Solution to HW12, ST544

Problem 1

The subject-specific logistic model for data in problem 9.15 is

$$\text{logit}[P(obesity|u_i)] = u_i + \alpha + \beta_1 male + \beta_2 time + \beta_3 male \times time.$$

```
data prob9_15;
input sex $ n1-n8;
datalines;
Male   119 7 8 3 13 4 11 16
Female 129 8 7 9 6 2 7 14
;

title "Recover individual data; Y=1/0 for normal/obese";
data prob9_15; set prob9_15;
array temp {8} n1-n8;

male=(sex="Male");
retain id;
if _n_=1 then id=0;

do k=1 to 8;
do i=1 to temp(k);
id = id + 1;
do j=1 to 3;
time=j-1; /*time = 0, 1, 2, for 77, 79, 81 */
if k=1 then y = 1;
if k=2 then y = (j ne 3);
if k=3 then y = (j ne 2);
if k=4 then y = (j = 1);
if k=5 then y = (j ne 1);
if k=6 then y = (j = 2);
if k=7 then y = (j = 3);
if k=8 then y = 0;
obese = 1-y;
output;
end;
end;
end;
run;

proc glimmix;
class id;
model obese = male time male*time / s dist=bin link=logit;
random int / subject=id;
run;
****************************************************************************

Covariance Parameter Estimates

Standard
Cov Parm     Subject    Estimate       Error

Intercept    id           2.3136      0.3444


Solutions for Fixed Effects

Standard
Effect       Estimate       Error       DF    t Value    Pr > |t|

Intercept     -1.8926      0.2349      361      -8.06      <.0001
male           0.5705      0.3183      724       1.79      0.0735
time          0.09343      0.1529      724       0.61      0.5415
male*time     -0.3934      0.2125      724      -1.85      0.0645
```

The estimated variance of u_i is $\hat{\sigma}^2 = 2.3136$. The estimated logit of the obesity probability over time for a girl is

$$\text{logit}[P(obesity|u_i)] = u_i - 1.8926 + 0.09343time,$$

indicating that for a (randomly selected) girl, her odds-ratio of obesity per year is $e^{0.09343} = 1.10$, a 10% increase in odds per year (even though statistically not significant).

The estimated logit of the obesity probability over time for a boy is

$$\text{logit}[P(obesity|u_i)] = u_i - 1.3221 - 0.3time.$$

indicating that for a (randomly selected) boy, his odds-ratio of obesity per year is $e^{-0.3} = 0.74$, a 26% decrease in odds per year.

Using the approximate formula on slide 482, the approximate population-averaged coefficient estimates are: -1.41, 0.43, 0.07, 0.29, somewhat similar to what we got from problem 9.15.

Problem 2

The sas program to fit the model on slide 521 after combining the last two categories as one category:

```
options ls=80 ps=200 nodate nodate;

data table9_6;
input trt y0 y1-y4;
cards;
1 1  7  4  1  0
1 2 11  5  2  2
1 3 13 23  3  1
1 4  9 17 13  8
0 1  7  4  2  1
0 2 14  5  1  0
0 3  6  9 18  2
0 4  4 11 14 22
;

title "Recover individual data";
data table9_6; set table9_6;
  array temp {4} y1-y4;

  retain id;
  if _n_=1 then id=0;

  do k=1 to 4;
    do i=1 to temp(k);
      id = id + 1;
      do time=0 to 1;
        if time=0 then y=y0;
        else y=k;

        output;
      end;
    end;
  end;
run;

data table9_6; set table9_6;
  if y=4 then y=3; * combine the last 2 categories;
run;
```

```
title "GLMM for new insomnia data";
  proc Glimmix data=table9_6;
  class id;
  model y = time trt time*trt / dist=multinomial link=clogit s;
  random int / subject=id type=vc;
run;
```

The estimates (SE) are: $\widehat{\alpha}_1 = -2.63$, $\widehat{\alpha}_2 = -0.99$, $\widehat{\beta}_1 = 1.18(0.28)$, $\widehat{\beta}_2 = -0.04(0.34)$, $\widehat{\beta}_3 = 0.90(0.39)$. The estimate of the random effect variance is $\widehat{\sigma}^2 = 1.48(0.35)$. These estimates are somewhat different from those on slide 523.

For any placebo individual, the odds of having a shorter time to falling asleep 2 weeks later (after receiving the placebo) is $e^{1.18} = 3.25$ times his/her odds of having a shorter time to falling asleep at baseline. For any treated individual, the odds of having a shorter time to falling asleep 2 weeks later (after receiving the treatment) is $e^{1.18+0.9} = 8.0$ times his/her odds of having a shorter time to falling asleep at baseline, which is significantly greater than that for the placebo group (p-value=0.02).

There is not much difference between the treatment group and placebo group at the baseline (p-value=0.90).

Problem 10.5

The SAS program for this problem is in the following:

```
options ls=80 ps=110;

data coin;
  input y @@;
  n=5; p = y/n; id=_n_;
  datalines;
  2 4 1 3 3 5 4 2 3 1
;
run;

title "Problem 10.5(a)";
proc genmod data=coin;
  class id;
  model y/n = id / noint link=identity;
run;

title "Problem 10.5(b) & (c)";
proc glimmix data=coin; * method=quad(qpoints=19);
  class id;
  model y/n = / dist=bin link=logit s;
  random int / subject=id type=vc s;
  output out=out pred(ilink)=pihat;
run;

data out; set out;
  dist1 = abs(p-0.5);
  dist2 = abs(pihat-0.5);
run;

title "Problem 10.5(d) & (e)";
proc print;
run;
```

(a) See the SAS program.

(b) See the SAS program. The estimated intercept is $\widehat{\alpha} = 0.25$, the estimated variance of the random intercpet is $\widehat{\sigma}^2 = 0.36$, indicating there is some variation in the head probabilities among those 10 coins.

(c) See the SAS program.

(d) We would prefer the estimated $\widehat{\pi}_i$ from a GLMM since they will have smaller variances.

(e) Except for one coin, the absolute distances for other coins using $\widehat{\pi}_i$ are much smaller than those using the sample proportions.

Problem 10.6

(a) The estimated $\widehat{\beta}_t$ indicates relative subject-specific magnitude of using substance t. For example, the subject-specific odds-ratio between cigarette use and marijuana use is

$$e^{\widehat{\beta}_1 - \widehat{\beta}_3} = e^{4.2227-(-0.7751)} = 148,$$

indicating that for any individual, the odds of using cigarette is 148 times the odds of marijuana use.

(b) The estimated $\widehat{\sigma} = 3.55$ ($\widehat{\sigma}^2 = 12.6$) indicates that there is a large variation in the subject-specific probabilities in substance use between individuals.

(c) A large positive value for u_i indicates that subject i has large probabilities of using all three substance.