

Homework problem, Chapter 11 problem 6:

Here's my SAS code:

```
data abc1;
  input seq a b c count @@;
  datalines;
1 0 0 0 0 1 0 0 1 2 1 0 1 0 2 1 0 1 1 9 1 1 0 0 0 1 1 0 1 0 1 1 1 0 1 1 1 1 1 1 1
2 0 0 0 2 2 0 0 1 0 2 0 1 0 0 2 0 1 1 9 2 1 0 0 1 2 1 0 1 0 2 1 1 0 0 2 1 1 1 4
3 0 0 0 0 3 0 0 1 1 3 0 1 0 1 3 0 1 1 8 3 1 0 0 1 3 1 0 1 3 3 1 1 0 0 3 1 1 1 1
4 0 0 0 0 4 0 0 1 1 4 0 1 0 1 4 0 1 1 8 4 1 0 0 1 4 1 0 1 0 4 1 1 0 0 4 1 1 1 1
5 0 0 0 3 5 0 0 1 0 5 0 1 0 0 5 0 1 1 7 5 1 0 0 0 5 1 0 1 1 5 1 1 0 2 5 1 1 1 1
6 0 0 0 1 6 0 0 1 5 6 0 1 0 0 6 0 1 1 4 6 1 0 0 0 6 1 0 1 3 6 1 1 0 1 6 1 1 1 0
;
data abc2; set abc1;
  case=0;
  do i=1 to count;
    case=case+1;
    pattern=4*a+2*b+c;
    y=a; treat=1; output;
    y=b; treat=2; output;
    y=c; treat=3; output;
  end;
proc print;
proc genmod descending; class pattern case treat seq;
  model y=treat seq / dist=bin link=logit;
  repeated subject=case(seq*pattern) / type=exch;
  estimate '3 vs 1' treat -1 0 1 / exp;
  estimate '2 vs 1' treat -1 1 0 / exp;
  estimate '3 vs 2' treat 0 -1 1 / exp;
```

with output:

GEE Model Information

Correlation Structure	Exchangeable
Subject Effect	case(pattern*seq) (86 levels)
Number of Clusters	86
Correlation Matrix Dimension	3
Maximum Cluster Size	3
Minimum Cluster Size	3

Exchangeable Working

Correlation

Correlation -0.04403048

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha	Confidence Limits	Chi-Square	Pr > ChiSq
3 vs 1	2.5076	0.4141	0.05	1.6959 3.3193	36.66	<.0001
Exp(3 vs 1)	12.2750	5.0836	0.05	5.4513 27.6400		
2 vs 1	1.9914	0.3876	0.05	1.2317 2.7511	26.39	<.0001
Exp(2 vs 1)	7.3257	2.8396	0.05	3.4270 15.6599		
3 vs 2	0.5162	0.3158	0.05	-0.1029 1.1352	2.67	0.1022
Exp(3 vs 2)	1.6756	0.5292	0.05	0.9023 3.1118		

Empirical Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > Z
Intercept		0.9554	0.3282	0.3121	1.5987	2.91	0.0036
treat	1	-2.5076	0.4141	-3.3193	-1.6959	-6.05	<.0001
treat	2	-0.5162	0.3158	-1.1352	0.1029	-1.63	0.1022
treat	3	0.0000	0.0000	0.0000	0.0000	.	.
seq	1	0.5200	0.3907	-0.2459	1.2858	1.33	0.1833
seq	2	0.7775	0.5352	-0.2715	1.8265	1.45	0.1463
seq	3	0.6454	0.3865	-0.1122	1.4029	1.67	0.0950
seq	4	0.5830	0.4230	-0.2460	1.4121	1.38	0.1681
seq	5	0.2384	0.5116	-0.7642	1.2410	0.47	0.6412
seq	6	0.0000	0.0000	0.0000	0.0000	.	.

I am nesting the subject (case) index within both the drug sequence $k = 1, \dots, 6$ and pattern type $p = 1, \dots, 8$ for $(0, 0, 0), (0, 0, 1), \dots, (1, 1, 1)$. The model looks like

$$\text{logit } P(Y_{i(k*p)j} = 1) = \gamma + \alpha_k + \beta_j,$$

where $\beta_3 = \alpha_6 = 0$ correspond to baseline, a bit different than what is asked for in your homework.

11.5 Markov chains for transitional modeling

When j indexes time, $Y_{i1}, Y_{i2}, \dots, Y_{iT_i}$ is a stochastic process, often termed a *time series*. Let's consider $Y_{ij} = 0, 1$ for now.

The series $Y_{i1}, Y_{i2}, \dots, Y_{iT_i}$ follows a first-order Markov chain if the distribution of Y_{ij} only cares about the previous value $Y_{i,j-1}$, formally $[Y_{ij}|Y_{i1}, \dots, Y_{i,j-1}] = [Y_{ij}|Y_{i,j-1}]$.

Time-varying covariates can be included:

$$\text{logit } P(Y_{ij} = 1|Y_{i,j-1}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma_1 Y_{i,j-1},$$

where γ_1 models the effect of the i^{th} subject's previous observation on the probability of a current (time j) success. e^{γ_1} has a nice interpretation in terms of how success odds changes based on what happened at last time point.

Second-order, and in general t -order, Markov chains can be considered by including the most previous t observations $(Y_{i,j-1}, \dots, Y_{i,j-t})$:

$$\text{logit } P(Y_{ij} = 1 | Y_{i,j-1}, \dots, Y_{i,j-t}) = \mathbf{x}'_{ij} \boldsymbol{\beta} + \sum_{s=1}^t \gamma_s Y_{i,j-s}.$$

Interactions between covariates \mathbf{x}_{ij} and previous values can also improve model fit.

For a first order Markov-chain with no interaction the likelihood is written

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^n f_1(y_{i1}) f_2(y_{i2}|y_{i1}) f_3(y_{i3}|y_{i2}) \cdots f_{T_i}(y_{iT_i}|y_{i,T_i-1}).$$

if we ignore the marginal contribution of the first observation $f_1(y_{i1})$ we get

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^n f_2(y_{i2}|y_{i1}) f_3(y_{i3}|y_{i2}) \cdots f_{T_i}(y_{iT_i}|y_{i,T_i-1}).$$

For each subject i we have the product of $T_i - 1$ conditional logistic regression kernels; the transitional model can be fit in PROC LOGISTIC as usual, but for observation Y_{ij} , treating $Y_{i,j-1}$ as an observed predictor!

Example (p. 480): Children were evaluated every year on whether they had a respiratory illness. A covariate of interest is whether the child's mom smoked at the beginning of the study; $s_i = 0$ indicates not and $s_i = 1$ indicates a smoker.

Each child has a sequence of 4 indicators $(Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ taken at 7, 8, 9, and 10 years. For each child we have covariates s_i and $t_j = j + 6$. The first order Markov model is fit

$$\text{logit } P(Y_{ij} = 1 | Y_{i,j-1} = y_{i,j-1}) = \beta_0 + \beta_1 s_i + \beta_2 t_j + \beta_3 y_{i,j-1},$$

for $i = 1, \dots, 537$ and $j = 2, 3, 4$.

SAS code to fit the Markov model:

```
data mm1;
input s y1 y2 y3 y4 count;
y=y2; yp=y1; sm=s; t=8; ct=count; output;
y=y3; yp=y2; sm=s; t=9; ct=count; output;
y=y4; yp=y3; sm=s; t=10; ct=count; output;
datalines;
0 0 0 0 0 237
0 0 0 0 1 10
0 0 0 1 0 15
0 0 0 1 1 4
0 0 1 0 0 16
0 0 1 0 1 2
0 0 1 1 0 7
0 0 1 1 1 3
etc...
1 1 1 0 0 4
1 1 1 0 1 2
1 1 1 1 0 4
1 1 1 1 1 7
;
proc logistic descending;
freq ct; model y=sm t yp / lackfit;
```

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-0.2926	0.8460	0.1196	0.7295	.	.
SM	1	0.2960	0.1563	3.5837	0.0583	0.077761	1.344
T	1	-0.2428	0.0947	6.5800	0.0103	-0.109336	0.784
YP	1	2.2111	0.1582	195.3589	0.0001	0.450688	9.126

Hosmer and Lemeshow Goodness-of-Fit Test

Goodness-of-fit Statistic = 1.1723 with 6 DF (p=0.9782)

We see both time and whether the child had a respiratory illness the previous year are important predictors. Smoking is *almost* significant at the 5% level (and is significant if we perform a one-sided test). Maternal smoking increases the odds of a respiratory illness by about 34%. As time goes on the child is less likely to have a respiratory illness. If a child had a respiratory illness last year, the odds of having one this year are nine times greater than if the child did not have one last year.

Chapter 12: Generalized Linear Mixed Models

- Observations often occur in related clusters. Phrases like *repeated measures* and *longitudinal data* get at the same thing: there's correlation among observations in a cluster.
- Chapter 11 dealt with an estimation procedure (GEE) that accounted for correlation in estimating population-averaged (marginal) effects.
- This chapter models cluster correlation explicitly through *random effects*, yielding a GLMM.

Let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iT_i})$ be T_i correlated responses in cluster i . Associated with each repeated measure Y_{ij} are fixed (population) effects β and cluster-specific random effects \mathbf{u}_i . As usual, $\mu_{ij} = E(Y_{ij})$.

In a GLMM the linear predictor is augmented to include random effects:

$$g(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i.$$

for logistic regression, this is

$$\text{logit } P(Y_{ij} = 1) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i.$$

Note that conditional on \mathbf{u}_i ,

$$E(Y_{ij}|\mathbf{u}_i) = \frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i}}{1 + e^{\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i}}.$$

Example: I ask a random sample of *the same* $n = 30$ graduate students “do you like statistics?” once a month for 4 months.

$Y_{ij} = 1$ if “yes” and $Y_{ij} = 0$ if no. Here, $i = 1, \dots, 30$ and $j = 1, \dots, 4$.

Covariates might include m_{ij} , the average mood of the student over the previous month ($m_{ij} = 0$ is bad, $m_{ij} = 1$ is good), the degree being sought ($d_i = 0$ doctoral, $d_i = 1$ masters), the month $t_j = j$, and p_j the number of homework problems assigned in PubH 7407 in the previous month.

A GLMM might be

$$\text{logit } P(Y_{ij} = 1) = \beta_0 + \beta_1 m_{ij} + \beta_2 d_i + \beta_3 p_j + \beta_4 j + u_i.$$

This model assumes that log-odds of liking statistics changes linearly in time, holding all else constant. Alternatively, we might fit a quadratic instead or treat time as categorical. Here, u_i represents a student’s *a priori* disposition towards statistics.

Let's compare month $j + 1$ to month j for individual i , holding all else (m , d , and p) constant. The difference in log odds is

$$(\beta_0 + \beta_1 m_{ij} + \beta_2 d_i + \beta_3 p_j + \beta_4(j + 1) + u_i) - (\beta_0 + \beta_1 m_{ij} + \beta_2 d_i + \beta_3 p_j + \beta_4 j + u_i) = \beta_4.$$

Not holding everything constant we get

$$\begin{aligned} &(\beta_0 + \beta_1 m_{i,j+1} + \beta_2 d_i + \beta_3 p_{j+1} + \beta_4(j + 1) + u_i) - (\beta_0 + \beta_1 m_{ij} + \beta_2 d_i + \beta_3 p_j + \beta_4 j + u_i) \\ &= \beta_1(m_{i,j+1} - m_{ij}) + \beta_3(p_{j+1} - p_j) + \beta_4. \end{aligned}$$

Either way, we are conditioning on individual i , or *the subpopulation of all individuals with predisposition u_i* ; i.e. everyone “like” individual i in terms of liking statistics to begin with.

How are e^{β_1} , e^{β_2} , e^{β_3} and e^{β_4} interpreted here?

The random effects are assumed to come from (in general) a multivariate normal distribution

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

The covariance $\text{cov}(\mathbf{u}_i) = \boldsymbol{\Sigma}$ can have special structure, e.g. exchangeability, AR(1), or be unstructured. The free elements of $\boldsymbol{\Sigma}$ are estimated along with $\boldsymbol{\beta}$.

- The \mathbf{u}_i can account for heterogeneity caused by omitting explanatory variables.
- They can also explicitly model overdispersion, e.g.

$$Y_i \sim \text{Pois}(\lambda_i), \quad \log \lambda_i = \mathbf{x}'_i \boldsymbol{\beta} + u_i, \quad u_i \stackrel{iid}{\sim} N(0, \sigma^2).$$

Logit model for binary matched pairs

Recall $j = 1, 2$ denotes a binary covariate; for the PMA data it's time.

$$\text{logit } P(Y_{ij} = 1) = \alpha + u_i + \beta I\{j = 2\}.$$

Here, e^β is a cluster-specific odds ratio. We further assume $u_i \stackrel{iid}{\sim} N(0, \sigma^2)$.

Example: PMA data. Although a closed form estimate of β exists (see p. 494), we'll fit this in SAS using two different data structures for illustrative purposes.

```
data Data1;
  do ID=1 to 794; ap=1; time=0; output; ap=1; time=1; output; end;
  do ID=795 to 944; ap=1; time=0; output; ap=0; time=1; output; end;
  do ID=945 to 1030; ap=0; time=0; output; ap=1; time=1; output; end;
  do ID=1031 to 1600; ap=0; time=0; output; ap=0; time=1; output; end;
proc logistic data=Data1; strata ID;
  model ap(event='1')=time;
proc genmod data=Data1 descending; class ID;
  model ap=time / link=logit dist=bin;
  repeated subject=ID / corr=exch corrw;
```

On previous slide, first is conditional logistic approach from Chapter 10, second is marginal GEE logistic approach from Chapter 11.

Here is the GLMM approach of Chapter 12 with $u_i \stackrel{iid}{\sim} N(0, \sigma^2)$:

```
proc nlmixed maxiter=100 qpoints=100;
  parms beta0=1.0 beta1=-0.556 sigma=5.2;
  eta = beta0+beta1*time+u; pi = exp(eta)/(1+exp(eta));
  model ap ~ binary(pi);
  random u ~ normal(0,sigma*sigma) subject=ID;
  estimate 'subject-specific odds at 6 months' exp(beta1);
data matched;
input case occasion response count @@; datalines;
1 0 1 794    1 1 1 794    2 0 1 150    2 1 0 150
3 0 0 86    3 1 1 86    4 0 0 570    4 1 0 570
;
proc nlmixed maxiter=100 qpoints=100;
  eta = alpha + beta*occasion + u; p = exp(eta)/(1 + exp(eta));
  model response ~ binary(p);
  random u ~ normal(0, sigma*sigma) subject = case; replicate count;
```

Output from the first fit:

Parameter Estimates

Parameter	Estimate	Standard	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
		Error							
beta0	1.2424	0.1857	1599	6.69	<.0001	0.05	0.8781	1.6067	-4.72E-7
beta1	-0.5563	0.1353	1599	-4.11	<.0001	0.05	-0.8216	-0.2910	-3.05E-7
sigma	5.1593	0.3527	1599	14.63	<.0001	0.05	4.4676	5.8510	8.779E-7

Additional Estimates

Label	Estimate	Standard	DF	t Value	Pr > t	Alpha	Lower
		Error					
subject-specific odds at 6 months	0.5733	0.07755	1599	7.39	<.0001	0.05	0.4212

Additional Estimates

Label	Upper
subject-specific odds at 6 months	0.7254

Read through **12.1.5**: random effects versus conditional approach.

12.2 Logistic normal model

A special, often-used case of the GLMM.

The logistic normal model is given by:

$$\text{logit } P(Y_{ij} = 1|u_i) = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_i, \quad u_i \stackrel{iid}{\sim} N(0, \sigma^2).$$

When $\sigma = 0$ we get the standard logistic regression model, when $\sigma > 0$ we account for extra heterogeneity in clustered responses (each i is a cluster with its own random u_i).

Connection between marginal and conditional models

In the GEE approach, the marginal means are explicitly modeled:

$$\mu_{ij} = E(Y_{ij}) = g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta}),$$

and correlation among $(Y_{i1}, \dots, Y_{iT_i})$ is accounted for in the estimation procedure.

The conditional approach models the means conditional on the random effects:

$$E(Y_{ij}|\mathbf{u}_i) = g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i).$$

The corresponding marginal mean is given by

$$E(Y_{ij}) = \int_{\mathbb{R}^q} g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i) f(\mathbf{u}_i; \boldsymbol{\Sigma}) d\mathbf{u}_i.$$

In general, this is a complicated function of β , however for the logistic-normal model when σ is “small,” we obtain (not obvious)

$$E(Y_{ij}) \approx \exp(c\mathbf{x}'_{ij}\beta)/[1 + \exp(c\mathbf{x}'_{ij}\beta)],$$

where $c = 1/\sqrt{1 + 0.6\sigma^2}$. The *marginal odds* change by approximately $e^{c\beta_s}$ when x_{ijs} is increased by unity.

Since $c < 1$, the marginal effect is smaller than the conditional effect, reflecting that we are averaging with respect to the population. Note that the larger σ is, the more subject-to-subject variability there is, and the *smaller* the averaged effect $\hat{c}\hat{\beta}_s$ becomes.

PMA data, a final look. Here, $\hat{c} = 1/\sqrt{1 + 0.6(5.16)^2} = 0.24$. Then $e^{-0.556(0.24)} = 0.87$. Recall that the GEE approach yields $e^{-0.163} = 0.85$; not a bad approximation! Also recall that the conditional approach yielded $e^{-0.556}$. Severely annotated output:

The LOGISTIC Procedure
 Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
time	1	-0.5563	0.1353	16.9152	<.0001

The GENMOD Procedure

Exchangeable Working
 Correlation

Correlation 0.7023650596

Analysis Of GEE Parameter Estimates
 Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > Z
Intercept	0.3640	0.0508	0.2643	0.4636	7.16	<.0001
time	-0.1633	0.0390	-0.2398	-0.0868	-4.18	<.0001

The NLMIXED Procedure
Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
beta0	1.2424	0.1857	1599	6.69	<.0001	0.05	0.8781	1.6067	-4.72E-7
beta1	-0.5563	0.1353	1599	-4.11	<.0001	0.05	-0.8216	-0.2910	-3.05E-7
sigma	5.1593	0.3527	1599	14.63	<.0001	0.05	4.4676	5.8510	8.779E-7

Text comments:

- In epi studies, often want to compare disease prevalence across groups. Then it's of interest to compute marginal odds ratios and compare them.
- We did not discuss MLE approach to marginal models; uses a huge multinomial distribution; can be unstable. See text.
- Direction and significance of effects usually the same across marginal/conditional models (e.g. PMA data).

- The more variability that's accounted for in the conditional model, the more we can “focus in” on the conditional effect of covariates. This is true in any situation where we block. This has the effect enlarging $\hat{\beta}_s$ estimates under a conditional model.
- When correlation is small, independence is approximately achieved, and marginal and conditional modeling yield similar results.
- GLMMs are being increasingly used, in part due to the availability of standard software to fit them!
- Bayesian approach is also natural here.