Duration of Unemployment - Logit Model

February 8, 2012

At first the "unemployment" data from the "catdata" package are loaded and attached.

```
> library(catdata)
> data(unemployment)
> attach(unemployment)
```

Now a frequency table is created and used to fit a Logit model based on grouped data.

```
> durbin <- as.factor(durbin)
> table.durbin <- ftable(subset(unemployment, select=c("age", "durbin")),
+ col.vars="durbin")
> rels<-table.durbin[,1]/rowSums(table.durbin)
> age.new <- min(age):max(age)
> model1 <- glm(table.durbin ~ age.new, family=binomial)
> summary(model1)

Call:
glm(formula = table.durbin ~ age.new, family = binomial)

Deviance Residuals:
     Min       1Q     Median        3Q        Max
-2.01336   -0.74344  -0.00988    0.60784    1.72252

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.473369   0.195891   7.521 5.42e-14 ***
age.new     -0.027458   0.005886  -4.665 3.08e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 54.809  on 45  degrees of freedom
Residual deviance: 32.938  on 44  degrees of freedom
AIC: 178.56

Number of Fisher Scoring iterations: 3
```
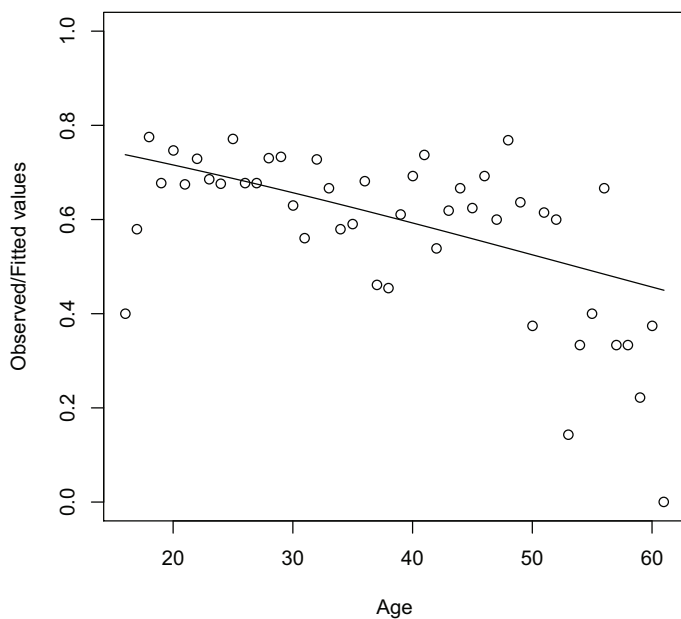
Here the observed frequencies are plotted against the fitted probabilities.

```
> plot(age.new, model1$fitted.values, xlab="Age", ylab="Observed/Fitted values",
+ type="l", ylim=c(0,1))
> points(age.new,table.durbin[,1]/rowSums(table.durbin))
```


The standardized deviance residuals are plotted against the predicted values
and a quantile plot is created.

```
> plot(model1$fitted.values,sqrt(rowSums(table.durbin))*rstandard(model1),
+ xlab="Predicted values", ylab="Residuals")
```
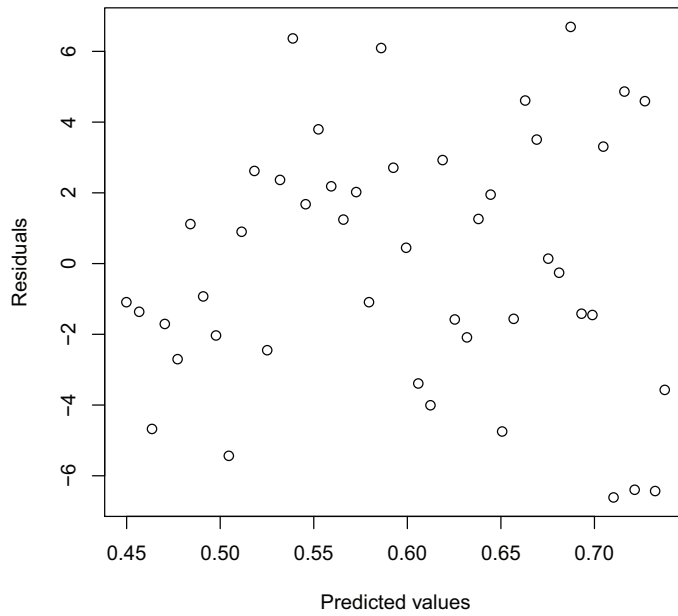
```
> qqnorm(sqrt(rowSums(table.durbin))*rstandard(model1), main="",
+         ylab="Standardized deviance residuals")
> qqline(sqrt(rowSums(table.durbin))*rstandard(model1), lwd=2.5,
+         lty="dashed", col="red")
```

3