Estimation of the dispersion parameter

The overdispersion parameter is unknown and therefore must be estimated. There are two suggested methods of estimating the overdispersion parameter.

McCullagh (1983) gives the estimate as

$$\hat{\phi} = \frac{(y-\hat{\mu})' V_{\mu}^{-1} (y-\hat{\mu})}{N-p} = \frac{\text{Pearsons } \chi^2}{N-p}$$

where N-p is the df for lack of fit and $\mathbf{V}_{\mu}$ is the diagonal matrix of variance functions.

McCullagh and Nelder (1989) suggest using the deviance

$$\hat{\phi} = \frac{\text{Deviance}}{N-p}$$

So for the two models that we have investigated so far, the estimates of the dispersion parameters using the two methods are:

Model	**Using Pearsons**	**Using Deviance**
1	1.01	.64
2	.79	.59

PROC GENMOD fixes the scale parameter at 1.0, unless you specify that you want to use the dispersion parameter. If the dispersion parameter is significantly greater than one, indicating overdispersion (variance greater than the mean), then the scale parameter should be used to adjust the variance. Failing to account for the overdispersion can result in inflated test statistics. However, when the dispersion parameter is less than one, then the test statistics become more conservative, which is not considered as much of a problem.

PROC GLIMMIX

SAS has available an experimental GLIMMIX procedure available for download. This GLIMMIX procedure replaces the GLIMMIX macro that was previously available. The GLIMMIX procedure is easier to use and has a manual available online. The manual is 220 pages long.

To download the GLIMMIX procedure, go to the following web address

http://support.sas.com/rnd/app/da/glimmix.html

Click on the **download** button. This will bring you to a page where you can download the procedure. Before you can download the software, SAS requires that you register.

There is a button on the right to click to go through the registration process. If you are already a registered SAS user, you will need to enter your email address and password. Once you are logged into SAS, you can download the software. Click on the **request download** button. This will download an executable file for installing the procedure into your SAS 9.1 software. If you do not have SAS 9.1 on your computer, you can purchase a license for SAS 9.1 from the Statistics Department for $40, where you will check out a set of CD's for installing SAS 9.1.

Model 3 - fixed effects of baseline and treatment, random center

The model that we are fitting with Model 3 is the following

$\log(\mu_{ij}/(1-\mu_{ij}))=a + \beta x_{ij}+\tau_i +C_j$

This looks just like Model 2, but now we are considering centers as random rather than fixed. The assumption is

$C_j \sim N(0,\sigma^2_c)$

Model 3 can be fit with the following SAS program

```
proc glimmix data=ldbp;
class trt center;
model cfb/one=cf1b trt/ddfm=satterth cl;
lsmeans trt/diff pdiff;
random center;
run;
```

In the model statement the CL option requests a t type CI be constructed for each fixed effect parameter. Below is selected output from the analysis.

```
                  The GLIMMIX Procedure

                           Model Information

             Data Set                          WORK.LDBP
             Response Variable (Events)        cfb
             Response Variable (Trials)        one
             Response Distribution             Binomial
             Link Function                     Logit
             Variance Function                 Default
             Variance Matrix                   Not blocked
             Estimation Technique              Residual PL
             Degrees of Freedom Method         Satterthwaite
```

First we look at the model information to make sure that we have set up the model correctly. The variance function is the default variance function for a bernoulli (μ(1- μ)). PROC GLIMMIX does allow you to define a different variance function. The variance function is not blocked. In other words, we did not do any grouping of the observations. The class information is listed next.

```
                    Class Level Information

 Class       Levels     Values

 trt              3     A B C
 center          29     1 2 3 4 5 6 7 8 9 11 12 13 14 15 18 23 24 25
                        26 27 29 30 31 32 35 36 37 40 41


             Number of Observations Read          283
             Number of Observations Used          283
             Number of Events                      41
             Number of Trials                     283


                           Dimensions

                 G-side Cov. Parameters         1
                 Columns in X                   5
                 Columns in Z                  29
                 Subjects (Blocks in V)         1
                 Max Obs per Subject          283
```

The dimension give information about the sizes of the X and Z design matrices. Note that X has one column for each fixed effect and Z has one column for each center effect. The next information presented is the optimization information. The default optimization is the Quisi-Newton method. The iteration history displays information about the progress of the optimization process. After the initial optimization, the GLIMMIX procedure performed 6 updates before the convergence criterion was met.

```
                    Optimization Information

           Optimization Technique         Dual Quasi-Newton
           Parameters in Optimization     1
           Lower Boundaries               1
           Upper Boundaries               0
           Fixed Effects                  Profiled
           Starting From                  Data
```

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Max Change	Gradient
0	0	1	1341.5683928	2.00000000	3.241048
1	0	3	1475.3792347	2.00000000	0.000068
2	0	3	1524.8573372	0.89553871	0.000114
3	0	3	1525.1622045	0.07024552	2.49E-6
4	0	2	1524.9247004	0.00481471	1.48E-8
5	0	1	1524.8876009	0.00004290	1.249E-9
6	0	0	1524.8872551	0.00000000	2.949E-7

The fit statistics are output next. The word "Pseudo" preceding the different statistics indicates that these are computed from a pseudo-likelihood. Just like with the previous models, the ratio of the Pearson Chi-Square over degrees of freedom gives us the information on the variability. The ratio close to 1 indicates that the variability has been properly modeled and that there is no residual overdispersion.

Fit Statistics

-2 Res Log Pseudo-Likelihood	1524.89
Pseudo-AIC (smaller is better)	1526.89
Pseudo-AICC (smaller is better)	1526.90
Pseudo-BIC (smaller is better)	1528.25
Pseudo-CAIC (smaller is better)	1529.25
Pseudo-HQIC (smaller is better)	1527.32
Pearson Chi-Square	269.88
Pearson Chi-Square / DF	0.97

The covariance parameter estimates are output next. We only have the random effect of center in our model. The amount of variation accounted for by center is very small and the estimate is not significantly different from 0.

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error
center	0.05830	0.2451

The solutions for the fixed effects follow

Solutions for Fixed Effects

Effect	trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
Intercept		-3.3346	0.5180	279	-6.44	<.0001	0.05
cf1b		2.9324	0.4896	279	5.99	<.0001	0.05
trt	A	0.9307	0.6002	279	1.55	0.1221	0.05
trt	B	1.7014	0.5720	279	2.97	0.0032	0.05
trt	C	0	.	.	.	.	.

Solutions for Fixed Effects

Effect	trt	Lower	Upper
Intercept		-4.3544	-2.3149
cf1b		1.9686	3.8961
trt	A	-0.2509	2.1123
trt	B	0.5754	2.8275
trt	C	.	.

Because the variance for center was close to zero, the solutions for the fixed effects from this model are very similar to the solutions that we obtained with Model 1, which was a fixed effect model with only treatment and the baseline covariate in the model.

The final output contains the Type III tests of the fixed effects and the treatment least squares means.

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
cf1b	1	279	35.88	<.0001
trt	2	279	4.78	0.0091

trt Least Squares Means

trt	Estimate	Standard Error	DF	t Value	Pr > |t|
A	-2.1242	0.3501	126.7	-6.07	<.0001
B	-1.3534	0.2888	49.93	-4.69	<.0001
C	-3.0549	0.5071	279	-6.02	<.0001

Differences of trt Least Squares Means

trt	_trt	Estimate	Standard Error	DF	t Value	Pr > |t|
A	B	-0.7707	0.4399	279	-1.75	0.0809
A	C	0.9307	0.6002	279	1.55	0.1221

```
   B       C          1.7014      0.5720      279        2.97       0.0032
```

Again, because the variance for center was close to zero, the tests and least squares means are very similar to the solutions that we obtained with Model 1.

Model 4 - fixed effects of baseline and treatment, random center and center*trt

The final model that we will look at is the model with both center and center*trt included as random effects.

The model that we are fitting with Model 3 is the following

$\log(\mu_{ij}/(1-\mu_{ij}))=a + \beta x_{ij}+\tau_i +C_j+C\tau_{ij}$

The assumptions are

$C_j \sim N(0,\sigma^2_c)$

$C\tau_{ij} \sim N(0,\sigma^2_{ct})$

Model 4 can be fit with the following SAS program

```
proc glimmix data=ldbp;
class trt center;
model cfb/one=cf1b trt/ddfm=satterth cl;
lsmeans trt/diff pdiff;
random center center*trt;
run;
```

The model information and class level information are the same as for Model 3.

```
                    The GLIMMIX Procedure

                      Model Information

         Data Set                        WORK.LDBP
         Response Variable (Events)      cfb
         Response Variable (Trials)      one
         Response Distribution           Binomial
         Link Function                   Logit
         Variance Function               Default
         Variance Matrix                 Not blocked
         Estimation Technique            Residual PL
         Degrees of Freedom Method       Satterthwaite


                   Class Level Information

    Class      Levels    Values

    trt             3    A B C
```

```
center         29    1 2 3 4 5 6 7 8 9 11 12 13 14 15 18 23 24 25
                        26 27 29 30 31 32 35 36 37 40 41

                   Number of Observations Read         283
                   Number of Observations Used         283
                   Number of Events                     41
                   Number of Trials                    283
```

The dimension of the X matrix is the same as for Model 3. However, the dimension of the Z matrix is now 107, which is equal to the number of center*treatment subclasses that have data (3*29-9).

Dimensions

G-side Cov. Parameters	2
Columns in X	5
Columns in Z	107
Subjects (Blocks in V)	1
Max Obs per Subject	283

The optimization information tells us that we have two parameters (center and center*trt) that we are optimizing and the iteration history shows that it took 11 rounds of iteration before the convergence criteria was met. Usually, the more parameters that you are trying to estimate, the more rounds of iteration before convergence.

Optimization Information

Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2
Upper Boundaries	0
Fixed Effects	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	0	2	1335.0964241	0.78925560	5.46922
1	0	6	1460.4378836	1.14998764	3.47755
2	0	3	1485.2181157	0.26645692	2.827735
3	0	3	1488.7860435	0.06082887	2.694558
4	0	3	1488.3706433	0.00802877	2.678094
5	0	2	1488.2777156	0.00083941	2.676403
6	0	1	1488.2657691	0.00011646	2.676353
7	0	1	1488.2640536	0.00004953	2.676192
8	0	1	1488.2647853	0.00002108	2.67626
9	0	1	1488.264474	0.00000897	2.676231
10	0	1	1488.2646065	0.00000382	2.676244
11	0	0	1488.2645501	0.00000000	2.676242

Convergence criterion (PCONV=1.11022E-8) satisfied.

This statement concerning the estimated G matrix indicates that one of the estimated variance parameters was negative. This is an underestimate of the true variance component. As you recall from the section on normal mixed models, one of the reasons that this can occur is when the number of observations per random effect category is small. Another reason is if the ratio of the true variance component to the residual is small. From Model 3, when we included just center as a random effect, our estimate of the variance component was not different from zero.

```
Estimated G matrix is not positive definite.
```

The fit statistics show the ratio of the Pearson Chi-Square over degrees of freedom is somewhat less than zero. PROC GLIMMIX automatically scales the variance with the overdispersion parameter for generalized linear mixed models.

Fit Statistics

-2 Res Log Pseudo-Likelihood	1488.26
Pseudo-AIC (smaller is better)	1490.26
Pseudo-AICC (smaller is better)	1490.28
Pseudo-BIC (smaller is better)	1491.63
Pseudo-CAIC (smaller is better)	1492.63
Pseudo-HQIC (smaller is better)	1490.69
Pearson Chi-Square	215.00
Pearson Chi-Square / DF	0.77

The estimates of the two variance parameters are presented next. Because the estimate for the center was less than 0, SAS set the estimate to zero. The center*trt variance estimate indicates that the center*treatment effect does not account for a significant source of variation. The solutions for the fixed effects, tests for fixed effects and least squares means from Model 4 are again very close to what was obtained with Model 1, which does not have the problems associated with the effects of uniform categories.

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error
center	0	.
trt*center	0.5470	0.4894

Solutions for Fixed Effects

Effect	trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
Intercept		-3.3545	0.5465	199.4	-6.14	<.0001	0.05
cf1b		2.9367	0.5206	279	5.64	<.0001	0.05
trt	A	1.0180	0.6520	114	1.56	0.1212	0.05
trt	B	1.6998	0.6314	94.64	2.69	0.0084	0.05
trt	C	0	.	.	.	.	.

Solutions for Fixed Effects

Effect	trt	Lower	Upper

Intercept		-4.4321	-2.2768
cf1b		1.9118	3.9615
trt	A	-0.2736	2.3095
trt	B	0.4463	2.9533
trt	C	.	.

The GLIMMIX Procedure

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
cf1b	1	279	31.82	<.0001
trt	2	66.99	3.69	0.0302

```
                        trt Least Squares Means
                     Standard
  trt     Estimate      Error        DF     t Value     Pr > |t|
  A        -2.0563     0.3874        65       -5.31       <.0001
  B        -1.3745     0.3424        35       -4.01       0.0003
  C        -3.0743     0.5354       190       -5.74       <.0001

                    Differences of trt Least Squares Means

                             Standard
trt     _trt     Estimate       Error        DF     t Value     Pr > |t|

A       B         -0.6818      0.5139     46.34       -1.33       0.1911
A       C          1.0180      0.6520       114        1.56       0.1212
B       C          1.6998      0.6314     94.64        2.69       0.0084
```

Using the results from Model 4 as an example, we can again look at the significance tests and construct confidence intervals from the linear scale to the odds ratio scale, which is the scale of measure. We do this by exponentiating the estimate of the treatment differences and also exponentiating the confidence intervals constructed on the linear scale. To calculate the confidence intervals, we first construct the 95% CI on a linear scale for (A-B), which would be

95% CI= $-0.6818 \pm t_{46,0.975} \times 0.5139 = -.6818 \pm 2.01 \times 0.5139 = (-1.715, 0.351)$

The following table prestents the estimated effects and confidence intervals on both the linear and as an odds ratio

Effect	Linear Scale	Odds Ratio
A-B	-0.6818 (-1.715,0.351)	.51 (0.18,1.42)
A-C	1.0180 (-0.280,2.309)	2.77 (.76,10.06)
B-C	1.6998 (.443,2.956)	5.47 (1.56,19.22)

When we look at confidence intervals on the linear scale, we look to see if the range includes 0. If the range does not include 0, such as with B-C, then the treatment response is different. However, it is hard to interpret these logit values. The confidence intervals on the odds ratio scale are based on an odds ratio of 1. If the confidence interval includes 1 in the range, then the odds of the treatment responses are not different. However, if 1 is not included in the range (either both ends of the confidence interval are above 1 or below 1, then the odds of treatment responses are different.