Regression_For_Oridinal_Outcomes

Roxanne Dimadi

2025-11-12

```r
data <- read.table("mentalimpairment-data.txt",header=TRUE)
attach(data)
```

#First, do not take into consideration the order

```r
library(nnet)

#Null model without any predicors
(M0 = multinom(Impairment ~ 1))
```

```
## # weights:  8 (3 variable)
## initial  value 55.451774
## final  value 54.521026
## converged
```

```
## Call:
## multinom(formula = Impairment ~ 1)
##
## Coefficients:
##          (Intercept)
## Mild       0.2876842
## Moderate  -0.2513121
## Well       0.2876842
##
## Residual Deviance: 109.0421
## AIC: 115.0421
```

```r
summary(M0)
```

```
## Call:
## multinom(formula = Impairment ~ 1)
##
## Coefficients:
##          (Intercept)
## Mild       0.2876842
## Moderate  -0.2513121
## Well       0.2876842
##
## Std. Errors:
##          (Intercept)
```

```
## Mild       0.4409587
## Moderate   0.5039527
## Well       0.4409587
##
## Residual Deviance: 109.0421
## AIC: 115.0421
```

```
#using economics status and life events as predictors

(M = multinom(Impairment ~ SES + Events))
```

```
## # weights:  16 (9 variable)
## initial  value 55.451774
## iter  10 value 48.350248
## final  value 48.349131
## converged
```

```
## Call:
## multinom(formula = Impairment ~ SES + Events)
##
## Coefficients:
##          (Intercept)        SES     Events
## Mild        1.247474  1.3575833 -0.3316398
## Moderate    1.240500 -0.3374756 -0.2671694
## Well        1.902708  1.6062981 -0.5601850
##
## Residual Deviance: 96.69826
## AIC: 114.6983
```

Now to preform a likelihood test to see if any of these variables can be deleted from the model For a likelihood ratio test between a reduced model that does not contain a certain predictor and a full model that contains that predictor, the statistics is: D=Deviancereduced−Deviancefull

```
#Model with only economic status
(Ms = multinom(Impairment ~ SES))
```

```
## # weights:  12 (6 variable)
## initial  value 55.451774
## final  value 52.642355
## converged
```

```
## Call:
## multinom(formula = Impairment ~ SES)
##
## Coefficients:
##            (Intercept)        SES
## Mild     -2.231483e-01  0.9163037
## Moderate -8.032857e-06 -0.6931261
## Well     -2.231480e-01  0.9163039
##
## Residual Deviance: 105.2847
## AIC: 117.2847
```

```r
#Model with only life events
(Me = multinom(Impairment ~ Events))
```

```
## # weights:  12 (6 variable)
## initial  value 55.451774
## iter  10 value 50.908223
## final  value 50.908198
## converged
```

```
## Call:
## multinom(formula = Impairment ~ Events)
##
## Coefficients:
##          (Intercept)     Events
## Mild        1.673461 -0.2673717
## Moderate    1.202394 -0.2835592
## Well        2.453627 -0.4847294
##
## Residual Deviance: 101.8164
## AIC: 113.8164
```

```r
#Test whether we can drop SES
1 - pchisq(Me$deviance - M$deviance,
           length(coef(M)) - length(coef(Me)))
```

```
## [1] 0.1633484
```

```r
#Test whether we can drop Events
1 - pchisq(Ms$deviance - M$deviance,
           length(coef(M)) - length(coef(Ms)))
```

```
## [1] 0.03532589
```

```r
BIC(Me)
```

```
## [1] 123.9497
```

```r
BIC(Ms)
```

```
## [1] 127.418
```

```r
BIC(M)
```

```
## [1] 129.8982
```

```r
AIC(Me)
```

```
## [1] 113.8164
```

```r
AIC(Ms)
```

```
## [1] 117.2847
```

```r
AIC(M)
```

```
## [1] 114.6983
```

Using likelihood ratio tests, I compared the full multinomial model containing both SES and Events to reduced models that removed each predictor in turn. When SES was removed (comparing the Events-only model to the full model), the p value was 0.163, indicating that SES does not significantly improve model fit once Events is included. In contrast, when Events was removed (comparing the SES-only model to the full model), the p value was 0.035, showing that Events provides a significant improvement in model fit. Therefore, SES can be dropped from the multinomial model, but Events must be retained, and the preferred reduced model is Impairment ~ Events.

Both AIC (113.8) and BIC (123.9) select this model because it achieves the best balance of goodness-of-fit and model simplicity. SES does not meaningfully improve the fit, consistent with the likelihood ratio test showing SES is not a significant predictor once Events is included.

When modeling the impairment outcome as multinomial (ignoring order), the best model is the Events-only model, Impairment ~ Events.

Choosing one model and determining the probability of being well, mild, moderate or impaired for John who has a low economic status and had seven important life events.

```r
john_dat <- data.frame(Events = 7)

# predicted probabilities for each impairment category
predict(Me, newdata = john_dat, type = "probs")
```

```
##  Impaired      Mild  Moderate      Well
## 0.3747652 0.3074026 0.1713622 0.1464701
```

#Now, considering the order

```r
library(MASS)

Impairment = factor(Impairment,levels=c("Well","Mild","Moderate","Impaired"))

(orderM0 = polr(Impairment ~ 1))
```

```
## Call:
## polr(formula = Impairment ~ 1)
##
## No coefficients
##
## Intercepts:
##         Well|Mild     Mild|Moderate Moderate|Impaired
##        -0.8472977         0.4054527         1.2367503
##
## Residual Deviance: 109.0421
## AIC: 115.0421
```

```r
(orderMs = polr(Impairment ~ SES))
```

```
## Call:
## polr(formula = Impairment ~ SES)
##
## Coefficients:
##        SES
## -0.8553378
##
## Intercepts:
##         Well|Mild     Mild|Moderate Moderate|Impaired
##       -1.36376017       -0.04172851        0.83061097
##
## Residual Deviance: 106.8744
## AIC: 114.8744
```

```r
(orderMe = polr(Impairment ~ Events))
```

```
## Call:
## polr(formula = Impairment ~ Events)
##
## Coefficients:
##  Events
## 0.28793
##
## Intercepts:
##         Well|Mild     Mild|Moderate Moderate|Impaired
##         0.2614334         1.6562752         2.5876265
##
## Residual Deviance: 102.5271
## AIC: 110.5271
```

```r
summary(orderM0)
```

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

## Call:
## polr(formula = Impairment ~ 1)
##
## No coefficients
##
## Intercepts:
##                   Value   Std. Error t value
## Well|Mild         -0.8473  0.3450    -2.4557
## Mild|Moderate      0.4055  0.3227     1.2562
## Moderate|Impaired  1.2368  0.3786     3.2663
##
## Residual Deviance: 109.0421
## AIC: 115.0421
```

```r
summary(orderMe)
```

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

## Call:
## polr(formula = Impairment ~ Events)
##
## Coefficients:
##         Value Std. Error t value
## Events 0.2879     0.1175   2.451
##
## Intercepts:
##                   Value  Std. Error t value
## Well|Mild         0.2614 0.5639     0.4636
## Mild|Moderate     1.6563 0.6171     2.6838
## Moderate|Impaired 2.5876 0.6938     3.7296
##
## Residual Deviance: 102.5271
## AIC: 110.5271
```

```r
summary(orderMs)
```

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

## Call:
## polr(formula = Impairment ~ SES)
##
## Coefficients:
##       Value Std. Error t value
## SES -0.8553     0.5865  -1.458
##
## Intercepts:
##                   Value   Std. Error t value
## Well|Mild         -1.3638  0.5038    -2.7070
## Mild|Moderate     -0.0417  0.4448    -0.0938
## Moderate|Impaired  0.8306  0.4659     1.7827
##
## Residual Deviance: 106.8744
## AIC: 114.8744
```

```r
BIC(orderM0)
```

```
## [1] 120.1087
```

```r
BIC(orderMe)
```

```
## [1] 117.2826
```

```r
BIC(orderMs)
```

```
## [1] 121.6299
```

```r
AIC(orderM0)
```

```
## [1] 115.0421
```

```r
AIC(orderMe)
```

```
## [1] 110.5271
```

```r
AIC(orderMs)
```

```
## [1] 114.8744
```

When the ordinal nature of the impairment outcome is taken into account using a proportional odds model, both AIC and BIC clearly select the model including only the Events predictor (Impairment ~ Events). The AIC for this model (110.53) is the lowest among all models, and the BIC (117.28) is also the lowest, indicating that Events provides meaningful explanatory power while SES does not. SES does not sufficiently improve model fit once Events is included. Therefore, the best ordinal regression model is Impairment ~ Events, and the coefficient for Events is positive, indicating that individuals with more stressful life events are more likely to fall into worse impairment categories.

```r
predict(orderMe, newdata = john_dat, type = "probs")
```

```
##      Well      Mild  Moderate  Impaired
## 0.1475337 0.2636110 0.2281066 0.3607486
```

Using the ordinal logistic regression model that includes only the number of important life events (the best-fitting model based on AIC and BIC), I estimated John's probability of falling into each impairment category. John has low SES and experienced seven major life events, but SES is not included in this model. The predicted probabilities are approximately 14.8 percent for being Well, 26.4 percent for Mild impairment, 22.8 percent for Moderate impairment, and 36.1 percent for being Impaired. Therefore, given his high number of stressful events, the model suggests that John is most likely to be in the Impaired category.