Hypothesis Tests for Multivariate Linear Models Using the car Package

by John Fox, Michael Friendly, and Sanford Weisberg

Abstract The *multivariate linear model* is

$$\underset{(n\times m)}{\mathbf{Y}} = \underset{(n\times p)}{\mathbf{X}}\ \underset{(p\times m)}{\mathbf{B}} + \underset{(n\times m)}{\mathbf{E}}$$

The multivariate linear model can be fit with the `lm` function in R, where the left-hand side of the model comprises a matrix of response variables, and the right-hand side is specified exactly as for a univariate linear model (i.e., with a single response variable). This paper explains how to use the `Anova` and `linearHypothesis` functions in the **car** package to perform convenient hypothesis tests for parameters in multivariate linear models, including models for repeated-measures data.

Basic ideas

The *multivariate linear model* accommodates two or more *response* variables. The theory of multivariate linear models is developed very briefly in this section, which is based on Fox (2008, Sec. 9.5). There are many texts that treat multivariate linear models and multivariate analysis of variance (*MANOVA*) more extensively: The theory is presented in Rao (1973); more generally accessible treatments include Hand and Taylor (1987) and Morrison (2005). A good brief introduction to the MANOVA approach to repeated-measures may be found in O'Brien and Kaiser (1985), from which we draw an example below. Winer (1971, Chap. 7) presents the traditional univariate approach to repeated-measures ANOVA.

The multivariate general linear model is

$$\underset{(n\times m)}{\mathbf{Y}} = \underset{(n\times p)}{\mathbf{X}}\ \underset{(p\times m)}{\mathbf{B}} + \underset{(n\times m)}{\mathbf{E}}$$

where $\mathbf{Y}$ is a matrix of n observations on m response variables; $\mathbf{X}$ is a model matrix with columns for p regressors, typically including an initial column of 1s for the regression constant; $\mathbf{B}$ is a matrix of regression coefficients, one column for each response variable; and $\mathbf{E}$ is a matrix of errors. The contents of the model matrix are exactly as in the univariate linear model, and may contain, therefore, dummy regressors representing factors, polynomial or regression-spline terms, interaction regressors, and so on. For brevity, we assume that $\mathbf{X}$ is of full column-rank p; allowing for less than full rank cases would only introduce additional notation but not fundamentally change any of the results presented here.

The assumptions of the multivariate linear model concern the behavior of the errors: Let $\boldsymbol{\varepsilon}_i'$ represent the ith row of $\mathbf{E}$. Then $\boldsymbol{\varepsilon}_i' \sim \mathbf{N}_m(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a nonsingular error-covariance matrix, constant across observations; $\boldsymbol{\varepsilon}_i'$ and $\boldsymbol{\varepsilon}_j'$ are independent for $i \neq j$; and $\mathbf{X}$ is fixed or independent of $\mathbf{E}$. We can write more compactly that $\text{vec}(\mathbf{E}) \sim \mathbf{N}_{nm}(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma})$. Here, $\text{vec}(\mathbf{E})$ ravels the error matrix row-wise into a vector, $\mathbf{I}_n$ is the order-n identity matrix, and $\otimes$ is the Kronecker-product operator.

The maximum-likelihood estimator of $\mathbf{B}$ in the multivariate linear model is equivalent to equation-by-equation least squares for the individual responses:

$$\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

Procedures for statistical inference in the multivariate linear model, however, take account of correlations among the responses.

Paralleling the decomposition of the total sum of squares into regression and residual sums of squares in the univariate linear model, there is in the multivariate linear model a decomposition of the total *sum-of-squares-and-cross-products* (*SSP*) matrix into regression and residual SSP matrices. We have

$$\underset{(m\times m)}{\mathbf{SSP}_T} = \mathbf{Y}'\mathbf{Y} - n\overline{\mathbf{y}}\,\overline{\mathbf{y}}'$$

$$= \widehat{\mathbf{E}}'\widehat{\mathbf{E}} + \left(\widehat{\mathbf{Y}}'\widehat{\mathbf{Y}} - n\overline{\mathbf{y}}\,\overline{\mathbf{y}}'\right)$$

$$= \mathbf{SSP}_R + \mathbf{SSP}_{\text{Reg}}$$

where $\overline{\mathbf{y}}$ is the $(m \times 1)$ vector of means for the response variables; $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\mathbf{B}}$ is the matrix of fitted values; and $\widehat{\mathbf{E}} = \mathbf{Y} - \widehat{\mathbf{Y}}$ is the matrix of residuals.

Many hypothesis tests of interest can be formulated by taking differences in $\mathbf{SSP}_{\text{Reg}}$ (or, equivalently, $\mathbf{SSP}_R$) for nested models, although the `Anova` function in the **car** package (Fox and Weisberg, 2011), described below, calculates SSP matrices for common hypotheses more cleverly, without refitting the model. Let $\mathbf{SSP}_H$ represent the incremental SSP matrix for a hypothesis—that is, the difference between $\mathbf{SSP}_{\text{Reg}}$ for the model unrestricted by the hypothesis and $\mathbf{SSP}_{\text{Reg}}$ for the model on which the hypothesis is imposed. Multivariate tests for the hypothesis are based on the m eigenvalues λ_j of $\mathbf{SSP}_H\mathbf{SSP}_R^{-1}$ (the hypothesis SSP matrix "divided by" the residual SSP matrix), that is, the values of λ for which

$$\det(\mathbf{SSP}_H\mathbf{SSP}_R^{-1} - \lambda\mathbf{I}_m) = 0$$

The several commonly used multivariate test statistics are functions of these eigenvalues:

$$
\begin{aligned}
&\text{Pillai-Bartlett Trace, } T_{PB} = \sum_{j=1}^{m}\frac{\lambda_j}{1-\lambda_j} \\
&\text{Hotelling-Lawley Trace, } T_{HL} = \sum_{j=1}^{m}\lambda_j \\
&\text{Wilks's Lambda, } \Lambda = \prod_{j=1}^{m}\frac{1}{1+\lambda_j} \\
&\text{Roy's Maximum Root, } \lambda_1
\end{aligned}
\tag{1}
$$

By convention, the eigenvalues of $\mathbf{SSP}_H\mathbf{SSP}_R^{-1}$ are arranged in descending order, and so λ_1 is the largest eigenvalue. The **car** package uses F approximations to the null distributions of these test statistics (see, e.g., Rao, 1973, p. 556, for Wilks's Lambda).

The tests apply generally to all linear hypotheses. Suppose that we want to test the linear hypothesis

$$H_0: \underset{(q\times p)}{\mathbf{L}}\underset{(p\times m)}{\mathbf{B}} = \underset{(q\times m)}{\mathbf{C}} \tag{2}$$

where $\mathbf{L}$ is a hypothesis matrix of full row-rank $q \leq p$, and the right-hand-side matrix $\mathbf{C}$ consists of constants, usually 0s. Then the SSP matrix for the hypothesis is

$$\mathbf{SSP}_H = \left(\widehat{\mathbf{B}}'\mathbf{L}' - \mathbf{C}'\right)\left[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}'\right]^{-1}\left(\mathbf{L}\widehat{\mathbf{B}} - \mathbf{C}\right)$$

The various test statistics are based on the $k = \min(q, m)$ nonzero eigenvalues of $\mathbf{SSP}_H\mathbf{SSP}_R^{-1}$.

When a multivariate response arises because a variable is measured on different occasions, or under different circumstances (but for the same individuals), it is also of interest to formulate hypotheses concerning comparisons among the responses. This situation, called a *repeated-measures design*, can be handled by linearly transforming the responses using a suitable "within-subjects" model matrix, for example extending the linear hypothesis in Equation 2 to

$$H_0: \underset{(q\times p)}{\mathbf{L}}\underset{(p\times m)}{\mathbf{B}}\underset{(m\times v)}{\mathbf{P}} = \underset{(q\times v)}{\mathbf{C}} \tag{3}$$

Here, the *response-transformation matrix* $\mathbf{P}$, assumed to be of full column-rank, provides contrasts in the responses (see, e.g., Hand and Taylor, 1987, or O'Brien and Kaiser, 1985). The SSP matrix for the hypothesis is

$$\underset{(q\times q)}{\mathbf{SSP}_H} = \left(\mathbf{P}'\widehat{\mathbf{B}}'\mathbf{L}' - \mathbf{C}'\right)\left[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}'\right]^{-1}\left(\mathbf{L}\widehat{\mathbf{B}}\mathbf{P} - \mathbf{C}\right)$$

and test statistics are based on the $k = \min(q, v)$ nonzero eigenvalues of $\mathbf{SSP}_H(\mathbf{P}'\mathbf{SSP}_R\mathbf{P})^{-1}$.

Fitting and testing multivariate linear models

Multivariate linear models are fit in R with the `lm` function. The procedure is the essence of simplicity: The left-hand side of the model formula is a matrix of responses, with each column representing a response variable and each row an observation; the right-hand side of the model formula and all other arguments to `lm` are precisely the same as for a univariate linear model (as described, e.g., in Fox and Weisberg, 2011, Chap. 4). Typically, the response matrix is composed from individual response variables via the `cbind` function. The `anova` function in the standard R distribution is capable of handling multivariate linear models (see Dalgaard, 2007), but the `Anova` and `linearHypothesis` functions in the **car** package may also be employed. We briefly demonstrate the use of these functions in this section.

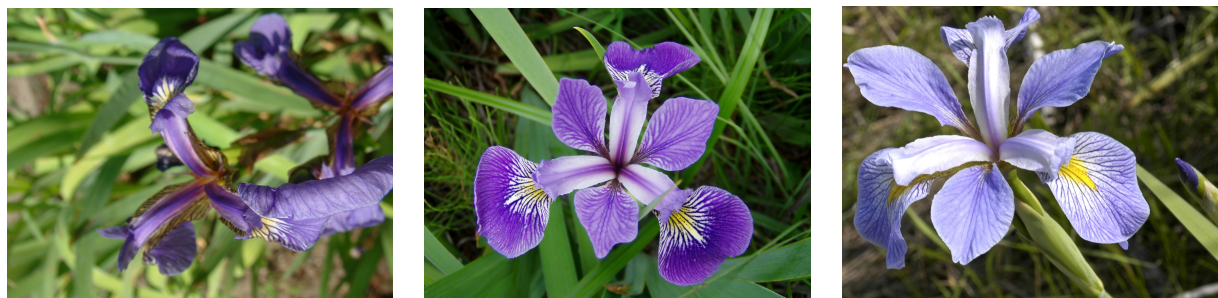Figure 1: Three species of irises in the Anderson/Fisher data set: setosa (left), versicolor (center), and virginica (right). *Source*: The photographs are respectively by Radomil Binek, Danielle Langlois, and Frank Mayfield, and are distributed under the Creative Commons Attribution-Share Alike 3.0 Unported license (first and second images) or 2.0 Creative Commons Attribution-Share Alike Generic license (third image); they were obtained from the Wikimedia Commons.

`Anova` and `linearHypothesis` are generic functions with methods for many common classes of statistical models with linear predictors. In addition to multivariate linear models, these classes include linear models fit by `lm` or `aov`; generalized linear models fit by `glm`; mixed-effects models fit by `lmer` or `glmer` in the **lme4** package (Bates et al., 2012) or `lme` in the **nlme** package (Pinheiro et al., 2012); survival regression models fit by `coxph` or `survreg` in the **survival** package (Therneau, 2012); multinomial-response models fit by `multinom` in the **nnet** package (Venables and Ripley, 2002); ordinal regression models fit by `polr` in the **MASS** package (Venables and Ripley, 2002); and generalized linear models fit to complex-survey data via `svyglm` in the **survey** package (Lumley, 2004). There is also a generic method that will work with many models for which there are `coef` and `vcov` methods. The `Anova` and `linearHypothesis` methods for `"mlm"` objects are special, however, in that they handle multiple response variables and make provision for designs on repeated measures, discussed in the next section.

To illustrate multivariate linear models, we will use data collected by Anderson (1935) on three species of irises in the Gaspé Peninsula of Québec, Canada. The data are of historical interest in statistics, because they were employed by R. A. Fisher (1936) to introduce the method of discriminant analysis. The data frame `iris` is part of the standard R distribution, and we load the **car** package now for the `some` function, which randomly samples the rows of a data set. We rename the variables in the `iris` data to make listings more compact:

```
> names(iris)

[1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"

> names(iris) <- c("SL", "SW", "PL", "PW", "SPP")
> library(car)
> some(iris, 3) # 3 random rows

     SL  SW  PL  PW        SPP
44   5.0 3.5 1.6 0.6    setosa
61   5.0 2.0 3.5 1.0 versicolor
118  7.7 3.8 6.7 2.2  virginica
```

The first four variables in the data set represent measurements (in cm) of parts of the flowers, while the final variable specifies the species of iris. (Sepals are the green leaves that comprise the calyx of the plant, which encloses the flower.) Photographs of examples of the three species of irises—setosa, versicolor, and virginica— appear in Figure 1. Figure 2 is a scatterplot matrix of the four measurements classified by species, showing within-species 50 and 95% concentration ellipses (see Fox and Weisberg, 2011, Sec. 4.3.8); Figure 3 shows boxplots for each of the responses by species. These graphs are produced by the `scatterplotMatrix` and `Boxplot` functions in the **car** package (see Fox and Weisberg, 2011, Sec. 3.2.2 and 3.3.2). As the photographs suggest, the scatterplot matrix and boxplots for the measurements reveal that versicolor and virginica are more similar to each other than either is to setosa. Further, the ellipses in the scatterplot matrix suggest that the assumption of constant within-group covariance matrices is problematic: While the shapes and sizes of the concentration ellipses for versicolor and virginica are reasonably similar, the shapes and sizes of the ellipses for setosa are different from the other two.

We proceed nevertheless to fit a multivariate one-way ANOVA model to the iris data:

```
> mod.iris <- lm(cbind(SL, SW, PL, PW) ~ SPP, data=iris)
> class(mod.iris)
```
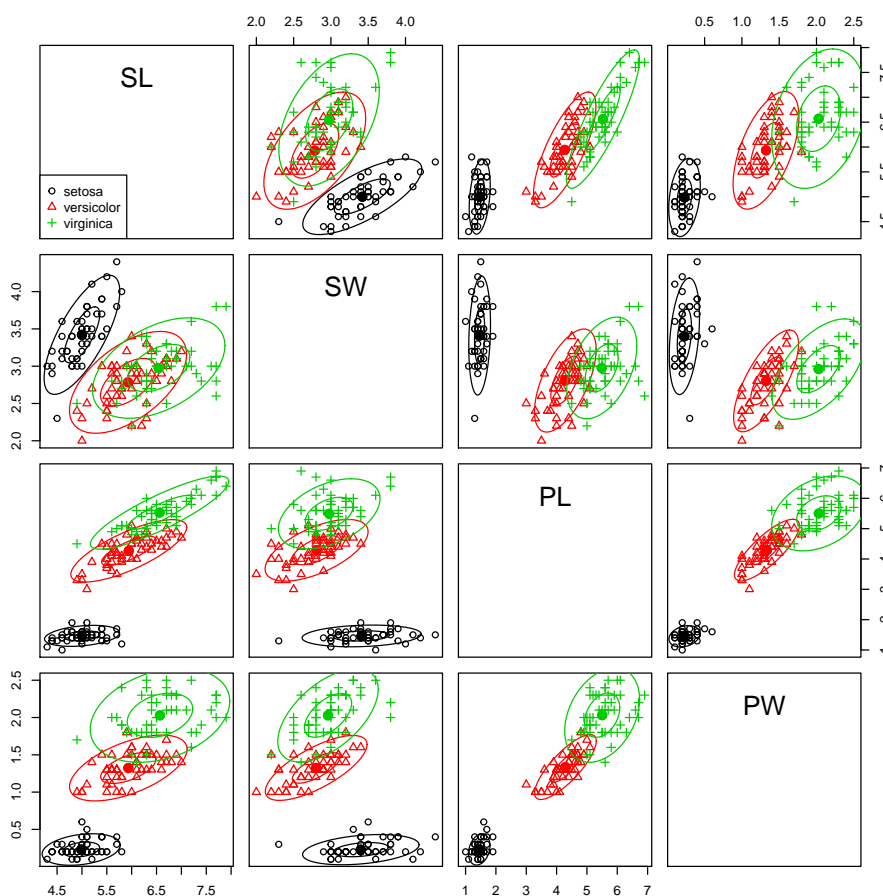
Figure 2: Scatterplot matrix for the Anderson/Fisher iris data, showing within-species 50 and 95% concentration ellipses.

```
[1] "mlm" "lm"
```

The `lm` function returns an S3 object of class `"mlm"` inheriting from class `"lm"`. The printed representation of the object (not shown) simply displays the estimated regression coefficients for each response, and the model summary (also not shown) is the same as we would obtain by performing separate least-squares regressions for the four responses.

We use the `Anova` function in the **car** package to test the null hypothesis that the four response means are identical across the three species of irises:

```
> manova.iris <- Anova(mod.iris)
> manova.iris

Type II MANOVA Tests: Pillai test statistic
    Df test stat approx F num Df den Df Pr(>F)
SPP  2     1.19     53.5      8    290 <2e-16

> class(manova.iris)

[1] "Anova.mlm"

> summary(manova.iris)

Type II MANOVA Tests:

Sum of squares and products for error:
        SL     SW     PL     PW
```
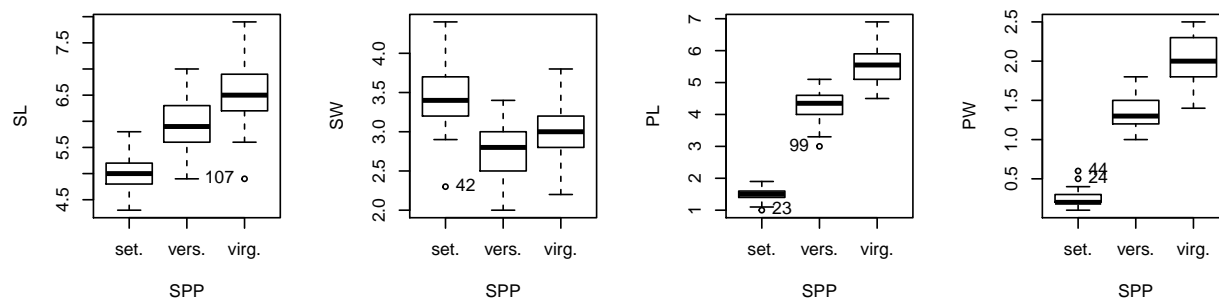
Figure 3: Boxplots for the response variables in the iris data set classified by species.

```
SL 38.956 13.630 24.625 5.645
SW 13.630 16.962  8.121 4.808
PL 24.625  8.121 27.223 6.272
PW  5.645  4.808  6.272 6.157


-----------------------------------------


Term: SPP


Sum of squares and products for the hypothesis:
       SL     SW      PL      PW
SL  63.21 -19.95 165.25   71.28
SW -19.95  11.34 -57.24  -22.93
PL 165.25 -57.24 437.10  186.77
PW  71.28 -22.93 186.77   80.41


Multivariate Tests: SPP
                Df test stat approx F num Df den Df Pr(>F)
Pillai           2     1.19     53.5      8    290 <2e-16
Wilks            2     0.02    199.1      8    288 <2e-16
Hotelling-Lawley 2    32.48    580.5      8    286 <2e-16
Roy              2    32.19   1167.0      4    145 <2e-16
```

The `Anova` function returns an object of class `"Anova.mlm"` which, when printed, produces a MANOVA table, by default reporting Pillai's test statistic;[1] summarizing the object produces a more complete report. Because there is only one term (beyond the regression constant) on the right-hand side of the model, in this example the "type-II" test produced by default by `Anova` is the same as the sequential ("type-I") test produced by the standard R `anova` function (output not show):

```
> anova(mod.iris)
```

The null hypothesis is soundly rejected.

The object returned by `Anova` may also be used in further computations, for example, for displays such as HE plots (Friendly, 2007; Fox et al., 2009; Friendly, 2010), as we illustrate below.

The `linearHypothesis` function in the **car** package may be used to test more specific hypotheses about the parameters in the multivariate linear model. For example, to test for differences between setosa and the average of versicolor and virginica, and for differences between versicolor and virginica:

```
> linearHypothesis(mod.iris, "0.5*SPPversicolor + 0.5*SPPvirginica")
```

```
. . .
Multivariate Tests:
```

---

[1]The `Manova` function in the **car** package may be used as a synonym for `Anova` applied to a multivariate linear model. The computation of the standard multivariate test statistics is performed via unexported functions from the standard R **stats** package, such as `stats:::Pillai`.

```
                 Df test stat approx F num Df den Df Pr(>F)
Pillai            1      0.967     1064      4    144 <2e-16
Wilks             1      0.033     1064      4    144 <2e-16
Hotelling-Lawley  1     29.552     1064      4    144 <2e-16
Roy               1     29.552     1064      4    144 <2e-16

> linearHypothesis(mod.iris, "SPPversicolor = SPPvirginica")


. . .
Multivariate Tests:
                 Df test stat approx F num Df den Df Pr(>F)
Pillai            1     0.7452    105.3      4    144 <2e-16
Wilks             1     0.2548    105.3      4    144 <2e-16
Hotelling-Lawley  1     2.9254    105.3      4    144 <2e-16
Roy               1     2.9254    105.3      4    144 <2e-16
```

Here and elsewhere in this paper, we use widely separated ellipses (. . .) to indicate abbreviated R output.

Setting the argument `verbose=TRUE` to `linearHypothesis` (not given here to conserve space) shows in addition the hypothesis matrix **L** and right-hand-side matrix **C** for the linear hypothesis in Equation 2 (page 2). In this case, all of the multivariate test statistics are equivalent and therefore translate into identical *F*-statistics. Both focussed null hypotheses are easily rejected, but the evidence for differences between setosa and the other two iris species is much stronger than for differences between versicolor and virginica. Testing that `"0.5*SPPversicolor + 0.5*SPPvirginica"` is **0** tests that the average of the mean vectors for these two species is equal to the mean vector for setosa, because the latter is the baseline category for the `Species` dummy regressors.

An alternative, equivalent, and in a sense more direct, approach is to fit the model with custom contrasts for the three species of irises, followed up by a test for each contrast:

```
> C <- matrix(c(1, -0.5, -0.5, 0, 1, -1), 3, 2)
> colnames(C) <- c("S:VV", "V:V")
> rownames(C) <- unique(iris$SPP)
> contrasts(iris$SPP) <- C
> contrasts(iris$SPP)

          S:VV V:V
setosa     1.0   0
versicolor -0.5   1
virginica  -0.5  -1

> mod.iris.2 <- update(mod.iris)
> coef(mod.iris.2)

                 SL      SW      PL      PW
(Intercept)  5.8433  3.0573   3.758  1.1993
SPPS:VV     -0.8373  0.3707  -2.296 -0.9533
SPPV:V      -0.3260 -0.1020  -0.646 -0.3500

> linearHypothesis(mod.iris.2, c(0, 1, 0)) # setosa vs. versicolor & virginica


. . .
Multivariate Tests:
                 Df test stat approx F num Df den Df Pr(>F)
Pillai            1      0.967     1064      4    144 <2e-16
Wilks             1      0.033     1064      4    144 <2e-16
Hotelling-Lawley  1     29.552     1064      4    144 <2e-16
Roy               1     29.552     1064      4    144 <2e-16

> linearHypothesis(mod.iris.2, c(0, 0, 1)) # versicolor vs. virginica


. . .
Multivariate Tests:
                 Df test stat approx F num Df den Df Pr(>F)
```
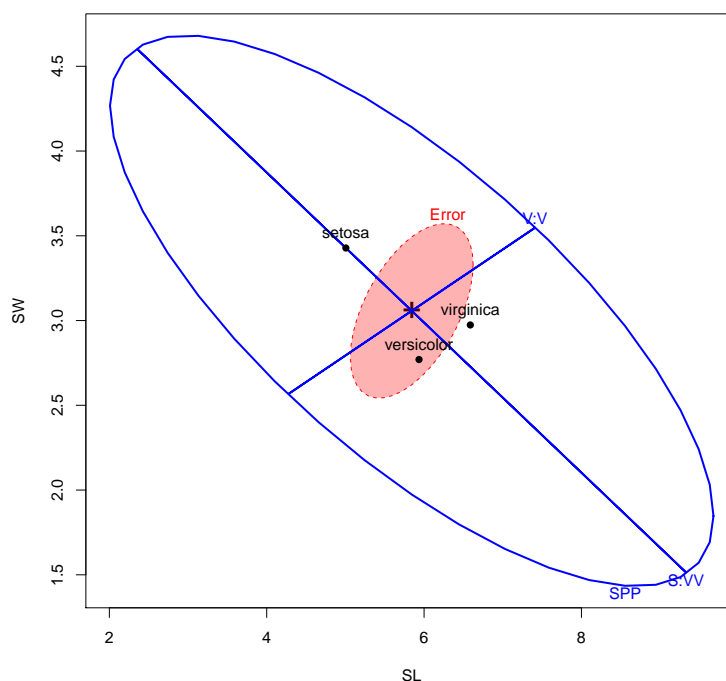
Figure 4: HE plot for the multivariate test of `Species` in the iris data, $\alpha = 0.05$, shown for the sepal length and sepal width response variables. Also shown are the multivariate `linearHypothesis` tests for two contrasts among species. The shaded red ellipse is the error ellipse, and the hypothesis ellipses (including the two lines) are blue.

Pillai	1	0.7452	105.3	4	144	<2e-16
Wilks	1	0.2548	105.3	4	144	<2e-16
Hotelling-Lawley	1	2.9254	105.3	4	144	<2e-16
Roy	1	2.9254	105.3	4	144	<2e-16

We note here briefly that the **heplots** package (Friendly, 2007; Fox et al., 2009) provides informative visualizations in 2D and 3D *hypothesis-error (HE) plots* of multivariate hypothesis tests and `"Anova.mlm"` objects based on Eqn. 2. These plots show direct visual representations of the $\mathbf{SSP}_H$ and $\mathbf{SSP}_E$ matrices as (possibly degenerate) ellipses and ellipsoids.

Using the default *significance scaling*, HE plots have the property that the $\mathbf{SSP}_H$ ellipsoid extends outside the $\mathbf{SSP}_E$ ellipsoid *if and only if* the corresponding multivariate hypothesis test is rejected by Roy's maximum root test at a given α level. See Friendly (2007) and Fox et al. (2009) for details of these methods, and Friendly (2010) for analogous plots for repeated measure designs.

To illustrate, Figure 4 shows the 2D HE plot of the two sepal variables for the overall test of `Species`, together with the tests of the contrasts among species described above. The $\mathbf{SSP}_H$ matrices for the contrasts have rank 1, so their ellipses plot as lines. All three $\mathbf{SSP}_H$ ellipses extend far outside the $\mathbf{SSP}_E$ ellipse, indicating that all tests are highly significant.

```
> library(heplots)
> hyp <- list("V:V"="SPPV:V", "S:VV"="SPPS:VV")
> heplot(mod.iris.2, hypotheses=hyp, fill=c(TRUE, FALSE), col=c("red", "blue"))
```

Finally, we can code the response-transformation matrix $\mathbf{P}$ in Equation 3 (page 2) to compute linear combinations of the responses, either via the `imatrix` argument to `Anova` (which takes a list of matrices) or the `P` argument to `linearHypothesis` (which takes a matrix). We illustrate trivially with a univariate ANOVA for the first response variable, sepal length, extracted from the multivariate linear model for all four responses:

```
> Anova(mod.iris, imatrix=list(Sepal.Length=matrix(c(1, 0, 0, 0))))

Type II Repeated Measures MANOVA Tests: Pillai test statistic
                Df test stat approx F num Df den Df Pr(>F)
```

```
Sepal.Length     1     0.992     19327     1     147 <2e-16
SPP:Sepal.Length 2     0.619       119     2     147 <2e-16
```

The univariate ANOVA for sepal length by species appears in the second line of the MANOVA table produced by `Anova`. Similarly, using `linearHypothesis`,

```
> linearHypothesis(mod.iris, c("SPPversicolor = 0", "SPPvirginica = 0"),
+   P=matrix(c(1, 0, 0, 0))) # equivalent

. . .
Multivariate Tests:
                 Df test stat approx F num Df den Df Pr(>F)
Pillai            2    0.6187    119.3      2    147 <2e-16
Wilks             2    0.3813    119.3      2    147 <2e-16
Hotelling-Lawley  2    1.6226    119.3      2    147 <2e-16
Roy               2    1.6226    119.3      2    147 <2e-16
```

In this case, the `P` matrix is a single column picking out the first response. We verify that we get the same F-test from a univariate ANOVA for `Sepal.Length`:

```
> Anova(lm(SL ~ SPP, data=iris))

Anova Table (Type II tests)

Response: SL
          Sum Sq  Df F value Pr(>F)
SPP         63.2   2     119 <2e-16
Residuals   39.0 147
```

Contrasts of the responses occur more naturally in the context of repeated-measures data, which we discuss in the following section.

Handling repeated measures

Repeated-measures data arise when multivariate responses represent the same individuals measured on a response variable (or variables) on different occasions or under different circumstances. There may be a more or less complex design on the repeated measures. The simplest case is that of a single repeated-measures or *within-subjects* factor, where the former term often is applied to data collected over time and the latter when the responses represent different experimental conditions or treatments. There may, however, be two or more within-subjects factors, as is the case, for example, when each subject is observed under different conditions on each of several occasions. The terms "repeated measures" and "within-subjects factors" are common in disciplines, such as psychology, where the units of observation are individuals, but these designs are essentially the same as so-called "split-plot" designs in agriculture, where plots of land are each divided into sub-plots, which are subjected to different experimental treatments, such as differing varieties of a crop or differing levels of fertilizer.

Repeated-measures designs can be handled in R with the standard `anova` function, as described by Dalgaard (2007), but it is considerably simpler to get common tests from the `Anova` and `linearHypothesis` functions in the **car** package, as we explain in this section. The general procedure is first to fit a multivariate linear model with all of the repeated measures as responses; then an artificial data frame is created in which each of the repeated measures is a row and in which the columns represent the repeated-measures factor or factors; finally, as we explain below, the `Anova` or `linearHypothesis` function is called, using the `idata` and `idesign` arguments (and optionally the `icontrasts` argument)—or alternatively the `imatrix` argument to `Anova` or `P` argument to `linearHypothesis`—to specify the intra-subject design.

To illustrate, we use data reported by O'Brien and Kaiser (1985), in what they (justifiably) bill as "an extensive primer" for the MANOVA approach to repeated-measures designs. Although the data are apparently not real, they are contrived cleverly to illustrate the computations for repeated-measures MANOVA, and we use the data for this reason, as well as to permit comparison of our results to those in an influential published source. The data set `OBrienKaiser` is provided by the **car** package:

```
> some(OBrienKaiser, 4)
```

```
    treatment gender pre.1 pre.2 pre.3 pre.4 pre.5 post.1 post.2 post.3 post.4
11          B      M     3     3     4     2     3      5      4      7      5
12          B      M     6     7     8     6     3      9     10     11      9
14          B      F     2     2     3     1     2      5      6      7      5
16          B      F     4     5     7     5     4      7      7      8      6
   post.5 fup.1 fup.2 fup.3 fup.4 fup.5
11      4     5     6     8     6     5
12      6     8     7    10     8     7
14      2     6     7     8     6     3
16      7     7     8    10     8     7
```

```
> contrasts(OBrienKaiser$treatment)

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

```
> contrasts(OBrienKaiser$gender)

   [,1]
F     1
M    -1
```

```
> xtabs(~ treatment + gender, data=OBrienKaiser)

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

There are two between-subjects factors in the O'Brien-Kaiser data: `gender`, with levels `F` and `M`; and `treatment`, with levels `A`, `B`, and `control`. Both of these variables have predefined contrasts, with $-1,1$ coding for `gender` and custom contrasts for `treatment`. In the latter case, the first contrast is for the `control` group vs. the average of the experimental groups, and the second contrast is for treatment `A` vs. treatment `B`. We have defined these contrasts, which are orthogonal in the row-basis of the between-subjects design, to reproduce the type-III tests that are reported in the original source.

The frequency table for `treatment` by `gender` reveals that the data are mildly unbalanced. We will imagine that the treatments `A` and `B` represent different innovative methods of teaching reading to learning-disabled students, and that the `control` treatment represents a standard method.

The 15 response variables in the data set represent two crossed within-subjects factors: *phase*, with three levels for the *pretest*, *post-test*, and *follow-up* phases of the study; and *hour*, representing five successive hours, at which measurements of reading comprehension are taken within each phase. We define the "data" for the within-subjects design as follows:

```
> phase <- factor(rep(c("pretest", "posttest", "followup"), each=5),
+   levels=c("pretest", "posttest", "followup"))
> hour <- ordered(rep(1:5, 3))
> idata <- data.frame(phase, hour)
> idata

      phase hour
1   pretest    1
2   pretest    2
3   pretest    3
. . .
14 followup    4
15 followup    5
```

Mean reading comprehension is graphed by hour, phase, treatment, and gender in Figure 5. It appears as if reading improves across phases in the two experimental treatments but not in the control group (suggesting
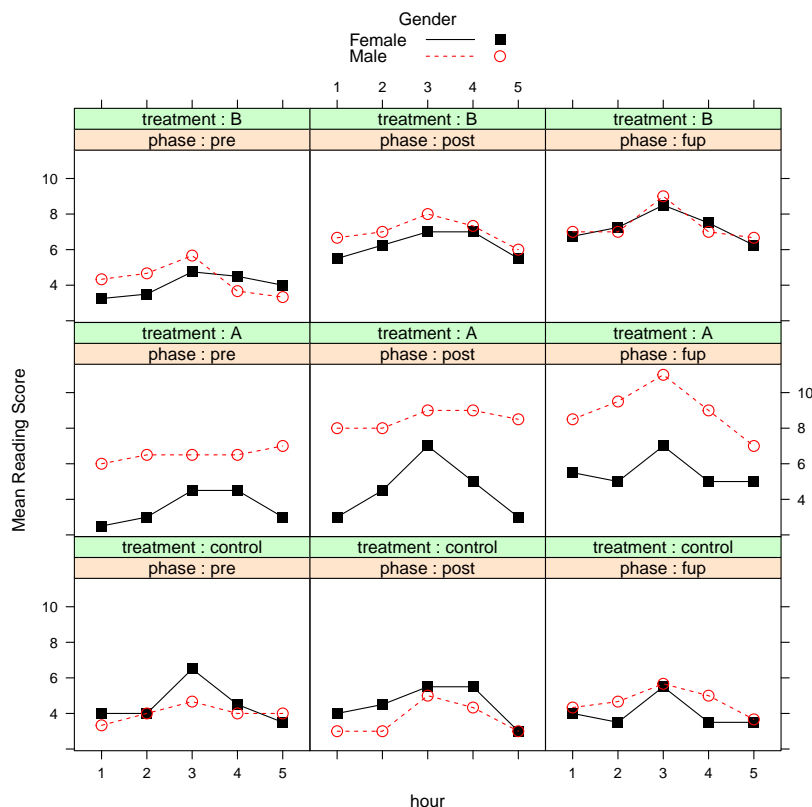
Figure 5: Mean reading score by gender, treatment, phase, and hour, for the O'Brien-Kaiser data.

a possible treatment-by-phase interaction); that there is a possibly quadratic relationship of reading to hour within each phase, with an initial rise and then decline, perhaps representing fatigue (suggesting an hour main effect); and that males and females respond similarly in the control and B treatment groups, but that males do better than females in the A treatment group (suggesting a possible gender-by-treatment interaction).

We next fit a multivariate linear model to the data, treating the repeated measures as responses, and with the between-subject factors `treatment` and `gender` (and their interaction) appearing on the right-hand side of the model formula:

```
> mod.ok <- 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)
```

We then compute the repeated-measures MANOVA using the `Anova` function in the following manner:

```
> av.ok <- Anova(mod.ok, idata=idata, idesign=~phase*hour, type=3)
> av.ok
```

```
Type III Repeated Measures MANOVA Tests: Pillai test statistic
                       Df test stat approx F num Df den Df  Pr(>F)
(Intercept)             1     0.967    296.4      1     10 9.2e-09
treatment               2     0.441      3.9      2     10 0.05471
gender                  1     0.268      3.7      1     10 0.08480
treatment:gender        2     0.364      2.9      2     10 0.10447
phase                   1     0.814     19.6      2      9 0.00052
treatment:phase         2     0.696      2.7      4     20 0.06211
gender:phase            1     0.066      0.3      2      9 0.73497
treatment:gender:phase  2     0.311      0.9      4     20 0.47215
hour                    1     0.933     24.3      4      7 0.00033
treatment:hour          2     0.316      0.4      8     16 0.91833
gender:hour             1     0.339      0.9      4      7 0.51298
```

```
treatment:gender:hour        2    0.570    0.8    8    16 0.61319
phase:hour                   1    0.560    0.5    8     3 0.82027
treatment:phase:hour         2    0.662    0.2   16     8 0.99155
gender:phase:hour            1    0.712    0.9    8     3 0.58949
treatment:gender:phase:hour  2    0.793    0.3   16     8 0.97237
```

- Following O'Brien and Kaiser (1985), we report type-III tests (partial tests violating marginality), by specifying the argument `type=3`. Although, as in univariate models, we generally prefer type-II tests (see Fox and Weisberg, 2011, Sec. 4.4.4, and Fox, 2008, Sec. 8.2), we wanted to preserve comparability with the original source. Type-III tests are computed correctly because the contrasts employed for `treatment` and `gender`, and hence their interaction, are orthogonal in the row-basis of the between-subjects design. We invite the reader to compare these results with the default type-II tests.

- When, as here, the `idata` and `idesign` arguments are specified, `Anova` automatically constructs orthogonal contrasts for different terms in the within-subjects design, using `contr.sum` for a factor such as `phase` and `contr.poly` (orthogonal polynomial contrasts) for an ordered factor such as `hour`. Alternatively, the user can assign contrasts to the columns of the intra-subject data, either directly or via the `icontrasts` argument to `Anova`. In any event, `Anova` checks that the within-subjects contrast coding for different terms is orthogonal and reports an error when it is not.

- By default, Pillai's test statistic is displayed; we invite the reader to examine the other three multivariate test statistics. Much more detail of the tests is provided by `summary(av.ok)` (not shown).

- The results show that the anticipated `hour` effect is statistically significant, but the `treatment` × `phase` and `treatment` × `gender` interactions are not quite significant. There is, however, a statistically significant `phase` main effect. Of course, we should not over-interpret these results, partly because the data set is small and partly because it is contrived.

Univariate ANOVA for repeated measures

A traditional univariate approach to repeated-measures (or split-plot) designs (see, e.g., Winer, 1971, Chap. 7) computes an analysis of variance employing a "mixed-effects" models in which subjects generate random effects. This approach makes stronger assumptions about the structure of the data than the MANOVA approach described above, in particular stipulating that the covariance matrices for the repeated measures transformed by the within-subjects design (within combinations of between-subjects factors) are *spherical*—that is, the transformed repeated measures for each within-subjects test are uncorrelated and have the same variance, and this variance is constant across cells of the between-subjects design. A sufficient (but not necessary) condition for sphericity of the errors is that the covariance matrix Σ of the repeated measures is *compound-symmetric*, with equal diagonal entries (representing constant variance for the repeated measures) and equal off-diagonal elements (implying, together with constant variance, that the repeated measures have a constant correlation).

By default, when an intra-subject design is specified, summarizing the object produced by `Anova` reports both MANOVA and univariate tests. Along with the traditional univariate tests, the summary reports tests for sphericity (Mauchly, 1940) and two corrections for non-sphericity of the univariate test statistics for within-subjects terms: the Greenhouse-Geisser correction (Greenhouse and Geisser, 1959) and the Huynh-Feldt correction (Huynh and Feldt, 1976). We illustrate for the O'Brien-Kaiser data, suppressing the output for brevity; we invite the reader to reproduce this analysis:

```
> summary(av.ok, multivariate=FALSE)
```

There are statistically significant departures from sphericity for *F*-tests involving `hour`; the results for the univariate ANOVA are not terribly different from those of the MANOVA reported above, except that now the `treatment` × `phase` interaction is statistically significant.

Using `linearHypothesis` with repeated-measures designs

As for simpler multivariate linear models (discussed previously in this paper), the `linearHypothesis` function can be used to test more focused hypotheses about the parameters of repeated-measures models, including for within-subjects terms.

As a preliminary example, to reproduce the test for the main effect of `hour`, we can use the `idata`, `idesign`, and `iterms` arguments in a call to `linearHypothesis`:

```
> linearHypothesis(mod.ok, "(Intercept) = 0", idata=idata,
+   idesign=~phase*hour, iterms="hour")

 Response transformation matrix:
       hour.L  hour.Q      hour.C  hour^4
pre.1  -0.6325  0.5345 -3.162e-01  0.1195
pre.2  -0.3162 -0.2673  6.325e-01 -0.4781
. . .
fup.5   0.6325  0.5345  3.162e-01  0.1195


. . .


Multivariate Tests:
                  Df test stat approx F num Df den Df   Pr(>F)
Pillai             1     0.933    24.32      4      7 0.000334
Wilks              1     0.067    24.32      4      7 0.000334
Hotelling-Lawley   1    13.894    24.32      4      7 0.000334
Roy                1    13.894    24.32      4      7 0.000334
```

Because `hour` is a within-subjects factor, we test its main effect as the regression intercept in the between-subjects model, using a response-transformation matrix for the `hour` contrasts.

Alternatively and equivalently, we can generate the response-transformation matrix P for the hypothesis directly:

```
> Hour <- model.matrix(~ hour, data=idata)
> dim(Hour)

[1] 15  5

> head(Hour, 5)

  (Intercept) hour.L  hour.Q      hour.C  hour^4
1           1 -0.6325  0.5345 -3.162e-01  0.1195
2           1 -0.3162 -0.2673  6.325e-01 -0.4781
3           1  0.0000 -0.5345 -4.096e-16  0.7171
4           1  0.3162 -0.2673 -6.325e-01 -0.4781
5           1  0.6325  0.5345  3.162e-01  0.1195

> linearHypothesis(mod.ok, "(Intercept) = 0", P=Hour[ , c(2:5)])

 Response transformation matrix:
       hour.L  hour.Q      hour.C  hour^4
pre.1  -0.6325  0.5345 -3.162e-01  0.1195
pre.2  -0.3162 -0.2673  6.325e-01 -0.4781
. . .
fup.5   0.6325  0.5345  3.162e-01  0.1195


Sum of squares and products for the hypothesis:
          hour.L    hour.Q    hour.C     hour^4
hour.L   0.01034     1.556    0.3672    -0.8244
hour.Q   1.55625   234.118   55.2469  -124.0137
hour.C   0.36724    55.247   13.0371   -29.2646
hour^4  -0.82435  -124.014  -29.2646    65.6907


. . .


Multivariate Tests:
                  Df test stat approx F num Df den Df   Pr(>F)
Pillai             1     0.933    24.32      4      7 0.000334
Wilks              1     0.067    24.32      4      7 0.000334
Hotelling-Lawley   1    13.894    24.32      4      7 0.000334
Roy                1    13.894    24.32      4      7 0.000334
```

As mentioned, this test simply duplicates part of the output from `Anova`, but suppose that we want to test the individual polynomial components of the `hour` main effect:

```
> linearHypothesis(mod.ok, "(Intercept) = 0", P=Hour[ , 2, drop=FALSE]) # linear

 Response transformation matrix:
        hour.L
pre.1  -0.6325
pre.2  -0.3162
. . .
fup.5   0.6325


. . .


Multivariate Tests:
                 Df test stat approx F num Df den Df Pr(>F)
Pillai            1    0.0001 0.001153      1     10  0.974
Wilks             1    0.9999 0.001153      1     10  0.974
Hotelling-Lawley  1    0.0001 0.001153      1     10  0.974
Roy               1    0.0001 0.001153      1     10  0.974

> linearHypothesis(mod.ok, "(Intercept) = 0", P=Hour[ , 3, drop=FALSE]) # quadratic

 Response transformation matrix:
        hour.Q
pre.1   0.5345
pre.2  -0.2673
. . .
fup.5   0.5345


. . .


Multivariate Tests:
                 Df test stat approx F num Df den Df    Pr(>F)
Pillai            1     0.834    50.19      1     10 0.0000336
Wilks             1     0.166    50.19      1     10 0.0000336
Hotelling-Lawley  1     5.019    50.19      1     10 0.0000336
Roy               1     5.019    50.19      1     10 0.0000336

> linearHypothesis(mod.ok, "(Intercept) = 0", P=Hour[ , c(2, 4:5)]) # all non-quadratic

 Response transformation matrix:
        hour.L      hour.C   hour^4
pre.1  -0.6325 -3.162e-01   0.1195
pre.2  -0.3162  6.325e-01  -0.4781
. . .
fup.5   0.6325  3.162e-01   0.1195


. . .


Multivariate Tests:
                 Df test stat approx F num Df den Df    Pr(>F)
Pillai            1     0.896    23.05      3      8 0.000272
Wilks             1     0.104    23.05      3      8 0.000272
Hotelling-Lawley  1     8.644    23.05      3      8 0.000272
Roy               1     8.644    23.05      3      8 0.000272
```

The `hour` main effect is more complex, therefore, than a simple quadratic trend.

Conclusions

In contrast to the standard R `anova` function, the `Anova` and `linearHypothesis` functions in the **car** package make it relatively simple to compute hypothesis tests that are typically used in applications of multivariate linear models, including to repeated-measures data. Although similar facilities for multivariate analysis of variance and repeated measures are provided by traditional statistical packages such as SAS and SPSS, we believe that the printed output from `Anova` and `linearHypothesis` is more readable, producing compact standard output and providing details when one wants them. These functions also return objects containing information—for example, SSP and response-transformation matrices—that may be used for further computations and in graphical displays, such as HE plots.

Acknowledgments

Bibliography

E. Anderson. The irises of the Gaspé Peninsula. *Bulletin of the American Iris Society*, 59:2–5, 1935. [p3]

D. Bates, M. Maechler, and B. Bolker. *lme4: Linear mixed-effects models using S4 classes*, 2012. R package version 0.999999-0. [p3]

P. Dalgaard. New functions for multivariate analysis. *R News*, 7(2):2–7, 2007. [p2, 8]

R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, Part II:179–188, 1936. [p3]

J. Fox. *Applied Regression Analysis and Generalized Linear Models*. Sage, Thousand Oaks, CA, second edition, 2008. [p1, 11]

J. Fox and S. Weisberg. *An R Companion to Applied Regression*. Sage, Thousand Oaks, CA, second edition, 2011. [p2, 3, 11]

J. Fox, M. Friendly, and G. Monette. Visualizing hypothesis tests in multivariate linear models: The heplots package for R. *Computational Statistics*, 24:233–246, 2009. [p5, 7]

M. Friendly. HE plots for multivariate linear models. *Journal of Computational and Graphical Statistics*, 16: 421–444, 2007. [p5, 7]

M. Friendly. HE plots for repeated measures designs. *Journal of Statistical Software*, 37(4):1–40, 2010. [p5, 7]

S. W. Greenhouse and S. Geisser. On methods in the analysis of profile data. *Psychometrika*, 24:95–112, 1959. [p11]

D. J. Hand and C. C. Taylor. *Multivariate Analysis of Variance and Repeated Measures: A Practical Approach for Behavioural Scientists*. Chapman and Hall, London, 1987. [p1, 2]

H. Huynh and L. S. Feldt. Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, 1:69–82, 1976. [p11]

T. Lumley. Analysis of complex survey samples. *Journal of Statistical Software*, 9(1):1–19, 2004. [p3]

J. W. Mauchly. Significance test for sphericity of a normal n-variate distribution. *The Annals of Mathematical Statistics*, 11:204–209, 1940. [p11]

D. F. Morrison. *Multivariate Statistical Methods*. Duxbury, Belmont CA, 4th edition, 2005. [p1]

R. G. O'Brien and M. K. Kaiser. MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin*, 97:316–333, 1985. [p1, 2, 8, 11]

J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2012. R package version 3.1-105. [p3]

C. R. Rao. *Linear Statistical Inference and Its Applications*. Wiley, New York, second edition, 1973. [p1, 2]

T. Therneau. *A Package for Survival Analysis in S*, 2012. R package version 2.36-14. [p3]

W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. ISBN 0-387-95457-0. [p3]

B. J. Winer. *Statistical Principles in Experimental Design*. McGraw-Hill, New York, second edition, 1971. [p1, 11]

John Fox
Department of Sociology
McMaster University
Canada
jfox@mcmaster.ca

Michael Friendly
Psychology Department
York University
Canada
friendly@yorku.ca

Sanford Weisberg
School of Statistics
University of Minnesota
USA
sandy@umn.edu