5.2.1 - R: Logistic Regression with Polytomous Response

Stat 5100: Dr. Bean

Example: Individuals were surveyed regarding how important they viewed AC and power steering in cars. The sex (M or W), age (1=18-23, 2=24-40, 3=40+), and response (1=little importance, 2=important, 3=very important) of each individual was recorded. The count of responses in each sex/age/response combination was summarized. We want to determine whether and how the sex and age of individuals affects their response.

Example 1: Nominal Logistic Regression

In this nominal logistic regression example, we assume that the classes of the response do not carry any particular order to them.

```
# Load and look at the data
library(stat5100)
data(car)
head(car)

##   sex age response count
## 1   W   1        1    26
## 2   W   1        2    12
## 3   W   1        3     7
## 4   W   2        1     9
## 5   W   2        2    21
## 6   W   2        3    15

# Create dummy variables (necessary for multinomial logistic regression)
# -------------------------
# This line of code sets all women to be coded as 0 and all men to be coded as 1
car <- cbind(car, S = as.numeric(car$sex == "M"))

# Because age has 3 different levels, we need 2 dummy variables here.
car <- cbind(car, A2 = as.numeric(car$age == 2))
car <- cbind(car, A3 = as.numeric(car$age == 3))

# Create a multinomial logistic regression model. Note that in the following
# code, we use the "weights" option inside the multinom() call. This is how
# we incorporate the "count" variable contained inside the car dataset which
# tells us how many observations are seen at each particular X-profile.
car_nom_mlr <- nnet::multinom(as.factor(response) ~ S + A2 + A3, data = car,
                              weights = car$count)

## # weights:  15 (8 variable)
## initial  value 329.583687
## iter  10 value 290.490920
## final  value 290.351098
## converged

summary(car_nom_mlr)
```

```
## Call:
## nnet::multinom(formula = as.factor(response) ~ S + A2 + A3, data = car,
##     weights = car$count)
##
## Coefficients:
##   (Intercept)          S       A2       A3
## 2  -0.5907992 -0.3881301 1.128268 1.587709
## 3  -1.0390726 -0.8130202 1.478104 2.916757
##
## Std. Errors:
##   (Intercept)         S        A2        A3
## 2   0.2839756 0.3005115 0.3416449 0.4028997
## 3   0.3305014 0.3210382 0.4009256 0.4229276
##
## Residual Deviance: 580.7022
## AIC: 596.7022
```

Example 2: Ordinal Logistic Regression

Here, we code the above slightly differently to account for an assumption that the orders of the categories of the response variable mean something. To do this, we can use the polr() function from the MASS package.

```
car_ord_mlr <- MASS::polr(as.factor(response) ~ S + A2 + A3, data = car,
                          weights = car$count)
summary(car_ord_mlr)
```

```
##
## Re-fitting to get Hessian
```

```
## Call:
## MASS::polr(formula = as.factor(response) ~ S + A2 + A3, data = car,
##     weights = car$count)
##
## Coefficients:
##      Value Std. Error t value
## S  -0.5762     0.2262  -2.548
## A2  1.1471     0.2776   4.132
## A3  2.2325     0.2915   7.659
##
## Intercepts:
##     Value   Std. Error t value
## 1|2  0.0435  0.2323     0.1874
## 2|3  1.6550  0.2556     6.4744
##
## Residual Deviance: 581.2956
## AIC: 591.2956
```

Example with predicting some data

Looking at the summaries of the fitted models can feel a little abstract, so let's try creating some hypothetical example data and then using our models to predict what the response variable will be.

We will suppose that we have a 50 year-old man. (Yes, this is very little information about a person, but it is what we have to work with on such a small dataset like this) Thus, we create his profile in the following way:

```r
the_man <- data.frame(S = c(1), A2 = c(0), A3 = c(1))
```

Using our two models, let's predict how much he will likely value AC and power steering in cars:

```r
predict(car_nom_mlr, newdata = the_man)

## [1] 3
## Levels: 1 2 3

predict(car_ord_mlr, newdata = the_man)

## [1] 3
## Levels: 1 2 3
```

In both cases, our models predict that this 50-year old man would very power steering and AC "very much."