

A Comparison of Denominator Degrees of Freedom Approximation Methods in the Unbalanced Two-Way Factorial Mixed Model

Karl B. Gregory

Texas A & M University

April 12, 2011

Chair:

Dr. Michael SPEED

Co-Chair:

Dr. Michael LONGNECKER

Committee Member:

Dr. Bruce LOWE

The Two-Way Factorial Mixed Model

Model Expression

Tests for Fixed Effects

Denominator Degrees of Freedom Approximation Methods

Purpose of Study

A Rule for Determining The Best ddfm Option

Indices of Unbalancedness

Simulation

Results

Simulation Output

Tables, Tests, and Plots

Conclusions

The Two-Way Factorial Mixed Model

The model is expressed

$$y_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + e_{ijk}$$

where

The Two-Way Factorial Mixed Model

The model is expressed

$$y_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + e_{ijk}$$

where

- ▶ y_{ijk} response

The Two-Way Factorial Mixed Model

The model is expressed

$$y_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + e_{ijk}$$

where

- ▶ y_{ijk} response
- ▶ μ overall mean

The Two-Way Factorial Mixed Model

The model is expressed

$$y_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + e_{ijk}$$

where

- ▶ y_{ijk} response
- ▶ μ overall mean
- ▶ α_i fixed factor effects

The Two-Way Factorial Mixed Model

The model is expressed

$$y_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + e_{ijk}$$

where

- ▶ y_{ijk} response
- ▶ μ overall mean
- ▶ α_i fixed factor effects
- ▶ $b_j \sim N(0, \sigma_b^2)$, $(\alpha b)_{ij} \sim N(0, \sigma_{\alpha b}^2)$, and $e_{ijk} \sim N(0, \sigma_e^2)$

The Two-Way Factorial Mixed Model

The model is expressed

$$y_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + e_{ijk}$$

where

- ▶ y_{ijk} response
- ▶ μ overall mean
- ▶ α_i fixed factor effects
- ▶ $b_j \sim N(0, \sigma_b^2)$, $(\alpha b)_{ij} \sim N(0, \sigma_{\alpha b}^2)$, and $e_{ijk} \sim N(0, \sigma_e^2)$
- ▶ $i = 1, \dots, a$, $j = 1, \dots, b$, and $k = 1, \dots, n_{ij}$

Tests for Fixed Effects

It is generally of interest to test

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$$

Test statistic when $n_{ij} = n$ for all ij :

$$\begin{aligned} F &= MS_A / MS_{AB} \\ &= \frac{\sum_{i=1}^a bn (\bar{y}_{i..} - \bar{y}_{...})^2 / (a-1)}{\sum_{i=1}^a \sum_{j=1}^b n (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 / (a-1)(b-1)} \\ &\sim F_{a-1, (a-1)(b-1)} \end{aligned}$$

Unbalancedness

The design is “unbalanced” when not all n_{ij} s are equal.
In this case

$$F = MS_A / MS_{AB}$$
$$\neq \frac{\sum_{i=1}^a bn (\bar{y}_{i..} - \bar{y}_{...})^2 / (a-1)}{\sum_{i=1}^a \sum_{j=1}^b n (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 / (a-1)(b-1)}$$

And now

$$F \sim F_{a-1, \hat{\nu}}$$

where $\hat{\nu}$, the denominator degrees of freedom, must be chosen carefully.

Denominator Degrees of Freedom Approximation Methods

SAS PROC MIXED offers several **ddfm** options.

This research considers four:

- ▶ Containment (default)
- ▶ Satterthwaite
- ▶ Kenward-Roger
- ▶ Kenward-Roger 1st order

The “best” **ddfm** option is that which best preserves the nominal Type I error rate.

What Is The Best ddfm Option?

- ▶ *Spilke et al.*: RCB, split-plot, and strip-plot designs under small sample sizes. Imposed minor unbalancedness. Favored Kenward-Roger.[1]

What Is The Best ddfm Option?

- ▶ *Spilke et al.*: RCB, split-plot, and strip-plot designs under small sample sizes. Imposed minor unbalancedness. Favored **Kenward-Roger**.^[1]
- ▶ *Waseem Alnosair*: Partially balanced incomplete block, BIBD, and RCB designs. Favored **Kenward-Roger** as well as his own modified Kenward-Roger over **Satterthwaite** and **containment**.^[2]

What Is The Best ddfm Option?

- ▶ *Spilke et al.*: RCB, split-plot, and strip-plot designs under small sample sizes. Imposed minor unbalancedness. Favored **Kenward-Roger**.^[1]
- ▶ *Waseem Alnosair*: Partially balanced incomplete block, BIBD, and RCB designs. Favored **Kenward-Roger** as well as his own modified Kenward-Roger over **Satterthwaite** and **containment**.^[2]
- ▶ *Schaalje et al.*: Split plot and repeated measures designs under various covariance structures. Favored **Kenward-Roger** over **Satterthwaite**. Identify the following as affecting **ddfm option** performance:

What Is The Best ddfm Option?

- ▶ *Spilke et al.*: RCB, split-plot, and strip-plot designs under small sample sizes. Imposed minor unbalancedness. Favored **Kenward-Roger**. [1]
- ▶ *Waseem Alnosair*: Partially balanced incomplete block, BIBD, and RCB designs. Favored **Kenward-Roger** as well as his own modified Kenward-Roger over **Satterthwaite** and **containment**. [2]
- ▶ *Schaalje et al.*: Split plot and repeated measures designs under various covariance structures. Favored **Kenward-Roger** over **Satterthwaite**. Identify the following as affecting **ddfm option** performance:
 1. covariance structure
 2. sample size
 3. unbalancedness [3]

A Rule for Determining The Best ddfm Option

According to these studies, the relative performance of the **ddfm options** is determined by

- ▶ the covariance structure,
- ▶ the degree of unbalancedness,

and the differences will be greater when sample sizes are small.

This research thus seeks a rule in the small sample setting for determining the best **ddfm option** based on the degree of unbalancedness in the design.

How to measure unbalancedness?

The Matrix of Cell Counts

Consider defining a function of the matrix of replication values that serves as an index of unbalancedness.

Let

$$\mathbf{N} = \begin{pmatrix} n_{11} & n_{12} & \dots & n_{1b} \\ n_{21} & n_{22} & \dots & n_{2b} \\ \vdots & \vdots & \ddots & \vdots \\ n_{a1} & n_{a2} & \dots & n_{ab} \end{pmatrix}$$

Shannon's Entropy

This is often used to measure species diversity in an ecosystem.[4]

Let

- ▶ N be the total number of observations,
- ▶ n_{ij} be the number of replicates in the ij^{th} treatment group,
- ▶ and let $p_{ij} = n_{ij}/N$

Then

$$H = - \sum_{i=1}^a \sum_{j=1}^b p_{ij} \log(p_{ij})$$

Higher values of H correspond to greater balancedness in the design, that is, to more uniform dispersion of the N observations across the ab treatment combinations.

Simpson's Diversity

This is also a common measure of species diversity.[5]

$$D = 1 - \frac{\sum_{i=1}^a \sum_{j=1}^b n_{ij}(n_{ij} - 1)}{N(N - 1)}$$

Likewise, larger values of D correspond to greater balancedness in the design.

ImbB (Imbalance Across Factor B)

Conceivably, the orientation of the unbalancedness with respect to the fixed and random factors has an effect on **ddfm option** performance.

This index treats each of the n_{ij} s as a response in a 2-way factorial design with a single replication. The index is defined as SSB/SST :

$$ImbB = \sum_{j=1}^b (\bar{n}_{.j} - \bar{n}_{..})^2 / \sum_{i=1}^a \sum_{j=1}^b (n_{ij} - \bar{n}_{..})^2$$

Goal of Simulation

Questions:

- ▶ Does any of the three indices provide a means of determining the best **ddfm option**?
- ▶ That is, does any of the **ddfm options** perform better than others at different values of an unbalancedness index?

E.g.: Does **Satterthwaite** outperform **Kenward-Roger** at high values of Shannon's Entropy but not at low values?

Nuts and Bolts

- ▶ For each of the four **ddfm** options,
 - ▶ Generate 10,000 datasets such that $n_{ij} \sim \frac{1}{5}I_{\{1,2,3,4,5\}}(n_{ij})$.

Nuts and Bolts

- ▶ For each of the four **ddfm** options,
 - ▶ Generate 10,000 datasets such that $n_{ij} \sim \frac{1}{5}I_{\{1,2,3,4,5\}}(n_{ij})$.
 - ▶ On each dataset, compute Shannon's, Simpson's, and ImbB indices.

Nuts and Bolts

- ▶ For each of the four **ddfm** options,
 - ▶ Generate 10,000 datasets such that $n_{ij} \sim \frac{1}{5}I_{\{1,2,3,4,5\}}(n_{ij})$.
 - ▶ On each dataset, compute Shannon's, Simpson's, and ImbB indices.
 - ▶ Test for significance of fixed effects on each dataset.

Nuts and Bolts

- ▶ For each of the four **ddfm options**,
 - ▶ Generate 10,000 datasets such that $n_{ij} \sim \frac{1}{5}I_{\{1,2,3,4,5\}}(n_{ij})$.
 - ▶ On each dataset, compute Shannon's, Simpson's, and ImbB indices.
 - ▶ Test for significance of fixed effects on each dataset.
- ▶ For each of the three indices
 - ▶ Pool the four sets of 10,000 datasets together, but then separate all 40,000 datasets into bins at deciles of the index values.

Nuts and Bolts

- ▶ For each of the four **ddfm** options,
 - ▶ Generate 10,000 datasets such that $n_{ij} \sim \frac{1}{5}I_{\{1,2,3,4,5\}}(n_{ij})$.
 - ▶ On each dataset, compute Shannon's, Simpson's, and ImbB indices.
 - ▶ Test for significance of fixed effects on each dataset.
- ▶ For each of the three indices
 - ▶ Pool the four sets of 10,000 datasets together, but then separate all 40,000 datasets into bins at deciles of the index values.
 - ▶ This results in 10 bins, each containing about 4,000 datasets, of which there are approximately 1,000 tested under each **ddfm** option.

Simulation

In sum. . .

The simulation produces 40,000 datasets, each having three properties of interest:

Simulation

In sum. . .

The simulation produces 40,000 datasets, each having three properties of interest:

- ▶ The **ddfm option** under which significance of fixed effects were tested

Simulation

In sum. . .

The simulation produces 40,000 datasets, each having three properties of interest:

- ▶ The **ddfm option** under which significance of fixed effects were tested
- ▶ The bin into which it fell for each of the three indices

Simulation

In sum. . .

The simulation produces 40,000 datasets, each having three properties of interest:

- ▶ The **ddfm option** under which significance of fixed effects were tested
- ▶ The bin into which it fell for each of the three indices
- ▶ The outcome of the test for fixed effects—correct or incorrect decision.

Example of Simulation Output

Shannon Index	Bin	1		2		...	10	
	Outcome	FP	TN	FP	TN	...	FP	TN
	Con	49	896	65	879	...	51	958
	Satterth	57	988	49	959	...	54	952
	KR	49	950	60	947	...	54	909
	KR 1st	39	933	44	935	...	47	934
Simpson Index	Bin	1		2		...	10	
	Outcome	FP	TN	FP	TN	...	FP	TN
	Con	51	881	59	870	...	49	973
	Satterth	58	988	49	982	...	53	969
	KR	50	979	69	947	...	52	997
	KR 1st	43	938	38	957	...	47	930
ImbB Index	Bin	1		2		...	10	
	Outcome	FP	TN	FP	TN	...	FP	TN
	Con	46	914	45	921	...	61	940
	Satterth	52	949	52	963	...	48	901
	KR	53	973	46	946	...	51	970
	KR 1st	62	940	39	990	...	52	983

Analyzing Output

Shannon Index	Bin	1		2		...	10	
	Outcome	FP	TN	FP	TN	...	FP	TN
	Con	49	896	65	879	...	51	958
	Satterth	57	988	49	959	...	54	952
	KR	49	950	60	947	...	54	909
	KR 1st	39	933	44	935	...	47	934

Analyzing Output

Shannon Index	Bin	1		2		...	10	
	Outcome	FP	TN	FP	TN	...	FP	TN
	Con	49	896	65	879	...	51	958
	Satterth	57	988	49	959	...	54	952
	KR	49	950	60	947	...	54	909
	KR 1st	39	933	44	935	...	47	934

- In any bin, does it matter which **ddfm option** is used?

Analyzing Output

Shannon Index	Bin	1		2		...	10	
	Outcome	FP	TN	FP	TN	...	FP	TN
	Con	49	896	65	879	...	51	958
	Satterth	57	988	49	959	...	54	952
	KR	49	950	60	947	...	54	909
	KR 1st	39	933	44	935	...	47	934

- In any bin, does it matter which **ddfm option** is used?
- I.e., is there an association between outcome and **ddfm option** in any of the bins?

Analyzing Output

Shannon Index	Bin	1		2		...	10	
	Outcome	FP	TN	FP	TN	...	FP	TN
	Con	49	896	65	879	...	51	958
	Satterth	57	988	49	959	...	54	952
	KR	49	950	60	947	...	54	909
	KR 1st	39	933	44	935	...	47	934

- ▶ In any bin, does it matter which **ddfm option** is used?
- ▶ I.e., is there an association between outcome and **ddfm option** in any of the bins?
- ▶ *Enter Cochran, Mantel, and Haenzel.*

The Cochran-Mantel-Haenzel (CMH) Test

Hypotheses of the CMH test:

$H_0 : X, Y$ are independent given Z

$H_1 : X, Y$ are not independent given Z

- ▶ X = outcome
- ▶ Y = **ddfm option**
- ▶ Z = index bin

If H_0 is rejected, then outcome and **ddfm option** are not independent conditional on the index bin.

Simulation Settings

The described simulation was carried out under 6 different settings:

Simulation Settings

The described simulation was carried out under 6 different settings:

- ▶ H_0 True. $\alpha_1 = \alpha_2 = \alpha_3 = 0$:

Simulation Settings

The described simulation was carried out under 6 different settings:

- ▶ H_0 True. $\alpha_1 = \alpha_2 = \alpha_3 = 0$:
 - ▶ Independent error terms.
 - ▶ AR(1) error covariance structure—*unspecified* in PROC MIXED
 - ▶ AR(1) error covariance structure—*specified* in PROC MIXED

Simulation Settings

The described simulation was carried out under 6 different settings:

- ▶ H_0 True. $\alpha_1 = \alpha_2 = \alpha_3 = 0$:
 - ▶ Independent error terms.
 - ▶ AR(1) error covariance structure—*unspecified* in PROC MIXED
 - ▶ AR(1) error covariance structure—*specified* in PROC MIXED
- ▶ H_0 False. $\alpha_1 = -2.5$, $\alpha_2 = 0$, $\alpha_3 = 2.5$:

Simulation Settings

The described simulation was carried out under 6 different settings:

- ▶ H_0 True. $\alpha_1 = \alpha_2 = \alpha_3 = 0$:
 - ▶ Independent error terms.
 - ▶ AR(1) error covariance structure—*unspecified* in PROC MIXED
 - ▶ AR(1) error covariance structure—*specified* in PROC MIXED
- ▶ H_0 False. $\alpha_1 = -2.5$, $\alpha_2 = 0$, $\alpha_3 = 2.5$:
 - ▶ Independent error terms.
 - ▶ AR(1) error covariance structure—*unspecified* in PROC MIXED
 - ▶ AR(1) error covariance structure—*specified* in PROC MIXED

Simulation Settings

The described simulation was carried out under 6 different settings:

- ▶ H_0 True. $\alpha_1 = \alpha_2 = \alpha_3 = 0$:
 - ▶ Independent error terms.
 - ▶ AR(1) error covariance structure—*unspecified* in PROC MIXED
 - ▶ AR(1) error covariance structure—*specified* in PROC MIXED
- ▶ H_0 False. $\alpha_1 = -2.5$, $\alpha_2 = 0$, $\alpha_3 = 2.5$:
 - ▶ Independent error terms.
 - ▶ AR(1) error covariance structure—*unspecified* in PROC MIXED
 - ▶ AR(1) error covariance structure—*specified* in PROC MIXED
- ▶ For all simulations, $a = 3$, $b = 5$, $\mu = 10$, $\sigma_b^2 = 9$, $\sigma_{\alpha b}^2 = 4$, $\sigma_e^2 = 1$, and $\rho = .5$ when applicable.

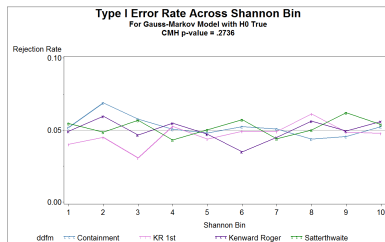
The Cochran-Mantel-Haenzel (CMH) Test

CMH test results for H_0 -true simulations:

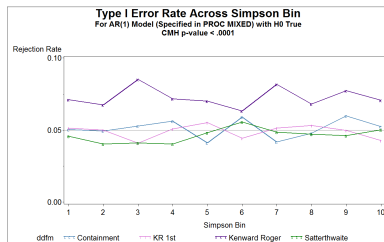
Situation	Index	CMH p-value
Gauss-Markov model with H_0 true	Shannon	.2736
	Simpson	.2831
	lmbB	.2730
AR(1) model with H_0 true (<i>AR(1) ignored in PROC MIXED</i>)	Shannon	.8661
	Simpson	.8652
	lmbB	.8713
AR(1) model with H_0 true (<i>AR(1) specified in PROC MIXED</i>)	Shannon	< .0001
	Simpson	< .0001
	lmbB	< .0001

Plotted Output (H_0 True)

H_0 rejection rate (type I error rate) across the bins of the index.



CMH p-value = .2736.



CMH p-value < .0001.

Note that **Kenward-Roger** has a higher rejection rate than the others when the AR(1) model is specified in PROC MIXED.

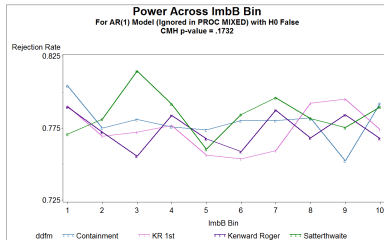
The Cochran-Mantel-Haenzel (CMH) Test

CMH test results for H_0 -false simulations:

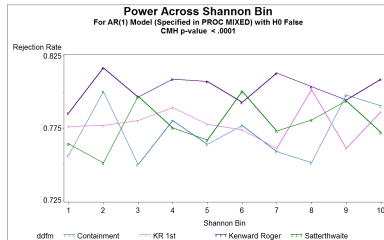
Situation	Index	CMH p-value
Gauss-Markov model with H_0 false	Shannon	.2892
	Simpson	.2946
	lmbB	.3055
AR(1) model with H_0 false (<i>AR(1) ignored in PROC MIXED</i>)	Shannon	.1772
	Simpson	.1762
	lmbB	.1732
AR(1) model with H_0 false (<i>AR(1) specified in PROC MIXED</i>)	Shannon	< .0001
	Simpson	< .0001
	lmbB	< .0001

Plotted Output (H_0 False)

H_0 rejection rate (power) across the bins of the index.



CMH p-value = .1732.



CMH p-value < .0001.

The **Kenward-Roger** method exhibits the same tendency.

Conclusions

1. Apart from the AR(1)-specified case, no significant associations between outcome and **ddfm option**, conditional on index bin, were found.

Conclusions

1. Apart from the AR(1)-specified case, no significant associations between outcome and **ddfm option**, conditional on index bin, were found.
2. In the AR(1)-specified case, the **Kenward-Roger** method tended to inflate the rejection rate for tests of fixed effects.

"When the covariance structure is complex, such as ANTE(1) and the sample size is small, the KR method produces inflated Type I error rates."

–SAS PROC MIXED documentation [6]

Conclusions

1. Apart from the AR(1)-specified case, no significant associations between outcome and **ddfm option**, conditional on index bin, were found.
2. In the AR(1)-specified case, the **Kenward-Roger** method tended to inflate the rejection rate for tests of fixed effects.

"When the covariance structure is complex, such as ANTE(1) and the sample size is small, the KR method produces inflated Type I error rates."

–SAS PROC MIXED documentation [6]

3. No rule for determining the best **ddfm option** was found.

Conclusions

1. Apart from the AR(1)-specified case, no significant associations between outcome and **ddfm option**, conditional on index bin, were found.
2. In the AR(1)-specified case, the **Kenward-Roger** method tended to inflate the rejection rate for tests of fixed effects.

"When the covariance structure is complex, such as ANTE(1) and the sample size is small, the KR method produces inflated Type I error rates."

–SAS PROC MIXED documentation [6]

3. No rule for determining the best **ddfm option** was found.
4. No evidence of **containment**'s inferiority to the other methods was found.

Alternate Approach

Rather than using the CMH test to test for conditional independence between `ddfm option` and outcome given the index bin, do a separate chi-square test of independence at each level of the bin. There would be 10 simultaneous tests of independence, so set the significance level to $.05/10 = .005$. Then, within each bin, you would be testing for association, rather than for a conditional association.

Future Research

- ▶ Increase the number of simulations.
- ▶ Try different numbers of bins.
- ▶ Consider other indices of unbalancedness.

References



Joachim Spilke, Hans-Peter Piepho, and Xiyuan Hu.

A simulation study on tests of hypotheses and confidence intervals for fixed effects in mixed models for blocked experiments with missing data.

Journal of Agricultural, Biological and Environmental Statistics, 10(3):374–389, Sep 2005.



Waseem S. Alnosaier.

Kenward-Roger Approximate F Test for Fixed Effects in Mixed Linear Models.

PhD thesis, Oregon State University, Corvallis, Oregon, 2007.



G. Bruce Schaalje, Justin B. McBride, and Gilbert W. Fellingham.

Adequacy of approximations to distributions of test statistics in complex mixed linear models.

Journal of Agricultural, Biological, and Environmental Statistics, 7(4):512–524, Dec 2002.



Robert K. Peet.

Relative diversity indices.

Ecology, 56(2):496–498, Early Spring 1975.



E. H. Simpson.

Measurement of diversity.

Nature, 163:688–688, April 1949.



Jill Tao and Catherine Truxillo.

Mixed Models Analyses Using SAS.

Acknowledgements

Dr. Michael SPEED
Dr. Michael LONGNECKER
Dr. Bruce LOWE