

# Generalized Linear Models (GLMs)

## Contents

<b>1</b>	<b>Logistic Regression</b>	<b>2</b>
<b>2</b>	<b>Possion Regression</b>	<b>5</b>
<b>3</b>	<b>References</b>	<b>6</b>

```
library(COUNT)

## Loading required package: msme
## Loading required package: MASS
## Loading required package: lattice
## Loading required package: sandwich

library(ISLR)
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

library(ROCR)
```

## 1 Logistic Regression

```
## fitting the logistic regression model
data("titanic")
titanic_glm = glm(survived ~ class + age + sex, family = binomial, data = titanic)
summary(titanic_glm)

##
## Call:
## glm(formula = survived ~ class + age + sex, family = binomial,
##      data = titanic)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0652  -0.6718  -0.4740   0.7930   2.1175
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    3.0619     0.2980  10.275 < 2e-16 ***
## class2nd class -1.0106     0.1949  -5.184 2.17e-07 ***
## class3rd class -1.7664     0.1707 -10.347 < 2e-16 ***
## ageadults     -1.0556     0.2427  -4.350 1.36e-05 ***
## sexman        -2.3695     0.1453 -16.313 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1746.8  on 1315  degrees of freedom
## Residual deviance: 1276.2  on 1311  degrees of freedom
## AIC: 1286.2
```

```
##
## Number of Fisher Scoring iterations: 4
# coefficients and exp(coef) denoting odds ratios
coef(titanic_glm)

##      (Intercept) class2nd class class3rd class      ageadults      sexman
##      3.061882    -1.010558    -1.766372    -1.055608    -2.369465

exp(coef(titanic_glm))

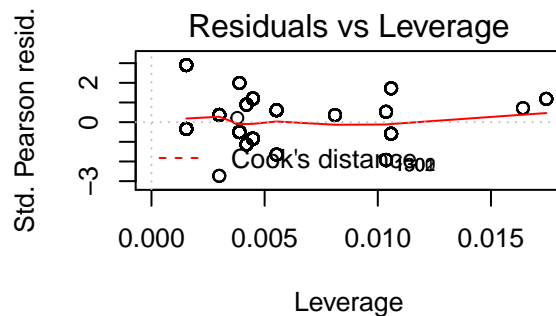
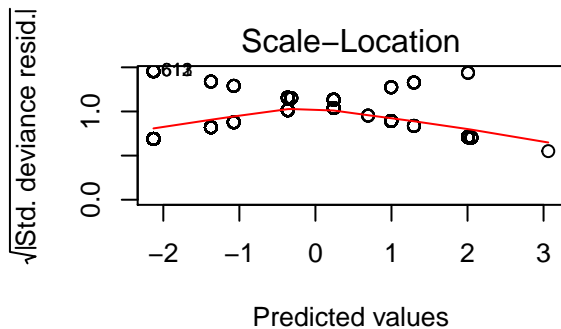
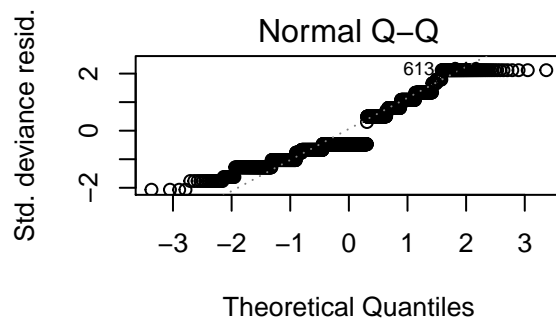
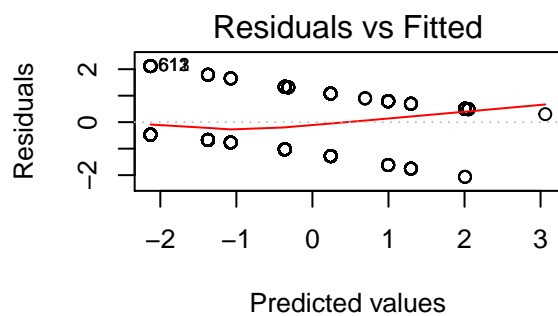
##      (Intercept) class2nd class class3rd class      ageadults      sexman
##      21.36772532    0.36401585    0.17095216    0.34798085    0.09353076

# confidenc interval for exp(coefficients)
confint(titanic_glm)

## Waiting for profiling to be done...

##              2.5 %      97.5 %
## (Intercept)    2.485799  3.6551482
## class2nd class -1.395694 -0.6310028
## class3rd class -2.104643 -1.4349492
## ageadults      -1.532907 -0.5804285
## sexman         -2.658198 -2.0884581

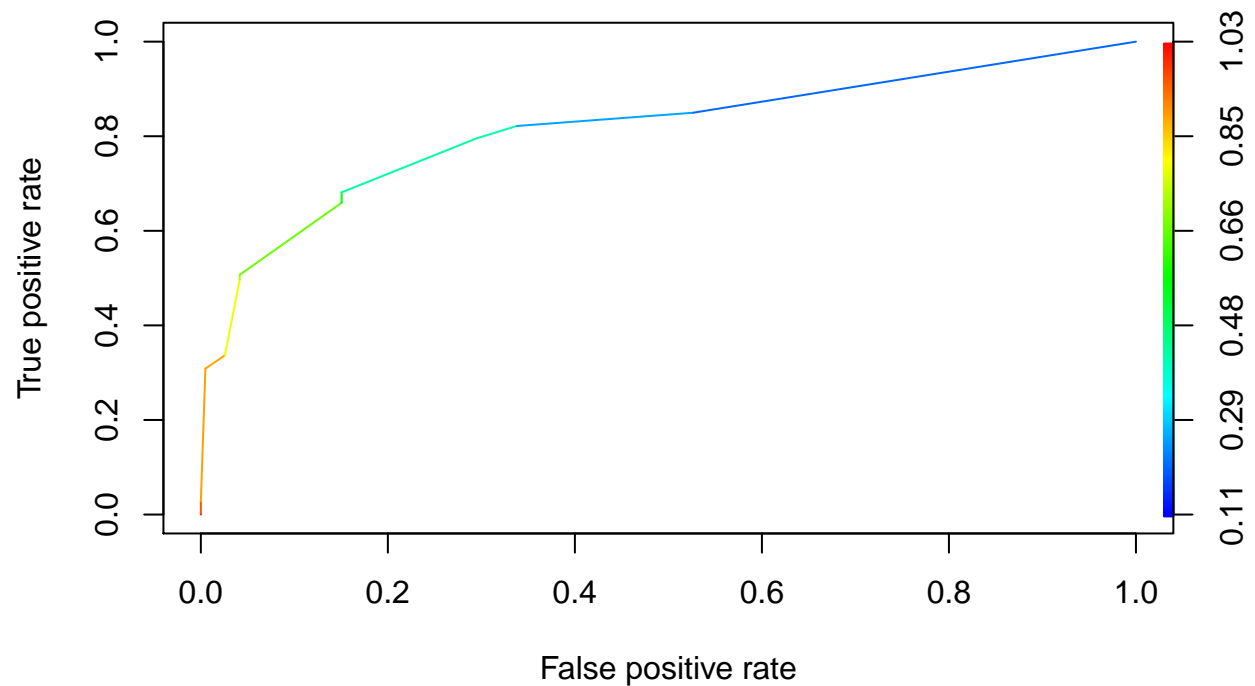
# plotting the model
par(mfrow = c(2,2))
plot(titanic_glm)
```



```
# make predictions
new_person = data.frame(class = "1st class", age = "child", sex = "man")
predict(titanic_glm, newdata = new_person, type = "link")
```

```
##          1
## 0.6924167
```

```
# ROC curve
par(mfrow = c(1,1))
titanic_pHat = predict(titanic_glm, type = 'response')
titanic_prediction = prediction(titanic_pHat, titanic$survived)
titanic_performance = performance(titanic_prediction,
measure = "tpr",
x.measure = "fpr")
plot(titanic_performance, colorize = TRUE)
```



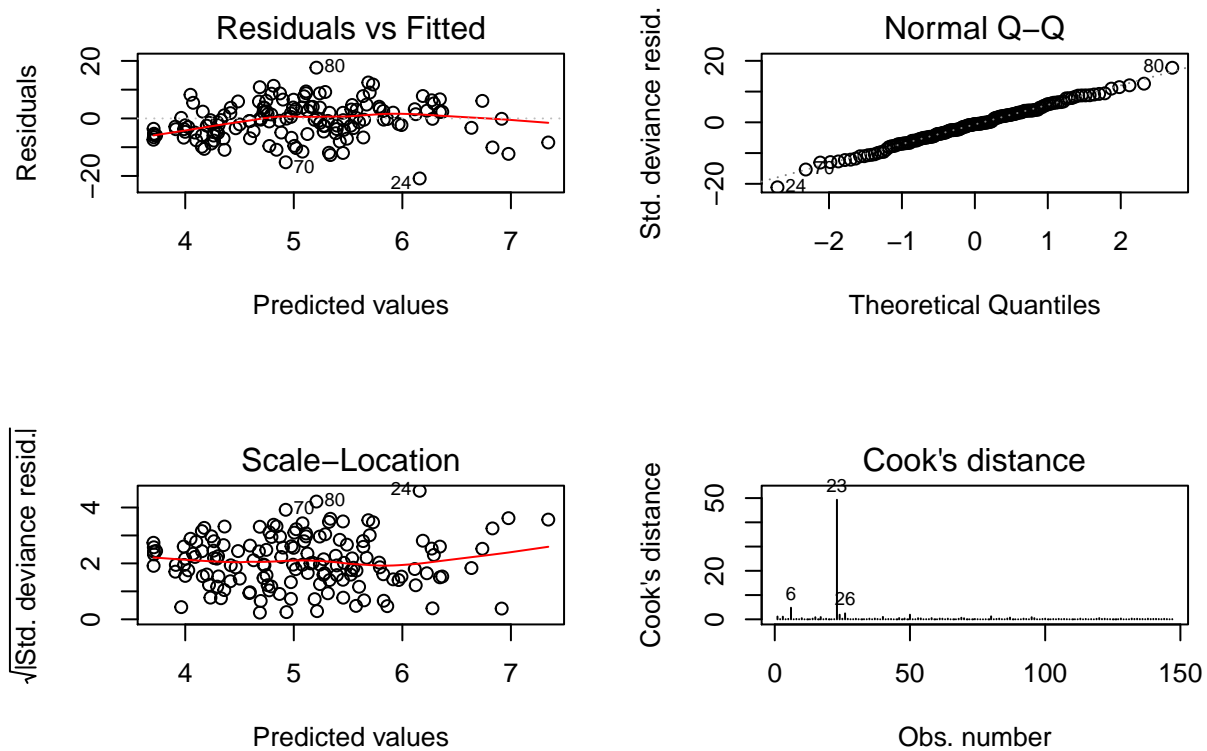
```
# AUC
performance(titanic_prediction, "auc")@y.values
```

```
## [[1]]
## [1] 0.8146035
```

## 2 Poisson Regression

```
data("fishing")
fish_dat = fishing
fish_glm = glm(totabund ~ meandepth + density + period,
               family = poisson, data = fish_dat)
summary(fish_glm)

##
## Call:
## glm(formula = totabund ~ meandepth + density + period, family = poisson,
##      data = fish_dat)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -20.8706  -4.9145  -0.5943   3.5663  17.6465
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    5.548e+00  1.915e-02  289.68  <2e-16 ***
## meandepth      -3.239e-04  7.423e-06  -43.64  <2e-16 ***
## density        7.984e+01  7.900e-01  101.07  <2e-16 ***
## period2000-2002 -2.842e-01  1.310e-02  -21.69  <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: 28049.8  on 146  degrees of freedom
## Residual deviance:  5873.3  on 143  degrees of freedom
## AIC: 6853.9
##
## Number of Fisher Scoring iterations: 5
par(mfrow = c(2,2))
plot(fish_glm, 1:4)
```



### 3 References

<https://www.datacamp.com/community/tutorials/generalized-linear-models>

Gareth James, Daniela Witten, Trevor Hastie Robert Tibshirani (2013), An Introduction to Statistical Learning with Applications in R.