stat consulting

#Quasi-indenpendence

```r
datagss <- data.frame(
  Now = c("NortheastN", "MidwestN", "SouthN", "WestN",
          "NortheastN", "MidwestN", "SouthN", "WestN",
          "NortheastN", "MidwestN", "SouthN", "WestN",
          "NortheastN", "MidwestN", "SouthN", "WestN"),
  Age16 =rep( c("Northeast16", "Midwest16", "South16", "West16"),each=4),
  Count = c(394,17,81,38,8,596,74,59,29,32,769,35,10,24,35,417)
)
fit1<-glm(Count ~ Now + Age16,family=poisson,data=datagss)
summary(fit1)
```

```
Call:
glm(formula = Count ~ Now + Age16, family = poisson, data = datagss)

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)    
(Intercept)       5.23821    0.04970 105.407  < 2e-16 ***
NowNortheastN    -0.41674    0.06133  -6.795 1.09e-11 ***
NowSouthN         0.36011    0.05037   7.149 8.76e-13 ***
NowWestN         -0.19769    0.05759  -3.433 0.000597 ***
Age16Northeast16 -0.32971    0.05695  -5.789 7.07e-09 ***
Age16South16      0.16014    0.05013   3.195 0.001400 ** 
Age16West16      -0.41638    0.05843  -7.126 1.03e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)
```

```
    Null deviance: 4236.9  on 15  degrees of freedom
Residual deviance: 3871.7  on  9  degrees of freedom
AIC: 3981

Number of Fisher Scoring iterations: 6
```

```
datagss_quasi <- data.frame(
  Now = c("NortheastN", "MidwestN", "SouthN", "WestN",
          "NortheastN", "MidwestN", "SouthN", "WestN",
          "NortheastN", "MidwestN", "SouthN", "WestN",
          "NortheastN", "MidwestN", "SouthN", "WestN"),
  Age16 =rep( c("Northeast16", "Midwest16", "South16", "West16"),each=4),
  Count = c( 394,17,81,38,8,596,74,59,29,32,769,35,10,24,35,417),
  diag_index = c(1,0,0,0,0,2,0,0,0,0,3,0,0,0,0,4) )
fit2 <- glm(Count ~ factor(Now) + factor(Age16) + factor(diag_index),family=poisson, data=da
summary(fit2)
```

```
Call:
glm(formula = Count ~ factor(Now) + factor(Age16) + factor(diag_index),
    family = poisson, data = datagss_quasi)

Coefficients:
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)               3.38251    0.16019  21.115  < 2e-16 ***
factor(Now)NortheastN    -0.48251    0.19270  -2.504  0.01228 *
factor(Now)SouthN         0.97109    0.15062   6.447 1.14e-10 ***
factor(Now)WestN          0.42567    0.15172   2.806  0.00502 **
factor(Age16)Northeast16 -0.11300    0.12406  -0.911  0.36237
factor(Age16)South16      0.03512    0.14716   0.239  0.81137
factor(Age16)West16      -0.59721    0.15254  -3.915 9.04e-05 ***
factor(diag_index)1       3.18935    0.18796  16.968  < 2e-16 ***
factor(diag_index)2       3.00773    0.16535  18.191  < 2e-16 ***
factor(diag_index)3       2.25637    0.15333  14.716  < 2e-16 ***
factor(diag_index)4       2.82212    0.17038  16.563  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 4236.927  on 15  degrees of freedom
Residual deviance:   29.803  on  5  degrees of freedom
```

```
AIC: 147.09

Number of Fisher Scoring iterations: 4
```

#bradley-Terry

```r
Tennis <- read.table("http://www.stat.ufl.edu/~aa/cat/data/Tennis.dat",header=TRUE)
Tennis
```

```
   Djokovic Federer Murray Nadal Wawrinka nij nji
1         1      -1      0     0        0   9   6
2         1       0     -1     0        0  14   3
3         1       0      0    -1        0   9   2
4         1       0      0     0       -1   4   3
5         0       1     -1     0        0   5   0
6         0       1      0    -1        0   5   1
7         0       1      0     0       -1   7   2
8         0       0      1    -1        0   2   4
9         0       0      1     0       -1   2   2
10        0       0      0     1       -1   4   3
```

```r
Tennis
```

```
   Djokovic Federer Murray Nadal Wawrinka nij nji
1         1      -1      0     0        0   9   6
2         1       0     -1     0        0  14   3
3         1       0      0    -1        0   9   2
4         1       0      0     0       -1   4   3
5         0       1     -1     0        0   5   0
6         0       1      0    -1        0   5   1
7         0       1      0     0       -1   7   2
8         0       0      1    -1        0   2   4
9         0       0      1     0       -1   2   2
10        0       0      0     1       -1   4   3
```

```r
data <- data.frame(
  Winner = c("Djokovic", "Federer", "Murray", "Nadal", "Wawrinka"),
  Djokovic = c(NA, 6, 3, 2, 3),
  Federer = c(9, NA, 0, 1, 2),
  Murray = c(14, 5, NA, 4, 2),
  Nadal = c(9, 5,2, NA, 3),
  Wawrinka = c(4, 7, 2, 4, NA)
)
table_result <- as.matrix(data[,-1])
rownames(table_result) <- data$Winner
table_result
```

```
         Djokovic Federer Murray Nadal Wawrinka
Djokovic       NA       9     14     9        4
Federer         6      NA      5     5        7
Murray          3       0     NA     2        2
Nadal           2       1      4    NA        4
Wawrinka        3       2      2     3       NA
```

```r
library(knitr)
kable(table_result, format = "markdown", col.names = c("Djokovic", "Federer", "Murray", "Nada
      caption = "Tennis Match Wins (Loser in Columns, Winner in Rows)")
```

Table 1: Tennis Match Wins (Loser in Columns, Winner in Rows)

	Djokovic	Federer	Murray	Nadal	Wawrinka
Djokovic	NA	9	14	9	4
Federer	6	NA	5	5	7
Murray	3	0	NA	2	2
Nadal	2	1	4	NA	4
Wawrinka	3	2	2	3	NA

```r
fit <- glm(nij/(nij+nji) ~ -1 + Djokovic + Federer + Murray + Nadal + Wawrinka, family=binomi
summary(fit)
```

```

Call:
glm(formula = nij/(nij + nji) ~ -1 + Djokovic + Federer + Murray +
    Nadal + Wawrinka, family = binomial, data = Tennis, weights = nij +
    nji)

Coefficients: (1 not defined because of singularities)
         Estimate Std. Error z value Pr(>|z|)
Djokovic  1.17612    0.49952   2.354   0.0185 *
Federer   1.13578    0.51095   2.223   0.0262 *
```

```
Murray   -0.56852    0.56833  -1.000   0.3172
Nadal    -0.06185    0.51487  -0.120   0.9044
Wawrinka       NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 26.8960  on 10  degrees of freedom
Residual deviance:  4.3958  on  6  degrees of freedom
AIC: 34.041

Number of Fisher Scoring iterations: 4
```

```
library(BradleyTerry2)
Head2Head<-countsToBinomial(table_result)
names(Head2Head)[3:4]<-c("Win", "Lose")
model<-BTm(cbind(Win, Lose), player1, player2, formula=~player, id="player", refcat="Wawrinka
summary(model)
```

```
Call:
BTm(outcome = cbind(Win, Lose), player1 = player1, player2 = player2,
    formula = ~player, id = "player", refcat = "Wawrinka", data = Head2Head)

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
playerDjokovic  1.17612    0.49952   2.354   0.0185 *
playerFederer   1.13578    0.51095   2.223   0.0262 *
playerMurray   -0.56852    0.56833  -1.000   0.3172
playerNadal    -0.06185    0.51487  -0.120   0.9044
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 26.8960  on 10  degrees of freedom
Residual deviance:  4.3958  on  6  degrees of freedom
AIC: 34.041

Number of Fisher Scoring iterations: 4
```

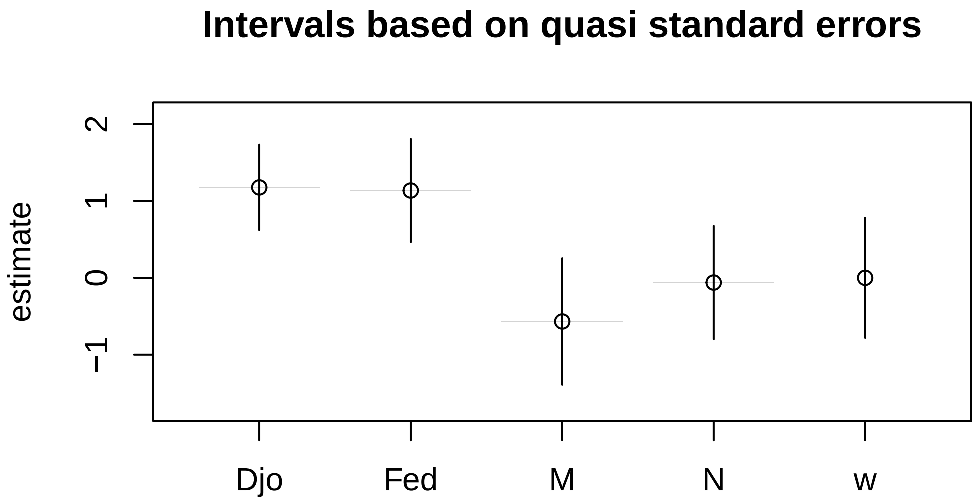
```
BTabilities(model)
```

```
             ability      s.e.
Djokovic  1.17612172 0.4995230
Federer   1.13578408 0.5109457
Murray   -0.56851913 0.5683333
Nadal    -0.06185141 0.5148698
Wawrinka  0.00000000 0.0000000
```

```
model2<-update(model, refcat="Nadal")
BTabilities(model2)
```

```
             ability      s.e.
Djokovic  1.23797313 0.4736563
Federer   1.19763549 0.5162229
Murray   -0.50666771 0.5367784
Nadal     0.00000000 0.0000000
Wawrinka  0.06185141 0.5148698
```

```
library("qvcalc")
tennis.qv <- qvcalc(BTabilities(model))
plot(tennis.qv,levelNames = c("Djo","Fed","M","N","w"))
```


Intervals based on quasi standard errors
estimate
Djo
Fed
M
N
w