

## *GLMM examples with bivariate random effects vectors*

### **5. Poisson regression with multivariate random effects**

Thall and Vail (1990) presented data from a clinical trial of  $n = 59$  epileptic patients who were randomized to take either a new drug  $d_i = 1$  or a placebo ( $d_i = 0$ ) in addition to standard chemotherapy. Other baseline data included  $a_i = \log(\text{age}_i)$  where  $\text{age}_i$  is age in years and  $b_i = \log(\text{base}_i/4)$ , where  $\text{base}_i$  is number of seizures in preceding 8-week period. The outcome is  $Y_{ij}$  the number of seizures within the following 4 2-week periods, up to 8 weeks. So  $j = 1, 2, 3, 4$ . The time variable used is actually  $t_j = 0.2(j - 2.5)$ . The model fit is

$$Y_{ij} \sim \text{Poisson}(\lambda_{ij}),$$

where

$$\begin{aligned}\log \lambda_{ij} &= \beta_0 + \beta_b b_i + \beta_d d_i + \beta_{bd} b_i d_i + \beta_a a_i + \beta_v t_j + u_{i1} + u_{i2} t_j \\ &= \begin{bmatrix} 1 \\ b_i \\ d_i \\ b_i d_i \\ a_i \\ t_j \end{bmatrix}' \begin{bmatrix} \beta_0 \\ \beta_b \\ \beta_d \\ \beta_{bd} \\ \beta_a \\ \beta_v \end{bmatrix} + \begin{bmatrix} 1 \\ t_j \end{bmatrix}' \begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix} \\ &= \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{u}_i\end{aligned}$$

We further assume

$$\mathbf{u}_1, \dots, \mathbf{u}_{59} \stackrel{iid}{\sim} N_2(\mathbf{0}, \mathbf{\Sigma}) = N_2 \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \right).$$

This assumes a log linear trend individual  $i$ 's seizure rate over the 8 weeks. Specifically,

$$\log \lambda_{ij} = \theta_{0i} + \theta_{1i} \text{weeks}_j,$$

where  $\text{cov}(\theta_{0i}, \theta_{1i}) = \sigma_{12}$ . This follows from properties of multivariate normal distributions.

```

data seiz1;
  input id$ seiz visit treat age base;
  age=log(age);
  base=log(base/4);
  if visit=1 then visit=-3;
  if visit=2 then visit=-1;
  if visit=3 then visit=1;
  if visit=4 then visit=3;
  visit=visit/10;
  datalines;
101  11    1  1  18  76
101  14    2  1  18  76
101   9    3  1  18  76
101   8    4  1  18  76
102   8    1  1  32  38
102   7    2  1  32  38
102   9    3  1  32  38
102   4    4  1  32  38
etc...
234   3    1  0  19  10
234   1    2  0  19  10
234   4    3  0  19  10
234   2    4  0  19  10
236   1    1  1  37  12
236   4    2  1  37  12
236   3    3  1  37  12
236   2    4  1  37  12
238  13    1  0  22  47
238  15    2  0  22  47
238  13    3  0  22  47
238  12    4  0  22  47

```

```

;
proc nlmixed qpoints=50;
parms b_const=-1.3 b_base=0.9 b_trt=-0.9 b_basetrt=0.3 b_age=0.2
      b_visit=-0.3 s11=0.25 s22=0.53 s12=0.003;
eta=b_const+b_base*base+b_trt*treat+b_age*age+b_basetrt*base*treat
    +b_age*age+b_visit*visit+u1+u2*visit;
lambda=exp(eta);
model seiz ~ poisson(lambda);
random u1 u2 ~ normal([0,0],[s11,s12,s22]) subject=id;

```

Here's the annotated output:

Parameter	Estimate	SE	Pr >  t	Lower	Upper
b_const	-1.3682	1.2007	0.2593	-3.7726	1.0363
b_base	0.8850	0.1313	<.0001	0.6221	1.1478
b_trt	-0.9287	0.4022	0.0246	-1.7340	-0.1233
b_basetrt	0.3380	0.2044	0.1038	-0.07142	0.7474
b_age	0.2384	0.1768	0.1830	-0.1157	0.5924
b_visit	-0.2664	0.1647	0.1113	-0.5962	0.06342
s11	0.2515	0.05879	<.0001	0.1338	0.3692
s22	0.5315	0.2294	0.0241	0.07214	0.9908
s12	0.002871	0.08870	0.9743	-0.1748	0.1805

Consider an individual from the population with covariates  $(a, b)$  at time  $t$  with random effect  $(u_1, u_2)$ . The ratio of seizure rates, within this individual, for drug versus placebo is:

$$\frac{\lambda(a, b, t, d = 1 | \mathbf{u})}{\lambda(a, b, t, d = 0 | \mathbf{u})} = \frac{e^{\beta_0 + \beta_b b + \beta_d + \beta_{bd} b + \beta_a a + \beta_v t + u_1 + u_2 t}}{e^{\beta_0 + \beta_b b + \beta_a a + \beta_v t + u_1 + u_2 t}} = e^{\beta_d + \beta_{bd} b}.$$

Within a subject, the mean number of seizures over a 2-week period is reduced by  $e^{-0.929 + 0.338 \log(\text{base}/4)} = (0.247) \text{base}^{0.338}$ .

This function crosses unity between 62 and 63 baseline seizures within the previous 8 weeks. It's about 0.5 when  $\text{base} = 8$ . So the drug significantly reduces seizures at any visit, but the reduction rate critically depends on the baseline seizure rate.

Would any other interactions be of interest here? How about a visit by treatment interaction?

## 6. Mice toxicity example: continuation ratio logit model

We have  $n = 94$  pregnant mice that were given one of four doses of a toxic substance, ethylene glycol (antifreeze, Table 12.9, page 518). The doses were  $x_i = 0.00, 0.75, 1.50, 3.00$  g/kg. There are three possible outcomes for each fetus: (dead/resorption, malformation, normal). Let  $Y_{ij}$  be the outcome for the  $j^{\text{th}}$  fetus of mouse  $i$ . The  $(Y_{i1}, Y_{i2}, \dots, Y_{in_i})$  are clustered within a litter, and each  $Y_{ij} = 1, 2, 3$  depending on the outcome.

The hazard, or continuation ratio, of remaining at stage  $k$  given that a fetus has made it to stage  $k$  is

$$\omega_{ik} = P(Y_{ij} = k | Y_{ik} \geq k).$$

One minus this is the probability of “moving on.”

$Y_{ij} = 1$  indicates death and  $Y_{ij} \geq 2$  indicates survival.

In your homework you will show the probability of death is  $P(Y_{ij} = 1) = \omega_{i1}$  and the probability of malformation given survival is  $P(Y_{ij} = 2) = \omega_{i2}(1 - \omega_{i1})$ . Then necessarily the probability of normal is  $P(Y_{ij} = 3) = 1 - [P(Y_{ij} = 1) + P(Y_{ij} = 2)]$ .

Let  $I\{Y_{ij} = k\}$  indicate that fetus  $j$  from mouse  $i$  made it to stage  $k$ . Let  $n_{i1} = \sum_{j=1}^{n_i} I\{Y_{ij} = 1\}$  be the number of fetuses in litter  $i$  that died/resorbed. Let  $n_{i2} = \sum_{j=1}^{n_i} I\{Y_{ij} = 2\}$  be the number that were malformed, and  $n_{i3} = \sum_{j=1}^{n_i} I\{Y_{ij} = 3\}$  be the number that survived. Then

$$(n_{i1}, n_{i2}, n_{i3}) \sim \text{mult}(n_i, \mathbf{p}_i),$$

where  $\mathbf{p}_i = (\omega_{i1}, \omega_{i2}(1 - \omega_{i1}), 1 - [\omega_{i1} + \omega_{i2}(1 - \omega_{i1})])$ .

The pmf of  $(n_{i1}, n_{i2}, n_{i3})$ , conditional on  $(\omega_{i1}, \omega_{i2})$  is

$$p(n_{i1}, n_{i2}, n_{i3}) \propto [\omega_{i1}]^{n_{i1}} [\omega_{i2}(1 - \omega_{i1})]^{n_{i2}} [1 - [\omega_{i1} + \omega_{i2}(1 - \omega_{i1})]]^{n_{i3}},$$

which will be explicitly used in PROC NLMIXED.

Further we assume the hazard model

$$\text{logit } \omega_{i1} = u_{i1} + \alpha_1 + \beta_1 x_i$$

$$\text{logit } \omega_{i2} = u_{i2} + \alpha_2 + \beta_2 x_i$$

The hazard odds depend on the stage the fetus is currently in (i.e. different  $\alpha_k$  for the two stages  $k = 1, 2$ ) and change differing amounts for different doses  $x_i$  (different  $\beta_k$ ). The  $\mathbf{u}_i = (u_{i1}, u_{i2})$  is a litter-specific random effect that allows heterogeneity in going from stage 1 to 2 ( $u_{i1}$ ) and stage 2 to 3 ( $u_{i2}$ ) across litters, beyond the population effects  $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ . We initially assume

$$\mathbf{u}_1, \dots, \mathbf{u}_{94} \stackrel{iid}{\sim} N_2(\mathbf{0}, \mathbf{\Sigma}),$$

where  $\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$ . We have the ingredients to fit the model.

```

data micel;
input dead malf norm treat;
datalines;
1 0 7 0
0 0 14 0
0 0 13 0
etc...
1 3 9 3
0 2 5 3
0 1 11 3
;
data mice2; set micel; id=_n_;
proc nlmixed;
parms a1=-4.0 a2=-4.3 b1=0.1 b2=1.8 s11=0.1 s22=2.5 s12=0;
eta1=u1+a1+b1*treat; eta2=u2+a2+b2*treat;
w1=exp(eta1)/(1+exp(eta1)); w2=exp(eta2)/(1+exp(eta2));
p1=w1; p2=w2*(1-w1); p3=1-(w1+w2*(1-w1));
ll=dead*log(p1)+malf*log(p2)+norm*log(p3);
model dead ~ general(ll);
random u1 u2 ~ normal([0,0],[s11,s12,s22]) subject=id;
run;

```

with output:

Parameter	Estimate	SE	DF	t Value	Pr >  t	Alpha	Lower	Upper	Gradient
a1	-4.1953	0.4679	92	-8.97	<.0001	0.05	-5.1245	-3.2660	0.000066
a2	-4.3582	0.4406	92	-9.89	<.0001	0.05	-5.2332	-3.4831	0.000205
b1	0.08270	0.2092	92	0.40	0.6935	0.05	-0.3328	0.4982	0.000033
b2	1.7804	0.2213	92	8.04	<.0001	0.05	1.3408	2.2200	0.000242
s11	0.3137	0.7011	92	0.45	0.6556	0.05	-1.0788	1.7062	8.121E-6
s22	2.5270	0.6904	92	3.66	0.0004	0.05	1.1558	3.8981	0.000016
s12	0.06583	0.4775	92	0.14	0.8906	0.05	-0.8825	1.0141	1.125E-6

The number of quadrature points used (this is adaptive) is 7. Fixing `qpnts=30` changes the results negligibly.

We see that the dose of ethylene glycol does not appreciably change the probability of death versus survival, but does significantly affect the probability of malformation given survival. The odds of malformation (given survival) increase by  $e^{1.78} = 5.9$  for every g/kg increase in ethylene glycol given.

## 7. Leading the crowd (Table 12.8, p. 516)

A sample of schoolboys were asked asked two questions several months apart:

- “Are you a member of the leading crowd?”  $Y_{ij1} = 0/1$  for no/yes.
- “Do you have to go against your principles to belong to your group?”  $Y_{ij2} = 0, 1$  for no/yes.

Here  $i$  is the boy and  $j = 1, 2$  denotes the time period when asked both questions. It is of interest to see how responses change within individuals. The model considered (page 517) is

$$\text{logit } P(Y_{ij1} = 1) = \beta_{j1} + u_{i1}$$

$$\text{logit } P(Y_{ij2} = 1) = \beta_{j2} + u_{i2}$$

$\mathbf{u}_i = (u_{i1}, u_{i2})$  is a schoolboy-specific random vector representing a boy's predisposed self-perceived idea of being in the leading crowd versus the "followers" ( $u_{i1}$ ) and going against his principles to belong in his respective group ( $u_{i2}$ ).

$e^{\beta_{21} - \beta_{11}}$  is how the odds of self-perceived membership in the leading crowd changes over time;  $e^{\beta_{22} - \beta_{12}}$  is how going against principles changes over time within a randomly selected boy.

Agresti's code:

```
data crowd;
input mem1 att1 mem2 att2 count;
datalines;
1 1 1 1 458
1 1 1 0 140
1 1 0 1 110
1 1 0 0 49
1 0 1 1 171
1 0 1 0 182
1 0 0 1 56
1 0 0 0 87
0 1 1 1 184
0 1 1 0 75
0 1 0 1 531
0 1 0 0 281
0 0 1 1 85
0 0 1 0 97
0 0 0 1 338
0 0 0 0 554
;
data new; set crowd;
  case=_n_;
  x1m=1; x1a=0; x2m=0; x2a=0; var=1; resp=mem1; output;
  x1m=0; x1a=1; x2m=0; x2a=0; var=0; resp=att1; output;
  x1m=0; x1a=0; x2m=1; x2a=0; var=1; resp=mem2; output;
  x1m=0; x1a=0; x2m=0; x2a=1; var=0; resp=att2; output;
drop mem1 att1 mem2 att2;
```

```

proc nlmixed data=new qpoints=30;
  eta=beta1m*x1m + beta1a*x1a + beta2m*x2m + beta2a*x2a + um*var + ua*(1-var);
  p=exp(eta)/(1+exp(eta));
  model resp ~ binary(p);
  random um ua ~ normal([0,0],[s1*s1, cov12, s2*s2]) subject=case;
  replicate count;
  estimate 'mem change' beta2m-beta1m; estimate 'att change' beta2a-beta1a;

```

## Output:

### Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper	Gradient
beta1m	-1.1882	0.08544	3396	-13.91	<.0001	0.05	-1.3558	-1.0207	0.001879
beta1a	0.2149	0.04876	3396	4.41	<.0001	0.05	0.1193	0.3105	-0.0015
beta2m	-0.8096	0.08097	3396	-10.00	<.0001	0.05	-0.9684	-0.6509	-0.00312
beta2a	0.3911	0.04950	3396	7.90	<.0001	0.05	0.2941	0.4882	-0.00193
s1	3.0807	0.1308	3396	23.55	<.0001	0.05	2.8242	3.3373	-0.00111
cov12	1.4729	0.1691	3396	8.71	<.0001	0.05	1.1415	1.8044	-0.00206
s2	1.4930	0.07385	3396	20.22	<.0001	0.05	1.3482	1.6378	-0.00029

Additional Estimates

Label	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper
mem change	0.3786	0.07469	3396	5.07	<.0001	0.05	0.2322	0.5250
att change	0.1762	0.05808	3396	3.03	0.0024	0.05	0.06232	0.2901

Does self-perceived membership change significantly over time within boys? How about their attitude towards going against their principles? In which direction? Note  $e^{0.379} = 1.46$  and  $e^{1.76} = 1.19$ .

Of interest: we reject  $H_0 : \text{cov}(u_{i1}, u_{i2}) = \sigma_{12} = 0$  in favor of  $\sigma_{12} > 0$ . This corresponds to  $\hat{\rho} = 1.47 / (3.08 \times 1.49) = 0.32$ . How is this interpreted?

Some final comments:

- 12.5.3 discusses multilevel modeling: different sets of random effects at different levels of a hierarchy (e.g. a student takes a battery of tests at a school: students within school, schools within state, state within country).
- Agresti has  $G^2$  for GOF (deviance-based) tests when looking at contingency tables (e.g. leading crowd example). We get the maximized log-likelihood out of PROC NLMIXED. If careful we might be able to get  $G^2$ .
- You will explore the toxicity data in more detail along the lines of Section 12.5.2.
- Did not discuss Bayesian approaches; very natural here.
- Can check normality assumption by looking at  $\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_n$  but problems with this when cluster sizes are small.

- 12.6.6 discusses testing  $H_0 : \sigma = 0$  versus  $H_1 : \sigma > 0$  in a simple model with univariate  $u_1, \dots, u_n \stackrel{iid}{\sim} N(0, \sigma^2)$ . Fit the full model with random effects compute  $L_f$  (maximized log-likelihood), fit simpler model without random effects  $\sigma = 0$  and get  $L_r$ . Let  $t = -2[L_r - L_f]$  be the LRT statistic. The  $p$ -value for the test is  $p = 0.5P(\chi_1^2 > t)$ . You can use this in 12.2(b).
- 12.1.5: random intercept model not appropriate in case-control study as clusters not randomly sampled. Apparently there are fixes to this.