

# Visualizing Tests for Equality of Covariance Matrices

Supplementary Materials

*Michael Friendly and Matthew Sigal*

*May 16, 2018*

## Contents

<b>Introduction</b>	<b>1</b>
<b>1 Visualizing mean differences: The HE plot framework</b>	<b>2</b>
1.1 HE plot example: Wine data . . . . .	3
1.2 Canonical views . . . . .	3
<b>2 Multivariate generalization of Levene-type tests</b>	<b>5</b>
2.1 HE plots for multivariate Levene-type tests . . . . .	7
2.2 Canonical views . . . . .	8
<b>3 Distance based approaches</b>	<b>9</b>
3.1 Example: Iris data . . . . .	10
3.2 Example: Skulls data . . . . .	13
<b>References</b>	<b>14</b>

## Introduction

This is the supplemental appendix to “Visualizing Tests for Equality of Covariance Matrices.” It covers topics of interest that were considered too long or not sufficiently essential to include in the paper. Section 1 gives a brief, capsule summary of the HE plot framework for visualizing multivariate tests of means using Hypothesis-Error plots and related canonical discriminant plots. An example is presented using the `Wine` data. Section 2 presents a multivariate generalization of Levene’s test for homogeneity of variance which gives new graphical displays within the HE plot framework. Section 3 presents a different approach to questions of multivariate homogeneity of variance using analyses of distances from a centroid to measure multivariate dispersion.

# 1 Visualizing mean differences: The HE plot framework

For univariate response models, there is a variety of effective graphical methods for visualizing differences among group means, ranging from simple dot charts (with error bars), to boxplots that provide more details, to effect plots (Fox, 2003) for complex designs.

In the MANOVA setting, a collection of such univariate plots for each of the responses can be useful, but they do not show how the responses vary *jointly*. In this section we introduce some *multivariate* graphical methods (HE plots and canonical views) for the comparison of means (for further details, see Fox, Friendly, & Monette, 2009; Friendly, 2007; and, Friendly & Sigal, 2014) that can also be usefully applied to the comparison of covariance matrices.

The essence of these methods is the Hypothesis–Error (HE) plot, and we call the collection of their applications the HE plot framework. The main ideas underlying this framework are:

- (a) All hypothesis tests of model terms or linear hypotheses in the MLM correspond to statistics based on the eigenvalues  $\lambda_i$  of the sum of squares and products (SSP) matrix for the hypothesis (the  $\mathbf{H}$  matrix) relative to the SSP matrix for error (the  $\mathbf{E}$  matrix), i.e., the  $s = \min(p, df_h)$  non-zero values in the eigenvalue decomposition,  $\mathbf{H}\mathbf{E}^{-1}\mathbf{v}_i = \lambda_i\mathbf{v}_i$ . They answer the question, “How big is  $\mathbf{H}$  relative to  $\mathbf{E}$ ?” and are analogous to univariate  $F$  statistics.
- (b) The  $\mathbf{H}$  matrix for any term  $t$  is essentially the  $p \times p$  SSP matrix of the *fitted* values,  $\widehat{\mathbf{Y}}_t^\top \widehat{\mathbf{Y}}_t$ , for that term.<sup>1</sup> The  $\mathbf{E}$  matrix is the SSP matrix of the *residuals* for the full model.
- (c) The  $\mathbf{H}$  and  $\mathbf{E}$  matrices can therefore be visualized as the data ellipsoids of the fitted values and residuals. In the `heplot()` function,  $\mathbf{H}$  and  $\mathbf{E}$  can be scaled to show the size of each effect; alternatively, the  $\mathbf{H}$  matrix can be scaled to provide a visual test of significance, in the sense that the  $\mathbf{H}$  ellipsoid projects outside the  $\mathbf{E}$  ellipsoid *iff* the effect is significant by Roy’s maximum root test.
- (d) For MANOVA designs, when there are more than a few response variables, all these effects may be readily viewed in the orthonormalized *canonical* space corresponding to the linear transformation  $\mathbf{Y}\mathbf{T} \mapsto \mathbf{Y}^* = \mathbf{Y}\mathbf{E}^{-1/2}\mathbf{V}$ . This is the view that shows the maximal differences among means.

Overall, this framework is incredibly versatile and can be utilized to visualize and summarize the important aspects from a plethora of designs.

---

<sup>1</sup>This uses so-called “Type II” sums of squares, calculated according to the principle of marginality, testing each term after all others, except ignoring the term’s higher-order relatives.

## 1.1 HE plot example: Wine data

This example illustrates the basic properties of Hypothesis-Error plots, using the Wine data analyzed in Section 4 of the main paper. Recall that this data set consists of 13 chemical properties of wine samples derived from three different cultivars of grapes grown in a region of Italy. The question is *how* do the cultivars differ in their means on these properties.

The standard MANOVA of differences in means is conducted as follows and shows a highly significant overall multivariate test of differences among regions on the response variables jointly.

```
Wine.mod <- lm(as.matrix(Wine[, -1]) ~ Cultivar, data=Wine)
Anova(Wine.mod, test="Roy")
```

Type II MANOVA Tests: Roy test statistic

	Df	test	stat	approx	F	num	Df	den	Df	Pr(>F)
Cultivar	2		9.0817		114.57		13		164	< 2.2e-16 ***

---

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

Follow-up univariate  $F$  tests for each of the responses show that the grape varieties differ substantially on all of these measures. However, we want to understand the *pattern* of these differences, individually and jointly. The boxplots of the data shown in Fig. 7 of the paper show the univariate differences, but don't help to understand the correlations of these differences.

```
pairs(Wine.mod, variables=1:5, fill=TRUE, fill.alpha=.1)
```

An HE plot for the first five response variables is shown in Figure 1 here as a scatterplot matrix for all pairs. In this plot, in each panel the co-variation of the means (the  $\mathbf{H}$  matrix) is shown by the blue ellipse. The eccentricity of the  $\mathbf{H}$  ellipse indicates the correlation of the means. For example, means on Alcohol are positively correlated with those on Ash and Mg. The red ellipse for the  $\mathbf{E}$  matrix shows the size and correlation of the residuals. As noted above, this plot has the property that the overall multivariate test of `Cultivar` differences is significant (by Roy's test) if the  $\mathbf{H}$  ellipse projects anywhere outside the  $\mathbf{E}$  ellipse. It can be seen that the group means are well-separated on all pairs of variables, but the pattern of these is somewhat complex, even for just five of the responses.

## 1.2 Canonical views

There are too many response variables to see them all at once in variable space, but with  $g = 3$  groups, the MANOVA has a very simple 2D representation in the canonical space.

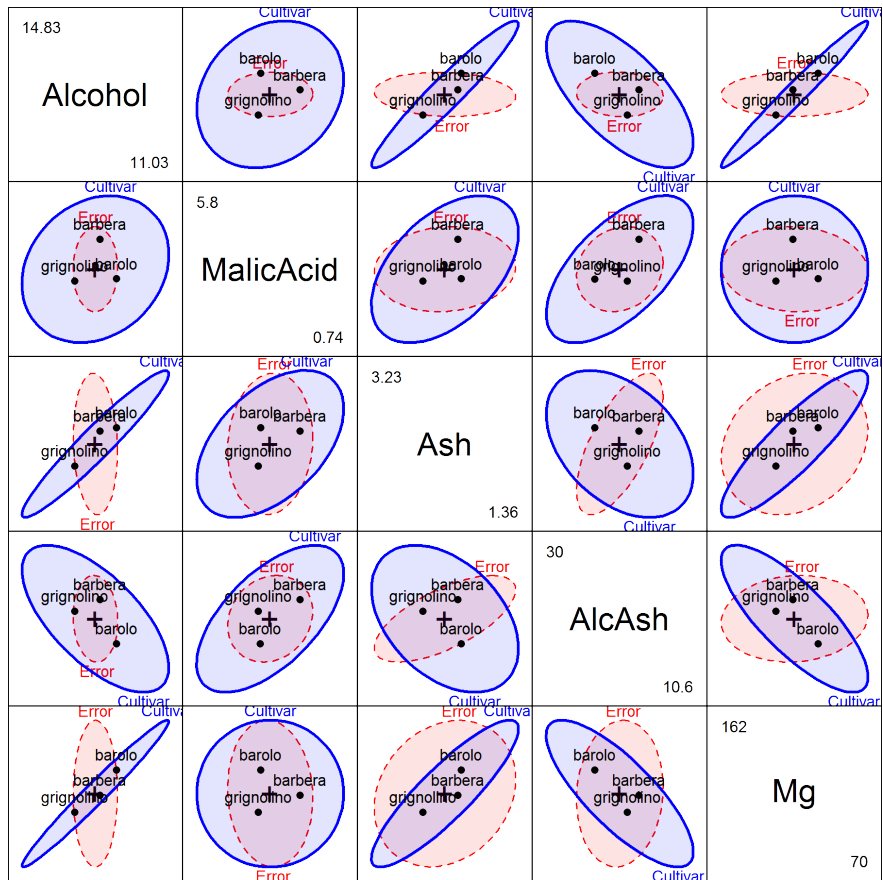


Figure 1: Pairwise HE plot for the first five response variables in the Wine data. The group means for each pair are labeled.

This is effectively a projection of the data into the space of the linear combinations of the responses that accounts for the greatest differences among the group means.

```
Wine.can <- candisc(Wine.mod)
Wine.can
```

Canonical Discriminant Analysis for Cultivar:

	CanRsq	Eigenvalue	Difference	Percent	Cumulative
1	0.90081	9.0817	4.9533	68.748	68.748
2	0.80501	4.1285	4.9533	31.252	100.000

Test of H0: The canonical correlations in the current row and all that follow are zero

LR test stat	approx F	numDF	denDF	Pr(> F)
--------------	----------	-------	-------	---------

```

1      0.019341   77.620    26   326 < 2.2e-16 ***
2      0.194990   56.422    12   164 < 2.2e-16 ***

```

```
---
```

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

The two canonical correlations are both large and highly significant. A 2D canonical plot (Figure 2) shows the canonical scores together with data ellipses for the scores and vectors for the structure coefficients. The second canonical dimension separates Grinolino wines from the other two regions. The first canonical dimension accounts for 69% of the group mean differences, and the groups are very well separated on this dimension. Longer vectors correspond to response variables that are better discriminated among the groups on this dimension. A guess at the interpretation of this dimension is that it contrasts those wines high in Malic acid and non-flavor phenols vs. those with high phenols and OD ratios.

```
plot(Wine.can, ellipse=TRUE, var.lwd=2)
```

We think of such canonical plots as a multivariate juicer for MANOVA problems: they show the data in the low-D view that extracts the most flavor regarding group differences, and also show the relations of the response variables to this space.

As well, note that in this canonical view, the data ellipses for the canonical scores for the groups have apparently different shapes, which speaks to the question of heterogeneity of covariance matrices. This is explored in the Section 5 of the paper in relation to our multivariate generalization of Levene-type tests.

## 2 Multivariate generalization of Levene-type tests

In the MANOVA setting, a multivariate version of Levene’s test (and robust Brown-Forsythe extensions) for the standard model  $\mathbf{y}_{ij} = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_{ij}$  considers instead the multivariate analog substituting  $\mathbf{z}_{ij}$  for  $\mathbf{y}_{ij}$ , where  $\mathbf{z}_{ij} = |\mathbf{y}_{ij} - \tilde{\mathbf{y}}_i|$  are the deviations from a “central” value (mean, median, trimmed mean)  $\tilde{\mathbf{y}}_i$  for group  $i, i = 1, 2, \dots, g; j = 1, 2, \dots, p$ . In matrix terms, this can be expressed as the linear model,

$$\mathbf{Z}_{N \times p} = \mathbf{X}_{(N \times g)} \mathbf{B}_{(g \times p)} + \mathbf{U}_{(N \times p)}$$

where  $\mathbf{X}$  is design matrix for the groups, typically composed of an intercept and  $g-1$  contrasts or dummy variables,  $\mathbf{B}$  is the matrix of parameters, and  $\mathbf{U}$  is the matrix of residuals.

The same multivariate tests (Wilks’  $\Lambda$ , Pillai trace, etc.) used to test differences in mean vectors  $\mathbf{y}$  can then be used to test for differences in the means of the centered deviations,  $\mathbf{z}$ . For example, this test applied to the Wine data gives Pillai trace = 0.7924, approx.  $F(26, 328) = 8.278, p < 0.000001$ . The overall multivariate test is highly significant, leading

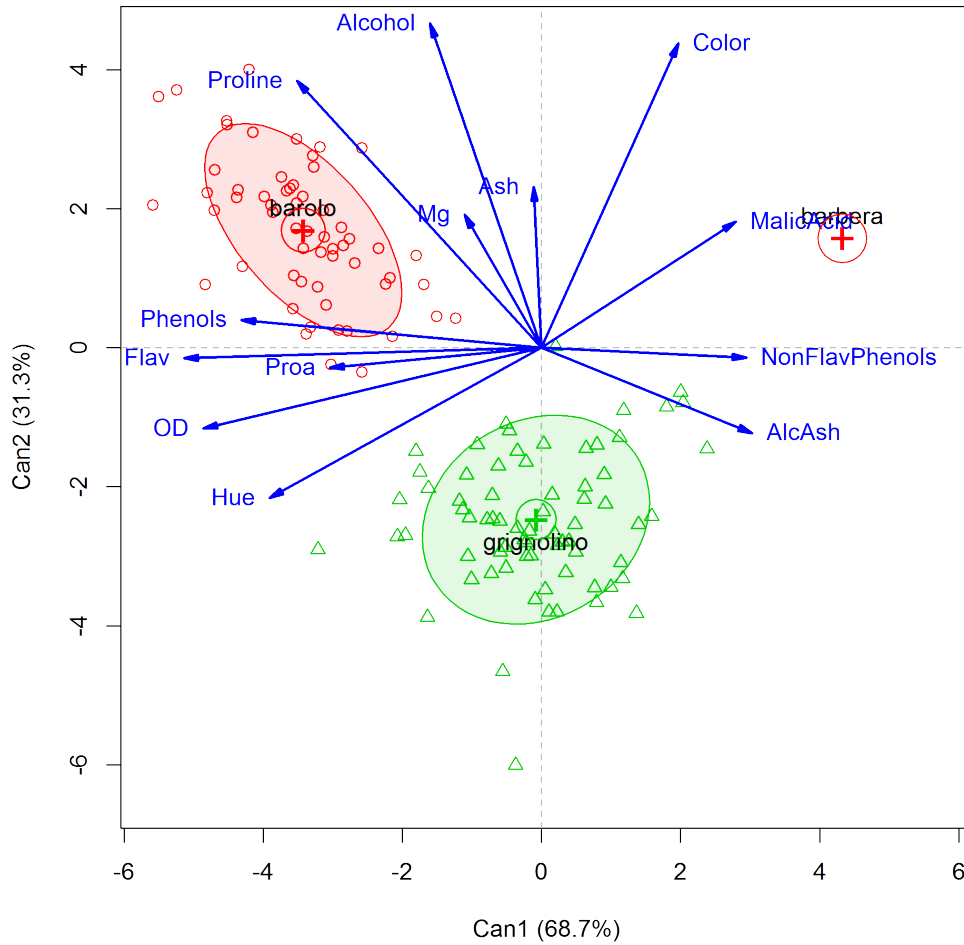


Figure 2: 2D canonical discriminant plot for the Wine data. Variable vectors show the structure coefficients (correlations) of the response variables with the canonical dimensions.

us to conclude that the covariance matrices differ, as we saw numerically in Box's M test and in Figure 9 in the paper. But the question is, *where* and *how* do they differ in relation to the response variables?

For multivariate tests of means, Hypothesis-Error (HE) plots and related canonical HE plots provide an ideal method to visualize the patterns of group mean differences in multivariate space. It would take us too far afield to describe these methods here; we give a brief summary and examples using the Wine data in the Supplemental Appendix. The essential ideas are that: (a) variation in means is captured by a matrix  $\mathbf{H}$  of sum of squares and products of the fitted values relative to an analogous matrix  $\mathbf{E}$  for error; (b) these can be visualized as suitably scaled data ellipsoids for  $\mathbf{H}$  and  $\mathbf{E}$  in variable space or in the canonical space of the linear combinations accounting for greatest differences in group means.

The fact that this extension of the Levene tests can be treated as a MANOVA of the deviations

in  $\mathbf{Z}$  suggests that HE plots can also be used to visualize the differences in variances and covariances across groups, as we illustrate in the remainder of this section. Note that in this formulation,

- The overall multivariate test by the General Linear Test of  $H_0 : \mathbf{LB} = \mathbf{0}$  gives rise to an  $\mathbf{H}$  matrix against the hypothesis, here of equal absolute deviations from the centering values, and an  $\mathbf{E}$  matrix, expressing within group sum of squares and cross products of these deviations.
- The diagonal elements of  $\mathbf{H}$  are exactly the same as would be used in univariate Levene-type tests. The off-diagonal entries reflect the covariances of the absolute deviations.
- The corresponding  $\mathbf{E}$  matrix expresses the within-group variations and their covariances in the absolute deviations.
- The univariate versions of this test simply use the diagonal elements of the  $\mathbf{H}$  and  $\mathbf{E}$  matrices, i.e.,  $F_i = \frac{h_{ii}/df_h}{e_{ii}/df_e}$ , with  $df_h = g - 1$  and  $df_e = N - g$ .
- HE plots in variable space show how the means of group variabilities differ in relation to their within-group variances and covariances.
- Comparable plots in canonical space show how means of (absolute) deviations differ in a space that accounts for the maximal differences among groups.

A similar test was proposed by O'Brien (1992), but there is a subtle difference between our proposal and his. O'Brien applied MANOVA to the matrix  $\mathbf{W}_{N \times p^2}$  whose  $(i, j)$ -th row is the vectorized version of the  $p \times p$  outer product,  $[(\mathbf{y}_{ij} - \tilde{\mathbf{y}}_i)(\mathbf{y}_{ij} - \tilde{\mathbf{y}}_i)^\top]$ . This  $\mathbf{W}$  directly reflects the contributions of each  $\mathbf{y}_{ij}$  to the variances and covariances of each group, but it is not amenable to visual display. In our proposal, the covariances are reflected in the cross-products of the the absolute deviations from the central values calculated in the  $\mathbf{H}$  and  $\mathbf{E}$  matrices and enjoy the same visualizations for scatter as for means.

## 2.1 HE plots for multivariate Levene-type tests

To return to the Wine dataset, a pairwise HE plot of all 13 responses in variable space is too complex to show here with any resolution. Instead, Figure 3 shows these for four selected variables. The means for the groups in the  $\mathbf{H}$  ellipses reflect the variances of the two responses across the wine cultivars. A positive orientation of the ellipse in a given panel means that the groups with larger variance on one variable also have larger variance on the other, as is true for the variables Flav and OD. The overall nature of these relations for all variables is more easily seen through a canonical HE plot, which is considered next.

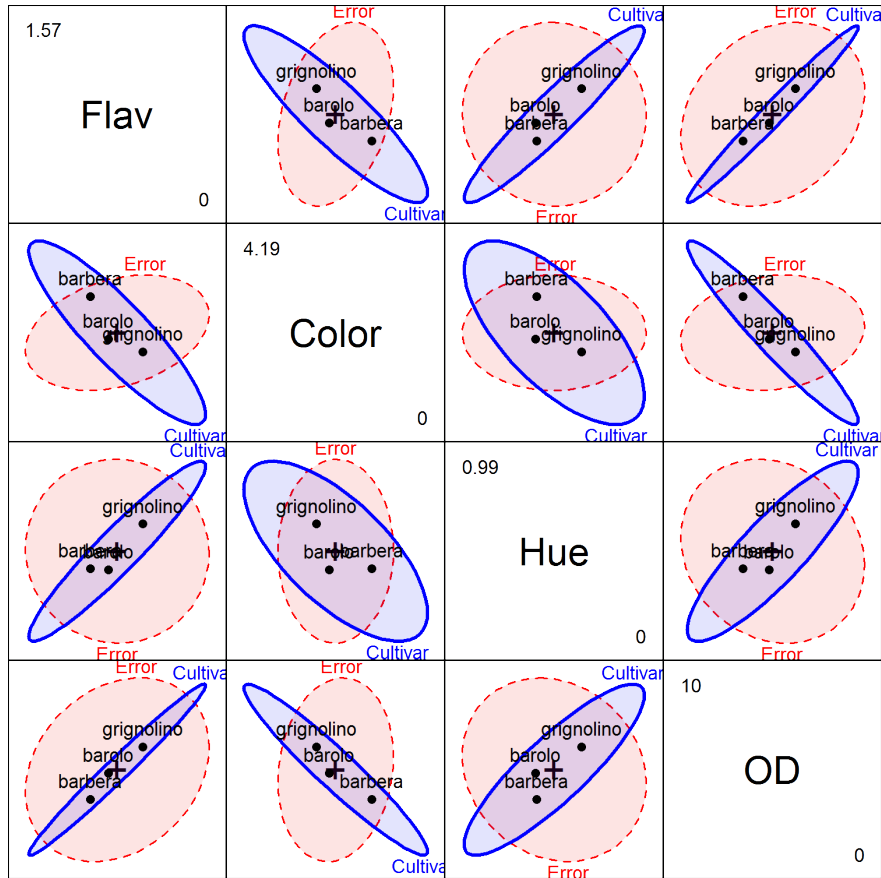


Figure 3: HE plot for selected variables in the multivariate extension of Levene's test showing significant differences in group scatter.

## 2.2 Canonical views

A canonical discriminant analysis of the MLM (`winedev.mod`) based on median-centered deviations was carried out (`candisc(winedev.mod)`). Because there are  $g = 3$  groups, only two canonical dimensions are needed to account for *all* group differences in the means of the absolute deviations; this provides a compact summary of heterogeneity in variances and covariances among the wine cultivars.

The two squared canonical correlations are  $\rho_1^2 = 0.528$  and  $\rho_2^2 = 0.265$ , accounting for respectively 75.6% and 24.4% of the between-group differences in scatter. The standard step-down likelihood ratio tests of these gave  $F(26, 326) = 8.74, p < 2.2 \times 10^{-16}$  and  $F(12, 164) = 4.92, p < 6.37 \times 10^{-7}$  respectively, both highly significant.

The 2D canonical plot is shown in Figure 4. The horizontal dimension (Can1) shows that Barbera wines differ most in variability from the other two cultivars. The variable **Color** stands out in this plot as the only variable on which Barbera wines have greater variance than the rest; this can also be seen in the boxplot for this variable (Figure 7 in the



paper). The second dimension mainly distinguishes Barolo wines from the other two cultivars. The variables Proline, Malic Acid and Non-flavor phenols contribute most strongly to this dimension.

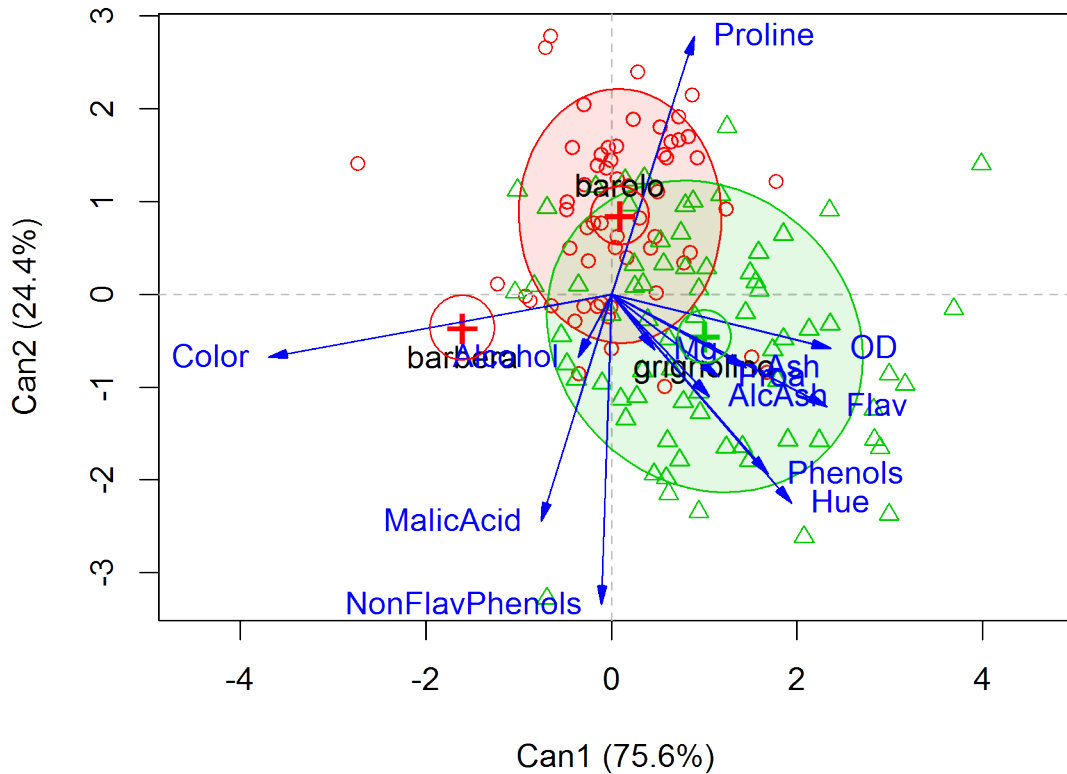


Figure 4: 2D canonical discriminant plot for differences in covariance matrices in the Wine data. The canonical scores (points) for the samples are summarized by their data ellipses. Variable vectors shown in this plot show the correlations of the responses with the two dimensions.

We admit that HE plots themselves are somewhat novel. This extension to tests for differences in variances takes a bit more thought to get used to. However, for this relatively complex problem, the canonical plot in Figure 4 provides a relatively simple description.

### 3 Distance based approaches

If we think of a covariance matrix as representing multivariate *dispersion* of points around a multivariate *centroid*, then a reasonable alternative to methods based on the covariance

matrices themselves is an approach based on distances. A set of multivariate samples,  $\mathbf{y}_{ij}, i = 1, \dots, g; j = 1, \dots, n_i$  can be said to be equally dispersed if an “average” distance,  $\Delta(\mathbf{y}_{ij}, \tilde{\mathbf{y}}_i)$  around a group “centroid”  $\tilde{\mathbf{y}}_i$  does not differ substantially across the groups.

Equality of multivariate dispersion is not exactly the same as equality of covariance matrices (it doesn’t account for orientation), but it has the advantage of transforming the  $p$ -dimensional problem into one based on an ANOVA of distances, in a way similar to what Levene-type tests do for variances.

This idea admits a variety of different forms:

- **distances** can be defined as standard Euclidean distances  $D(\mathbf{y}) = [(\mathbf{y} - \bar{\mathbf{y}})^\top (\mathbf{y} - \bar{\mathbf{y}})]^{1/2}$  in an orthogonal  $\mathbb{R}^p$  variable space, Mahalanobis distances,  $D_M(\mathbf{y}) = [(\mathbf{y} - \bar{\mathbf{y}})^\top \mathbf{S}^{-1} (\mathbf{y} - \bar{\mathbf{y}})]^{1/2}$ , which take the variances and covariances in  $\mathbf{S}$  into account, or a wide variety of other distance measures (e.g., Manhattan,  $L_1$  or taxi-cab distance, based on the sum of absolute differences).
- **centroids** can be defined as multivariate means, medians, trimmed means, “spatial medians”, and so forth.
- **average** can be defined as the arithmetic average, a root-mean-square or robust mean.
- **tests** for differences in dispersion can be based on a standard ANOVA, or a non-parametric permutation test.

Several versions of this distance based approach to multivariate dispersion were suggested by Anderson (2006; extending Van Valen, 1978). Her methods (designed for biological/ecological applications) take this idea a bit further, in that it starts with the  $N \times N$  pairwise distance measures among all observations  $\mathbf{y}_{ij}$ , carries out a metric multidimensional scaling (MDS, equivalent to a PCA of the doubly-centered distance matrix), and then applies the distance to centroid calculations to the principal coordinates. This is largely to accommodate a variety of non-Euclidean dissimilarity measures of interest in ecological studies.

These methods have been implemented in the `betadisper()` function in the `vegan` package (Oksanen et al., 2016). The  $F$ -test provided by the `anova()` method is analogous to a Levene-type test based on average distance from the group centroids. The package also provides a permutation test, `permutest()` that assesses the  $p$ -value of the  $F$ -statistic relative to the permutation distribution under random assignment to groups. The examples below are merely illustrative of this approach and the graphical methods they provide.

### 3.1 Example: Iris data

It is easiest to illustrate this approach using the Iris data. Here, we calculate Euclidean distances among all observations and use `betadisper()` to carry out analysis of these

distances by Species.

```
library(vegan)
dst <- dist(iris[,1:4])
iris.bd <- betadisper(dst, iris$Species)
```

Both the `anova()` and `permutest()` methods indicate that there are significant differences in the dispersion by Species.

```
anova(iris.bd)
```

Analysis of Variance Table

Response: Distances

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Groups	2	2.9092	1.45458	10.748	4.4e-05 ***
Residuals	147	19.8941	0.13533		

---

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

It is easy to see this in the principal coordinates plot of the first two MDS dimensions. The plot method (Figure 5) highlights the distances of each observation to the group centroid and also draws a 68% data ellipse. The configuration of the group means is similar to that of the PCA (Fig. 3 of the paper, left panel), however the orientation of the within-group data ellipses are in the opposite direction here. (The orientation of MDS axes is arbitrary.)

```
labs <- paste("Dimension", 1:4, "(",
             round(100*iris.bd$eig / sum(iris.bd$eig), 2), "%)")
plot(iris.bd, cex=1, pch=15:17,
     main="Iris data: MDS coordinates", cex.lab=1.25,
     xlab=labs[1], ylab=labs[2],
     hull=FALSE, ellipse=TRUE, conf=0.68, lwd=2)
```

The boxplot method (Figure 6) applied to the distances confirms what we can see from the ANOVA tests and the plot in principal coordinates space: There are substantial differences in dispersion among the species, particularly for *Setosa* versus the other two species, which do not differ from each other.

```
boxplot(iris.bd, xlab="Species", notch=TRUE, col=c("gray", "red", "green"))
```

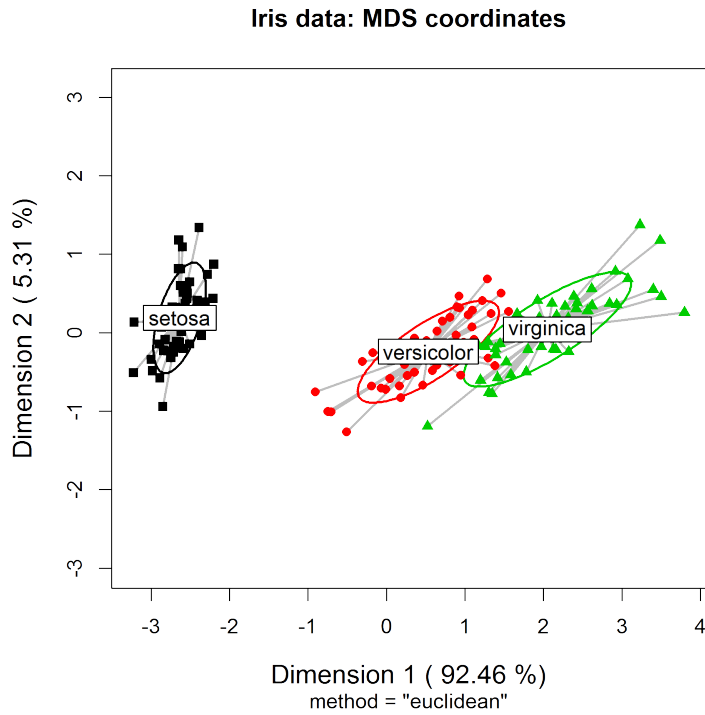


Figure 5: Principal coordinates plot of the MDS solution for the Iris data. Ellipses are 68% data ellipses for the coordinate scores.

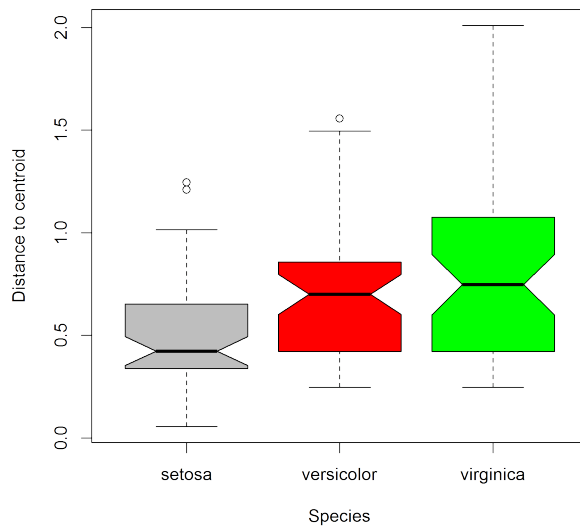


Figure 6: Boxplot of distances to the group centroid for the Iris data. Notches indicate 95% confidence intervals for group medians.

## 3.2 Example: Skulls data

For the Skulls data, we saw earlier that there were no differences among covariance matrices in any plots or analyses. Here, we calculate Euclidean distances among all observations and use `betadisper()` to carry out analysis of these distances by Epoch.

```
dst <- dist(Skulls[, -1])
skulls.bd <- betadisper(dst, Skulls$epoch)
```

Both the `anova()` and `permutest()` method indicate that there are non-significant differences in the dispersion by Epoch.

```
anova(skulls.bd)
```

Analysis of Variance Table

Response: Distances

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Groups	4	35.43	8.8565	0.8211	0.5137
Residuals	145	1563.96	10.7859		

Then, a simple display of the differences among groups in their dispersion is a boxplot of the distances to the centroid (Figure 7). The pattern of the medians seen here is similar to that shown in the plot for Box's M test in Fig. 8 (right).

```
boxplot(skulls.bd, notch=TRUE, col="lightblue", xlab="Epoch")
```

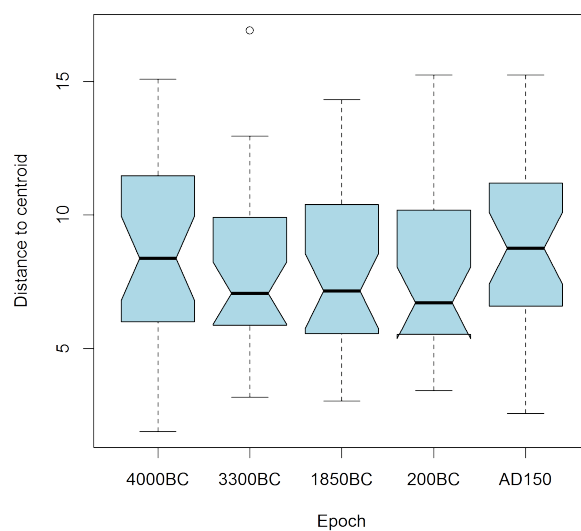


Figure 7: Boxplot of distances to the group centroid for the Skulls data

## References

- Anderson, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*, *62*(1), 245–253. <http://doi.org/10.1111/j.1541-0420.2005.00440.x>
- Fox, J. (2003). Effect displays in R for generalized linear models. *Journal of Statistical Software*, *8*(15), 1–27.
- Fox, J., Friendly, M., & Monette, G. (2009). Visualizing hypothesis tests in multivariate linear models: The *heplots* package for R. *Computational Statistics*, *24*(2), 233–246. Retrieved from <http://dx.doi.org/10.1007/s00180-008-0120-1>
- Friendly, M. (2007). HE plots for multivariate general linear models. *Journal of Computational and Graphical Statistics*, *16*(2), 421–444. <http://doi.org/10.1198/106186007X208407>
- Friendly, M., & Sigal, M. (2014). Recent advances in visualizing multivariate linear models. *Revista Colombiana de Estadística*, *37*(2), 261–283. <http://doi.org/http://dx.doi.org/10.15446/rce.v37n2spe.47934>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., . . . Wagner, H. (2016). *Vegan: Community ecology package*. Retrieved from <https://CRAN.R-project.org/package=vegan>
- O'Brien, P. C. (1992). Robust procedures for testing equality of covariance matrices. *Biometrics*, *48*(3), 819–827. Retrieved from <http://www.jstor.org/stable/2532347>
- Van Valen, L. (1978). The statistics of variation. *Evolutionary Theory*, *4*, 33–43.