New Functions for Multivariate Analysis

Peter Dalgaard

R (and S-PLUS) used to have limited support for multivariate tests. We had the `manova` function, which extended the features of `aov` to multivariate responses, but like `aov`, this effectively assumed a balanced design, and was not capable of dealing with the within-subject transformations that are commonly used in repeated measurement modelling.

Although the methods encoded in procedures available in SAS and SPSS can seem somewhat old-fashioned, they do have some added value relative to analysis by mixed model methodology, and they have a strong tradition in several applied areas. It was thus worthwhile to extend R's capabilities to handle contrast tests, as well as Greenhouse-Geisser and Huynh-Feldt epsilons. The extensions also provide flexible ways of dealing with linear models with a multivariate response.

Theoretical setting

The general setup is given by

$$Y \sim N(\Xi B, I \otimes \Sigma)$$

Here, Y is $N \times p$ matrix and Σ is a $p \times p$ covariance matrix. The rows y_i of Y are independent with the same covariance Σ.

Ξ is a $N \times k$ design matrix (the reader will have to apologise that I am not using the traditional X, but that symbol is reserved for other purposes later on) and B is a $k \times p$ matrix of regression coefficients.

Thus, we have the same linear model for all p response coordinates, with separate parameters contained in the columns of B.

Standard test procedures

From classical univariate and multivariate theory, a number of standard tests are available. We shall focus on three of them:

1. Testing hypotheses of simplified mean value structure: This reduction is required to be the same for all coordinates, effectively replacing the design matrix Ξ with one spanning a subspace. Such tests take the form of generalized F tests, replacing the variance ratio by

   $$R = \mathrm{MS}_{\mathrm{res}}^{-1}\mathrm{MS}_{\mathrm{eff}}$$

   in which the MS terms are matrices which generalize the mean square terms from analysis of variance. Under the hypothesis, R should be distributed around the unit matrix (in the sense that the two MS matrices both have mean Σ; notice, however, that $\mathrm{MS}_{\mathrm{eff}}$ will be rank deficient whenever the degrees of freedom for the effect is less than p), but for a test statistic we need to reduce R to a scalar measure. Four such measures are cited in the literature, namely Wilks's Λ, the Pillai trace, the Hotelling-Lawley trace, and Roy's greatest root. These are all based on combinations of the eigenvalues of R. Details can be found in, e.g., Hand and Taylor (1987).

   Wilks's Λ is equivalent to the likelihood ratio test, but R and S-PLUS have traditionally favoured the Pillai trace based on the (rather vague) recommendations cited in Hand and Taylor (1987). Each test can be converted to an approximately F distributed statistic. If the tests are for a single-degree-of-freedom hypothesis, the matrix R has only one non-zero eigenvalue and all four tests are equivalent.

2. Testing whether Σ is proportional to a given matrix, say Σ_0 (which is usually the unit matrix I): This is known as Mauchly's test of sphericity. It is based on a comparison of the determinant and the trace of $U = \Sigma_0^{-1}S$ where S is the SSD (deviation sum-of-squares-and-products) matrix ($\mathrm{MS}_{\mathrm{res}}$ instead of S is equivalent). Specifically

   $$W = \det(U)/\mathrm{tr}(U/p)^p$$

   is close to 1 if U is close to a diagonal matrix of dimension p with a constant value along the diagonal. The test statistic $-f \log W$ is an asymptotic χ^2 on $p(p+1)/2 - 1$ degrees of freedom (where f is the degrees of freedom for the covariance matrix. An improved approximation is found in Anderson (1958).

3. Testing linear hypotheses as in point 1 above, but *assuming* that sphericity holds. In this case, we are assuming that the covariance is known up to a constant, and thus are effectively in a univariate setting of (weighted) least squares theory. The relevant F statistic will have (pf_1, pf_2) degrees of freedom if the coordinatewise tests have (f_1, f_2) degrees of freedom.

Within-subject transformations

It is often necessary to consider a transformation of responses: If, for instance, coordinates are repeated measures, we might wish to test for "no change over time" or "same changes over time in different groups" (profile analysis). This leads to analysis of

 ISSN 1609-3631

within-subjects contrasts, rather than of the observations themselves.

If we assume a covariance structure with random effects both between and within subject, the contrasts will cancel out the between-subject variation. They will effectively behave *as if* the between-subject term wasn't there and satisfy a sphericity condition.

Hence we consider the transformed response YT' (which has rows Ty_i). In many cases T is chosen to annihilate another matrix X, i.e. it satisfies $TX = 0$; by rotation invariance, different such choices of T will lead to the same inference as long as they have maximal rank. In such cases it may well be more convenient to specify X rather than T.

For profile analysis, X could be a p-vector of ones. In that case, the hypothesis of *compound symmetry* implies sphericity of $T\Sigma T'$ w.r.t. TT' (but sphericity does not imply compound symmetry), and the hypothesis $\mathsf{E}YT' = 0$ is equivalent to a flat (constant level) profile.

More elaborate within-subject designs exist. As an example, consider a a complete two-way layout, in which case we might want to look at the sets of (a) Contrasts between row means, (b) Contrasts between column means, and (c) Interaction contrasts,

These contrasts connect to the theory of balanced mixed-effects analysis of variance. A model with "all interactions with subject are considered random", i.e. random effects of subjects, as well as random interaction between subject and rows and between subjects and columns, implies sphericity of each of the above contrasts. The proportionality constants will be different linear combinations of the random effect variances.

A common pattern for such sets of contrasts is as follows: Define two subspaces of $\mathbb{R}^p$, defined by matrices X and M, such that $\mathrm{span}(X) \subset \mathrm{span}(M)$. The transformation is calculated as $T = P_M - P_X$ where P denotes orthogonal projection. Actually, T defined like this would be singular and therefore T is thinned by deletion of linearly dependent rows (it can fairly easily be seen that it does not matter exactly how this is done). Putting $M = I$ (the identity matrix) will make T satisfy $TX = 0$ which is consistent with the situation described earlier.

The choice of X and M is most easily done by viewing them as design matrices for linear models in p-dimensional space. Consider again a two-way intrasubject design. To make T a transformation which calculates contrasts between column means, choose M to describe a model with different means in each column and X to describe a single mean for all data. Preferably, but equivalently for a balanced design, let M describe an additive model in rows and columns and let X be a model in which there is only a row effect.

There is a hidden assumption involved in the choice of the *orthogonal* projection onto $\mathrm{span}(M)$. This calculates (for each subject) an ordinary least squares (OLS) fit and for some covariance structures, a weighted fit may be more appropriate. However, OLS is not actually biased, and the efficiency that we might gain is offset by the need to estimate a rather large number of parameters to find the optimal weights.

The epsilons

Assume that the conditions for a balanced analysis of variance are roughly valid. We may decide to compute, say, column means for each subject and test whether the contrasts are zero on average. There could be a loss of efficiency in the use of unweighted means, but the critical issue for an F test is whether the sphericity condition holds for the contrasts between column means.

If the sphericity condition is not valid, then the F test is in principle wrong. Instead, we could use one of the multivariate tests, but they will often have low power due to the estimation of a large number of parameters in the empirical covariance matrix.

For this reason, a methodology has evolved in which "near-sphericity" data are analyzed by F tests, but applying the so-called epsilon corrections to the degrees of freedom. The theory originates in Box (1954a) in which it is shown that F is approximately distributed as $F(\varepsilon f_1, \varepsilon f_2)$, where

$$\varepsilon = \frac{\sum \lambda_i^2/p}{(\sum \lambda_i/p)^2}$$

and the λ_i are the eigenvalues of the true covariance matrix (after contrast transformation, and with respect to a similarly transformed identity matrix). One may notice that $1/p \leq \varepsilon \leq 1$; the upper limit corresponds to sphericity (all eigenvalues equal) and the lower limit corresponds to the case where there is one dominating eigenvalue, so that data are effectively one-dimensional. Box (1954b) details the result as a function of the elements of the covariance matrix in the two-way layout.

The Box correction requires knowledge of the true covariance matrix. Greenhouse and Geisser (1959) suggest the use of $\varepsilon_{\mathrm{GG}}$, which is simply the empirical verson of ε, inserting the empirical covariance matrix for the true one. However, it is easily seen that this estimator is biased, at least when the true epsilon is close to 1, since $\varepsilon_{\mathrm{GG}} < 1$ almost surely when $p > 1$.

The Huynh-Feldt correction (Huynh and Feldt, 1976) is

$$\varepsilon_{\mathrm{HF}} = \frac{(f+1)p\varepsilon_{\mathrm{GG}} - 2}{p(f - p\varepsilon_{\mathrm{GG}})}$$

where f is the number of degrees of freedom for the empirical covariance matrix. (The original paper has N instead of $f + 1$ in the numerator for the split-plot design. A correction was published 15 years later

(Lecoutre, 1991), but another 15 years later, this error is still present in SAS and SPSS.)

Notice that ε_{HF} can be larger than one, in which case you should use the uncorrected F test. Also, ε_{HF} is obtained by bias-correcting the numerator and denominator of ε_{GG}, which is not guaranteed to be helpful, and simulation studies in (Huynh and Feldt, 1976) indicate that it actually makes the approximations worse when $\varepsilon_{GG} < 0.7$ or so.

Implementation in R

Basics

Quite a lot of the calculations could be copied from the existing `manova` and `summary.manova` code for balanced designs. Also, objects of class `mlm`, inheriting from `lm`, were already defined as the output from `lm(Y~....)` when `Y` is a matrix.

It was necessary to add the following new functions

- `SSD` creates an object of (S3) class `SSD` which is the sums of squares and products matrix augmented by degrees of freedom and information about the call. This is a generic function with a method for `mlm` objects.
- `estVar` calculates the estimated variance-covariance matrix. It has methods for `SSD` and `mlm` objects. In the former case, it just normalizes the SSD by the degrees of freedom, and in the latter case it calculates `SSD(object)` and then applies the `SSD` method.
- `anova.mlm` is used to compare two multivariate linear models or partition a single model in the usual cumulative way. The various multivariate tests can be specified using `test="Pillai"`, etc., just as in `summary.manova`. In addition, it is possible to specify `test="Spherical"` which gives the F test under assumption of sphericity.
- `mauchly.test` tests for sphericity of a covariance matrix with respect to another.
- `sphericity` calculates the ε_{GG} and ε_{HF} based on the SSD matrix and its degrees of freedom. This function is private to the `stats` namespace (at least currently, as of R version 2.6.0) and used to generate the headings and adjusted p-values in `anova.mlm`.

Representing transformations

As previously described, sphericity tests and tests of linear hypotheses may make sense only for a transformed response. Hence we need to be able to specify transformations in `anova.mlm` and `mauchly.test`. The code has several ways to deal with this:

- The transformation matrix can be given directly using the argument `T`.
- It is possible to specify the arguments `X` and `M` as the matrices described previously. The defaults are set so that the default transformation is the identity.
- It is also possible to specify `X` and/or `M` using model formulas. In that case, they usually need to refer to a data frame which describes the intra-subject design. This can be given in the `idata` argument. The default for `idata` is a data frame consisting of the single variable `index=1:p` which may be used to specify, e.g., a polynomial response for equispaced data.

Example

These data from the book by Maxwell and Delaney (1990) are also used by Baron and Li (2006). They show reaction times where ten subjects respond to stimuli in the absence and presence of ambient noise, and using stimuli tilted at three different angles.

```
> reacttime <- matrix(c(
+ 420, 420, 480, 480, 600, 780,
+ 420, 480, 480, 360, 480, 600,
+ 480, 480, 540, 660, 780, 780,
+ 420, 540, 540, 480, 780, 900,
+ 540, 660, 540, 480, 660, 720,
+ 360, 420, 360, 360, 480, 540,
+ 480, 480, 600, 540, 720, 840,
+ 480, 600, 660, 540, 720, 900,
+ 540, 600, 540, 480, 720, 780,
+ 480, 420, 540, 540, 660, 780),
+ ncol = 6, byrow = TRUE,
+ dimnames=list(subj=1:10,
+       cond=c("deg0NA", "deg4NA", "deg8NA",
+              "deg0NP", "deg4NP", "deg8NP")))
```

The same data are used by `example(estVar)` and `example(anova.mlm)`, so you can load the `reacttime` matrix just by running the examples. The following is mainly an expanded explanation of those examples.

First let us calculate the estimated covariance matrix:

```
> mlmfit <- lm(reacttime~1)
> estVar(mlmfit)
         cond
cond      deg0NA deg4NA deg8NA deg0NP deg4NP deg8NP
  deg0NA    3240   3400   2960   2640   3600   2840
  deg4NA    3400   7400   3600    800   4000   3400
  deg8NA    2960   3600   6240   4560   6400   7760
  deg0NP    2640    800   4560   7840   8000   7040
  deg4NP    3600   4000   6400   8000  12000  11200
  deg8NP    2840   3400   7760   7040  11200  13640
```

In this case there is no between-subjects structure, except for a common mean, so the result is equivalent to `var(reacttime)`.

Next we consider tests for whether the response depends on the design at all. We generate a contrast transformation which is orthogonal to an intercept-only within-subject model; this is equivalent to any full-rank set of within-subject contrasts. A test based on multivariate normal theory is performed as follows.

```
> mlmfit0 <- update(mlmfit, ~0)
> anova(mlmfit, mlmfit0, X=~1)
Analysis of Variance Table

Model 1: reacttime ~ 1
Model 2: reacttime ~ 1 - 1

Contrasts orthogonal to
~1

  Res.Df Df Gen.var.  Pillai approx F
1      9     1249.57
2     10  1  2013.16    0.95    17.38
  num Df den Df   Pr(>F)
1
2      5      5 0.003534 **
```

This gives the default Pillai test, but actually, you get the same result with the other multivariate tests since the degrees of freedom (per coordinate) only changes by one.

To perform the same test, but assuming sphericity of the covariance matrix, just add an argument to the anova call:

```
> anova(mlmfit, mlmfit0, X=~1, test="Spherical")
Analysis of Variance Table

Model 1: reacttime ~ 1
Model 2: reacttime ~ 1 - 1

Contrasts orthogonal to
~1

Greenhouse-Geisser epsilon: 0.4855
Huynh-Feldt epsilon:        0.6778

  Res.Df Df Gen.var.      F num Df
1      9      1249.6
2     10  1   2013.2 38.028      5
  den Df    Pr(>F)    G-G Pr    H-F Pr
1
2     45 4.471e-15 2.532e-08 7.393e-11
```

It can be noticed that the epsilon corrections in this case are rather strong, but that the corrected p-values nevertheless are considerably more significant with this approach, reflecting the low power of the multivariate test. This is commonly the case when the number of replications is low.

To test the hypothesis of sphericity, we employ Mauchly's criterion:

```
> mauchly.test(mlmfit, X=~1)

        Mauchly's test of sphericity
        Contrasts orthogonal to
        ~1

data:  SSD matrix from lm(formula = reacttime ~ 1)

W = 0.0311, p-value = 0.04765
```

Accordingly, the hypothesis of sphericity is rejected at the 0.05 significance level.

However, the analysis of contrasts completely disregards the two-way intrasubject design. To generate tests for overall effects of `deg` and `noise`, as well as interaction between the two, we do as follows: First we need to set up the `idata` data frame to describe the design.

```
> idata <- expand.grid(
+             deg=c("0", "4", "8"),
+             noise=c("A", "P"))
> idata
  deg noise
1   0     A
2   4     A
3   8     A
4   0     P
5   4     P
6   8     P
```

Then we can specify the tests using model formulas to specify the X and M matrices. To test for an effect of `deg` we let `M` specify an additive model in `deg` and `noise` and let `X` be the model with `noise` alone.

```
> anova(mlmfit, mlmfit0, M = ~ deg + noise,
+       X = ~ noise,
+       idata = idata, test="Spherical")
Analysis of Variance Table

Model 1: reacttime ~ 1
Model 2: reacttime ~ 1 - 1

Contrasts orthogonal to
~noise


Contrasts spanned by
~deg + noise

Greenhouse-Geisser epsilon: 0.9616
Huynh-Feldt epsilon:        1.2176

  Res.Df Df Gen.var.      F num Df
1      9      1007.0
2     10  1   2703.2 40.719      2
  den Df    Pr(>F)    G-G Pr    H-F Pr
1
2     18 2.087e-07 3.402e-07 2.087e-07
```

It might at this point be helpful to explain the roles of `X` and `M` a bit further. To do this, we need to access an internal function in the `stats` namespace, namely `proj.matrix`. This function calculates

the projection matrix for a given design matrix, i.e. $P_X = X(X'X)^{-1}X'$. In the above context, we have

```
M <- model.matrix(~deg+noise, data=idata)
P1 <- stats:::proj.matrix(M)
X <- model.matrix(~noise, data=idata)
P2 <- stats:::proj.matrix(X)
```

The two design matrices are (eliminating attributes)

```
> M
  (Intercept) deg4 deg8 noiseP
1           1    0    0      0
2           1    1    0      0
3           1    0    1      0
4           1    0    0      1
5           1    1    0      1
6           1    0    1      1
> X
  (Intercept) noiseP
1           1      0
2           1      0
3           1      0
4           1      1
5           1      1
6           1      1
```

For printing the projection matrices, it is useful to include the `fractions` function from `MASS`.

```
> library("MASS")
> fractions(P1)
  1    2    3    4    5    6
1  2/3  1/6  1/6  1/3 -1/6 -1/6
2  1/6  2/3  1/6 -1/6  1/3 -1/6
3  1/6  1/6  2/3 -1/6 -1/6  1/3
4  1/3 -1/6 -1/6  2/3  1/6  1/6
5 -1/6  1/3 -1/6  1/6  2/3  1/6
6 -1/6 -1/6  1/3  1/6  1/6  2/3
> fractions(P2)
  1   2   3   4   5   6
1 1/3 1/3 1/3   0   0   0
2 1/3 1/3 1/3   0   0   0
3 1/3 1/3 1/3   0   0   0
4   0   0   0 1/3 1/3 1/3
5   0   0   0 1/3 1/3 1/3
6   0   0   0 1/3 1/3 1/3
```

Here, `P2` is readily recognized as the operator that replaces each of the first three values by their average, and similarly the last three. The other one, `P1`, is a little harder, but if you look long enough, you will recognize the formula $\hat{x}_{ij} = \bar{x}_{i\cdot} + \bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot}$ from two-way ANOVA.

The transformation T is the difference between `P1` and `P2`

```
> fractions(P1-P2)
  1    2    3    4    5    6
1  1/3 -1/6 -1/6  1/3 -1/6 -1/6
2 -1/6  1/3 -1/6 -1/6  1/3 -1/6
3 -1/6 -1/6  1/3 -1/6 -1/6  1/3
4  1/3 -1/6 -1/6  1/3 -1/6 -1/6
5 -1/6  1/3 -1/6 -1/6  1/3 -1/6
6 -1/6 -1/6  1/3 -1/6 -1/6  1/3
```

This is the representation of $\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot}$ where i and j index levels of `deg` and `noise`, respectively. The matrix has (row) rank 2; for the actual computations, only the first two rows are used.

The important aspect of this matrix is that it assigns equal weights to observations at different levels of `noise` and that it constructs a maximal set of within-`deg` differences. Any other such choice of transformation leads to equivalent results, since it will be a full-rank transformation of T. For instance:

```
> T1 <- rbind(c(1,-1,0,1,-1,0),c(1,0,-1,1,0,-1))
> anova(mlmfit, mlmfit0, T=T1,
+     idata = idata, test = "Spherical")
Analysis of Variance Table

Model 1: reacttime ~ 1
Model 2: reacttime ~ 1 - 1

Contrast matrix
 1 -1  0  1 -1  0
 1  0 -1  1  0 -1
Greenhouse-Geisser epsilon: 0.9616
Huynh-Feldt epsilon:        1.2176

  Res.Df Df Gen.var.      F num Df
1      9       12084
2     10  1    32438 40.719      2
1 den Df    Pr(>F)    G-G Pr    H-F Pr
2     18 2.087e-07 3.402e-07 2.087e-07
```

This differs from the previous analysis only in the `Gen.var.` column, which is not invariant to scaling and transformation of data.

Returning to the interpretation of the results, notice that the epsilons are much closer to one than before. This is consistent with a covariance structure induced by a mixed model containing a random interaction between subject and `deg`. If we perform the similar procedure for `noise` we get epsilons of exactly 1, because we are dealing with a single-df contrast.

```
> anova(mlmfit, mlmfit0, M = ~ deg + noise,
+       X = ~ deg,
+       idata = idata, test="Spherical")
Analysis of Variance Table

Model 1: reacttime ~ 1
Model 2: reacttime ~ 1 - 1

Contrasts orthogonal to
~deg


Contrasts spanned by
~deg + noise

Greenhouse-Geisser epsilon: 1
Huynh-Feldt epsilon:        1

  Res.Df Df Gen.var.      F num Df
1      9        1410
```

```
2     10  1     6030 33.766      1
  den Df    Pr(>F)    G-G Pr    H-F Pr
1
2      9 0.0002560 0.0002560 0.0002560
```

However, we really shouldn't be doing these tests on individual effects of `deg` and `noise` if there is interaction between the two, and as the following output shows, there is:

```
> anova(mlmfit, mlmfit0, X = ~ deg + noise,
+       idata = idata, test = "Spherical")
Analysis of Variance Table

Model 1: reacttime ~ 1
Model 2: reacttime ~ 1 - 1

Contrasts orthogonal to
~deg + noise

Greenhouse-Geisser epsilon: 0.904
Huynh-Feldt epsilon:        1.118

  Res.Df Df Gen.var.     F num Df
1      9      316.58
2     10  1   996.34 45.31      2
  den Df    Pr(>F)    G-G Pr    H-F Pr
1
2     18 9.424e-08 3.454e-07 9.424e-08
```

Finally, to test between-within interactions, rephrase them as effects of between-factors on within-contrasts. To illustrate this, we introduce a fake grouping `f` of the `reacttime` data into two groups. The interaction between `f` and the (implied) column factor is obtained by testing whether the contrasts orthogonal to `~1` depend on `f`.

```
> f <- factor(rep(1:2, 5))
> mlmfit2 <- update(mlmfit, ~f)
> anova(mlmfit2, X = ~1, test = "Spherical")
Analysis of Variance Table


Contrasts orthogonal to
~1

Greenhouse-Geisser epsilon: 0.4691
Huynh-Feldt epsilon:        0.6758

            Df       F num Df den Df
(Intercept)  1 34.9615      5     40
f            1  0.2743      5     40
Residuals    8
               Pr(>F)    G-G Pr    H-F Pr
(Intercept) 1.382e-13 2.207e-07 8.254e-10
f             0.92452   0.79608   0.86456
Residuals
```

More detailed tests of interactions between `f` and `deg`, `f` and `noise`, and the three-way interaction can be constructed using `M` and `X` settings as described previously.

Bibliography

T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, 1958.

J. Baron and Y. Li. Notes on the use of R for psychology experiments and questionnaires, 2006. URL `http://www.psych.upenn.edu/~baron/rpsych/rpsych.html`.

G. E. P. Box. Some theorems on quadratic forms applied in the study of analysis of variance problems, i. effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics*, 25 (2):290–302, 1954a.

G. E. P. Box. Some theorems on quadratic forms applied in the study of analysis of variance problems, ii.effect of inequality of variance and of correlation between errors in the two-way classification. *Annals of Mathematical Statistics*, 25(3):484–498, 1954b.

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

D. J. Hand and C. C. Taylor. *Multivariate Analysis of Variance and Repeated Measures*. Chapman and Hall, 1987.

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(1):69–82, 1976.

B. Lecoutre. A correction for the $\tilde{\epsilon}$ approximate test in repeated measures designs with two or more independent groups. *Journal of Educational Statistics*, 16(4):371–372, 1991.

S. E. Maxwell and H. D. Delaney. *Designing Experiments and Analyzing Data: A model comparison perspective*. Brooks/Cole, Pacific Grove, CA, 1990.

Peter Dalgaard
Deptartment of Biostatistics
University of Copenhagen
`Peter.Dalgaard@R-project.org`