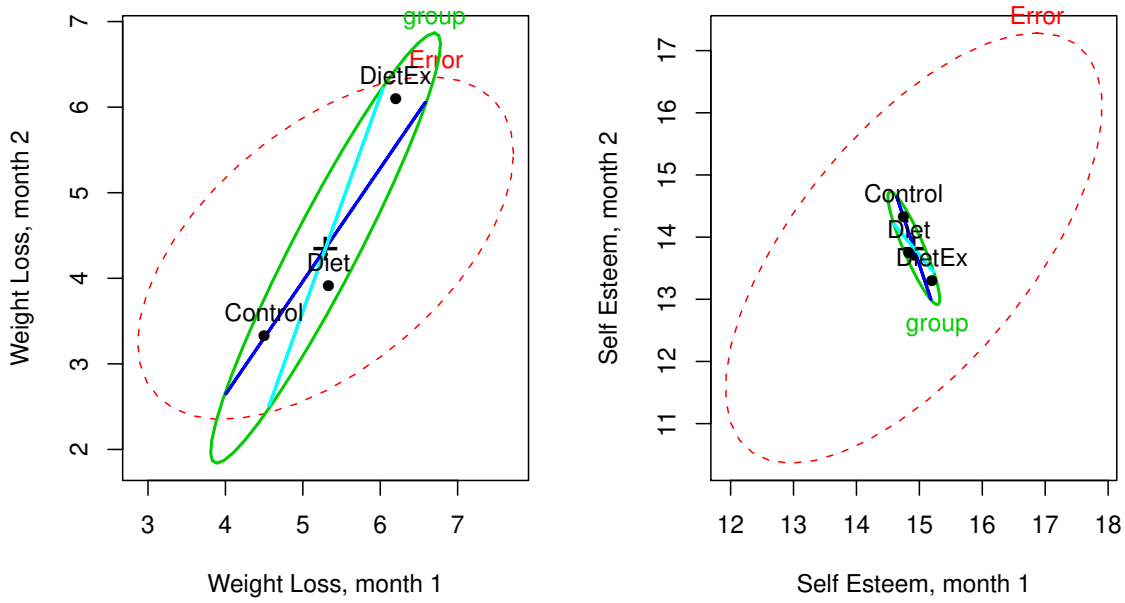

Figure 10: HE plot for the `wl.mod` MANOVA model, showing between-S effects for weight loss (left) and self esteem (right) at months 1 and 2.

or month. In the following discussion, we will assume that the researcher is particularly interested in understanding the relation between weight loss and self esteem as it is expressed in changes over time and differences among groups.

To carry out the doubly-multivariate analysis, we proceed as follows. First, we define the $\boldsymbol{M}$ matrix for the measures, used in the between-S analysis. We use $\boldsymbol{M} = \boldsymbol{I}_2 \otimes \mathbf{1}/3$ so that the resulting scores are the means (not sums) for weight loss and self esteem.

```
R> measure <- kronecker(diag(2), unit(3, "M")/3, make.dimnames = TRUE)
R> colnames(measure) <- c("WL", "SE")
R> measure

          WL       SE
:M1 0.333333 0.000000
:M2 0.333333 0.000000
:M3 0.333333 0.000000
:M1 0.000000 0.333333
:M2 0.000000 0.333333
:M3 0.000000 0.333333

R> between <- as.matrix(WeightLoss[, -1]) %*% measure
R> between.mod <- lm(between ~ group, data = WeightLoss)
R> Anova(between.mod, test = "Roy", type = "III")

Type III MANOVA Tests: Roy test statistic
```

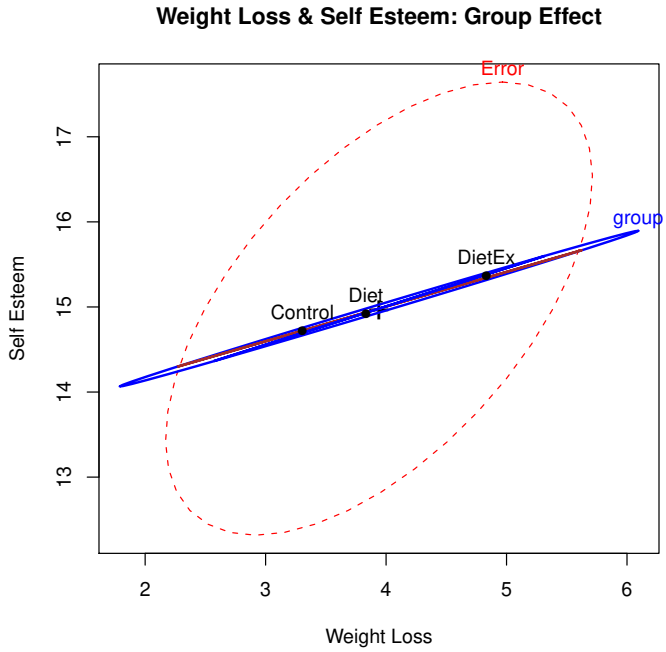

Figure 11: HE plot for the `between.mod` doubly-multivariate design, showing overall between-S effects for weight loss and self esteem.

```
            Df test stat approx F num Df den Df Pr(>F)
(Intercept)  1     85.62   1284.3      2     30 <2e-16 ***
group        2      0.36      5.5      2     31 0.0089 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The HE plot for this component of the analysis (Figure 11) shows a striking effect: Averaging over all three months, the means for the `Control`, `Diet` and `DietEx` group on both weight loss and self esteem are highly correlated and in the expected direction. This is something that is not at all obvious in Figure 9.

```
R> heplot(between.mod, hypotheses=c("group1", "group2"), xlab = "Weight Loss",
+    ylab = "Self Esteem", col = c("red", "blue", "brown"),
+    main = "Weight Loss & Self Esteem: Group Effect")
```

Next, for the within-S analysis, we define the $\boldsymbol{M}$ matrix for months, using orthogonal polynomials representing linear and quadratic trends. As before, the test of the `(Intercept)` term in these trend scores corresponds to the month effect in the doubly-multivariate model, and the `group` effect tests the group $\times$ month interaction.

```
R> month <- kronecker(diag(2), poly(1:3, degree = 2), make.dimnames = TRUE)
R> colnames(month) <- c("WL1", "WL2", "SE1", "SE2")
R> round(month, digits = 4)
```

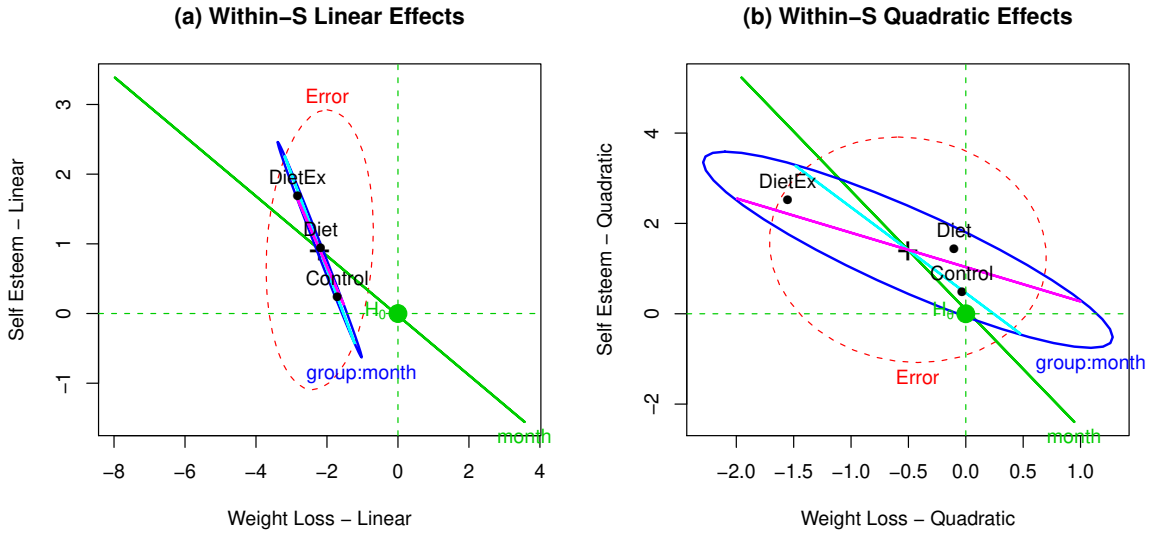

Figure 12: HE plots for the `within.mod` doubly-multivariate design, showing the effects of `month` and the interaction `group:month` for weight loss vs. self esteem. (a) Linear effects; (b) Quadratic effects.

```
      WL1     WL2     SE1     SE2
: -0.7071  0.4082  0.0000  0.0000
:  0.0000 -0.8165  0.0000  0.0000
:  0.7071  0.4082  0.0000  0.0000
:  0.0000  0.0000 -0.7071  0.4082
:  0.0000  0.0000  0.0000 -0.8165
:  0.0000  0.0000  0.7071  0.4082
```

```
R> trends <- as.matrix(WeightLoss[, -1]) %*% month
R> within.mod <- lm(trends ~ group, data = WeightLoss)
R> Anova(within.mod, test = "Roy", type = "III")

Type III MANOVA Tests: Roy test statistic
            Df test stat approx F num Df den Df   Pr(>F)
(Intercept)  1     9.928    69.50      4     28 3.96e-14 ***
group        2     1.772    12.84      4     29 3.91e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

HE plots corresponding to this model (Figure 12) can be produced as follows. The $\boldsymbol{H}$ and $\boldsymbol{E}$ matrices are all 4×4, but the $\boldsymbol{H}$ matrices for the `month` and `group:month` effects are rank 1 and 2 respectively.

```
R> op <- par(mfrow = c(1, 2))
R> heplot(within.mod, hypotheses = c("group1", "group2"), variables = c(1, 3),
+    xlab = "Weight Loss - Linear", ylab = "Self Esteem - Linear",
+    type = "III", remove.intercept = FALSE, term.labels =
```

```
+    c("month", "group:month"), main = "(a) Within-S Linear Effects")
R> mark.H0()
R> heplot(within.mod, hypotheses=c("group1", "group2"), variables = c(2,4),
+    xlab = "Weight Loss - Quadratic", ylab = "Self Esteem - Quadratic",
+    type = "III", remove.intercept = FALSE, term.labels =
+    c("month", "group:month"), main = "(b) Within-S Quadratic Effects")
R> mark.H0()
R> par(op)
```

Figure 12 shows the plots for the linear and quadratic effects separately for weight loss vs. self esteem. The plot of linear effects (Figure 12a) shows that the effect of `month` can be be described as negative slopes for weight loss combined with positive slopes for self esteem– all groups lose progressively less weight over time, but generally feel better about themselves. Differences among groups in the `group:month` effect are in the same direction, but with greater differences among groups in the slopes for self esteem. The interpretation of the quadratic effects (Figure 12b) is similar, except here, differences in curvature over months are driven largely by the difference between the `DietEx` group from the others on weight loss.

The interested reader might wish to compare the standard univariate plots of means in Figure 9 with the HE plots in Figure 11 and Figure 12. The univariate plots have the advantage of showing the data directly, but cannot show the sources of significant effects in multivariate repeated measures models. HE plots have the advantage that they show directly what is expressed in the multivariate tests for relevant hypotheses.

6. Simplified interface: heplots 0.9 and car 2.0

It sometimes happens that the act of describing and illustrating software spurs development to make both simpler, and such is the case here. At the beginning, the stable version of **car** on CRAN provided the computation for multivariate linear hypotheses including repeated measures designs, but could not handle doubly-multivariate designs directly; the CRAN version of **heplots** could only repeated measures by explicitly transforming $\boldsymbol{Y} \mapsto \boldsymbol{Y}\boldsymbol{M}$ and re-fitting submodels in terms of the transformed responses.

The new versions of these packages on CRAN (http://CRAN.R-project.org/) now handle these cases *directly* from the basic `mlm` object. `heplot()` now provides the arguments `idata`, `idesign`, `icontrasts`, or, for the doubly-multivariate case, `imatrix`, which are passed to `Anova()` to calculate the appropriate $\boldsymbol{H}$ and $\boldsymbol{E}$ matrices.

Omitting these arguments in the call to `heplot()` gives an HE plot for all between-S effects (or the subset specified by the `terms` argument), just as before. For the within-S effects, $\boldsymbol{E}$ matrices differ for for different within-S terms, so it is necessary to specify the intra-subject term (`iterm`, corresponding to $\boldsymbol{M}$) for which HE plots are desired. Several examples are given below.

For the `VocabularyGrowth` data, Figure 3(b) can be produced by

```
R> (Vocab.mod <- lm(cbind(grade8, grade9, grade10, grade11) ~ 1,
+    data = VocabGrowth))
R> idata <- data.frame(grade = ordered(8:11))
```

```
R> heplot(Vocab.mod, type = "III", idata = idata, idesign = ~ grade,
+    iterm = "grade", main = "HE plot for Grade effect")
```

For the `OBrienKaiser` data, the code for plots of between-S effects is the same as shown above for Figure 4 and Figure 5. The HE plot for within-S effects involving `session` (Figure 6) can be produced using `iterm="session"`:

```
R> idata <- data.frame(session)
R> heplot(mod.OBK, idata=idata, idesign = ~ session, iterm = "session",
+    col = c("red", "black", "blue", "brown"),
+    main = "Within-S effects: Session * (Treat*Gender)")
```

Similarly, HE plots for terms involving `hour` can be obtained using the expanded model (`mod.OBK2`) for the 15 combinations of `hour` and `session`:

```
R> mod.OBK2 <- lm(cbind(pre.1, pre.2, pre.3, pre.4, pre.5, post.1,
+    post.2, post.3, post.4, post.5, fup.1, fup.2, fup.3, fup.4,
+    fup.5) ~  treatment * gender, data = OBrienKaiser)
R> heplot(mod.OBK2, idata = within, idesign = ~ hour, iterm = "hour")
R> heplot(mod.OBK2, idata = within, idesign = ~ session * hour,
+    iterm = "session:hour")
```

7. Comparison with other approaches

The principal goals of this paper have been (a) to describe the extension of the classical MVLM to repeated measures designs; (b) to explain how HE plots provide compact and understandable visual summaries of the effects shown in typical numerical tables of MANOVA tests; and (c) illustrate these in a variety of contexts ranging from single-sample designs to complex doubly-multivariate designs.

In the context of repeated measures designs, I mentioned earlier that mixed models for longitudinal data provide an attractive alternative to the MVLM (because the former easily accommodate missing or unbalanced data over intra-subject measurements, time-varying covariates, and often allows the residual covariation to be modeled with fewer parameters).

Here we consider a classic data set (Potthoff and Roy 1964) used in the first application of the MVLM to growth-curve analysis. These data are often used as illustrations of longitudinal models, e.g., Verbeke and Molenberghs (2000, Chapter 17.4).

Investigators at the University of North Carolina Dental School followed the growth of 27 children (16 males, 11 females) from age 8 until age 14 in a study designed to establish typical patterns of jaw size useful for orthodontic practice. Every two years they measured the distance between the pituitary and the pterygomaxillary fissure, two points that are easily identified on x-ray exposures of the side of the head. The questions of interest include: (a) Over this age range, can growth be adequately represented as linear in time, or is some more complex function necessary? (b) Are separate growth curves needed for boys and girls, or can both be described by the same growth curve?

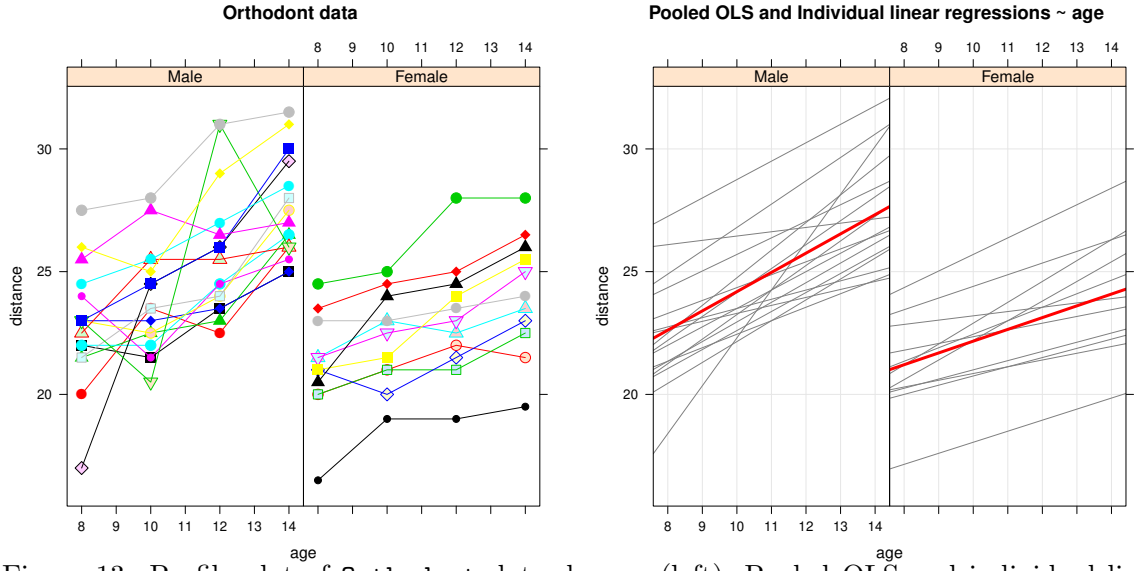

Figure 13: Profile plot of `Orthodont` data, by sex (left); Pooled OLS and individual linear regressions on age, by sex (right).

7.1. Longitudinal, mixed model approach

The mixed model for longitudinal data is very general and flexible for the reasons noted above, but it is inappropriate here to relate any more than the barest of details necessary for this example. We begin with simple plots of the data: A profile plot grouped by `Sex` (Figure 13 left),

```
R> data("Orthodont", package = "nlme")
R> library("lattice")
R> xyplot(distance ~ age | Sex, data = Orthodont, type = "b",
+     groups = Subject, pch = 15:25, col = palette(), cex = 1.3,
+     main = "Orthodont data")
```

and also a summary plot showing fitted lines for each individual, together with the pooled ordinary least squares regression of `distance` on `age` (Figure 13 right).

```
R> xyplot(distance ~ age | Sex, data = Orthodont, groups = Subject,
+     main = "Pooled OLS and Individual linear regressions ~ age",
+     type = c("g", "r"), panel = function(x, y, ...) {
+       panel.xyplot(x, y, ..., col = gray(0.5))
+       panel.lmline(x, y, ..., lwd = 3, col = "red")
+     })
```

From these plots, we can see that boys generally have larger jaw distances than girls, and the rate of growth (slopes) for boys is generally larger than for girls. It is difficult to discern any patterns within the sexes, except that one boy seems to stand out, with a lower intercept and steeper slope.

With the longitudinal mixed model, contemplate fitting two models describing an individual's pattern of growth: a model fitting only linear growth and a model fitting each person's trajectory exactly by including quadratic and cubic trends in time. For the sake of interpretation

of coefficients in these models, it is common to recenter the time variable so that time = 0 corresponds to initial status. Using year = (age −8), we have:

$$\text{m1}: \quad y_{it} = \beta_{0i} + \beta_{1i}\text{Year}_{it} + e_{it} \tag{11}$$

$$\text{m3}: \quad y_{it} = \beta_{0i} + \beta_{1i}\text{Year}_{it} + \beta_{2i}\text{Year}_{it}^2 + \beta_{3i}\text{Year}_{it}^3 + e_{it} \ , \tag{12}$$

where the vector of residuals for subject i is $\boldsymbol{e}_{i\bullet} \sim N(\mathbf{0}, \boldsymbol{R}_i)$. (For this example, we take $\boldsymbol{R}_i$ to be unstructured, even though other specifications require fewer parameters.) For the linear model (m1), we entertain the possibility that the person-level intercepts (β_{0i}) and slopes (β_{1i}) depend on Sex, and so specify them as random coefficients,

$$\beta_{0i} = \gamma_{00} + \gamma_{01}\text{Sex}_i + u_{0i} \ , \tag{13}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}\text{Sex}_i + u_{1i} \ . \tag{14}$$

In these equations the γs are the fixed effects, while the u (along with the errors e_{it}) are random effects. Note that Sex is coded 0=Male, 1=Female, so γ_{00} and γ_{10} are the intercept and slope for Males; γ_{01} pertains to the difference in intercepts for Females relative to Males, while γ_{11} is the difference in slopes.

The linear growth model can be fit using `lme` as follows:

```
R> Ortho <- Orthodont
R> Ortho$year <- Ortho$age - 8
R> Ortho.mix1 <- lme(distance ~ year * Sex, data = Ortho,
+    random = ~1 + year | Subject, method = "ML")
R> anova(Ortho.mix1)

            numDF denDF F-value p-value
(Intercept)     1    79 4197.05  <.0001
year            1    79  103.42  <.0001
Sex             1    25    8.34  0.0079
year:Sex        1    79    5.32  0.0237
```

Similarly, the model (m3) allowing cubic growth at level 1 can be fit using:

```
R> Ortho.mix3 <- lme(distance ~ year * Sex + I(year^2) + I(year^3),
+    data = Ortho, random = ~1 + year | Subject, method = "ML")
R> anova(Ortho.mix3)

            numDF denDF F-value p-value
(Intercept)     1    77 4116.30  <.0001
year            1    77  101.43  <.0001
Sex             1    25    8.18  0.0084
I(year^2)       1    77    0.81  0.3703
I(year^3)       1    77    0.22  0.6414
year:Sex        1    77    5.22  0.0251
```

A likelihood ratio test confirms that the quadratic and cubic components of year do not improve the model,

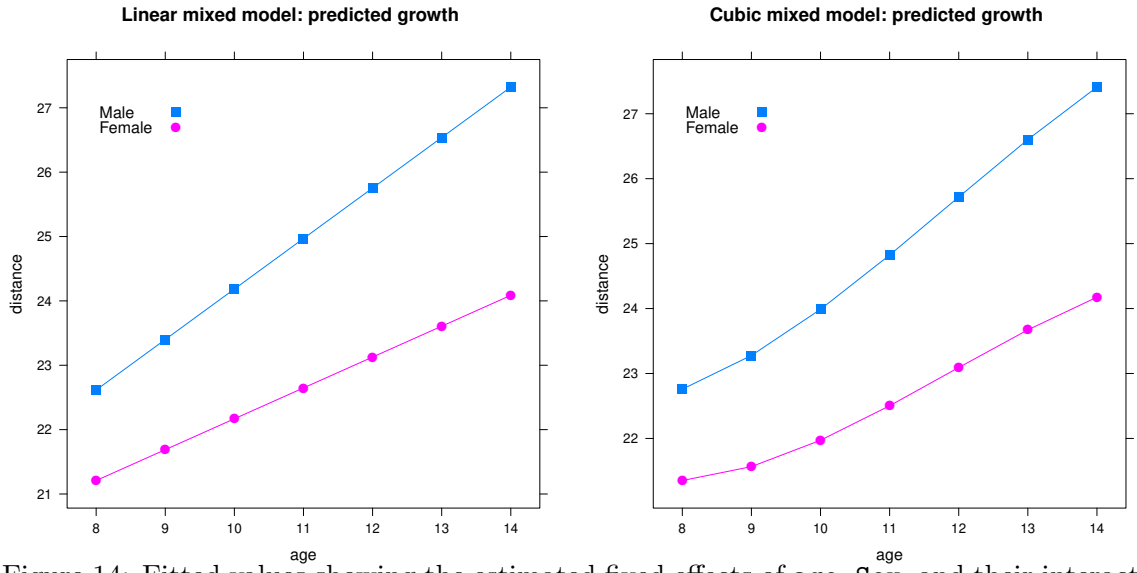

Figure 14: Fitted values showing the estimated fixed effects of `age`, `Sex`, and their interaction in the linear growth model (left) and cubic growth model (right).

```
R> anova(Ortho.mix1, Ortho.mix3)

           Model df      AIC      BIC    logLik   Test L.Ratio p-value
Ortho.mix1     1  8 443.806 465.263 -213.903
Ortho.mix3     2 10 446.725 473.547 -213.363 1 vs 2 1.08061  0.5826
```

To aid interpretation, we can plot the estimated fixed effects from the linear model (m1) as follows, using the `predict` method for `lme` objects to calculate the the fitted values for boys and girls over the range of years (0 to 6) corresponding to ages 8 to 14, as in Figure 14(left). Similar code produces a plot of the cubic model Figure 14(right).

```
R> grid <- expand.grid(year = 0:6, Sex = c("Male", "Female"))
R> grid$age <- grid$year + 8
R> fm.mix1 <-cbind(grid, distance = predict(Ortho.mix1, newdata = grid,
+    level = 0))
R> xyplot(distance ~ age, data = fm.mix1, groups = Sex, type = "b",
+    par.settings = list(superpose.symbol = list(cex = 1.2, pch = c(15, 16))),
+    auto.key = list(text = levels(fm.mix1$Sex), points = TRUE, x = 0.05,
+    y = 0.9, corner = c(0,1)), main = "Linear mixed model: predicted growth")
```

For this simple example with only two predictors, such plots provide a direct visual summary of the fitted fixed effects in the model, as far as they go.

7.2. MVLM approach

For the multivariate approach, the `Orthodont` data must first be reshaped to wide format with the distance values as separate columns.

```
R> library("nlme")
R> Orthowide <- reshape(Orthodont, v.names = "distance",
```

```
+    idvar = c("Subject", "Sex"), timevar = "age", direction = "wide")
R> some(Orthowide, 4)

   Subject    Sex distance.8 distance.10 distance.12 distance.14
1      M01   Male         26        25.0        29.0        31.0
9      M03   Male         23        22.5        24.0        27.5
65     F01 Female         21        20.0        21.5        23.0
93     F08 Female         23        23.0        23.5        24.0
```

The MVLM is then fit as follows, with `Sex` as the between-S factor. Age is quantitative, so the intra-subject data frame (`idata`) is created with `age` as an ordered factor.

```
R> Ortho.mod <- lm(cbind(distance.8, distance.10, distance.12,
+    distance.14) ~ Sex, data = Orthowide)
R> idata <- data.frame(age = ordered(seq(8, 14, 2)))
R> Ortho.aov <- Anova(Ortho.mod, idata = idata, idesign = ~ age)
R> Ortho.aov

Type II Repeated Measures MANOVA Tests: Pillai test statistic
            Df test stat approx F num Df den Df   Pr(>F)
(Intercept)  1    0.9940     4123      1     25  < 2e-16 ***
Sex          1    0.2710        9      1     25  0.00538 **
age          1    0.8256       36      3     23 6.88e-09 ***
Sex:age      1    0.2601        3      3     23  0.06960 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that both `Sex` and `age` are highly significant and their interaction is nearly significant.

Figure 15 shows HE plots for the between- and within-S effects, produced as shown below. The left panel plots the effect of `Sex` for ages 8 and 14, with a green line of unit slope. Males clearly show greater growth by age 14, and the difference between males and females is greater at at 14 than at age 8. The right panel shows the linear and quadratic trends with age, reflecting the overall `age` main effect and the `Sex:age` interaction. Recalling that the contributions of each displayed variable to each effect in an HE plot can be seen by their horizontal and vertical shadows relative to the $\boldsymbol{E}$ ellipse, we see that the main effect of `age` is essentially linear, and the overall `Sex:age` effect is nearly significant due to a difference in slopes, but not curvature.

```
R> op <- par(mfrow = c(1, 2))
R> heplot(Ortho.mod, variables = c(1, 4), asp = 1, col = c("red", "blue"),
+    xlim = c(18, 30), ylim = c(18, 30), main = "Orthodont data: Sex effect")
R> abline(0,1, col = "green")
R> heplot(Ortho.mod, idata = idata, idesign = ~ age, iterm = "age",
+    col = c("red", "blue", "brown"),
+    main = "Orthodont data: Within-S effects")
R> par(op)
```

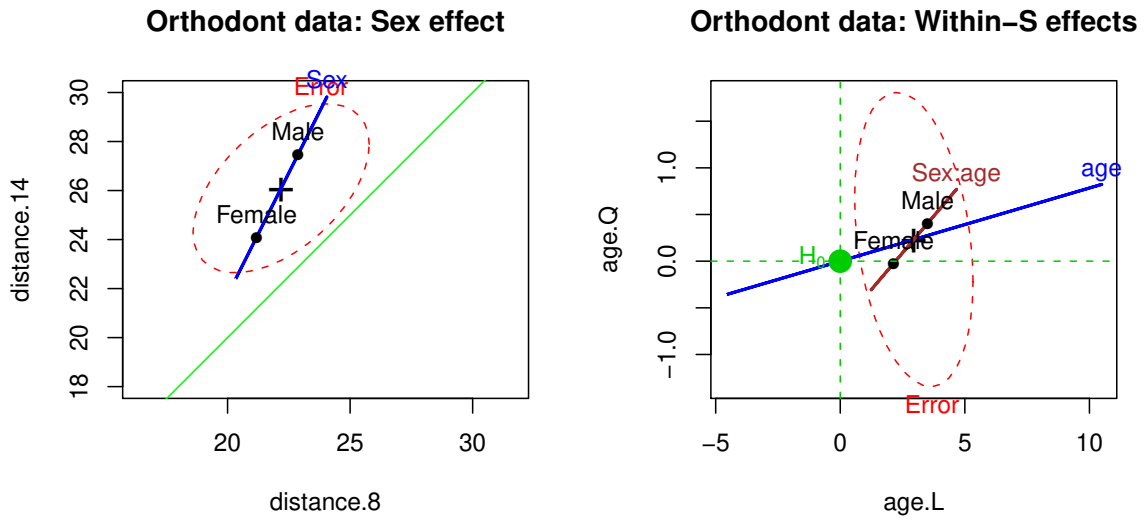

Figure 15: HE plots for the `Ortho.mod` MANOVA model, showing between-S effects for distance at ages 8 and 14 (left) and the linear and quadratic within-S effects for `age` and `Sex:age` (right).

To examine the questions of interest here in more detail, we focus on the intra-subject design and ask if linear growth is sufficient to explain both average development over time and differences between boys and girls. This is easily answered visually from the `pairs()` plot (not shown here),

```
R> pairs(Ortho.mod, idata = idata, idesign = ~ age, iterm = "age",
+    col = c("red", "blue", "brown"))
```

and, in particular, the panel corresponding to the nonlinear (quadratic and cubic) components of trend, shown in Figure 16.

```
R> heplot(Ortho.mod, idata = idata, idesign = ~ age, iterm = "age",
+    col = c("red", "blue", "brown"), variables = c(2, 3),
+    main = "Orthodont data: Nonlinear Within-S effects")
```

We can confirm the impression that no nonlinear effects are important by testing linear hypotheses. To explain this, we first show the details of the test of the overall `Sex:age` effect, as tested with `linearHypothesis()`. The "response transformation matrix" shown below is equivalent to $\boldsymbol{M}_{\text{poly}}$ described earlier (Section 3.1) for a 4-level factor with linear, quadratic and cubic trend components. The univariate tests for individual contrasts in age are then based on the diagonal elements of the $\boldsymbol{H}$ and $\boldsymbol{E}$ matrices.

```
R> linearHypothesis(Ortho.mod, hypothesis = "SexFemale", idata = idata,
+    idesign = ~ age, iterms = "age", title = "Sex:age effect")

 Response transformation matrix:
               age.L age.Q     age.C
distance.8  -0.670820   0.5 -0.223607
```

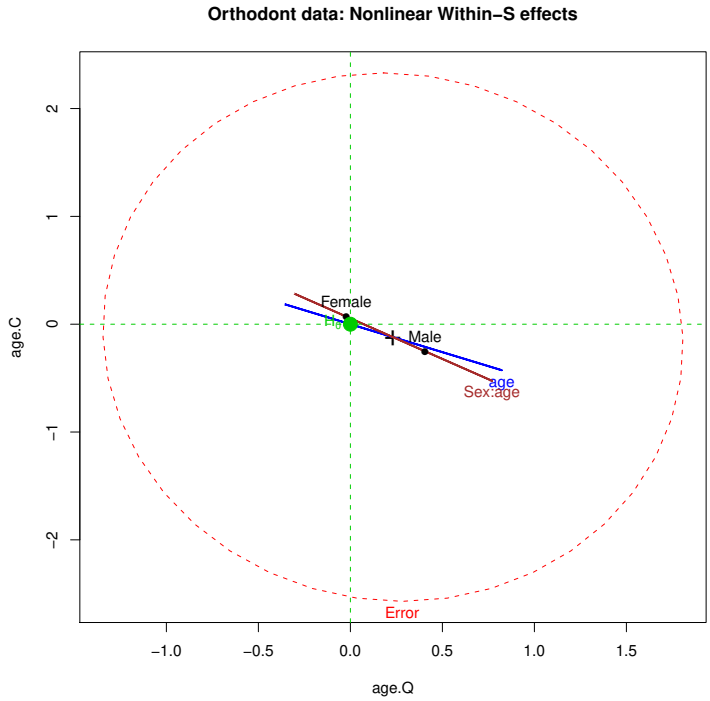

Figure 16: Nonlinear components of the within-S effects of `age` and `Sex:age`, showing that they are quite small in relation to within-S error.

```
distance.10 -0.223607  -0.5  0.670820
distance.12  0.223607  -0.5 -0.670820
distance.14  0.670820   0.5  0.223607

Sum of squares and products for the hypothesis:
         age.L    age.Q    age.C
age.L 12.11415  3.81202 -2.86766
age.Q  3.81202  1.19955 -0.90238
age.C -2.86766 -0.90238  0.67883

Sum of squares and products for error:
          age.L     age.Q    age.C
age.L  59.16733 -11.22417  4.52784
age.Q -11.22417  26.04119 -1.28193
age.C   4.52784  -1.28193 62.91932

Multivariate Tests: Sex:age effect
                 Df test stat approx F num Df den Df   Pr(>F)
Pillai            1  0.260113  2.69527      3     23 0.069604 .
Wilks             1  0.739887  2.69527      3     23 0.069604 .
Hotelling-Lawley  1  0.351557  2.69527      3     23 0.069604 .
Roy               1  0.351557  2.69527      3     23 0.069604 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From this, the linear and nonlinear terms can be tested by selecting the appropriate columns of $\boldsymbol{M}_{\text{poly}}$ supplied as the contrasts associated with the age effect. For example, for tests of the linear effect of age and the Sex:age interaction (differences in slopes),

```
R> linear <- idata
R> contrasts(linear$age, 1) <- contrasts(linear$age)[, 1]
R> print(linearHypothesis(Ortho.mod, hypothesis = "(Intercept)",
+    idata = linear, idesign = ~ age, iterms = "age", title = "Linear age"),
+    SSP = FALSE)

Multivariate Tests: Linear age
                 Df test stat approx F num Df den Df     Pr(>F)
Pillai            1   0.76892   83.187      1     25 1.9862e-09 ***
Wilks             1   0.23108   83.187      1     25 1.9862e-09 ***
Hotelling-Lawley  1   3.32748   83.187      1     25 1.9862e-09 ***
Roy               1   3.32748   83.187      1     25 1.9862e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R> print(linearHypothesis(Ortho.mod, hypothesis = "SexFemale", idata = linear,
+    idesign = ~ age, iterms = "age", title = "Linear Sex:age"), SSP = FALSE)

Multivariate Tests: Linear Sex:age
                 Df test stat approx F num Df den Df   Pr(>F)
Pillai            1  0.169948   5.1186      1     25 0.032614 *
Wilks             1  0.830052   5.1186      1     25 0.032614 *
Hotelling-Lawley  1  0.204744   5.1186      1     25 0.032614 *
Roy               1  0.204744   5.1186      1     25 0.032614 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Similarly, the nonlinear effects of `age` and the `Sex:age` interaction can be tested as follows, using the contrasts for quadratic and cubic trends in age.

```
R> nonlin <- idata
R> contrasts(nonlin$age, 2) <- contrasts(nonlin$age)[, 2:3]
R> print(linearHypothesis(Ortho.mod, hypothesis = "(Intercept)",
+    idata = nonlin, idesign = ~ age, iterms = "age",
+    title = "Nonlinear age"), SSP = FALSE)

Multivariate Tests: Nonlinear age
                 Df test stat approx F num Df den Df  Pr(>F)
Pillai            1  0.103180  1.38061      2     24 0.27069
Wilks             1  0.896820  1.38061      2     24 0.27069
Hotelling-Lawley  1  0.115051  1.38061      2     24 0.27069
Roy               1  0.115051  1.38061      2     24 0.27069
```

```
R> print(linearHypothesis(Ortho.mod, hypothesis = "SexFemale", idata = nonlin,
+    idesign = ~ age, iterms = "age", title = "Nonlinear Sex:age"),
+    SSP = FALSE)

Multivariate Tests: Nonlinear Sex:age
                 Df test stat approx F num Df den Df  Pr(>F)
Pillai            1  0.052578 0.665952      2     24 0.52302
Wilks             1  0.947422 0.665952      2     24 0.52302
Hotelling-Lawley  1  0.055496 0.665952      2     24 0.52302
Roy               1  0.055496 0.665952      2     24 0.52302
```

These examples show that, in simple cases, traditional plots of fitted values from mixed models (Figure 14) have the advantage of simple visual interpretation in terms of slopes and intercepts, at least for linear models. But such profile plots are generic: comparable plots could equally well be drawn for the fitted values from the MVLM in this section. HE plots for repeated measure designs have the additional advantage of showing the *nature* of significance tests and linear hypotheses, though the structure of the MVLM requires separate plots of between-S and within-S effects. In more complex designs with multiple between-S and within-S effects, and designs with multivariate responses (where mixed models do not apply), HE plots gain greater advantage.

8. Discussion

Graphical methods for univariate response models are well-developed, but analogous methods for multivariate responses are still developing. Indeed, this is a fruitful area for new research (Friendly 2007). HE plots provide one new direction, providing direct visualizations of effects in MVLMs, in the space of response variables (or in the the reduced-rank canonical space (**candisc**) displaying maximal differences).

This paper has shown how these methods can be extended to repeated measures designs, by displaying effects in the transformed space of contrasts or linear combinations for within-S effects. As we hope to have shown, these plots can provide insights into the relations among effects and interpretations of those that are significant (or not) which go beyond what is available in traditional, univariate displays.

Acknowledgments

I am grateful to John Fox for collaboration, advice and continued development of the **car** which provides the computational infrastructure for the **heplots**. The Associate Editor and one reviewer helped considerably in clarifying the presentation of these methods.

References

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

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

Fox J (2003). "Effect Displays in R for Generalised Linear Models." *Journal of Statistical Software*, **8**(15), 1–18. URL http://www.jstatsoft.org/v08/i15/.

Fox J, Friendly M, Monette G (2009a). ***heplots**: Visualizing Tests in Multivariate Linear Models*. R package version 0.8-11, URL http://CRAN.R-project.org/package=heplots.

Fox J, Friendly M, Monette G (2009b). "Visualizing Hypothesis Tests in Multivariate Linear Models: The **heplots** Package for R." *Computational Statistics*, **24**(2), 233–246.

Fox J, Hong J (2009). "Effect Displays in R for Multinomial and Proportional-Odds Logit Models: Extensions to the **effects** Package." *Journal of Statistical Software*, **32**(1), 1–24. URL http://www.jstatsoft.org/v32/i01/.

Fox J, Weisberg S (2009). ***car**: Companion to Applied Regression*. R package version 1.2-16, URL http://CRAN.R-project.org/package=car.

Friendly M (2006). "Data Ellipses, HE Plots and Reduced-Rank Displays for Multivariate Linear Models: SAS Software and Examples." *Journal of Statistical Software*, **17**(6), 1–42. URL http://www.jstatsoft.org/v17/i06/.

Friendly M (2007). "HE plots for Multivariate General Linear Models." *Journal of Computational and Graphical Statistics*, **16**(2), 421–444.

Friendly M, Fox J (2010). ***candisc**: Generalized Canonical Discriminant Analysis*. R package version 0.5-19, URL http://CRAN.R-project.org/package=candisc.

Hsu PL (1938). "Notes on Hotelling's Generalized T." *Annals of Mathematical Statistics*, **9**, 231–243.

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

O'Brien RG, Kaiser MK (1985). "MANOVA Method for Analyzing Repeated Measures Designs: An Extensive Primer." *Psychological Bulletin*, **97**, 316–333.

Pinheiro JC, Bates DM (2000). *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag, New York, NY.

Potthoff RF, Roy SN (1964). "A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems." *Biometrika*, **51**, 313–326.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rencher AC (1995). *Methods of Multivariate Analysis*. John Wiley & Sons, New York, NY.

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

Timm NH (1980). "Multivariate Analysis of Variance of Repeated Measurements." In PR Krishnaiah (ed.), *Handbook of Statistics*, volume 1, chapter 2, pp. 41–87. North-Holland, Amsterdam.

Verbeke G, Molenberghs G (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York, NY.

Warnes GR (2010). ***gplots**: Various R Programming Tools for Plotting Data*. R package version 2.8.0, URL http://CRAN.R-project.org/package=gplots.

Affiliation:

Michael Friendly
Psychology Department
York University
Toronto, ON, M3J 1P3, Canada
E-mail: friendly@yorku.ca
URL: http://datavis.ca/

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 37, Issue 4 *Submitted:* 2009-12-22
November 2010 *Accepted:* 2010-06-19