

GEE and GLMM; interpretation of marginal parameters in logistic regression models; missing data

We'll fit models with general estimating equations (`gee`) and general linear mixed models (`lme4`).

```
In [1]: library(data.table)
library(gee)
library(lme4)
library(tools)
library(xtable)
```

Loading required package: Matrix

Fluoride Data

```
In [2]: head(fluoride.data <- data.table(read.csv('fluoride.csv'), key='id'))
summary(fluoride.data)
```

id	age	income	fluoride	fl
2	3.0	1	0.00000000	FALSE
2	6.0	1	0.05063998	TRUE
2	9.0	1	0.04779446	FALSE
3	1.5	0	0.11742604	TRUE
3	3.0	0	0.08832044	TRUE
3	6.0	0	0.06216184	TRUE

id		age		income		fluoride	
Min.	: 2	Min.	:1.500	Min.	:0.0000	Min.	:0.000000
1st Qu.:	444	1st Qu.:	1.500	1st Qu.:	0.0000	1st Qu.:	0.008185
Median :	934	Median :	3.000	Median :	1.0000	Median :	0.048175
Mean :	929	Mean :	4.675	Mean :	0.6382	Mean :	0.067876
3rd Qu.:	1409	3rd Qu.:	6.000	3rd Qu.:	1.0000	3rd Qu.:	0.104724
Max.	:1886	Max.	:9.000	Max.	:1.0000	Max.	:1.794320

fl	
Mode :	logical
FALSE:	1966
TRUE :	1898

Fluoride Data with Missing Entries

```
In [3]: head(fluoride.miss.data <- data.table(read.csv('fluoride.miss.csv'), key='id'))
summary(fluoride.miss.data)
```

id	age	income	fluoride	fl
2	3.0	1	0.00000000	FALSE
2	6.0	1	0.05063998	TRUE
2	9.0	1	0.04779446	FALSE
3	3.0	0	0.08832044	TRUE
3	6.0	0	0.06216184	TRUE
4	1.5	1	0.03531871	FALSE

id		age		income		fluoride	
Min.	: 2.0	Min.	:1.500	Min.	:0.0000	Min.	:0.000000
1st Qu.:	485.0	1st Qu.:	3.000	1st Qu.:	1.0000	1st Qu.:	0.006707
Median :	975.0	Median :	3.000	Median :	1.0000	Median :	0.042219
Mean :	954.9	Mean :	4.709	Mean :	0.7811	Mean :	0.064560
3rd Qu.:	1431.0	3rd Qu.:	6.000	3rd Qu.:	1.0000	3rd Qu.:	0.100249
Max.	:1886.0	Max.	:9.000	Max.	:1.0000	Max.	:1.794320


```
fl
Mode :logical
FALSE:1679
TRUE :1478
```

Models

General Estimating Equations (GEE)

```
In [4]: gee.age.independent <- gee(fl ~ age, id=id,
                                   family=binomial,
                                   data=fluoride.data)
gee.age.exchangeable <- update(gee.age.independent, corstr='exchangeable')
gee.interaction.independent <- update(gee.age.independent, formula=~. + income + age:income)
gee.interaction.exchangeable <- update(gee.interaction.independent, corstr='exchangeable')
```

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

```
(Intercept)          age
-0.024537225 -0.002280917
```

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

```
(Intercept)          age
-0.024537225 -0.002280917
```

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

```
(Intercept)          age          income  age:income
0.57645733 -0.04872948 -0.96444671  0.07683365
```

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

```
(Intercept)          age          income  age:income
0.57645733 -0.04872948 -0.96444671  0.07683365
```

General Linear Mixed Models (GLMM)

```
In [5]: glmm.age <- glmer(fl ~ age + (1|id), family=binomial, data=fluoride.data)
glmm.interaction <- update(glmm.age, formula=~. + income + age:income)
```

Missing Data and GEE

```
In [6]: gee.age.independent.miss <- update(gee.age.independent, data=fluoride.miss.data)
gee.age.exchangeable.miss <- update(gee.age.exchangeable, data=fluoride.miss.data)
gee.interaction.independent.miss <- update(gee.interaction.independent, data=fluoride.miss.data)
gee.interaction.exchangeable.miss <- update(gee.interaction.exchangeable, data=fluoride.miss.data)
```

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

```
(Intercept)          age
-0.165918518  0.008153993
```

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

```
(Intercept)          age
-0.165918518  0.008153993
```

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

```
(Intercept)          age          income  age:income
0.59166829 -0.05778955 -0.97965768  0.08589372
```

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

```
(Intercept)          age          income  age:income
0.59166829 -0.05778955 -0.97965768  0.08589372
```

Missing Data and GLMM

```
In [7]: glmm.age.miss <- update(glmm.age, data=fluoride.miss.data)
glmm.interaction.miss <- update(glmm.interaction, data=fluoride.miss.data)
```

Estimates and Standard Errors

```

In [8]: summarize.model <- function(model) {
  coefficients <- summary(model)$coefficients
  standard.error <- if (is(model, 'gee')) {
    coefficients[, 'Robust S.E.']
  } else if (is(model, 'glmerMod')) {
    coefficients[, 'Std. Error']
  }
  data.frame(coefficient=row.names(coefficients),
             estimate=coefficients[, 'Estimate'],
             standard.error=standard.error,
             row.names=NULL)
}

key.model <- function(model) {
  data.frame(
    correlation.structure=if (is(model, 'gee')) {
      if (is.null(getCall(model)$corstr)) {
        'GEE Independent'
      } else {
        paste('GEE', toTitleCase(getCall(model)$corstr))
      }
    } else if (is(model, 'glmerMod')) {
      'Mixed Model'
    },
    has.interaction=nrow(summary(model)$coefficients) == 4,
    is.missing=getCall(model)$data == quote(fluoride.miss.data)
  )
}

model.summaries <- do.call(rbind, lapply(list(
  gee.age.independent, gee.age.exchangeable, glmm.age,
  gee.interaction.independent, gee.interaction.exchangeable, glmm.interaction,
  gee.age.independent.miss, gee.age.exchangeable.miss, glmm.age.miss,
  gee.interaction.independent.miss, gee.interaction.exchangeable.miss, glmm.interaction.miss
),
  function(model) {
    cbind(key.model(model), summarize.model(model))
  }
))

write.csv(model.summaries, file='model_summaries.csv', row.names=FALSE)

```