Assign8_Choi_GLM

Serena Choi

3/21/2022

Setting up

```
duncan <- read.csv("duncan.csv")
colnames(duncan)<-c("occtitle", "occtype", "income", "educ", "prestige")
attach(duncan)
summary(duncan)
```

```
##    occtitle            occtype              income           educ       
##  Length:45          Length:45          Min.   : 7.00   Min.   :  7.00  
##  Class :character   Class :character   1st Qu.:21.00   1st Qu.: 26.00  
##  Mode  :character   Mode  :character   Median :42.00   Median : 45.00  
##                                        Mean   :41.87   Mean   : 52.56  
##                                        3rd Qu.:64.00   3rd Qu.: 84.00  
##                                        Max.   :81.00   Max.   :100.00  
##     prestige    
##  Min.   : 3.00  
##  1st Qu.:16.00  
##  Median :41.00  
##  Mean   :47.69  
##  3rd Qu.:81.00  
##  Max.   :97.00
```

```
library(nnet)
library(dplyr)
```

```
## 
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union
```

```
table(occtype)
```

```
## occtype
##   bc prof   wc 
##   21   18    6
```

```
help(multinom)
library(car)
```

```
## Loading required package: carData

## Warning: package 'carData' was built under R version 4.1.2

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode
```

```
library(carData)
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 4.1.2

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select
```

Using the Duncan data, estimate the effect of prestige scores on occupational type using either a multinomial logistic regression OR an ordered logit model.

Ordered logit model

```
help(polr)
ordered <- polr (as.ordered(occtype) ~ prestige  , method="logistic")
summary(ordered)
```

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

## Call:
## polr(formula = as.ordered(occtype) ~ prestige, method = "logistic")
##
## Coefficients:
##            Value Std. Error t value
## prestige 0.03782     0.0109   3.471
##
## Intercepts:
##         Value  Std. Error t value
## bc|prof 1.8497 0.6798     2.7209
## prof|wc 4.3452 0.8987     4.8349
##
## Residual Deviance: 74.3773
## AIC: 80.3773
```

According to the ordered logit model above, for a one-unit change in prestige score, the logit of observing higher level is expected to change by 0.038.

Multinomial logistic regression.

```
multinomial <- multinom(factor(occtype) ~ prestige , data=duncan, method="logistic")
```

```
## # weights:  9 (4 variable)
```

```
## initial  value 49.437553
## iter  10 value 20.019561
## final  value 20.015658
## converged
```

```r
summary(multinomial)
```

```
## Call:
## multinom(formula = factor(occtype) ~ prestige, data = duncan,
##     method = "logistic")
##
## Coefficients:
##      (Intercept)   prestige
## prof   -9.318950 0.17392619
## wc     -2.552742 0.04446319
##
## Std. Errors:
##      (Intercept)   prestige
## prof    3.187237 0.05665328
## wc      1.012640 0.02761336
##
## Residual Deviance: 40.03132
## AIC: 48.03132
```

According to the mutlinomial logistic model, the effect of the prestige on logit of being in the professional category relative to the blue collar category is -9.319. The effect of the prestige on logit of being in the white-collar category relative to the blue collar category is -2.553.

##Interpret the effect of prestige scores on occupational type using odds (in one sentence).

ordered

```r
or1 <- exp(coef(ordered))
or1
```

```
## prestige
## 1.038546
```

For a one-unit change in prestige, the effect of one-unit change in prestige on the odd of being in a higher category is 1.039.

multinomial

```r
or2 <- exp(coef(multinomial))
or2
```

```
##       (Intercept) prestige
## prof 8.970806e-05 1.189968
## wc   7.786785e-02 1.045466
```

The coefficients above reflect the odds of being in a given category relative to the blue-collar category.

What is the predicted probability of being a professional for those with a prestige score of 10? What about for those with a prestige score of 90?

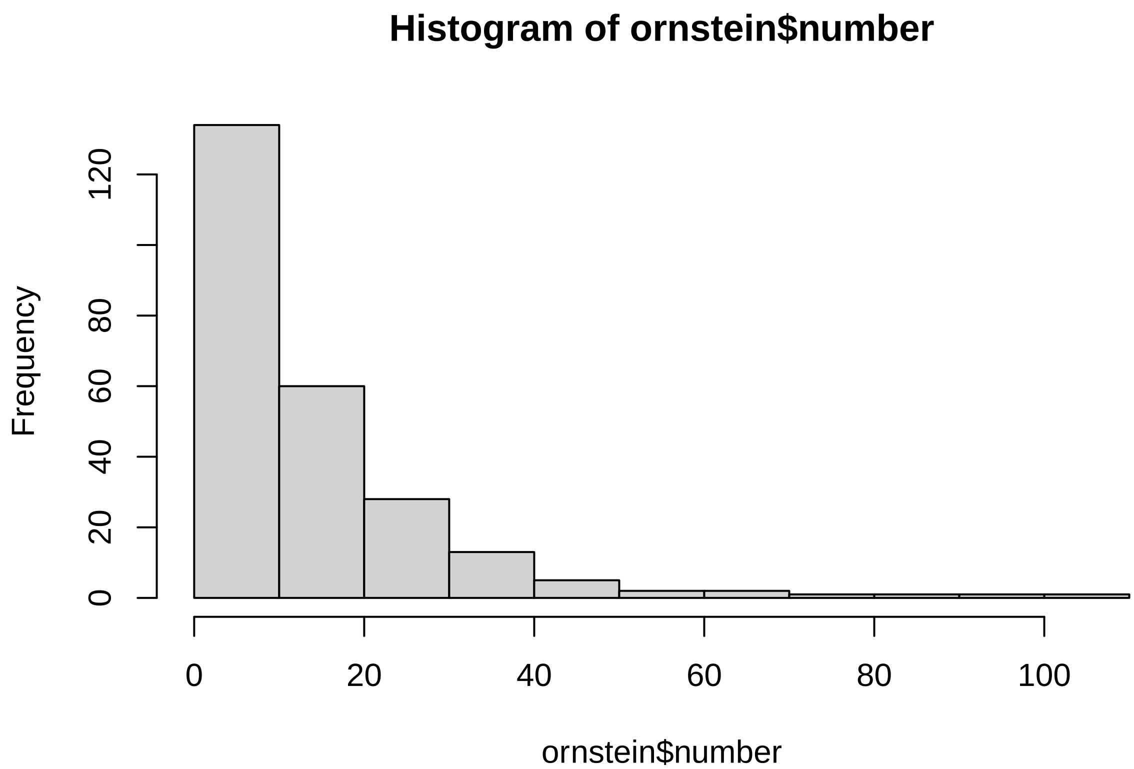
I left the sections below as a comment as I kept getting non-conformable arguments below.

```r
###ordered
#beta1 <- coef(ordered)
#tau1 <- ordered$zeta  # store the intercepts in an object called tau
#X<-cbind(seq(from = 10, to = 90, by = 10)) # looking at the prestige score from 10 to 90
#logit.prob <- function(eta){exp(eta)/(1+exp(eta))}  #the equation for calculating predicted probabilit
#p1 <- logit.prob(tau[1] - X %*% beta)  # calculate the pred probability for blue collar (or category 1
#p2 <- logit.prob(tau[2] - X %*% beta) - logit.prob(tau[1] - X %*% beta) # calculate the pred probabili
#p3 <- 1.0 - logit.prob(tau[2] - X %*% beta) # calculate the pred prob for white collar (or category 3)

#p1 #blue collar
#p2 #professional
#p3     #white collar
```

```r
### multinomial
#beta <- coef(multinomial) # store the coefficients in an object called beta
#tau <- multinomial$zeta  # store the intercepts in an object called tau
#X<-cbind(seq(from = 10, to = 90, by = 10)) # looking at the prestige score from 10 to 90
#logit.prob <- function(eta){exp(eta)/(1+exp(eta))}  #the equation for calculating predicted probabilit
#p1 <- logit.prob(tau[1] - X %*% beta)  # calculate the pred probability for blue collar (or category 1
#p2 <- logit.prob(tau[2] - X %*% beta) - logit.prob(tau[1] - X %*% beta) # calculate the pred probabili
#p3 <- 1.0 - logit.prob(tau[2] - X %*% beta) # calculate the pred prob for white collar (or category 3)

#p1 #blue collar
#p2 #professional
#p3     #white collar
```

##Using the data on interlocking directorates, estimate the effect of assets and country on the number of interlocking directorates among firms.

```r
ornstein <- read.csv("ornstein.csv")
hist(ornstein$number)   #note the Poisson distribution
```




```
table(ornstein$nation)
```

```
##
## CAN OTH  UK  US
## 117  18  17  96
```

```
model3 <- glm(number ~ assets + nation, family=poisson (link ="log"), data=ornstein)
summary(model3)
```

```
##
## Call:
## glm(formula = number ~ assets + nation, family = poisson(link = "log"),
##     data = ornstein)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -5.798  -2.808  -0.939   1.776   9.127
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.715e+00  2.436e-02 111.450  < 2e-16 ***
## assets       1.520e-05  4.368e-07  34.801  < 2e-16 ***
## nationOTH   -1.013e-01  6.687e-02  -1.515     0.13
## nationUK    -5.724e-01  8.565e-02  -6.683 2.34e-11 ***
## nationUS    -8.101e-01  4.516e-02 -17.939  < 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: 3737.0  on 247  degrees of freedom
```

```
## Residual deviance: 2248.9  on 243  degrees of freedom
## AIC: 3156.9
##
## Number of Fisher Scoring iterations: 5
```

```
exp(model3$coef)
```

```
## (Intercept)      assets   nationOTH    nationUK    nationUS
##  15.1091134   1.0000152   0.9036390   0.5641921   0.4448041
```

For every one million dollar increase in assets, the expected count of interlocking directorates increase by 1.00, net of country.