Generalized-Linear-Model_GLM_.R

Shraddha Somani

```r
# Use the Faraway gala data to first a Poisson regresssion model to Species
using all the predictors.
# Compare the coefficients form this model to the Box-Cox Tranformed Normal
Errors Linear Model that we used in the class notes.
require(faraway)

## Loading required package: faraway

require(arm)

## Loading required package: arm

## Loading required package: MASS

## Loading required package: Matrix

## Loading required package: lme4

##
## arm (Version 1.9-1, built: 2016-8-21)

## Working directory is C:/Users/Shraddha Somani/Desktop/GitHub

##
## Attaching package: 'arm'

## The following objects are masked from 'package:faraway':
##
##     fround, logit, pfround

require(car)

## Loading required package: car

##
## Attaching package: 'car'

## The following object is masked from 'package:arm':
##
##     logit

## The following objects are masked from 'package:faraway':
##
##     logit, vif

require(ggplot2)
```

```
## Loading required package: ggplot2

head(gala)

##             Species Endemics  Area Elevation Nearest Scruz Adjacent
## Baltra           58       23 25.09       346     0.6   0.6     1.84
## Bartolome        31       21  1.24       109     0.6  26.3   572.33
## Caldwell          3        3  0.21       114     2.8  58.7     0.78
## Champion         25        9  0.10        46     1.9  47.4     0.18
## Coamano           2        1  0.05        77     1.9   1.9   903.82
## Daphne.Major     18       11  0.34       119     8.0   8.0     1.84

pois.mod = glm(Species ~ ., data = gala, family = poisson(link = "log"))
display(pois.mod)

## glm(formula = Species ~ ., family = poisson(link = "log"), data = gala)
##             coef.est coef.se
## (Intercept) 2.83     0.06
## Endemics    0.03     0.00
## Area        0.00     0.00
## Elevation   0.00     0.00
## Nearest     0.01     0.00
## Scruz       0.00     0.00
## Adjacent    0.00     0.00
## ---
##   n = 30, k = 7
##   residual deviance = 313.4, null deviance = 3510.7 (difference = 3197.4)

g = lm(Species ~ ., data = gala)
lambda = powerTransform(g)
lambda

## Estimated transformation parameters
##        Y1
## 0.5803311

lam = lambda$lambda
glam = lm(Species^lam ~ ., gala)
modlam = fortify(glam)
p1 = qplot(sample = scale(.resid), data = modlam) + geom_abline(intercept =
0, slope = 1, color = "red") + labs(title = "Normal QQ-Plot", y = "Residuals
Box-Cox-Transform")
p1
```
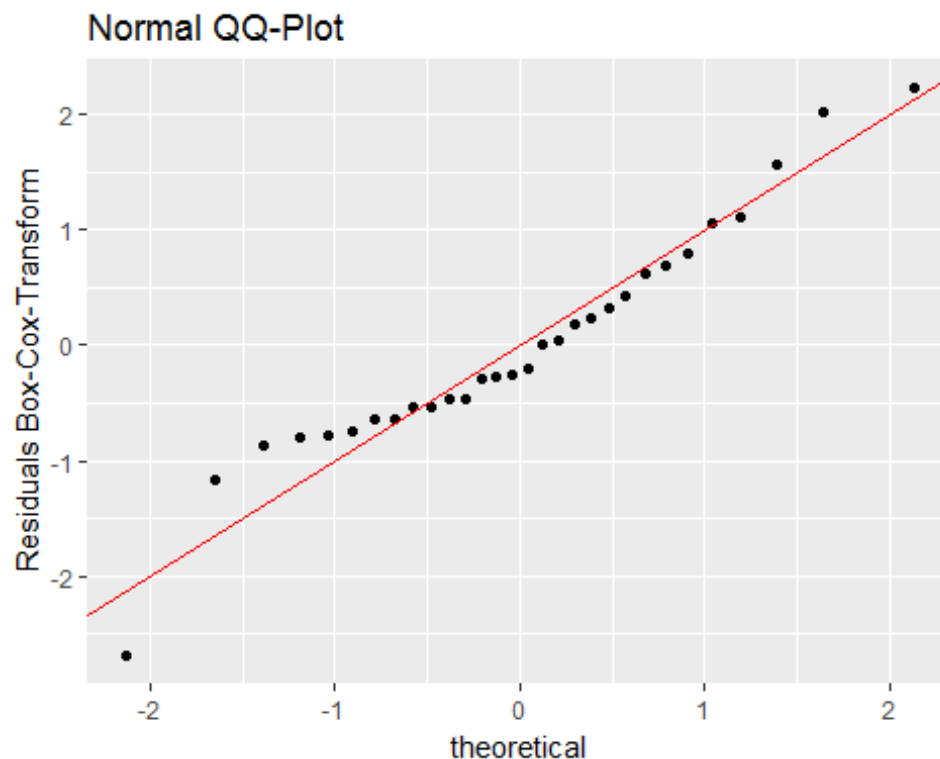
Normal QQ-Plot


```r
summary(glam)
```

```
## 
## Call:
## lm(formula = Species^lam ~ ., data = gala)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6863 -1.0684 -0.3981  1.0026  3.8970
## 
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.3931431  0.6368955   3.758  0.00103 **
## Endemics     0.3350357  0.0325224  10.302 4.35e-10 ***
## Area        -0.0002086  0.0007707  -0.271  0.78907
## Elevation   -0.0008701  0.0032168  -0.270  0.78920
## Nearest      0.0248540  0.0338516   0.734  0.47024
## Scruz       -0.0028778  0.0071562  -0.402  0.69130
## Adjacent     0.0001690  0.0008028   0.210  0.83516
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.958 on 23 degrees of freedom
## Multiple R-squared:  0.9623, Adjusted R-squared:  0.9525
## F-statistic: 97.87 on 6 and 23 DF,  p-value: 3.33e-15
```

```r
compareCoefs(pois.mod, glam)
```

```
## 
## Call:
## 1: glm(formula = Species ~ ., family = poisson(link = "log"), data =
##    gala)
## 2: lm(formula = Species^lam ~ ., data = gala)
##               Est. 1       SE 1     Est. 2      SE 2
## (Intercept)  2.83e+00  5.96e-02  2.39e+00  6.37e-01
## Endemics     3.39e-02  1.74e-03  3.35e-01  3.25e-02
## Area        -1.07e-04  3.74e-05 -2.09e-04  7.71e-04
## Elevation    2.64e-04  1.93e-04 -8.70e-04  3.22e-03
## Nearest      1.05e-02  1.61e-03  2.49e-02  3.39e-02
## Scruz       -6.83e-04  5.80e-04 -2.88e-03  7.16e-03
## Adjacent     4.54e-05  4.80e-05  1.69e-04  8.03e-04
```

```
# Fit the fpe data using a GLM-Poisson model.
# Compare coefficients form this model to the Weighted Least Squares Normal
Errors Linear Model that we used in the class notes.
head(fpe)
```

```
##                      EI   A   B   C   D  E  F  G  H  J K  A2  B2  N
## Ain                 260  51  64  36  23  9  5  4  4  3 3 105 114 17
## Alpes                75  14  17   9   9  3  1  2  1  1 1  32  31  5
## Ariege              107  27  18  13  17  2  2  2  1  1 1  57  33  6
## Bouches.du.Rhone   1036 191 204 119 205 29 13 13 10 10 6 466 364 30
## Charente.Maritime   367  71  76  47  37  8 34  5  4  4 2 163 142 17
## Cotes.du.Nord       396  93  90  57  54 13  5  9  4  3 5 193 155 15
```

```
pois.mod = glm(EI ~ ., data = fpe, family = poisson(link = "log"))
display(pois.mod)
```

```
## glm(formula = EI ~ ., family = poisson(link = "log"), data = fpe)
##             coef.est coef.se
## (Intercept)  4.55     0.05
## A           -0.04     0.01
## B            0.00     0.01
## C            0.01     0.01
## D           -0.04     0.01
## E            0.01     0.02
## F           -0.02     0.01
## G            0.00     0.02
## H           -0.07     0.03
## J            0.02     0.04
## K           -0.13     0.02
## A2           0.04     0.01
## B2           0.00     0.01
## ---
##   n = 24, k = 13
##   residual deviance = 113.2, null deviance = 3389.1 (difference = 3275.9)
```