Marginal Model for Categorical Data: Case Studies

A 2×2 Crossover Trial

- A **crossover trial** of efficacy of two treatments on cerebrovascular deficiency. Sixty-seven subjects from one center was used in the analysis for illustration.
- Two treatment arms (A: active drug, B: placebo) in the trial.
- Thirty-four patients received the active drug (A) followed by placebo (B); another 33 patients were treated in the reverse order.
- Binary outcome, 1: normal electrocardiogram reading, 0: abnormal reading.

Crossover design is one in which subjects are given a sequence of treatments with the objective of studying the difference between individual treatment.

- In crossover design, a subject can be considered as his/her own control to eliminate between subject variation, hence crossover design is more powerful than similar size parallel design.
- Period-by-treatment interactions may indicate **carry over effect**. A reasonable **wash out period** is needed.

```
>xover <- read.table ("../data/xover1.data", col.names = c("id", "class", "y", "intercept",
+                      "trt", "period", "xover", "BA"))
> xover$trtA <- 1-xover$trt
> xover$trtAP <- xover$trtA*xover$period
> xoverw <- reshape (xover[,c("id", "y","period","BA")],
+                   direction = "wide", v.names = "y", timevar = "period",idvar = "id")
> xoverw$respat <- ifelse(xoverw$y.0==0,2,3)
> xoverw$respat[(xoverw$y.0+xoverw$y.1)==2] <- 1
> xoverw$respat[(xoverw$y.0+xoverw$y.1)==0] <- 4

> #Table 8.1 in DHLZ book
> tab8.1 <- cbind(table(xoverw$BA,xoverw$respat),table(xoverw$BA),
+           table(xover$BA[xover$y==1],xover$period[xover$y==1]))
> dimnames(tab8.1) <- list(c("AB","BA"),c("(1,1)","(0,1)","(1,0)","(0,0)",
+                      "Total","Period 1 Effective","Period 2 Effective"))
> tab8.1
   (1,1) (0,1) (1,0) (0,0) Total Period 1 Effective Period 2 Effective
AB    22     0     6     6    34                 28                 22
BA    18     4     2     9    33                 20                 22
```

- What is the treatment effect if only period 1 data are considered?

```
> # Log odds ratio comparing the chance of being normal for drug vs. placebo using period 1 data
> log((28*13)/(20*6))
[1] 1.109662

> # Standard error for the estimated log-odds ratio
> sqrt(1/28+1/6+1/20+1/13)
[1] 0.5738502
> # Z
> 1.109662/0.5738502
[1] 1.933714
```

  The estimated treatment effect based on the first period's data is not significant. Combining the data from both periods could improve efficiency.

- Any within-subject correlation?
  - For patients treated with AB
  - For patients treated with BA
- Any carry-over effect (treatment-by-period interaction)?

Fit a GEE marginal model for the 2×2 Crossover Trial.

```
> library(gee)
> summary (gee (y ~ trtA+period+trtAP, data = xover, cor = "exchangeable",
+               id = id, family = binomial, scale.fix = TRUE))
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
(Intercept)        trtA      period       trtAP
  0.4307829   1.1096621   0.1753529  -1.0226507

 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Exchangeable

Call:
gee(formula = y ~ trtA + period + trtAP, id = id, data = xover,
    family = binomial, corstr = "exchangeable", scale.fix = TRUE)

Summary of Residuals:
       Min         1Q     Median         3Q        Max
-0.8235294 -0.6060606  0.1764706  0.3529412  0.3939394
```

```
Coefficients:
              Estimate Naive S.E.    Naive z Robust S.E.   Robust z
(Intercept)  0.4307829  0.3562627  1.2091723   0.3562627  1.2091723
trtA         1.1096621  0.5738502  1.9337140   0.5738502  1.9337140
period       0.1753529  0.5056787  0.3467674   0.5056787  0.3467674
trtAP       -1.0226507  0.9846776 -1.0385641   0.9789663 -1.0446231

Estimated Scale Parameter:  1
Number of Iterations:  1

Working Correlation
          [,1]      [,2]
[1,] 1.0000000 0.6401548
[2,] 0.6401548 1.0000000

> # Drop the interaction term
> summary (gee (y ~ trtA+period, data = xover, cor = "exchangeable",
+                id = id, family = binomial, scale.fix = TRUE))
...
Coefficients:
              Estimate Naive S.E.   Naive z Robust S.E.  Robust z
(Intercept)  0.6660113  0.2842188  2.343304   0.2878956  2.313378
trtA         0.5690305  0.2287782  2.487258   0.2327207  2.445123
period      -0.2953176  0.2271589 -1.300048   0.2311211 -1.277761
```

Fit a GEE marginal model with working independent correlation.

```
> summary (gee (y ~ trtA+period, data = xover, cor = "independence",
+               id = id, family = binomial, scale.fix = TRUE))
...
Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Independent
...
Coefficients:
              Estimate Naive S.E.    Naive z Robust S.E.  Robust z
(Intercept)  0.6603929  0.3212796  2.0555081   0.2874920  2.297083
trtA         0.5581594  0.3784107  1.4750094   0.2332530  2.392936
period      -0.2743154  0.3768179 -0.7279787   0.2322731 -1.181004

Working Correlation
     [,1] [,2]
[1,]    1    0
[2,]    0    1
```

Fit a GEE marginal model with odds ratio and alternating logistic regression.

The `alr` package of R implements the ALR algorithm (`http://www.biostat.harvard.edu/~carey/vcwww_4.html`).

```
> library(alr)
> # Model 1 in Table 8.2 of DHLZ
> x1 <-  as.matrix (xover[,c("trtA", "period","trtAP")])
> xover.alr1 <- alr (xover$y ~ x1, id = xover$id,
+                    ainit = 0.01, depmodel = "exchangeable")
[1] "alternating logistic regression - Splus, @(#) alr.q 4.4 98/02/24"
[1] "Running glm to get initial estimates"
[1]  0.4307829  1.1096621  0.1753529 -1.0226507
[1] "nobs"
[1] 134
> summary (xover.alr1)

ALR:  ALTERNATING LOGISTIC REGRESSION
alr S-function, version 4.4 98/02/24

Call:
alr(formula = xover$y ~ x1, id = xover$id, ainit = 0.01, depmodel = "exchangeable")

Summary of Residuals:
       Min         1Q     Median          3Q         Max
-0.8235294 -0.6060606  0.1764706  0.3529412  0.3939394
```

```
Coefficients:
              Estimate Robust S.E.   Robust z
(Intercept)  0.4307829   0.3562627  1.2091723
x1trtA       1.1096621   0.5738502  1.9337140
x1period     0.1753529   0.5056787  0.3467674
x1trtAP     -1.0226507   0.9789663 -1.0446231


Alpha:
   Estimate Robust S.E. Robust z
a1 3.537803   0.8200298 4.314238


Number of observations :  134
Number of Iterations   :  5

> # Odds ratio between the two periods
> exp(3.54)
[1] 34.46692
```

Drop the interaction term from the above ALR model.

```
> # Model 2 in Table 8.2 of DHLZ
> x2 <-  as.matrix (xover[,c("trtA", "period")])
> xover.alr2 <- alr (xover$y ~ x2, id = xover$id,
+                    ainit = 0.01, depmodel = "exchangeable")

> summary (xover.alr2)
ALR:  ALTERNATING LOGISTIC REGRESSION
alr S-function, version 4.4 98/02/24

Call:
alr(formula = xover$y ~ x2, id = xover$id, ainit = 0.01, depmodel = "exchangeable")

Summary of Residuals:
       Min         1Q     Median         3Q        Max
-0.7761458 -0.5937025  0.2238542  0.3375072  0.4062975

Coefficients:
              Estimate Robust S.E.  Robust z
(Intercept)  0.6744228   0.2882568  2.339659
x2trtA       0.5689228   0.2335157  2.436336
x2period    -0.2951299   0.2318499 -1.272935

Alpha:
   Estimate Robust S.E. Robust z
a1 3.561692   0.8147993  4.37125
```

A 3×3 Crossover Trial

- Three-period crossover trial of an analgesic drug for pain relieving.
- Three levels of analgesic (placebo, low, and high) were given to each of the 86 women.
- Women were randomized to one of the six possible orders for administering the three treatment levels.
- Ignoring the order of treatment, pain was relieved for 26% with placebo, 73% with low dose, and 78% with high dose
- A cross-over study, the treatment changes each time.
- Binary outcome, 1: relief, 0: no relief.

```
> xover3 <- read.table ("../data/xover3new.txt",col.names = c("id", "class",
+              "relief","intercept", "tx2", "tx3", "p2", "p3","ptx1", "ptx2", "ptx3"))

> xover3$period <- ifelse (xover3$p2 == 1, 2,ifelse (xover3$p3 == 1, 3, 1))
> xover3$tr <- ifelse (xover3$tx2 == 1, 2,ifelse (xover3$tx3 == 1, 3, 1))
>
> xover3 <- xover3[order(xover3$id,xover3$period),]
> xover3w <- reshape (xover3[,c("id", "relief","period","tr")],
+                  direction = "wide", timevar = "period",idvar = "id")
> xover3w$grp[xover3w$tr.1==1] <- ifelse(xover3w$tr.2[xover3w$tr.1==1] == 2, 1, 2)
> xover3w$grp[xover3w$tr.1==2] <- ifelse(xover3w$tr.2[xover3w$tr.1==2] == 1, 3, 4)
> xover3w$grp[xover3w$tr.1==3] <- ifelse(xover3w$tr.2[xover3w$tr.1==3] == 1, 5, 6)
>
> xover3w$pat <- 1
> xover3w$pat[apply(xover3w[,c(2,4,6)],1,sum)==3] <- 8
> xover3w$pat[apply(xover3w[,c(2,4,6)],1,sum)==1] <- 2
> ind34 <- xover3w$pat==2&xover3w$relief.1==0
> xover3w$pat[ind34] <- ifelse(xover3w$relief.2[ind34] == 1, 3, 4)
> xover3w$pat[apply(xover3w[,c(2,4,6)],1,sum)==2] <- 7
> ind56 <- xover3w$pat==7&xover3w$relief.1==1
> xover3w$pat[ind56] <- ifelse(xover3w$relief.2[ind56] == 1, 5, 6)
>
> #Table 8.3 in DHLZ book
> tab8.3 <- cbind(table(xover3w$grp,xover3w$pat),table(xover3w$grp))
> dimnames(tab8.3) <- list(c("ABC","ACB","BAC","BCA","CAB","CBA"),
+     c("000","100","010","001","110","101","011","111","Total"))
```

```
> tab8.3
    000 100 010 001 110 101 011 111 Total
ABC   0   0   2   2   1   0   9   1    15
ACB   2   1   0   0   0   0   9   4    16
BAC   0   1   1   1   0   8   3   1    15
BCA   0   1   1   1   8   0   0   1    12
CAB   3   0   0   0   1   7   2   1    14
CBA   1   5   0   0   4   3   1   0    14
> # Look at all periods, ignorning within-subject correlation
> xover3tab <- table(xover3$relief,xover3$tr)
> xover3tab[2,]/apply(xover3tab,2,sum)
        1         2         3
0.2558140 0.7093023 0.8023256
> # Look at the data stratified by period
> with (xover3, ftable (period, relief, tr))
              tr  1  2  3
period relief
1      0         24  7  7
       1          7 20 21
2      0         20 11  5
       1          9 18 23
3      0         20  7  5
       1          6 23 25
```

- Just from the above tables, it appears that the two different doses of the analgesic treatment (B and C) both did better than the placebo (A). The high dose may be slightly better than the low dose.

- Is there any carry-over effect?

```
> xover3$ptx <- ifelse (xover3$ptx1 == 1, 1, ifelse (xover3$ptx2 == 1, 2, 3))
> xover3$ptx[xover3$period == 1] <- 0
> with (xover3, ftable (ptx, relief, tr))
           tr  1  2  3
ptx relief
0   0         24  7  7
    1          7 20 21
1   0          0  6  5
    1          0 23 26
2   0         20  0  5
    1          9  0 22
3   0         20 12  0
    1          6 18  0

> round(matrix(c(20*24/(7*7),21*7/(7*20),24*21/(7*7),
+                 18*20/(6*12),26*6/(23*5),22*20/(9*5)), nrow=2,byrow=T),2)
     [,1] [,2]  [,3]
[1,]  9.8 1.05 10.29
[2,]  5.0 1.36  9.78
```

We fit four models, assuming working independence (model 1) or unstructured correlation (model 2) and using odds ratios to characterize the associations (model 3, 4).

Model 1: Working independence correlation.

```
> xover.gee <- gee (relief ~ p2 + p3 + tx2 + tx3 + ptx2 + ptx3,
+                    data = xover3, scale.fix = TRUE, id = id,
+                    family = binomial)
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
(Intercept)          p2          p3         tx2         tx3        ptx2        ptx3
 -1.0866229   0.4141734   0.5885481   1.9493789   2.2222895  -0.1922158  -0.8308649
> summary (xover.gee)

 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
 Link:                         Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Independent

Call:
gee(formula = relief ~ p2 + p3 + tx2 + tx3 + ptx2 + ptx3, id = id,
    data = xover3, family = binomial, scale.fix = TRUE)
```

```
Summary of Residuals:
       Min         1Q     Median         3Q        Max
-0.8486709 -0.2522547  0.1751094  0.2431169  0.8180683


Coefficients:
              Estimate Naive S.E.    Naive z Robust S.E.   Robust z
(Intercept) -1.0866229  0.3280932 -3.3119335   0.3171169 -3.4265694
p2           0.4141734  0.4609959  0.8984318   0.4188751  0.9887755
p3           0.5885481  0.4752612  1.2383678   0.4556855  1.2915665
tx2          1.9493789  0.3888244  5.0135193   0.4139646  4.7090470
tx3          2.2222895  0.3945988  5.6317689   0.4203606  5.2866265
ptx2        -0.1922158  0.5070045 -0.3791205   0.5120981 -0.3753495
ptx3        -0.8308649  0.4818407 -1.7243562   0.4199874 -1.9783093

Estimated Scale Parameter:  1
Number of Iterations:  1

Working Correlation
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

Model 2: Working unstructured correlation.

```
> xover.gee <- gee (relief ~ p2 + p3 + tx2 + tx3 + ptx2 + ptx3,
+                   data = xover3, scale.fix = TRUE,
+                   family = binomial, corstr = "unstructured")
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
(Intercept)          p2          p3         tx2         tx3        ptx2        ptx3
 -1.0866229   0.4141734   0.5885481   1.9493789   2.2222895  -0.1922158  -0.8308649
> summary (xover.gee)

 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Unstructured

Call:
gee(formula = relief ~ p2 + p3 + tx2 + tx3 + ptx2 + ptx3, data = xover3,
    family = binomial, corstr = "unstructured", scale.fix = TRUE)

Summary of Residuals:
       Min         1Q     Median         3Q        Max
-0.8453150 -0.2494076  0.1717796  0.2401164  0.8222795
```

```
Coefficients:
              Estimate Naive S.E.    Naive z Robust S.E.   Robust z
(Intercept) -1.1017745  0.3259071 -3.3806403   0.3215724 -3.4262094
p2           0.3759339  0.4753671  0.7908287   0.4167114  0.9021444
p3           0.5462774  0.4608093  1.1854738   0.4504195  1.2128191
tx2          1.9884546  0.3801905  5.2301535   0.4163466  4.7759601
tx3          2.2538159  0.3878265  5.8114026   0.4252588  5.2998688
ptx2        -0.1252514  0.4811531 -0.2603150   0.5122957 -0.2444904
ptx3        -0.8060278  0.4598286 -1.7528876   0.4167318 -1.9341644

Estimated Scale Parameter:  1
Number of Iterations:  3

Working Correlation
            [,1]        [,2]       [,3]
[1,]  1.00000000 -0.17040570 0.03317488
[2,] -0.17040570  1.00000000 0.03673587
[3,]  0.03317488  0.03673587 1.00000000
```

- The results are similar. The two doses of treatments seemed to be highly effective and the difference between them is small (odds ratios ~ 8).
- There is a slight evidence for the carry-over effect after the high dose analgesic (OR is $exp(-0.81) = 0.4$).
- DHLZ Example 8.2 used different models, i.e. ALR, (models 1 and 2 in Table 8.4, p152) with similar conclusions.

Model 3: ALR with exchangeable correlation for the association (model 1 in Table 8.4 of DHLZ).

```
> library (alr)
> x <-  as.matrix (xover3[,c("p2", "p3", "tx2",
+                             "tx3", "ptx2", "ptx3")])
> y <- xover3$relief
> xover.alr <- alr (y ~ x, id = xover3$id,
+                      ainit = 0.01, depmodel = "exchangeable")
[1] "alternating logistic regression - Splus, @(#) alr.q 4.4 98/02/24"
[1] "Running glm to get initial estimates"
[1] -1.0866229  0.4141734  0.5885481  1.9493789  2.2222895 -0.1922158 -0.8308649
[1] "nobs"
[1] 258
> summary (xover.alr)

ALR:  ALTERNATING LOGISTIC REGRESSION
alr S-function, version 4.4 98/02/24

Call:
alr(formula = y ~ x, id = xover3$id, ainit = 0.01, depmodel = "exchangeable")

Summary of Residuals:
       Min         1Q     Median         3Q        Max
-0.8493048 -0.2529760  0.1745086  0.2431256  0.8169703
```

```
Coefficients:
              Estimate Robust S.E.   Robust z
(Intercept) -1.0828026   0.3166356 -3.4197123
xp2          0.4183858   0.4206120  0.9947072
xp3          0.5935401   0.4582208  1.2953146
xtx2         1.9444196   0.4135581  4.7016844
xtx3         2.2184216   0.4190099  5.2944374
xptx2       -0.2095632   0.5156085 -0.4064387
xptx3       -0.8315376   0.4196590 -1.9814602


Alpha:
     Estimate Robust S.E.   Robust z
a1 -0.2235135   0.3803329 -0.5876787


Number of observations :  258
Number of Iterations   :  3
```

Model 4: ALR with unstructured correlation for the association (Model 2, Table 8.4).

For an unstructured correlation matrix, there are $n_i(n_i - 1)/2$ parameters. For the balanced design here we have $n_i = 3$ and 3 odds ratios to estimate. We need first make a design matrix with $n_i(n_i - 1) \times q$ (6×3) elements where q is the number of parameters to be estimated.

```
> correctz <- matrix (c(1, 0, 0,
+                       0, 1, 0,
+                       1, 0, 0,
+                       0, 0, 1,
+                       0, 1, 0,
+                       0, 0, 1),
+             ncol = 3, byrow = TRUE)
> xover.alr <- alr (y ~ x, id = xover3$id,
+                   z = correctz,
+                   zmast = 1,
+                   zlocs = rep (1:3, 86),
+                   ainit = rep (0.01, 3),
+                   depmodel = "general")
```

- The rows correspond to pairs (1,2), (1,3), (2,1), (2,3), (3,1), (3,2).
- Pairs of (1,2) and (2,1) share the same parameter, so they have the same values in their rows. So are (1,3) and (3,1); (2,3) and (3,2).

```
> summary (xover.alr)
...
Summary of Residuals:
       Min         1Q      Median          3Q         Max
-0.8445614 -0.2503970  0.1722254  0.2416136  0.8218775

Coefficients:
              Estimate Robust S.E.   Robust z
(Intercept) -1.0964962   0.3203766 -3.4225233
xp2          0.3696734   0.4138135  0.8933333
xp3          0.5487134   0.4480590  1.2246452
xtx2         1.9818621   0.4145504  4.7807504
xtx3         2.2403497   0.4240615  5.2830774
xptx2       -0.1226300   0.5078993 -0.2414455
xptx3       -0.8022969   0.4159715 -1.9287305

Alpha:
     Estimate Robust S.E.   Robust z
a1 -0.9643585   0.6067443 -1.5893986
a2  0.1366401   0.7075102  0.1931282
a3  0.2622779   0.6956535  0.3770237
```

ALR in SAS

The implementation of ALR is perhaps more mature in SAS:

```
data xover3;
    infile '../xover3new.txt';
    input id class relief inter tx2 tx3 p2 p3 ptx1 ptx2 ptx3;
run;
```

ALR with exchangeable correlation for the association.

```
proc genmod data = xover3 descending;
    class id;
    model relief = p2 p3 tx2 tx3 ptx2 ptx3 / dist = bin;
    repeated subject = id / logor = exch modelse;
run;
```

```
                         GEE Model Information

            Log Odds Ratio Structure          Exchangeable
            Subject Effect                  id (86 levels)
            Number of Clusters                          86
            Correlation Matrix Dimension                 3
            Maximum Cluster Size                         3
            Minimum Cluster Size                         3


    Algorithm converged.
```

GEE Fit Criteria

QIC	297.1377
QICu	297.1526

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	-1.0828	0.3166	-1.7034	-0.4622	-3.42	0.0006
p2	0.4184	0.4206	-0.4060	1.2428	0.99	0.3199
p3	0.5935	0.4582	-0.3046	1.4916	1.30	0.1952
tx2	1.9444	0.4136	1.1339	2.7550	4.70	<.0001
tx3	2.2184	0.4190	1.3972	3.0397	5.29	<.0001
ptx2	-0.2096	0.5156	-1.2201	0.8010	-0.41	0.6844
ptx3	-0.8315	0.4197	-1.6541	-0.0090	-1.98	0.0475
Alpha1	-0.2235	0.3803	-0.9690	0.5219	-0.59	0.5567

ALR with unstructured correlation for the association.

```
proc genmod data = xover3 descending;
    class id;
    model relief = p2 p3 tx2 tx3 ptx2 ptx3 / dist = bin;
    repeated subject = id / logor = fullclust modelse;
run;
```

```
                              GEE Model Information

                  Log Odds Ratio Structure        Fully Parameterized Clusters
                  Subject Effect                                id (86 levels)
                  Number of Clusters                                        86
                  Correlation Matrix Dimension                               3
                  Maximum Cluster Size                                       3
                  Minimum Cluster Size                                       3

                                  Log Odds Ratio
                              Parameter Information

                              Parameter       Group

                              Alpha1          (1, 2)
                              Alpha2          (1, 3)
                              Alpha3          (2, 3)
           Algorithm converged.
                                 GEE Fit Criteria
```

```
                QIC             297.0066
                QICu            297.1717

         Analysis Of GEE Parameter Estimates
          Empirical Standard Error Estimates

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

Intercept  -1.0965    0.3204  -1.7244  -0.4686   -3.42   0.0006
p2          0.3697    0.4138  -0.4414   1.1807    0.89   0.3717
p3          0.5487    0.4481  -0.3295   1.4269    1.22   0.2207
tx2         1.9819    0.4146   1.1693   2.7944    4.78   <.0001
tx3         2.2404    0.4241   1.4092   3.0715    5.28   <.0001
ptx2       -0.1226    0.5079  -1.1181   0.8729   -0.24   0.8092
ptx3       -0.8023    0.4160  -1.6176   0.0130   -1.93   0.0538
Alpha1     -0.9644    0.6067  -2.1536   0.2248   -1.59   0.1120
Alpha2      0.1366    0.7075  -1.2501   1.5233    0.19   0.8469
Alpha3      0.2623    0.6957  -1.1012   1.6257    0.38   0.7062
```

- Other possible choices for `logor` are: `logorvar` (allows the log OR to depend on another variable, e.g., center); `ZFULL` (fully specified z-matrix specified in ZDATA= data set) etc.

Seizure Data

- Clinical trial of 59 epileptics.
- For each patients, the number of epileptic seizures was recorded during a baseline period of eight weeks.
- Patients were randomized to be treated with the anti-epileptic drug progabide or placebo.
- Number of seizures was then recorded in four consecutive two-week intervals.
- Does progabide treatment reduce the rate of epileptic seizures?
- Count data.

A little data manipulation such that each person will have an additional row of baseline (or pre-treatment) data.

```
> seize <- read.table ("../data/seize.data",
+                      col.names = c("id", "seizure", "week",
+                      "progabide", "baseline8", "age"))
> seize.base <- data.frame (id = seize$id, seizure = seize$baseline8, week = seize$week,
+                           progabide = seize$progabide, age = seize$age)
> seize.base <- seize.base[seize.base$week == 1,]
> seize.base$week <- 0
> seize.full <- rbind (seize[,-5], seize.base)
> seize.full <- seize.full[order(seize.full$id, seize.full$week),]
> seize.full$time <- ifelse (seize.full$week == 0, 8, 2)
> seize.full$post <- as.numeric (seize.full$week != 0)
> seize.full[1:10,]
      id seizure week progabide age time post
1131 101      76    0         1  18    8    0
113  101      11    1         1  18    2    1
114  101      14    2         1  18    2    1
115  101       9    3         1  18    2    1
116  101       8    4         1  18    2    1
1171 102      38    0         1  32    8    0
117  102       8    1         1  32    2    1
118  102       7    2         1  32    2    1
119  102       9    3         1  32    2    1
120  102       4    4         1  32    2    1
```

Fit a Poisson family GEE model with exchangeable correlation (Table 8.11, text book).

- Offset (a term with constant coefficient) is used to take into account that the time periods are different (8 weeks vs 2 weeks).

$$\log(\mu_{ij}) = \log(t_{ij}\lambda_{ij}) = \log(t_{ij}) + \log \lambda_{ij} = \log(t_{ij}) + \boldsymbol{X}_{ij}^T\boldsymbol{\beta}.$$

```
> sg1 <- gee (seizure ~ post + progabide + post:progabide +
+              offset (log(time)),
+              data = seize.full, id = id, family = "poisson",
+              cor = "exchangeable")
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
   (Intercept)          post      progabide post:progabide
    1.34760922    0.11079814     0.02651461    -0.10368067
> summary (sg1)

 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
 Link:                      Logarithm
 Variance to Mean Relation: Poisson
 Correlation Structure:     Exchangeable

Call:
gee(formula = seizure ~ post + progabide + post:progabide + offset(log(time)),
```

```
    id = id, data = seize.full, family = "poisson", corstr = "exchangeable")

Summary of Residuals:
       Min         1Q     Median         3Q        Max
 -4.299107  -1.299107   2.020161  10.374640 147.048387

Coefficients:
                  Estimate Naive S.E.    Naive z Robust S.E.   Robust z
(Intercept)     1.34760922  0.1511851  8.9136359   0.1573571  8.5640166
post            0.11079814  0.1547038  0.7161956   0.1160997  0.9543358
progabide       0.02651461  0.2072721  0.1279217   0.2218539  0.1195138
post:progabide -0.10368067  0.2199500 -0.4713830   0.2136100 -0.4853736

Estimated Scale Parameter:  19.70269
Number of Iterations:  1

Working Correlation
         [,1]     [,2]     [,3]     [,4]     [,5]
[1,] 1.000000 0.771588 0.771588 0.771588 0.771588
[2,] 0.771588 1.000000 0.771588 0.771588 0.771588
[3,] 0.771588 0.771588 1.000000 0.771588 0.771588
[4,] 0.771588 0.771588 0.771588 1.000000 0.771588
[5,] 0.771588 0.771588 0.771588 0.771588 1.000000
```

A negative coefficient for post:progabide suggests a greater reduction in number of seizures for the treatment group compared with the control group.

Illustration of *over-dispersed* seizure count data.

```
> seizew <- reshape(seize[, c("id", "seizure", "week", "progabide")],
+                   direction = "wide", v.names = "seizure", timevar = "week",idvar = "id")
>
> Treated <- apply(seizew[seizew[,"progabide"]==1,3:6],2,var)/apply(
+                  seizew[seizew[,"progabide"]==1,3:6],2,mean)
>
> Placebo <- apply(seizew[seizew[,"progabide"]==0,3:6],2,var)/apply(
+                  seizew[seizew[,"progabide"]==0,3:6],2,mean)
>
> #Table 8.8 in DHLZ book, variance-to-mean ratios
>
> round(rbind(Treated,Placebo),2)
        seizure.1 seizure.2 seizure.3 seizure.4
Treated     38.78     16.71     23.75     18.91
Placebo     10.98      8.04     24.50      7.31
```

We can also allow the dispersion parameter to differ between the treatment and placebo group by fitting Yan and Fine (2004)'s model.

```
> library(geepack)
> sg2 <- geese (seizure ~ post + progabide + post:progabide +
+                  offset (log(time)), sformula = ~ progabide,
+                  data = seize.full, id = id, family = "poisson",
+                  corstr = "exchangeable")
> summary (sg2)

Call:
geese(formula = seizure ~ post + progabide + post:progabide +
    offset(log(time)), sformula = ~progabide, id = id, data = seize.full,
    family = "poisson", corstr = "exchangeable")

Mean Model:
 Mean Link:                 log
 Variance to Mean Relation: poisson

 Coefficients:
                  estimate    san.se        wald            p
(Intercept)     1.34760922 0.1620151 69.18579843 1.110223e-16
post            0.11079814 0.1203986  0.84688095 3.574362e-01
progabide       0.02651461 0.2251817  0.01386450 9.062676e-01
post:progabide -0.10368067 0.2159766  0.23045286 6.311882e-01
```

```
Scale Model:
 Scale Link:                identity

 Estimated Scale Parameters:
             estimate    san.se     wald           p
(Intercept) 14.228486  4.592945 9.596986 0.001948970
progabide    9.910183 16.657605 0.353947 0.551887297

Correlation Model:
 Correlation Structure:     exchangeable
 Correlation Link:          identity

 Estimated Correlation Parameters:
       estimate     san.se     wald p
alpha 0.7451173 0.08256208 81.44944 0

Returned Error Value:    0
Number of clusters:   59   Maximum cluster size: 5
```

For illustration, we assume an exchangeable correlation structure where the correlation may depend on age.

$$\mathrm{Cor}(Y_{ij}, Y_{ik}) = \rho_i$$
$$\log\left(\frac{1+\rho_i}{1-\rho_i}\right) = \alpha_1 + \mathrm{Age}_i\alpha_2$$

This model can be fitted using **geese**. The design matrix for the correlation model has to be constructed by hand. The matrix Z has the same number of rows as the number of clusters (the covariate should be invariant within a cluster).

```
> z <- cbind (1, seize.full$age[seize.full$week == 0])
> sg2 <- geese (seizure ~ progabide + post + post:progabide +
+               offset (log(time)), sformula = ~ progabide,
+               data = seize.full, id = id, family = "poisson",
+               cor.link = "fisherz", zcor = z, corstr = "exchangeable")
> summary (sg2)

Call:
geese(formula = seizure ~ progabide + post + post:progabide +
    offset(log(time)), sformula = ~progabide, id = id, data = seize.full,
    zcor = z, family = "poisson", cor.link = "fisherz", corstr = "exchangeable")

Mean Model:
 Mean Link:                 log
 Variance to Mean Relation: poisson
```

```
 Coefficients:
                   estimate    san.se         wald            p
(Intercept)     1.413723297 0.1697429 6.936582e+01 1.110223e-16
progabide       0.003521332 0.2227135 2.499890e-04 9.873851e-01
post            0.119463923 0.1448016 6.806551e-01 4.093612e-01
progabide:post -0.321152168 0.2880464 1.243073e+00 2.648795e-01

Scale Model:
 Scale Link:                identity

 Estimated Scale Parameters:
            estimate    san.se       wald           p
(Intercept) 13.30317  4.188812 10.0862304 0.001493814
progabide   14.30612 19.040456  0.5645324 0.452439804

Correlation Model:
 Correlation Structure:     exchangeable
 Correlation Link:          fisherz

 Estimated Correlation Parameters:
          estimate     san.se     wald          p
alpha:1  5.8652470 2.80875050 4.360600 0.03677934
alpha:2 -0.1370331 0.08139154 2.834606 0.09225342

Returned Error Value:    0
Number of clusters:   59   Maximum cluster size: 5
```

Further Reading

- Chapter 8 of DHLZ.