Chapter 13: Random Effects Models & more - II

Dipankar Bandyopadhyay

Department of Biostatistics,
Virginia Commonwealth University

BIOS 625: Categorical Data & GLM

13.3.2 Opinion on legalized abortion

	Response sequence							
Gender	(1,1,1)	(1,1,0)	(0,1,1)	(0,1,0)	(1,0,1)	(1,0,0)	(0,0,1)	(0,0,0)
Male	342	26	6	21	11	32	19	356
Female	440	25	14	18	14	47	22	547

Let (Y_{i1}, Y_{i2}, Y_{i3}) be the response to three questions asked of the same individual, "Do you support legalized abortion under three scenarios: (1) if the family has very low income, (2) the woman is unmarried & doesn't want to get married, (3) woman wants it for any reason?" $Y_{ij} = 1$ indicates "yes." A covariate of interest is gender: $x_i = 0$ for male $x_i = 1$ for female. A logistic-normal model is

$$\text{logit } P(Y_{ij} = 1) = \alpha + \beta_1 I\{j = 1\} + \beta_2 I\{j = 2\} + \gamma x_i + u_i, \quad u_i \overset{iid}{\sim} N(0, \sigma^2).$$

Within the same individual:

- e^{β_1} compares the odds of "support legalized abortion" comparing "poor" to "any reason."
- e^{β_2} compares the odds of "support legalized abortion" comparing "single" to "any reason."
- $e^{\beta_2 - \beta_1}$ compares the odds of "support legalized abortion" of "single" to "poor."
- e^{γ} compares the odds of "support legalized abortion" comparing females to males.

Agresti's SAS code:

```
data new;
    input sex poor single any count;
    datalines ;
    1 1 1 1 342
    1 1 1 0 26
    1 1 0 1 11
    1 1 0 0 32
    1 0 1 1 6
    1 0 1 0 21
    1 0 0 1 19
    1 0 0 0 356
    2 1 1 1 440
    2 1 1 0 25
    2 1 0 1 14
    2 1 0 0 47
    2 0 1 1 14
    2 0 1 0 18
    2 0 0 1 22
    2 0 0 0 457
;
```

```
data new1;  set new;
    sex = sex-1; case = _n_;
    q1=1; q2=0; resp = poor; output;
    q1=0; q2=1; resp = single; output;
    q1=0; q2=0; resp = any; output;
    drop poor single any;
proc nlmixed data=new1 qpoints = 50;
    parms alpha=0 beta1=.8 beta2=.3 gamma=0 sigma=8.6;
    eta = alpha + beta1*q1 + beta2*q2 + gamma*sex + u;
    p = exp(eta)/(1 + exp(eta));
    model resp ~ binary(p);
    random u ~ normal(0,sigma*sigma) subject = case;
     replicate count;
```

I added the following to get estimates of interest:

```
    estimate 'odds:   poor vs. any ' exp(beta1);
    estimate 'odds: single vs. any ' exp(beta2);
    estimate 'odds: single vs. poor' exp(beta2-beta1);
    estimate 'odds: female vs. male' exp(gamma);
```

The output looks like:

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
alpha	−0.6222	0.3812	1849	−1.63	0.1028	0.05	−1.3698	0.1255	0.000588
beta1	0.8358	0.1602	1849	5.22	<.0001	0.05	0.5217	1.1500	−0.0004
beta2	0.2929	0.1568	1849	1.87	0.0619	0.05	−0.01465	0.6004	0.000506
gamma	0.01272	0.4936	1849	0.03	0.9794	0.05	−0.9554	0.9809	0.000306
sigma	8.7878	0.5565	1849	15.79	<.0001	0.05	7.6964	9.8791	−0.00032

Additional Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
odds: poor vs. any	2.3068	0.3695	1849	6.24	<.0001	0.05	1.5821	3.0314
odds: single vs. any	1.3403	0.2102	1849	6.38	<.0001	0.05	0.9281	1.7525
odds: single vs. poor	0.5810	0.09137	1849	6.36	<.0001	0.05	0.4018	0.7602
odds: female vs. male	1.0128	0.5000	1849	2.03	0.0429	0.05	0.03226	1.9933

According to this (additive) model, there are significant differences within individuals on how they feel about legalized abortion depending on the circumstance. There is no significant difference due to gender. Under which circumstance is one's position on legalized abortion most favorable? Least?

The estimate of $\hat{\sigma} = 8.8$ is quite large relative to the magnitude of the fixed effects (which are all less than unity). This reflects extreme heterogeneity in subject-to-subject response clusters (Y_{i1}, Y_{i2}, Y_{i3}). 1595 of 1850 subjects answered either $(0, 0, 0)$ or $(1, 1, 1)$. Does this also agree with what we know about abortion as a "polarizing issue?"
Code to fit the marginal exchangeable model via GEE looks like:

```
data new2;  set new;
 case=0; seq=_n_; * nesting case within sequence type (y1,y2,y3);
 do i=1 to count;
  case=case+1;
  q1=1; q2=0; resp = poor;   output;
  q1=0; q2=1; resp = single; output;
  q1=0; q2=0; resp = any;    output;
 end;
 drop poor single any i count;
proc genmod data=new2; class case sex seq;
 model resp=q1 q2 sex / dist=bin link=logit;
 repeated subject=case(seq) / type=exch;
```

This code makes use of nesting. Instead of having one case index $i = 1, \ldots, 1850$ for each individual, I have case nested within the type of sequence (Y_1, Y_2, Y_3), $i = 1, \ldots, j(i)$ where $j(1) = 342$, $j(2) = 26$, etc., $j(16) = 457$. This allows me to quickly get the data into a form SAS can use in PROC GENMOD. Output:

GEE Model Information

Correlation Structure	Exchangeable
Subject Effect	case(seq) (1850 levels)
Number of Clusters	1850
Correlation Matrix Dimension	3
Maximum Cluster Size	3
Minimum Cluster Size	3

Exchangeable Working Correlation

Correlation 0.8173308153

Empirical Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		−0.1219	0.0607	−0.2408	−0.0030	−2.01	0.0446
q1		0.1493	0.0297	0.0911	0.2076	5.02	<.0001
q2		0.0520	0.0270	−0.0010	0.1050	1.92	0.0544
sex	1	−0.0034	0.0878	−0.1756	0.1687	−0.04	0.9688
sex	2	0.0000	0.0000	0.0000	0.0000	.	.

As before, we see attenuation of the effects towards zero in the marginal model. From the conditional model we compute $\hat{c} = 1/\sqrt{1 + 0.346(8.79)^2} = 0.190$. Note that 0.149 is very close to $0.159 = 0.190(0.836)$.

We can estimate the *population* ratio of odds for "poor" versus "single" by adding the command:

```
estimate "odds poor vs.  single" q1 1 q2 -1 / exp;
```

to the PROC GENMOD statement yielding:

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
odds poor vs. single	0.0973	0.0275	0.05	0.0434	0.1513	12.50	0.0004
Exp(odds poor vs. single)	1.1022	0.0303	0.05	1.0443	1.1633		

13.3.3. Longitudinal study of mental health

Table 11.2 (p. 459) houses data from a longitudinal study comparing a new drug with a standard drug for treatment of subjects suffering mental depression. $n = 340$ Patients were either mildly or severely depressed upon admission into the study. At weeks 1, 2, and 4, corresponding to $j = 1, 2, 3$, patient i's suffering Y_{ij} was classified as normal $Y_{ij} = 1$ or abnormal $Y_{ij} = 0$. Let $s_i = 0, 1$ be the severity of the diagnosis (mild, severe) and $d_i = 0, 1$ denote the drug (standard, new).
We treat time as a categorical predictor and fit a marginal logit model with an exchangeable correlation structure:

```
data depress;
     infile "E:/CategoricalDataAnalysis/Spring2013/Chapter13/depress.txt";
    input case diag treat time outcome; time=time+1;
  q1=0; q2=0; if time=1 then q1=1; if time=2 then q2=1;

proc genmod descending; class case time;
     model outcome = diag treat time treat*time
     / dist=bin link=logit type3;
     repeated subject=case / type=exch corrw;
```

GEE Model Information

Correlation Structure	Exchangeable
Subject Effect	case (340 levels)
Number of Clusters	340
Correlation Matrix Dimension	3

Working Correlation Matrix

	Col1	Col2	Col3
Row1	1.0000	−0.0034	−0.0034
Row2	−0.0034	1.0000	−0.0034
Row3	−0.0034	−0.0034	1.0000

Empirical Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		0.9812	0.1841	0.6203	1.3421	5.33	<.0001
diag		−1.3117	0.1453	−1.5964	−1.0269	−9.03	<.0001
treat		2.0427	0.3061	1.4428	2.6426	6.67	<.0001
time	1	−0.9601	0.2379	−1.4265	−0.4938	−4.04	<.0001
time	2	−0.6207	0.2372	−1.0855	−0.1559	−2.62	0.0089
time	3	0.0000	0.0000	0.0000	0.0000	.	.
treat*time	1	−2.0975	0.3923	−2.8663	−1.3287	−5.35	<.0001
treat*time	2	−1.0958	0.3900	−1.8602	−0.3314	−2.81	0.0050
treat*time	3	0.0000	0.0000	0.0000	0.0000	.	.

Score Statistics For Type 3 GEE Analysis

Source	DF	Chi-Square	Pr > ChiSq
diag	1	70.83	<.0001
treat	1	40.38	<.0001
time	2	15.73	0.0004
treat*time	2	29.52	<.0001

We see a severe diagnosis ($s = 1$) significantly decreases the odds of a normal classification by a factor of $e^{-1.31} = 0.27$. The odds (or normal classification) ratio comparing the new drug to the standard drug changes with time because of the interaction. At 1 week it's $e^{2.04-2.09} = 0.95$, and week 2 it's $e^{2.04-1.10} = 2.6$, and at 4 weeks it's $e^{2.04-0} = 7.7$. The new drug is better, but takes time to work.

Here, the focus is on whole populations of patients at 1, 2, and 4 weeks, and on the new drug versus the standard drug. These interpretations are not within the individual.
We now consider a conditional analysis

$$\begin{aligned} \text{logit } P(Y_{ij}=1) &= \alpha+\beta_1 s_i+\beta_2 d_i+\beta_3 I\{j=1\}+\beta_4 I\{j=2\} \\ &\quad +\beta_5 I\{j=1\}d_i+\beta_6 I\{j=2\}d_i+u_i \\ \text{where } u_i &\sim N(0,\sigma^2). \end{aligned}$$

I round parameter estimates from the GEE approach to use as starting values and fix `qpoints=200` (more on this later):

```
proc nlmixed  qpoints=200;
  parms a=1 b1=-1 b2=2 b3=-1 b4=-0.5 b5=-2 b6=-1 sig=.1;
  eta  = a+b1*diag+b2*treat+b3*q1+b4*q2+b5*q1*treat+b6*q2*treat+u;
  p = exp(eta)/(1+exp(eta));
  model outcome ~ binary(p);
  random u ~ normal(0, sig*sig ) subject=case;
```

The NLMIXED Procedure

AIC (smaller is better) 1176.8

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
a	0.9822	0.1844	339	5.33	<.0001	0.05	0.6194	1.3450	0.000363
b1	−1.3131	0.1543	339	−8.51	<.0001	0.05	−1.6165	−1.0097	0.000909
b2	2.0450	0.3129	339	6.54	<.0001	0.05	1.4296	2.6605	0.000101
b3	−0.9610	0.2313	339	−4.15	<.0001	0.05	−1.4160	−0.5060	−0.00049
b4	−0.6213	0.2256	339	−2.75	0.0062	0.05	−1.0650	−0.1775	0.000303
b5	−2.1002	0.3958	339	−5.31	<.0001	0.05	−2.8788	−1.3217	0.00004
b6	−1.0971	0.3852	339	−2.85	0.0047	0.05	−1.8548	−0.3394	−0.00046
sig	0.07027	1.1428	339	0.06	0.9510	0.05	−2.1777	2.3182	0.002123

The estimate $\hat{\sigma} = 0.07$ is small relative to the magnitude of the fixed effects. Let's refit the model without the random effects part:

```
proc nlmixed;
   parms a=1 b1=-1 b2=1 b3=-1.5 b4=-1 b5=-0.5 b6=-0.5;
   eta = a+b1*diag+b2*treat+b3*q1+b4*q2+b5*q1*treat+b6*q2*treat;
   p = exp(eta)/(1+exp(eta));
   model outcome ~ binary(p);
```

with output:

AIC (smaller is better) 1174.8

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
a	0.9812	0.1809	1020	5.43	<.0001	0.05	0.6263	1.3360	0.000029
b1	−1.3116	0.1462	1020	−8.97	<.0001	0.05	−1.5985	−1.0247	0.000048
b2	2.0430	0.3056	1020	6.68	<.0001	0.05	1.4432	2.6427	6.903E−6
b3	−0.9600	0.2290	1020	−4.19	<.0001	0.05	−1.4093	−0.5107	6.676E−6
b4	−0.6206	0.2245	1020	−2.76	0.0058	0.05	−1.0612	−0.1800	0.000017
b5	−2.0980	0.3893	1020	−5.39	<.0001	0.05	−2.8619	−1.3342	−4.79E−6
b6	−1.0961	0.3838	1020	−2.86	0.0044	0.05	−1.8491	−0.3431	0.000018

- The AIC *drops* without the random effects! We have rather strong evidence that observations within a cluster (an individual here, taken at 1, 2, and 4 weeks) are essentially independent when adjusted for baseline covariates.
- Note that the regression coefficients are essentially the same as those obtained from PROC GENMOD using the GEE approach. The absence of subject-to-subject heterogeneity implies that the marginal and conditional models are essentially the same.

13.3.5. Clinical trial example

Clinical trial with 8 centers; two creams compared to cure infection.

Center $Z = k$	Treatment X	Response Y		$\widehat{\theta}_{XY(k)}$
		Success	Failure	
1	Drug	11	25	1.2
	Control	10	27	
2	Drug	16	4	1.8
	Control	22	10	
3	Drug	14	5	4.8
	Control	7	12	
4	Drug	2	14	2.3
	Control	1	16	
5	Drug	6	11	∞
	Control	0	12	
6	Drug	1	10	∞
	Control	0	10	
7	Drug	1	4	2.0
	Control	1	8	
8	Drug	4	2	0.3
	Control	6	1	

Center-to-center variability in how people respond to treatment can be incorporated in the conditional model

$$\text{logit } P(Y_{ij} = 1) = \alpha + \beta x_{ij} + u_i, \quad u_1, \ldots, u_8 \overset{iid}{\sim} N(0, \sigma^2),$$

where $x_{ij} = 0$ for drug and $x_{ij} = 1$ for control.

SAS code:

```
data ctr1;
     input center$ treat s n @@; f=n-s; treat=treat-1;
     datalines;
     a 1 11 36 a 2 10 37  b 1 16 20 b 2 22 32
     c 1 14 19 c 2 7  19  d 1 2  16 d 2 1  17
     e 1 6  17 e 2 0  12  f 1 1  11 f 2 0  10
     g 1 1   5 g 2 1   9  h 1 4   6 h 2 6   7
    ;
data ctr2; set ctr1;
     do i=1 to n; if i<=s then y=1; else y=0; output; end;
proc nlmixed data=ctr2 qpoints=100;
     eta=alpha+beta*treat+u;
     p=exp(eta)/(1+exp(eta));
     model y ~ binary(p);
     random u ~ normal(0,sig*sig) subject=center; run;
```

with output:

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
alpha	−0.4591	0.5508	7	−0.83	0.4320	0.05	−1.7616	0.8433	0.000013
beta	−0.7385	0.3004	7	−2.46	0.0436	0.05	−1.4489	−0.02808	2.115E−6
sig	1.4008	0.4261	7	3.29	0.0133	0.05	0.3934	2.4083	0.000033

Within a given clinic, the odds of curing the infection is estimated to be (significantly) $1/e^{-0.739} = 2.1$ times greater on the drug versus the control. SAS will output empirical Bayes estimates of $u_1, \ldots, u_8$ by adding `out=re` (or whatever you want to call the new data set) to the `random` statement. Here they are:

Obs	center	Effect	Estimate	StdErr Pred	DF	tValue	Probt	Alpha	Lower	Upper
1	a	u	−0.09886	0.57554	7	−0.17177	0.86848	0.05	−1.45980	1.26208
2	b	u	1.85011	0.60147	7	3.07598	0.01792	0.05	0.42786	3.27235
3	c	u	0.99147	0.60198	7	1.64702	0.14355	0.05	−0.43199	2.41493
4	d	u	−1.29471	0.69606	7	−1.86006	0.10520	0.05	−2.94062	0.35121
5	e	u	−0.55775	0.64815	7	−0.86052	0.41800	0.05	−2.09038	0.97488
6	f	u	−1.60169	0.81836	7	−1.95719	0.09120	0.05	−3.53681	0.33343
7	g	u	−0.70444	0.76815	7	−0.91706	0.38961	0.05	−2.52081	1.11194
8	h	u	1.73721	0.74864	7	2.32047	0.05336	0.05	−0.03306	3.50747

Which clinic has the best overall success? Is it significant? Multiple testing needs to be considered, e.g. by controlling false discovery rate, particularly when the number of random effects is large.

Marginal model Using GEE:

```
proc genmod data=ctr2 descending; class center;
        model y = treat / dist=bin link=logit type3;
        repeated subject=center / type=exch corrw; run;
```

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	−0.3201	0.4111	−1.1259	0.4858	−0.78	0.4363
treat	−0.5540	0.2330	−1.0106	−0.0974	−2.38	0.0174

As expected, the marginal effect of -0.554 is similar to the corresponding marginal effect of $-0.7385/\sqrt{1+0.346*1.4008^2} = -0.5699$ computed from GLMM.

The GLMM is a hierarchical model, e.g. with logit link:

$$Y_{ij}|\mathbf{u}_i \overset{ind.}{\sim} \text{Bern}\left(\frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\mathbf{u}_i}}{1+e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\mathbf{u}_i}}\right),$$

$$\mathbf{u}_1,\ldots,\mathbf{u}_n \overset{iid}{\sim} N_q(\mathbf{0},\boldsymbol{\Sigma}).$$

Conditional on the random effect $\mathbf{u}_i$, the elements in $\mathbf{Y}_i = (Y_{i1},\ldots,Y_{iT_i})$ are independent. So the conditional PDF of $\mathbf{Y}_i|\mathbf{u}_i$ is

$$p(\mathbf{y}_i|\mathbf{u}_i) = \prod_{j=1}^{T_i}\left(\frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\mathbf{u}_i}}{1+e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\mathbf{u}_i}}\right)^{y_{ij}}\left(\frac{1}{1+e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\mathbf{u}_i}}\right)^{1-y_{ij}}.$$

However, the $\mathbf{u}_1,\ldots,\mathbf{u}_n$ are not model parameters. The model parameters are $(\boldsymbol{\beta},\boldsymbol{\Sigma})$. We need the to maximize the marginal likelihood

$$\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\Sigma}) = p(\mathbf{y}_1,\ldots,\mathbf{y}_n|\boldsymbol{\beta},\boldsymbol{\Sigma}).$$

The *unconditional* (or marginal) PDF of $\mathbf{Y}_i$ is

$$p(\mathbf{y}_i) = \int_{\mathbb{R}^q} \left[\prod_{j=1}^{T_i} \frac{(e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\mathbf{u}_i})^{y_{ij}}}{1+e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\mathbf{u}_i}} \right] p(\mathbf{u}_i|\boldsymbol{\Sigma})\mathrm{d}\mathbf{u}_i,$$

where $p(\mathbf{u}_i|\boldsymbol{\Sigma})$ is a $N_q(\mathbf{0}, \boldsymbol{\Sigma})$ PDF. The $\mathbf{u}_i$ is integrated out and this is a function of $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ only. The likelihood is the product of these

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \prod_{i=1}^{n} \int_{\mathbb{R}^q} \left[\prod_{j=1}^{T_i} \frac{(e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\mathbf{u}_i})^{y_{ij}}}{1+e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\mathbf{u}_i}} \right] p(\mathbf{u}_i|\boldsymbol{\Sigma})\mathrm{d}\mathbf{u}_i.$$

This involves n q-dimensional integrals that do not have closed-form.

PROC NLMIXED estimates the integrals (for a "current" quasi-Newton value of $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$) using adaptive Gauss-Hermite quadrature. This approach approximates the integrals above by sums

$$\int_{\mathbb{R}^q} h(\mathbf{u}_i) p(\mathbf{u}_i|\boldsymbol{\Sigma}) \mathrm{d}\mathbf{u}_i \approx \sum_{k=1}^{Q} c_k h(\mathbf{s}_k),$$

for arbitrary $h(\cdot)$ where Q is the number of quadrature points $\mathbf{s}_1, \ldots, \mathbf{s}_Q$ and $c_1, \ldots, c_Q$ are weights. The (adaptive) quadrature points and weights are chosen from a theory on integral approximations; we don't need to worry about that here.

Marginal model estimates the population-averaged effects of treatment while GLMM estimates subject-specific effects of treatment. Assuming a logit link, and let $Trt_{ij} = 0, 1$ for placebo or active drug, and $Race_i = 0, 1$ for Race = black or white. Consider a marginal model with the form of

$$\text{logit}(P(Y_{ij} = 1)) = \alpha_0 + \alpha_1 Trt_{ij} + \alpha_2 Race_i,$$

and a random effects model with the form of

$$\text{logit}(P(Y_{ij} = 1)) = \beta_0 + \beta_1 Trt_{ij} + \beta_2 Race_i + \mu_i,$$
$$\mu_i \sim N(0, \sigma^2).$$

- Interpretation of α_1 and β_1:
  - The estimated treatment effect from the marginal model (using GEE) describes how the odds of an outcome would increase (or decrease) in the study population if individuals were treated with the active drug (versus placebo) (or comparing treated individuals versus untreated individuals);
  - The estimated treatment effect from the GLMM describes how the odds of an outcome increases (or decreases) for a typical (or any) individual if treated with the active drug (versus placebo) (or comparing a typical treated individual versus a typical untreated individual);
  - Here "treatment" variable is considered as manageable or controllable, a counterfactual interpretation is thus easy to understand (i.e., comparing a typical subject if treated vs. if not treated).

- Interpretation of α_2 and β_2:
  - The estimated "Race" effect from the marginal model (using GEE) describes how the odds of an outcome would increase (or decrease) in the study population if individuals were white (versus if they were black) (or comparing white individuals versus black individuals);
  - The estimated "Race" effect from the GLMM describes how the odds of an outcome increases (or decreases) for a typical white individual versus a typical black individual (by using word "typical", it emphasizes the "conditional" effect that we are comparing a white individual with a black individual who have similar random effect μ_i).
  - Here "Race" variable is considered as not manageable/controllable, a counterfactual interpretation is a bit difficult to understand (i.e., conditional effect of a subject being "white" vs the subject being "black").

- 13.6.5 discusses testing $H_0 : \sigma = 0$ versus $H_1 : \sigma > 0$ in a simple model with univariate $u_1, \ldots, u_n \overset{iid}{\sim} N(0, \sigma^2)$. 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 $p = 0.5P(\chi^2_1 > t)$.
- Note that can have `model success~binomial(trials,prob);` in NLMIXED as well as other distributions; see the documentation.