STATISTICS
TEXAS A&M UNIVERSITY

Handout 16

Residual and Influence

Diagnostics and Troubleshooting

Residual and Influence Diagnostics

Objectives

- Use ODS Statistical Graphics.
- Perform linear mixed models residual and influence diagnostics.

General Syntax of ODS Graphics

ODS GRAPHICS ON;

statistical procedure code

<ODS GRAPHICS OFF;>

- ODS Graphics are turned on in the SAS Windowing Environment and can be turned off from the Preferences.
- Some procedures produce certain graphs by default.
- Other procedures require options (such as PLOTS=) to produce graphics of your choice.
- You can find more details at

http://support.sas.com/documentation/cdl/en/odsug/67325/HTML/default/viewer.htm#p0kroq43yu0lspn16hk1u4c65lti.htm or

http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_mixed_sect027.htm

Example

```
ods graphics on;
proc mixed data=fev1uni
           plots=boxplot;
   class drug patient hour;
   model fev1=drug basefev1 hour drug*hour
              / ddfm=kr residual;
   repeated hour / type=TOEP
                   subject=patient(drug);
run;
ods graphics off;
```

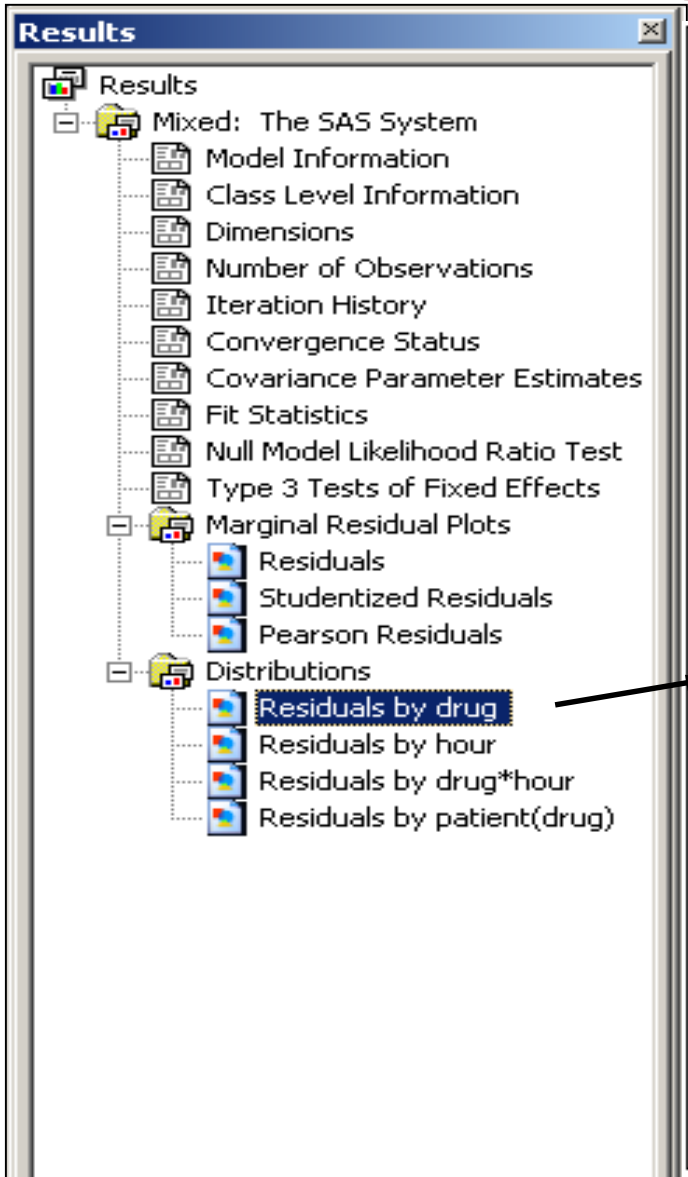
PROC MIXED ODS Graphics

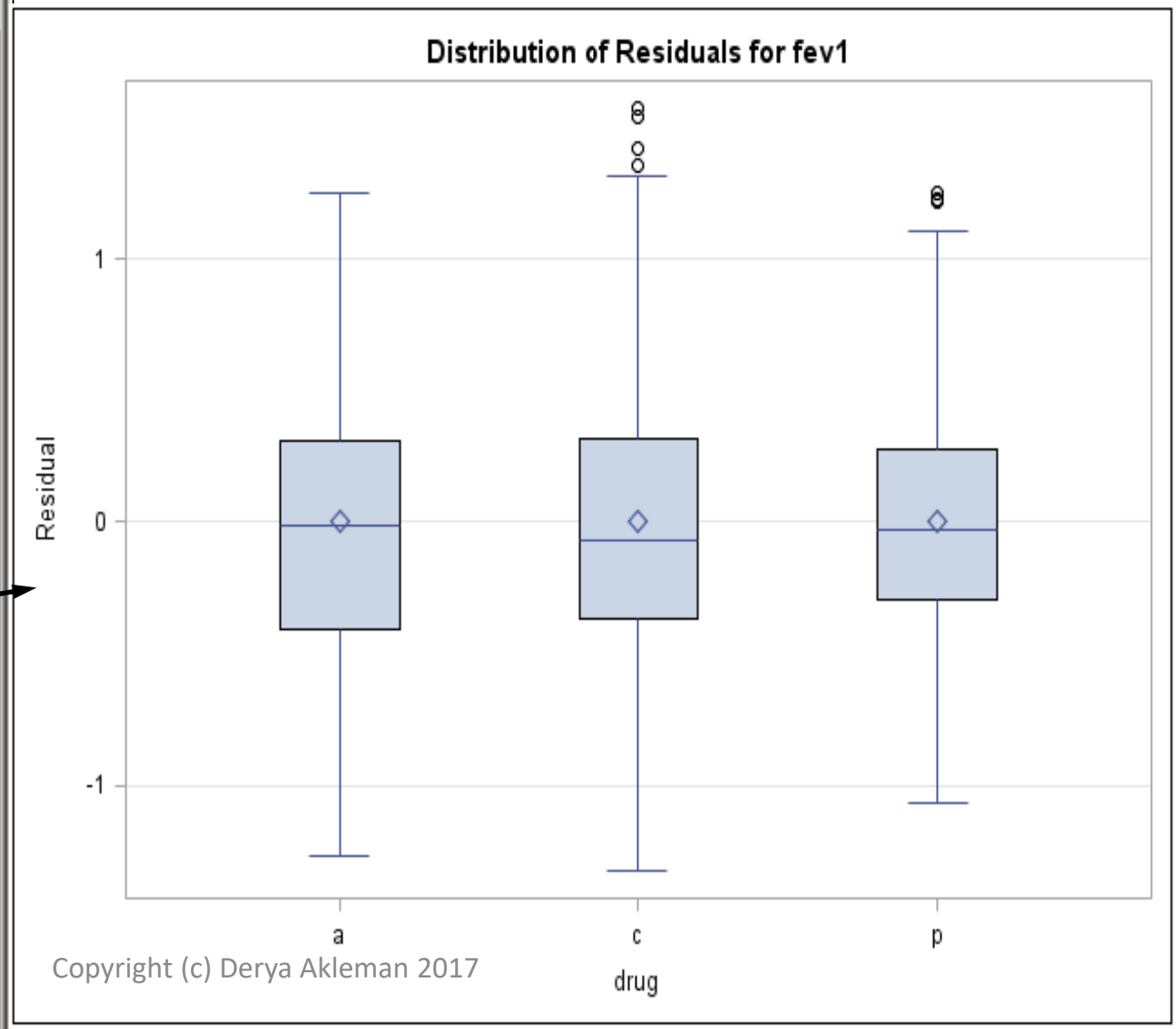
Table 56.24 ODS Graphics Produced by PROC MIXED

ODS Graph Name	Plot Description	Statement or Option
Boxplot	Box plots	PLOTS=BOXPLOT
CovRatioPlot	CovRatio statistics for fixed effects or covariance parameters	PLOTS=INFLUENCESTATPANEL(UNPACK) and MODEL / INFLUENCE
CooksDPlot	Cook's D for fixed effects or covariance parameters	PLOTS=INFLUENCESTATPANEL(UNPACK) and MODEL / INFLUENCE
DistancePlot	Likelihood or restricted likelihood distance	MODEL / INFLUENCE
InfluenceEstPlot	Panel of deletion estimates	MODEL / INFLUENCE(EST) or PLOTS=INFLUENCEESTPLOT and MODEL / INFLUENCE
InfluenceEstPlot	Parameter estimates after removing observation or sets of observations	PLOTS=INFLUENCEESTPLOT(UNPACK) and MODEL / INFLUENCE
InfluenceStatPanel	Panel of influence statistics	MODEL / INFLUENCE
PearsonBoxPlot	Box plot of Pearson residuals	PLOTS=PEARSONPANEL(UNPACK BOX)
PearsonByPredicted	Pearson residuals vs. predicted	PLOTS=PEARSONPANEL(UNPACK)
PearsonHistogram	Histogram of Pearson residuals	PLOTS=PEARSONPANEL(UNPACK)
PearsonPanel	Panel of Pearson residuals	MODEL / RESIDUAL
PearsonQQplot	***Q-Q*** plot of Pearson residuals	PLOTS=PEARSONPANEL(UNPACK)
PressPlot	Plot of PRESS residuals or PRESS statistic	PLOTS=PRESS and MODEL / INFLUENCE
ResidualBoxplot	Box plot of (raw) residuals	PLOTS=RESIDUALPANEL(UNPACK BOX)
ResidualByPredicted	Residuals vs. predicted	PLOTS=RESIDUALPANEL(UNPACK)
ResidualHistogram	Histogram of raw residuals	PLOTS=RESIDUALPANEL(UNPACK)
ResidualPanel	Panel of (raw) residuals	MODEL / RESIDUAL
ResidualQQplot	***Q-Q*** plot of raw residuals	PLOTS=RESIDUALPANEL(UNPACK)
ScaledBoxplot	Box plot of scaled residuals	PLOTS=VCIRYPANEL(UNPACK BOX)
ScaledByPredicted	Scaled residuals vs. predicted	PLOTS=VCIRYPANEL(UNPACK)
ScaledHistogram	Histogram of scaled residuals	PLOTS=VCIRYPANEL(UNPACK)
ScaledQQplot	***Q-Q*** plot of scaled residuals	PLOTS=VCIRYPANEL(UNPACK)
StudentBoxplot	Box plot of studentized residuals	PLOTS=STUDENTPANEL(UNPACK BOX)
StudentByPredicted	Studentized residuals vs. predicted	PLOTS=STUDENTPANEL(UNPACK)
StudentHistogram	Histogram of studentized residuals	PLOTS=STUDENTPANEL(UNPACK)
StudentPanel	Panel of studentized residuals	MODEL / RESIDUAL
StudentQQplot	***Q-Q*** plot of studentized residuals	PLOTS=STUDENTPANEL(UNPACK)
VCIRYPanel	Panel of scaled residuals	MODEL / VCIRY

STATISTICS
TEXAS A&M UNIVERSITY

Accessing Individual Graphs from the Results Window





Modifying Your Graphs

- Use the ODS Graph Editor, a point-and-click interface
  - for data and graph-specific changes
  - to customize titles and labels, annotate data points, add text, and change the properties of graph elements.
- Make persistent changes by modifying the ODS graph template for a particular plot.

Mixed Model Diagnostics

- You can request raw residuals, Pearson residuals, studentized residuals, and scaled residuals.
- You can request influence statistics by noniterative or iterative methods.
- ODS Graphics can be used to display the results.

Raw Residuals—Conditional and Marginal

– OUTP= option in the MODEL statement computes the raw conditional residuals

$$r_{ci} = y_i - \mathbf{x}_i'\hat{\beta} - \mathbf{z}_i'\hat{\gamma}$$

Conditional residuals are helpful for detecting outlying subjects or diagnose whether the random effects are reasonably specified.

– OUTPM= option in the MODEL statement computes the raw marginal residuals

$$r_{mi} = y_i - \mathbf{x}_i'\hat{\beta}$$

Marginal residuals are helpful for diagnose whether the fixed effects are reasonably specified.

Residuals in SAS® 9.4

- Pearson Residuals and Studentized Residuals
  - are requested by the RESIDUAL option in the MODEL statement
  - are added to OUTP= and/or OUTPM= data sets.
- Scaled Residuals
  - are requested by the VCIRY option in the MODEL statement
  - are added to the OUTPM= data set.
- The RESIDUAL option or the VCIRY option has no effect unless the OUTP= or OUTPM= option is specified or you request the ODS Graphics.

Residuals in SAS® 9.4

Type of Residuals	Marginal (OUTPM=)	Conditional (OUTP=)
Raw (default)	$r_{mi} = Y_i - \mathbf{x}'_i \hat{\beta}$	$r_{ci} = r_{mi} - z'_i \hat{\gamma}$
Studentized (the RESIDUAL option)	$r_{mi}^{student} = \frac{r_{mi}}{\sqrt{\hat{\text{var}}(r_{mi})}}$	$r_{ci}^{student} = \frac{r_{ci}}{\sqrt{\hat{\text{var}}(r_{ci})}}$
Pearson (the RESIDUAL option)	$r_{mi}^{Pearson} = \frac{r_{mi}}{\sqrt{\hat{\text{var}}(Y_i)}}$	$r_{ci}^{Pearson} = \frac{r_{ci}}{\sqrt{\hat{\text{var}}(Y_i \mid \gamma)}}$
Scaled (the VCIRY option)	$\hat{\mathbf{C}}^{-1} r_{mi}$ where $\hat{\mathbf{C}}\hat{\mathbf{C}}' = \mathbf{V}(\hat{\theta})$	

Scaled Residuals

Scaled Residuals

For correlated data, a set of scaled quantities can be defined through the Cholesky decomposition of the variance-covariance matrix. Since fitted residuals in linear models are rank-deficient, it is customary to draw on the variance-covariance matrix of the data. If $\text{Var}[\mathbf{Y}] = \mathbf{V}$ and $\mathbf{C}'\mathbf{C} = \mathbf{V}$, then $\mathbf{C}'^{-1}\mathbf{Y}$ has uniform dispersion and its elements are uncorrelated.

Scaled residuals in a mixed model are meaningful for quantities based on the marginal distribution of the data. Let $\widehat{\mathbf{C}}$ denote the Cholesky root of $\widehat{\mathbf{V}}$, so that $\widehat{\mathbf{C}}'\widehat{\mathbf{C}} = \widehat{\mathbf{V}}$, and define

$$\mathbf{Y}_c = \widehat{\mathbf{C}}'^{-1}\mathbf{Y}$$

$$\mathbf{r}_{m(c)} = \widehat{\mathbf{C}}'^{-1}\mathbf{r}_m$$

By analogy with other scalings, the inverse Cholesky decomposition can also be applied to the residual vector, $\widehat{\mathbf{C}}'^{-1}\mathbf{r}_m$, although $\mathbf{V}$ is not the variance-covariance matrix of $\mathbf{r}_m$.

To diagnose whether the covariance structure of the model has been specified correctly can be difficult based on $\mathbf{Y}_c$, since the inverse Cholesky transformation affects the expected value of $\mathbf{Y}_c$. You can draw on $\mathbf{r}_{m(c)}$ as a vector of (approximately) uncorrelated data with constant mean.

When the OUTPM= option in the MODEL statement is specified in addition to the VCIRY option, $\mathbf{Y}_c$ is added as variable *ScaledDep* and $\mathbf{r}_{m(c)}$ is added as *ScaledResid* to the data set.

Residual Analysis

- More mathematics and explanation for residual analysis can be found at

http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_mixed_sect027.htm

- Studentized residuals and the Pearson residuals are useful for detecting potential outliers.

- Scaled residuals are useful for evaluating the appropriateness of the covariance structure of your model.

Influence Diagnostics

The INFLUENCE option in the MODEL statement

1. Fits the model to the data and obtains estimates of all parameters

2. Removes one or more data points from the analysis and computes updated estimates of model parameters

3. Based on full- and reduced-data estimates, contrasts quantities of interest to determine how the absence of the observations changes the analysis.

The Nature of the Influence

The Observation is Influential on		Statistics
the overall objective function		Likelihood distance
the fitted and predicted values		DFFITS and PRESS residuals
fixed effects	the estimates	Cook's D or Multivariate DFFITS
	the precision	COVTRACE or COVRATIO
covariance parameters	the estimates	Cook's D or Multivariate DFFITS
	the precision	COVTRACE or COVRATIO

Iterative and Noniterative Influence Analysis

Iterative influence analysis

- refits the model and iteratively re-estimates the covariance parameters when the observations in questions are removed
- generally is a better approach but is computationally intensive.

Noniterative influence analysis

- relies on closed-form update formulas for the fixed effects without updating the (unprofiled) covariance parameters
- is computationally efficient and is the default analysis.

Question

Which of the following is **false** about the model diagnostics in PROC MIXED?

a. You use the RESIDUAL or the VCIRY option in the MODEL statement to request residual analysis.
b. You use the INFLUENCE option in the MODEL statement to request influential analysis.
c. The RESIDUAL option should be specified together with OUTP= or OUTPM= option, or with ODS Graphics.
d. The default influential analysis is an iterative method.

Residual and Influence Diagnostics in PROC MIXED

This demonstration illustrates the concepts discussed previously.

Handout16_fev1uniExample.sas

Troubleshooting Convergence Problems

Objectives

- explain common causes of convergence problems.
- suggest ways of dealing with nonconvergence.
- deal with a nonconvergence situation in the MIXED procedure.
- understand the note in the LOG window about the G matrix not being positive definite.

Common Causes of Nonconvergence

- Two covariance parameters that are several orders of magnitude apart
- Values that are extremely large or extremely small in scale
- Infinite likelihood caused by nonpositive definite or singular **R** or **V** matrix, or not enough data to estimate the specified covariance structure
- Linear dependencies among covariance parameters, and/or confounding of mean and covariance parameters
- Over-specified or incorrectly specified model
- Violation of model assumptions

Ways of Dealing with Nonconvergence

- Rescale the data to improve stability; add ITDETAILS and LOGNOTE options to obtain more information.
- Plot your data and check for extreme or unusual observations. Adjust or delete them if appropriate.
- Use the PARMS statement to input initial values. Add boundary constraints (LOWERB= and/or UPPERB=) for parameters that might be unstable. Search over a grid of parameters.
- Specify the SCORING= option to invoke the Fisher scoring estimation method.

continued...

Ways of Dealing with Nonconvergence

- Make sure no observations from the same subject are producing identical rows in the **R** or **V** matrix.
- Rearrange effects in the MODEL statement so that the most significant ones are first to improve stability.
- Make sure you have enough data and the parameters are not linearly dependent. Respecify the model, if necessary.
- Try fitting a simple model and then gradually increase complexity.

continued...

Ways of Dealing with Nonconvergence

- Tune the singularity options
  - SINGULAR=, tunes the sensitivity in sweeping
  - SINGCHOL=, tunes the sensitivity in cholesky roots
  - SINGRES= sets the tolerance for which Var(residual)=0

  in the MODEL statement.
- Tune the MAXITER= (default is 50) and MAXFUNC= (default is 150) options in the PROC MIXED statement.
- Use the NOPROFILE (residual variance as a part og Newton-Raphson) and NOBOUND options in the PROC MIXED statement.
- Try CONVF= or CONVG=, possibly along with the ABSOLUTE option, as a convergence criterion in the PROC MIXED statement.

Example 1

```
proc mixed data=fev1uni;
   class drug patient hour;
   model fev1=drug basefev1 drug*basefev1
              hour drug*hour / ddfm=kr;
   random patient(drug);
   repeated hour / type=cs
                   subject=patient(drug);
run;
```

The Log and the Output Windows

```
NOTE: Convergence criteria met but final hessian is not positive
      definite.
NOTE: Asymptotic variance matrix of covariance parameter estimates
      has been found to be singular and a generalized inverse was
      used. Covariance parameters with zero variance do not
      contribute to degrees of freedom computed by DDFM=KENWARDROGER.
```

Convergence criteria met but final hessian is not positive definite.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
patient(drug)		0.2088
CS	patient(drug)	0
Residual		0.06313

Example 1 – What Went Wrong?

– It is an over-parameterized model.

- The REPEATED statement fits a compound symmetry structure for the residual covariance.
- The RANDOM statement is redundant because the resulting **V** matrix is also a compound symmetry.
- Similar issues exist for TYPE=UN, TYPE=TOEP, and some of other structures.
- You might want to delete the RANDOM statement, or keep it and specify serial correlations such as AR(1) for the REPEATED statement.

Example 2

```
proc mixed data=fev1test;
   class drug patient hour;
   model fev1=drug basefev1 hour drug*hour
                / ddfm=kr;
   repeated hour / type=TOEP
                   subject=patient(drug);
run;
```

Example 2

```
proc mixed data=fev1test;
   class drug patient hour;
   model fev1=drug basefev1 hour drug*hour
                / ddfm=kr;
   repeated hour / type=TOEP
                   subject=patient(drug);
run;
```

```
NOTE: An infinite likelihood is assumed in iteration 0 because of a
      nonpositive definite estimated R matrix for patient(drug) 214 a.
```

Example 2 – What Went Wrong?

```
proc print data=fev1test;
   where patient=214 and drug='a';
run;
```

obs	patient	basefev1	drug	hour	fev1
97	214	2.77	a	1	3.36
98	214	2.77	a	2	3.42
99	214	2.77	a	3	3.28
100	214	2.77	a	4	3.30
101	214	2.77	a	5	3.31
102	214	2.77	a	6	2.99
103	214	2.77	a	7	3.01
104	214	2.77	a	7	3.08

Question

Can nonpositive definite R matrix be the result of erroneous data or misspecification of the SUBJECT= effect?

- Yes
- No

Example 3

When specifying TYPE=UN in the RANDOM or the REPEATED statement and the program failed to converge, try to specify TYPE=FA0(*q*) where *q* corresponds to the dimension of the UN matrix.

```
proc mixed data=wheat;
    class variety;
    model yield=moisture / ddfm=kr;
    random int moisture / type=FA0(2)
                          subject=variety;
    parms / lowerb=.,.,0.001;
run;
```

Example 4

```
proc mixed data=mydata;
    class network hospital;
    model y=network / ddfm=kr;
    random hospital;
run;
```

```
ERROR: Out of memory.
NOTE: The SAS System stopped processing this step because of
      insufficient memory.
```

Example 4 – What Went Wrong?

Dimensions

Covariance Parameters	2
Columns in X	5
Columns in Z	378
Subjects	1
Max Obs Per Subject	7560

Number of Observations

Number of Observations Read	7560
Number of Observations Used	7560
Number of Observations Not Used	0

- There are 378 levels for the random effect **`hospital`**.
- Large **Z** matrix is resource intensive.

Example 4 – Modify Your Program

```
proc mixed data=mydata;
    class network hospital;
    model y=network / ddfm=kr;
    random int / subject=hospital;
run;
```

```
NOTE: Convergence criteria met.
NOTE: PROCEDURE MIXED used (Total process time):
      real time             0.35 seconds
      cpu time              0.33 seconds
```

Example 4 – Other Possible Specifications

```
proc sort data=mydata;
   by hospital;
run;

proc mixed data=mydata;
   class network;
   model y=network / ddfm=kr;
   random int / subject=hospital;
run;
```

```
proc mixed data=mydata;
   class network hospital;
   model y=network / ddfm=kr;
   repeated/type=cs subject=hospital;
run;
```

Equivalent Marginal Models

```
random int / subject=hospital;
```

$$\rightarrow \quad \mathbf{G} = \sigma_1^2 \mathbf{I}_m, \mathbf{R} = \sigma^2 \mathbf{I}_n$$

It follows that the **V** matrix has a compound symmetry structure (**V**=**ZGZ**'+**R**).

```
repeated/type=cs subject=hospital;
```

→ No **G** matrix. **R** is compound symmetry.

It follows that the **V** matrix has the same compound symmetry structure (**V=R**).

Question

Lack of convergence should be evaluated on a case-by-case basis because no panacea exists.

- ❍ True
- ❍ False

The LOG Window

```
NOTE: Convergence criteria met.
NOTE: Estimated G matrix is not positive definite.
NOTE: Asymptotic variance matrix of covariance parameter estimates
      has been found to be singular and a generalized inverse was
      used. Covariance parameters with zero variance do not contribute
      to degrees of freedom computed by DDFM=KENWARDROGER.
```

The Partial Output

Covariance Parameter Estimates

Cov Parm	Estimate
Clinic	0
Trt*Clinic	75.3629
Residual	447.57

Zero Variance Component Estimates

Zero variance component estimates might 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 to be sure the model is not overparameterized.
- Use the NOBOUND option in the MODEL statement.
- Investigate significant variance components, if appropriate.

Question

Select all that apply. Which statements are true?

a. The F-tests on covariance parameters can be obtained by using the METHOD=TYPE3 option for models with no REPEATED statement and with no SUBJECT= option in the RANDOM statement.
b. The note in the LOG window about the G matrix being nonpositive definite might be a result of one of more variance components hitting the boundary.
c. Negative (or zero) variance component estimates might indicate model misspecifications.
d. The unbiased estimates for variance components can occasionally be negative.