

Categorical Data Analysis: Marginal Modeling of Depression Data via GEE

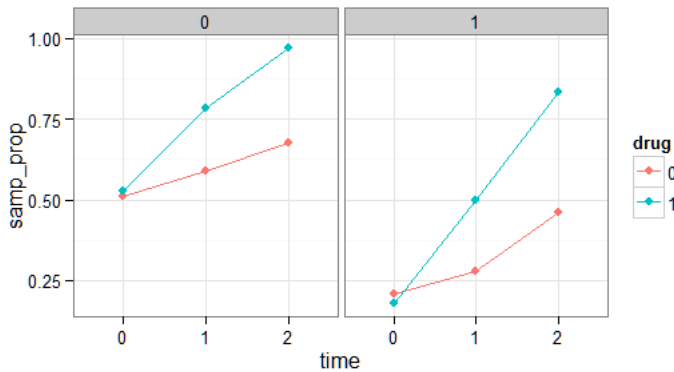
```
>
depr=read.table("http://sites.williams.edu/bklingen/files/2012/02/depression.
txt", header=TRUE)

> head(depr,10)
  case severity drug time outcome
1     1         0   0     0         1
2     1         0   0     1         1
3     1         0   0     2         1
4     2         0   0     0         1
5     2         0   0     1         1
6     2         0   0     2         1
7     3         0   0     0         1
8     3         0   0     1         1
9     3         0   0     2         1
10    4         0   0     0         1
> tail(depr,10)
  case severity drug time outcome
1011  332         1   1     2         0
1012  333         1   1     0         0
1013  333         1   1     1         0
1014  333         1   1     2         0
1015  334         1   1     0         0
1016  334         1   1     1         0
1017  334         1   1     2         0
1018  335         1   1     0         0
1019  335         1   1     1         0
1020  335         1   1     2         0
>
> ## Exploring:
> tabledat <- table(depr$severity,depr$drug,depr$time,depr$outcome)
> tabledat1 <- as.data.frame(tabledat)
> names(tabledat1) <- c("severity", "drug", "time", "outcome",
"count")
> tabledat1
  severity drug time outcome count
1         0   0   0         0     39
2         1   0   0         0     79
3         0   1   0         0     33
4         1   1   0         0     74
5         0   0   1         0     33
6         1   0   1         0     72
7         0   1   1         0     15
8         1   1   1         0     45
9         0   0   2         0     26
10        1   0   2         0     54
11        0   1   2         0      2
12        1   1   2         0     15
13        0   0   0         1     41
14        1   0   0         1     21
15        0   1   0         1     37
16        1   1   0         1     16
17        0   0   1         1     47
18        1   0   1         1     28
```

```

19      0      1      1      1      55
20      1      1      1      1      45
21      0      0      2      1      54
22      1      0      2      1      46
23      0      1      2      1      68
24      1      1      2      1      75
> n <- tabledat1$count[1:12]+tabledat1$count[13:24]
> tabledat2 <- cbind(tabledat1[13:24,], n, tabledat1$count[13:24]/n)
> names(tabledat2)[7] <- "samp_prop"
> tabledat2
  severity drug time outcome count   n samp_prop
13      0      0      0         1   41  80 0.5125000
14      1      0      0         1   21 100 0.2100000
15      0      1      0         1   37  70 0.5285714
16      1      1      0         1   16  90 0.1777778
17      0      0      1         1   47  80 0.5875000
18      1      0      1         1   28 100 0.2800000
19      0      1      1         1   55  70 0.7857143
20      1      1      1         1   45  90 0.5000000
21      0      0      2         1   54  80 0.6750000
22      1      0      2         1   46 100 0.4600000
23      0      1      2         1   68  70 0.9714286
24      1      1      2         1   75  90 0.8333333
> require(ggplot2)
> ggplot(data=tabledat2, aes(x=time,y=samp_prop, color=drug, group=drug)) +
  geom_point() + geom_line() + facet_grid(~severity) + theme_bw()

```



```

> ## Assuming all observation on the same subject are independent:
> fit_logistic <-
glm(outcome~severity+drug+time+drug*time,family=binomial,data=depr)
> summary(fit_logistic)

```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.02799	0.16391	-0.171	0.864
severity	-1.31391	0.14641	-8.974	< 2e-16 ***
drug	-0.05960	0.22221	-0.268	0.789
time	0.48241	0.11476	4.204	2.63e-05 ***
drug:time	1.01744	0.18879	5.389	7.08e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1411.9 on 1019 degrees of freedom
Residual deviance: 1161.9 on 1015 degrees of freedom
AIC: 1171.9

Number of Fisher Scoring iterations: 4

```
> anova(fit_logistic, test="LRT")  
Analysis of Deviance Table
```

Model: binomial, link: logit

Response: outcome

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			1019	1412.0	
severity	1	72.295	1018	1339.7	< 2.2e-16 ***
drug	1	35.641	1017	1304.0	2.373e-09 ***
time	1	111.302	1016	1192.7	< 2.2e-16 ***
drug:time	1	30.767	1015	1161.9	2.909e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>

```
> require(gee) #load GEE package  
> fit_exch <-  
gee(outcome~severity+drug+time+drug*time,family=binomial,id=case,data=depr,co  
rstr = "exchangeable")  
> summary(fit_exch)  
GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
```

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

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	-0.02809866	0.1625499	-0.1728617	0.1741791	-0.1613205
severity	-1.31391033	0.1448627	-9.0700418	0.1459630	-9.0016667
drug	-0.05926689	0.2205340	-0.2687427	0.2285569	-0.2593091
time	0.48246420	0.1141154	4.2278625	0.1199383	4.0226037
drug:time	1.01719312	0.1877051	5.4191018	0.1877014	5.4192084

Estimated Scale Parameter: 0.985392
Number of Iterations: 2

Working Correlation

	[,1]	[,2]	[,3]
[1,]	1.000000000	-0.003432732	-0.003432732
[2,]	-0.003432732	1.000000000	-0.003432732
[3,]	-0.003432732	-0.003432732	1.000000000

```

> fit_unstr <-
gee(outcome~severity+drug+time+drug*time,family=binomial,id=case,data=depr,co
rstr = "unstructured")
> summary(fit_unstr)
GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA

```

```

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

```

```

Coefficients:
      Estimate Naive S.E.   Naive z Robust S.E.   Robust z
(Intercept) -0.02552611  0.1664840 -0.1533247   0.1726392 -0.1478581
severity     -1.30484850  0.1448724 -9.0068787   0.1450136 -8.9981088
drug         -0.05438636  0.2261876 -0.2404480   0.2271321 -0.2394481
time         0.47587182  0.1150534  4.1360955   0.1190418  3.9975178
drug:time    1.01297603  0.1870636  5.4151419   0.1865407  5.4303205

```

```

Estimated Scale Parameter: 0.9823364
Number of Iterations: 3

```

```

Working Correlation
      [,1] [,2] [,3]
[1,] 1.0000000 0.07393977 -0.02741128
[2,] 0.07393977 1.0000000 -0.05669559
[3,] -0.02741128 -0.05669559 1.0000000

```

```

> fit_ind <-
gee(outcome~severity+drug+time+drug*time,family=binomial,id=case,data=depr,co
rstr = "independence")
> summary(fit_ind)
GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA

```

```

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

```

```

Coefficients:
      Estimate Naive S.E.   Naive z Robust S.E.   Robust z
(Intercept) -0.02798843  0.1627083 -0.1720160   0.1741865 -0.1606808
severity     -1.31391092  0.1453432 -9.0400569   0.1459845 -9.0003423
drug         -0.05960381  0.2205812 -0.2702126   0.2285385 -0.2608042
time         0.48241209  0.1139224  4.2345663   0.1199350  4.0222784
drug:time    1.01744498  0.1874132  5.4288855   0.1876938  5.4207709

```

```

Estimated Scale Parameter: 0.9854113
Number of Iterations: 1

```

```

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

```

```

> #Let's go with with exchangeable model
> # Plotting fitted Model
> # get design matrix
> X <- tabledat2[,1:3]
> X <- apply(X,2,as.numeric) #changes factors to numerical values
> X
      severity drug time
[1,]         0    0    0
[2,]         1    0    0
[3,]         0    1    0
[4,]         1    1    0
[5,]         0    0    1
[6,]         1    0    1
[7,]         0    1    1
[8,]         1    1    1
[9,]         0    0    2
[10,]        1    0    2
[11,]        0    1    2
[12,]        1    1    2
> X <- cbind(1, X, X[,2]*X[,3]) #design matrix
> coef <- as.matrix(fit_exch$coefficients,ncol=1) #estimated beta vector
> eta <- X * coef #fitted linear predictor
> fit_prob <- exp(eta)/(1+exp(eta)) #fitted probabilities
> tabledat2$fit_prob <- fit_prob
> tabledat2
  severity drug time outcome count   n samp_prop fit_prob
13         0    0    0         1   41  80 0.5125000 0.4929758
14         1    0    0         1   21 100 0.2100000 0.2071799
15         0    1    0         1   37  70 0.5285714 0.4781725
16         1    1    0         1   16  90 0.1777778 0.4051426
17         0    0    1         1   47  80 0.5875000 0.6116767
18         1    0    1         1   28 100 0.2800000 0.2974345
19         0    1    1         1   55  70 0.7857143 0.5975095
20         1    1    1         1   45  90 0.5000000 0.5245755
21         0    0    2         1   54  80 0.6750000 0.7184588
22         1    0    2         1   46 100 0.4600000 0.4068312
23         0    1    2         1   68  70 0.9714286 0.7063169
24         1    1    2         1   75  90 0.8333333 0.6412620
> ggplot(data=tabledat2, aes(x=time,y=fit_prob, color=drug, group=drug)) +
geom_point() + geom_line() + facet_grid(~severity) + theme_bw()

```

