

Duration of Unemployment - Analysis of Deviance Table for Nested Models

November 1, 2024

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
```

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)
```

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)
```

Tests for hierarchies and corresponding effects:

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

```

> anova(unemp_1, unemp_L)
> anova(unemp_LG, unemp_sat)
> anova(unemp_G, unemp_LG)
> anova(unemp_1, unemp_G)

```

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

```

> anova(unemp_1, unemp_sat)
> anova(unemp_L, unemp_sat)
> anova(unemp_G, unemp_sat)
> anova(unemp_LG, unemp_sat)

```

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)
> #Corresponding un-grouped model:
> summary(unemp_1)
> groupgender<-glm(cbind(Long.term, Short.term) ~ Gender, family=binomial,
+ data=genderleveldat)
> summary(groupgender)
> #Corresponding un-grouped model:
> summary(unemp_G)
> grouplevel<-glm(cbind(Long.term, Short.term) ~ as.factor(Level), family=binomial,
+ data=genderleveldat)
> summary(grouplevel)
> #Corresponding un-grouped model:
> summary(unemp_L)
> groupgenderlevel<-glm(cbind(Long.term, Short.term) ~ as.factor(Gender) +
+ as.factor(Level), family=binomial, data=genderleveldat)
> summary(groupgenderlevel)
> #Corresponding un-grouped model:
> summary(unemp_LG)
> groupsat<-glm(cbind(Long.term, Short.term) ~ as.factor(Gender) * as.factor(Level),
+ family=binomial, data=genderleveldat)
> summary(groupsat)
> #Corresponding un-grouped model:
> summary(unemp_sat)

```

ANOVA for grouped data:

```

> anova(groupgenderlevel, groupsat)
> anova(grouplevel, groupgenderlevel)
> anova(groupintercept, grouplevel)
> anova(groupgenderlevel, groupsat)
> anova(groupgender, groupgenderlevel)
> anova(groupintercept, groupgender)

```