7.5 Generalized Mixed-effects Regression Models for Binary and Count Responses

Binary and count responses in longitudinal studies can be modeled via **generalized mixed-effects regression models**. In the setting of each regression, we assume that the data were collected longitudinally at times $t_1, \ldots, t_m$, and that for the i-th individual at time t_j the observations are $x_{1ij}, \ldots, x_{kij}$, and y_{ij} where $i = 1, ..., n$, and $j = 1, ..., m$, and y_{ij}'s have either binary or Poisson distribution. The generalized mixed-effects model in each case has a random intercept term in the linear regression expression: $\beta_0 + \beta_1 x_{1ij} + \cdots + \beta_k x_{kij} + \beta_{k+1} t_j + u_i$ where u_i's are independent $\mathcal{N}(0, \sigma_u^2)$ random intercepts.

Example (Binary logistic random intercept model). A study was conducted on 49 school-aged children for the presence of respiratory tract infection. The subjects were evaluated by a doctor during four visits with a 3-month interval between consecutive visits. The file "respiratory_infection.csv" contains information on children's gender, baseline age, baseline BMI, and a presence of xerophthalmia (vitamin A deficiency).

Note: Xerophthalmia /ziro-fthal-mia/ (from Ancient Greek "xeros" meaning "dry" and "ophthalmos" meaning "eye").

We fit a binary logistic random intercept model (random slope and intercept model doesn't converge).

In SAS:

```
proc import out=resp_data datafile="./respiratory_infection.csv"
dbms=csv replace;

/*creating longform dataset*/
data longform;
set resp_data;
 array x[4] xerophthalmia1-xerophthalmia4;
 array i[4] infection1-infection4;
  do visit=1 to 4;
   xerophthalmia=x[visit];
     infection=i[visit];
   output;
   end;
```

```
keep medID gender age BMI xerophthalmia infection visit;
run;

/*fitting random intercept logistic regression model*/
proc glimmix method=Laplace;
 class gender(ref="M") xerophthalmia(ref="0");
  model infection = gender age BMI xerophthalmia visit /
     solution dist=binomial link=logit;
       random intercept/ subject=medID;
  covtest/wald;
run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	medID	6.0890	3.2813	1.86	0.0318

Solutions for Fixed Effects

Effect	gender	xerophthalmia	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			-3.9270	3.1707	45	-1.24	0.2219
gender	F		0.2631	0.9179	145	0.29	0.7748
gender	M		0	.	.	.	.
age			0.6434	0.2309	145	2.79	0.0060
BMI			-0.05979	0.1401	145	-0.43	0.6702
xerophthalmia		1	3.4378	0.8297	145	4.14	<.0001
xerophthalmia		0	0	.	.	.	.
visit			-1.0533	0.2798	145	-3.76	0.0002

From the output, the variance of the random intercept is statistically significant at the 5% level. Age, presence of xerophthalmia, and visit number are significant predictors at the 1% or smaller level. The fitted regression model

has the form

$$\ln \frac{\hat{P}(respiratory\ infection)}{1-\hat{P}(resperatory\ infection)} = \exp\big\{-3.9270+0.2631\cdot female+0.6434\cdot age$$
$$-0.05979\cdot BMI+3.4378\cdot xerophthalmia-1.0533\cdot visit\big\},$$

with the estimated variance of the random intercept $\hat{\sigma}^2_{u_1} = 6.0890$. For a one-year increase in age at the baseline, the estimated odds of respiratory tract infection increase by $(\exp(0.6434)-1)\cdot 100\% = 90.29399\%$. The estimated odds for subjects with xerophthalmia are $\exp(3.4378)\cdot 100\% = 3,111.842\%$ of those for subjects without it. As the visit number increases, the estimated odds change by $(\exp(-1.0533)-1)\cdot 100\% = -65.1215\%$, that is, decrease by 65.1215%.

In R:

```
infection.data<- read.csv(file="./respiratory_infection.csv", header=TRUE,
sep=",")

#creating long-form data set
library(reshape2)
data1<- melt(infection.data[,c("medID","gender","age","BMI","xerophthalmia1",
"xerophthalmia2","xerophthalmia3","xerophthalmia4")], id.vars=c("medID",
"gender","age","BMI"), variable.name="xero_visit", value.name="xerophthalmia")

data2<- melt(infection.data[,c("medID","infection1","infection2",
"infection3","infection4")], id.vars="medID",
variable.name="infection_visit", value.name="infection")
data2<- data2[c("infection")]

longform.data<- cbind(data1,data2)

#creating variable for visit
longform.data$visit<- ifelse(longform.data$xero_visit=="xerophthalmia1",1,
ifelse(longform.data$xero_visit=="xerophthalmia2",2,
ifelse(longform.data$xero_visit=="xerophthalmia3",3,4))

#specifying reference category
longform.data$gender.rel<- relevel(as.factor(longform.data$gender), ref="M")
longform.data$xerophthalmia.rel<- relevel(as.factor(longform.data$xerophthalmia),
```

```
ref="0")

#cleaning long-form data
longform.data<- longform.data[!names(longform.data) %in%
c("gender", "xero_visit","xerophthalmia")]

#fitting logistic model with random intercept
#install.packages("Rcpp")
library(lme4)
summary(fitted.model<- glmer(infection ~ gender.rel + age+ BMI
+ xerophthalmia.rel + visit + (1 | medID), data=longform.data,
family=binomial(link="logit")))
```

```
Random effects:
 Groups Name        Variance Std.Dev.
 medID  (Intercept) 6.069    2.463

Fixed effects:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)        -3.91752    3.15707  -1.241 0.214652
gender.relF         0.26254    0.91580   0.287 0.774360
age                 0.64225    0.22899   2.805 0.005035
BMI                -0.05974    0.13982  -0.427 0.669197
xerophthalmia.rel1  3.43409    0.82409   4.167 3.08e-05
visit              -1.05238    0.27849  -3.779 0.000158
```

Finally, we predict the probability of a respiratory infection for a 10-year old boy with a BMI of 15.6, and with symptoms of xerophthalmia at the second visit. We compute

$$\mathbb{P}^0(infection) = \frac{\exp\big\{-3.9270 + 0.6434 \cdot 10 - 0.05979 \cdot 15.6 + 3.4378 - 1.0533 \cdot 2\big\}}{1 + \exp\big\{-3.9270 + 0.6434 \cdot 10 - 0.05979 \cdot 15.6 + 3.4378 - 1.0533 \cdot 2\big\}}$$
$$= 0.948116.$$

In SAS:

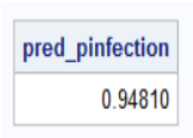
```
/*using fitted model for prediction*/
data predict;
input medID gender$ age BMI xerophthalmia visit;
cards;
```

```
999999 M 10 15.6 1 2
;

data longform;
set longform predict;
run;

proc glimmix method=Laplace;
  class gender xerophthalmia;
  model infection = gender age BMI xerophthalmia visit /
    dist=binomial link=logit;
      random intercept/ subject=medID;
   output out=outdata pred(ilink)=pred_pinfection;
run;

proc print data=outdata(firstobs=197) noobs;
 var pred_pinfection;
run;
```



pred_pinfection
0.94810

In R:

#using fitted model for prediction
new.data<- rbind(longform.data, data.frame(medID="999999",
age=10, BMI=15.6, infection=NA, visit=2, gender.rel="M",
xerophthalmia.rel=1))

pred<- predict(fitted.model, new.data, re.form=NA, type="response")
pred[length(pred)]

```
0.9479643
```

The argument **re.form=NA** requests predicted response not conditioned on any values of the random-effects terms. □

Example (Poisson random intercept model). The Skin Cancer Prevention Study was a multi-center randomized clinical trial to evaluate beta-carotene as a treatment for preventing non-melanoma skin cancer in high-risk subjects. Subjects in the trial were randomized to receive beta-carotene or placebo for five years, and each subject was to return once a year to be examined for new skin cancers. The data are given in the file "skin_cancer_data.csv" with the following columns: patient ID, group, age at baseline, gender, the number of previous (pre-study) skin cancers, year, and the number of new cancers that year. We use a Poisson regression with random intercept to model the number of new cancers.

In SAS:

```
proc import out=cancer_data
datafile="./skin_cancer_data.csv" dbms=csv replace;

/*fitting Poisson model with random slope and intercept*/
 proc glimmix method=Laplace;
   class group(ref="Tx") gender(ref="F");
  model nnewcancers = group age gender nprevcancers
            year / solution dist=poisson link=log;
   random intercept / subject=ptID type=un;
    covtest/wald;
run;
```

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
UN(1,1)	ptID	0.4013	0.1154	3.48	0.0003

Solutions for Fixed Effects							
Effect	group	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			-0.5293	0.6639	63	-0.80	0.4283
group	Cx		0.8362	0.2081	255	4.02	<.0001
group	Tx		0	.	.	.	.
age			0.003414	0.009569	255	0.36	0.7216
gender		M	0.1616	0.1929	255	0.84	0.4029
gender		F	0	.	.	.	.
nprevcancers			0.06861	0.03401	255	2.02	0.0447
year			-0.07585	0.03142	255	-2.41	0.0165

From the output, we see that the variance of the random intercept is significant, indicating that the model is appropriate. Group, number of previous cancers, and year are significant predictors. The fitted model is

$$\hat{\mathbb{E}}(number\ of\ new\ cancers) = \exp\big(-0.5293 + 0.8362 \cdot control + 0.003414 \cdot age + 0.1616 \cdot male + 0.06861 \cdot nprevcancers - 0.07585 \cdot year\big),$$

and $\hat{\sigma}_u^2 = 0.4013$. For the subjects in the control group, the estimated mean number of new cancers is $\exp(0.8362) \cdot 100\% = 230.7581\%$ of that for subjects in the treatment group. As the number of pre-study cancers increases by 1, the estimated mean number of new cancers increases by $(\exp(0.06861) - 1) \cdot 100\% = 7.101843\%$. As the study year increases by one, the estimated mean number of new cancers changes by $\exp(-0.07585) - 1) \cdot 100\% = -7.30448\%$, that is, decreases by 7.30448%.

In R:
cancer.data<- read.csv(file="./skin_cancer_data.csv", header=TRUE, sep=",")

#specifying reference category
cancer.data$group.rel<- relevel(as.factor(cancer.data$group), ref="Tx")
cancer.data$gender.rel<- relevel(as.factor(cancer.data$gender), ref="F")

#cleaning the data set
cancer.data<- cancer.data[!names(cancer.data) %in% c("group", "gender")]

#fitting random intercept Poisson model
#install.packages("Rcpp")
library(lme4)

summary(fitted.model<- glmer(nnewcancers ∼ group.rel + age + gender.rel
+ nprevcancers + year + (1|ptID), data=cancer.data,
family=poisson(link="log")))

```
Random effects:
 Groups Name        Variance Std.Dev.
 ptID   (Intercept) 0.4005   0.6329

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.527276   0.658085  -0.801   0.4230
group.relCx   0.835065   0.206445   4.045 5.23e-05
age           0.003410   0.009497   0.359   0.7196
gender.relM   0.161469   0.191521   0.843   0.3992
nprevcancers  0.068573   0.033890   2.023   0.0430
year         -0.075862   0.030940  -2.452   0.0142
```

Next, we predict the number of new cancers in year 3 for a 57-year old female patient in the treatment group who had 8 previous cancers. We calculate

$$(number\ of\ new\ cancers)^0 = \exp\big(-0.5293 + 0.003414 \cdot 57 + 0.06861 \cdot 8 - 0.07585 \cdot 3\big) = 0.986717.$$

In SAS:

```
/*using the fitted model for prediction
data predict;
```

```
input ptID group$ age gender$ nprevcancers year;
cards;
111111 Tx 57 F 8 3
;

data cancer_data;
set cancer_data predict;
run;

proc glimmix method=Laplace;
 class group gender;
  model nnewcancers = group age gender nprevcancers
            year / dist=poisson link=log;
   random intercept / subject=ptID type=un;
     output out=outdata pred(ilink)=pred_nnewcancers;
run;

proc print data=outdata (firstobs=325) noobs;
var pred_nnewcancers;
run;
```

pred_nnewcancers
0.98668

In R:
#using fitted model for prediction
new.data<- rbind(cancer.data, data.frame(ptID="111111",
age=57, nprevcancers=8, year=3, nnewcancers=NA, group.rel="Tx",
gender.rel="F"))
pred<- predict(fitted.model, new.data, re.form=NA, type="response")
pred[length(pred)]

```
0.988165
```

□