STATISTICS
TEXAS A&M UNIVERSITY

Handout 17

Additional Topics: Statistical Analysis with MIXED, GLIMMIX, GENMOD Procedures

Repeated Measures with Discrete Response (resp1 data) Using PROC GENMOD: Code

```
proc genmod data=resp desc;
   class id center sex treatment visit;
   model outcome = sex|treatment visit age
                   baseline center
                 / link=logit dist=binomial;
   repeated subject=id(center) /
                    type=ar(1) corrw;
run;
```

Repeated Measures with Discrete Response
Using PROC GENMOD: Partial Output

Working Correlation Matrix

	Col1	Col2	Col3	Col4
Row1	1.0000	0.3890	0.1513	0.0588
Row2	0.3890	1.0000	0.3890	0.1513
Row3	0.1513	0.3890	1.0000	0.3890
Row4	0.0588	0.1513	0.3890	1.0000

STATISTICS
TEXAS A&M UNIVERSITY

Using PROC GENMOD: Partial Output

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		-0.5020	0.5803	-1.6393	0.6354	-0.87	0.3870
sex	F	-0.5530	0.6301	-1.7880	0.6821	-0.88	0.3802
sex	M	0.0000	0.0000	0.0000	0.0000	.	.
treatment	A	0.8950	0.3838	0.1428	1.6471	2.33	0.0197
treatment	P	0.0000	0.0000	0.0000	0.0000	.	.
sex*treatment	F A	1.9076	1.0385	-0.1278	3.9431	1.84	0.0662
sex*treatment	F P	0.0000	0.0000	0.0000	0.0000	.	.
sex*treatment	M A	0.0000	0.0000	0.0000	0.0000	.	.
sex*treatment	M P	0.0000	0.0000	0.0000	0.0000	.	.
visit	1	0.3516	0.2564	-0.1510	0.8542	1.37	0.1703
visit	2	0.0976	0.2541	-0.4003	0.5956	0.38	0.7007
visit	3	0.3498	0.2390	-0.1187	0.8183	1.46	0.1434
visit	4	0.0000	0.0000	0.0000	0.0000	.	.
age		-0.0160	0.0131	-0.0416	0.0096	-1.22	0.2211
baseline		1.8630	0.3470	1.1829	2.5431	5.37	<.0001
center	1	-0.5950	0.4542	-1.4853	0.2952	-1.31	0.1902
center	2	-0.4154	0.4820	-1.3602	0.5294	-0.86	0.3888
center	3	0.5920	0.5858	-0.5561	1.7401	1.01	0.3122
center	4	0.0000	0.0000	0.0000	0.0000	.	.

Repeated Measures with Discrete Response
PROC GLIMMIX: Code

```
proc glimmix data=resp empirical;
   class id center sex treatment visit;
   model outcome (event='1') =
         sex|treatment visit age baseline
         center / link=logit dist=binary
                  solution;
   random _residual_ / subject=id(center)
                       type=ar(1);
run;
```

PROC GLIMMIX: Partial Output

Fit Statistics

-2 Res Log Pseudo-Likelihood	2017.45
Generalized Chi-Square	451.49
Gener. Chi-Square / DF	1.05

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
AR(1)	id(center)	0.3965	0.04897
Residual		1.0451	0.08064

PROC GLIMMIX: Partial Output

Solutions for Fixed Effects

Effect	sex	treatment	center	visit	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept					-0.5026	0.5801	102	-0.87	0.3884
sex	F				-0.5543	0.6301	102	-0.88	0.3811
sex	M				0	.	.	.	.
treatment		A			0.8927	0.3839	102	2.33	0.0220
treatment		P			0	.	.	.	.
sex*treatment	F	A			1.9060	1.0385	102	1.84	0.0694
sex*treatment	F	P			0	.	.	.	.
sex*treatment	M	A			0	.	.	.	.
sex*treatment	M	P			0	.	.	.	.
visit				1	0.3515	0.2564	330	1.37	0.1713
visit				2	0.09754	0.2540	330	0.38	0.7013
visit				3	0.3497	0.2390	330	1.46	0.1444
visit				4	0	.	.	.	.
age					-0.01592	0.01306	102	-1.22	0.2256
baseline					1.8621	0.3470	102	5.37	<.0001
center			1		-0.5943	0.4544	102	-1.31	0.1939
center			2		-0.4175	0.4821	102	-0.87	0.3885
center			3		0.5943	0.5856	102	1.01	0.3126
center			4		0	.	.	.	.

Automatic Variables

Automatic variables can be used to specify your own link function and/or variance function. They include the following:

- _LINP_ linear predictor
- _MU_ expresses the mean of an observation as a function of the linear predictor
- _VARIANCE_ estimate of the variance function
- _XBETA_ equals $\mathbf{X}\hat{\beta}$
- _ZGAMMA_ equals $\mathbf{Z}\hat{\gamma}$
- _N_ the observation number in the sequence of the data read.

Automatic Variables in PROC GLIMMIX

```
proc glimmix data=yourdata;
   class A B;
   model y/n = A B / dist=binomial link=logit;
run;
```

```
proc glimmix data=yourdata;
   class A B;
   prob = 1 / (1+exp(-_linp_));
   _mu_ = n * prob ;
   _variance_ = n * prob *(1-prob);
   model y = A B;
run;
```

Modeling Multivariate Responses

(herniodata.sas)

Patient: patient ID **Age:** patient age **Gender:** patient gender

Leave_status: patient's condition upon leaving the operation room (1: experience routine recovery, 0:required post operative intensive care). Binary data

LOS: length of hospital stay after the operation (days). Poisson Data.

patient	age	gender	leave_status	LOS
1	78	m	0	7
2	60	m	0	2
3	68	m	1	5
4	62	m	1	20
...				

Use the DIST=BYOBS() option in the MODEL statement in PROC GLIMMIX.

Modeling Multivariate Responses

First, convert your data into the following format:

patient	age	gender	y	variable	distrib
1	78	m	0	leave_status	Binary
1	78	m	7	LOS	Poisson
2	60	m	0	leave_status	Binary
2	60	m	2	LOS	Poisson
3	68	m	1	leave_status	Binary
3	68	m	5	LOS	Poisson
4	62	m	1	leave_status	Binary
4	62	m	20	LOS	Poisson
...					

Univariate Analysis for Leave_status

```
proc glimmix data=new(where=(distrib="Binary"));
   model y (event='1') = age
                       / s dist=binary link=logit;
run;
```

Parameter Estimates

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	5.2154	2.2641	30	2.30	0.0284
age	-0.07258	0.03449	30	-2.10	0.0438

Univariate Analysis for LOS

```
proc glimmix data=new
                    (where=(distrib="Poisson"));
   model y = age / s dist=Poisson link=log;
run;
```

Parameter Estimates

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	-0.1962	0.3797	30	-0.52	0.6093
age	0.02739	0.005766	30	4.75	<.0001

Multivariate Analysis Assuming Independence

```
proc glimmix data=new;
   class dist;
   model y(event='1') = distrib distrib*age
         / noint s dist=byobs(distrib);
run;
```

Parameter Estimates

Effect	distrib	Estimate	Standard Error	DF	t Value	Pr > |t|
distrib	Binary	5.2153	2.2640	60	2.30	0.0247
distrib	Poisson	-0.1962	0.3797	60	-0.52	0.6074
age*distrib	Binary	-0.07258	0.03449	60	-2.10	0.0395
age*distrib	Poisson	0.02739	0.005766	60	4.75	<.0001

Multivariate Analysis with Correlation

Modeling correlations through G-side random effects:

```
proc glimmix data=new;
   class patient distrib;
   model y(event='1') = distrib distrib*age
       / noint s dist=byobs(distrib);
   random int / subject=patient;
run;
```

Multivariate Analysis with Correlation

Partial output from the G-side random effect model:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Intercept	patient	0.3419	0.1397

Solutions for Fixed Effects

Effect	distrib	Estimate	Standard Error	DF	t Value	Pr > |t|
distrib	Binary	5.1482	2.2688	30	2.27	0.0306
distrib	Poisson	-0.4323	0.5538	30	-0.78	0.4412
age*distrib	Binary	-0.07141	0.03490	30	-2.05	0.0496
age*distrib	Poisson	0.02941	0.008782	30	3.35	0.0022

Multivariate Analysis with Correlation

```
proc glimmix data=new;
   class patient distrib;
   model y(event='1') = distrib distrib*age
                      / noint s dist=byobs(distrib);
   random _residual_ / subject=patient
                       type=chol;
run;
```

Multivariate Analysis with Correlation

Partial output from the R-side random effect model:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
CHOL(1,1)	patient	1.0075	0.1301
CHOL(2,1)	patient	0.2550	0.3336
CHOL(2,2)	patient	1.8201	0.2350

Solutions for Fixed Effects

Effect	distrib	Estimate	Standard Error	DF	t Value	Pr > |t|
distrib	Binary	5.1895	2.2695	28	2.29	0.0300
distrib	Poisson	-0.2006	0.6979	28	-0.29	0.7759
age*distrib	Binary	-0.07232	0.03460	28	-2.09	0.0458
age*distrib	Poisson	0.02746	0.01060	28	2.59	0.0150

Parameter Estimation Methods in PROC GLIMMIX

- For generalized linear models (GzLM)
  - The data is never correlated
  - No G side random effect
  - Normal distribution: restricted maximum likelihood
  - All other known distributions: maximum likelihood
  - Unknown distributions: quasi-likelihood

- For generalized linear models with overdispersion
  - Parameters are estimated using maximum likelihood.
  - An overdispersion parameter can be estimated from the Pearson statistic. Need to include random _residual_to estimate the overdispersion in the model.

- For generalized linear mixed models (GzLMM)
  - Might have random effect
  - Might have correlated data. This will change the R matrix
  - Pseudo-likelihood (default)
  - Likelihood-based (Quadrature or Laplace)

Pseudo-Likelihood versus Maximum-Likelihood

The challenge in fitting GzLMMs is how to obtain the marginal log likelihood function:

$$\int p(y \mid x, \beta, \gamma) q(\gamma) d\gamma$$

difficult to have closed-form solutions

Compute the marginal log-likelihood for a different but similar models (LMM), whose marginal log-likelihood has a closed-form solution

Pseudo-likelihood (Linearization)

Approximate the integral numerically to obtain the likelihood function

Maximum-likelihood with quadrature or Laplace approximation

Pseudo-Likelihood Method

There are four pseudo-likelihood estimation techniques: RSPL (the default), MSPL, RMPL, and MMPL.

- RSPL is based on the residual likelihood, so it should have less bias than MSPL, which is based on maximum likelihood.
- RSPL is also based on the BLUPs, so it has a subject-specific interpretation.
- For linear mixed models, RSPL=REML and MSPL=ML.

First letter: M is for ML and R is for Residual likelihood
Second Letter: S is for vector of random effect solutions and M for mean of random effects.

Pseudo-Likelihood (Linearization) Method

The pseudo-likelihood method

- can fit complex models.
- is for an approximated linear mixed model. There is no true likelihood, so there are no likelihood ratio tests.
- might have biased variance estimates for random effects, especially for binary outcome and few clusters or small number of observations per cluster.

Maximum Likelihood Method

Use numerical technique to approximate the integral to obtain the marginal log-likelihood.

- METHOD=QUAD (Gauss-Hermite quadrature)
- METHOD=LAPLACE (Laplace approximation)
  - More computationally efficient
  - Possible for unbounded covariance parameter estimate

Adaptive Gaussian Quadrature

The term *quadrature* is more or less a synonym for *numerical integration*, especially as applied to one-dimensional integrals. Two-dimensional integration is sometimes described as *cubature*, although this term is much less frequently used and the meaning of quadrature is understood for higher dimensional integration as well.

The basic problem considered by numerical integration is to compute an approximate solution to a definite integral $\int_a^b f(x)dx$. If $f(x)$ is a smooth, well-behaved function and the limits of integration are fixed, the method of Gaussian quadrature is very effective. An n-point **Gaussian quadrature rule**, named after Carl Friedrich Gauss, is a quadrature rule constructed to yield an exact result for polynomials of degree $2n-1$, by a suitable choice of the n points x_i and n weights w_i. The domain of integration for such a rule is conventionally taken as [-1, 1], so the rule is stated as shown below:

$$\int_{-1}^{1} f(x)dx \approx \sum_{i=1}^{n} w_i f(x_i)$$

It can be shown (Stoer and Bulirsch 1980) that the evaluation points are only the roots of a polynomial belonging to a class of orthogonal polynomials.

The adaptive algorithm calculates the definite integral of a function adaptively by choosing smaller steps near problematic points.

Maximum Likelihood (ML) Method

The maximum likelihood method

- computes the log-likelihood of the data so model comparisons are possible based on information criteria
- avoids the pseudo-likelihood bias
- cannot be implemented for models with the R-side random effects.
- presents some limitations for the G-side random effects for the quadrature method
  - G-side random effect must be processed by subjects (use the SUBJECT= option)
  - Nested subjects can only be accommodated with limitations.

ML in PROC GLIMMIX versus ML in PROC NLMIXED

- PROC GLIMMIX determines starting values.
  - Fixed effects: GLM
  - Covariance parameters: MIVQUE(0) step
  - Can run PL iterations prior to quadrature
- PROC NLMIXED assumes a starting value of 1.
- PROC GLIMMIX has a dedicated algorithm for METHOD=LAPLACE.
  - Larger class of models
  - NOBOUND option
- PROC NLMIXED treats LAPLACE as a special case of quadrature (QPOINTS=1).

Question

The maximum likelihood estimation method in the GLIMMIX procedure enables you to

a. Fit any type of generalized linear mixed models using this estimation method.
b. Perform likelihood ratio tests for nested models.
c. Use PROC GLIMMIX rather than PROC NLMIXED for all models that can be fit in PROC NLMIXED.
d. Model R-side random effects.

Processing by Subjects

random A*B;

$\Leftrightarrow$ random intercept / subject=A*B;

$\Leftrightarrow$ random A / subject=B;

$\Leftrightarrow$ random B / subject=A;

processing by subject, which is more numerically efficient

If there are multiple subject effects (in multiple RANDOM statements) in one PROC GLIMMIX, then processing by subjects is possible if the effects are equal or contained in each other.

```
random int / subject=A;
random int / subject=A*B;
```

Processing By Subjects

The following specifications are not processed by subjects:

```
Other PROC GLIMMIX statements;
random A;
random B / subject=A;
```

or

```
Other PROC GLIMMIX statements;
random int / subject=A;
random A / subject=B;
```

You need to write the following to process by subjects

```
random int / subject=A;
random B / subject=A;
```

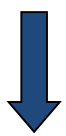
or

```
random int / subject=A;
random int / subject=A*B;
```

Equivalent Subject Variable

Nesting or crossing of interaction effects in subject effects is equivalent.

```
class A B;
model y=;
random intercept / subject=A*B;
```



is equivalent to

```
class A B;
model y=;
random intercept / subject=B(A);
```

Additional Information about Linear Mixed Models

- Unbalanced Data Issues
- Estimation of Covariance Parameters
- Degrees of Freedom Estimates

Objectives of Unbalanced Data Issues

– Issues associated with unbalanced data in linear mixed models. Balanced vs Unbalanced:

	I	II
A	6	6
B	6	6
C	6	6

	I	II
A	8	4
B	6	7
C	2	5

– Issues associated with data with empty cells.

Clinic	Trt A	Trt B
41	0	16
42	2	1
43	14	14
…	…	…

Issues Associated with Unbalanced Data

- In general, there is not a problem for one-way mean models.
- For two-way or higher mean models, the critical issue is constructing meaningful linear combinations of model parameters for hypothesis testing and estimation.
- In mixed model situations, the additional issue for unbalanced data analysis is to account for multiple sources of random variations.

Using PROC MIXED for Unbalanced Data

- You need to determine which type of tests (Type I, Type II, or Type III) to use for fixed effects.
- The procedure takes into account the multiple sources of random variations for inferences about the fixed effects in unbalanced data.
- The procedure provides various methods to estimate the denominator degrees of freedom for fixed effects.

Different Types of Tests, Reduction Notation

Model: $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$, $i = 1, 2$, $j = 1$ to 5.

```
model y = A B A*B;
```

Effects	Type I	Type II	Type III
A	$R(\alpha\|\mu)$	$R(\alpha\|\ \mu, \beta)$	$R(\alpha\|\ \mu, \beta, \alpha\beta)$
B	$R(\beta\|\mu, \alpha)$	$R(\beta\|\ \mu, \alpha)$	$R(\beta\|\ \mu, \alpha, \alpha\beta)$
A*B	$R(\alpha\beta\|\ \mu, \alpha, \beta)$	$R(\alpha\beta\|\ \mu, \alpha, \beta)$	$R(\alpha\beta\|\ \mu, \alpha, \beta)$
	Model-order dependent	**Model-order independent**	**Model-order independent**

Which Type to Use?

- Type I is useful for nested models, polynomial models, and certain tests in analysis of covariance models.
- Type II is useful for models with no significant interactions.
- Type III is useful for comparing main effects in the presence of interactions and situations where unequal sample sizes were not obtained by design.

A Multicenter Clinical Trial

A pharmaceutical company compared the effects of two drugs, A and B, on a clinical measurement called flush.

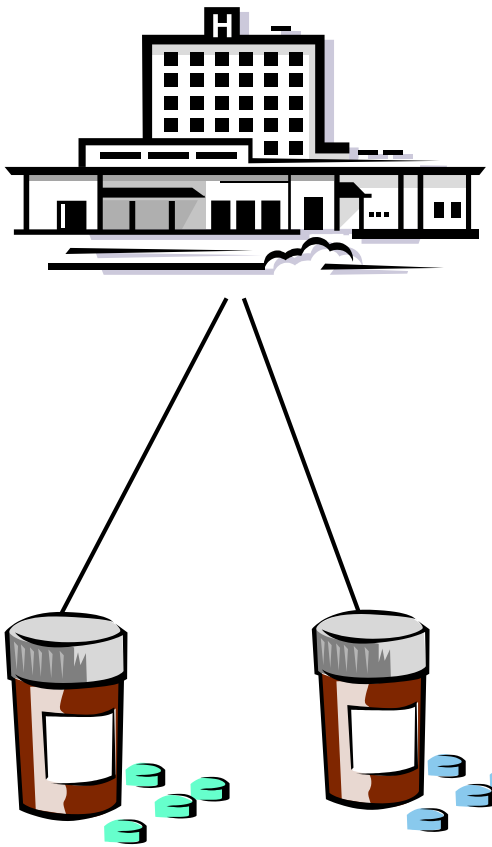
The study utilized patients in 10 clinics.

Multiple clinics were used in order to obtain representation of diverse patient populations.

The original plan called for each drug to be randomly assigned to the same number of patients within each clinic. This was basically done, but a few patients abandoned the trial before completion, leaving unequal numbers of patients on the two drugs within some of the clinics.

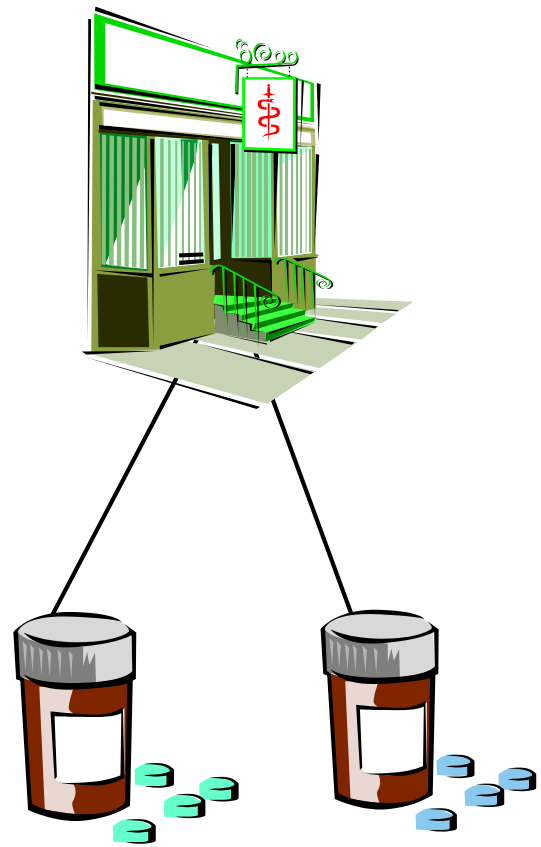
In addition, the availability of patients varied between the clinics, ranging from 3 to 28. Therefore, although this was a designed experiment, the realities of the situation resulted in unbalanced data.

A Multicenter Clinical Trial



Clinic

Treatment



The Data

Clinic: the clinics where the clinical trial was carried out
Trt: the treatment, drug A or B
Patient: patient identification
Flush0: the flush measurements obtained before administering the drugs
Flush: the flush measurements after administering the drugs

Obs	Clinic	Trt	Patient	Flush0	Flush
1	41	B	102	77.5	72.0000
2	41	B	104	23.5	5.6250
3	41	B	105	63.5	81.8750
4	41	B	106	72.5	83.5000
5	41	B	107	58.0	75.5000
6	41	B	108	49.0	13.7500
7	41	B	109	7.5	9.3750
8	41	B	110	13.5	7.8750
9	41	B	111	13.5	6.0000
10	41	B	112	76.5	61.6000
11	41	B	113	78.5	98.1250
12	41	B	114	56.5	46.1250
13	41	B	115	61.0	24.2500
14	41	B	116	91.0	64.4000
15	41	B	117	13.5	7.3333
16	41	B	118	63.5	79.2500
17	42	A	201	50.5	70.3333
18	42	A	203	84.5	16.1429
19	42	B	202	33.5	28.3333
20	43	A	302	22.0	14.5000
...	...	...	...	...	...

Complete Frequency Table of Clinic*trt

The FREQ Procedure

Table of Trt by Clinic

		Clinic										
		41	42	43	44	45	46	47	48	49	50	Total
Trt												
A	Frequency	0	2	14	6	12	6	6	8	10	4	68
	Percent	0.00	1.32	9.27	3.97	7.95	3.97	3.97	5.30	6.62	2.65	45.03
	Row Pct	0.00	2.94	20.59	8.82	17.65	8.82	8.82	11.76	14.71	5.88	
	Col Pct	0.00	66.67	50.00	50.00	52.17	46.15	46.15	50.00	50.00	57.14	
B	Frequency	16	1	14	6	11	7	7	8	10	3	83
	Percent	10.60	0.66	9.27	3.97	7.28	4.64	4.64	5.30	6.62	1.99	54.97
	Row Pct	19.28	1.20	16.87	7.23	13.25	8.43	8.43	9.64	12.05	3.61	
	Col Pct	100.00	33.33	50.00	50.00	47.83	53.85	53.85	50.00	50.00	42.86	
Total	Frequency	16	3	28	12	23	13	13	16	20	7	151
	Percent	10.60	1.99	18.54	7.95	15.23	8.61	8.61	10.60	13.25	4.64	100.00

Two-Way Mixed Model

Trt*Clinic effect, random

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

Trt effect, fixed

Clinic effect, random

i = 1, 2 (**Trt**)
j = 1 to 10 (**Clinic**)
k = 1 to n_{ij} (**Patient**)

$E(y_{ijk})=\mu+\alpha_i=\mu_i$ and $Var(y_{ijk})=Var(b_j)+Var((\alpha b)_{ij})+Var(\varepsilon_{ijk})$
$Cov(y_{ijk}, y_{ijk'})=Var(b_j)+Var((\alpha b)_{ij})$, $k \neq k'$
$Cov(y_{ijk}, y_{i'jk'})=Var(b_j)$, $i \neq i'$ and $k \neq k'$

Analyze the Unbalanced Multicenter Clinical Trial Data

This demonstration illustrates the concepts discussed previously using only the unbalanced portion of the data.

In PROC MIXED, are the three different types of tests always identical for the highest order term when it is specified as the last term in your model?

- ❍ Yes
- ❍ No

multicenterExample.sas

Effects of Empty Cells in Mixed Models

CONTRAST, ESTIMATE, and LSMEANS statements can produce nonestimable linear combinations of parameters.

Clinic	Trt A	Trt B
41	0	16
42	2	1
43	14	14
…	…	…

Analyze Data with Empty Cells

This demonstration illustrates the concepts discussed previously.

multictr2Example.sas

E Option Output

Trt	Clinic										marginal
	41	42	43	44	45	46	47	48	49	50	
A		.1111	.1111	.1111	.1111	.1111	.1111	.1111	.1111	.1111	1
B	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-1
marginal	0	0	0	0	0	0	0	0	0	0	0

Estimation of Covariance Parameters

Objectives

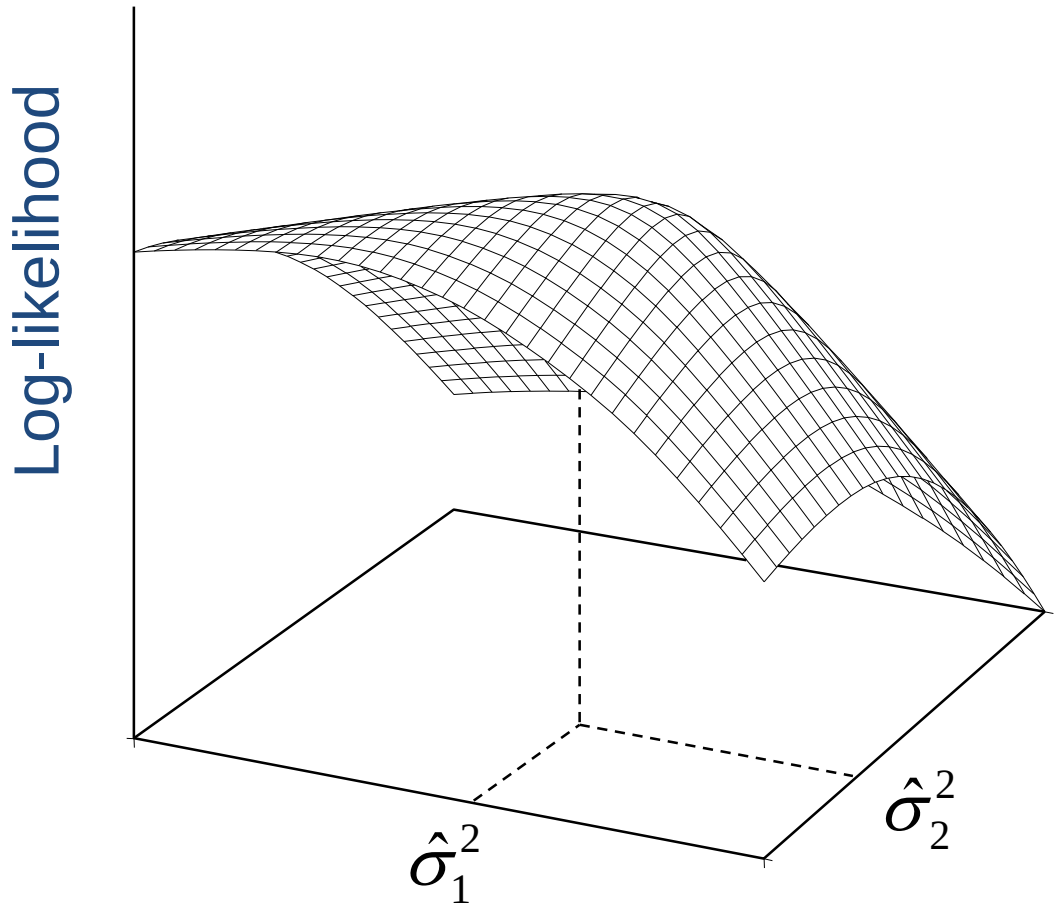
- Maximum Likelihood (ML) versus Restricted/Residual Maximum Likelihood (REML).
- Issues associated with zero or negative covariance parameter estimates.
- Inferences about the covariance parameters.

Estimation Methods for Covariance Parameters

- Method of Moments:
  - MIVQUE0
  - Type 1
  - Type 2
  - Type 3

- Likelihood-based Methods:
  - ML
  - REML (default)

Maximum Likelihood Estimation



ML versus REML

- Both methods are likelihood-based and therefore are consistent, asymptotically normal, and efficient.
- Both methods require numerical optimization.
- In general, REML estimators of the variance components are unbiased; ML estimators are biased low.
- In general, REML solutions are the ANOVA estimators for balanced data.
- REML is less sensitive to outliers in the data than ML is.
- The fit statistics based on REML can be used to compare different covariance models based on the same mean model; the fit statistics based on ML can be used to compare different mean models based on the same covariance model.

Estimation of Variance Components

REML: Var(clinic)=0, Var(trt*clinic)=75.3629,
Var(ε)=447.57

ML : Var(clinic)=0, Var(trt*clinic)=48.5949,
Var(ε)=452.81

MIVQUE0: Var(clinic)=0.004578, Var(trt*clinic)=29.0592,
Var(ε)=475.87

TYPE3: Var(clinic)=-12.6728, Var(trt*clinic)=89.0063,
Var(ε)=435.83

multicenterExample

Zero Variance Component Estimates

Zero variance component estimates can arise for a variety of reasons:

- The variability in your data might be large enough to produce a negative estimate, even though the true value of the variance component is positive.
- Your data might contain outliers.
- A different model for interpreting your data might be appropriate.

What Can Be Done?

- Plot your data and check for extreme or unusual values.
- Respecify your model.
- Investigate significant variance components, if appropriate.

Making Inferences about Covariance Parameters

- Use the COVTEST option or the CL option in the PROC MIXED statement. Covtest option computes the Wald-Z statistics which is based on large sample theory and is only asymptotically valid. CL option computes the confidence limits based on chisquare distribution and is also asymptotically valid.

- Use the METHOD=TYPE3 option in the PROC MIXED statement and examine the *F*-test results. This is better to use when you do not have repeated or random statements with subject. It computes the F test and it is an exact test for balanced data and computes reasonable approximations for unbalanced data.

- Use likelihood ratio tests. When one or more variance components hit the boundary, the NOBOUND option in the PROC MIXED statement should be used.

Negative Variance Component Estimates

This demonstration illustrates the concepts discussed previously.

Is the variance for trt*clinic significant at alpha=0.05?

a. The F test shows that trt*clinic variance is significant.
b. The likelihood ratio test shows that trt*clinic variance is significant.
c. The F test and the likelihood ratio test produced inconsistent results, so you must respecify the model until both tests are consistent.
d. I do not know.

multicenterExample

Degrees of Freedom Estimates

Objectives

- Discuss issues with degrees of freedom estimates.
- Explain different degrees of freedom estimation methods in the MIXED procedure.
- Use different degrees of freedom estimation methods in PROC MIXED and compare the results.

Issues with Degrees of Freedom Estimates

- For balanced data sets only and for a number of hypotheses, the determination of the denominator degrees of freedom is simple.

- For unbalanced data and/or more complex linear hypotheses about the fixed effects, the denominator degrees of freedom must be estimated from the data.

- The default degrees of freedom provided by the MIXED procedure might not always be appropriate.

Containment Method

DDFM=CONTAIN

- is the default method when you specify the RANDOM statement
- uses the degrees of freedom for the random effect syntactically containing the effect of interest.

```
model drug hour drug*hour basefev1;
random patient(drug);
```

DDF for **drug** is DF for **patient(drug)**

DDF for **hour**, **drug*hour**, and **basefev1** is DF for residual

DF for residual =n-Rank(X,Z)

- This method is computationally expensive when Z is large

Between-Within Method

DDFM=BETWITHIN / BW

- is the default method when you specify the REPEATED statement with no RANDOM statement
- divides the residual degrees of freedom into between-subject and within-subject portions, and then determines the appropriate degrees of freedom as one of the two based on whether fixed effect changes within any subject
- is a fairly reliable, quick way to obtain DDFM for models with a REPEATED statement.

Residual Method

DDFM=RESIDUAL

- performs all tests using the residual degrees of freedom.
  residual df=n-rank(X)
- typically overestimates the optimal denominator degrees of freedom
- is computationally cheap and might be appropriate when you are interested in predictions not the inferences. When the df is very large, it provides a reasonable approximation for inferences.

Satterthwaite's Method

DDFM=SATTERTH

- more computer intensive but advantages are worth it
- provides a better approximation that usually lies between the lowest and highest values of degrees of freedom associated with variance components that contribute to the background variability
- is syntactically invariant
- is a better choice in unbalanced situations and/or when the default methods are not appropriate.

Satterthwaite's Formula

Let $MS_1, MS_2, \ldots, MS_k$ be independent mean squares with degrees of freedom $df_1, df_2, \ldots, df_k$, respectively.

Then the linear combination

$MS = a_1MS_1 + a_2MS_2 + \ldots + a_kMS_k$

has the approximate degrees of freedom:

$$df = \frac{(MS)^2}{\frac{(a_1MS_1)^2}{df_1} + \cdots + \frac{(a_kMS_k)^2}{df_k}}$$

Example

Type 3 Analysis of Variance

Source	Error Term	Error DF	F Value	Pr > F
Trt	0.6604 MS(Trt*Clinic) + 0.3396 MS(Residual)	11.689	2.17	0.1674
Clinic	MS(Trt*Clinic)	8	0.83	0.5995
Trt*Clinic	MS(Residual)	114	2.44	0.0178
Residual	.	.	.	.

$$\underline{MS(Error)} = \underline{0.6604}*\underline{MS(Trt*Clinic)} + \underline{0.3396}*\underline{MS(Residual)}$$

MS a1 MS1 a2 MS2

KENWARDROGER Option

DDFM=KENWARDROGER / KR

- corrects the downward bias of the variance-covariance matrix of the fixed and random effects
- computes the Satterthwaite-type degrees of freedom based on the adjusted covariance matrix
- is computationally expensive
- is considered by many to be the most appropriate method.

DDFM=KR Option

- One issue with the DDFM=KR option is it might have undesirable consequences for the adjustment when covariance matrices have nonzero second derivatives
  - Adjustment can lead to shrinkage of standard errors
  - Adjusted covariance matrix might not be positive definite
  - Results are not invariant under reparameterization.

- Affected covariance structures include
  - ANTE(1), AR(1), ARH(1), ARMA(1,1)
  - CHOL, CSH, FA0(), TOEPH, UNR, and all SP().

The DDFM=KR(FIRSTORDER) Option

The FIRSTORDER suboption

- eliminates the second derivatives from the calculation of the covariance matrix adjustment
- might be preferred for covariance structures that have nonzero second derivatives.

Estimating Degrees of Freedom

Denominator df (right) Method (below)	For Drug	For hour	For drug*hour	For basefev1
Containment	68=3*(24-1)-1	483	483	483=576-(1+2+7+14+1+68)
Between-within	68	483	483	68
Residual	551=576-(1+2+7+14+1)	551	551	551
Satterthwaite	68.8	278	278	68
KenwardRoger	68.8	395	424	68

fev1uniExample.sas

Denominator Degrees-of-Freedom Methods

The DDFM= option in the MODEL statement specifies the method for computing the df for the tests of fixed effects resulting from the MODEL, CONTRAST, ESTIMATE, LSMEANS, LSMESTIMATE and has the following methods:

- CONTAIN default for models containing G-side random effects
- BETWITHIN default for models containing only the R-side random effects and models processed by subjects
- RESIDUAL default for generalized linear models
- NONE produce the chi-squared tests
- SATTERTHWAITE
- KENWARDROGER.

Recommendations for DDF Methods

- For linear mixed models and moderate sized problems, use of DDFM=KR is recommended in most cases.

- For generalized linear models with no random effects, you might want to use DDFM=NONE to obtain the chi-squared tests.

- For a nonnormal response with random effects, the recommendation might not be available due to lack of studies thus far.

One Issue with the DDFM=KR Option

- The DDFM=KR option might have undesirable consequences for the adjustment when covariance matrices have second derivatives
  - adjustment can lead to shrinkage of standard errors
  - adjusted covariance matrix may not be positive definite
  - results are not invariant under reparameterization.
- Affected covariance structures include
  - ANTE(1), AR(1), ARH(1), ARMA(1,1)
  - CHOL, CSH, FA0(), TOEPH, UNR, and all SP().

The FIRSTORDER suboption (DDFM=KR(FIRSTORDER))

- eliminates the second derivatives from the calculation of the covariance matrix adjustment
- might be preferred for covariance structures that have nonzero second derivatives.

Automatic Variables

Automatic variables can be used to specify your own link function and/or variance function. They include the following:

- _LINP_ linear predictor
- _MU_ expresses the mean of an observation as a function of the linear predictor
- _VARIANCE_ estimate of the variance function
- _XBETA_ equals $\mathbf{X}\hat{\beta}$
- _ZGAMMA_ equals $\mathbf{Z}\hat{\gamma}$
- _N_ the observation number in the sequence of the data read.

Automatic Variables in PROC GLIMMIX

```
proc glimmix data=yourdata;
   class A B;
   model y/n = A B / dist=binomial link=logit;
run;
```

```
proc glimmix data=yourdata;
   class A B;
   prob = 1 / (1+exp(-_linp_));
   _mu_ = n * prob ;
   _variance_ = n * prob *(1-prob);
   model y = A B;
run;
```

Convergence Problems

Lack of convergence can be caused by the following:

- widely varying scales
  - for data values
  - for parameters
- incorrectly specified models
- over-specified models
- poor starting values
- poor optimization algorithms
- problematic parameterizations
- not enough data for a specified model.

Dealing with Convergence Problems

- Plot your data and check for extreme or unusual observations. Adjust or delete them if appropriate.
- Be certain that your model is appropriate for your data.
- Rescale the data to improve stability.
- Try to fit a simple model and then gradually increase the complexity.
- Provide different starting values or attempt a grid search of values.
- Use **`TYPE=CHOL`** rather than the **`TYPE=UN`** option.
- Try a different optimization technique. For example, **`NLOPTIONS tech=nrridg;`** often works better for binary data.

continued...

Dealing with Convergence Problems

- Use a different estimation method for GzLMMs; for example, **`method=RMPL`** for pseudo-likelihood method or **`method=LAPLACE`** for likelihood method.
- Increase the maximum number of optimizations, **`maxopt=100`**, for example.
- Tune the singularity criterion; for example, **`singular=1e-8`**.
- Tune the convergence criterion for the individual optimizations; for example, **`nloptions gconv=0`**.
- Tune the convergence criterion for the doubly iterative estimation method; for example, **`pconv=1e-10`**.