BIOS6643. L12 GEE and Marginal Models

Fitting population-averaged models using the gee() function and the geeglm() function in geepack. We will need to install.packages("gee") and install.packages("geepack").

The syntax for each is similar. A general call looks like

fit.object <- gee(model formula, id, corstr, family, data)

- **corstr** is a specification for the working correlation matrix
- **family** is a specification for the scaled exponential family that would be relevant under independence; the canonical link is the default

Dental data

Example of continuous outcome

```
##  Read in the data
dat.den <- read.csv("/Users/juarezce/Documents/OneDrive - The University of Colorado Denver/BIOS6643/BI
head(dat.den,3)
```

```
##   id age distance gender
## 1  1   8     21.0  Girls
## 2  1  10     20.0  Girls
## 3  1  12     21.5  Girls
```

```
##dat.den$id <- as.factor(dat.den$id)
##dat.den$gender <- as.factor(dat.den$gender)

ind.gee.den <- gee(distance ~ age*gender,
             id = id,
             data = dat.den,
             family = gaussian,
             corstr = "independence"
             )
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
```

```
## running glm to get initial regression estimate
```

```
##     (Intercept)             age     genderGirls age:genderGirls
##      16.3406250       0.7843750       1.0321023      -0.3048295
```

```
summary(ind.gee.den)
```

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

```
##  Link:                       Identity
##  Variance to Mean Relation: Gaussian
##  Correlation Structure:     Independent
##
## Call:
## gee(formula = distance ~ age * gender, id = id, data = dat.den,
##     family = gaussian, corstr = "independence")
##
## Summary of Residuals:
##        Min         1Q     Median         3Q        Max
## -5.6156250 -1.3218750 -0.1681818  1.3299006  5.2468750
##
##
## Coefficients:
##                   Estimate Naive S.E.   Naive z Robust S.E.   Robust z
## (Intercept)     16.3406250  1.4162242 11.538163  1.17148092 13.9486906
## age              0.7843750  0.1261673  6.216945  0.09834755  7.9755416
## genderGirls      1.0321023  2.2187969  0.465163  1.37778506  0.7491025
## age:genderGirls -0.3048295  0.1976661 -1.542143  0.11686730 -2.6083390
##
## Estimated Scale Parameter:  5.093818
## Number of Iterations:  1
##
## Working Correlation
##      [,1] [,2] [,3] [,4]
## [1,]    1    0    0    0
## [2,]    0    1    0    0
## [3,]    0    0    1    0
## [4,]    0    0    0    1
```

Recall the interpretation of the parameter estimates is at the population level. For instance for boys, the $\hat{\beta}_{age} = 0.78$ may be interpreted as the average (expected) increase in distance per year of age increased.

Seizure data using gee()

Epileptic Seizure Study of a randomized trial reported in Thall and Vail (1990).

- 59 subjects with epilepsy suffering from simple or partial seizures were assigned at random to receive either the anti-epileptic drug progabide or a placebo in addition to a standard chemotherapy regimen all were taking.
- Because each individual might be prone to different rates of experiencing seizures, the investigators first tried to get a sense of this by recording the number of seizures suffered by each subject over the 8-week period prior to the start of administration of the assigned treatment.
  - It is common in such studies to record such baseline measurements, so that the effect of treatment for each subject can be measured relative to how that subject behaved prior to treatment.
- Following initiation of treatment, the number of seizures for each subject was counted for each of 4 consecutive 2-week periods.
- The age of each subject at the start of the study was also recorded, as it was suspected that subject age might be associated with the effect of the treatment.

```r
#  Read in the data

dat.sz <- read.table("/Users/juarezce/Documents/OneDrive - The University of Colorado Denver/BIOS6643/B
colnames(dat.sz) <- c("subj","seize","visit","trt","base","age")
## trt=0 corresponds to placebo
head(dat.sz,3)
```

```
##   subj seize visit trt base age
## 1  104    11     0   0   11  31
## 2  104     5     1   0   11  31
## 3  104     3     2   0   11  31
```

```r
#  Create other covariates
dat.sz$o <- 8*(dat.sz$visit==0)+2*(dat.sz$visit>0)
dat.sz$logo <- log(dat.sz$o)
dat.sz$vm0 <- as.numeric(dat.sz$visit>0)

#  Basic models -- the unstructured fit using gee() does not converge
#  even with maxiter set to be much larger than the default

##un.gee <- gee(seize ~ trt + offset(logo), id=subj, family=poisson,
##              corstr="unstructured", data=dat.sz, maxiter=100)

cs.gee <- gee(seize ~ trt + offset(logo),id=subj,family=poisson,
              corstr="exchangeable", data=dat.sz)
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
```

```
## running glm to get initial regression estimate
```

```
## (Intercept)         trt
##  1.40454203 -0.02685314
```

```r
summary(cs.gee)
```

```
##
##  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 = seize ~ trt + offset(logo), id = subj, data = dat.sz,
##     family = poisson, corstr = "exchangeable")
##
## Summary of Residuals:
##         Min          1Q      Median          3Q         Max
##  -3.9392609  -0.9392609   2.3491032  10.7049211 147.0607391
##
##
## Coefficients:
##               Estimate Naive S.E.   Naive z Robust S.E.  Robust z
## (Intercept) 1.29497285  0.1392018 9.3028451   0.1718363 7.5360855
```

```
## trt         0.07602026  0.1886740 0.4029185   0.2139380 0.3553378
##
## Estimated Scale Parameter:  20.73326
## Number of Iterations:  3
##
## Working Correlation
##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] 1.0000000 0.7692888 0.7692888 0.7692888 0.7692888
## [2,] 0.7692888 1.0000000 0.7692888 0.7692888 0.7692888
## [3,] 0.7692888 0.7692888 1.0000000 0.7692888 0.7692888
## [4,] 0.7692888 0.7692888 0.7692888 1.0000000 0.7692888
## [5,] 0.7692888 0.7692888 0.7692888 0.7692888 1.0000000
```

```
## AR1
ar1.gee <- gee(seize ~ trt + offset(logo),id=subj,family=poisson,
               corstr="AR-M",Mv=1,data=dat.sz)
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
## running glm to get initial regression estimate
```

```
## (Intercept)         trt
##  1.40454203 -0.02685314
```

```
 summary(ar1.gee)
```

```
##
##  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:     AR-M , M = 1
##
## Call:
## gee(formula = seize ~ trt + offset(logo), id = subj, data = dat.sz,
##     family = poisson, corstr = "AR-M", Mv = 1)
##
## Summary of Residuals:
##         Min          1Q      Median          3Q         Max
##  -3.7578756  -0.7578756   2.4656263  10.8538753 147.2421244
##
##
## Coefficients:
##               Estimate Naive S.E.  Naive z Robust S.E.  Robust z
## (Intercept) 1.26253613  0.1420223 8.889702   0.1750352 7.2130401
## trt         0.06131768  0.1931454 0.317469   0.2118369 0.2894571
##
## Estimated Scale Parameter:  21.68155
## Number of Iterations:  3
##
## Working Correlation
##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] 1.0000000 0.8098380 0.6558376 0.5311222 0.4301229
## [2,] 0.8098380 1.0000000 0.8098380 0.6558376 0.5311222
## [3,] 0.6558376 0.8098380 1.0000000 0.8098380 0.6558376
```

```
## [4,] 0.5311222 0.6558376 0.8098380 1.0000000 0.8098380
## [5,] 0.4301229 0.5311222 0.6558376 0.8098380 1.0000000
```

Seizure data using geeglm()

```
## UN using geeglm
 un.geeglm <- geeglm(seize ~ trt, id=subj,family=poisson("log"),
             offset=logo, corstr="unstructured",data=dat.sz)
 summary(un.geeglm)
```

```
##
## Call:
## geeglm(formula = seize ~ trt, family = poisson("log"), data = dat.sz,
##     offset = logo, id = subj, corstr = "unstructured")
##
##  Coefficients:
##             Estimate Std.err Wald Pr(>|W|)
## (Intercept)    -1846       0  Inf   <2e-16 ***
## trt             1852       0  Inf   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation structure = unstructured
## Estimated Scale Parameters:
##
##              Estimate   Std.err
## (Intercept) 9.954e+17 3.302e+17
##   Link = identity
##
## Estimated Correlation Parameters:
##           Estimate Std.err
## alpha.1:2   1.0262 0.14746
## alpha.1:3   0.9145 0.07836
## alpha.1:4   0.9718 0.18927
## alpha.1:5   0.8637 0.09313
## alpha.2:3   0.3005 0.04790
## alpha.2:4   0.3326 0.08001
## alpha.2:5   0.2680 0.04504
## alpha.3:4   0.3202 0.09746
## alpha.3:5   0.2423 0.02407
## alpha.4:5   0.3068 0.10557
## Number of clusters:   59  Maximum cluster size: 5
```

```
 ## CS using geeglm
 cs.geeglm <- geeglm(seize ~ trt,id=subj,family=poisson("log"),
             offset=logo, corstr="exchangeable",data=dat.sz)
summary(cs.geeglm)
```

```
##
## Call:
## geeglm(formula = seize ~ trt, family = poisson("log"), data = dat.sz,
##     offset = logo, id = subj, corstr = "exchangeable")
##
```

```
##  Coefficients:
##             Estimate Std.err  Wald Pr(>|W|)
## (Intercept)   1.2943  0.1721 56.54  5.5e-14 ***
## trt           0.0766  0.2141  0.13     0.72
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation structure = exchangeable
## Estimated Scale Parameters:
##
##             Estimate Std.err
## (Intercept)     20.6    9.81
##   Link = identity
##
## Estimated Correlation Parameters:
##       Estimate Std.err
## alpha    0.772  0.0953
## Number of clusters:   59  Maximum cluster size: 5
```

```
ar1.geeglm <- geeglm(seize ~ trt,id=subj,family=poisson("log"),
             offset=logo, corstr="ar1",data=dat.sz)
summary(ar1.geeglm)
```

```
##
## Call:
## geeglm(formula = seize ~ trt, family = poisson("log"), data = dat.sz,
##     offset = logo, id = subj, corstr = "ar1")
##
##  Coefficients:
##             Estimate Std.err  Wald Pr(>|W|)
## (Intercept)    1.216   0.195 38.78  4.7e-10 ***
## trt            0.095   0.237  0.16     0.69
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation structure = ar1
## Estimated Scale Parameters:
##
##             Estimate Std.err
## (Intercept)     22.2    10.9
##   Link = identity
##
## Estimated Correlation Parameters:
##       Estimate Std.err
## alpha    0.889  0.0645
## Number of clusters:   59  Maximum cluster size: 5
```

Investigate if there is a different different effect after baseline visit 0

This means include an interaction between the indicator variable for visit>0 and the treatment indicator. Use two different working correlation structures and compare the results.