

Random and Mixed Effects Models - Part III

Statistics 149

Spring 2006



Quasi-F Tests

When we get to more than two categorical factors, some times there are not nice F tests available. For example, consider the 3-way model with 3 random effects (A: a levels, B: b levels, C: c levels) and m observations per treatment combination. The expected mean squares are

MS	df	$E[MS]$
MSA	$a - 1$	$\sigma^2 + mbc\sigma_\alpha^2 + mc\sigma_{\alpha\beta}^2 + mb\sigma_{\alpha\gamma}^2 + m\sigma_{\alpha\beta\gamma}^2$
MSB	$b - 1$	$\sigma^2 + mac\sigma_\beta^2 + mc\sigma_{\alpha\beta}^2 + ma\sigma_{\beta\gamma}^2 + m\sigma_{\alpha\beta\gamma}^2$
MSC	$c - 1$	$\sigma^2 + mab\sigma_\gamma^2 + mb\sigma_{\alpha\gamma}^2 + ma\sigma_{\beta\gamma}^2 + m\sigma_{\alpha\beta\gamma}^2$
MSAB	$(a - 1)(b - 1)$	$\sigma^2 + mc\sigma_{\alpha\beta}^2 + m\sigma_{\alpha\beta\gamma}^2$
MSAC	$(a - 1)(b - 1)$	$\sigma^2 + mc\sigma_{\alpha\gamma}^2 + m\sigma_{\alpha\beta\gamma}^2$
MSBC	$(a - 1)(c - 1)$	$\sigma^2 + ma\sigma_{\beta\gamma}^2 + m\sigma_{\alpha\beta\gamma}^2$
MSABC	$(a - 1)(b - 1)(c - 1)$	$\sigma^2 + m\sigma_{\alpha\beta\gamma}^2$
MSE	$(m - 1)abc$	σ^2

If we want to test $H_0 : \sigma_{\alpha\beta}^2 = 0$, we can use

$$F_{AB} = \frac{MSAB}{MSABC}$$

as $E[MSAB] = E[MSABC]$ under H_0 .

Now consider testing $H_0 : \sigma_{\alpha}^2 = 0$. Looking through the previous table there is no entry that has the same expectation as $E[MSA]$ has under H_0 , i.e. there is nothing satisfying

$$MSX = \sigma^2 + mc\sigma_{\alpha\beta}^2 + mb\sigma_{\alpha\gamma}^2 + m\sigma_{\alpha\beta\gamma}^2$$

So we can't use a test statistic of the form

$$F = \frac{MSA}{MSX}$$

One alternative is to use a Quasi- F test (also known as Satterthwaite's Approximate F test).

The idea is to find a linear combination of mean squares

$$\hat{L} = c_1 MS_1 + \dots + c_h MS_h$$

that has the desired expectation, i.e.

$$\begin{aligned} E[\hat{L}] &= c_1 E[MS_1] + \dots + c_h E[MS_h] \\ &= MSX \end{aligned}$$

Then the hypothesis can be tested with the test statistic

$$F_A^{**} = \frac{MSA}{\hat{L}}$$

In this example,

$$\hat{L} = MSAB + MSAC - MSABC$$

will work. Note that there may be multiple possibilities for the denominator in the F ratio.

This test statistic has an approximate F distribution with $a - 1$ and df degrees of freedom where

$$df = \frac{(c_1MS_1 + \dots + c_hMS_h)^2}{\frac{(c_1MS_1)^2}{df_1} + \dots + \frac{(c_hMS_h)^2}{df_h}}$$

Example: A study on daily output of a factory examined three factors, operators (3 levels), machines (2 levels), and batches (5 levels). For each treatment combination, 3 observations were taken. The resulting ANOVA table was generated for this table

Source of Variation	<i>SS</i>	<i>df</i>	<i>MS</i>
A (operators)	17.3	2	8.65
B (machines)	4.2	1	4.20
C (batches)	24.8	4	6.20
AB	4.8	2	2.40
AC	31.7	8	3.96
BC	12.5	4	3.13
ABC	11.9	8	1.49
Error	137.7	60	2.30

For example to examine the effect of operators, we can use the Quasi- F statistic

$$\begin{aligned}
 F_A^{**} &= \frac{MSA}{MSAB + MSAC - MSABC} \\
 &= \frac{8.65}{2.40 + 3.96 + 1.49} = 1.78
 \end{aligned}$$

The denominator degrees of freedom for this case is

$$df = \frac{4.87^2}{\frac{2.40^2}{2} + \frac{3.96^2}{8} + \frac{-1.49^2}{8}} = 4.63$$

Note that this formula will usually give non-integer degrees of freedom. Rounding to the closest integer is usually fine when working with tables. Note that many packages will allow non-integer degrees in their probability functions.

The p -value for this F statistic is 0.267 (=pf(1.78,2,4.63, lower.tail=F))

Note that exactly how you deal with these non-integer degrees of freedom usually doesn't matter, as exhibited by

```
> pf(1.78, 2, 4.63, lower.tail=F)
[1] 0.2670347
```

```
> pf(1.78, 2, 4, lower.tail=F)
[1] 0.2799474
```

```
> pf(1.78, 2, 5, lower.tail=F)
[1] 0.2607598
```

For the example, if anybody is interested, none of the test are significant. The smallest p -value (0.094) occurs with the test on the AC interaction.

As mentioned before, it is possible that there is more than one linear combination L that can be chosen. In the case where that occurs, you generally want to choose the one that has the least number of terms involved. If there are two possibilities and they both have the same number of terms involved, pick the one that avoids low order interactions. This test tends to work better when the mean square components are close to estimating the measurement error variance σ^2 .

One way of figuring out a possible L is to consider the estimate of the variance component of interest. In the example, to estimate σ_α^2 , one estimate is

$$\hat{\sigma}_\alpha^2 = \frac{MSA - MSAB - MSAC + MSABC}{mbc} = \frac{MSA - \hat{L}}{mbc}$$

Often the L you want to use will come from a relationship like this.

General Form of the Linear Mixed Model

All the the examples involving random effects, the models can be written in the general matrix form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$$
$$\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} G & \mathbf{0} \\ \mathbf{0} & \sigma^2 R \end{bmatrix} \right)$$

where $\boldsymbol{\beta}$ is the vector of fixed effects and \mathbf{u} is the vector of random effects, \mathbf{X} and \mathbf{Z} are known design matrices, and G and R are scaled covariance matrices.

Elements of the matrix G are the variances of the random effects, plus possibly covariances between the random effects. R is the correlation matrix of the measurement errors. In the examples discussed so far, both matrices have been diagonal matrices, but in general they don't need to be.

For example, consider a mixed model with one fixed effect (A: 4 levels), and two random effects (C: 2 levels, D: 3 levels) and 2 observations per treatment combination and no interactions. Then

$$\beta^T = \left[\mu \quad \alpha_1 \quad \alpha_2 \quad \alpha_3 \quad \alpha_4 \right]$$

$$\mathbf{u}^T = \left[\gamma_1 \quad \gamma_2 \quad \delta_1 \quad \delta_2 \quad \delta_3 \right]$$

\mathbf{Z} is a 48×5 matrix of 0s and 1s. In each row there will be two 1s, one in the first two columns (indicating the C random effect) and one in the last three columns (indicating the D random effect)

$$G = \text{diag} \left(\sigma_\gamma^2 \quad \sigma_\gamma^2 \quad \sigma_\delta^2 \quad \sigma_\delta^2 \quad \sigma_\delta^2 \right)$$

$$R = I_{48}$$

Restricted Maximum Likelihood - REML

The likelihood function in the random effects case has two components as the model being fit can be considered as a hierarchical model. The two components are

$$\begin{aligned}\mathbf{Y}|\mathbf{u} &\sim N(\mathbf{X}\beta + \mathbf{Z}\mathbf{u}, \sigma^2 R) \\ \mathbf{u} &\sim N(0, G)\end{aligned}$$

So the likelihood function factors as

$$L(\beta, \sigma^2, G) = f(\mathbf{Y}|\mathbf{u})f(\mathbf{u})$$

So the maximum likelihood approach would choose the values of β , σ^2 , and G (actually the variances that make up G) that maximize $L(\beta, \sigma^2, G)$

One known “problem” with maximum likelihood in this case is that the estimates of the variance components are known to have a negative bias.

Consider the case when there is one variance component σ^2 and β has p components. The the MLE of σ^2 is

$$\tilde{\sigma}^2 = \frac{SSE}{n}$$

However its expectation is

$$E[\tilde{\sigma}^2] = \frac{n-p}{n}\sigma^2 < \sigma^2$$

As we have seen before, the easy solutions here is to use

$$\hat{\sigma}^2 = \frac{SSE}{n-p} = MSE$$

This solution is a special case of a more general method, Restricted Maximum Likelihood (REML).

In REML, we use structure that must occur in the residuals (like $\sum e_i = 0$ in standard regression) to create a correction term in the likelihood and correct for the bias problems.

For example, for known G and R we get

$$\mathbf{X}^T R^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{u}}) = 0$$

and

$$\mathbf{X}^T R^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{u}}) + G^{-1}\hat{\mathbf{u}} = 0$$

From these, we get

$$\hat{\mathbf{u}} = (G^{-1} - \mathbf{Z}^T R^{-1} \mathbf{Z})^{-1} \mathbf{Z}^T R^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

and

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}^T (R + \mathbf{Z}G\mathbf{Z}^T)^{-1} \mathbf{X}]^{-1} \mathbf{X}^T (R + \mathbf{Z}G\mathbf{Z}^T)^{-1} \mathbf{Y}$$

Plugging this into the previous equation gives \hat{u} and the information for the correction term.

The default estimation method for `lmer` is REML (`method="REML"`). It will also fit by maximum likelihood (`method="ML"`).

To see the effect of the two estimation methods, let's examine the orthodontic example from the last class.

```
> orthodont.reml
Linear mixed-effects model fit by REML
Formula: distance ~ Sex + age + (1 | Subject)
      Data: Orthodont
           AIC      BIC    logLik MLdeviance REMLdeviance
445.5125 456.241 -218.7563   434.8982    437.5125
Random effects:
Groups   Name      Variance Std.Dev.
Subject (Intercept) 3.2667   1.8074
Residual                2.0495   1.4316
number of obs: 108, groups: Subject, 27
```

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	15.385690	0.895983	17.1718
SexMale	2.321023	0.761412	3.0483
age	0.660185	0.061606	10.7162

Correlation of Fixed Effects:

	(Intr)	SexMal
SexMale	-0.504	
age	-0.756	0.000

> orthodont.mle

Linear mixed-effects model fit by maximum likelihood

Formula: distance ~ Sex + age + (1 | Subject)

Data: Orthodont

AIC	BIC	logLik	MLdeviance	REMLdeviance
442.8565	453.585	-217.4282	434.8565	437.5526

Random effects:

Groups	Name	Variance	Std.Dev.
Subject	(Intercept)	2.9932	1.7301
Residual		2.0242	1.4227

number of obs: 108, groups: Subject, 27

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	15.385690	0.878447	17.5146
SexMale	2.321023	0.732672	3.1679
age	0.660185	0.061225	10.7830

Correlation of Fixed Effects:

	(Intr)	SexMal
SexMale	-0.494	
age	-0.767	0.000

In this case, both estimated variance are smaller with maximum likelihood, particularly the subject to subject variance. Also note that the estimates of the fixed effects are the same, though their standard errors and correlations are different.