Generalized Linear Mixed Models (GLMM)

GLMMs are based on GLM, extended to include random effects, random coefficients and covariance patterns.

GLMMs are a fairly new class of models. Research is still being carried out and there is not much software available to analyse.

Linear mixed models and when implied assumptions not appropriate

$$y|\alpha \sim N(X\beta + Z\alpha, R)$$

$\alpha \sim N(0,G)$

G and R covariance matrices may depend on a set of unknown variance components.

Linear mixed model assumes

1. the relationship between the mean of the dependent variable y and the fixed and random effects can be modeled as a linear function
2. the variance is not a function of the mean
3. the random effects follow a normal distribution

Any or all of these assumptions may be violated for certain traits

An example of this, pregnancy rate, would probably violate the assumption of a linear relationship between the dependent variable and the fixed and random effects.

Pregnancy is a zero/one trait, at any given timepoint. Pregnancy rate is a herd measure of the number of cows pregnant/total number of cows and can only range between 0 and 1. If a change in management practice in a herd with a pregnancy rate of .5, increases that rate by .1 to .6, you would not expect that same change in management practice in a herd with a pregnacy rate of .9 to increase the rate by that same amount. In other words, a treatment effect or environmental effect would be expected to have a greater effect when the mean rate is smaller, than when the mean rate is closer to 1. The second assumption, that the variance is not a function of the mean, is also questionable with pregnancy rate. If the predicted pregnancy rate, μ, for a cow is .5, the variance $\mu(1-\mu)=.25$. If the predicted pregnancy rate is .8, the variance=.16. So for some production traits, the variance increases as the mean level of production increases.

Historically, a number of options have been used to try and address the problem of using linear mixed models, even when the use is not correct. These include log transformations, linear and multiplicate adjustments, or just ignoring the fact that the linear mixed model is not correct and using it anyway. These options are appealing because they are

relatively simple and cheap to implement. However, they sidestep the issue that the linear mixed model is not the correct model for the data.

The GLMM gives extra flexibiltiy in developing an appropriate model.

GLMM definition

$y=\mu + e$

Where μ is the vector of expected means of the y observations and is linked to the parameter by a link function, *g*.

With the GLM, *g* was defined

$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$

With GLMM, the link function, *g*

$\mathbf{g(\mu) = X\beta + Z\alpha}$

y=dependent variable
μ=expected values
e=residual error
X=design matrix for fixed effects
Z=design matrix for random effects
β=fixed effect parameters
α=random effect parameters

The random effects are assumed to follow a normal distiribution,
$\boldsymbol{\alpha \sim N(0,G)}$

Where **G** is the same as we defined under normal mixed models.

So just like the normal mixed models, we can write the variance matrix

$var(y)=\mathbf{V}=var(\boldsymbol{\mu}) + \mathbf{R}$

R = var(e), which is dependent on **μ.**

Note that both the var(**μ**) and **R** do not have closed form solutions. As a result, the sampling properties of the test statistics and estimators will only be approximate. In other words, a 95% confidence interval may in fact be either an 80% or 99% confidence interval. P-values and standard errors may also be too large or too small.

Inverse link function

In the GLM section, we were introduced to canonical link function as a way to map the original data to the linear predictor of the model ($g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$). The linear predictor can be transformed to the observed scale through an inverse link function. In other words, the inverse link function is used to map the value of the linear predictor for observation *i*, to the conditional mean for observation *i*, μ_i. To get the inverse link function start out with the link function

$\log(\mu/1-\mu) = g(\mu) = X\beta$

If we exponentiate each side

$\mu/1-\mu = e^{X\beta}$

$\mu = e^{X\beta}(1-\mu) = e^{X\beta} - e^{X\beta}\mu$

$\mu + e^{X\beta}\mu = e^{X\beta}$

$\mu(1+e^{X\beta}) = e^{X\beta}$

$\mu = e^{X\beta}/(1+e^{X\beta})$ which is the inverse link function

which will be denoted as $g^{-1}(g(\mathbf{\mu}))= g^{-1}(\mathbf{X\beta})$.

Therefore, **μ** depends on the linear predictor through an inverse link function and the covariance matrix **R** depends on **μ** through the variance function.

The following table presents the link, inverse link and variance functions

Distribution	**$g(\mu)=b'^{-1}(\mu)$**	**$g^{-1}(g(\mu))$**	**var(μ)**	**Name**
Normal	μ	μ	1	Identity
Bernoulli	log(μ/(1-μ))	$e^{g(\mu)}/(1+e^{g(\mu)})$	μ(1-μ)	Logit
Binomial	log(μ/(1-μ))	$e^{g(\mu)}/(1+e^{g(\mu)})$	μ(1-μ)/n	Logit
Poisson	log(μ)	$e^{g(\mu)}$	μ	Log
Poisson with offset	log(μ)	$e^{g(\mu)}$	μ/t	Log

Looking at the logit link function further:

Linear Predictor (g(μ))	**Corresponding Mean**	**Difference**
-4= log(μ/(1-μ))	.02	
-3= log(μ/(1-μ))	.05	.03
-2= log(μ/(1-μ))	.12	.07
-1= log(μ/(1-μ))	.27	.15
0= log(μ/(1-μ))	.5	.23
1= log(μ/(1-μ))	.73	.23
2= log(μ/(1-μ))	.88	.15
3= log(μ/(1-μ))	.95	.07
4= log(μ/(1-μ))	.98	.03

We see that an increase in the linear predictor results in an increase in the mean, but not at a constant rate. Also note that the logit link function will always yield estimated means in the range of 0 to 1.

Variance Function

The variance function is used to model non-systematic variability. With GLMs, residual variability arises from two sources, the variability from the sampling distribution and the variability due to over-dispersion. The over-dispersion can be modeled in a number of ways. When we covered GLMs, we discussed the scale or dispersion parameter, ϕ, which can increase or decrease the variance in the model from the observation variances

$Var(y_i|\alpha) = \phi v(\mu_i)$.

A second approach is to add an additional random effect, $e_i \sim N(0, \phi)$, to the linear predictor for each observation.

A third approach is to select another distribution. For example, using a two parameter (μ,ϕ) negative binomail distribution in place of a one parameter Poisson distribution for count data.

Notice that all three of these approaches involve the estimation of an additional parameter, ϕ.

Summary of the parts

Generalized linear mixed models are composed of three parts

1. Linear predictor, $g(\boldsymbol{\mu})=\mathbf{X\beta + Z\alpha}$, used to model the relationship between the fixed and random effects. The residual variability contained in the residual, $\boldsymbol{e}$, of the linear mixed model equation is incorporated in the variance function of the GLMM.

2. An inverse link function, $\mu_i = g^{-1}(g(\boldsymbol{\mu}))$, is used to model the relationship between the linear predictor and the conditional mean of the observed trait. The link function is selected to be both simple and reasonable.

3. A variance function, $v(\mu_i, \phi)$, is used to model the residual variability. Although we looked at three possible approaches, the simpliest approach and the one used the most is using the dispersion parameter as a scaling factor; $Var(y_i|\alpha) = \phi v(\mu_i)$.

Example

We are going to look at a portion of the adverse event data associated with the multi-center trial that we have been working with. The adverse event is “cold feet”. In the study the occurrence of “cold feet” was recorded at each visit on a 1-5 scale, but for this example, the data will be analyzed as a binary variable from the observation from the last visit. “Cold feet” was also recorded at baseline. In order to include a baseline covariate in the model (reduce between-patient variation), the data will be analyzed in Bernoulli form, where each patient is recorded as either a "success" (cold feet) or a "failure" (no cold feet). Remember that the Bernoulli is a special case of the Binomial, where n=1.

The results presented here will be different from what is presented in the book. The dataset that the book uses to test the models is not the same dataset that they describe. The dataset described has 41 events of cold feet out of 283 patients. The dataset that the analyze has 39 events out of 279 patients, with 4 missing values. I have not been able to determine where the discrepancies occurs and have emailed the author to see if he can clarify.

The following SAS program reads in the data, finds the last observation for each patient, recodes the 1-5 scale to the 0,1 scale for both the baseline value and the final value for "cold feet"

```
options ps=80 ls=64;
filename bp 'C:\users\kathy\statistics department\statistics 892- mixed
models\downloaded stuff\brown and prescott\bp.dat';

data dbp;
infile bp;
input pat visit center trt $ dbp dbp0 cf cf1;
if cf ne .;
run;

*sort the data by pat and visit for the last record carried forward;

proc sort;
by pat visit;
run;

*get the last record for each patient for the last record carried forward;
*and also code the cold feet 1,2 to 0 and cold feet 3,4,5 to 1;
*add a dummy variable one=1 for all observations;

*if the baseline value cf1 is missing, drop that patient;

data ldbp;
set dbp;
by pat;
if last.pat;
one=1;
if cf1=. then delete;
if cf in (1,2)then cfb=0;
else if cf in (3,4,5) then cfb=1;
if cf1 in (1,2)then cf1b=0;
else if cf1 in (3,4,5) then cf1b=1;
run;

*print out the first 16 observations of the dataset;

data prntl;
set ldbp;
if _n_<17;

proc print ;
run;
```

Obs	pat	visit	center	trt	dbp	dbp0	cf	cf1	one	cfb	cf1b
1	1	5	29	C	89	97	1	1	1	0	0
2	3	5	5	B	111	117	5	5	1	1	1
3	4	6	5	A	87	100	3	1	1	1	0
4	5	6	29	A	85	105	3	3	1	1	1
5	7	6	3	A	100	114	1	2	1	0	0
6	8	5	3	B	85	105	2	1	1	0	0
7	9	6	3	B	90	100	1	1	1	0	0
8	10	3	3	A	100	102	1	1	1	0	0
9	11	6	3	C	94	105	1	1	1	0	0
10	12	5	3	C	80	105	1	1	1	0	0
11	13	6	36	B	80	100	4	1	1	1	0
12	14	6	36	A	85	100	1	1	1	0	0
13	15	6	36	C	80	100	1	1	1	0	0
14	18	6	36	A	100	100	1	1	1	0	0
15	19	6	5	B	102	100	1	1	1	0	0
16	21	5	5	B	96	106	5	1	1	1	0

The following SAS code summarizes by treatment and center, the frequency of cold feet, which are presented in the next table.

```
proc sort data=ldbp;
by center;

proc freq data=ldbp;
by center;
table (cfb)*trt/list;
run;
```

Center	Treatment A	Treatment B	Treatment C	Total
1	3/13	5/14	1/12	9/13
2	2/3	0/4	0/3	2/10
3	0/3	0/3	0/2	0/8
4	1/ 4	1 /4	0/4	2/12
5	1/ 4	3/5	0/2	4/11
6	0/2	1/1	1 /2	2/5
7	0/6	1/6	0/6	1/18
8	1/ 2	0/1	1/ 2	2/5
9	-	-	0/1	0/1
11	0/4	1 /4	0/4	1/12
12	0/3	1 /3	0/4	1/10
13	1/1	0/1	0/2	1 /4
14	0/8	2/8	1 /8	3/24
15	1 /4	0/4	0/3	1/11
18	0/2	0/2	0/2	0/6
23	1/1	-	0/2	1/3
24	-	-	0/1	0/1
25	0/3	0/2	0/2	0/7
26	0/3	1 /4	0/3	1/10
27	-	1/1	0/1	1 /2
29	1/1	-	0/1	1/ 2
30	0/1	0/2	0/2	0/5
31	0/12	0/12	0/12	0/36
32	1 /2	0/1	0/1	1 /4
35	0/2	0/1	-	0/3
36	0/9	5/6	0/8	5/23
37	0/2	0/1	1/ 2	1/5
40	0/1	1/1	-	1 /2
41	0/2	0/1	0/1	0/4
total	13/98	23/92	5/93	41/283

Note that there are several zero frequencies, which will lead to uniform center and center*treatment categories, which in turn may cause variance component bias and therefore it is not clear whether a random effects model will fit well. Just as we did with the normal linear models, we will fit a variety of models which are presented in the following table. The book also looks at a fixed effect model with basline, treatment, center and the center*treatment interaction term. Because so many of the centers have uniform effects, this model does not work with these data. Models 3 and 4 presented here are the same as the Models 4 and 5 presented in the book.

Model	Fixed effects	Random effects	Method
1	baseline,treatment	---	GLM
2	baseline,treatment,center	---	GLM
3	baseline,treatment	center	pseudo-likelihood
4	baseline,treatment	center,center*treatment	pseudo-likelihood

The pseudo-likelihood is a method for fitting the GLMM that was proposed by Wolfinger and O'Connell (1993) and is used by GLIMMIX in SAS. It is an iterative procedure and is called 'pseudo-likelihood' because the likelihood function maximized at each iteration is that of the pseudo variable and not the original data. See Section 3.2.3 in the text book for more details.

Model 1 - fixed effects of baseline and treatment

The model that we are fitting with Model 1 is the following

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

This model assumes that the slopes are the same for each treatment. The following SAS code is used for Model 1. There will be 5 parameters; 1 for the intercept, one for the slope and one for each of the 3 treatments. Contrast statements to test for treatment differences are constructed. The treatment LSMs and associated standard errors are given in the logit scale. If you want the equivalent of the LSM for the probability of a favorable outcome, you need to use the inverse link function. This can be accomplished in SAS by using the ODS to output the treatment LSMs and then using the inverse link function. You can also approximate the standard error by taking the $(\mu_i/(1-\mu_1))$ times the estimated standard error.

```
*model 1 fixed effect model including baseline and treatment;
proc genmod data=ldbp;
class trt;
model cfb/one=cf1b trt/dist=b type1;
contrast 'A-B' trt 1 -1 0;
contrast 'A-C' trt 1 0 -1;
contrast 'B-C' trt 0 1 -1;
lsmeans trt/pdiff;
ods output lsmeans=lsm;

data prob_hat;
set lsm;
phat=exp(estimate)/(1+exp(estimate));
se_phat=phat*(1-phat)*stderr;
proc print data=prob_hat;
run;
```

Note that we have designated a Binomial distribution. However, because n=1, we are really using a Bernoulli distribution.

The GENMOD Procedure

Data Set	WORK.LDBP
Distribution	Binomial
Link Function	Logit
Response Variable (Events)	cfb
Response Variable (Trials)	one
Number of Observations Read	283
Number of Observations Used	283
Number of Events	41
Number of Trials	283

Check to make sure that the model information is what you expect, including the distribution, link function, variables, and number of events and trials.

Class Level Information

Class	Levels	Values
trt	3	A B C

Parameter Information

Parameter	Effect	trt
Prm1	Intercept	
Prm2	cf1b	
Prm3	trt	A
Prm4	trt	B
Prm5	trt	C

Check the above to make sure the class information is what you expect. Note that, as expected, there are 5 parameters.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	279	178.1744	0.6386
Scaled Deviance	279	178.1744	0.6386
Pearson Chi-Square	279	282.4176	1.0122
Scaled Pearson X2	279	282.4176	1.0122
Log Likelihood		-89.0872	

You can see from the goodness-of-fit that there is no evidence of lack of fit.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-3.3532	0.5156	-4.3637	-2.3427	42.30	<.0001

cf1b		1	2.9697	0.4858	2.0176	3.9218	37.37	<.0001
trt	A	1	0.9361	0.5999	-0.2397	2.1120	2.43	0.1187
trt	B	1	1.7043	0.5717	0.5837	2.8248	8.89	0.0029
trt	C	0	0.0000	0.0000	0.0000	0.0000	.	
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
cf1b	1	40.97	<.0001
trt	2	10.91	0.0043

The covariate of the baseline of "cold feet" has a significant effect on the proportion of cold feet at the end of the study. There is also a significant treatment effect. The probability reported here is the Wald statistic.

Least Squares Means

Effect	trt	Estimate	Standard Error	DF	Chi-Square	Pr > ChiSq
trt	A	-2.1337	0.3445	1	38.35	<.0001
trt	B	-1.3656	0.2812	1	23.58	<.0001
trt	C	-3.0699	0.5046	1	37.01	<.0001

These estimates are in the logit scale. Later we will see the output from the code where we applied the inverse link function to the estimates.

Differences of Least Squares Means

Effect	trt	_trt	Estimate	Standard Error	DF	Chi-Square	Pr > ChiSq
trt	A	B	-0.7681	0.4392	1	3.06	0.0803
trt	A	C	0.9361	0.5999	1	2.43	0.1187
trt	B	C	1.7043	0.5717	1	8.89	0.0029

There is a significant treatment differences between treatments B and C. You cannot convert the estimates of the treatment differences to the probability scale using the inverse link. You just get nonsense. You can interpret the treatment differences as the log of the odds-ratio between the two treatments. The odds ratios for the three differences are

A-B	.46389
A-C	2.55002
B-C	5.49753

To interpret these, you would say for the B-C, that the odds of having "cold feet" are approximately five times higher for treatment B than for treatment C.

Contrast Results

Contrast	DF	Chi-Square	Pr > ChiSq	Type
A-B	1	3.16	0.0757	LR
A-C	1	2.62	0.1056	LR
B-C	1	10.75	0.0010	LR

Obs	Effect	trt	Estimate	StdErr	DF	ChiSq	Prob ChiSq	phat	se_phat
1	trt	A	-2.1337	0.3445	1	38.35	<.0001	0.10586	0.032612
2	trt	B	-1.3656	0.2812	1	23.58	<.0001	0.20333	0.045557
3	trt	C	-3.0699	0.5046	1	37.01	<.0001	0.04437	0.021396

These are the least-squares mean equivalents for the probability of a favorable outcome for each treatment. The estimated probability of "cold feet" under treatment A is .10.

Model 1 does not take into account differences among centers. In the next model the centers will be treated as a fixed effect.

Model 2 - Treatment and Center as fixed effects with baseline covariate

The model for Model 2 is

The model that we are fitting with Model 2 is the following

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

Although Model 2 appears like a logical model, there will be some problems with this model because of the large number of uniform categories within each center. In other words, there are many centers, where the response to all three treatments is the same. When there are uniform fixed effects a corresponding effect estimate on the linear scale cannot be estimated. The consequences of this will be noted in the output.

```
proc genmod data=ldbp;
class trt center;
model cfb/one=cf1b trt center/dist=b type3;
contrast 'A-B' trt 1 -1 0;
contrast 'A-C' trt 1 0 -1;
contrast 'B-C' trt 0 1 -1;
lsmeans trt/pdiff;
run;
```

```
                    The GENMOD Procedure

                      Model Information

              Data Set                     WORK.LDBP
              Distribution                  Binomial
              Link Function                    Logit
              Response Variable (Events)         cfb
              Response Variable (Trials)         one


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


                     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
```

A check of the model information and class information shows that our SAS statements were correct.

Just as with Model 1, we get a listing of all of the parameters. In Model 1, we just had 5 parameters. With Model 2, with the addition of Center as a fixed effect, we now have 34 parameters. Following is the first several records of the SAS output for the parameter information.

Parameter Information

Parameter	Effect	trt	center
Prm1	Intercept		
Prm2	cf1b		
Prm3	trt	A	
Prm4	trt	B	
Prm5	trt	C	
Prm6	center		1
Prm7	center		2
Prm8	center		3
Prm9	center		4
Prm10	center		5

etc.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	251	147.7812	0.5888
Scaled Deviance	251	147.7812	0.5888
Pearson Chi-Square	251	199.3547	0.7942
Scaled Pearson X2	251	199.3547	0.7942
Log Likelihood		-73.8906	

Again, the goodness-of-fit statistics shows no evidence of lack of fit. As we look at the output further, we find the following warning.

```
WARNING: Negative of Hessian not positive definite.
```

This is the first indication that there is a problem with the analyses. If we look at the log file for this run, we see the following warning

```
WARNING: The negative of the Hessian is not positive definite. The
         convergence is questionable.
WARNING: The procedure is continuing but the validity of the model
         fit is questionable.
WARNING: The specified model did not converge.
WARNING: Negative of Hessian not positive definite.
```

This was caused by the problem of having uniform fixed effects. The impact of this, is that the information on treatments is lost from all centers where there were no cold feet and from any center with only one treatment. This causes the treatment estimates (presented below) to be different from those of Model 1 and the standard errors to be larger.

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	-26.5793	2.1509	-30.7950	-22.3635	152.70	<.0001
cf1b		1	2.6711	0.5548	1.5838	3.7584	23.18	<.0001
trt	A	1	1.0470	0.6477	-0.2224	2.3164	2.61	0.1060
trt	B	1	2.0302	0.6294	0.7967	3.2637	10.41	0.0013
trt	C	0	0.0000	0.0000	0.0000	0.0000	.	.
center	1	1	23.4900	2.1070	19.3604	27.6197	124.29	<.0001
center	2	1	23.2142	2.2258	18.8516	27.5768	108.77	<.0001
center	3	1	-0.1467	110060.4	-215715	215714.3	0.00	1.0000
center	4	1	23.2940	2.2582	18.8680	27.7201	106.40	<.0001
center	5	1	24.1606	2.1989	19.8508	28.4704	120.72	<.0001
center	6	1	25.3009	2.2938	20.8051	29.7966	121.66	<.0001
center	7	1	21.8392	2.3609	17.2120	26.4665	85.57	<.0001
center	8	1	24.7352	2.2914	20.2442	29.2262	116.53	<.0001
center	9	1	1.2139	322114.2	-631331	631333.5	0.00	1.0000
center	11	1	22.6362	2.3220	18.0852	27.1872	95.04	<.0001
center	12	1	22.1269	2.4024	17.4183	26.8355	84.83	<.0001
center	13	1	24.5423	2.4163	19.8066	29.2781	103.17	<.0001
center	14	1	23.1055	2.1751	18.8423	27.3687	112.84	<.0001
center	15	1	22.9148	2.3279	18.3521	27.4775	96.89	<.0001
center	18	1	-0.0486	126467.5	-247872	247871.8	0.00	1.0000
center	23	1	24.3186	2.7168	18.9938	29.6434	80.12	<.0001
center	24	1	1.2139	322114.2	-631331	631333.5	0.00	1.0000
center	25	1	-0.0211	117643.3	-230577	230576.5	0.00	1.0000
center	26	1	22.9963	2.3328	18.4242	27.5685	97.18	<.0001
center	27	1	25.5642	2.6333	20.4029	30.7254	94.24	<.0001
center	29	1	24.7202	2.9282	18.9811	30.4594	71.27	<.0001
center	30	1	-0.0857	137672.4	-269833	269832.9	0.00	1.0000
center	31	1	-0.0486	51630.15	-101193	101193.2	0.00	1.0000
center	32	1	23.3668	2.4567	18.5518	28.1817	90.47	<.0001
center	35	1	-0.2451	183098.8	-358867	358866.7	0.00	1.0000

Also, we did not get the contrasts or LSMs that were requested.
Because of the uniform effects across many of the centers, Model 2 is not recommended for estimating treatment effect. However, the model can be used to test for fixed center effect by using a likelihood ratio test. To test center effect we calculate twice the difference in the minus log likelihoods of the two models and compare that to a χ^2 with 28 degrees of freedom, which is the difference in the number of parameters between the two models.

2(89.0872)-2(73.8906)=178.1744-147.7812=30.3932. .50>p>.25

χ^2_{28} =27.3 for .50
χ^2_{28} =32.5 for .25

So treating center as a fixed effect, there is no significant center effect.