Duration of Unemployment - Analysis of Deviance Table for Nested Models

February 5, 2020

The data unemployment is included as a contingency table. The response is the duration of unemployment, gender and the level of education are predictors.

```
> unemployment <- matrix(c(97, 216, 56, 34, 105, 91, 31, 11,
+                   45, 81, 32, 9, 51, 81, 34, 9), nrow=8, ncol=2)
> rownames(unemployment) <- c(paste("male", 1:4), paste("female", 1:4))
> colnames(unemployment) <- c("Short term","Long term")
> unemployment
```

```
         Short term Long term
male 1           97        45
male 2          216        81
male 3           56        32
male 4           34         9
female 1        105        51
female 2         91        81
female 3         31        34
female 4         11         9
```

In the first part the data are considered as ungrouped. Thus, first the dataset is transformed into single observations on the variables y (duration of unemployment, binary), L (level of education) and G (gender).

```
> y <- c(rep(1, sum(97, 216, 56, 34, 105, 91, 31, 11)),
+        rep(0, sum(45, 81, 32, 9, 51, 81, 34, 9)))
> G <- c(rep(1, sum(97, 216, 56, 34)), rep(0, sum(105, 91, 31, 11)),
+        rep(1, sum(45, 81, 32, 9)), rep(0, sum(51, 81, 34, 9)))
> L <- factor(c(rep(1, 97), rep(2, 216), rep(3, 56), rep(4, 34),
+        rep(1, 105), rep(2, 91), rep(3, 31), rep(4, 11),
+        rep(1, 45), rep(2, 81), rep(3, 32), rep(4, 9),
+        rep(1, 51), rep(2, 81), rep(3, 34), rep(4, 9)))
> table(G,L,y)
```

```
, , y = 0

   L
G     1   2   3   4
  0  51  81  34   9
```

```
  1  45  81  32   9

, , y = 1

   L
G     1   2   3   4
  0 105  91  31  11
  1  97 216  56  34
```

Fitting of various logit models; in particular, the saturated model (model with both covariates and their interaction), the model with main effects, the two models with only one covariate and the intercept model. Deviances are for ungrouped data

```
> unemp_1 <- glm(y ~ 1,family=binomial)
> unemp_G <- glm(y ~ G,family=binomial)
> unemp_L <- glm(y ~ L,family=binomial)
> unemp_LG <- glm(y ~ G + L,family=binomial)
> unemp_sat <- glm(y ~ G * L,family=binomial)
> summary(unemp_sat)

Call:
glm(formula = y ~ G * L, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7686  -1.2272   0.7981   0.8898   1.2169

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.72213    0.17068   4.231 2.33e-05 ***
G            0.04591    0.24832   0.185  0.85331
L2          -0.60572    0.22906  -2.644  0.00818 **
L3          -0.81451    0.30133  -2.703  0.00687 **
L4          -0.52146    0.48078  -1.085  0.27809
G:L2         0.81851    0.31933   2.563  0.01037 *
G:L3         0.60608    0.41526   1.460  0.14442
G:L4         1.08255    0.63577   1.703  0.08862 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1270.3  on 982  degrees of freedom
Residual deviance: 1237.4  on 975  degrees of freedom
AIC: 1253.4

Number of Fisher Scoring iterations: 4
```

Tests for hierarchies and corresponding effects:

```
> anova(unemp_LG, unemp_sat)
```

```
Analysis of Deviance Table

Model 1: y ~ G + L
Model 2: y ~ G * L
  Resid. Df Resid. Dev Df Deviance
1       978     1245.0
2       975     1237.4  3   7.5213
```

```
> anova(unemp_L, unemp_LG)

Analysis of Deviance Table

Model 1: y ~ L
Model 2: y ~ G + L
  Resid. Df Resid. Dev Df Deviance
1       979     1263.8
2       978     1245.0  1   18.808
```

```
> anova(unemp_1, unemp_L)

Analysis of Deviance Table

Model 1: y ~ 1
Model 2: y ~ L
  Resid. Df Resid. Dev Df Deviance
1       982     1270.3
2       979     1263.8  3   6.5573
```

```
> anova(unemp_LG, unemp_sat)

Analysis of Deviance Table

Model 1: y ~ G + L
Model 2: y ~ G * L
  Resid. Df Resid. Dev Df Deviance
1       978     1245.0
2       975     1237.4  3   7.5213
```

```
> anova(unemp_G, unemp_LG)

Analysis of Deviance Table

Model 1: y ~ G
Model 2: y ~ G + L
  Resid. Df Resid. Dev Df Deviance
1       981     1252.4
2       978     1245.0  3   7.4063
```

```
> anova(unemp_1, unemp_G)

Analysis of Deviance Table
```

```
Model 1: y ~ 1
Model 2: y ~ G
  Resid. Df Resid. Dev Df Deviance
1       982     1270.3
2       981     1252.4  1   17.959
```

Tests that can be used to obtain the deviances for the grouped data.

```
> anova(unemp_1, unemp_sat)

Analysis of Deviance Table

Model 1: y ~ 1
Model 2: y ~ G * L
  Resid. Df Resid. Dev Df Deviance
1       982     1270.3
2       975     1237.4  7   32.886

> anova(unemp_L, unemp_sat)

Analysis of Deviance Table

Model 1: y ~ L
Model 2: y ~ G * L
  Resid. Df Resid. Dev Df Deviance
1       979     1263.8
2       975     1237.4  4   26.329

> anova(unemp_G, unemp_sat)

Analysis of Deviance Table

Model 1: y ~ G
Model 2: y ~ G * L
  Resid. Df Resid. Dev Df Deviance
1       981     1252.4
2       975     1237.4  6   14.928

> anova(unemp_LG, unemp_sat)

Analysis of Deviance Table

Model 1: y ~ G + L
Model 2: y ~ G * L
  Resid. Df Resid. Dev Df Deviance
1       978     1245.0
2       975     1237.4  3   7.5213
```

In the second part the model are fitted as grouped data, which directly yields the deviances for the grouped data case. The parameter estimates remain the same, but the deviances and the AIC differ from the ungrouped case.

```
> genderleveldat<-data.frame("Long term"=unemployment[,1],
+ "Short term"=unemployment[,2],"Level"=rep(1:4,2),"Gender"=rep(c(1,0),each=4))
> groupintercept<-glm(cbind(Long.term, Short.term) ~ 1, family=binomial,
+                     data=genderleveldat)
> summary(groupintercept)

Call:
glm(formula = cbind(Long.term, Short.term) ~ 1, family = binomial,
    data = genderleveldat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.3163  -1.4275   0.1223   1.0837   2.7745

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.62822    0.06696   9.382   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 32.886  on 7  degrees of freedom
Residual deviance: 32.886  on 7  degrees of freedom
AIC: 73.818

Number of Fisher Scoring iterations: 4

> #Corresponding un-grouped model:
> summary(unemp_1)

Call:
glm(formula = y ~ 1, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4531  -1.4531   0.9247   0.9247   0.9247

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.62822    0.06696   9.382   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1270.3  on 982  degrees of freedom
Residual deviance: 1270.3  on 982  degrees of freedom
AIC: 1272.3

Number of Fisher Scoring iterations: 4
```

```
> groupgender<-glm(cbind(Long.term, Short.term) ~ Gender, family=binomial,
+                  data=genderleveldat)
> summary(groupgender)

Call:
glm(formula = cbind(Long.term, Short.term) ~ Gender, family = binomial,
    data = genderleveldat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6098  -1.2923  -0.4293   0.8908   2.4806

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.30748    0.09958   3.088  0.00202 **
Gender       0.57346    0.13559   4.229 2.34e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 32.886  on 7  degrees of freedom
Residual deviance: 14.928  on 6  degrees of freedom
AIC: 57.859

Number of Fisher Scoring iterations: 3

> #Corresponding un-grouped model:
> summary(unemp_G)

Call:
glm(formula = y ~ G, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5669  -1.3105   0.8327   0.8327   1.0499

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.30748    0.09958   3.088  0.00202 **
G            0.57346    0.13559   4.229 2.34e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1270.3  on 982  degrees of freedom
Residual deviance: 1252.4  on 981  degrees of freedom
AIC: 1256.4

Number of Fisher Scoring iterations: 4
```

```
> grouplevel<-glm(cbind(Long.term, Short.term) ~ as.factor(Level), family=binomial,
+                 data=genderleveldat)
> summary(grouplevel)

Call:
glm(formula = cbind(Long.term, Short.term) ~ as.factor(Level),
    family = binomial, data = genderleveldat)

Deviance Residuals:
  male 1    male 2    male 3    male 4  female 1  female 2  female 3  female 4
  0.1340    2.6858    1.2933    1.1442   -0.1275   -3.3867   -1.4844   -1.5577

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)         0.7439     0.1240   6.001 1.96e-09 ***
as.factor(Level)2  -0.1047     0.1575  -0.665   0.5063
as.factor(Level)3  -0.4677     0.2050  -2.282   0.0225 *
as.factor(Level)4   0.1724     0.3052   0.565   0.5722
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 32.886  on 7  degrees of freedom
Residual deviance: 26.329  on 4  degrees of freedom
AIC: 73.261

Number of Fisher Scoring iterations: 4

> #Corresponding un-grouped model:
> summary(unemp_L)

Call:
glm(formula = y ~ L, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5829  -1.4581   0.8819   0.9206   1.0626

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.7439     0.1240   6.001 1.96e-09 ***
L2           -0.1047     0.1575  -0.665   0.5063
L3           -0.4677     0.2050  -2.282   0.0225 *
L4            0.1724     0.3052   0.565   0.5722
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1270.3  on 982  degrees of freedom
```

```
Residual deviance: 1263.8  on 979  degrees of freedom
AIC: 1271.8

Number of Fisher Scoring iterations: 4
```

```
> groupgenderlevel<-glm(cbind(Long.term, Short.term) ~ as.factor(Gender) +
+   as.factor(Level), family=binomial, data=genderleveldat)
> summary(groupgenderlevel)
```

```
Call:
glm(formula = cbind(Long.term, Short.term) ~ as.factor(Gender) +
    as.factor(Level), family = binomial, data = genderleveldat)

Deviance Residuals:
  male 1    male 2    male 3    male 4  female 1  female 2  female 3  female 4
 -1.6508    0.8595    0.1123    0.5655    1.4615   -1.0290   -0.1260   -0.7171

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)         0.47594    0.13803   3.448 0.000564 ***
as.factor(Gender)1  0.59585    0.13780   4.324 1.53e-05 ***
as.factor(Level)2  -0.20203    0.16073  -1.257 0.208790
as.factor(Level)3  -0.53702    0.20792  -2.583 0.009801 **
as.factor(Level)4   0.04949    0.30918   0.160 0.872833
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 32.8863  on 7  degrees of freedom
Residual deviance:  7.5213  on 3  degrees of freedom
AIC: 56.453

Number of Fisher Scoring iterations: 4
```

```
> #Corresponding un-grouped model:
> summary(unemp_LG)
```

```
Call:
glm(formula = y ~ G + L, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6753  -1.2957   0.8367   0.9603   1.2035

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.47594    0.13803   3.448 0.000564 ***
G            0.59585    0.13780   4.324 1.53e-05 ***
L2          -0.20203    0.16073  -1.257 0.208790
L3          -0.53702    0.20792  -2.583 0.009801 **
```

```
L4           0.04949    0.30918   0.160 0.872833
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1270.3  on 982  degrees of freedom
Residual deviance: 1245.0  on 978  degrees of freedom
AIC: 1255

Number of Fisher Scoring iterations: 4
```

```
> groupsat<-glm(cbind(Long.term, Short.term) ~ as.factor(Gender) * as.factor(Level),
+               family=binomial, data=genderleveldat)
> summary(groupsat)
```

```
Call:
glm(formula = cbind(Long.term, Short.term) ~ as.factor(Gender) *
    as.factor(Level), family = binomial, data = genderleveldat)

Deviance Residuals:
[1]  0  0  0  0  0  0  0  0

Coefficients:
                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                          0.72213    0.17068   4.231 2.33e-05 ***
as.factor(Gender)1                   0.04591    0.24832   0.185  0.85331
as.factor(Level)2                   -0.60572    0.22906  -2.644  0.00818 **
as.factor(Level)3                   -0.81451    0.30133  -2.703  0.00687 **
as.factor(Level)4                   -0.52146    0.48078  -1.085  0.27809
as.factor(Gender)1:as.factor(Level)2  0.81851    0.31933   2.563  0.01037 *
as.factor(Gender)1:as.factor(Level)3  0.60608    0.41526   1.460  0.14442
as.factor(Gender)1:as.factor(Level)4  1.08255    0.63577   1.703  0.08862 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3.2886e+01  on 7  degrees of freedom
Residual deviance: 1.4211e-14  on 0  degrees of freedom
AIC: 54.932

Number of Fisher Scoring iterations: 3
```

```
> #Corresponding un-grouped model:
> summary(unemp_sat)
```

```
Call:
glm(formula = y ~ G * L, family = binomial)

Deviance Residuals:
```

```
    Min       1Q   Median       3Q      Max
-1.7686  -1.2272   0.7981   0.8898   1.2169

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.72213    0.17068   4.231 2.33e-05 ***
G            0.04591    0.24832   0.185  0.85331
L2          -0.60572    0.22906  -2.644  0.00818 **
L3          -0.81451    0.30133  -2.703  0.00687 **
L4          -0.52146    0.48078  -1.085  0.27809
G:L2         0.81851    0.31933   2.563  0.01037 *
G:L3         0.60608    0.41526   1.460  0.14442
G:L4         1.08255    0.63577   1.703  0.08862 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1270.3  on 982  degrees of freedom
Residual deviance: 1237.4  on 975  degrees of freedom
AIC: 1253.4

Number of Fisher Scoring iterations: 4
```

ANOVA for grouped data:

```
> anova(groupgenderlevel, groupsat)

Analysis of Deviance Table

Model 1: cbind(Long.term, Short.term) ~ as.factor(Gender) + as.factor(Level)
Model 2: cbind(Long.term, Short.term) ~ as.factor(Gender) * as.factor(Level)
  Resid. Df Resid. Dev Df Deviance
1         3     7.5213
2         0     0.0000  3   7.5213

> anova(grouplevel, groupgenderlevel)

Analysis of Deviance Table

Model 1: cbind(Long.term, Short.term) ~ as.factor(Level)
Model 2: cbind(Long.term, Short.term) ~ as.factor(Gender) + as.factor(Level)
  Resid. Df Resid. Dev Df Deviance
1         4    26.3290
2         3     7.5213  1   18.808

> anova(groupintercept, grouplevel)

Analysis of Deviance Table

Model 1: cbind(Long.term, Short.term) ~ 1
```

```
Model 2: cbind(Long.term, Short.term) ~ as.factor(Level)
  Resid. Df Resid. Dev Df Deviance
1         7     32.886
2         4     26.329  3   6.5573
```

```
> anova(groupgenderlevel, groupsat)

Analysis of Deviance Table

Model 1: cbind(Long.term, Short.term) ~ as.factor(Gender) + as.factor(Level)
Model 2: cbind(Long.term, Short.term) ~ as.factor(Gender) * as.factor(Level)
  Resid. Df Resid. Dev Df Deviance
1         3     7.5213
2         0     0.0000  3   7.5213
```

```
> anova(groupgender, groupgenderlevel)

Analysis of Deviance Table

Model 1: cbind(Long.term, Short.term) ~ Gender
Model 2: cbind(Long.term, Short.term) ~ as.factor(Gender) + as.factor(Level)
  Resid. Df Resid. Dev Df Deviance
1         6    14.9275
2         3     7.5213  3   7.4063
```

```
> anova(groupintercept, groupgender)

Analysis of Deviance Table

Model 1: cbind(Long.term, Short.term) ~ 1
Model 2: cbind(Long.term, Short.term) ~ Gender
  Resid. Df Resid. Dev Df Deviance
1         7     32.886
2         6     14.928  1   17.959
```