

Generalized Linear Mixed Model: Examples

Biostatistics 653

Applied Statistics III: Longitudinal Data Analysis

Estimation using Maximum Likelihood

- The joint probability density function is given by

$$P(Y_i|X_i, b_i)P(b_i)$$

- However, recall that the b_i are unobserved. Typically, we base inferences on the marginal (integrated) likelihood function, given by

$$\prod_{i=1}^N \int P(Y_i|X_i, b_i)P(b_i) db_i$$

- Unfortunately, in most cases, this integral is not analytically tractable and is in a closed form.

Estimation using Maximum Likelihood

- For example, consider a log-linear Poisson model with a random intercept

$$\begin{aligned}Y_{ij} &\sim \text{Poi}(\mu_{ij}) \\ \log(\mu_{ij}) &= X_{ij}^T \beta + b_i \\ b_i &\sim N(0, \sigma^2)\end{aligned}$$

- The Likelihood is

$$\begin{aligned}L(\beta, \alpha) &= \prod_{i=1}^N \int \prod_{j=1}^{n_i} \frac{e^{-\mu_{ij}} \mu_{ij}^{y_{ij}}}{y_{ij}!} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} b_i^2} db_i \\ &= \prod_{i=1}^N \frac{1}{y_{ij}! \sqrt{2\pi\sigma^2}} \int \prod_{j=1}^{n_i} e^{-e^{(X_{ij}^T \beta + b_i)}} e^{y_{ij}(X_{ij}^T \beta + b_i)} e^{-\frac{1}{2\sigma^2} b_i^2} db_i \\ &= \prod_{i=1}^N \frac{e^{\sum_j y_{ij} X_{ij}^T \beta}}{y_{ij}! \sqrt{2\pi\sigma^2}} \int \prod_{j=1}^{n_i} e^{-\frac{1}{2\sigma^2} b_i^2 + y_{ij} b_i - e^{(X_{ij}^T \beta + b_i)}} db_i\end{aligned}$$

- Which is an integral with respect to a normal random variable b_i and is analytically intractable.

Estimation using Maximum Likelihood

There are several general approaches to approximate the integral

- Gaussian-Hermite quadrature
- Importance sampling or other Markov chain Monte Carlo (MCMC) approaches
- Laplace approximation
- Penalized quasi-likelihood (PQL)
- More recently, the integrated nested Laplace approximation (INLA)

Model Fitting in SAS

- One limitation of using generalized linear mixed models is the computational burden involved. Because there are no simple closed-form solutions available in general, numerical integration techniques are required to obtain the marginal likelihood.
- Maximum likelihood estimation in this setting has only quite recently been implemented in standard software (such as PROC NL MIXED in SAS). PROC NL MIXED directly maximizes an approximate integrated likelihood, using numerical quadrature.

Model Fitting in SAS

- PROC NLMIXED requires completely different model specifications from most other SAS procedures.
- The main advantage of this is that the user is given much flexibility in the way the model is specified and parameterized; a drawback is that the user needs to specify both the model and names for all the parameters in the model.
- In addition, the user must specify starting values for all parameters in the model. This is one of the main drawbacks of PROC NLMIXED: it does not automatically generate smart starting values, except for the default value of 1 for all parameters that are not specified by the user. In complex models, the convergence of the optimization algorithms may depend strongly on the specified starting values.
- PROC NLMIXED only produces MLE.

Example: Skin Cancer Prevention Study

- We consider data from the Skin Cancer Prevention Study, a randomized, double-blind, placebo-controlled clinical trial of beta-carotene to prevent non-melanoma skin cancer in high risk subjects. A total of 1805 subjects were randomized to either placebo or 50mg of beta-carotene per day for 5 years.
- Subjects were examined once a year and biopsied if a cancer was suspected to determine the number of new skin cancers occurring since the last exam. The outcome variable Y_{ij} is a binary variable that takes value 1 if new skin cancers were detected at time j and 0 otherwise.

Example: Skin Cancer Prevention Study

- The categorical variable treatment is coded 1=beta-carotene, 0=placebo. The variable year denotes the year of follow-up. The categorical variable gender is coded 1=male, 0=female. The categorical variable skin denotes skin type and is coded 1=burns, 0=otherwise. The variable age is the age (in years) of each subject at randomization. In addition, a variable exposure contains the number of previously-diagnosed skin cancers. Complete data are available on 1683 subjects comprising a total of 7081 measurements.

Example: Skin Cancer Prevention Study

- Investigators in this randomized clinical trial are interested in whether betacarotene helps to prevent skin cancer. However, because doses in the same range had been anecdotally linked to development of other cancers, the investigators do not wish to conduct one-sided hypothesis tests. In addition, the investigators wish to know whether covariates gender, skin type, and age are also related to cancer risk.

PROC NLMIXED

- We fit a random intercept model to the Skin Cancer Prevention Data using PROC NLMIXED. In particular, we will fit the model

$$\begin{aligned} & \text{logit} \left(P(Y_{ij} = 1 | b_i) \right) \\ &= \beta_0 + \beta_1 \text{trt}_i + \beta_2 \text{year}_{ij} + \beta_3 (\text{trt}_i)(\text{year}_{ij}) + \beta_4 (\text{age}_i - 63) \\ &+ \beta_5 \text{gender}_i + \beta_6 \text{skin}_i + \beta_7 \text{exposure}_i + \beta_8 \text{exposure}_i^2 + b_i \end{aligned}$$

where b_i is the random intercept for subject i. We assume $b_i \sim N(0, \sigma^2)$ and that conditional on b_i , Y_{ij} follows a Bernoulli distribution.

PROC NLMIXED

- In order to fit the model in PROC NLMIXED, we must specify starting values. Though we realize that the parameters in marginal and random effects models for categorical data in general do not represent the same unknown quantities, as starting values for β we will use $\hat{\beta}$ from the GEE fit to the data in the previous lecture. Alternatively, we could obtain initial values based on a prior analysis assuming no random effects (i.e., with an independence covariance structure).

PROC NL MIXED

- We must also specify a starting value for σ^2 . A good way to do this is by specifying a grid of feasible values; when given a grid of values, PROC NL MIXED evaluates the marginal likelihood at each grid value and then selects the grid point that produces the largest value of the marginal likelihood as the initial value for the covariance parameters. Starting values are very important for this SAS procedure. Experience indicates the procedure can be very sensitive to poor choices of starting values or the numerical accuracy of the quadrature used.

PROC NLMIXED

- Our starting values for β will be

$$\begin{aligned}\beta^{(0)} \\ = (-3.28, -0.02, -0.01, 0.03, 0.02, 0.22, 0.66, 0.41, -0.01)^T\end{aligned}$$

The intercept value of -3.28 from the GEE corresponds to a baseline (year 0, which isn't observed in the data, though there was not much effect due to year; placebo, age 63, female, skin that doesn't burn, with no prior cancers, which is also not observed in data) probability of cancer occurrence of roughly 4%.

PROC NLMIXED

- If we consider a point observed in the data, say year 1 on placebo for a 63-year old woman who has skin that does not burn and who had one prior cancer, we get $X_{i1}\hat{\beta} = -3.28 - 0.01(1) + 0.41(1) - 0.01(1^2) = -2.89$ from the GEE, which corresponds to an estimated cancer probability of $\frac{\exp(-2.89)}{1+\exp(-2.89)}$ or 5%.

PROC NLMIXED

- In order to pick a starting value for σ^2 , we think about how much variability we expect in a baseline cancer rate across subjects. If $b_i \sim N(0, \sigma^2)$, then $\sigma^2 = 1$ implies that our prior belief is that around 95% of such subjects as the type above would have b_i in the range (-2, 2). This corresponds to the estimated cancer probabilities across women ranging from 1-29% (obtained by taking the inverse logit of $-2.89-2$ and $-2.89+2$). Perhaps we do not expect such variability; $\sigma^2 = 0.25$ would reduce the woman-specific susceptibilities to the range of 2-13%. If we had $\sigma^2 = 2$, then the susceptibilities would be in the range of 0.3-48%. We will consider a grid of starting values in the range of 0.25 to 2.00 for σ^2 .

PROC NLMIXED

- The code follows. The Fitzmaurice et al. book gives a nice explanation of PROC NLMIXED. In the code, notice the starting values, specified by PARMs, the number of quadrature points specified (50 is a lot), the way the linear predictor must be explicitly specified (don't forget the random effects!), the specification of the link, and the specification of distributional assumptions of Y conditional on random effects and of the random effects themselves.

SAS Code

```
proc nlmixed data=new qpoints=50;
    PARMS beta0=-3.28 beta1=-0.02 beta2=-0.01
        beta3=0.03 beta4=0.02 beta5=0.22
        beta6=0.66 beta7=0.41 beta8=-0.01
        d11=0.25 to 2. by 0.25;
    eta=beta0+beta1*treatment+beta2*yearcont
        +beta3*treatment*yearcont
        +beta4*agecent +beta5*skin
        +beta6*gender + beta7*exposure
        +beta8*exposure*exposure+b1;
    mu=exp(eta)/(1.+exp(eta));
    MODEL ybin~BINARY(mu);
    RANDOM b1 ~ NORMAL(0,d11) SUBJECT=id;
    PREDICT b1 OUT=predbi;
run;
```

SAS Output

The NL MIXED Procedure

Specifications

Data Set	WORK.NEW
Dependent Variable	ybin
Distribution for Dependent Variable	Binary
Random Effects	b1
Distribution for Random Effects	Normal
Subject Variable	ID
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Dimensions

Observations Used	7081
-------------------	------

SAS Output

Observations Not Used	0
Total Observations	7081
Subjects	1683
Max Obs Per Subject	5
Parameters	10
Quadrature Points	50

Parameters

beta0	beta1	beta2	beta3	beta4	beta5	beta6	beta7
-3.28	-0.02	-0.01	0.03	0.02	0.22	0.66	0.41
-3.28	-0.02	-0.01	0.03	0.02	0.22	0.66	0.41

Parameters

beta8	d11	NegLogLike
-0.01	0.25	2818.47721
-0.01	0.5	2796.90981

SAS Output

The NL MIXED Procedure

Parameters

beta0	beta1	beta2	beta3	beta4	beta5	beta6	beta7
-3.28	-0.02	-0.01	0.03	0.02	0.22	0.66	0.41
-3.28	-0.02	-0.01	0.03	0.02	0.22	0.66	0.41
-3.28	-0.02	-0.01	0.03	0.02	0.22	0.66	0.41
-3.28	-0.02	-0.01	0.03	0.02	0.22	0.66	0.41
-3.28	-0.02	-0.01	0.03	0.02	0.22	0.66	0.41
-3.28	-0.02	-0.01	0.03	0.02	0.22	0.66	0.41

Parameters

beta8	d11	NegLogLike
-0.01	0.75	2789.20354
-0.01	1	2789.23647
-0.01	1.25	2793.75171
-0.01	1.5	2800.92822
-0.01	1.75	2809.6992
-0.01	2	2819.41498

SAS Output

Iter	Calls	Iteration History			
		NegLogLike	Diff	MaxGrad	Slope
1	6	2759.40506	29.79849	320.1533	-1926573
2	8	2750.70187	8.703182	440.6332	-2025.95
3	11	2749.16076	1.541115	74.87779	-1241.25
4	12	2746.66149	2.499274	67.59016	-26.5175
5	13	2742.7522	3.909289	374.1694	-20.4089
6	15	2741.58213	1.170066	537.8357	-21.0111
7	17	2738.65512	2.927008	616.5656	-6.05498
8	19	2738.09829	0.556835	284.6616	-0.80047
9	21	2737.89135	0.206942	90.01861	-0.17327
10	23	2737.82855	0.062797	33.88309	-0.06947
11	25	2737.82716	0.001385	2.319042	-0.00237
12	27	2737.82714	0.000027	1.18131	-0.00004
13	29	2737.82714	1.33E-6	0.113793	-2.12E-6

NOTE: GCONV convergence criterion satisfied.

SAS Output

Fit Statistics

-2 Log Likelihood	5475.7
AIC (smaller is better)	5495.7
AICC (smaller is better)	5495.7

The NLMIXED Procedure

Fit Statistics

BIC (smaller is better)	5549.9
-------------------------	--------

SAS Output

Parameter Estimates							
Standard							
Parameter	Estimate	Error	DF	t Value	Pr > t	Alpha	Lower
beta0	-3.9088	0.1911	1682	-20.46	<.0001	0.05	-4.2836
beta1	-0.04831	0.1794	1682	-0.27	0.7878	0.05	-0.4002
beta2	-0.01712	0.04052	1682	-0.42	0.6727	0.05	-0.09660
beta3	0.04590	0.05643	1682	0.81	0.4162	0.05	-0.06479
beta4	0.02110	0.005131	1682	4.11	<.0001	0.05	0.01103
beta5	0.2705	0.09747	1682	2.77	0.0056	0.05	0.07929
beta6	0.7628	0.1122	1682	6.80	<.0001	0.05	0.5428
beta7	0.4875	0.03635	1682	13.41	<.0001	0.05	0.4162
beta8	-0.01653	0.001926	1682	-8.58	<.0001	0.05	-0.02030
d11	1.1958	0.1549	1682	7.72	<.0001	0.05	0.8921

Parameter Estimates		
Parameter	Upper	Gradient
beta0	-3.5340	-0.0006
beta1	0.3036	-0.001
beta2	0.06236	-0.00039
beta3	0.1566	-0.00293
beta4	0.03116	0.004754
beta5	0.4616	-0.00081
beta6	0.9828	0.000397
beta7	0.5588	0.005575
beta8	-0.01275	0.113793
d11	1.4995	-0.00002

Interpretation of the Results

- The interpretation of these parameters is conditional on the value of b_i . For example, $\exp(0.2705) = 1.31$ is the odds ratio for cancer occurrence for someone whose skin burns easily, compared to if the person's skin did not burn easily. This is an individual-specific OR and is not a population-average OR.
- Recall that the population average OR estimated by the GEE in the previous lecture was $\exp(0.22) = 1.25$, so that we see the expected attenuation discussed by Diggle et al. in the population average parameters for the binary logistic regression model with random intercept.
- To see the estimated random effects, we can print the dataset `predbi` that we generated, and then create subject-specific growth curves from the subject-specific intercepts.

Modeling Count

- In addition, we can also fit a random effects model to the skin cancer counts over time. In this case, we should specify the Poisson distribution and change the link function used. Code to do this follows. Note that we are also now testing whether the treatment effect and past cancer history are significant predictors of the cancer occurrence rate. We use starting values for the betas from the Poisson GEE.
- To find starting values for σ^2 , again consider the 63 year-old woman on placebo with 1 prior cancer at year 1 of the study whose skin does not burn. Her estimated cancer rate if $b_i = 0$ would be $\exp(-2.89 + 0.00(1) + 0.32(1) - 0.01(1^2)) = \exp(-2.58) = 0.08$, so $\sigma^2 = 0.25$ on the lower end of the range would allow most estimated rates in the range (0.03, 0.21) while $\sigma^2 = 2$ would put most estimated rates in (0.004, 1.282). (The latter seems rather large.)

SAS Code

```
proc nlmixed data=new qpoints=50;
    PARMS beta0=-2.89 beta1=0.02 beta2=0.
        beta3=0.03 beta4=0.01 beta5=0.10
        beta6=0.58 beta7=0.32 beta8=-0.01
        d11=0.25 to 2. by 0.25;
    eta=beta0+beta1*treatment+beta2*yearcont
        +beta3*treatment*yearcont+beta4*agecent
        +beta5*skin + beta6*gender + beta7*exposure
        +beta8*exposure*exposure+b1;
    mu=exp(eta);
    MODEL y~POISSON(mu);
    RANDOM b1 ~ NORMAL(0,d11) SUBJECT=id;
    CONTRAST 'beta3' beta3 +0;
    CONTRAST 'beta3 no +0' beta3;
    CONTRAST 'exposure' beta7 +0, beta8 +0;
    CONTRAST 'trt' beta1 +0, beta3 +0;
    PREDICT b1 OUT=predbipois;
run;
```

SAS Output

Specifications

Data Set	WORK.NEW
Dependent Variable	Y
Distribution for Dependent Variable	Poisson
Random Effects	b1
Distribution for Random Effects	Normal
Subject Variable	ID
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Dimensions

Observations Used	7081
Observations Not Used	0
Total Observations	7081
Subjects	1683
Max Obs Per Subject	5
Parameters	10
Quadrature Points	50

SAS Output

Parameters							
beta0	beta1	beta2	beta3	beta4	beta5	beta6	beta7
-2.89	0.02	0	0.03	0.01	0.1	0.58	0.32
-2.89	0.02	0	0.03	0.01	0.1	0.58	0.32
Parameters							
beta8	d11	NegLogLike					
-0.01	0.25	4102.50873					
-0.01	0.5	4050.42642					

SAS Output

The NL MIXED Procedure

Parameters

beta0	beta1	beta2	beta3	beta4	beta5	beta6	beta7
-2.89	0.02	0	0.03	0.01	0.1	0.58	0.32
-2.89	0.02	0	0.03	0.01	0.1	0.58	0.32
-2.89	0.02	0	0.03	0.01	0.1	0.58	0.32
-2.89	0.02	0	0.03	0.01	0.1	0.58	0.32
-2.89	0.02	0	0.03	0.01	0.1	0.58	0.32
-2.89	0.02	0	0.03	0.01	0.1	0.58	0.32

Parameters

beta8	d11	NegLogLike
-0.01	0.75	4047.30572
-0.01	1	4059.90734
-0.01	1.25	4078.25489
-0.01	1.5	4098.59513
-0.01	1.75	4119.34207
-0.01	2	4139.78723

SAS Output

Iter	Calls	Iteration History			
		NegLogLike	Diff	MaxGrad	Slope
1	6	4045.93436	1.371363	693.1229	-168720
2	9	4017.49027	28.44408	722.1195	-10742
3	12	4012.65177	4.838508	92.11763	-5502.25
4	13	4002.25752	10.39424	74.85663	-154.701
5	14	3992.21513	10.04239	137.9533	-217.342
6	16	3987.75945	4.455681	98.83598	-106.176
7	17	3980.90742	6.852036	75.68521	-23.6668
8	19	3977.55147	3.355942	36.31041	-6.10859
9	21	3976.83625	0.715226	33.94884	-0.95738
10	23	3976.79624	0.040006	2.533237	-0.07241
11	25	3976.79246	0.003786	3.637925	-0.00805
12	27	3976.79234	0.000113	0.925929	-0.0002
13	29	3976.79234	2.039E-6	0.037134	-3.35E-6

NOTE: GCONV convergence criterion satisfied.

Fit Statistics

-2 Log Likelihood	7953.6
AIC (smaller is better)	7973.6
AICC (smaller is better)	7973.6
BIC (smaller is better)	8027.9

SAS Output

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower
beta0	-3.8010	0.1493	1682	-25.46	<.0001	0.05	-4.0938
beta1	0.001901	0.1265	1682	0.02	0.9880	0.05	-0.2462
beta2	0.003868	0.02609	1682	0.15	0.8821	0.05	-0.04729
beta3	0.03527	0.03576	1682	0.99	0.3241	0.05	-0.03487
beta4	0.01684	0.004409	1682	3.82	0.0001	0.05	0.008194
beta5	0.2529	0.08344	1682	3.03	0.0025	0.05	0.08926
beta6	0.6377	0.09674	1682	6.59	<.0001	0.05	0.4479
beta7	0.4319	0.02920	1682	14.79	<.0001	0.05	0.3746
beta8	-0.01432	0.001520	1682	-9.42	<.0001	0.05	-0.01730
d11	1.0859	0.09608	1682	11.30	<.0001	0.05	0.8975

Parameter Estimates

Parameter	Upper	Gradient
beta0	-3.5081	0.001595
beta1	0.2500	0.00055
beta2	0.05503	-0.00028
beta3	0.1054	0.00035

Parameter Estimates

Parameter	Upper	Gradient
beta0	-3.5081	0.001595
beta1	0.2500	0.00055
beta2	0.05503	-0.00028
beta3	0.1054	0.00035
beta4	0.02549	0.003908
beta5	0.4166	-0.00042
beta6	0.8274	0.000714
beta7	0.4891	0.006202
beta8	-0.01133	0.037134
d11	1.2744	0.001063

Contrasts

Label	Num DF	Den DF	F Value		Pr > F
			F	Value	
beta3	1	1682		0.97	0.3241
beta3	1	1682		0.97	0.3241
exposure	2	1682		201.58	<.0001
trt	2	1682		1.18	0.3088

Distribution of Random Effects

- We have used the normal distribution as a convenient model for the random effects. When we are primarily interested in the subject-specific regression coefficients, the specific form of the random effects distribution is less important. However, if the random effects themselves are the focus, then inferences are more dependent on the assumptions about their distribution. Researchers have been developing non-parametric approaches to estimating the random effects distribution with non-linear models, but these approaches are not yet available in standard software.

Multilevel Models

- Multilevel generalized linear models extend in a natural way the conceptual approach we discussed previously for multilevel models for normal response data. They no longer require the level 1 observations to have a normal distribution; instead, they are assumed to have a distribution belonging to the exponential family.

Two-level Models

- The basic idea of multilevel models is that clustering among units can be thought of as arising from their sharing a set of random effects. Conditional on the random effects, the level 1 observations are assumed to be independent.
- Using the same (“reverse”) notation as before, we let Y_{ij} be the response of the i 'th level 1 unit in the j 'th level 2 cluster. We specify a two-level model as follows.

Two-level Models

- Assume the conditional distribution of each Y_{ij} given b_j belongs to the exponential family and that $\text{Var}(Y_{ij}|b_j) = \nu(E(Y_{ij}|b_j))\phi$, where $\nu(\cdot)$ is a known variance function, and ϕ is a scale parameter. In addition, conditional on the b_j , the Y_{ij} are independent of each other.
- The conditional mean of Y_{ij} depends on fixed and random effects through the linear predictor

$$\eta_{ij} = X_{ij}^T \beta + Z_{ij}^T b_j.$$

with

$$g(E(Y_{ij}|b_j)) = \eta_{ij} = X_{ij}^T \beta + Z_{ij}^T b_j$$

for some known link function $g(\cdot)$.

- The random effects are assumed to have some probability distribution; usually, we assume $b_j \sim MVN(0, D)$.

Two-level Models

- While higher-level models may be considered, the difficulty in computing the marginal log-likelihood is increased due to the introduction of additional random effects. Currently, PROC NLMIXED does not handle three-level or higher level models, though specialized software specifically developed for multilevel models (for example, MLwinN or gllamm), can do so.
- Note that PROC GLIMMIX has now been released and can handle a much wider variety of model structures than PROC NLMIXED; it has the potential to handle higher-order multilevel models. However, the approximation used in model fitting for PROC GLIMMIX (restricted pseudo-likelihood, which is a linearization scheme based on Taylor expansion and discussed in Wolfinger and O'Connell, 1993) is less accurate than that used by PROC NLMIXED.

Example: Malignant Melanoma and UV Light Exposure

- Following Fitzmaurice et al., we consider a study of the effects of ultraviolet (UV) light exposure on malignant melanoma mortality. Counts of the number of deaths due to malignant melanoma were recorded for males in the UK; these counts were aggregated over 70 counties or shires, which were nested within 11 regions of the UK. Thus the county is the level 1 unit, and the region is the level 2 unit. The predictor of interest is exposure to UV light in the B band (UVB), and an index of UVB dose reaching the surface of the earth for each county was calculated. Note that the mean UVB index in the UK was 10.9 ($sd=1.5$) during the time of the study.

Malignant Melanoma Study

- Let Y_{ij} denote the count of deaths in county i of region j due to malignant melanoma. For each county, T_{ij} is the number of deaths that we would expect to see in the county (based on UK national death rates). Including $\log(T_{ij})$ in the model as an offset allows us to assume a linear relationship between the log standardized mortality ratio (SMR), Y/T , due to malignant melanoma and county-level UV radiation exposure.
- Without the offset, we would just be modeling raw counts, so that results would not be meaningful in light of different population sizes and dynamics in the different counties. Offsets are also useful in other settings in which modeling rates is our goal; these include settings in which the counts are taken over differing person-time values, or in which counts are obtained over a geographical space and the area covered is different for different units.

Malignant Melanoma Study

- We assume that the counts, conditional on the random effects, follow a Poisson distribution with conditional mean related to UVB dose through a log link function, so that

$$\log(E(Y_{ij}|b_j)) = \log(T_{ij}) + \beta_1 + \beta_2 UVB_{ij} + b_j$$

and we assume $b_j \sim N(0, \sigma^2)$.

- When modeling using PROC NL MIXED, we need to supply starting values. We can get these from PROC GENMOD.

SAS Code and Results

```
proc genmod data=new2;
model obsdeath=centuvb/dist=poisson offset=logexp;
run;
/* got starting values of -0.0194 for int and 0.1405 for slope */

proc nlmixed data=new2 qpoints=50;
PARMS beta0=-0.0194 beta1=0.1405
d11=0.001 to 2. by 0.05;
eta=logexp+beta0+beta1*centuvb+b1;
mu=exp(eta);
MODEL obsdeath~POISSON(mu);
RANDOM b1 ~ NORMAL(0,d11) SUBJECT=region;
PREDICT b1 OUT=predbi;
run;
```

Table 1: Estimates for malignant melanoma mortality data (NLMIXED)

Variable	Estimate	SE	Z
Intercept	-0.0365	0.0352	-1.04
UVB Slope	0.1301	0.0279	4.67
Level 2 Variance	0.00622	0.005087	

Malignant Melanoma Study

- There is a significant positive relationship between UVB exposure level and malignant mortality. Recalling that the standard deviation of UVB exposure in the UK was 1.5, we see that the SMR is about 1.5 (1.25, 1.74) times larger when the county has UVB index 1 standard deviation above the UK average compared to the same county if it had UVB index 1 sd below (a two standard deviation difference is $1.5+1.5$ so we take $\exp((1.5 + 1.5)(0.13)) \approx 1.5$).

GLIMMIX Code and Results

- While we don't need PROC GLIMMIX in this example, it is useful when we need to specify more complex covariance structures. Here is the code for PROC GLIMMIX, as well as results, for this example. Note that The Level 2 variance estimates differ, but other estimates are quite similar.

```
proc glimmix data=new2;
  class region;
  model obsdeath=centuvb/dist=poisson offset=logexp solution;
  random intercept/subject=region;
run;
```

Table 2: Estimates for malignant melanoma mortality data (NLMIXED)

Variable	Estimate	SE	Z
Intercept	-0.0372	0.0378	-0.98
UVB Slope	0.1308	0.0299	4.37
Level 2 Variance	0.00855	0.006677	