

STAT 565 – Fall 2005 – Solutions to Assignment 7

1. (a) The following SAS code was used. Note that x1 is transformed to be 1 for the active drug and zero for the placebo. Glimmix estimates the log odds of an abnormal electrocardiogram because Y is coded 1 for abnormal and zero for a normal electrocardiogram.

```
/* Analyze the data from a crossover
   trial on cerebrovascular deficiency
   from example 8.1 in DHLZ */

/* THE VARIABLES WERE CODED AS FOLLOWS:

   ID = CLUSTER VARIABLE

   CLASS = 1, NEEDED TO RUN GEE2

   Y = 1 abnormal electrocardiogram
       0 normal electrocardiogram

   Int = 1 (INTERCEPT)

   X1 = 1 placebo (treatment B)
       0 active drug A

   X2 = 1 if period 2
       0 if period 1

   X12 = X1*X2

   order = 1 if B before A
           0 if A before B

REFERENCE: JONES AND KENWARD(1989)
           CHAPMAN AND HALL, P.9 */

data set1;
  infile "c:\st565\dhlz.example8_1.dat";
  input patient class y int x1 x2 x12 order ;
  /* Recode the variables */
  x1=abs(1-x1);
  x12=x1*x2;
  run;

/* Fit a model use GLMMIX to include
   a random location effect */
```

```

%inc 'c:\mydocuments\courses\st557\sas\glmm800.sas' /
      nosource;
      run;

/* Logistic regression with random
   effects for transect and sub-transects */

%glimmix(data=set1,
  stmts=%str(
    class patient ;
    model y = x1 x2 x12 / covb
    solution ddfm=kr;
    random intercept / subject=patient g
              gcorr solution;),
  error=binomial, link=logit,
  converge=1e-8, maxit=20, out=setp
  )
run;

proc print data=setp (obs=5); run;

```

Estimates for fixed effects parameters are given below:

Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	0.8109	0.8703	65	0.93	0.3549
x1	4.4901	1.3409	65	3.35	0.0014
x2	-0.1688	1.2320	65	-0.14	0.8915
x12	-3.2772	2.4097	65	-1.36	0.1785

$\beta_1$  is the logarithm of the ratio of conditional odds obtained by dividing the odds of having an abnormal electrocardiogram when the active drug is given in the first period by the odds of having an abnormal electrocardiogram when the placebo is given in the first period. The estimate of the corresponding conditional odds ratio is  $\exp(\hat{\beta}_1) = \exp(4.4901) = 89.13$ , and an approximate 95% confidence interval for this conditional odds ratio is (6.436, 1234.3). This conditional odds ratio is not well estimated, but there is evidence that the active drug increases the risk of an abnormal electrocardiogram.

$\beta_2$  is the logarithm of the ratio of conditional odds obtained by dividing the odds of having an abnormal electrocardiogram in period 1 when the placebo is given by the odds of having an abnormal electrocardiogram in the second period when the placebo is given. The estimate of the corresponding conditional odds ratio is  $\exp(-.1688) = 0.84$  and an approximate 95% confidence interval for this conditional odds ratio is (.076, 9.45). There is no significant period effect for the placebo treatment.

$\beta_3 = [(\beta_0 + \beta_1 + \beta_2 + \beta_3) - \beta_0 + \beta_2] - [(\beta_0 + \beta_1) - \beta_0]$  is a difference in two log-odds ratios. The logarithm of the ratio of conditional odds obtained by dividing the odds of having an abnormal electrocardiogram when the active drug is given in the second period divided by the odds of having an abnormal electrocardiogram when the placebo is given in the second period minus the logarithm of the ratio of conditional odds obtained by dividing the odds of having an abnormal electrocardiogram when the active drug is given in the first period divided by the odds of having an abnormal electrocardiogram when the placebo is given in the first period. The estimate of the corresponding ratio of conditional odds ratios is  $\exp(\hat{\beta}_3) = \exp(-3.2772) = .038$ , and an approximate 95% confidence interval for this ratio of conditional odds ratios is (.000335, 4.25). The estimate of this ratio of conditional odds ratios is not significantly different from one.

Including the insignificant interaction between drug effects and period in the model seems to distort the estimation of the effect of the active drug. It is worthwhile in this case to fit a logistic regression model with conditionally additive period and treatment effects, i.e., fit the model

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \eta_i$$

The estimated parameters for this model are shown below:

Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	1.6141	0.6388	72	2.53	0.0137
x1	2.7607	0.4559	68.9	6.06	<.0001
x2	-1.7089	0.4548	69.2	-3.76	0.0004

These results indicate that in either period the odds of an abnormal electrocardiogram with the active drug are about  $\exp(2.7607) = 15.8$  times greater than the the odds of an abnormal electrocardiogram with the placebo, with approximate 95% confidence interval (6.5, 38.6). These results also indicate a significant period effect. Within either treatment group, the odds of an abnormal electrocardiogram in the second period are about 18% of the odds of an abnormal electrocardiogram in the first period.

Take a look at the data. When the placebo is given the observed proportions of abnormal electrocardiograms are .606 and .647 in the first and second periods, respectively. When the active drug is given the observed proportions of abnormal electrocardiograms are .824 and .667 in the first and second periods, respectively. Is there an overall increase in the incidence of abnormal electrocardiograms when the active drug is used or does the increase only occur in the first period? Do the results from either of the models fit about accurately reflect the pattern in the data?

- (b) Given the value of the random effect for the  $i$ -th patient, the conditional distribution for the binary responses is

$$y_{ij}|\eta_i \sim \text{Bin}(1, \pi_{ij})$$

It follows that the conditional variance is

$$\text{var}(y_{ij}|\eta_i) = \pi_{ij}(1 - \pi_{ij})$$

- (c) The GLIMMIX macro used to fit the models in part (a) assumes that the random patient effects are independent and identically distributed according to a normal distribution with mean zero and unknown variance  $\sigma_\eta^2$ . Define

$$f(\eta_i; \sigma_\eta^2) = \frac{1}{\sqrt{2\pi\sigma_\eta}} e^{-(\eta_i^2)/(2\sigma_\eta^2)}$$

Then

$$E(Y_{ij}) = Pr(Y_{ij} = 1) = \int Pr(Y_{ij} = 1|\eta_i)f(\eta_i; \sigma_\eta^2) d\eta_i = \int \pi_{ij}f(\eta_i; \sigma_\eta^2) d\eta_i$$

where

$$\pi_{ij} = \frac{\exp(\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij1} X_{ij2} + \eta_i)}{1 + \exp(\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij1} X_{ij2} + \eta_i)}$$

Then,

$$\text{var}(y_{ij}) = E(y_{ij}^2) - (E(y_{ij}))^2 = p(y_{ij} = 1) - p(y_{ij} = 1)^2 = \int \pi_{ij}f(\eta_i; \sigma_\eta^2) d\eta_i - \left( \int \pi_{ij}f(\eta_i; \sigma_\eta^2) d\eta_i \right)^2$$

- (d) For the first model fit in part(a), the approximate REML estimate of  $\sigma_\eta^2$  is 19.16. For the second model fit in part (a), the approximate REML estimate of the variance of the patient effects is 18.91, almost the same estimate.
- (e) The results for fitting marginal logistic models on pages 820 and 823 of the notes seem to better reflect the effects of the active drug on incidence of abnormal electrocardiograms. The GLIMMIX results for the first model fit in part (a) seem to overstate drug and period effects. GLIMMIX estimates a very large variance for the random patient effects. This may occur because random patient effects cannot be distinguished from drug and period effects in this case. For each of the four distinct sets of values for  $(X_1, X_2)$  there are only two types of patients, those with abnormal electrocardiograms ( $Y=1$ ) and those with normal electrocardiograms ( $Y=0$ ). When the response is binary and there are only a few repeated measures on each subject, the estimation procedure used by GLIMMIX can produce biased results (see Breslow and Clayton, 1993, JASA, 9-25). It was not a good idea to use GLIMMIX for this analysis.
- (f) In this case,  $X_{ij1} = 1, X_{ij2} = 1, X_{ij1}X_{ij2} = 1$ . Since you do not know the value of the random effect for any subject, the best thing to do is set it equal to its mean, which is zero. Then

$$\hat{\pi}_{ij} = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3)} = \frac{\exp(1.855)}{1 + \exp(1.855)} = 0.865$$

To obtain a standard error, first obtain the standard error of the logit. Do not ignore the correlations among the parameter estimates or the variation in the random patient effect. The reported REML estimate of the covariance matrix of the parameter estimates is

Covariance Matrix for Fixed Effects

Row	Effect	Col1	Col2	Col3	Col4
1	Intercept	0.7588	-0.7588	-0.7588	1.4234
2	x1	-0.7588	1.8125	1.3295	-3.0478
3	x2	-0.7588	1.3295	1.5199	-2.7552
4	x12	1.4234	-3.0478	-2.7552	5.8267

The sum of the variances and covariances in the table shown above provides an estimate of the variance of  $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$ . This value is 0.7825. Since the variance of the random patient effect is estimated as 19.16, an estimate of

$$Var(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + \eta_{36})$$

is  $0.7825+19.16=19.9425$  and the estimated standard error is 4.466. Applying the delta method, the standard error the conditional probability of success is estimated as  $\hat{\pi}_{ij}(1 - \hat{\pi}_{ij}) * 4.466 = 0.52$ . This is a large standard error for a predicted probability.

(g) When the active drug is given in the first period, we estimated

$$\beta_0 + \beta_1 + \beta_2 + \beta_3 + \eta_i$$

as

$$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + 0 = 1.855$$

The REML estimate of the variance of the random patient effects is 19.16. If the random patient effects have a normal distribution, then values of the random patient effects have probability .95 of being within two standard deviations of zero, the mean of the random patient effects. Consequently,, random patient effects are likely to be between -8.579 and 8.579 and

$$\beta_0 + \beta_1 + \beta_2 + \beta_3 + \eta_i$$

is likely to be between  $1.855-8.579=-6.724$  and  $1.855+8.579=10.434$ . Probabilities are likely to be between .0012 and .9997, which is a result of no practical value. Using a random effects model was not a good idea for this cross-over study, at least when the interaction between period and treatment is included in the model. DHLZ were able to get reasonable results using a random effects model, on pages 176-177, when they considered a model with just a drug effect or a model where the logits were linked to additive drug and period effects. They did not consider a model that also included a drug\*period interaction like they did in Chapter 8.

2. (a) Fit independent poisson counts at the each time point. SAS code for plotting mean counts across time for the six treatments and fitting a separate log-linear model at each time point is shown below,

```
/* Enter the papalloma data */

data set1;
  infile 'c:\mydocuments\mpap.dat';
  input mouse trt diet surgery y week;
run;

proc format; value trt
  1 = 'adlib/no surgery'
  2 = 'adlib/ADX'
  3 = 'DER/no surgery'
  4 = 'DER/ADX'
  5 = 'adlib/ADX-GCH'
  6 = 'DER/ADX-GCH';
run;

proc sort data=set1; by trt week; run;

proc means data=set1 noprint;
  by trt week;
  var y;
  output out=means mean=y;
run;

proc print data=means;
  run;

axis1 label=(f=swiss h=1.8 a=90 r=0 "Papalloma Rate")
  order = 0 to 15 by 5
  value=(f=swiss h=1.8) w=3.0
  length= 4.0in;

axis2 label=(f=swiss h=2.0 "Time(weeks)")
  order = 10 to 30 by 4
  value=(f=swiss h=1.8) w= 3.0
  length = 6.5 in;

SYMBOL1 V=CIRCLE H=1.7 w=3 l=1 i=join ;
SYMBOL2 V=DIAMOND H=1.7 w=3 l=3 i=join ;
SYMBOL3 V=square H=1.7 w=3 l=9 i=join ;
SYMBOL4 V=dot H=1.7 w=3 l=2 i=join ;
SYMBOL5 V=triangle H=1.7 w=3 l=4 i=join ;
SYMBOL6 V=' ' H=1.7 w=3 l=10 i=join ;
```

```

PROC GPLOT DATA=means;
  PLOT y*week=trt /
        vaxis=axis1 haxis=axis2;
  TITLE1 ls=0.01in H=2.0 F=swiss "Average Papalloma Counts";
        footnote ls=0.01in;
  format trt trt.;
  RUN;
  proc sort data=set1;
  by week;

/* Fit a separate log-linear model to
  compare mean counts for the six
  treatments at each time point. */

proc genmod data=set1;
by week;
  class mouse trt;
  model y = trt / itprint
        dist=poisson link=log covb;
run;

```

Week14 is shown as an example.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr>ChiSq
Intercept	1	-2.5902	0.7284	-4.0178	-1.1627	12.65	0.0004
trt 1	1	4.9152	0.7386	3.4676	6.3627	44.29	<.0001
trt 2	1	4.5550	0.7510	3.0831	6.0270	36.79	<.0001
trt 3	1	2.8381	0.8207	1.2295	4.4467	11.96	0.0005
trt 4	1	3.3231	0.7936	1.7677	4.8786	17.53	<.0001
trt 5	1	1.7917	0.8422	0.1411	3.4424	4.53	0.0334
trt 6	0	0.0000	0.0000	0.0000	0.0000		

This table shows that the first five treatments are significantly different from treatment 6 with respect to the mean number of papallomas at week 16. Treatment 6 is protective. Papalloma counts for treatment 6 are relatively low across all time points. You could examine comparisons with other treatments by re-running the analysis 5 times with each of the other treatments given the highest code on one of the runs.

(b) Results from the analysis of the counts at week 14 are shown below:

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	204	839.0620	4.1130

Scaled Deviance	204	167.0390	0.8188
Pearson Chi-Square	204	1024.7226	5.0232
Scaled Pearson X2	204	204.0000	1.0000
Log Likelihood		105.8918	

If the expected counts are large enough for a normal distribution to be a good approximation to the Poisson distribution and the counts have independent Poisson distributions for the mice in the six treatment groups, then the Pearson goodness of fit statistic should have an approximate central chi-square distribution with 204 df. At week 14, the observed value of the Pearson statistics is five times larger than the mean of a central chi-square random variable with 204 degrees of freedom. There is more variation among papalloma counts within treatment groups than can be attributed to counts with Poisson distributions, i.e., the within treatment group variances are much larger than the corresponding mean counts.

To allow for the extra variance, you could abandon the Poisson distribution and use a model where the within treatment variance in the papalloma counts is allowed to be 5.0232 times larger than the mean count for each treatment. This can be achieved by putting the `pscale` option in the `GENMOD` as shown below

```
proc genmod data=set1;
by week;
  class mouse trt;
  model y = trt / itprint covb
  dist=poisson link=log;
run;
```

Results for week 14 are shown below. Note that the parameter estimates are unchanged. They are still obtained by maximizing a Poisson likelihood function (which is incorrect, but it produces consistent estimates of the parameters in this case). Standard errors of the parameter estimates have been inflated by a factor of  $\sqrt{5.0232} = 2.2412$ . This changes the p-values for the tests of significance. Diets 5 and 6 are no longer significantly different at week 14, for example.

#### Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald	95% Confidence Limits
Intercept	1	-2.5902	1.2940	-5.1264	-0.0541
trt	1	4.9152	1.3001	2.3671	7.4633
trt	2	4.5550	1.3030	2.0012	7.1089
trt	3	2.8381	1.3405	0.2108	5.4654
trt	4	3.3231	1.3189	0.7381	5.9082
trt	5	1.7917	1.3976	-0.9476	4.5311

trt	6	0	0.0000	0.0000	0.0000	0.0000
Scale		0	2.2412			

Analysis Of Parameter Estimates

Parameter		Chi-Square	Pr > ChiSq
Intercept		4.01	0.0453
trt	1	14.29	0.0002
trt	2	12.22	0.0005
trt	3	4.48	0.0342
trt	4	6.35	0.0118
trt	5	1.64	0.1999
trt	6	.	.

Another alternative is to use the negative binomial distribution for the counts. The negative binomial distribution contains another parameter that allows the variance of the counts to be different from the mean. GENMOD code that specifies the negative binomial distribution is shown below:

```
proc genmod data=set1;
by week;
class mouse trt;
model y = trt / itprint covb
dist=negbin link=log;
run;
```

Note that the parameter estimates are slightly changed because they are obtained from maximizing a likelihood function for independent negative binomial counts. Also note the values of the standard errors.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	
Intercept	1	-2.5903	0.6274	-3.8199	-1.3606
trt	1	4.9152	0.6889	3.5650	6.2654
trt	2	4.5550	0.6919	3.1990	5.9111
trt	3	2.8381	0.7024	1.4614	4.2148
trt	4	3.3232	0.6869	1.9769	4.6694
trt	5	1.7918	0.7138	0.3928	3.1907
trt	6	0.0000	0.0000	0.0000	0.0000

Dispersion            1        2.4117        0.4269        1.7047        3.4120

Analysis Of Parameter Estimates

Parameter		Chi-Square	Pr > ChiSq
Intercept		17.05	<.0001
trt	1	50.91	<.0001
trt	2	43.34	<.0001
trt	3	16.33	<.0001
trt	4	23.41	<.0001
trt	5	6.30	0.0121
trt	6	.	.

- (c) Code for fitting some models is shown below. Note that in week 10 there were no papallomas observed for some diets. This cause numerical problems with a Poisson distribution because the variance of the counts is taken to be a multiple of the mean. With time\*diet interactions in the model, we can fit a different mean for each diet at each time point, and the program tries to fit a zero mean when the sample mean is zero for any diet at any time point. The you get a zero on the diagonal of the information matrix and the program cannot proceed when it tries to invert that matrix. To avoid this we added .01 to each count, so the total number of papallomas would never be exactly zero for any diet at any time point.

```
data set2;
  set set1;
  y=y+0.01;
run;

proc genmod data=set2 descending;
class mouse trt week;
model y = trt|week /
  dist=poisson link=log
  itprint pscale
converge=1e-8 maxit=50;
repeated subject=mouse(trt) / type=un
  modelse covb corrw;
run;

proc genmod data=set2 descending;
class mouse trt week;
model y = trt|week /
  dist=poisson link=log
  itprint pscale
converge=1e-8 maxit=50;
```

```

repeated subject=mouse(trt) / type=ar(1)
           modelse covb corrw;
run;

proc genmod data=set2 descending;
class mouse trt week;
model y = trt|week /
      dist=poisson link=log
      itprint pscale
converge=1e-8 maxit=50;
repeated subject=mouse(trt) / type=arh(1)
           modelse covb corrw;
run;

proc genmod data=set2 descending;
class mouse trt week;
model y = trt|week /
      dist=poisson link=log
      itprint pscale
converge=1e-8 maxit=50;
repeated subject=mouse(trt) / type=toep
           modelse covb corrw;
run;

proc genmod data=set2 descending;
class mouse trt week;
model y = trt|week /
      dist=poisson link=log
      itprint pscale
converge=1e-8 maxit=50;
repeated subject=mouse(trt) / type=cs
           modelse covb corrw;
run;

```

You do not get any statistic from GENMOD like the AIC criterion that you can use to choose between correlation structures. One thing you could do is examine the estimate of the correlation matrix for papalloma counts recorded on the same mouse across the six points in time. This matrix is shown below. The pattern does not seem to correspond to an AR(1) structure or an exchangeable structure. Robust estimation yields about the same results regardless of what correlation structure you select.

#### Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5	Col6
Row1	1.0000	0.3780	0.4003	0.3174	0.2894	0.2234
Row2	0.3780	1.0000	0.6826	0.6456	0.6715	0.5341

Row3	0.4003	0.6826	1.0000	0.9420	0.9420	0.8212
Row4	0.3174	0.6456	0.9420	1.0000	0.9420	0.9287
Row5	0.2894	0.6715	0.9420	0.9420	1.0000	0.9420
Row6	0.2234	0.5341	0.8212	0.9287	0.9420	1.0000

- (d) Given the results from part (c), this part was done assuming a general correlation structure for papalloma counts recorded on the same animal across time. We attempted to fit some time trends using polynomial curves. The following code was used.

```

proc genmod data=set2 descending;
class mouse trt;
model y = trt week trt*week week*week
      trt*week*week week*week*week / noint
      dist=poisson link=log
      itprint pscale
converge=1e-8 maxit=50;
repeated subject=mouse(trt) / type=un
      modelse covb corrw;
run;

```

The estimated coefficients are shown below. Obviously, this model could be simplified.

#### Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1241	6225.9401	5.0169
Scaled Deviance	1241	994.0933	0.8010
Pearson Chi-Square	1241	7772.3005	6.2629
Scaled Pearson X2	1241	1241.0000	1.0000
Log Likelihood		537.2774	

#### Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5	Col6
Row1	1.0000	0.3551	0.3398	0.2501	0.2428	0.1682
Row2	0.3551	1.0000	0.6457	0.5654	0.6252	0.4360
Row3	0.3398	0.6457	1.0000	0.9360	0.9360	0.7150
Row4	0.2501	0.5654	0.9360	1.0000	0.9360	0.7823
Row5	0.2428	0.6252	0.9360	0.9360	1.0000	0.9106
Row6	0.1682	0.4360	0.7150	0.7823	0.9106	1.0000

Analysis Of GEE Parameter Estimates  
Empirical Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z
Intercept		0.0000	0.0000	0.0000	0.0000	.
trt	1	-5.8949	1.7946	-9.4122	-2.3776	-3.28
trt	2	-4.0337	1.4202	-6.8173	-1.2501	-2.84
trt	3	-9.2992	2.7210	-14.6322	-3.9662	-3.42
trt	4	-12.1873	4.9231	-21.8365	-2.5382	-2.48
trt	5	-8.8121	2.6214	-13.9499	-3.6742	-3.36
trt	6	-13.5836	4.4573	-22.3197	-4.8475	-3.05
week		1.3407	0.4140	0.5293	2.1521	3.24
week*trt	1	-0.3043	0.3322	-0.9553	0.3468	-0.92
week*trt	2	-0.4683	0.3125	-1.0808	0.1442	-1.50
week*trt	3	-0.1745	0.3152	-0.7923	0.4434	-0.55
week*trt	4	0.0880	0.4835	-0.8596	1.0356	0.18
week*trt	5	-0.3471	0.3068	-0.9484	0.2542	-1.13
week*trt	6	0.0000	0.0000	0.0000	0.0000	.
week*week		-0.0445	0.0135	-0.0709	-0.0180	-3.30
week*week*trt	1	0.0045	0.0065	-0.0082	0.0172	0.70
week*week*trt	2	0.0073	0.0060	-0.0045	0.0190	1.21
week*week*trt	3	0.0033	0.0061	-0.0087	0.0152	0.53
week*week*trt	4	-0.0034	0.0098	-0.0226	0.0158	-0.34
week*week*trt	5	0.0071	0.0059	-0.0043	0.0186	1.22
week*week*trt	6	0.0000	0.0000	0.0000	0.0000	.
week*week*week		0.0005	0.0002	0.0001	0.0008	2.80

3. (a) Fit a marginal Cox proportional hazards model:

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trt	5	1020	32.50	<.0001
week	5	1020	13.66	<.0001
trt*week	25	1020	3.75	<.0001

Solution for Fixed Effects

Effect	trt	week	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept			-2.2119	0.4815	204	-4.59	<.0001
trt	1		4.0654	0.5631	1020	7.22	<.0001
trt	2		2.9173	0.5844	1020	4.99	<.0001
trt	3		2.8258	0.5814	1020	4.86	<.0001

trt	4		0.8526	0.5874	1020	1.45	0.1470
trt	5		0.6045	0.6191	1020	0.98	0.3291
trt	6		0	.	.	.	.
week		10	-3.5115	1.8996	1020	-1.85	0.0648
week		14	-1.3715	0.7189	1020	-1.91	0.0567
week		18	-111E-16	0.4574	1020	-0.00	1.0000
week		22	0.07197	0.4494	1020	0.16	0.8728
week		26	0.1391	0.4423	1020	0.31	0.7532
week		30	0	.	.	.	.
trt*week	1	10	2.5448	1.9055	1020	1.34	0.1820
trt*week	1	14	1.7183	0.7263	1020	2.37	0.0182
trt*week	1	18	0.6584	0.4676	1020	1.41	0.1594
trt*week	1	22	0.5440	0.4599	1020	1.18	0.2372
trt*week	1	26	0.2806	0.4538	1020	0.62	0.5364
trt*week	1	30	0	.	.	.	.
trt*week	2	10	2.7347	1.9083	1020	1.43	0.1522
trt*week	2	14	1.8463	0.7307	1020	2.53	0.0117
trt*week	2	18	0.8767	0.4734	1020	1.85	0.0643
trt*week	2	22	0.5895	0.4668	1020	1.26	0.2068
trt*week	2	26	0.2169	0.4620	1020	0.47	0.6389
trt*week	2	30	0	.	.	.	.
trt*week	3	10	1.5568	1.9237	1020	0.81	0.4186
trt*week	3	14	0.2862	0.7498	1020	0.38	0.7028
trt*week	3	18	-0.1402	0.4835	1020	-0.29	0.7719
trt*week	3	22	-0.00862	0.4734	1020	-0.02	0.9855
trt*week	3	26	-0.01618	0.4660	1020	-0.03	0.9723
trt*week	3	30	0	.	.	.	.
trt*week	4	10	1.7335	1.9407	1020	0.89	0.3719
trt*week	4	14	1.6032	0.7468	1020	2.15	0.0321
trt*week	4	18	0.5500	0.4952	1020	1.11	0.2669
trt*week	4	22	0.4686	0.4879	1020	0.96	0.3371
trt*week	4	26	0.4746	0.4804	1020	0.99	0.3234
trt*week	4	30	0	.	.	.	.
trt*week	5	10	1.4991	1.9755	1020	0.76	0.4481
trt*week	5	14	0.5850	0.7922	1020	0.74	0.460
trt*week	5	18	-0.3524	0.5415	1020	-0.65	0.5153
trt*week	5	22	-0.1491	0.5235	1020	-0.28	0.7758
trt*week	5	26	-0.1642	0.5156	1020	-0.32	0.7502
trt*week	5	30	0	.	.	.	.
trt*week	6	10	0	.	.	.	.
trt*week	6	14	0	.	.	.	.
trt*week	6	18	0	.	.	.	.
trt*week	6	22	0	.	.	.	.
trt*week	6	26	0	.	.	.	.
trt*week	6	30	0	.	.	.	.

The overall effects of time and treatment are significant. There is also an indicate of significant interaction between time and treatments with most of the interaction occurring at the earlier time points. Standard errors are smaller than the robust standard errors provided by the GENMOD procedure in SAS. There is much to sort out here. It may be useful to fit trends across time. Otherwise, a separate analysis comparing the treatments at each time point tells you most of what you need to know about the diet effects.

- (b) Code for fitting a random effect model is shown below. Predictions of the random effects for individual animals can indicate which animals are unusually high or low responders. In this case you can select the mice that appear to be most susceptible to developing papillomas by selecting the mice with the largest positive predictions of their random effects. For this study mice 518, 534, 622, 635, 676, 690, 693, 722, 769, 786, 2623, 2653, 2264, 2681 all have predictions larger than 3.0. If you were really selecting mice for susceptibility to papillomas, it is generally better to treat them all the same. Here you had to rely on a model to adjust for diet effects.

```

%inc 'C:\mydocuments\glmm800.sas' /nosource;
      run;
%glimmix(data=set2,
  stmts=%str(
    class mouse trt week;
    model y = trt|week/
          solution;
    random intercept / solution type=un subject=mouse;),
error=poisson, link=log,
  converge=1e-8, maxit=20, out=setp)
run;

```

Solution for Random Effects

Effect	mouse	Estimate	Std Err		t Value	Pr >  t
			Pred	DF		
Intercept	2664	4.0249	0.3898	1020	10.33	<.0001