536HW2

Coco_Luo

2022-10-06

R Markdown

Problem 1 Mental Impairment Data

```
M0 = multinom(Impairment ~ 1, data=mental)
MS = multinom(Impairment ~ SES, data = mental)
ME = multinom(Impairment ~ Events, data = mental)
M = multinom(Impairment ~ SES + Events, data = mental)
```

After fitting three models, where: ME include only life event (Events) MS include only economic status (SES) and M include both Events and SES, I found that the AIC for ME is the smallest, but the AIC of M is larger, so I guess the effect of Events is relevant to the response and economic status might not.

```
# manual method
1 - pchisq(MS$deviance - M$deviance, length(coef(M))-length(coef(MS)))
1 - pchisq(ME$deviance - M$deviance, length(coef(M))-length(coef(ME)))
```

When we compare between MS and M, we get a p-value equal to $0.035 < 0.05$. So we reject the null that the coefficients of Events are equal to 0. When we compare between ME and M, we get a p-value equal to 0.163 > 0.05. So we fail to reject the null that the coefficients of SES are equal to 0. Therefore, we can use ME as our model of performance.

$$\pi\text{Mild(events)} = P(Impairment = Mild|Events)$$
$$\pi\text{Moderate(events)} = P(Impairment = Moderate|Events)$$
$$\pi\text{Well(events)} = P(Impairment = Well|Events)$$

And

$$log\frac{\pi Mild(numevents)}{\pi Imparied(num_events)} = 1.673 - 0.267 \cdot Events$$
$$log\frac{\pi moderate(numevents)}{\pi Imparied(numevents)} = 1.202 - 0.284 \cdot Events$$
$$log\frac{\pi Well(numevents)}{\pi Imparied(numevents)} = 2.454 - 0.485 \cdot Events$$

```
coef(ME)%*%c(1,7)
# probability of impair
1/(1+sum(exp(coef(ME)%*%c(1,7))))
exp(coef(ME)%*%c(1,7))/(1+sum(exp(coef(ME)%*%c(1,7))))
```

Since we do not have SES included in our model, we do not need to consider about this variable. We know that John has been through 7 life events, so $log\frac{\pi Mild(7)}{\pi Imparied(7)} = -0.198$,$log\frac{\pi moderate(7)}{\pi Imparied(7)} = -0.784$,$log\frac{\pi Well(7)}{\pi Imparied(7)} = -0.939$. The probability that John's is mentally impaired is 0.375. The probability that his mental impairment status is "Moderate"is 0.171, "Mild" is 0.307 and "Well" is 0.146. From that, I conclude that John is more likely to be impaired due to life events than economic status which does not make much a difference.

Part B: take ordering into consideration

```
attach(mental)
rment = factor(Impairment,levels=c("Well","Mild","Moderate","Impaired"))
orderM0 = polr(Impairment ~ 1)
orderMs = polr(Impairment ~ SES)
orderMe = polr(Impairment ~ Events)
orderM = polr(Impairment ~ SES + Events)
```

```
BIC(orderM0)
BIC(orderMs)
BIC(orderMe)
BIC(orderM)

AIC(orderM0)
AIC(orderMs)
AIC(orderMe)
AIC(orderM)
```

The AIC for orderM (include both SES and Events) is the lowest at this time, and the BIC is lowest for orderME (only events). Both orderM's BIC does not differ much from orderME. Now, let's perform a likelihood ratio test:

```
# given likelihood ratio test
1 - pchisq(orderMs$deviance - orderM$deviance, length(coef(orderM))-length(coef(orderMs)))
```

```
## [1] 0.00625561
```

```
1 - pchisq(orderMe$deviance - orderM$deviance, length(coef(orderM))-length(coef(orderMe)))
```

```
## [1] 0.3204756
```

The p-value of the test between orderMs and orderM < 0.05, hence we reject the null and include variable life events in our model. The likelihood ratio test between orderME and orderM gives us a p-value > 0.05, hence we fail to reject the null hypothesis so we will exclude the variable SES. As a result, orderMe is preferred. However, since orderM gives us a smaller AIC and similar low BIC, I would like to continue analyzing orderM.

```
coefficients(orderM)
```

```
##        SES     Events
##  0.5919468 -0.3179840
```

```
orderM$zeta
```

```
## Impaired|Mild Mild|Moderate Moderate|Well
##   -2.42133194   -0.89349539   -0.05552578
```

Our model is

$$logit[P(impairment = Well)|Events, SES] = -0.282 - 0.319Events - 1.111SES$$

$$logit[P(impairment \in \{Well, Mild\})|Events, SES] = 1.213 - 0.319Events - 1.111SES$$

$$logit[P(impairment = \in \{Well, Mild, Moderate\})|Events, SES] = 2.209 - 0.319Events - 1.111SES$$

Since events = 7, and SES = 0

$$logit[P(impairment = Well)|Events = 7, SES = 0] = -2.514$$
$$logit[P(impairment \in \{Well, Mild\})|Events = 7, SES = 0] = -1.02$$
$$logit[P(impairment = \in \{Well, Mild, Moderate\})|Events = 7, SES = 0] = -0.023$$

In order to convert logit function back to probability, we need to use $logit^{-1} = \frac{e^x}{1+e^x}$ to transfer back.

```
log1 = c(orderM$zeta[1],-coef(orderM))%*%c(1,0,7)
log2 = c(orderM$zeta[2],-coef(orderM))%*%c(1,0,7)
log3 = c(orderM$zeta[3],-coef(orderM))%*%c(1,0,7)
inv <- function(x) { exp(x)/(1+exp(x)) }
prob1 = inv(log1)
prob12 = inv(log2)
prob123 = inv(log3)
prob1
```

```
##          [,1]
## [1,] 0.451294
```

```
prob12
```

```
##           [,1]
## [1,] 0.7912361
```

```
prob123
```

```
##           [,1]
## [1,] 0.8975563
```

$$P(impairment = Well|Events = 7, SES = 0) = 0.075$$
$$P(impairment \in \{Well, Mild\}|Events = 7, SES = 0) = 0.265$$
$$P(impairment \in \{Well, Mild, Moderate\}|Events = 7, SES = 0) = 0.494$$

From the above, John has a 1-0.075-0.265-0.494=0.506 probability of being mentally impaired.

Problem 2 Alligator Food Choice Data

```
attach(food)
y = cbind(Fish,Invertebrate,Reptile,Bird,Other)
M1 = multinom(y ~ 1)
M2 = multinom(y ~ Lake)
M3 = multinom(y ~ Gender)
M4 = multinom(y ~ Size)
M5 = multinom(y ~ Lake + Gender)
M6 = multinom(y ~ Lake + Size)
M7 = multinom(y ~ Gender + Size)
M8 = multinom(y ~ Lake+Gender+Size)
```

```
Mfull = multinom(y ~ Lake + Gender + Size, data = food)
Mempty = multinom(y ~ 1, data = food)
M = step(Mfull); M
```

By doing a stepwise model selection, I found that variables lake and size seem to influence alligator's food choice. And model 6 contains only lake and size with the lowest AIC 580.08.

```
mod = multinom(formula = y ~ Lake + Size)
```

From the above result, with Fish being baseline and only consider LakeTrafford, I got a model:

$$log\frac{\pi Invertebrate(Lake = LakeTrafford, size > 2.3)}{\pi Fish(Lake = LakeTrafford, size > 2.3)} = 0.007 + 1.024 - 1.458 = -0.427$$

$$log\frac{\pi Reptile(Lake = LakeTrafford, size > 2.3)}{\pi Fish(Lake = LakeTrafford, size > 2.3)} = -1.605 + 0.875 + 0.352 = -0.378$$

$$log\frac{\pi Bird(Lake = LakeTrafford, size > 2.3)}{\pi Fish(Lake = LakeTrafford, size > 2.3)} = -1.953 + 0.317 + 0.631 = -1.005$$

$$log\frac{\pi Other(Lake = LakeTrafford, size > 2.3)}{\pi Fish(Lake = LakeTrafford, size > 2.3)} = -0.789 + 0.732 - 0.331 = -0.388$$

Thus,

$$log\frac{1 - \pi Fish(Lake = LakeTrafford, size > 2.3)}{\pi Fish(Lake = LakeTrafford, size > 2.3)} = e^{-0.427} + e^{-0.378} + e^{-1.005} + e^{-0.388} = 2.38$$

$$\pi Fish(Lake = LakeTrafford, size > 2.3) = \frac{1}{1 + 2.38} = 0.296$$

Thus,

$$\pi Invertebrate(Lake = LakeTrafford, size > 2.3) = 0.296 * e^{-0.427} = 0.193$$

$$\pi Reptile(Lake = LakeTrafford, size > 2.3) = 0.296 * e^{-0.378} = 0.203$$

$$\pi Bird(Lake = LakeTrafford, size > 2.3) = 0.296 * e^{-1.005} = 0.108$$

$$\pi Other(Lake = LakeTrafford, size > 2.3) = 0.296 * e^{-0.388} = 0.201$$

Therefore the probability that a male alligator from Lake Trafford that is 3 meters long prefers reptiles for dinner is 0.203.