Solution to HW11, ST544

Problem 9.4: We fit the following GEE model to the depression data with the original time (see the model on slide 457):

$$\text{logit}\{\pi(s,t,d)\} = \alpha + \beta_1 s + \beta_2 d + \beta_3 t + \beta_4 (d \times t).$$

The SAS program for fitting the above model is:

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

data table9_1;
input severity $ treatment $ y1-y8;
cards;
Mild   Standard 16 13 9 3 14  4 15  6
Mild   Newdrug  31  0 6 0 22  2  9  0
Severe Standard  2  2 8 9  9 15 27 28
Severe Newdrug   7  2 5 2 31  5 32  6
;

title "Recover individual data";
data table9_1; set table9_1;
array temp {8} y1-y8;

trt = (treatment="Newdrug");
sev = (severity="Severe");
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;
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;
output;
end;
end;
end;
run;


data newdata; set table9_1;
time = 2**(time);
run;


title "Treatment for Depression: Original time and Exch working corr";
proc genmod descending;
class id;
model y = sev trt  time trt*time / dist=bin link=logit;
repeated subject=id / type=exch corrw;
run;
```

The correlation is estimated to be -0.0044. The estimates of β's (SE of the estimate) are: $\widehat{\beta}_1 = -1.30(0.14)$, $\widehat{\beta}_2 = -0.64(0.32)$, $\widehat{\beta}_3 = 0.32(0.08)$, $\widehat{\beta}_4 = 0.71(0.14)$. The estimate of β_1 is almost the same as that in the textbook. However, other estimates are very different.

The treatment effect (odds-ratio) at different time points is estimated as from this model:

$$\begin{aligned}\widehat{\theta}(s,t) &= e^{\widehat{\beta}_4 t+\widehat{\beta}_2} = e^{0.71t-0.64} \\ &= \begin{cases} e^{0.71-0.64} = 1.07 & t=1 \\ e^{0.71\times 2-0.64} = 2.18 & t=2 \\ e^{0.71\times 4-0.64} = 9.03 & t=4 \end{cases}\end{aligned}$$

giving a somewhat different estimate of $\widehat{\theta}(s,t)$ at $t=4$.

Problem 9.8

(a) All those estimated correlation coefficients are very similar. this suggests that an exchangeable working correlation matrix for the data is reasonable.

(b) For people with the same gender, the odds of answering yes to question 1 is $e^{0.1493} = 1.16$ (significant at 0.05) times the odds of answering yes to question 3; the odds of answering yes to question 2 is $e^{0.052} = 1.05$ (marginally significant) times the odds of answering yes to question 3. Regardless of questions, females are a little more likely to say yes (very insignificant).

Problem 9.12

(a) The model with an interaction term is

$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta_1 trt + \beta_2 y_1 + \beta_3 trt \times y_1.$$

```
data prob9_12;
  input trt y1 n1-n4;
  cards;
  1 45 13 23  3  1
  1 75  9 17 13  8
  0 10  7  4  2  1
  0 25 14  5  1  0
  0 45  6  9 18  2
  0 75  4 11 14 22
  ;

 title "Recover individual data";
data prob9_12; set prob9_12;
  array temp {4} n1-n4;

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

  do k=1 to 4;
    do i=1 to temp(k);
      if k=1 then y2=10;
      else if k=2 then y2=25;
      else if k=3 then y2=45;
      else y2=75;
      id = id + 1;
      output;
    end;
  end;
run;
```

```
title "Problem 9.12 (a)";
proc logistic;
  model y2 = trt y1 trt*y1;
run;
```

```
*****************************************************************************
                 Score Test for the Proportional Odds Assumption

                       Chi-Square       DF     Pr > ChiSq

                           7.1886        6         0.3038
                    Analysis of Maximum Likelihood Estimates

                                                  Standard          Wald
     Parameter        DF     Estimate       Error    Chi-Square    Pr > ChiSq

     Intercept 10      1       1.1075      0.4164        7.0728        0.0078
     Intercept 25      1       2.8106      0.4488       39.2143        <.0001
     Intercept 45      1       4.3268      0.5019       74.3151        <.0001
     trt               1      -0.2172      0.5963        0.1327        0.7157
     y1                1      -0.0528     0.00793       44.3845        <.0001
     trt*y1            1       0.0217      0.0107        4.1129        0.0426
```

The log odds-ratio of having shorter time to fall asleep between the active drug and placebo is

$$\text{log-OR} = -0.2172 + 0.0217 y_1,$$

which is 0 when $y_1 = 10$ and 1.41 when $y_1 = 75$.

(b) The SAS program and part of output is

```
title "Problem 9.12 (b)";
proc logistic;
  class y1 / param=ref;
  model y2 = trt y1;
run;
```

```
*****************************************************************************
                    Analysis of Maximum Likelihood Estimates

                                                  Standard          Wald
     Parameter        DF     Estimate       Error    Chi-Square    Pr > ChiSq

     Intercept 10      1      -2.6123      0.2884       82.0357        <.0001
     Intercept 25      1      -0.8812      0.2353       14.0319        0.0002
     Intercept 45      1       0.5980      0.2376        6.3316        0.0119
     trt               1       0.9106      0.2473       13.5571        0.0002
     y1         10     1       2.3068      0.4433       27.0788        <.0001
     y1         25     1       2.6726      0.3989       44.8830        <.0001
     y1         45     1       1.1524      0.2892       15.8802        <.0001
```

The estimated log odds-ratio of having shorter time to fall asleep between the active drug and placebo is log-OR = 0.911, with SE=0.25. The odds-ratio is 2.45. That is, regardless of initial level, the odds of having shorter time to fall asleep for patients receiving the active drug is 2.45 time the odds of having shorter time to fall asleep for patients receiving the placebo.

(c) The SAS program and part of output is

```
title "Problem 9.12 (c)";
```

```
proc logistic;
  class y1 / param=ref;
  model y2 = trt y1 trt*y1;
run;
******************************************************************************
                    Analysis of Maximum Likelihood Estimates

                                          Standard          Wald
     Parameter        DF    Estimate        Error     Chi-Square    Pr > ChiSq

     Intercept 10      1     -2.8706       0.3291        76.0605        <.0001
     Intercept 25      1     -1.0888       0.2825        14.8579        0.0001
     Intercept 45      1      0.4404       0.2684         2.6914        0.1009
     trt               1      1.3059       0.3795        11.8416        0.0006
     y1        10      1      2.7247       0.5886        21.4295        <.0001
     y1        25      1      3.7452       0.5820        41.4125        <.0001
     y1        45      1      1.1672       0.4080         8.1852        0.0042
     trt*y1    10      1     -0.7775       0.8509         0.8349        0.3609
     trt*y1    25      1     -2.1510       0.7512         8.1986        0.0042
     trt*y1    45      1    -0.00441       0.5663         0.0001        0.9938
```

From the output, we see that the odds-ratios of having shorter time to fall asleep between patients receiving the active and placebo are

$$
\begin{aligned}
e^{1.3059-0.7775} &= 1.70 \ \text{ when } y_1 = 10 \\
e^{1.3059-2.1510} &= 0.43 \ \text{ when } y_1 = 25 \\
e^{1.3059-0.0044} &= 3.67 \ \text{ when } y_1 = 45 \\
e^{1.3059} &= 3.69 \ \text{ when } y_1 = 75
\end{aligned}
$$

The above estimates indicate that the active drug seems relatively more successful at the two highest initial response levels.

Problem 9.15

We model the obesity probability as a function of gender and time and their interaction using the following marginal logistic model:

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

where *male* is the dummy variable for male, *time* is time (in year) sice 1977.

```
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;
        output;
      end;
    end;
  end;
run;

proc genmod;
  class id;
  model y = male time male*time / dist=bin link=logit;
  repeated subject=id / type = exch corrw;
run;
************************************************************************
```

```
                          Working Correlation Matrix

                               Col1          Col2          Col3

                  Row1       1.0000        0.5041        0.5041
                  Row2       0.5041        1.0000        0.5041
                  Row3       0.5041        0.5041        1.0000


                       Analysis Of GEE Parameter Estimates
                        Empirical Standard Error Estimates

                            Standard   95% Confidence
         Parameter Estimate    Error       Limits            Z Pr > |Z|

         Intercept  -1.5826   0.1916  -1.9582  -1.2070   -8.26   <.0001
         male        0.4619   0.2567  -0.0412   0.9649    1.80   0.0720
         time        0.0741   0.1013  -0.1244   0.2726    0.73   0.4645
         male*time  -0.3118   0.1404  -0.5869  -0.0367   -2.22   0.0263
```

From the output, we have:

$$\text{logit}[\widehat{P}(obesity)] = -1.5826 + 0.4619male + 0.0741time - 0.3118male \times time.$$

So the logit of obesity probability over time for girls is:

$$\text{logit}[\widehat{P}(obeseity)] = -1.5826 + 0.0741time$$

and the logit of obesity probability over time for boys is:

$$\text{logit}[\widehat{P}(obeseity)] = -1.5826 + 0.4619 + (0.0741 - 0.3118)time = -1.1207 - 0.2377time.$$

The estimated obesity probabilities over time for boys are: 0.25, 0.20, 0.17. The estimated obesity probabilities over time for girls are: 0.17, 0.18, 0.19.

The GEE model indicates that the obesity probability patterns over time between boys are girls are very different (P-value=0.0263). Even though there is a slight increase in the obesity probability over time for girls, the increase is not significant (P-value = 0.4645). However, the obesity probability for boys decreases significantly over time, from 0.25 at 1977 to 0.17 at 1979. The boys had a much higher obesity probability at 1977 than girls. Two years later, the obesity probabilities between boys and girls are very similar.