Discrete Response Model
Lecture 3

datascience@berkeley

Variable Transformation, Part 1: Interactions Among Explanatory Variables–An Example

Example

A specification with distance, wind, and the distance × wind interaction:

```
# Load the data
setwd("/Users/jeffrey/Documents/JStuff/AdvStat/pgms/CatData/Chapter2")
#list.files("/Users/jeffrey/Documents/JStuff/AdvStat/pgms/CatData/Chapter2")

placekick<-read.table(file = "placekick.csv", header = TRUE, sep = ",")
str(placekick)
head(placekick)

# Estimate a GLM() model with distance and wind interaction
mod.fit.Ha<-glm(formula = good ~ distance + wind +
    distance:wind, family = binomial(link = logit), data =
    placekick)
summary(mod.fit.Ha)
```

Example

```
Call:
glm(formula = good ~ distance + wind + distance:wind, family = binomial(link = logit),
    data = placekick)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7291   0.2465   0.2465   0.3791   1.8647

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    5.684181   0.335962  16.919   <2e-16 ***
distance      -0.110253   0.008603 -12.816   <2e-16 ***
wind           2.469975   1.662144   1.486   0.1373
distance:wind -0.083735   0.043301  -1.934   0.0531 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1013.43  on 1424  degrees of freedom
Residual deviance:  767.42  on 1421  degrees of freedom
AIC: 775.42

Number of Fisher Scoring iterations: 6
```

Example (Log-Likehilood Ratio Test)

- To perform a LRT, we can fit the model under the null hypothesis and then use the anova() function:

```
# Likelihood-Ratio Test
mod.fit.Ho<-glm(formula = good ~ distance + wind, family
    = binomial(link = logit), data = placekick)
anova(mod.fit.Ho, mod.fit.Ha, test = "Chisq")
```

```
> mod.fit.Ho<-glm(formula = good ~ distance + wind, family
+     = binomial(link = logit), data = placekick)
> anova(mod.fit.Ho, mod.fit.Ha, test = "Chisq")
Analysis of Deviance Table

Model 1: good ~ distance + wind
Model 2: good ~ distance + wind + distance:wind
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      1422     772.53
2      1421     767.42  1   5.1097  0.02379 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The test statistic is $-2\log(\Lambda) = 5.1097$, and the p-value is 0.0238, which provides an empirical evidence of a distance and wind interaction.

Berkeley SCHOOL OF INFORMATION