

Visualizing hypothesis tests in multivariate linear models: the *heplots* package for R

John Fox · Michael Friendly · Georges Monette

Received: 1 March 2007 / Accepted: 29 April 2008 / Published online: 15 May 2008
© Springer-Verlag 2008

Abstract Hypothesis-error (or “HE”) plots, introduced by Friendly (J Stat Softw 17(6):1–42, 2006a; J Comput Graph Stat 16:421–444, 2006b), permit the visualization of hypothesis tests in multivariate linear models by representing hypothesis and error matrices of sums of squares and cross-products as ellipses. This paper describes the implementation of these methods in the **heplots** package for R, as well as their extension, for example from two to three dimensions and by scaling hypothesis ellipses and ellipsoids in a natural manner relative to error.

Keywords Multivariate linear model · MANOVA · Linear hypotheses · Data ellipses · Statistical graphics

1 Introduction

This paper introduces the **heplots** package for R, which implements and extends the methods described in Friendly (2006a,b) for visualizing hypothesis tests in multivariate linear models. The paper begins with a brief description of multivariate linear

This is a paper for the proceedings of the Directions in Statistical Computing conference.

J. Fox (✉)
Department of Sociology, McMaster University, 1280 Main Street West,
Hamilton, ON, Canada L8S 4M4
e-mail: jfox@mcmaster.ca

M. Friendly
Department of Psychology, York University, Toronto, ON, Canada M3J 1P3

G. Monette
Department of Mathematics and Statistics, York University, Toronto, ON, Canada M3J 1P3

models; proceeds to explain how dispersion matrices can be represented by ellipses or ellipsoids; and illustrates the use of the **heplots** package for two and three-dimensional visualization of hypothesis tests in multivariate analysis of variance and regression.

2 Multivariate linear models

The multivariate linear model (e.g., [Timm 1975](#)) is given by the equation $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$, where \mathbf{Y} is an $n \times m$ matrix of responses in which each column represents a distinct response variable; \mathbf{X} is an $n \times p$ model matrix of full column rank that is either fixed or, if random, independent of the $n \times m$ matrix of errors \mathbf{E} ; and \mathbf{B} is a $p \times m$ matrix of regression coefficients. Under the assumption that the rows of \mathbf{E} are independent, and that each row is multivariately normally distributed with zero expectation and common covariance matrix, $\boldsymbol{\varepsilon}_i^T \sim \mathbf{N}_m(\mathbf{0}, \boldsymbol{\Sigma})$ or equivalently $\text{vec}(\mathbf{E}) \sim \mathbf{N}_{n \times m}(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma})$, the least squares estimator $\widehat{\mathbf{B}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ is the maximum-likelihood estimator of \mathbf{B} . Here, the $\mathbf{0}$ vectors are respectively of order $n \times 1$ and $np \times 1$, and \otimes represents the Kronecker product.

Hypothesis tests for multivariate linear models closely parallel those for univariate linear models. Consider the linear hypothesis

$$H_0: \mathbf{L}\mathbf{B} = \mathbf{0} \tag{1}$$

where \mathbf{L} is a $q \times p$ hypothesis matrix of pre-specified constants and $\mathbf{0}$ is the $q \times m$ zero matrix. We compute the $m \times m$ hypothesis sum of squares and products matrix $\mathbf{SSP}_H = \widehat{\mathbf{B}}^T \mathbf{L}^T [\mathbf{L}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{L}^T]^{-1} \mathbf{L} \widehat{\mathbf{B}}$ and the $m \times m$ error sum of squares and products matrix $\mathbf{SSP}_E = \widehat{\mathbf{E}}^T \widehat{\mathbf{E}}$, where $\widehat{\mathbf{E}} = \mathbf{Y} - \mathbf{X} \widehat{\mathbf{B}}$ is the matrix of residuals. Multivariate tests of the hypothesis in Eq. 1 are based on the $s = \min(q, m)$ nonzero latent roots $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s$ of the matrix \mathbf{SSP}_H relative to the matrix \mathbf{SSP}_E , that is, the values of λ for which $\det(\mathbf{SSP}_H - \lambda \mathbf{SSP}_E) = 0$. These are also the ordinary latent roots of $\mathbf{SSP}_H \mathbf{SSP}_E^{-1}$, that is, the values of λ for which $\det(\mathbf{SSP}_H \mathbf{SSP}_E^{-1} - \lambda \mathbf{I}_m) = 0$. 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 Eq. 1. The several commonly employed multivariate test statistics—Pillai's trace, the Hotelling-Lawley trace, Wilks's Lambda, and Roy's maximum root—are functions of the latent roots. There is an F approximation to the null distribution of each of these test statistics.

In a univariate 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 for these tests can be expressed as differences in the error sums of squares 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 multivariate analysis-of-variance (MANOVA) tables can be formed similarly by taking differences in error sum of squares and products matrices. We base the examples in Sect. 4 on Type-II

tests because we feel that these are most generally sensible, but the methods presented in the paper apply to any linear hypotheses.

3 Data ellipses and ellipsoids

The data ellipse, described by [Dempster \(1969\)](#) and [Monette \(1990\)](#), is a device for visualizing the relationship between two variables, Y_1 and Y_2 . Let $D_M^2(\mathbf{y}) = (\mathbf{y} - \bar{\mathbf{y}})^T \mathbf{S}^{-1} (\mathbf{y} - \bar{\mathbf{y}})$ represent the squared Mahalanobis distance of the point $\mathbf{y} = (y_1, y_2)^T$ from the centroid of the data $\bar{\mathbf{y}} = (\bar{Y}_1, \bar{Y}_2)^T$. The data ellipse \mathcal{E}_c of size c is the set of all points \mathbf{y} with $D_M^2(\mathbf{y})$ less than or equal to c^2 :

$$\mathcal{E}_c(\mathbf{y}; \mathbf{S}, \bar{\mathbf{y}}) \equiv \left\{ \mathbf{y}: (\mathbf{y} - \bar{\mathbf{y}})^T \mathbf{S}^{-1} (\mathbf{y} - \bar{\mathbf{y}}) \leq c^2 \right\} \tag{2}$$

Here, $\mathbf{S} = \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T / (n - 1)$ is the sample covariance matrix, where \mathbf{y}_i is the i th row of \mathbf{Y} written as a column vector.

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, $D_M^2(\mathbf{y})$ 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.

The generalization of the data ellipse to more than two variables is immediate: applying Eq. 2 to $\mathbf{y} = (y_1, y_2, y_3)^T$, for example, produces a data ellipsoid in three dimensions. For m multivariate-normal variables, selecting $c^2 = \chi_m^2(1 - \alpha)$ encloses approximately $100(1 - \alpha)$ percent of the data.

4 Hypothesis-error (HE) plots

Hypothesis-error (or HE) plots use ellipses to represent hypothesis and error sums of squares and product matrices. The plots are implemented in two and three dimensions in the *heplots* package for R.

The error ellipse is obtained by dividing \mathbf{SSP}_E by the error degrees of freedom $n - p$, producing a data ellipse for the residuals. The \mathbf{SSP}_E ellipse is also centered at the grand means, allowing individual factor means to be shown on the same plot, facilitating interpretation. Similarly, the hypothesis ellipse is formed from the sum-of-squares-and-products matrix for the hypothesis, \mathbf{SSP}_H . We consider two scalings of the hypothesis ellipse:

1. ‘‘Evidence-based’’ scaling, the default, in which the hypothesis ellipse protrudes from the error ellipse if and only if the hypothesis can be rejected by the Roy maximum-root criterion. The directions in which the hypothesis ellipse exceed the error ellipse are informative about the responses or their linear combinations that depart significantly from H_0 . This scaling is produced by dividing \mathbf{SSP}_H by $\lambda_\alpha(n - p)$, where λ_α is the critical value of Roy’s statistic for a test at level α .

- Scaling by “effect size,” where the hypothesis ellipse is put on the same scale as the error ellipse, and approximately represents the data ellipse of fitted values under the alternative hypothesis. Here, SSP_H is simply divided by $n - p$.

All of this extends straightforwardly to the three dimensional case.

4.1 One-way MANOVA: Romano-British Pottery

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

The data for this example are in the data frame `Pottery` in the `car` package:

```
> library(car)
> Pottery

      Site  Al  Fe  Mg  Ca  Na
1  Llanedyrn 14.4 7.00 4.30 0.15 0.51
2  Llanedyrn 13.8 7.08 3.43 0.12 0.17
3  Llanedyrn 14.6 7.09 3.88 0.13 0.20
. . .
25 AshleyRails 14.8 2.74 0.67 0.03 0.05
26 AshleyRails 19.1 1.64 0.60 0.10 0.03

> table(Pottery$Site)

AshleyRails    Caldicot  IsleThorns    Llanedyrn
           5             2             5             14
```

The ellipses in the output (`. . .`) represent elided lines.

In R, multivariate linear models are fit by the `lm` function, returning an object of class `mlm`. Here, we fit a one-way MANOVA model to the `Pottery` data. The `print` method for the object returned by the `Anova` function gives a brief display of the multivariate test for `Site`, using the Pillai trace statistic by default. A more detailed display (not shown), including the SSP_H and SSP_E matrices, all four multivariate tests, and univariate tests for each response, is provided by the `summary` method for `Anova.mlm` objects:

```
> pottery.mod <- lm(cbind(Al, Fe, Mg, Ca, Na) ~ Site, data=Pottery)
> Anova(pottery.mod)

Type II MANOVA Tests: Pillai test statistic
      df test stat approx F num df den df    Pr(>F)
Site  3    1.5539    4.2984    15     60 2.413e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

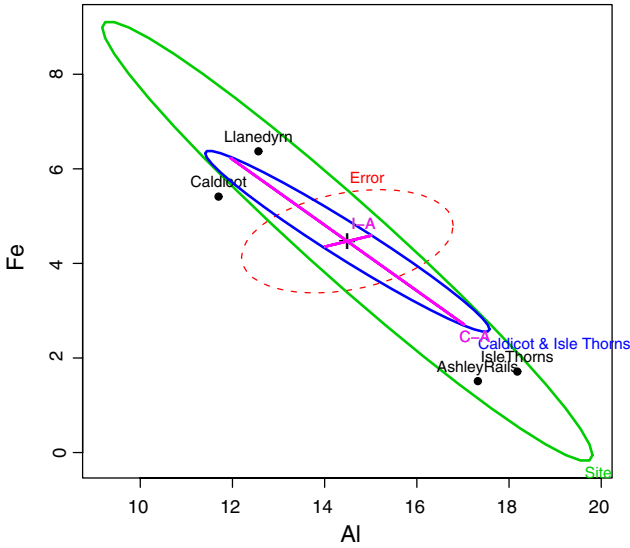


Fig. 1 HE plot for pottery data, showing iron and aluminum. H (i.e., SSP_H) matrices *solid lines*; E (i.e., SSP_E) matrix *dashed lines*

There is, therefore, strong evidence against the null hypothesis of no differences in mean vectors across sites. In this instance, we get the same test from the `anova` function in the standard `stats` package, because (as mentioned) for this one-factor design, the sequential test provided by `anova` is the same as the Type-II test provided by default by `Anova` (but objects returned by `anova` do not contain the information necessary for constructing HE plots).

For two response variables, we can use `heplot` to visualize the (co)variation due to the `Site` hypothesis—(co)variation of the group means—relative to error—(co)variation of the residuals. As illustrated below, we can also visualize the results of linear hypotheses related to subsets of the parameters.

```
> # Figure 1
> heplot(pottery.mod)
> # add a 2 df hypothesis
> heplot(pottery.mod, terms=FALSE, add=TRUE, col="blue",
+ hypotheses=list("Caldicot & Isle Thorns" = c("SiteCaldicot = 0", "SiteIsleThorns=0")))
> # add two 1 df hypotheses
> heplot(pottery.mod, terms=FALSE, add=TRUE, col="magenta",
+ hypotheses=list("C-A" = "SiteCaldicot", "I-A" = "SiteIsleThorns"))
```

Figure 1 shows the plot of the SSP_H and SSP_E matrices for two of the variables, iron and aluminum, and the means for the four sites. In addition, it shows the SSP_H matrices for linear hypotheses related to Caldicot and IsleThorns, where Ashley Rails is the baseline category. It is clear that the sites differ primarily in terms of a contrast between Caldicot and Llanedryn versus Ashley Rails and Isle Thorns, and that the means on these two chemical components are negatively related, while the pooled within group scatter shows a weak positive relation.

Figure 2 illustrates the difference between effect-size scaling and (the default) evidence-based scaling of SSP_H relative to SSP_E , produced using

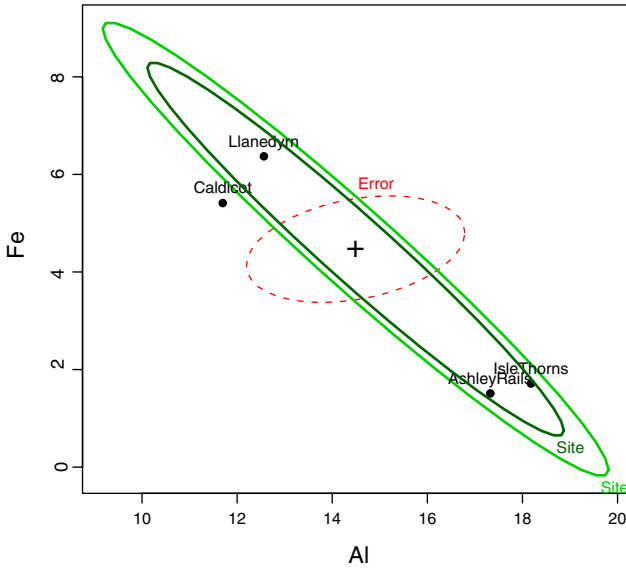


Fig. 2 HE plot for pottery data: effect scaling (*dark green*) versus evidence scaling (*light green*)

```
> # Figure 2: compare evidence and effect scaling
> heplot(pottery.mod)
> heplot(pottery.mod, add=TRUE, size="effect", col="darkgreen")
```

The evidence-scaled hypothesis ellipse for this one-way MANOVA model is the data ellipse for the group means weighted by group sample sizes.

Of course, other pairs of response variables can also be displayed (e.g., `heplot(pottery.mod, variables=c("Mg", "Fe"))`), or subsets of three response variables can be examined (by `heplot3d(pottery.mod, variables=c(...))`—see below). Alternatively, the variation across sites on *all* chemical components may be seen in the pairwise 2D projections of the HE plot matrix in Fig. 3. This graph was produced by the `pairs` method for `mlm` objects, reordering the variables to produce a more coherent display (Friendly and Kwan 2003):

```
> # Figure 3
> pairs(pottery.mod, variables=c("Mg", "Fe", "Ca", "Na", "Al"))
```

Quite a lot may be read directly from this plot. For example: the site means for magnesium (Mg) and iron (Fe) are nearly perfectly correlated, and have the same pattern with all other variables, while all mean differences for aluminium (Al) are in the opposite direction. The relations for calcium (Ca) and sodium (Na) also differ somewhat from those for magnesium and iron in that Caldicot samples are quite high on calcium, while Llanedryn is high on sodium.

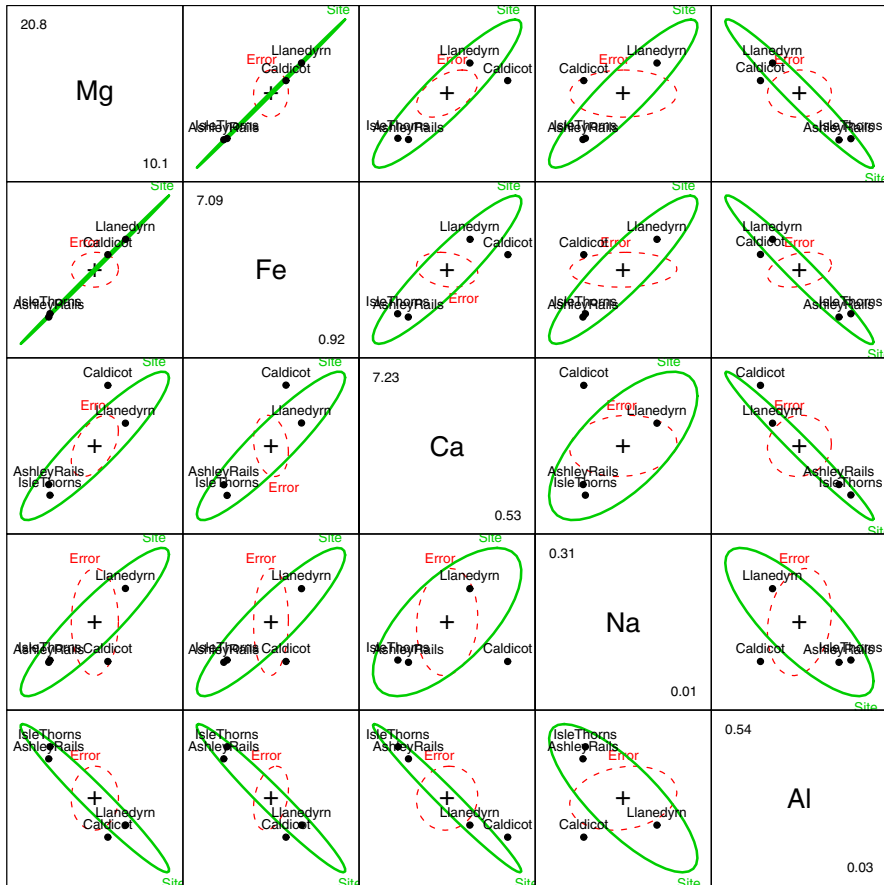


Fig. 3 pairs HE plot matrix for the pottery data

4.2 Two-way MANOVA: plastic film data

HE plots are particularly instructive when there are multiple sources of hypothesis variation to be tested in a multivariate linear model. The simplest case is for a 2×2 MANOVA, where the main effects and interaction each have 1 df (and so, the SSP_H ellipses or ellipsoids collapse to lines), but where the response variable space is 2 or more dimensional. To illustrate, we use textbook data from [Johnson and Wichern \(1992, 266\)](#) on an experiment conducted to determine the optimum conditions for extruding plastic film. Three responses (tear resistance, film gloss, and opacity) were measured in relation to two factors: rate of extrusion (Low/High) and amount of an additive (Low/High). The data are in the **heplots** package:

```
> library(heplots)
> Plastic
```

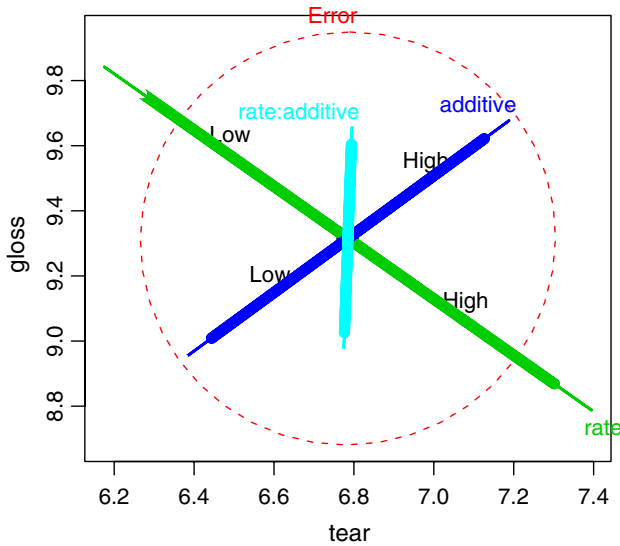


Fig. 4 HE plot for tear and gloss in the plastic-film data. *Thick lines* evidence scaling; *thin lines* effect scaling

	tear	gloss	opacity	rate	additive
1	6.5	9.5	4.4	Low	Low
2	6.2	9.9	6.4	Low	Low
3	5.8	9.6	3.0	Low	Low
. . .					
19	7.5	10.1	2.7	High	High
20	7.6	9.2	1.9	High	High

We fit the two-way MANOVA model and display the Anova results, using Roy’s maximum root test. Both main effects are significant, but their interaction is not:

```
> plastic.mod <- lm(cbind(tear, gloss, opacity) ~ rate*additive, data=Plastic)
> Anova(plastic.mod, test.statistic="Roy")
```

```
Type II MANOVA Tests: Roy test statistic
              df test stat approx F num df den df Pr(>F)
rate          1  1.6188   7.5543     3    14 0.003034 **
additive      1  0.9119   4.2556     3    14 0.024745 *
rate:additive 1  0.2868   1.3385     3    14 0.301782
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Again, we get the same tests from anova, this time because the data are balanced, with five observations per cell (so that sequential and Type-II tests coincide).

In Fig. 4, we show the HE plot for the first two response variables (tear and gloss). In this plot, we overlay the size="evidence" and size="effect" scalings, varying line width.

```
> # Figure 4: Compare evidence and effect scaling
> heplot(plastic.mod, size="evidence")
> heplot(plastic.mod, size="effect", add=TRUE, lwd=8, term.labels=FALSE)
```

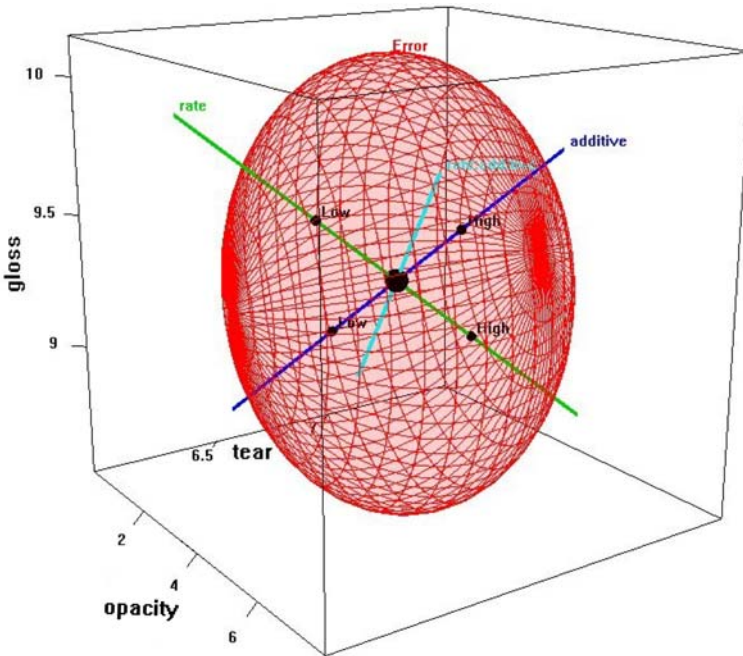



Fig. 5 `heplot3d` plot for the plastic-film data

Note that, in this view, the effect for `additive` does not extend outside the error ellipse. Indeed, the effect of `additive`, while significant in the multivariate test, does not quite protrude beyond the SSP_E ellipse in *any* of the 2D marginal projections of the multidimensional ellipsoids (as revealed, e.g., in a `pairs` plot, not shown). Using `heplot3d`, however, we can easily find 3D views of all effects that show the significant effects of both `additive` and `rate`, as in Fig. 5. These views correspond to linear combinations of the responses.

```
> # Figure 5
> heplot3d(plastic.mod)
```

The view in Fig. 5 facilitates interpretation of the multivariate factor effects. For example, for the `rate` effect, higher rate is associated with greater opacity and greater tear but lower gloss.

4.3 Multivariate multiple regression and MANCOVA: Rohwer data

The ideas behind HE plots extend naturally to multivariate multiple regression and multivariate analysis of covariance. In multivariate multiple regression, the \mathbf{X} matrix contains quantitative predictors, while in multivariate analysis of covariance (MANCOVA), there is a mixture of factors and quantitative predictors (covariates). To illustrate, we use data from a study by Rohwer (given in Timm 1975: Ex. 4.3, 4.7, and 4.23) on kindergarten children, designed to determine how well a set of paired-associate (PA)

tasks predicted performance on the Peabody Picture Vocabulary test (PPVT), a student achievement test (SAT), and the Raven Progressive matrices test (Raven). The PA tasks varied in how the stimuli were presented, and are called *named* (n), *still* (s), *named still* (ns), *named action* (na), and *sentence still* (ss). Two groups were tested: a group of $n = 37$ children from a low socioeconomic status (SES) school, and a group of $n = 32$ high SES children from an upper-class, white residential school. The data are in the data frame `Rohwer` in the **heplots** package:

```
> Rohwer
```

```
      group SES SAT PPVT Raven  n  s ns na ss
1         1  Lo  49   48     8  1  2  6 12 16
2         1  Lo  47   76    13  5 14 14 30 27
3         1  Lo  11   40    13  0 10 21 16 16
. . .
68        2  Hi  98   74    15  2  6 14 25 17
69        2  Hi  50   78    19  5 10 18 27 26
```

Initially (and optimistically), we fit the MANCOVA model that allows different means for the two SES groups on the responses, but constrains the slopes for the PA covariates to be equal.

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

Type II MANOVA Tests: Pillai test statistic

	df	test stat	approx F	num df	den df	Pr(>F)
SES	1	0.3785	12.1818	3	60	2.507e-06 ***
n	1	0.0403	0.8400	3	60	0.477330
s	1	0.0927	2.0437	3	60	0.117307
ns	1	0.1928	4.7779	3	60	0.004729 **
na	1	0.2313	6.0194	3	60	0.001181 **
ss	1	0.0499	1.0504	3	60	0.376988

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

This multivariate linear model is of interest because, although the multivariate tests for two of the covariates (ns and na) are highly significant, univariate multiple regression tests for the separate responses [from `summary(rohwer.mod)`] are relatively weak. We can test the 5 *df* hypothesis that *all* covariates have null effects for all responses as a linear hypothesis (suppressing display of the error and hypothesis SSP matrices),

```
> Reqr <- linear.hypothesis(rohwer.mod, diag(7)[3:7,])
> print(Reqr, digits=5, SSP=FALSE)
```

Multivariate Tests:

	df	test stat	approx F	num df	den df	Pr(>F)
Pillai	5.00	0.6658	3.5369	15.00	186.00	2.309e-05 ***
Wilks	5.00	0.4418	3.8118	15.00	166.03	8.275e-06 ***
Hotelling-Lawley	5.00	1.0309	4.0321	15.00	176.00	2.787e-06 ***

```
Roy          5.00    0.7574    9.3924    5.00    62.00 1.062e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As explained, in the MANCOVA model `rohwer.mod` we have assumed homogeneity of slopes for the predictors, and the test of SES relies on this assumption. We can test this as follows, adding interactions of SES with each of the covariates:

```
> rohwer.mod2 <- lm(cbind(SAT, PPVT, Raven) ~ SES * (n + s + ns + na + ss),
+                   data=Rohwer)
> Anova(rohwer.mod2)
```

Type II MANOVA Tests: Pillai test statistic

	df	test stat	approx F	num df	den df	Pr(>F)
SES	1	0.3912	11.7822	3	55	4.55e-06 ***
n	1	0.0790	1.5727	3	55	0.2063751
s	1	0.1252	2.6248	3	55	0.0595192 .
ns	1	0.2541	6.2461	3	55	0.0009995 ***
na	1	0.3066	8.1077	3	55	0.0001459 ***
ss	1	0.0602	1.1738	3	55	0.3281285
SES:n	1	0.0723	1.4290	3	55	0.2441738
SES:s	1	0.0994	2.0240	3	55	0.1211729
SES:ns	1	0.1176	2.4425	3	55	0.0738258 .
SES:na	1	0.1480	3.1850	3	55	0.0308108 *
SES:ss	1	0.0573	1.1150	3	55	0.3509357

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It appears from the above that there is only weak evidence of unequal slopes from the separate SES: terms. The evidence for heterogeneity is stronger, however, when these terms are tested collectively using the `linear.hypothesis` function:

```
> (coefs <- rownames(coef(rohwer.mod2)))

[1] "(Intercept)" "SESLo"      "n"          "s"          "ns"
[6] "na"          "ss"        "SESLo:n"    "SESLo:s"    "SESLo:ns"
[11] "SESLo:na"    "SESLo:ss"

> print(linear.hypothesis(rohwer.mod2, coefs[grep(":", coefs)]), SSP=FALSE)
```

Multivariate Tests:

	df	test stat	approx F	num df	den df	Pr(>F)
Pillai	5.0000	0.417938	1.845226	15.0000	171.0000	0.0320861 *
Wilks	5.0000	0.623582	1.893613	15.0000	152.2322	0.0276949 *
Hotelling-Lawley	5.0000	0.538651	1.927175	15.0000	161.0000	0.0239619 *
Roy	5.0000	0.384649	4.384997	5.0000	57.0000	0.0019053 **

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2D views of the additive MANCOVA model that we fit to the Rohwer data and the overall test for all covariates are provided in Fig. 6, produced using `pairs` as follows:

```
> # Figure 6
> colors <- c("red", "blue", rep("black",5), "darkgrey")
> pairs(rohwer.mod, col=colors,
+       hypotheses=list("Regr" = c("n", "s", "ns", "na", "ss")))
```

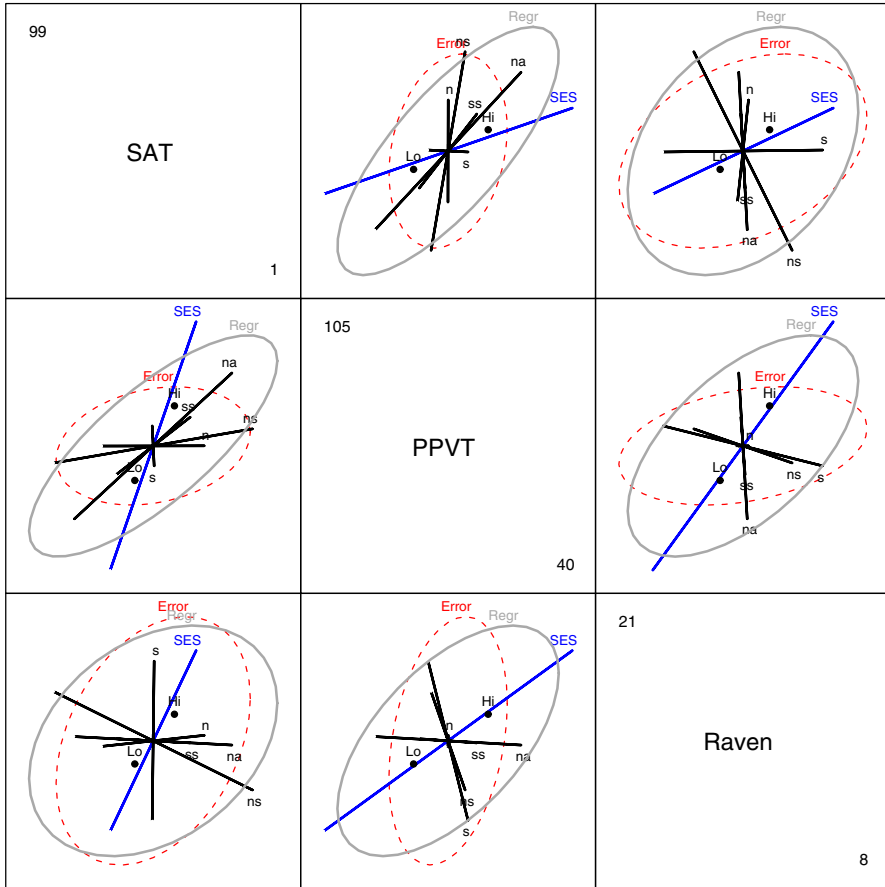


Fig. 6 HE pairs plot for the Rohwer data, MANCOVA model *rohwer.mod*.

It may be seen that the predicted values for all three responses are positively correlated, and that the hypothesized effects of the covariates span the full three dimensions of the responses. As well, the High SES group is higher on all responses than the Low SES group.

We also fit a model to the Rohwer data relaxing the assumption of equal slopes—that is, permitting interactions between the covariates and SES. There are several options for visualization: Either we can fit and display separate models for the High and Low SES groups (which also allows the within-groups error-covariance matrices to differ); we can fit a combined model with separate intercept and slopes for the two groups, which assumes a common within-groups error-covariance matrix; or we can try to visualize the slope differences in the heterogeneous-slopes model *rohwer.mod2*. Choosing the last option, we examine the HE pairs plot in Fig. 7. To simplify this display, we show the hypothesis ellipses for the overall effects of the PA tests in the baseline high-SES group, and a single combined ellipse for all the SES_{Lo} : interaction terms that we tested previously, representing differences in slopes between the low and high-SES groups. Because SES is “treatment-coded” in this model, the ellipse for

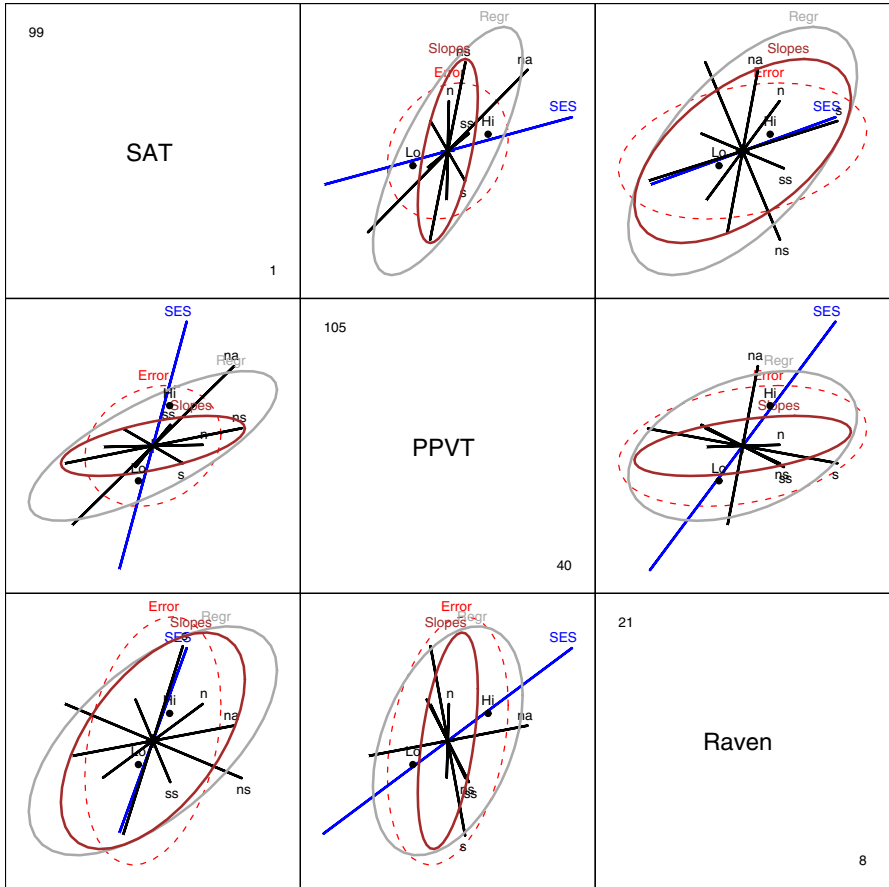


Fig. 7 pairs plot for the heterogeneous regression model `rohwer.mod2`. The ellipses labeled “Slopes” show the covariation of all terms for slope differences between the High and Low SES groups

each covariate represents the hypothesis that the slopes for that covariate are zero in the high-SES baseline category.

```
> # Figure 7
> pairs(rohwer.mod2, col=c(colors, "brown"),
+       terms=c("SES", "n", "s", "ns", "na", "ss"),
+       hypotheses=list("Regr" = c("n", "s", "ns", "na", "ss"),
+                         "Slopes" = coefs[grep(":", coefs)]))
```

Comparing Figs. 6 and 7 for the homogeneous and heterogeneous-slopes models, it may be seen that most of the covariates have ellipses of similar size and orientation, reflecting similar evidence against the null hypotheses in the baseline high-SES group and for both groups in the common-slopes model; so too does the effect of SES, with the High SES group performing better on all measures. The error covariation is noticeably smaller in some of the panels of Fig. 7 (those for SAT and PPVT), reflecting additional variation accounted for by differences in slopes.

5 Discussion

As we hope the previous examples illustrate, HE plots provide a remarkably simple, yet powerful visual tool for understanding the *nature* of effects expressed in multivariate test statistics for the multivariate linear model. We can visualize the size, dimensionality and orientation of hypothesis and error variation for two or more response variables by plotting what are essentially data ellipses (or ellipsoids) for the fitted values (from \mathbf{SSP}_H) versus residuals (from \mathbf{SSP}_E) corresponding to any model term. This is precisely the information that is *summarized* in multivariate test statistics, but the HE plot *exposes* the pattern of model effects in relation to the response variables. As we described, Roy's test provides a particularly convenient "evidence" scaling that yields a visual test of significance.

An attractive feature of our implementation in the **heplots** package is its generality. The method applies to any collection of terms—factors, covariates, interactions and so forth—in any multivariate linear model fit with `lm`. In addition, we have found it useful to superimpose hypothesis ellipses corresponding to general linear hypotheses, such as the contrasts shown in Fig. 1 and the composite tests for all regressors ("Regr") and all slope differences ("Slopes") shown in Figs. 6 and 7. For three responses, we can visualize these effects in 3D with `heplot3d`; for three or more responses, we provide a scatterplot matrix analog with a `pairs` method for `mlm` objects.

References

- Dempster AP (1969) Elements of continuous multivariate analysis. Addison-Wesley, Reading, MA
- Friendly M, Kwan E (2003) Effect ordering for data displays. *Computational Statistics and Data Analysis* 43:509–539
- Friendly M (2006a) Data ellipses, HE plots and reduced-rank displays for multivariate linear models: SAS software and examples. *J Stat Softw* 17(6):1–42
- Friendly M (2006b) HE plots for multivariate general linear models. *J Comput Graph Stat* 16:421–444
- Hand D, Daly F, Lunn AD, McConway KJ, Ostrowski E (1994) *A Handbook of Small Data Sets*. Chapman and Hall, London
- Johnson RA, Wichern DW (1992) *Applied multivariate statistical analysis*, 3rd edn. Prentice Hall, Englewood Cliffs
- Monette G (1990) Geometry of multiple regression and interactive 3-D graphics. In: Fox J, Long JS (eds) *Modern methods of data analysis*. Sage, Beverly Hills, pp 209–256
- Timm NH (1975) *Multivariate analysis with applications in education and psychology*. Wadsworth (Brooks/Cole), Belmont
- Tubb A, Parker A, Nickless G (1980) The analysis of Romano-British pottery by atomic absorption spectrophotometry. *Archaeometry* 22:153–171