GLMM Short Course **Spring 2008**

LAB 4: Notes

The dataset ratlitter gives number of rat pups surviving (numsurv) at day 21 out of the number in the litter (littersize) at day 5 after birth. Each row of the dataset is a litter. Litters with treatment=1 were born to a mother fed a diet with a possibly toxic chemical. Those with treatment=0 are a control group.

1) Using NLMIXED, fit a logistic model with random effects for litter. Is there clustering by litter?

The model is parameterized in terms of log(σ) in order to improve convergence, so it is not possible to look at the Wald test to test if there is clustering. The Wald test is testing if log(σ) is zero, i.e., σ=1. A likelihood ratio test gives a value for the change in –2loglik of 131.3-109.0 = 22.3 and recall that the distribution is ½ of a chi-square with 1 d.f. So, even before cutting the p-value in half the p-value is approximately zero. So, yes, evidence of clustering.

2) Does the treatment decrease the number surviving at day 21?

The Wald test gives a marginal p-value of 0.0861, but the Wald test is known to be less reliable in small sample sizes than the likelihood ratio test. Fitting without the treatment effect gives a –2loglik of 112.1 for a difference of 112.1-109.0 = 2.1, with a p-value of 0.078. So in this case the two tests agree – there is a borderline treatment effect – but hold on.

3) The treatment may also increase the variability from mom to mom. Use NLMIXED to write a model and test for a different litter-to-litter variance in the treatment group (as compared to the control group).

In the parameterization given in the lab handout, the parameter lnratio is the log of the ratio of the mouse to mouse standard deviation in treatment 2 as compared to treatment 1. So a value of zero indicates no difference in standard deviations. The Wald test does not indicate a difference, but the change in –2loglik is 109.0 - 105.4 = 3.6 , which gives a p-value of 0.06, so there is a borderline result. Note that the fixed treatment effect is no longer statistically significant. One way to think about this is that we can describe the treatment effect either as increasing the variance or affecting the fixed effects.

Recall the backpain dataset from Lab 2 with the following variables:

Doctor – Doctor ID number.
Cost – Cost of treatment in dollars.
Logcost – Logarithm of cost.
Actlim – Number of days in the past six months in which activity has been

limited.
Undrstnd – Did the patient understand the doctor's advice (1=yes, 0=no)
Age – Age of the patient.
Educ – Education of the patient (0 means <12 years, 1 means 13-16 years, 2 means > 16 years education).
Thoraic – Whether the back pain was cervical/thoraic or other (1=yes, 0=no).
Pracstyl – Practice style of the doctor (0=low, 1=medium, 2=high frequency of prescription of medicines and hospitalization for treatment of back pain).

1) Analyze the response Undrstnd using SAS Proc NLMIXED.
   a) You will need to code dummy variables for Educ and Pracstyl.
   b) Fit a model with Age, Educ and Pracstyle as predictors
   c) Provide an interpretation of the coefficient for age.
2) How do the results differ from those in Lab 2? How does the interpretation of the age coefficient differ?

The coefficient for age is now -.02277, which, when exponentiated gives .9775 or about a 2.5 percent decrease in the odds of understanding associated with each increase in age of one year.

Here are the coefficients, SEs and p-values from the two fits:

	NLmixed			GENMOD		
Parameter	b	SE	p-value	b	SE	p-value
Age	-0.0227	0.0053	<0.0001	-0.0218	0.0053	<0.0001
Educ 0 vs 2	0.1484	0.194	0.45	0.1421	0.132	0.28
Educ 1 vs 2	-0.0426	0.1885	.82	-0.0415	0.135	0.76
Prac style 0 vs 2	0.05689	0.2709	.83	0.0116	0.303	0.48
Prac style 1 vs 2	-0.2267	0.2359	.34	-0.2034	.285	0.48
ln(σ)	-2.0365	0.8112	.01	Doesn't fit a r. effects variance		

The coefficients, SEs and p-values are all qualitatively the same, though the coefficients in the conditionally specified models are larger in absolute value, as predicted by theory. The age coefficients have a slightly different interpretation. The coefficient from NLMIXED has the interpretation as the change in odds for an individual person as that person gets older. The interpretation from GENMOD is the change in odds averaged over all subjects associated with an increase of one year.

3) Repeat questions 1) and 2) using the outcome Actlim.

The coefficient of age is 0.005864, with exponential of exp(0.005864) = 1.0059. So the number of activity limitation days increases by about 0.6% as a person increases in age by one year. The numerical value for the coefficient from GENMOD was similar: exp(0.0075) = 1.0075 but the interpretation is the increase in activity limitation days associated with a one year increase in age, averaged over people.

A major difference is the p-values: for NLMIXED there is a t-statistic of 9.51 with a p-value approximately zero. For GENMOD the z-statistic was 1.82 with a p-value of 0.07. Which to believe?

The distribution is overdispersed compared to a Poisson distribution, so the Poisson assumption in NLMIXED is likely wrong. Unfortunately, it is harder to incorporate an overdispersed count distribution in NLMIXED – you have to program in an overdispersed distribution yourself using the GENERAL distribution.