YK_Final_P5

Yinan Kang

5/14/2019

```
rm(list=ls())
df <- read.csv("/cloud/project/Question 5.csv")

require(ResourceSelection)
```

```
## Loading required package: ResourceSelection
```

```
## ResourceSelection 0.3-4   2019-01-08
```

(a) Fit model with first orders plus all interactions

```
model <- glm(Y~X1+X2+X3+X4+X1*X2+X1*X3+X1*X4+X2*X3+X2*X4+X3*X4, family="binomial",data=df)
summary(model)
```

```
##
## Call:
## glm(formula = Y ~ X1 + X2 + X3 + X4 + X1 * X2 + X1 * X3 + X1 *
##     X4 + X2 * X3 + X2 * X4 + X3 * X4, family = "binomial", data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3440  -0.9089   0.4209   0.7961   2.0514
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.551402   1.590243   0.347    0.729
## X1           0.056505   0.045956   1.230    0.219
## X2          -1.250346   0.674971  -1.852    0.064 .
## X3           0.613566   1.008733   0.608    0.543
## X4          -1.780009   1.830666  -0.972    0.331
## X1:X2        0.001105   0.011955   0.092    0.926
## X1:X3       -0.019941   0.022211  -0.898    0.369
## X1:X4        0.020090   0.025639   0.784    0.433
## X2:X3        0.177436   0.430328   0.412    0.680
## X2:X4       -0.055230   0.476060  -0.116    0.908
## X3:X4        0.908936   0.818335   1.111    0.267
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 270.06  on 195  degrees of freedom
## Residual deviance: 213.50  on 185  degrees of freedom
## AIC: 235.5
##
## Number of Fisher Scoring iterations: 5
```

(b) Likelihood Ratio Test for dropping Interaction Terms

Full Model: Pi = [1+exp(-X'BetaFull)] ^-1
where X*BetaFull* = *Beta0* + *Beta1*X1 + ... Beta10*(X3*X4)

Reduced Model: Pi = [1+exp(-X'BetaReduced)] ^-1
where X*BetaFull* = *Beta0* + *Beta1*X1 + ... Beta4*X4

Hypotheses:
H0: Beta 5 = Beta 6 .... = Beta 10 = 0 (All Beta's with Interaction terms = 0)
Ha: Not All Betas in H0 = 0

Decision Rule:
If G2 <= Chi-Sq(1-alpha, p-q), conclude H0
If G2 > Chi-Sq(1-alpha, p-1), conclude Ha

Assume alpha = 0.05

```
model.r <- glm(Y~X1+X2+X3+X4, family = "binomial", data=df)
anova(model, model.r, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: Y ~ X1 + X2 + X3 + X4 + X1 * X2 + X1 * X3 + X1 * X4 + X2 * X3 +
##     X2 * X4 + X3 * X4
## Model 2: Y ~ X1 + X2 + X3 + X4
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       185     213.50
## 2       191     215.65 -6  -2.1492   0.9055
```

Decision:

We use the 'anova' function to compare our full and reduced models. We see that the p-value associated with the Chi-Sq distribution is > 0.05.
This interpretation is different than the formal one stated in the Decision Rule above. This p-value lets us know that there is no statistical significance in adding the interactions, so they can be DROPPED.

(c) Backward elimination

```
step(model, scope = formula(model), direction="backward")
```

```
## Start:  AIC=235.5
## Y ~ X1 + X2 + X3 + X4 + X1 * X2 + X1 * X3 + X1 * X4 + X2 * X3 +
##     X2 * X4 + X3 * X4
##
##         Df Deviance    AIC
## - X1:X2  1   213.51 233.51
## - X2:X4  1   213.52 233.52
## - X2:X3  1   213.67 233.67
## - X1:X4  1   214.14 234.14
## - X1:X3  1   214.30 234.30
## - X3:X4  1   214.75 234.75
## <none>       213.50 235.50
##
## Step:  AIC=233.51
## Y ~ X1 + X2 + X3 + X4 + X1:X3 + X1:X4 + X2:X3 + X2:X4 + X3:X4
```

```
##
##         Df Deviance    AIC
## - X2:X4  1   213.52 231.52
## - X2:X3  1   213.68 231.68
## - X1:X4  1   214.15 232.15
## - X1:X3  1   214.40 232.40
## - X3:X4  1   214.78 232.78
## <none>       213.51 233.51
##
## Step:  AIC=231.52
## Y ~ X1 + X2 + X3 + X4 + X1:X3 + X1:X4 + X2:X3 + X3:X4
##
##         Df Deviance    AIC
## - X2:X3  1   213.68 229.68
## - X1:X4  1   214.15 230.15
## - X1:X3  1   214.41 230.41
## - X3:X4  1   214.84 230.84
## <none>       213.52 231.52
##
## Step:  AIC=229.68
## Y ~ X1 + X2 + X3 + X4 + X1:X3 + X1:X4 + X3:X4
##
##         Df Deviance    AIC
## - X1:X4  1   214.32 228.32
## - X1:X3  1   214.53 228.53
## - X3:X4  1   215.04 229.04
## <none>       213.68 229.68
## - X2     1   240.51 254.51
##
## Step:  AIC=228.32
## Y ~ X1 + X2 + X3 + X4 + X1:X3 + X3:X4
##
##         Df Deviance    AIC
## - X1:X3  1   214.81 226.81
## - X3:X4  1   215.44 227.44
## <none>       214.32 228.32
## - X2     1   240.74 252.74
##
## Step:  AIC=226.81
## Y ~ X1 + X2 + X3 + X4 + X3:X4
##
##         Df Deviance    AIC
## - X3:X4  1   215.65 225.65
## <none>       214.81 226.81
## - X1     1   229.00 239.00
## - X2     1   240.83 250.83
##
## Step:  AIC=225.65
## Y ~ X1 + X2 + X3 + X4
##
##        Df Deviance    AIC
## - X4    1   215.66 223.66
## <none>      215.65 225.65
## - X3    1   220.39 228.39
```

```
## - X1    1   230.16 238.16
## - X2    1   241.92 249.92
##
## Step:  AIC=223.66
## Y ~ X1 + X2 + X3
##
##        Df Deviance    AIC
## <none>      215.66 223.66
## - X3    1   220.67 226.67
## - X1    1   230.85 236.85
## - X2    1   242.00 248.00
```

```
##
## Call:  glm(formula = Y ~ X1 + X2 + X3, family = "binomial", data = df)
##
## Coefficients:
## (Intercept)           X1           X2           X3
##     0.20085      0.03584     -0.97722      0.77005
##
## Degrees of Freedom: 195 Total (i.e. Null);  192 Residual
## Null Deviance:       270.1
## Residual Deviance: 215.7     AIC: 223.7
```

Analysis: Using Backward step, we see that X1, X2, X3 are retained.

Re-fitting model for rest of problem !!

```
model <- glm(Y~X1+X2+X3, family="binomial", data=df)
```

(d) Hoslem-Lameshow Goodness-of-Fit

Hypotheses:

H0: Logistic response is appropriate
Ha: Logistic Response is not appropriate

Alternatives:

If X.test <= Chi-Sq(0.95,c-2), conclude H0
If X.test > Chi-Sq(0.95,c-2), conclude Ha

```
hoslem.test(model$y,fitted(model),g=5)
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  model$y, fitted(model)
## X-squared = 1.223, df = 3, p-value = 0.7475
```

Decision: As X.test < Chi-Sq, we conclude H0, that a Logistic response is appropriate.

(e) Prediction

```
pred.data <- data.frame(X1=c(33,6), X2 =c(1,1), X3 = c(1,1), X4 = c(0,0))

pred <- predict(model, newdata=pred.data)
print(pred)
```

```
##         1         2
## 1.1764354 0.2087265
```

```
# Ran out of time to do joint confidence interval
```