

Homework7

Anshuman Narayan

12/5/2018

Q1

```
mmod=multinom( response ~ treatment*gender,data=lung)
```

```
## # weights:  20 (12 variable)
## initial  value 414.502014
## iter   10 value 392.989195
## final   value 391.744479
## converged
```

Q1a

We begin reducing the model by examining the parameters in the model and testing their significance in fitting the model on the given dataset. We first begin by considering a null model that does not contain the coefficient for the interaction term between treatment and gender. We state the following hypotheses for our test: H_0 :The coefficient for the interaction terms between treatment and gender are 0 H_a :The full model, without the restrictions of H_0 holds.

```
mod.no.int=multinom(response ~ treatment + gender , data = lung)
```

```
## # weights:  16 (9 variable)
## initial  value 414.502014
## iter   10 value 393.034226
## final   value 393.017781
## converged
```

```
anova(mod.no.int,mmod)
```

| ## | Model | Resid. df | Resid. Dev | Test | Df | LR stat. | Pr(Chi) |
|------|--------------------|-----------|------------|--------|----|----------|-----------|
| ## 1 | treatment + gender | 888 | 786.0356 | | NA | NA | NA |
| ## 2 | treatment * gender | 885 | 783.4890 | 1 vs 2 | 3 | 2.546603 | 0.4669277 |

Since the p-value is larger than our significance level we can safely replace our model with the reduced model. We now look at further reducing our model by considering removing the gender term from the model. Our null hypothesis model consists of a model with the gender term coefficients being 0. Our alternative hypothesis model consists of the full model without the restrictions of the

```
mod.no.gender=multinom(response~treatment,data=lung)
```

```
## # weights:  12 (6 variable)
## initial  value 414.502014
## iter   10 value 395.641305
## final   value 395.641300
## converged
```

```
anova(mod.no.gender,mod.no.int)
```

| ## | Model | Resid. df | Resid. Dev | Test | Df | LR stat. | Pr(Chi) |
|------|--------------------|-----------|------------|--------|----|----------|-----------|
| ## 1 | treatment | 891 | 791.2826 | | NA | NA | NA |
| ## 2 | treatment + gender | 888 | 786.0356 | 1 vs 2 | 3 | 5.247038 | 0.1545762 |

Since the p-value is larger than our significance level we can safely replace our model with the reduced model.

We now look at reducing the model further. We now remove the treatment term as well i.e we reduce to a null model. our null hypothesis model is a null model where the coefficients of all terms except the intercept are 0 and the alternative model is the reduced model i.e the full model without the restrictions of the null model.

```
null.mod=multinom(response~1,data=lung)
```

```
## # weights:  8 (3 variable)
## initial  value 414.502014
## final   value 399.983982
## converged
```

```
anova(null.mod,mod.no.gender)
```

| | Model | Resid. df | Resid. Dev | Test | Df | LR stat. | Pr(Chi) |
|------|-----------|-----------|------------|--------|----|----------|------------|
| ## 1 | 1 | 894 | 799.9680 | | NA | NA | NA |
| ## 2 | treatment | 891 | 791.2826 | 1 vs 2 | 3 | 8.685365 | 0.03378026 |

```
summary(mod.no.gender)
```

```
## Call:
## multinom(formula = response ~ treatment, data = lung)
##
## Coefficients:
##              (Intercept) treatmentSequential
## no change             -0.03846754             0.6157805
## partial remission    -0.83479810             0.8954216
## complete remission  -0.92576928             0.7922358
##
## Std. Errors:
##              (Intercept) treatmentSequential
## no change             0.1961524             0.2954143
## partial remission     0.2496922             0.3507248
## complete remission    0.2578506             0.3653097
##
## Residual Deviance: 791.2826
## AIC: 803.2826
```

The p-value is lesser than the 0.05 that means that gender is significant to the model. Thus our reduced model consists of of 1 paramter that consists of just treatment.

Q1b

We predict for probabilities for a male receiving sequential therapy using the reduced model from the previous question.

```
predict(mod.no.gender,newdata=data.frame(gender='male',treatment='Sequential'),type="probs")
```

| | | |
|------------------------|-----------|-------------------|
| ## progressive disease | no change | partial remission |
| ## 0.2119209 | 0.3774831 | 0.2251657 |
| ## complete remission | | |
| ## 0.1854304 | | |

Q1c

Yes the response is ordinal since it has a natural order. The order is the following, progressive disease, no change, parital remission, complete remission. #####Q1d We first fit the prop odds model.

```
library(MASS)
pmod=polr(response ~ treatment*gender,data=lung)
summary(pmod)

##
## Re-fitting to get Hessian
## Call:
## polr(formula = response ~ treatment * gender, data = lung)
##
## Coefficients:
##
##               Value Std. Error t value
## treatmentSequential    1.0785    0.5336   2.021
## genderMale             0.8646    0.4316   2.003
## treatmentSequential:genderMale -0.5904    0.5792  -1.019
##
## Intercepts:
##
##               Value Std. Error t value
## progressive disease|no change    0.0770  0.3988   0.1931
## no change|partial remission    1.6484  0.4099   4.0211
## partial remission|complete remission 2.6978  0.4257   6.3380
##
## Residual Deviance: 788.0098
## AIC: 800.0098
```

We use the drop1() method to use the likelihood ratio test backward elimination procedure.

```
drop1(pmod,test="Chi")

## Single term deletions
##
## Model:
## response ~ treatment * gender
##           Df    AIC    LRT Pr(>Chi)
## <none>           800.01
## treatment:gender  1 799.06 1.0468  0.3062
```

the p-value for the interaction term is larger than the significant level. It is insignificant to the model. We remove it from the model and continue reducing.

```
reduced_mod=polr(response~treatment+gender,data=lung)
drop1(reduced_mod,test="Chi")
```

```
## Single term deletions
##
## Model:
## response ~ treatment + gender
##           Df    AIC    LRT Pr(>Chi)
## <none>           799.06
## treatment  1 804.63 7.5702 0.005934 **
## gender     1 800.65 3.5965 0.057901 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The gender term is insignificant to the model as it has a p-value larger than 0.05. treatment however is lesser than the significant level and thus we can keep it in the reduced model. The reduced model consists of just one parameter that i.e treatment.

Q1e

We predict for a male receiving sequential therapy using the reduced model.

```
reduced_mod=polr(response~treatment,data=lung)
predict(reduced_mod,newdata=data.frame(gender='Male',treatment='Sequential'),
        type="probs")
```

```
## progressive disease      no change    partial remission
##           0.2285142        0.3552451        0.2152800
## complete remission
##           0.2009607
```

Q2

Q2a

```
mod.s=glm(count~first*second,data=freethrow,family=poisson)
summary(mod.s)
```

```
##
## Call:
## glm(formula = count ~ first * second, family = poisson, data = freethrow)
##
## Deviance Residuals:
## [1]  0  0  0  0  0
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      5.023881   0.081111  61.939 < 2e-16 ***
## firstMissed      -1.412963   0.183319  -7.708 1.28e-14 ***
## secondMissed     -1.527373   0.192047  -7.953 1.82e-15 ***
## firstMissed:secondMissed -0.004103   0.434637  -0.009  0.992
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance:  1.9469e+02 on 3  degrees of freedom
## Residual deviance: -2.0984e-27 on 0  degrees of freedom
## AIC: 29.594
##
## Number of Fisher Scoring iterations: 3
exp(confint.default(mod.s))[4,]
```

```
##      2.5 %      97.5 %
## 0.4248685 2.3344323
```

This interval agrees with the interval generated from the model for Homework 3. #####Q2b

```
mod.i=glm(count~first+second,family=poisson,data=freethrow)
summary(mod.i)
```

```
##
## Call:
```

```
## glm(formula = count ~ first + second, family = poisson, data = freethrow)
##
## Deviance Residuals:
##      1      2      3      4
## -0.001763  0.003785  0.003575 -0.007679
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   5.02402    0.07968  63.054 <2e-16 ***
## firstMissed  -1.41369    0.16622  -8.505 <2e-16 ***
## secondMissed -1.52817    0.17228  -8.870 <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: 1.9469e+02  on 3  degrees of freedom
## Residual deviance: 8.9181e-05  on 1  degrees of freedom
## AIC: 27.594
##
## Number of Fisher Scoring iterations: 3
```

```
gsq=deviance(mod.i,1)
cat("The probabilitly the g-squared variable is",1-pchisq(gsq,1))
```

```
## The probabilitly the g-squared variable is 0.9924652
```

H₀ : The two explanatory variables are independent H_a : THe null model is false. Since the p-value is larger than the significance level, we fail to reject the the null model that the variables are independent.
 ###Q3 #####Q3a

```
mod.sat=glm(formula=count~information*health*gender,data=opinion,family=poisson)
summary(mod.sat)
```

```
##
## Call:
## glm(formula = count ~ information * health * gender, family = poisson,
##      data = opinion)
##
## Deviance Residuals:
## [1]  0  0  0  0  0  0  0  0  0
##
## Coefficients:
##              Estimate Std. Error z value
## (Intercept)   3.87120    0.14434  26.820
## informationsupport  1.32730    0.16235   8.175
## healthsupport    -1.47331    0.33428  -4.407
## gendermale       -0.65233    0.24664  -2.645
## informationsupport:healthsupport  1.01101    0.35502   2.848
## informationsupport:gendermale    0.52900    0.26946   1.963
## healthsupport:gendermale    0.04619    0.56428   0.082
## informationsupport:healthsupport:gendermale -0.32833    0.59339  -0.553
##              Pr(>|z|)
## (Intercept) < 2e-16 ***
## informationsupport  2.95e-16 ***
## healthsupport    1.05e-05 ***
```

```

## gendermale                                0.00817 **
## informationsupport:healthsupport           0.00440 **
## informationsupport:gendermale              0.04962 *
## healthsupport:gendermale                   0.93476
## informationsupport:healthsupport:gendermale 0.58005
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance:  4.4582e+02 on 7 degrees of freedom
## Residual deviance: -2.1094e-14 on 0 degrees of freedom
## AIC: 61.382
##
## Number of Fisher Scoring iterations: 3
drop1(mod.sat,test="Chi")

## Single term deletions
##
## Model:
## count ~ information * health * gender
##               Df Deviance    AIC    LRT Pr(>Chi)
## <none>                0.00000 61.382
## information:health:gender 1  0.30072 59.683 0.30072  0.5834
mod.red=update(mod.sat,~.-information:health:gender)
drop1(mod.red,test="Chi")

## Single term deletions
##
## Model:
## count ~ information + health + gender + information:health +
##         information:gender + health:gender
##               Df Deviance    AIC    LRT Pr(>Chi)
## <none>                0.3007 59.683
## information:health  1 11.6657 69.048 11.3650 0.0007484 ***
## information:gender  1  4.1267 61.509  3.8260 0.0504635 .
## health:gender       1  2.3831 59.765  2.0824 0.1490040
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
mod.red=update(mod.red,~.-health:gender)
drop1(mod.red,test="Chi")

## Single term deletions
##
## Model:
## count ~ information + health + gender + information:health +
##         information:gender
##               Df Deviance    AIC    LRT Pr(>Chi)
## <none>                2.3831 59.765
## information:health  1 13.1201 68.502 10.7369  0.00105 **
## information:gender  1  5.5810 60.963  3.1979  0.07373 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

mod.red=update(mod.red,~.-information:gender)
drop1(mod.red,test="Chi")

## Single term deletions
##
## Model:
## count ~ information + health + gender + information:health
##           Df Deviance   AIC    LRT Pr(>Chi)
## <none>           5.581 60.963
## gender           1  17.810 71.192 12.229 0.0004706 ***
## information:health 1  16.318 69.700 10.737 0.0010502 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(mod.red)

```

```

##
## Call:
## glm(formula = count ~ information + health + gender + information:health,
##      family = poisson, data = opinion)
##
## Deviance Residuals:
##      1       2       3       4       5       6       7       8
## -0.6372  1.0893 -0.4999 -1.1823  0.5421 -0.9715  0.4116  0.9662
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      3.72843    0.12212  30.531 < 2e-16 ***
## informationsupport  1.54142    0.12896  11.953 < 2e-16 ***
## healthsupport     -1.45725    0.26930  -5.411 6.26e-08 ***
## gendermale        -0.28205    0.08106  -3.480 0.000502 ***
## informationsupport:healthsupport  0.87239    0.28411   3.071 0.002136 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 445.823  on 7  degrees of freedom
## Residual deviance:   5.581  on 3  degrees of freedom
## AIC: 60.963
##
## Number of Fisher Scoring iterations: 4

```

We now test the goodness of fit of our reduced model by looking at its deviance and then looking at the probability that it follows the distribution of a chi-squared variable with 3 degrees of freedom.

```

g.sq=deviance(mod.red)
cat("The probability that this variable is following this chi-squared variable is",1-pchisq(g.sq,3))

```

```
## The probability that this variable is following this chi-squared variable is 0.1338714
```

Since the p-value is larger than the significance level we cannot reject the above reduced model.

Q3b

The association structure can be defined as follows, gender is jointly independent of health and information.