xtmixed & denominator degrees of freedom: myth or magic

2011 Chicago Stata Conference

Phil Ender

UCLA Statistical Consulting Group

July 2011

Terminology

Here are two abbreviations I will be using:

- ddf – Denominator degrees of freedom.
- ddfm – Denominator degrees of freedom method.

Consider this Simple Randomized Block Example

Randomized block design with 16 subjects and 3 treatment levels.

```
. anova y trt id

         Number of obs =       48     R-squared     =  0.7592
         Root MSE      = 3.23265     Adj R-squared =  0.6227

  Source |  Partial SS    df       MS           F     Prob > F
  -------+----------------------------------------------------
   Model |  988.166667    17   58.127451       5.56     0.0000
     trt |  63.1666667     2  31.5833333       3.02     0.0637
      id |         925    15  61.6666667       5.90     0.0000
Residual |       313.5    30       10.45
---------+----------------------------------------------------
   Total |  1301.66667    47  27.6950355
```

Computing the F-ratio

$$
\begin{aligned}
F &= \frac{(SS_n)/(ndf)}{(SS_d)/(ddf)} \\
&= \frac{(63.1666667)/(2)}{(313.5)/(\mathbf{30})} \\
&= \frac{31.583}{10.45} \\
&= 3.02
\end{aligned}
$$

The denominator degrees of freedom is 30.

Same data using xtmixed

```
. xtmixed y i.trt || id:, var

Mixed-effects REML regression            Number of obs      =   48
Group variable: id                       Number of groups   =   16

                                         Obs per group: min =    3
                                                        avg =  3.0
                                                        max =    3

                                             Wald chi2(2)   =   6.04
Log restricted-likelihood = -134.12322   Prob > chi2    = 0.0487
------------------------------------------------------------------
      y |      Coef.   Std. Err.       z    P>|z|      [95% CI]
--------+---------------------------------------------------------
2.trt   |      -.875    1.142913   -0.77    0.444    -3.12   1.37
3.trt   |      -2.75    1.142913   -2.41    0.016    -4.99   -.51
 _cons  |     15.625    1.311541   11.91    0.000    13.05  18.2
------------------------------------------------------------------
```

xtmixed con't

```
------------------------------------------------------------------------------
Random-effects Parameters | Estimate Std. Err.  [95% CI]
--------------------------+---------------------------------------------------
id: Identity              |
               var(_cons) |17.07223 7.559526   7.17  40.66
--------------------------+---------------------------------------------------
            var(Residual) |   10.45 2.698177   6.3   17.33
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) = 16.95
                           Prob >= chibar2 =  0.0000
```

Testing main effect of trt

Omnibus test for treatment.

```
. test 2.trt 3.trt

 ( 1)  [y]2.trt = 0
 ( 2)  [y]3.trt = 0

           chi2(  2) =    6.04
         Prob > chi2 =    0.0487
```

```
Scale chi-square as F-ratio.

. display r(chi2)/r(df)

3.0223293
```

F-ratio from **xtmixed** is the same as the F-ratio from **anova**.

Assumming...

Assuming that the ddf for this simple balanced model is,

$$\begin{aligned} ddf &= obs - df(trt) - df(id) - 1 \\ &= 48 - 2 - 15 - 1 \\ &= 30 \end{aligned}$$

Then, the p-value equals,

Ftail(2, 30, 3.022) = 0.06372709

Comparing p-values

The p-value for the chi-square is 0.0487

The p-value for the anova F-ratio is 0.0637

Chi-square is a large sample normal based statistic, so for small experimental designs we prefer the p-values obtained from the F-distribution. If **xtmixed** provided denominator degrees of freedom this would be a very simple matter.

What's your problem, just use anova. Stop Complaining.

There are many situations that **anova** does not handle well. Here are three examples.

- Incomplete data within subject
- Unequally spaced time intervals
- Level 1 covariance structures other than compound symmetry

UCLA has many researchers working within traditional anova frameworks with relatively small experimental designs. Reviewers and editors of journals in these fields are familiar with experimental designs and with F-ratios.

However, it is common for data to be unbalanced within subject, as is the need for alternative level 1 covariance structures. **Xtmixed** would be ideal for these situations if it could produce probabilities adjusted for smaller samples.

MIssing Observations Example

Consider a modification of our randomized block example with one missing observation for each of four subjects.

Same **xtmixed** command.

```
. xtmixed y i.trt || id:, var
```

xtmixed with missing observations

```
Mixed-effects REML regression           Number of obs      =     44
Group variable: id                      Number of groups   =     16

                                      Obs per group: min   =      2
                                                     avg   =    2.8
                                                     max   =      3

                                           Wald chi2(2)    =     6.51
Log restricted-likelihood = -120.42308  Prob > chi2     =   0.0385
---------------------------------------------------------------
      y |      Coef.  Std. Err.      z    P>|z|      [95% CI]
-------+-------------------------------------------------------
2.trt   | -1.358164  1.133608   -1.20   0.231    -3.58    .86
3.trt   | -2.821488  1.105687   -2.55   0.011    -4.99   -.65
 _cons |  15.37974  1.275609   12.06   0.000    12.88  17.878
---------------------------------------------------------------
```

xtmixed with missing observations – Continued

```
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% CI]
-----------------------------+------------------------------------------------
id: Identity                 |
                  var(_cons) |   16.27504   7.162905     6.87  38.56
-----------------------------+------------------------------------------------
               var(Residual) |   8.929243   2.465748     5.2   15.34
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =   16.66
                            Prob >= chibar2 = 0.0000
```

Testing trt

Omnibus test for main effect for treatment.

```
. test 2.trt 3.trt

 ( 1)  [y]2.trt = 0
 ( 2)  [y]3.trt = 0

           chi2(  2) =    6.51
         Prob > chi2 =    0.0385
```

Scale chi-square as F-ratio.

```
. display r(chi2)/r(df)

3.2572416
```

Comparing p-values

The p-value for the chi-square is 0.0385

The p-value for the F-ratio is Ftail(2, ?, 3.257) $=$?

Even thought the chi-square has been rescaled as an F-ratio, there is no p-value for the F-ratio because we don't know the denominator degrees of freedom.

So, why doesn't xtmixed provide the ddf?

The simple answer:

xtmixed does not know the denominator degrees of freedom. It does not have mean squares or numerators or denominators in the anova sense. And, it does not compute F-ratios at all. **xtmixed** performs statistical tests by dividing parameter estimates by their standard errors.

What can be done?

Since there is no actual denominator degrees of freedom, we need an approximation of an F-distribution that has appropriate control over the Type I Error and has adequate power.

This is not an easy task. There does not seem to be a single F-approximation that works for all possible mixed models. It may be difficult, but it doesn't mean that no one ever tried.

Suvery of Major Stat Packages

Package	Command	ddf method	Philosophy
Stata	xtmixed	none	Statistical
R	lmer	none	Purity
R	lme	containment	Empirical
SPSS	mixed	Satterthwaite	Pragmatism
SAS	proc mixed	Satterthwaite Kenward-Roger* between-within residual containment	

* SAS' current favorite.

Residual, Containment & Between-within ddf

$$\begin{aligned} \textit{Residual df} &= N - \textit{rank}(X) \\ &= 44 - 3 \\ &= 41 \end{aligned}$$

$$\begin{aligned} \textit{Containment df} &= N - \textit{rank}(X, Z) \\ &= 44 - 3 - 15 \\ &= 44 - 18 \\ &= 26 \end{aligned}$$

$$\begin{aligned} \textit{Betwithin df} &= \textit{Residual df} - \textit{rank}(Z) \\ &= 41 - 15 \\ &= 26 \end{aligned}$$

Satterthwaite Approximation

The Satterthwaite approximation is intended as an accurate F-test approximation, and hence accurate p-values for the F-test. SAS does warn that the small-sample properties of the Satterthwaite approximation have not been thoroughly investigated for all models.

Kenward-Roger Approximation

The Kenward-Roger method is an attempt to make a further adjustment to the F-statistic, to take into account the fact that the REML estimates of the covariance parameters are estimates and not known quantities. This method inflates the marginal variance-covariance matrix and then applies the Satterthwaite method on the resulting matrix.

Computational Issues

Residual, containment and between-within methods are fairly simple to compute. However, Satterthwaite and Kenward-Rogers are both computationally and resource intensive.

The computational overhead increases with the complexity of the design and with the complexity of the unbalancedness.

The RB-3 example with missing observations

Various F-approximations with our RB-3 example with 4 missing observations using SAS.

Statistic	Value	ddf	p-value	ddfm
F	3.26	26.7	0.0542	Satterthwaite
F	3.25	26.7	0.0546	Kenward-Roger
F	3.26	26	0.0547	between-within
F	3.26	26	0.0547	contain
F	3.26	41	0.0487	residual
chi2	6.514		0.0385	from Stata

Exceptions in Stata

xtmixed does not provide adjusted ddf's, however **anova** with the **repeated** option will adjust both the numerator and denominator degrees of freedom.

We will return the the original randomized block data, the one without any missing observations and rerun **anova** using **repeated(trt)**.

anova repeated option

```
. anova y trt id, repeated(trt)
...
Between-subjects error term:  id
                     Levels:  16        (15 df)
     Lowest b.s.e. variable:  id

Repeated variable: trt
                        Huynh-Feldt epsilon        =  1.0847
                        *Huynh-Feldt epsilon reset to 1.0000
                        Greenhouse-Geisser epsilon =  0.9505
                        Box's conservative epsilon =  0.5000

                                  ------------ Prob > F ------------
  Source |     df      F    Regular    H-F      G-G      Box
 -----------+------------------------------------------------------
     trt |      2     3.02   0.0637   0.0637   0.0668   0.1026
Residual |     30
```

ddf with repeated option

$$\begin{aligned}
Ftail(2, 30, 3.022) &= 0.06372709 // Regular & (1)\\
Ftail(2 * 1, 30 * 1, 3.022) &= // Huynh - Feldt & (2)\\
Ftail(2, 30, 3.022) &= 0.06372709 &\\
Ftail(2 * .9505, 30 * .9505, 3.022) &= // Greenhouse - Geisser & (3)\\
Ftail(1.901, 28.515, 3.022) &= 0.0668668 &\\
Ftail(2 * .5, 30 * .5, 3.022) &= // Box's\ Conservative & (4)\\
Ftail(1, 15, 3.0222) &= 0.10261965 &
\end{aligned}$$

Use Three-Step Procedure to determine statistical significance.

And, of course, t-test with unequal variances

both Satterthwaite and Welch degrees of freedom for t-tests with unequal variances produce latent ddf.

. ttest y, by(grp)

Two-sample t test with equal variances:
$t = -2.0325$ $df = 38$ p-value $= 0.0491$

Two-sample t test with unequal variances using Satterthwaite's df:
$t = -2.0325$ $df = 26.7921$ p-value $= 0.0521$

Two-sample t test with unequal variances using Welch's df:
$t = -2.0325$ $df = 27.6124$ p-value $= 0.0518$

What can you do short of running SAS?

Consider a split-plot design with **a** between subjects and **b** within subjects and with missing observations within subject:

. xtmixed y a##b || id:

Use the ddf from the following **anova** models with the chi-squares rescaled as F-ratios from **xtmixed**:

Between-within ddf: (two error terms)
. anova y a / id|a b a#b /

Containment ddf: (one error term)
. anova y a id|a b a#b

Residual ddf: (one error term)
. anova y a b a#b

Conclusion – Myth or Magic?

Mostly myth

References

Giesbrecht, F.G. and Burns, J.C. (1985). Two-stage analysis based on a mixed model: Large-sample asymtotic theory and small-sample simulation results. *Biometrics*, *41*, 477-486.

Gould, W. (2009). How are the chi-squared and F distributions related? *Stata FAQ*, www.stata.com/support/faqs/stat/wald.html.

Kenward, M.G. and Roger, J. H. (1997). Small sample Inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*, 983997.

(2009). SAS/STAT 9.2 User's Guide, Second Edition, SAS Institute Inc, Cary, NC.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*, 110-114.