Multivariate Linear Models in R*

An Appendix to *An R Companion to Applied Regression*, third edition

John Fox & Sanford Weisberg

last revision: 2018-09-21

Abstract

The *multivariate linear model* is

$$\underset{(n \times m)}{\mathbf{Y}} = \underset{(n \times k+1)}{\mathbf{X}} \underset{(k+1 \times m)}{\mathbf{B}} + \underset{(n \times m)}{\mathbf{E}}$$

where $\mathbf{Y}$ is a matrix of n cases on m response variables; $\mathbf{X}$ is a model matrix with columns for $k+1$ 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. This 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 appendix to Fox and Weisberg (2019) explains how to use the `Anova()` and `linearHypothesis()` functions in the **car** package to test hypotheses for parameters in multivariate linear models, including models for repeated-measures data.

1 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. Much more extensive treatments may be found in the recommended reading for this appendix.

The multivariate general linear model is

$$\underset{(n \times m)}{\mathbf{Y}} = \underset{(n \times k+1)}{\mathbf{X}} \underset{(k+1 \times m)}{\mathbf{B}} + \underset{(n \times m)}{\mathbf{E}}$$

where $\mathbf{Y}$ is a matrix of n cases on m response variables; $\mathbf{X}$ is a model matrix with columns for $k+1$ 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.[1] The contents of the model matrix are exactly as in the univariate linear model (as described in Chapter 4 of *An R Companion to Applied Regression*, Fox and Weisberg, 2019—hereafter, the "*R Companion*"), and may contain, therefore, dummy regressors representing factors, polynomial or regression-spline terms, interaction regressors, and so on.

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 cases; $\boldsymbol{\varepsilon}_i'$ and $\boldsymbol{\varepsilon}_{i'}'$ are independent for $i \neq i'$; 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, and $\otimes$ is the Kronecker-product operator.

[1] A typographical note: $\mathbf{B}$ and $\mathbf{E}$ are, respectively, the upper-case Greek letters Beta and Epsilon. Because these are indistinguishable from the corresponding Roman letters B and E, we will denote the estimated regression coefficients as $\widehat{\mathbf{B}}$ and the residuals as $\widehat{\mathbf{E}}$.

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 the fact that there are several, generally correlated, 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

$$\begin{aligned}
\underset{(m\times m)}{\mathbf{SSP}_T} &= \mathbf{Y}'\mathbf{Y} - n\overline{\mathbf{y}}\,\overline{\mathbf{y}}' \\
&= \widehat{\mathbf{E}}'\widehat{\mathbf{E}} + \left(\widehat{\mathbf{Y}}'\widehat{\mathbf{Y}} - n\overline{\mathbf{y}}\,\overline{\mathbf{y}}'\right) \\
&= \mathbf{SSP}_R + \mathbf{SSP}_{\text{Reg}}
\end{aligned}$$

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. Let $\mathbf{SSP}_H$ represent the incremental SSP matrix for a hypothesis. 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 employed 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. There are F approximations to the null distributions of these test statistics. For example, for Wilks's Lambda, let s represent the degrees of freedom for the term that we are testing (i.e., the number of columns of the model matrix $\mathbf{X}$ pertaining to the term). Define

$$\begin{aligned}
r &= n - k - 1 - \frac{m-s+1}{2} \qquad (2) \\
u &= \frac{ms-2}{4} \\
t &= \begin{cases} \dfrac{\sqrt{m^2s^2-4}}{m^2+s^2-5} & \text{for } m^2+s^2-5>0 \\[2ex] 0 & \text{otherwise} \end{cases}
\end{aligned}$$

Rao (1973, p. 556) shows that under the null hypothesis,

$$F_0 = \frac{1-\Lambda^{1/t}}{\Lambda^{1/t}} \times \frac{rt-2u}{ms} \tag{3}$$

follows an approximate F-distribution with ms and $rt - 2u$ degrees of freedom, and that this result is exact if $\min(m, s) \le 2$ (a circumstance under which all four test statistics are equivalent).

Even more generally, suppose that we want to test the linear hypothesis

$$H_0\colon \underset{(q\times k+1)}{\mathbf{L}}\underset{(k+1\times m)}{\mathbf{B}} = \underset{(q\times m)}{\mathbf{C}} \tag{4}$$

where $\mathbf{L}$ is a hypothesis matrix of full-row rank $q \le k+1$, and the right-hand-side matrix $\mathbf{C}$ consists of constants (usually 0s).[2] Then the SSP matrix for the hypothesis is

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

and the various test statistics are based on the $p = \min(q, m)$ nonzero eigenvalues of $\mathbf{SSP}_H\mathbf{SSP}_R^{-1}$ (and the formulas in Equations 1, 2, and 3 are adjusted by substituting p for m).

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 model matrix, for example extending the linear hypothesis in Equation 4 to

$$H_0\colon \underset{(q\times k+1)}{\mathbf{L}}\underset{(k+1\times m)}{\mathbf{B}}\underset{(m\times v)}{\mathbf{P}} = \underset{(q\times v)}{\mathbf{C}} \tag{5}$$

Here, the *response-transformation matrix* $\mathbf{P}$ 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 $p = \min(q, v)$ nonzero eigenvalues of $\mathbf{SSP}_H(\mathbf{P}'\mathbf{SSP}_R\mathbf{P})^{-1}$.

2 Fitting and Testing Multivariate Linear Models in R

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 is a matrix of responses, with each column representing a response variable and each row a case; the right-hand side of the model and all other arguments to `lm` are precisely the same as for a univariate linear model (as described in Chapter 4 of the *R Companion*). 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, in a manner entirely analogous to that described in the *R Companion* (Section 5.3) for univariate linear models. We briefly demonstrate the use of these functions in this 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:

[2]Cf., Section 5.3.5 of the *R Companion* for linear hypotheses in univariate linear models.

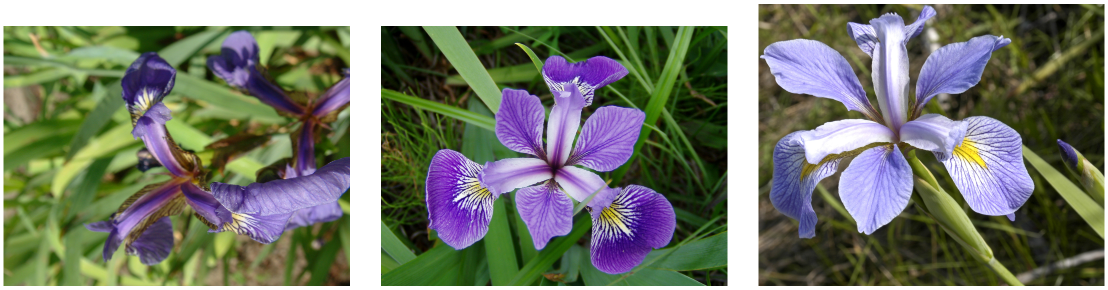

Figure 1: Three species of irises in the Anderson/Fisher data set: setosa (left), versicolor (center), and virginica (right). *Source*: The photographs are respectively by Radomil Binek, Danielle Langlois, and Frank Mayfield, and are distributed under the Creative Commons Attribution-Share Alike 3.0 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.

```
library(car)
```

```
Loading required package: carData
```

```
some(iris)
```

```
    Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
25           4.8         3.4          1.9         0.2     setosa
47           5.1         3.8          1.6         0.2     setosa
67           5.6         3.0          4.5         1.5 versicolor
73           6.3         2.5          4.9         1.5 versicolor
104          6.3         2.9          5.6         1.8  virginica
109          6.7         2.5          5.8         1.8  virginica
113          6.8         3.0          5.5         2.1  virginica
131          7.4         2.8          6.1         1.9  virginica
140          6.9         3.1          5.4         2.1  virginica
149          6.2         3.4          5.4         2.3  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 Section 5.2.3 of the *R Companion*); Figure 3 shows boxplots for each of the responses by species:

```
scatterplotMatrix(~ Sepal.Length + Sepal.Width + Petal.Length
    + Petal.Width | Species,
  data=iris, smooth=FALSE, regLine=FALSE, ellipse=TRUE,
  by.groups=TRUE, diagonal=FALSE, legend=list(coords="bottomleft"))
```

```
par(mfrow=c(2, 2))
for (response in c("Sepal.Length", "Sepal.Width",
                   "Petal.Length", "Petal.Width"))
  Boxplot(iris[, response] ~ Species, data=iris, ylab=response)
```

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

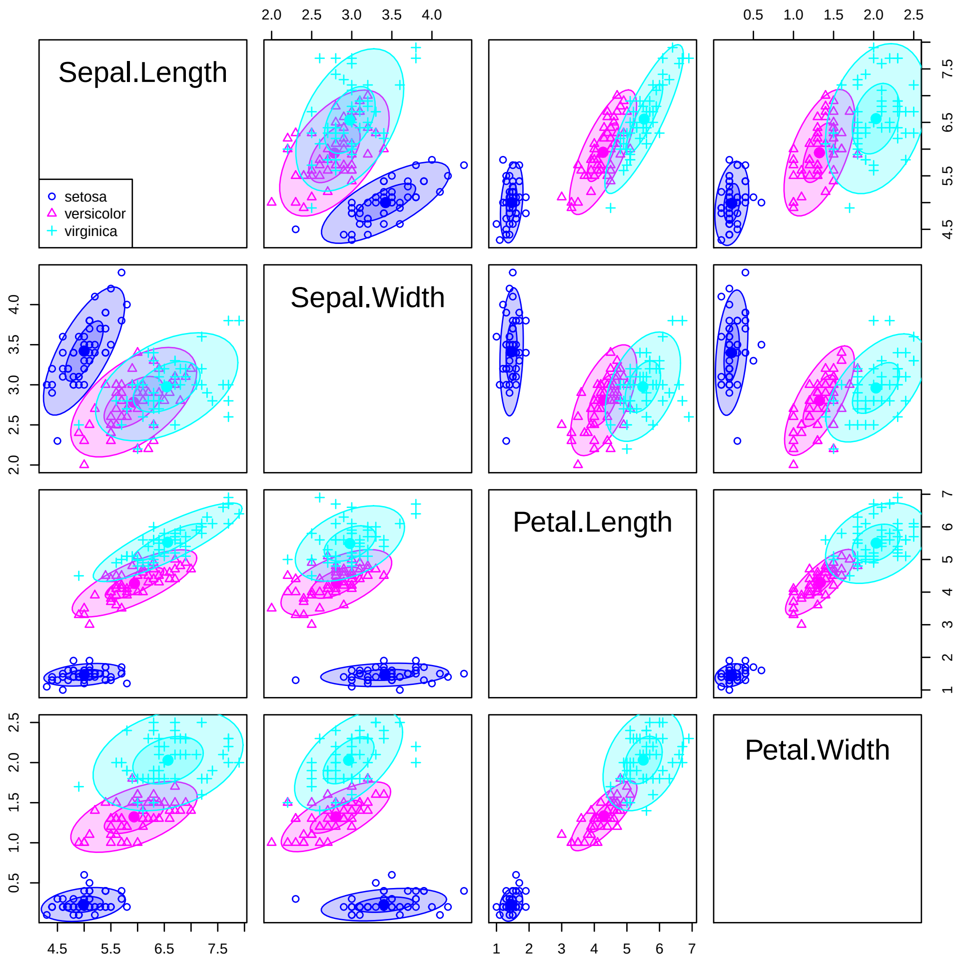

Figure 2: Scatterplot matrix for the Anderson/Fisher iris data, showing within-species 50 and 95% concentration ellipses.

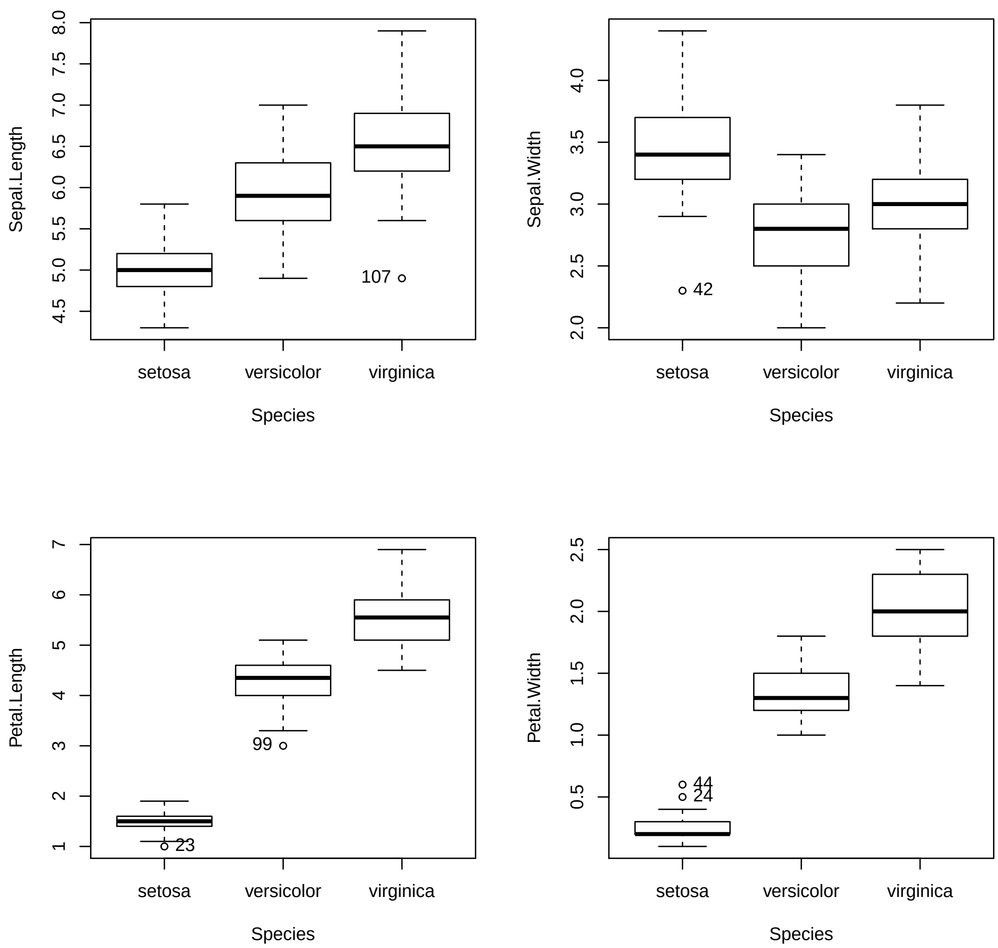

Figure 3: Boxplots for the response variables in the iris data set classified by species.

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(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width)
    ~ Species, data=iris)
class(mod.iris)
```

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

```
mod.iris
```

```

Call:
lm(formula = cbind(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width) ~
    Species, data = iris)

Coefficients:
                   Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
(Intercept)         5.006         3.428        1.462         0.246
Speciesversicolor   0.930        -0.658        2.798         1.080
Speciesvirginica    1.582        -0.454        4.090         1.780
```

```
summary(mod.iris)
```

```
Response Sepal.Length :

Call:
lm(formula = Sepal.Length ~ Species, data = iris)

Residuals:
   Min     1Q Median     3Q    Max
-1.688 -0.329 -0.006  0.312  1.312

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         5.0060     0.0728   68.76  < 2e-16
Speciesversicolor   0.9300     0.1030    9.03  8.8e-16
Speciesvirginica    1.5820     0.1030   15.37  < 2e-16

Residual standard error: 0.515 on 147 degrees of freedom
Multiple R-squared:  0.619,       Adjusted R-squared:  0.614
F-statistic:  119 on 2 and 147 DF,  p-value: <2e-16


Response Sepal.Width :

Call:
lm(formula = Sepal.Width ~ Species, data = iris)

Residuals:
   Min     1Q Median     3Q    Max
```

```
-1.128 -0.228  0.026  0.226  0.972

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         3.4280     0.0480   71.36  < 2e-16
Speciesversicolor  -0.6580     0.0679   -9.69  < 2e-16
Speciesvirginica   -0.4540     0.0679   -6.68  4.5e-10

Residual standard error: 0.34 on 147 degrees of freedom
Multiple R-squared:  0.401,	Adjusted R-squared:  0.393
F-statistic: 49.2 on 2 and 147 DF,  p-value: <2e-16


Response Petal.Length :

Call:
lm(formula = Petal.Length ~ Species, data = iris)

Residuals:
   Min     1Q Median     3Q    Max
-1.260 -0.258  0.038  0.240  1.348

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         1.4620     0.0609    24.0   <2e-16
Speciesversicolor   2.7980     0.0861    32.5   <2e-16
Speciesvirginica    4.0900     0.0861    47.5   <2e-16

Residual standard error: 0.43 on 147 degrees of freedom
Multiple R-squared:  0.941,	Adjusted R-squared:  0.941
F-statistic: 1.18e+03 on 2 and 147 DF,  p-value: <2e-16


Response Petal.Width :

Call:
lm(formula = Petal.Width ~ Species, data = iris)

Residuals:
   Min     1Q Median     3Q    Max
-0.626 -0.126 -0.026  0.154  0.474

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.2460     0.0289     8.5    2e-14
Speciesversicolor   1.0800     0.0409    26.4   <2e-16
Speciesvirginica    1.7800     0.0409    43.5   <2e-16

Residual standard error: 0.205 on 147 degrees of freedom
Multiple R-squared:  0.929,	Adjusted R-squared:  0.928
F-statistic:  960 on 2 and 147 DF,  p-value: <2e-16
```

The `lm()` function returns an S3 object of class `c("mlm", "lm")`. The printed representation of the object simply shows the estimated regression coefficients for each response, and the model summary 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:[3]

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


Type II MANOVA Tests: Pillai test statistic
        Df test stat approx F num Df den Df Pr(>F)
Species  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:
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length       38.956     13.6300      24.6246      5.6450
Sepal.Width        13.630     16.9620       8.1208      4.8084
Petal.Length       24.625      8.1208      27.2226      6.2718
Petal.Width         5.645      4.8084       6.2718      6.1566

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

Term: Species

Sum of squares and products for the hypothesis:
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length       63.212     -19.953       165.25      71.279
Sepal.Width       -19.953      11.345       -57.24     -22.933
Petal.Length      165.248     -57.240       437.10     186.774
Petal.Width        71.279     -22.933       186.77      80.413

Multivariate Tests: Species
                 Df test stat approx F num Df den Df    Pr(>F)
Pillai            2     1.192    53.47      8    290 < 2.2e-16
Wilks             2     0.023   199.15      8    288 < 2.2e-16
Hotelling-Lawley  2    32.477   580.53      8    286 < 2.2e-16
Roy               2    32.192  1166.96      4    145 < 2.2e-16
```

The `Anova()` function returns an object of class `"Anova.mlm"` which, when printed, produces a multivariate-analysis-of-variance ("MANOVA") table, by default reporting Pillai's test statistic; summarizing the object produces a more complete report. The object returned by `Anova()` may also be used in further computations, for example, for displays such as HE plots (Friendly, 2007; Fox

[3]The `Manova()` function in the **car** package is equivalent to `Anova()` applied to a multivariate linear model.

et al., 2009; Friendly, 2010). 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 test produced by the standard R `anova()` function:

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

```
Analysis of Variance Table

             Df Pillai approx F num Df den Df Pr(>F)
(Intercept)   1  0.993     5204      4    144 <2e-16
Species       2  1.192       53      8    290 <2e-16
Residuals   147
```

The null hypothesis is soundly rejected.

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*Speciesversicolor + 0.5*Speciesvirginica",
    verbose=TRUE)
```

```
Hypothesis matrix:
                                             (Intercept) Speciesversicolor
0.5*Speciesversicolor + 0.5*Speciesvirginica           0               0.5
                                             Speciesvirginica
0.5*Speciesversicolor + 0.5*Speciesvirginica              0.5

Right-hand-side matrix:
                                             Sepal.Length Sepal.Width Petal.Length
0.5*Speciesversicolor + 0.5*Speciesvirginica            0           0            0
                                             Petal.Width
0.5*Speciesversicolor + 0.5*Speciesvirginica           0

Estimated linear function (hypothesis.matrix %*% coef - rhs):
Sepal.Length  Sepal.Width Petal.Length  Petal.Width
       1.256       -0.556        3.444        1.430


Sum of squares and products for the hypothesis:
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length       52.585     -23.278      144.189      59.869
Sepal.Width       -23.278      10.305      -63.829     -26.503
Petal.Length      144.189     -63.829      395.371     164.164
Petal.Width        59.869     -26.503      164.164      68.163

Sum of squares and products for error:
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length       38.956     13.6300      24.6246      5.6450
Sepal.Width        13.630     16.9620       8.1208      4.8084
Petal.Length       24.625      8.1208      27.2226      6.2718
Petal.Width         5.645      4.8084       6.2718      6.1566
```

```
Multivariate Tests:
                 Df test stat approx F num Df den Df    Pr(>F)
Pillai            1    0.9673   1063.9      4    144 < 2.2e-16
Wilks             1    0.0327   1063.9      4    144 < 2.2e-16
Hotelling-Lawley  1   29.5520   1063.9      4    144 < 2.2e-16
Roy               1   29.5520   1063.9      4    144 < 2.2e-16
```

```
linearHypothesis(mod.iris, "Speciesversicolor = Speciesvirginica",
    verbose=TRUE)
```

```
Hypothesis matrix:
                                     (Intercept) Speciesversicolor Speciesvirginica
Speciesversicolor = Speciesvirginica           0                 1               -1

Right-hand-side matrix:
                                     Sepal.Length Sepal.Width Petal.Length
Speciesversicolor = Speciesvirginica            0           0            0
                                     Petal.Width
Speciesversicolor = Speciesvirginica           0

Estimated linear function (hypothesis.matrix %*% coef - rhs):
Sepal.Length  Sepal.Width Petal.Length  Petal.Width
      -0.652       -0.204       -1.292       -0.700


Sum of squares and products for the hypothesis:
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length      10.6276      3.3252      21.0596       11.41
Sepal.Width        3.3252      1.0404       6.5892        3.57
Petal.Length      21.0596      6.5892      41.7316       22.61
Petal.Width       11.4100      3.5700      22.6100       12.25

Sum of squares and products for error:
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length       38.956     13.6300      24.6246      5.6450
Sepal.Width        13.630     16.9620       8.1208      4.8084
Petal.Length       24.625      8.1208      27.2226      6.2718
Petal.Width         5.645      4.8084       6.2718      6.1566

Multivariate Tests:
                 Df test stat approx F num Df den Df    Pr(>F)
Pillai            1   0.74525   105.31      4    144 < 2.2e-16
Wilks             1   0.25475   105.31      4    144 < 2.2e-16
Hotelling-Lawley  1   2.92535   105.31      4    144 < 2.2e-16
Roy               1   2.92535   105.31      4    144 < 2.2e-16
```

The argument `verbose=TRUE` to `linearHypothesis()` shows the hypothesis matrix $\mathbf{L}$ and right-hand-side matrix $\mathbf{C}$ for the linear hypothesis in Equation 4 (page 3). 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*Speciesversicolor + 0.5*Speciesvirginica" 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 ccategory 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("setosa vs. versicolor & virginica", "versicolor & virginica")
contrasts(iris$Species) <- C
contrasts(iris$Species)
```

```
           setosa vs. versicolor & virginica versicolor & virginica
setosa                                   1.0                      0
versicolor                              -0.5                      1
virginica                               -0.5                     -1
```

```
(mod.iris.2 <- update(mod.iris))
```

```

Call:
lm(formula = cbind(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width) ~
    Species, data = iris)

Coefficients:
                                          Sepal.Length  Sepal.Width  Petal.Length
(Intercept)                                5.843         3.057        3.758
Speciessetosa vs. versicolor & virginica  -0.837         0.371       -2.296
Speciesversicolor & virginica             -0.326        -0.102       -0.646
                                          Petal.Width
(Intercept)                                1.199
Speciessetosa vs. versicolor & virginica  -0.953
Speciesversicolor & virginica             -0.350
```

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

```

Sum of squares and products for the hypothesis:
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length       52.585     -23.278      144.189      59.869
Sepal.Width       -23.278      10.305      -63.829     -26.503
Petal.Length      144.189     -63.829      395.371     164.164
Petal.Width        59.869     -26.503      164.164      68.163

Sum of squares and products for error:
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length       38.956     13.6300      24.6246      5.6450
Sepal.Width        13.630     16.9620       8.1208      4.8084
Petal.Length       24.625      8.1208      27.2226      6.2718
Petal.Width         5.645      4.8084       6.2718      6.1566

Multivariate Tests:
                 Df test stat approx F num Df den Df    Pr(>F)
```

```
Pillai            1    0.9673   1063.9      4    144 < 2.2e-16
Wilks             1    0.0327   1063.9      4    144 < 2.2e-16
Hotelling-Lawley  1   29.5520   1063.9      4    144 < 2.2e-16
Roy               1   29.5520   1063.9      4    144 < 2.2e-16
```

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

```
Sum of squares and products for the hypothesis:
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length      10.6276      3.3252      21.0596       11.41
Sepal.Width        3.3252      1.0404       6.5892        3.57
Petal.Length      21.0596      6.5892      41.7316       22.61
Petal.Width       11.4100      3.5700      22.6100       12.25

Sum of squares and products for error:
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length       38.956     13.6300      24.6246      5.6450
Sepal.Width        13.630     16.9620       8.1208      4.8084
Petal.Length       24.625      8.1208      27.2226      6.2718
Petal.Width         5.645      4.8084       6.2718      6.1566

Multivariate Tests:
                 Df test stat approx F num Df den Df    Pr(>F)
Pillai            1   0.74525   105.31      4    144 < 2.2e-16
Wilks             1   0.25475   105.31      4    144 < 2.2e-16
Hotelling-Lawley  1   2.92535   105.31      4    144 < 2.2e-16
Roy               1   2.92535   105.31      4    144 < 2.2e-16
```

Finally, we can code the response-transformation matrix $\mathbf{P}$ in Equation 5 (page 3) 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
Species: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("Speciesversicolor = 0", "Speciesvirginica = 0"),
  P=matrix(c(1, 0, 0, 0))) # equivalent
```

```
 Response transformation matrix:
             [,1]
Sepal.Length    1
```

```
Sepal.Width     0
Petal.Length    0
Petal.Width     0

Sum of squares and products for the hypothesis:
       [,1]
[1,] 63.212

Sum of squares and products for error:
       [,1]
[1,] 38.956

Multivariate Tests:
                 Df test stat approx F num Df den Df    Pr(>F)
Pillai            2   0.61871   119.27      2    147 < 2.2e-16
Wilks             2   0.38129   119.27      2    147 < 2.2e-16
Hotelling-Lawley  2   1.62265   119.27      2    147 < 2.2e-16
Roy               2   1.62265   119.27      2    147 < 2.2e-16
```

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

```
Anova(lm(Sepal.Length ~ Species, data=iris))
```

```
Anova Table (Type II tests)

Response: Sepal.Length
          Sum Sq  Df F value Pr(>F)
Species     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.

3 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 term "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 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 models 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, 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 employ contrived 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. The data set `OBrienKaiser` is provided by the **carData** package:

```
some(OBrienKaiser)
```

```
   treatment gender pre.1 pre.2 pre.3 pre.4 pre.5 post.1 post.2 post.3 post.4 post.5
2    control      M     4     4     5     3     4      2      2      3      5      3
4    control      F     5     4     7     5     4      2      2      3      5      3
5    control      F     3     4     6     4     3      6      7      8      6      3
6          A      M     7     8     7     9     9      9      9     10      8      9
7          A      M     5     5     6     4     5      7      7      8     10      8
11         B      M     3     3     4     2     3      5      4      7      5      4
12         B      M     6     7     8     6     3      9     10     11      9      6
13         B      F     5     5     6     8     6      4      6      6      8      6
14         B      F     2     2     3     1     2      5      6      7      5      2
16         B      F     4     5     7     5     4      7      7      8      6      7
   fup.1 fup.2 fup.3 fup.4 fup.5
2      4     5     6     4     1
4      4     4     5     3     4
5      4     3     6     4     3
6      9    10    11     9     6
7      8     9    11     9     8
11     5     6     8     6     5
12     8     7    10     8     7
13     7     7     8    10     8
14     6     7     8     6     3
16     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 versus the average of the experimental groups, and the second contrast is for treatment `A` versus treatment `B`. The frequency table for `treatment` by `sex` 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 teaching 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"), c(5, 5, 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
4   pretest    4
5   pretest    5
6  posttest    1
7  posttest    2
8  posttest    3
9  posttest    4
10 posttest    5
11 followup    1
12 followup    2
13 followup    3
14 followup    4
15 followup    5
```

We begin by reshaping the data set from "wide" to "long" format to facilitate graphing the data; we will eventually use the original wide version of the data set for repeated-measures analysis.

```
OBrien.long <- reshape(OBrienKaiser,
    varying=c("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"),
    v.names="score",
    timevar="phase.hour", direction="long")
OBrien.long$phase <- ordered(
      c("pre", "post", "fup")[1 + ((OBrien.long$phase.hour - 1) %/% 5)],
    levels=c("pre", "post", "fup"))
OBrien.long$hour <- ordered(1 + ((OBrien.long$phase.hour - 1) %% 5))
dim(OBrien.long)
```

```
[1] 240   7
```

```
head(OBrien.long, 25) # first 25 rows
```

```
     treatment gender phase.hour score id phase hour
1.1    control      M          1     1  1   pre    1
2.1    control      M          1     4  2   pre    1
3.1    control      M          1     5  3   pre    1
4.1    control      F          1     5  4   pre    1
5.1    control      F          1     3  5   pre    1
6.1          A      M          1     7  6   pre    1
7.1          A      M          1     5  7   pre    1
8.1          A      F          1     2  8   pre    1
9.1          A      F          1     3  9   pre    1
10.1         B      M          1     4 10   pre    1
11.1         B      M          1     3 11   pre    1
12.1         B      M          1     6 12   pre    1
13.1         B      F          1     5 13   pre    1
14.1         B      F          1     2 14   pre    1
15.1         B      F          1     2 15   pre    1
16.1         B      F          1     4 16   pre    1
1.2    control      M          2     2  1   pre    2
2.2    control      M          2     4  2   pre    2
3.2    control      M          2     6  3   pre    2
4.2    control      F          2     4  4   pre    2
5.2    control      F          2     4  5   pre    2
6.2          A      M          2     8  6   pre    2
7.2          A      M          2     5  7   pre    2
8.2          A      F          2     3  8   pre    2
9.2          A      F          2     3  9   pre    2
```

We then compute mean reading scores for combinations of `gender`, `treatment`, `phase`, and `hour`:

```
Means <- as.data.frame(ftable(with(OBrien.long,
    tapply(score,
        list(treatment=treatment, gender=gender, phase=phase, hour=hour),
        mean))))
names(Means)[5] <- "score"
dim(Means)
```

```
[1] 90  5
```

```
head(Means, 25) # first 25 means
```

```
   treatment gender phase hour  score
1    control      F   pre    1 4.0000
2          A      F   pre    1 2.5000
3          B      F   pre    1 3.2500
4    control      M   pre    1 3.3333
5          A      M   pre    1 6.0000
6          B      M   pre    1 4.3333
7    control      F  post    1 4.0000
8          A      F  post    1 3.0000
9          B      F  post    1 5.5000
10   control      M  post    1 3.0000
```

```
11         A      M  post    1 8.0000
12         B      M  post    1 6.6667
13   control      F   fup    1 4.0000
14         A      F   fup    1 5.5000
15         B      F   fup    1 6.7500
16   control      M   fup    1 4.3333
17         A      M   fup    1 8.5000
18         B      M   fup    1 7.0000
19   control      F   pre    2 4.0000
20         A      F   pre    2 3.0000
21         B      F   pre    2 3.5000
22   control      M   pre    2 4.0000
23         A      M   pre    2 6.5000
24         B      M   pre    2 4.6667
25   control      F  post    2 4.5000
```

Finally, we employ the `xyplot` function in the **lattice** package to graph the means:[4]

```
library(lattice)
xyplot(score ~ hour | phase + treatment, groups=gender, type="b",
    strip=function(...) strip.default(strip.names=c(TRUE, TRUE), ...),
    lty=1:2, pch=c(15, 1), col=1:2, cex=1.25,
    ylab="Mean Reading Score", data=Means,
    key=list(title="Gender",  cex.title=1,
      text=list(c("Female", "Male")), lines=list(lty=1:2, col=1:2),
      points=list(pch=c(15, 1), col=1:2, cex=1.25)))
```

The resulting graph is shown in Figure 4. It appears as if reading improves across phases in the two experimental treatments but not in the control group (suggesting 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 to 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)
mod.ok


Call:
lm(formula = 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)

Coefficients:
```

[4] Lattice graphics are described in Section 9.3.1 of the *R Companion*, and in more detail in Sarkar (2008).

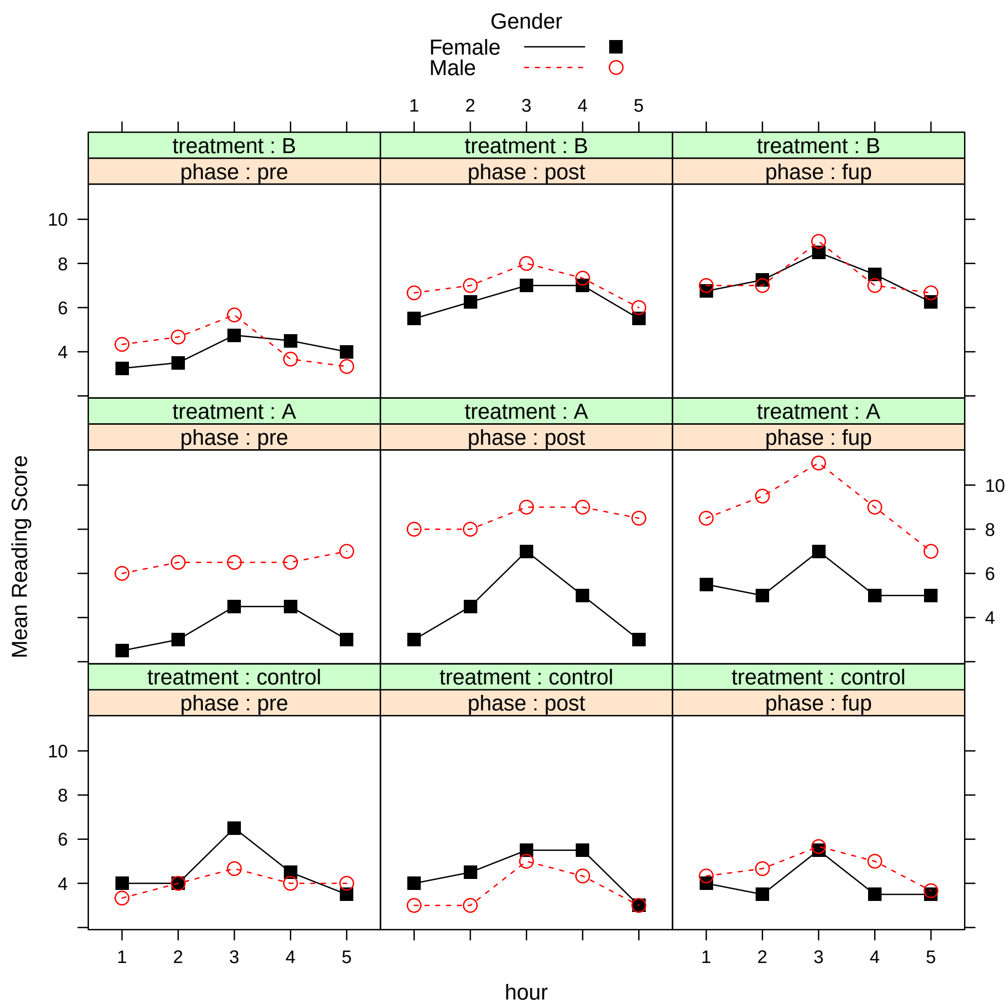

Figure 4: Mean reading score by gender, treatment, phase, and hour, for the O'Brien-Kaiser data.

```
                    pre.1     pre.2     pre.3     pre.4     pre.5     post.1
(Intercept)          3.90e+00  4.28e+00  5.43e+00  4.61e+00  4.14e+00  5.03e+00
treatment1           1.18e-01  1.39e-01 -7.64e-02  1.81e-01  1.94e-01  7.64e-01
treatment2          -2.29e-01 -3.33e-01 -1.46e-01 -7.08e-01 -6.67e-01  2.92e-01
gender1             -6.53e-01 -7.78e-01 -1.81e-01 -1.11e-01 -6.39e-01 -8.61e-01
treatment1:gender1  -4.93e-01 -3.89e-01 -5.49e-01 -1.81e-01 -1.94e-01 -6.81e-01
treatment2:gender1   6.04e-01  5.83e-01  2.71e-01  7.08e-01  1.17e+00  9.58e-01
                    post.2    post.3    post.4    post.5    fup.1     fup.2
(Intercept)          5.54e+00  6.92e+00  6.36e+00  4.83e+00  6.01e+00  6.15e+00
treatment1           8.96e-01  8.33e-01  7.22e-01  9.17e-01  9.24e-01  1.03e+00
treatment2           1.87e-01 -2.50e-01  8.33e-02  8.70e-18 -6.25e-02 -6.25e-02
gender1             -4.58e-01 -4.17e-01 -5.28e-01 -1.00e+00 -5.97e-01 -9.03e-01
treatment1:gender1  -6.04e-01 -3.33e-01 -5.56e-01 -5.00e-01 -2.15e-01 -1.60e-01
treatment2:gender1   6.88e-01  2.50e-01  9.17e-01  1.25e+00  6.88e-01  1.19e+00
                    fup.3     fup.4     fup.5
(Intercept)          7.78e+00  6.17e+00  5.35e+00
treatment1           1.10e+00  9.58e-01  8.82e-01
treatment2          -1.25e-01  1.25e-01  2.29e-01
gender1             -7.78e-01 -8.33e-01 -4.31e-01
treatment1:gender1  -3.47e-01 -4.17e-02 -1.74e-01
treatment2:gender1   8.75e-01  1.12e+00  3.96e-01
```

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))
```

```
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, by specifying the argument `type=3`. Although, as in univariate models, we generally prefer type-II tests (see Section 5.3.4 of the *R Companion*), 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 if it is not.

- By default, Pillai's test statistic is displayed; we invite the reader to examine the other three multivariate test statistics.

- The results show that the anticipated `hour` effect has a small p-value, but the `treatment` $\times$ `phase` and `treatment` $\times$ `gender` interactions have p-values that exceed 0.05. There is, however, a small p-values for the `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.

3.1 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 $\mathbf{\Sigma}$ 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-Geiser correction (Greenhouse and Geisser, 1959) and the Huynh-Feldt correction (Huynh and Feldt, 1976). We illustrate for the O'Brien-Kaiser data, suppressing the multivariate tests:

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

```
Univariate Type III Repeated-Measures ANOVA Assuming Sphericity

                        Sum Sq num Df Error SS den Df F value  Pr(>F)
(Intercept)               6759      1    228.1     10  296.39 9.2e-09
treatment                  180      2    228.1     10    3.94  0.0547
gender                      83      1    228.1     10    3.66  0.0848
treatment:gender           130      2    228.1     10    2.86  0.1045
phase                      130      2     80.3     20   16.13 6.7e-05
treatment:phase             78      4     80.3     20    4.85  0.0067
gender:phase                 2      2     80.3     20    0.28  0.7566
treatment:gender:phase      10      4     80.3     20    0.64  0.6424
hour                       104      4     62.5     40   16.69 4.0e-08
treatment:hour               1      8     62.5     40    0.09  0.9992
```

```
gender:hour                   3     4    62.5    40    0.45  0.7716
treatment:gender:hour         8     8    62.5    40    0.62  0.7555
phase:hour                   11     8    96.2    80    1.18  0.3216
treatment:phase:hour          7    16    96.2    80    0.35  0.9901
gender:phase:hour             9     8    96.2    80    0.93  0.4956
treatment:gender:phase:hour  14    16    96.2    80    0.74  0.7496


Mauchly Tests for Sphericity

                            Test statistic p-value
phase                                0.749   0.273
treatment:phase                      0.749   0.273
gender:phase                         0.749   0.273
treatment:gender:phase               0.749   0.273
hour                                 0.066   0.008
treatment:hour                       0.066   0.008
gender:hour                          0.066   0.008
treatment:gender:hour                0.066   0.008
phase:hour                           0.005   0.449
treatment:phase:hour                 0.005   0.449
gender:phase:hour                    0.005   0.449
treatment:gender:phase:hour          0.005   0.449


Greenhouse-Geisser and Huynh-Feldt Corrections
 for Departure from Sphericity

                            GG eps Pr(>F[GG])
phase                         0.80    0.00028
treatment:phase               0.80    0.01269
gender:phase                  0.80    0.70896
treatment:gender:phase        0.80    0.61162
hour                          0.46    9.8e-05
treatment:hour                0.46    0.97862
gender:hour                   0.46    0.62843
treatment:gender:hour         0.46    0.64136
phase:hour                    0.45    0.33452
treatment:phase:hour          0.45    0.93037
gender:phase:hour             0.45    0.44908
treatment:gender:phase:hour   0.45    0.64634

                             HF eps Pr(>F[HF])
phase                       0.92786 1.1247e-04
treatment:phase             0.92786 8.4388e-03
gender:phase                0.92786 7.4086e-01
treatment:gender:phase      0.92786 6.3200e-01
hour                        0.55928 2.3009e-05
treatment:hour              0.55928 9.8866e-01
gender:hour                 0.55928 6.6455e-01
treatment:gender:hour       0.55928 6.6930e-01
```

```
phase:hour                  0.73306 3.2966e-01
treatment:phase:hour        0.73306 9.7523e-01
gender:phase:hour           0.73306 4.7803e-01
treatment:gender:phase:hour 0.73306 7.0801e-01
```

The non-sphericity tests have small p-values 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` $\times$ `phase` interaction is associated with a p-value smaller than 0.05.

3.2 Using `linearHypothesis()` with repeated-measures designs

As for simpler multivariate linear models (discussed in Section 2), 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 `iterm` arguments in a call to `linearHypothesis()`:

```
linearHypothesis(mod.ok, "(Intercept) = 0", idata=idata,
  idesign=~phase*hour, iterms="hour") # test hour main effect
```

```
 Response transformation matrix:
        hour.L   hour.Q      hour.C   hour^4
pre.1  -0.63246  0.53452 -3.1623e-01  0.11952
pre.2  -0.31623 -0.26726  6.3246e-01 -0.47809
pre.3   0.00000 -0.53452 -4.0960e-16  0.71714
pre.4   0.31623 -0.26726 -6.3246e-01 -0.47809
pre.5   0.63246  0.53452  3.1623e-01  0.11952
post.1 -0.63246  0.53452 -3.1623e-01  0.11952
post.2 -0.31623 -0.26726  6.3246e-01 -0.47809
post.3  0.00000 -0.53452 -4.0960e-16  0.71714
post.4  0.31623 -0.26726 -6.3246e-01 -0.47809
post.5  0.63246  0.53452  3.1623e-01  0.11952
fup.1  -0.63246  0.53452 -3.1623e-01  0.11952
fup.2  -0.31623 -0.26726  6.3246e-01 -0.47809
fup.3   0.00000 -0.53452 -4.0960e-16  0.71714
fup.4   0.31623 -0.26726 -6.3246e-01 -0.47809
fup.5   0.63246  0.53452  3.1623e-01  0.11952

Sum of squares and products for the hypothesis:
         hour.L    hour.Q    hour.C     hour^4
hour.L  0.010345    1.5562   0.36724   -0.82435
hour.Q  1.556250  234.1182  55.24686 -124.01365
hour.C  0.367241   55.2469  13.03707  -29.26455
hour^4 -0.824354 -124.0137 -29.26455   65.69068

Sum of squares and products for error:
        hour.L   hour.Q  hour.C  hour^4
hour.L  89.7333  49.6106 -9.7167 -25.418
hour.Q  49.6106  46.6429  1.3522 -17.409
hour.C  -9.7167   1.3522 21.8083  16.111
hour^4 -25.4181 -17.4094 16.1107  29.315
```

```
Multivariate Tests:
                 Df test stat approx F num Df den Df    Pr(>F)
Pillai            1    0.9329   24.315      4      7 0.0003345
Wilks             1    0.0671   24.315      4      7 0.0003345
Hotelling-Lawley  1   13.8944   24.315      4      7 0.0003345
Roy               1   13.8944   24.315      4      7 0.0003345
```

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))
```

```
   (Intercept)   hour.L   hour.Q      hour.C   hour^4
1            1 -0.63246  0.53452 -3.1623e-01  0.11952
2            1 -0.31623 -0.26726  6.3246e-01 -0.47809
3            1  0.00000 -0.53452 -4.0960e-16  0.71714
4            1  0.31623 -0.26726 -6.3246e-01 -0.47809
5            1  0.63246  0.53452  3.1623e-01  0.11952
6            1 -0.63246  0.53452 -3.1623e-01  0.11952
7            1 -0.31623 -0.26726  6.3246e-01 -0.47809
8            1  0.00000 -0.53452 -4.0960e-16  0.71714
9            1  0.31623 -0.26726 -6.3246e-01 -0.47809
10           1  0.63246  0.53452  3.1623e-01  0.11952
11           1 -0.63246  0.53452 -3.1623e-01  0.11952
12           1 -0.31623 -0.26726  6.3246e-01 -0.47809
13           1  0.00000 -0.53452 -4.0960e-16  0.71714
14           1  0.31623 -0.26726 -6.3246e-01 -0.47809
15           1  0.63246  0.53452  3.1623e-01  0.11952
attr(,"assign")
[1] 0 1 1 1 1
attr(,"contrasts")
attr(,"contrasts")$hour
[1] "contr.poly"
```

```
linearHypothesis(mod.ok, "(Intercept) = 0",
  P=Hour[ , c(2:5)]) # test hour main effect (equivalent)
```

```
 Response transformation matrix:
        hour.L   hour.Q      hour.C   hour^4
pre.1  -0.63246  0.53452 -3.1623e-01  0.11952
pre.2  -0.31623 -0.26726  6.3246e-01 -0.47809
pre.3   0.00000 -0.53452 -4.0960e-16  0.71714
pre.4   0.31623 -0.26726 -6.3246e-01 -0.47809
pre.5   0.63246  0.53452  3.1623e-01  0.11952
post.1 -0.63246  0.53452 -3.1623e-01  0.11952
post.2 -0.31623 -0.26726  6.3246e-01 -0.47809
post.3  0.00000 -0.53452 -4.0960e-16  0.71714
post.4  0.31623 -0.26726 -6.3246e-01 -0.47809
```

```
post.5  0.63246  0.53452  3.1623e-01  0.11952
fup.1  -0.63246  0.53452 -3.1623e-01  0.11952
fup.2  -0.31623 -0.26726  6.3246e-01 -0.47809
fup.3   0.00000 -0.53452 -4.0960e-16  0.71714
fup.4   0.31623 -0.26726 -6.3246e-01 -0.47809
fup.5   0.63246  0.53452  3.1623e-01  0.11952

Sum of squares and products for the hypothesis:
          hour.L    hour.Q    hour.C     hour^4
hour.L  0.010345    1.5562   0.36724   -0.82435
hour.Q  1.556250  234.1182  55.24686 -124.01365
hour.C  0.367241   55.2469  13.03707  -29.26455
hour^4 -0.824354 -124.0137 -29.26455   65.69068

Sum of squares and products for error:
         hour.L   hour.Q  hour.C  hour^4
hour.L  89.7333  49.6106 -9.7167 -25.418
hour.Q  49.6106  46.6429  1.3522 -17.409
hour.C  -9.7167   1.3522 21.8083  16.111
hour^4 -25.4181 -17.4094 16.1107  29.315

Multivariate Tests:
                 Df test stat approx F num Df den Df    Pr(>F)
Pillai            1    0.9329   24.315      4      7 0.0003345
Wilks             1    0.0671   24.315      4      7 0.0003345
Hotelling-Lawley  1   13.8944   24.315      4      7 0.0003345
Roy               1   13.8944   24.315      4      7 0.0003345
```

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  -0.63246
pre.2  -0.31623
pre.3   0.00000
pre.4   0.31623
pre.5   0.63246
post.1 -0.63246
post.2 -0.31623
post.3  0.00000
post.4  0.31623
post.5  0.63246
fup.1  -0.63246
fup.2  -0.31623
fup.3   0.00000
fup.4   0.31623
fup.5   0.63246
```

```
Sum of squares and products for the hypothesis:
         hour.L
hour.L 0.010345

Sum of squares and products for error:
       hour.L
hour.L 89.733

Multivariate Tests:
                 Df test stat  approx F num Df den Df Pr(>F)
Pillai            1   0.00012 0.0011528      1     10 0.9736
Wilks             1   0.99988 0.0011528      1     10 0.9736
Hotelling-Lawley  1   0.00012 0.0011528      1     10 0.9736
Roy               1   0.00012 0.0011528      1     10 0.9736
```

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

```
 Response transformation matrix:
         hour.Q
pre.1   0.53452
pre.2  -0.26726
pre.3  -0.53452
pre.4  -0.26726
pre.5   0.53452
post.1  0.53452
post.2 -0.26726
post.3 -0.53452
post.4 -0.26726
post.5  0.53452
fup.1   0.53452
fup.2  -0.26726
fup.3  -0.53452
fup.4  -0.26726
fup.5   0.53452

Sum of squares and products for the hypothesis:
       hour.Q
hour.Q 234.12

Sum of squares and products for error:
       hour.Q
hour.Q 46.643

Multivariate Tests:
                 Df test stat approx F num Df den Df    Pr(>F)
Pillai            1    0.8339   50.194      1     10 3.356e-05
Wilks             1    0.1661   50.194      1     10 3.356e-05
Hotelling-Lawley  1    5.0194   50.194      1     10 3.356e-05
Roy               1    5.0194   50.194      1     10 3.356e-05
```

linearHypothesis(mod.ok, "(Intercept) = 0", P=Hour[, 4, drop=FALSE]) # cubic

```
 Response transformation matrix:
           hour.C
pre.1  -3.1623e-01
pre.2   6.3246e-01
pre.3  -4.0960e-16
pre.4  -6.3246e-01
pre.5   3.1623e-01
post.1 -3.1623e-01
post.2  6.3246e-01
post.3 -4.0960e-16
post.4 -6.3246e-01
post.5  3.1623e-01
fup.1  -3.1623e-01
fup.2   6.3246e-01
fup.3  -4.0960e-16
fup.4  -6.3246e-01
fup.5   3.1623e-01

Sum of squares and products for the hypothesis:
       hour.C
hour.C 13.037

Sum of squares and products for error:
       hour.C
hour.C 21.808

Multivariate Tests:
                 Df test stat approx F num Df den Df  Pr(>F)
Pillai            1   0.37414    5.978      1     10 0.03455
Wilks             1   0.62586    5.978      1     10 0.03455
Hotelling-Lawley  1   0.59780    5.978      1     10 0.03455
Roy               1   0.59780    5.978      1     10 0.03455
```

linearHypothesis(mod.ok, "(Intercept) = 0", P=Hour[, 5, drop=FALSE]) # quartic

```
 Response transformation matrix:
        hour^4
pre.1   0.11952
pre.2  -0.47809
pre.3   0.71714
pre.4  -0.47809
pre.5   0.11952
post.1  0.11952
post.2 -0.47809
post.3  0.71714
post.4 -0.47809
post.5  0.11952
fup.1   0.11952
fup.2  -0.47809
fup.3   0.71714
```

```
fup.4  -0.47809
fup.5   0.11952

Sum of squares and products for the hypothesis:
       hour^4
hour^4 65.691

Sum of squares and products for error:
       hour^4
hour^4 29.315

Multivariate Tests:
                 Df test stat approx F num Df den Df    Pr(>F)
Pillai            1   0.69144   22.408      1     10 0.0007997
Wilks             1   0.30856   22.408      1     10 0.0007997
Hotelling-Lawley  1   2.24082   22.408      1     10 0.0007997
Roy               1   2.24082   22.408      1     10 0.0007997
```

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  -0.63246 -3.1623e-01  0.11952
pre.2  -0.31623  6.3246e-01 -0.47809
pre.3   0.00000 -4.0960e-16  0.71714
pre.4   0.31623 -6.3246e-01 -0.47809
pre.5   0.63246  3.1623e-01  0.11952
post.1 -0.63246 -3.1623e-01  0.11952
post.2 -0.31623  6.3246e-01 -0.47809
post.3  0.00000 -4.0960e-16  0.71714
post.4  0.31623 -6.3246e-01 -0.47809
post.5  0.63246  3.1623e-01  0.11952
fup.1  -0.63246 -3.1623e-01  0.11952
fup.2  -0.31623  6.3246e-01 -0.47809
fup.3   0.00000 -4.0960e-16  0.71714
fup.4   0.31623 -6.3246e-01 -0.47809
fup.5   0.63246  3.1623e-01  0.11952

Sum of squares and products for the hypothesis:
          hour.L    hour.C    hour^4
hour.L  0.010345   0.36724  -0.82435
hour.C  0.367241  13.03707 -29.26455
hour^4 -0.824354 -29.26455  65.69068

Sum of squares and products for error:
        hour.L  hour.C  hour^4
hour.L  89.7333 -9.7167 -25.418
hour.C  -9.7167 21.8083  16.111
hour^4 -25.4181 16.1107  29.315

Multivariate Tests:
```

```
                 Df test stat approx F num Df den Df    Pr(>F)
Pillai            1    0.8963    23.05      3      8 0.0002724
Wilks             1    0.1037    23.05      3      8 0.0002724
Hotelling-Lawley  1    8.6439    23.05      3      8 0.0002724
Roy               1    8.6439    23.05      3      8 0.0002724
```

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

4 Complementary Reading and References

The material in the first section of this appendix is based on Fox (2016, Sec. 9.5).

There are many texts that treat MANOVA and multivariate linear models: 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). As mentioned, Winer (1971, Chap. 7) presents the traditional univariate approach to repeated-measures.

References

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

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

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

Fox, J. (2016). *Applied Regression Analysis and Generalized Linear Models*. Sage, Thousand Oaks CA, third edition.

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

Fox, J. and Weisberg, S. (2019). *An R Companion to Applied Regression*. Sage, Thousand Oaks, CA, third edition.

Friendly, M. (2007). HE plots for multivariate linear models. *Journal of Computational and Graphical Statistics*, 16:421–444.

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

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

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

Huynh, H. and Feldt, L. S. (1976). 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.

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

Morrison, D. F. (2005). *Multivariate Statistical Methods*. Duxbury, Belmont CA, 4th edition.

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

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* Wiley, New York, second edition.

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R.* Springer, New York.

Winer, B. J. (1971). *Statistical Principles in Experimental Design.* McGraw-Hill, New York, second edition.