

Introduction to Generalized Linear Mixed Models

A Count Data Example

Jerry W. Davis, University of Georgia, Griffin Campus

Analysis of variance rests on three basic assumptions: response variables are normally distributed, individual observations are independent and the variances between experimental units are homogeneous. Data from agricultural experiments do not always follow these assumptions. Traditional analysis of variance techniques are very robust, so some deviation from these assumptions does not necessarily lead to erroneous results, and the Central Limit Theorem implies that data from experiments with many observations have means that are approximately normal. Traditionally, transformations were used to “normalize” categorical response variables and minimize the effect of heterogeneous variances. In the 1990s and 2000s, advances in statistical methods and computing technology allowed analysts to model data from non-normal distributions in an analysis of variance framework. The distribution of the response variable is part of the model so the normality assumptions are unnecessary. These models are called generalized linear models because they extend linear model theory to model categorical response variables. Finally, mixed model theory was incorporated, which led to generalized linear mixed models.

Analysis of Variance Models

Linear models (**LM**) are for normally distributed (Gaussian) data and only model fixed effects. SAS (SAS/STAT® Software, 2017) procedures **reg**, **glm** or **anova** fit these models.

Linear mixed models (**LMM**) are for normally distributed (Gaussian) data and can model random and / or repeated effects. The **mixed** procedure fits these models.

Generalized linear models (**GLM**) are for non-normal data and only model fixed effects. SAS procedures **logistic**, **genmod**¹ and others fit these models.

Generalized linear mixed models (**GLMM**) are for normal or non-normal data and can model random and / or repeated effects. The **glimmix** procedure fits these models.

GLMM is the general model, with **LM**, **LMM**, and **GLM** being special cases of the generalized model (Stroup, 2013).

Distributions

- Selecting the proper distribution is key when fitting a **GLMM**.² Fortunately, there are guidelines to follow which match types of response data to known distributions. Here are some common non-normal data types, their characteristics and distribution.

¹ GENMOD can model repeated effects for some types of data.

² Analysis of variance models are referred to by their abbreviations.

- **Counts.** Count data comes from counting events of interest in an experimental unit. Counts are non-negative integers, often right skewed, with a **Poisson** or **Negative Binomial** distribution. Number of insects, weeds, diseased plants, etc., within each plot are common response variables. Count data is unbounded, i.e., there are no predetermined limits imposed on the range of values. For example, the number of flowers on a plant, cotton bolls per plant or number of insects in an area may have biological limits, but the limits are not predetermined. Whereas the number of plants in a plot that survived without water is bounded by the original number of plants in the plot. It is better to treat these counts as having a binomial distribution rather than a Poisson or negative binomial.
- **Binomial.** Binomial data are discrete positive integers between 0 and n . It is the standard distribution for the number of successes from n independent trials with only two outcomes. Usual outcomes are success or failure, 1 or 0, alive or dead, etc. Also called discrete proportions, example data include eggs hatched from the total number of eggs, seeds germinated from the total number of seeds or number of plants that survived without water from the total number of plants within a plot. When n equals one, it is known as a Bernoulli trial. A coin toss is an example of a Bernoulli trial.
- **Continuous proportions.** These data are percentages that represent the proportion of affected subjects, areas, etc., within an experimental unit. These proportions have a **Beta** distribution. Example data include the percent of a plot with insect or disease infestation, damaged leaf area, lesion size as a percent, etc.
- **Ratings and ranks.** Subjective measurements based on a discrete scale or criteria. The scale may or may not be linear. A linear scale implies the difference between ratings is equal. With a non-linear scale, the difference between one and two on a six-point scale; for example, is not equal to the difference between five and six. These data follow a **Multinomial** distribution. Example data include disease ratings, sensory evaluations, herbicide efficacy ratings, etc. Multinomial response data need not be numeric; **proc glimmix** will analyze data consisting of either numeric or character ratings.

Poisson and Negative Binomial Distributions

For a **Poisson** distribution the mean equals the variance ($\mu = \sigma^2$). This relationship implies that the events are randomly and evenly distributed within the experimental units. This is often an unrealistic assumption for agricultural data where events, such as insect outbreaks or patches of weeds, may be clustered with much variability between plots. When this happens, the variance may be larger than the mean. This condition is called **over-dispersion**. Over-dispersion may affect the fit and results of a **GLMM** so remedial steps are recommended to alleviate the problem. The data are under-dispersed when the variance is smaller than the mean.

The **negative binomial** distribution is similar to the Poisson except it has an additional parameter called a scale parameter. The scale parameter (δ) allows the variance to be larger (or smaller) than the mean

and may reduce or remedy the over-dispersion problem. Some argue that the negative binomial should always be used for agricultural data while others disagree.

Pseudo-likelihoods

Like linear mixed models, generalized linear mixed models use maximum likelihood techniques to estimate model parameters. The default estimation technique for **proc glimmix** is residual pseudo likelihood (**RSPL**) when the data are non-normal. However, **RSPL** does not produce a true log-likelihood when modeling non-normal data. It can only calculate a quasi or pseudo-likelihood. This affects the model in several ways.

- The model is not **conditioned** by the **random effects**. This may affect tests for the fixed effects and LS-means.
- Only a conditional model has the correct fit statistics for diagnosing over-dispersion. (Gbur et al, 2012).
- The information criteria (AIC, AICC, BIC, etc.), which are important for assessing the relative goodness of fit, cannot be calculated.

Changing the estimation method to **adaptive quadrature (quad)** or **Laplace** will solve these problems by fitting a true log-likelihood function. Therefore, when fitting count data, use **method=quad** or **method=laplace** to minimize the $-\log$ likelihood function.

Adaptive quadrature and Laplace methods have side effects as well. Some common ones are.

- The **ddfm** options **kr2** or **satterthwaite** are not available. Letting **ddfm** default to the **containment** method by omitting the **ddfm** option is acceptable.
- When using **method=quad**, random effects must be processed by subjects. For example, if a model has **block** as a random effect, **random block** will generate an error message. The correct syntax is, **random intercept / subject=block**.
- If there are two random effects, such as **block** and **year**, both affects must appear in the same **random** statement i.e., **random intercept / subject=block*year**. For a split-plot, only use the interaction term (**block*main plot**) as a random effect.
- The **laplace** method is less restrictive and doesn't complain about the simple syntax. Processing random effects by subject is more efficient and given the complexity of these models, writing random statements that way is probably best.
- Unlike adaptive quadrature, multiple **random statements** using the **subject=** syntax are allowed.
- Based upon limited testing, using the interaction random effect for **method=quad** gave the same results as using two random statements for **method=laplace**.

Adaptive quadrature is said to be more accurate, but I am beginning to prefer Laplace due to the fewer restrictions and often equal or better results.

Another, more problematic side effect is that **proc glimmix** cannot model categorical response variables from a repeated measures experiment the same way it models normally distributed response variables. When **method=quad** or **laplace** the **R-side** effects cannot be modeled. Recall that **R** or **R-side** effects

refer to the repeated or residual effects matrix and the **G** or **G-side** effects refer to the random effects matrix. Details about fitting a repeated measures GLMM for count data appear in a separate section along with an example program.

Link Functions

When fitting a GLMM the data remain on the original measurement scale (data scale). Yet when the means are estimated from a linear function of the explanatory variables, they are on the model scale. A link function is used to *link* the model scale means back to the original data scale. This is not the same approach as transforming the original measurements to a different measurement scale. “For example, application of the log transformation for counts followed by a normal theory analysis of variance is not the same as a generalized linear model assuming a Poisson distribution and a log link.” (Gbur et al, 2012). In a situation where the LS-means would normally be equal to the arithmetic means, the inverse linked means on the data scale may not equal the original sample means.

Distributions in **proc glimmix** have default **link** functions, but I always explicitly code the link function and encourage others to do so as well. This way there is no question as to which function was used. Using the wrong link will lead to unpredictable results. Here is a table of common distributions, the appropriate link function and the proper syntax for each. For a complete list, see tables 46.12 and 46.13 in version 14.2 of the SAS/STAT online documentation for PROC GLIMMIX.

Table 1. Common distributions and link functions for agricultural experiments.

Distribution	Link Function	Syntax dist=	Syntax Link=
Beta	Logit	dist=beta	link=logit
Binomial	Logit	dist=binomial bin b	link=logit
Normal	Identity	dist=gaussian g normal n	link=identity id
Multinomial	Cumulative logit	dist=multinomial multi mult	link=cumlogit clogit
Negative binomial	Log	dist=negbinomial negbin nb	link=log
Poisson	Log	dist=poisson poi	link=log

Fitting the Model

The mathematics behind fitting a GLMM are very complex. Using constructs like distributions, link functions, log likelihoods and quasi-likelihoods to fit a model is difficult to conceptualize. Perhaps the following points will help explain the modeling process.

- An analysis of variance model is a vector of linear predictors (equation) with unknown parameter estimates.
- Every distribution has a corresponding likelihood function.

- The vector of linear predictors is substituted into the likelihood function.
- Solutions for the parameter estimates are found by minimizing the negative of the log likelihood function (-log likelihood).
- LS-means are derived from the parameter estimates and are on the model scale.
- The **link** function converts the model scale LS-mean estimates back to the original data scale.

Two key concepts are:

- **Proc glimmix** uses a distribution to estimate model parameters; it does not fit the data to a distribution.
- The data values are not transformed by the **link** function; the **link** function converts the LS-means back to the data scale after being estimated on the model scale.

A Poisson Distribution Model for Count Data

This data comes from a randomized complete block experiment with five treatments (solvent) and five replications (colony). Two solvents were used to wash pheromones off argentine ant pupae onto pieces of paper. These pieces of paper, along with paper from three controls, were presented to five colonies of ants. The number of ants attracted to each treatment was recorded.³ Solvent is a fixed effect and colony is a random effect.

PROC GLIMMIX code to analyze the count data appears below. The complete program is listed in Appendix A. and available online.

```
proc glimmix method=quad;
```

Use the **method=** option to specify the method used to minimize the $-\log$ likelihood function. Adaptive quadrature (**quad**) or Laplace (**laplace**) are the preferred methods for categorical response variables. These two methods fit a conditional model.

```
class colony solvent;
model counts = solvent / dist=poisson link=log;
```

The **dist** option is where you specify the probability distribution that matches the response variable. The **link** option is for the distribution's link function. Omit the **ddfm** option so GLIMMIX will default to the **containment** method.

```
random intercept / subject=colony;
```

When **method=quad** random effects must be processed by subjects. Also, processing random effects by subjects is more efficient than using the **random colony** syntax.

```
lsmeans solvent / ilink;
```

The **ilink** (inverse link) option converts model scale LS-mean estimates back to the data scale.

³ Data is courtesy of Ben Gochnour and Dan Suiter.

```
store quad1;
run;
```

Store results for post processing.

```
proc plm restore=quad1;
  lsmeans solvent / ilink lines;
run;
```

While PROC GLIMMIX supports the **lines** option, which adds letter groups to LS-mean differences, I recommend and prefer using PROC PLM. It has more options, more flexibility and works with many procedures. Including GLM, MIXED, GENMOD, GLIMMIX, etc.

The **ilink** (inverse link) option is also supported in PROC PLM.

Output Tables for the Poisson Model

The following output tables display the results from modeling the response data using a Poisson distribution. Most of the PROC GLIMMIX output tables are the same as the output tables from PROC MIXED but some of the statistics have different interpretations. Some tables were omitted for brevity.

The GLIMMIX Procedure

Model Information	
Data Set	WORK.ONE
Response Variable	counts
Response Distribution	Poisson
Link Function	Log
Variance Function	Default
Variance Matrix Blocked By	colony
Estimation Technique	Maximum Likelihood
Likelihood Approximation	Gauss-Hermite Quadrature
Degrees of Freedom Method	Containment

The Model Information table identifies the response variable and summarizes relevant options used to fit the model. The variance matrix is blocked by colony because colony is a random effect.

Fit Statistics	
-2 Log Likelihood	278.04
AIC (smaller is better)	290.04
AICC (smaller is better)	294.70
BIC (smaller is better)	287.69
CAIC (smaller is better)	293.69
HQIC (smaller is better)	283.75

These fit statistics are calculable because **method=quad** (or **laplace**). If **method=rspl**, the default, this table doesn't appear in the output. Remember, these values are relative and, in general, only useful when comparing different model options.

Fit Statistics for Conditional Distribution	
-2 log L(counts r. effects)	252.59
Pearson Chi-Square	124.36
Pearson Chi-Square / DF	4.97

When modeling count data with a Poisson distribution, the **Pearson Chi-Square / DF** value should be roughly 1.0. A value greater than 1.0 indicates the data are over-dispersed, i.e., the variance is greater than the mean. In general, a value greater than 2.0 requires remedial action.

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	colony	0.1539	0.09886

Colony is a **random** effect and a non-zero estimate indicates that it affects the results.

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
solvent	4	16	210.20	<.0001

The ANOVA table shows that the effect of **solvent** on insect counts is highly significant. Remedial steps to correct the over-dispersion problem are unlikely to change the inference for **solvent** when the **F** value is this high, but may affect other tests.

solvent Least Squares Means							
solvent	Estimate	Standard Error	DF	t Value	Pr > t	Mean	Standard Error Mean
BlankPaperControl	3.4685	0.1913	16	18.13	<.0001	32.0900	6.1399
Hexane	4.9336	0.1794	16	27.50	<.0001	138.87	24.9155
Methanol	5.1229	0.1788	16	28.66	<.0001	167.83	30.0030
MethanolControl	3.8429	0.1866	16	20.60	<.0001	46.6597	8.7059
NoItemControl	0.8744	0.3283	16	2.66	0.0170	2.3975	0.7871

This column of LS-mean estimates is on the model scale. They are derived from the $-\log$ likelihood function. SAS always lists the model scale LS-means when creating tables of LS-mean tests.

This column of LS-mean estimates have been converted back to the data scale by the **inverse link** function. These values are estimates of the average counts for each treatment level. When reporting results, replace the corresponding model LS-mean values in the test tables with these estimates.

T Grouping for solvent Least Squares Means (Alpha=0.05)		
LS-means with the same letter are not significantly different.		
solvent	Estimate	
Methanol	5.1229	A
Hexane	4.9336	B
MethanolControl	3.8429	C
BlankPaperControl	3.4685	D
NoItemControl	0.8744	E

Letter groups for LS-mean differences. Note that the estimates are on the model scale.

A Negative Binomial Distribution Model

While these results look good, the **Pearson Chi-Square / DF** value of **4.97** is problematic. Ideally, this value should be roughly 1.0. In general, a value higher than 2.0 is evidence of over-dispersion (Stroup, 2015). Over-dispersed data can lead to underestimated standard errors and inflated test statistics (Gbur, et al, 2012).⁴

A **scale** parameter can be added to the Poisson distribution that may account for over-dispersion, but this is not a recommended course of action. Some may have seen or used the GLIMMIX statement that adds the parameter, but I am not going to discuss it. In my experience, it rarely works and re-introduces the problems associated with pseudo-likelihoods.

The best remedial option is to switch to a negative binomial distribution. As mentioned previously, some think the negative binomial distribution should be the first choice when modeling count data. The program listed below analyzes the same data set using a negative binomial distribution.

```
proc glimmix method=quad;
  class colony solvent;
  model counts = solvent / dist=negbin link=log;

  random intercept / subject=colony;
  lsmeans solvent / lines ilink;
run;
```

Change the **dist=** option to **negbin** to use the negative binomial distribution. Both Poisson and negative binomial use the **log** link function.

The **lsmeans** statement in GLIMMIX will produce the same output as PROC PLM for these two options.

⁴ Pages 149 – 160 contain very good information about over-dispersion.

Output Tables for the Negative Binomial Distribution

Fit Statistics	
-2 Log Likelihood	214.76
AIC (smaller is better)	228.76
AICC (smaller is better)	235.35
BIC (smaller is better)	226.03
CAIC (smaller is better)	233.03
HQIC (smaller is better)	221.42

The Fit Statistics are much smaller for the negative binomial (NB) model. AICC is 235.5 versus 294.7 for the Poisson. AICC is the preferred fit statistic for model comparisons because it **corrects** for the number of model parameters.

Fit Statistics for Conditional Distribution	
-2 log L(counts r. effects)	202.62
Pearson Chi-Square	23.24
Pearson Chi-Square / DF	0.93

The measure of over-dispersion dropped from 4.97 to 0.93. Indicating that over-dispersion is no longer a problem.

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	colony	0.1144	0.09052
Scale		0.1023	0.04496

The scale parameter measures the magnitude of over-dispersion and is analogous to the mean squared error in a normal theory analysis of variance. (Gbur, et al).

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
solvent	4	16	49.25	<.0001

The F value is smaller than for the Poisson model, but still highly significant. This is supporting evidence that over-dispersion can lead to inflated F values.

solvent Least Squares Means							
solvent	Estimate	Standard Error	DF	t Value	Pr > t	Mean	Standard Error Mean
BlankPaperControl	3.5482	0.2237	16	15.86	<.0001	34.7502	7.7732
Hexane	4.9668	0.2123	16	23.40	<.0001	143.56	30.4769
Methanol	5.1321	0.2118	16	24.23	<.0001	169.38	35.8701
MethanolControl	3.7738	0.2210	16	17.07	<.0001	43.5473	9.6253
NoltemControl	0.9237	0.3504	16	2.64	0.0180	2.5187	0.8825

5

The LS-mean estimates are slightly different for the two models. Small differences for the model scale estimates are magnified after applying the **inverse link** function to get the data scale estimates. The standard errors are slightly higher for the NB model. Which is not surprising because over-dispersion can underestimate standard errors.

T Grouping for solvent Least Squares Means (Alpha=0.05)		
LS-means with the same letter are not significantly different.		
solvent	Estimate	
Methanol	5.1321	A
		A
Hexane	4.9668	A
MethanolControl	3.7738	B
		B
BlankPaperControl	3.5482	B
NoltemControl	0.9237	C

The NB model separates the LS-means differently than the Poisson model does. Instead of each value receiving a different letter, the means are in groups of the same magnitude (143.56, 169.38), (34.75, 43.55), (2.52). Perhaps these groups are more logical given the nature of the treatments.

The improvement in the Pearson Chi-Squared / DF and AICC values indicate that it is better to model data from this experiment with the negative binomial distribution. Using the correct distribution gives unbiased F values and standard error estimates, and better LS-mean separation tests. While some think that the negative binomial should be the first choice for count data, I usually model the Poisson

⁵ The **log** inverse link function is equivalent to taking the inverse log (e^x) of the model scale LS-mean estimates. Converting standard error values from the model scale to the data scale is much more difficult.

distribution first and then switch if the diagnostics warrant it. This helps me better understand the data set and possibly the cause of other problems that may arise.

What if specifying the negative binomial distribution does not decrease over-dispersion to an acceptable level?

- Examine the data to see if any treatment levels or replications are sparse (few non-zero values), or have constant values. It may be necessary to sacrifice some treatment levels with poor data to get tests for the levels with better data.
- If there does not appear to be a data problem, perhaps selecting the model with better diagnostics and interpretable results is the best course of action.

Describing the analysis:

The response data were analyzed with an analysis of variance model using a negative binomial distribution in PROC GLIMMIX (SAS/STAT, 2017).

A Repeated Measures Analysis for Count Data

PROC GLIMMIX uses the **residual** option in a **random** statement to model data from a repeated measures experiment with normal (Gaussian) response data. This doesn't work for count data because the statement:

```
random day / residual type=cs subject=rep*trt;
```

is incompatible with methods **quadrature** and **laplace**, and one of those two methods should be used to minimize the $-\log$ likelihood function to fit a **conditional** model. If **method** is changed to **quad** or **laplace**, GLIMMIX will abort because it cannot model **R-side** effects and prints this error message:

```
ERROR: R-side random effects are not supported for METHOD=LAPLACE. [or QUAD]
```

The correct approach is to model the repeated effects from the **G-side**. For this example, removing **residual** from the statement above will model the repeated effects in the **G** matrix. The following code analyzes a repeated measures experiment with reps, treatments and days.⁶

```
proc glimmix method=laplace plots=residualpanel;  
  class rep trt day;  
  model counts = trt day trt*day / dist=poisson link=log;  
  random intercept / subject=rep;  
  random day / subject=rep*trt type=cs;  
run;
```

This is similar to treating the repeated measure as a **split-plot in time**, which is how the effects were often modeled with PROC GLM.

⁶ The complete program appears in Appendix B.

Other strategies for handling repeated measures might be to sum (not average) the counts over the time or space variable and analyze the totals, or analyze the counts within each time or space measurement. Summing the counts may have additional benefits if the responses are sparse.

Common Problems and Some Solutions

Notes about zero covariance parameter estimates were discussed in the PROC MIXED seminar for modeling random effects. PROC GLIMMIX prints a similar note, but prints it in the output instead of the SASLOG as PROC MIXED does.

Estimated G matrix is not positive definite.

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	rep	0	.

This message...

Results from the zero covariance estimate for this random effect.

In general, a zero estimate for a random effect indicates that it does not change how the fixed effects influence the response variable. The covariance estimate for **rep** probably came up negative, so GLIMMIX sets it to zero because variance estimates should not be negative. The zero estimate does not adversely affect fixed effect tests or tests for the LS-means. The results should be the same if **rep** is left in the random statement or removed.

Unusual LS-mean tests and lines display. Treatment levels or combinations that have all zero counts can lead to unexpected and often unpredictable results. One outcome that is particularly troublesome and may be hard to recognize is shown below.⁷

The first indication that something is wrong is that the LS-mean test table is very wide when using the default HTML output. The second indication is the note printed along the bottom of the table explaining that there are more differences between LS-means than are shown. Third, the main effect may be significant, but all of the means receive the same letter.

⁷ The complete program appears in Appendix C.

Conservative T Grouping for trt1*trt2 Least Squares Means (Alpha=0.05)			
LS-means with the same letter are not significantly different.			
trt1	trt2	Estimate	
b	13	2.7515	A
			A
a	13	2.5123	A
			A
b	7	2.0369	A
			A
b	9	1.8971	A
			A
a	7	1.8971	A
			A
a	9	-15.8836	A

The LINES display does not reflect all significant comparisons. The following additional pairs are significantly different: (b 13,b 7), (b 13,b 9), (b 13,a 7).

These results are contradictory. LS-means (b 13 and b 7) are assigned the same letter, but the note at the bottom says the means are different. The same goes for (b 13, b9) and (b 13, a 7). The note at the bottom of the table should be a red flag that this is an unusual mean separation table.

Obs	rep	trt1	trt2	counts
1	1	a	7	8
2	2	a	7	5
3	3	a	7	7
4	1	a	9	0
5	2	a	9	0
6	3	a	9	0
7	1	a	13	12
8	2	a	13	16
9	3	a	13	9

The Data Set.

The data values for trt1='a' and trt2=9 are zero. These zeros are the cause of the unusual mean separation table shown above and the tables and results that follow. Sparse data is a common problem with count data and it can lead to all sorts of problems. Some of which are hard to spot, as we will see.

While the unusual interaction means table is the easiest one to notice, other output tables give hints that the results are suspicious.

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	rep	-173E-20	.

-173E-20 is a very small number. One would expect such a small value to be set to zero instead of -0.0000000000000000173. When running the analysis without changing the values to zero, this estimate *is* zero.

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt1	1	10	0.00	0.9913
trt2	2	10	6.24	0.0174
trt1*trt2	2	10	0.03	0.9658

The ANOVA table looks OK. Note that trt2 is significant at alpha=0.05.

trt2 Least Squares Means							
trt2	Estimate	Standard Error	DF	t Value	Pr > t	Mean	Standard Error Mean
7	1.9670	0.1529	10	12.87	<.0001	7.1492	1.0929
9	-6.9932	811.85	10	-0.01	0.9933	0.000918	0.7453
13	2.6319	0.1099	10	23.95	<.0001	13.9004	1.5275

The LS-mean estimate for trt2=9 is very small (0.0009) and has a very large model scale standard error (811.85).

Differences of trt2 Least Squares Means						
trt2	_trt2	Estimate	Standard Error	DF	t Value	Pr > t
7	9	8.9602	811.85	10	0.01	0.9914
7	13	-0.6649	0.1883	10	-3.53	0.0054
9	13	-9.6252	811.85	10	-0.01	0.9908

The pairwise t-tests table shows that treatment 7 and 13 are different. This is consistent with the significant main effect test for **trt2**.

Conservative T Grouping for trt2 Least Squares Means (Alpha=0.05)		
LS-means with the same letter are not significantly different.		
trt2	Estimate	
13	2.6319	A
		A
7	1.9670	A
		A
9	-6.9932	A
The LINES display does not reflect all significant comparisons. The following additional pairs are significantly different: (13,7).		

The note along the bottom says the ls-means for treatments 13 and 7 are different, which agrees with the pairwise t-test. Yet all the means share the same letter, indicating not difference.

The lesson from this example is to check all the output tables to make sure the results are consistent. Do not look at one or two tables and assume the analysis is correct just because SAS assigned letters for mean separation tests. If something in the output looks unusual, there may be some underlying problems with the data, analysis or both.

In my experience, sparse data is the most common problem when analyzing count data. When treatment levels or combinations have data values that are all zeros, maximum likelihood estimation techniques cannot estimate zero for those levels. The estimates may be very small, but cannot be zero. This can also lead to large standard errors and suspicious or unusual results. Even if there are a few non-zero data values, be aware that the results may still be suspicious.

I think that deleting treatment levels with zero counts is a good solution. LS-means for the other treatment levels are always tested to see if they are different from zero. So, this test will determine if the estimable means are different from the treatment level that had zero counts.

Summing counts over dates, locations or treatment levels is another possible tactic for dealing with sparse data. This may decrease the scope of the experiment, but gain testable hypotheses.

Final notes

These questionable tables of LS-mean tests are not limited to count data and PROC GLIMMIX. I have seen similar results from PROC MIXED and PROC GLM when testing LS-means for normal (Gaussian) data.

It can be difficult to diagnose problems from large multi-factor experiments with messy data. Keep this in mind when designing an experiment, especially when working with non-normal response variables.

Convergence

As mentioned before, the mixed model procedures use maximum likelihood estimation to estimate parameters that minimize a $-\log$ likelihood function. The method works by iteratively trying different estimates and checking for improvement in the likelihood function. When no more improvement is found, the procedure has converged and the last set of estimates are the model parameters. Sometimes the procedure encounters mathematical problems because of data problems, a wrongly specified model or both and fails to converge. ***Non-convergence is like a lack-of-fit test – it implies that the data does not support the model.***

When MIXED or GLIMMIX fails to converge, the procedure stops and usually prints a cryptic message in the SASLOG or output file. Some common ones are:

NOTE: An infinite likelihood is assumed in iteration 0 because of a nonpositive definite estimated R matrix for 'variable names' 'place'.

NOTE: Did not converge.

Here is a new one I saw recently:

ERROR: Insufficient resources to determine number of quadrature points adaptively. The last successful evaluation was for 3 points and achieved relative accuracy of 0.000528.

Anyway, in the event of non-convergence, the procedure will stop, maybe print a message, and SAS will continue executing the remaining (if any) program statements. Always check that each procedure completed successfully if you are in the habit of submitting large programs with many procedures.

Here are some problems that may cause non-convergence and some possible remedies.

- If it is a repeated measures analysis, make sure there is one observation for each subject at each time point. A data coding error often causes this problem.
- Response values are very large numbers, e.g., 10,000 or 1,000,000. Divide all these values by a common factor to reduce their scale. Dividing by a constant will not affect the results, only the measurement scale.
- An over parameterized model. If there are many random effects, remove some random interaction terms to try to stabilize the variance. I never include 3-way interactions in model or random statements. Three-way interactions may be needed to identify subjects for the **type=** option however.
- The model is not specified correctly. For a standard mixed model ANOVA, always list class variables in the class statement; list fixed effects in the class and model statements; list random effects in the class and random statements; list repeated effects in the class, model and repeated statements. *These guidelines may not apply when fitting complex models.*
- If there are many subjects, combining them into similar groups (making group the subject) may be helpful. For example, instead of individual cows, maybe pens or lots can characterize a herd and there may be fewer parameters to estimate.
- Check that the values for a treatment level are not all zeros or have identical values. Maximum likelihood cannot estimate zero (to my knowledge), and treatment levels with zero variance are

problematic. This usually only happens with discrete response data, like counts or binomial data, not with continuous (normal) response data.

- Change the maximum likelihood estimation method. This is more useful for GLIMMIX than for MIXED and will be discussed in an upcoming seminar.
- Sometimes the procedure will stop after reaching the maximum number of iterations. Increasing the number of iterations is rarely helpful because some underlying problem is causing the convergence failure.

Sometimes the model just will not converge - finding no solution. It may be necessary to redefine the problem or sacrifice data to salvage parts of an experiment. Be careful about manipulating the data too much and always be prepared to explain what changes were made and the rationale for making them.

References

Gbur, Edward E., Walter W. Stroup, Kevin S. McCarter, Susan Durham, Linda J. Young, Mary Christman, Mark West, Matthew Kramer. 2012. Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America. Madison WI. doi: 10.2134/2012.generalized-linear-mixed-models.c2.

Kiernan, Kathleen, Jill Tao, Phil Gibbs. 2016. "Tips and Strategies for Mixed Modeling with SAS/STAT® Procedures," Proceedings of the SAS Global Forum 2016 Conference, Cary NC: SAS Institute Inc.

SAS/STAT Software version 14.1. 2017. SAS Institute Inc., Cary NC.

Stroup, Walter W. (2015), "Rethinking the Analysis of Non-Normal Data in Plant and Soil Science," *Agron. J.* 107:811-827. Doi:10.2134/agronj2013.0342

Stroup, Walter W. 2013. *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. CRC Press.

Tobias, Randy, Weiji Cai. (2010), "Introducing PROC PLM and Postfitting Analysis for Very General Linear Models in SAS/STAT® 9.22," Proceedings of the SAS Global Forum 2010 Conference, Cary NC: SAS Institute Inc.

Appendix A

CountSeminar1.sas – Analysis count data (ants) from an entomology experiment. The initial model used a Poisson distribution, but the negative binomial distribution provided a better fit.

```

/*
/  Two solvents were used to wash pheromones off ant pupae onto
/  pieces of paper. These pieces of paper, along with paper from
/  three controls, were presented to five colonies of ants and
/  the number of ants attracted to each treatment was recorded.
/
/  A randomized complete block with five treatments and five replications.
/
/  A good example demonstrating how a negative binomial distribution is
/  better than a Poisson for over dispersed count data.
/
/  Data courtesy of Ben Gochnour and Dan Suiter
/
/  Mar 27, 2018
/=====*/

```

```
ods html close;
ods html;
```

```

data one;
  length solvent $17;
  infile datalines firstobs=2;
  input solvent $ colony counts;
datalines;

```

Solvent	Colony	counts
NoItemControl	2	2
NoItemControl	3	1
NoItemControl	4	5
NoItemControl	5	2
NoItemControl	6	3
BlankPaperControl	2	27
BlankPaperControl	3	48
BlankPaperControl	4	16
BlankPaperControl	5	39
BlankPaperControl	6	44
MethanolControl	2	35
MethanolControl	3	14
MethanolControl	4	27
MethanolControl	5	49

MethanolControl	6	128
Methanol	2	162
Methanol	3	80
Methanol	4	128
Methanol	5	214
Methanol	6	326
Hexane	2	185
Hexane	3	118
Hexane	4	81
Hexane	5	116
Hexane	6	253

```
;
```

```
proc print;
run;
```

```
proc glimmix method=quad plots=residualpanel;
  class colony solvent;
  model counts = solvent / dist=poisson link=log;
  random intercept / subject=colony;
  lsmeans solvent / ilink;
  store quad1;
```

```
run;
```

```
proc plm restore=quad1;
  lsmeans solvent / ilink lines;
run;
```

```
proc glimmix method=quad plots=residualpanel;
  class colony solvent;
  model counts = solvent / dist=negbin link=log;
  random intercept / subject=colony;
  lsmeans solvent / lines ilink;
run;
```

Appendix B

CountRepeatedSeminar1.sas – A repeated measures analysis for simulated count data.

```

/*
/  Simulated data from a Poisson distribution. Three replications (rep),
/  three treatments (trt)
/  and three evenly spaced repeated measurements (day).
/
/  This program models the repeated effect covariance from the G-side.
/
/  Mar 27, 2018
/=====*/

ods html close;
ods html;

data one;
  call streaminit(213); /* initialize RN stream so that it returns */
                       /* the same numbers with each run          */
  do day = 4, 8, 12;
    do rep = 1, 2, 3;
      do trt = 7, 9, 13;
        counts = rand('poisson', trt);
        output;
      end;
    end;
  end;
run;

proc print;
run;

/* Add day as a repeated effect.          */
/* R-side effects don't work with Quad or Laplace */

proc glimmix method=laplace plots=residualpanel;
  class rep trt day;
  model counts = trt day trt*day / dist=poisson link=log;
  random intercept / subject=rep;
  random day / subject=rep*trt type=cs;
*  random day / residual type=cs subject=rep*trt;
*  lsmeans trt / ilink lines;
run;

```

Appendix C

PoissonBadMeans.sas – This program has zeros for data values in one treatment combination. It produces the wide, contradictory LS-mean test tables.

```

/*
/   Simulated data from a Poisson distribution.
/
/   Was GoodPoisson.sas. Trying to generate data that will cause
/   The wide, bogus lines display.
/   Added trt2 to test two-way means.
/
/   Mar 27, 2018
/=====*/

ods html close;
ods html;

data one;
  call streaminit(213); /* initialize RN stream so it          */
  do rep = 1, 2, 3;     /* returns the same numbers with each run */
    do trt1 = 'a', 'b';
      do trt2 = 7, 9, 13;
        counts = rand('poisson', trt2);
        if trt1='a' and trt2=9 then counts=0; * create zero counts;
        output;
      end;
    end;
  end;
run;

proc sort;
  by trt1 trt2;
run;
proc print;
run;

proc glimmix method=quad;
  class rep trt1 trt2;
  model counts = trt1 trt2 trt1*trt2 / dist=poisson link=log;
  random intercept / subject=rep;
  lsmeans trt1 trt2 trt1*trt2 / pdiff ilink lines;
run;
/*
proc plm gives the same results as GLIMMIX;
proc plm restore=bad1;
  lsmeans trt1 trt2 trt1*trt2 / pdiff ilink line;
run; */

```