Data 621 Blog 5

Bryan Persaud

12/20/2020

Generalized Linear Models

For my fifth blog I will be talking about generalized linear models. Generalized linear models(GLMs) consist of two factors, an exponential family of distributions and a link function.

Load Dataset

To demonstrate GLMs I will be using the Boston dataset.

```
library(MASS)
```

```
head(Boston)
```

```
##      crim zn indus chas   nox    rm  age    dis rad tax ptratio  black lstat
## 1 0.00632 18  2.31    0 0.538 6.575 65.2 4.0900   1 296    15.3 396.90  4.98
## 2 0.02731  0  7.07    0 0.469 6.421 78.9 4.9671   2 242    17.8 396.90  9.14
## 3 0.02729  0  7.07    0 0.469 7.185 61.1 4.9671   2 242    17.8 392.83  4.03
## 4 0.03237  0  2.18    0 0.458 6.998 45.8 6.0622   3 222    18.7 394.63  2.94
## 5 0.06905  0  2.18    0 0.458 7.147 54.2 6.0622   3 222    18.7 396.90  5.33
## 6 0.02985  0  2.18    0 0.458 6.430 58.7 6.0622   3 222    18.7 394.12  5.21
##   medv
## 1 24.0
## 2 21.6
## 3 34.7
## 4 33.4
## 5 36.2
## 6 28.7
```

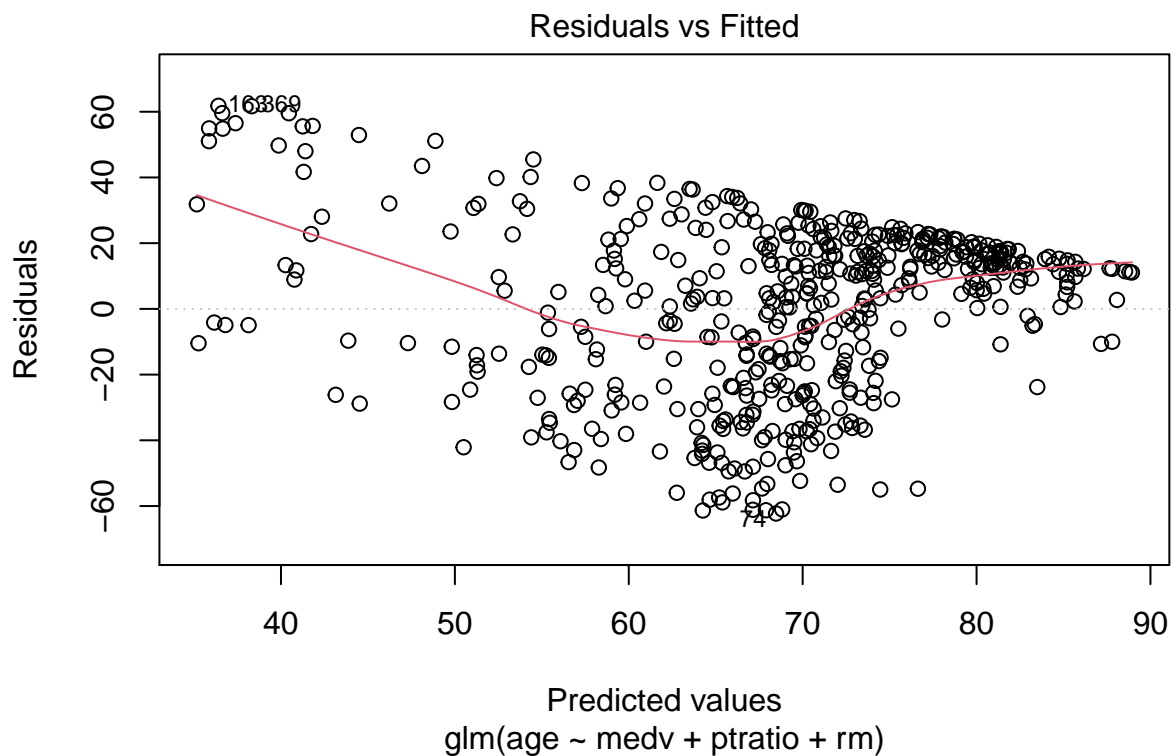## GLM

To fit the model the glm function is used.

```
model <- glm(age ~ medv + ptratio + rm, family = gaussian, data = Boston)
model
```
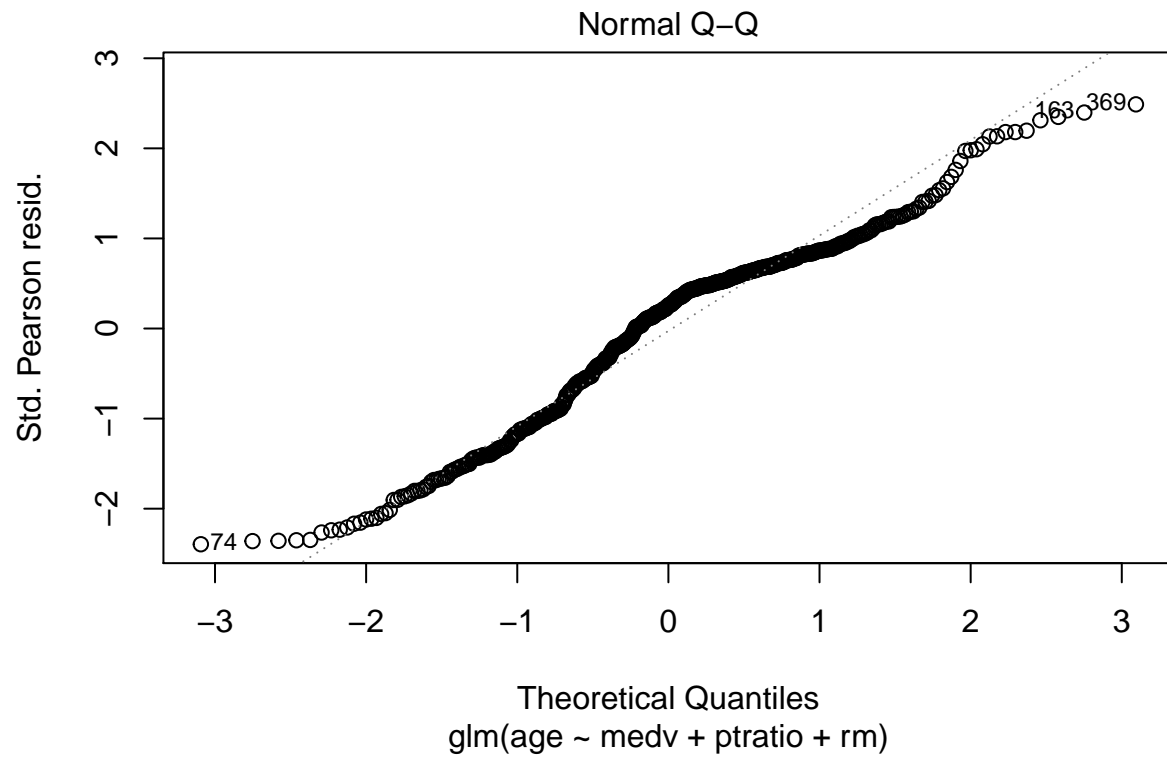
```
## 
## Call:  glm(formula = age ~ medv + ptratio + rm, family = gaussian, data = Boston)
## 
## Coefficients:
## (Intercept)          medv       ptratio            rm
##      59.845        -1.098         1.230         1.713
## 
## Degrees of Freedom: 505 Total (i.e. Null);   502 Residual
## Null Deviance:        400100
## Residual Deviance: 340300     AIC: 4740
```
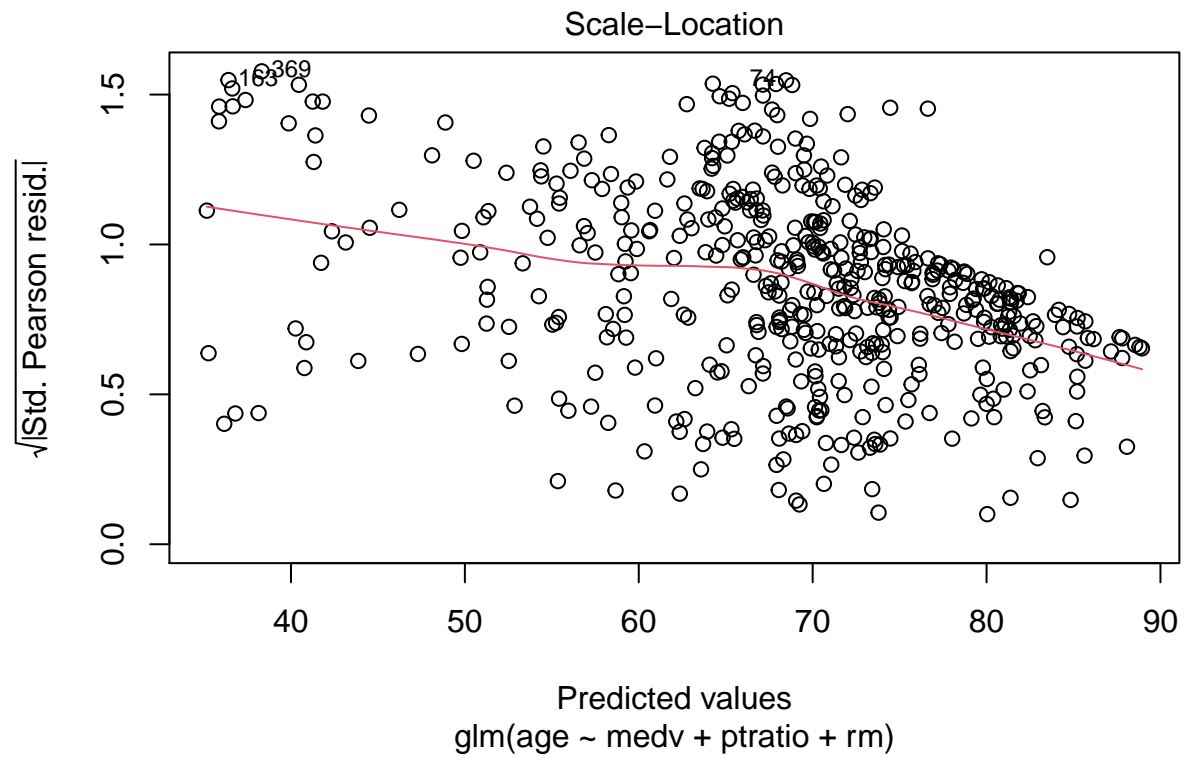
Here the glm function is made up of formula, family, link, and data. The formula is created based on the variables you use. The family is the type of distribution used, in this case gaussian is used. The family type has a default link function, gaussian's link function is "identity". The data used is the boston dataset.
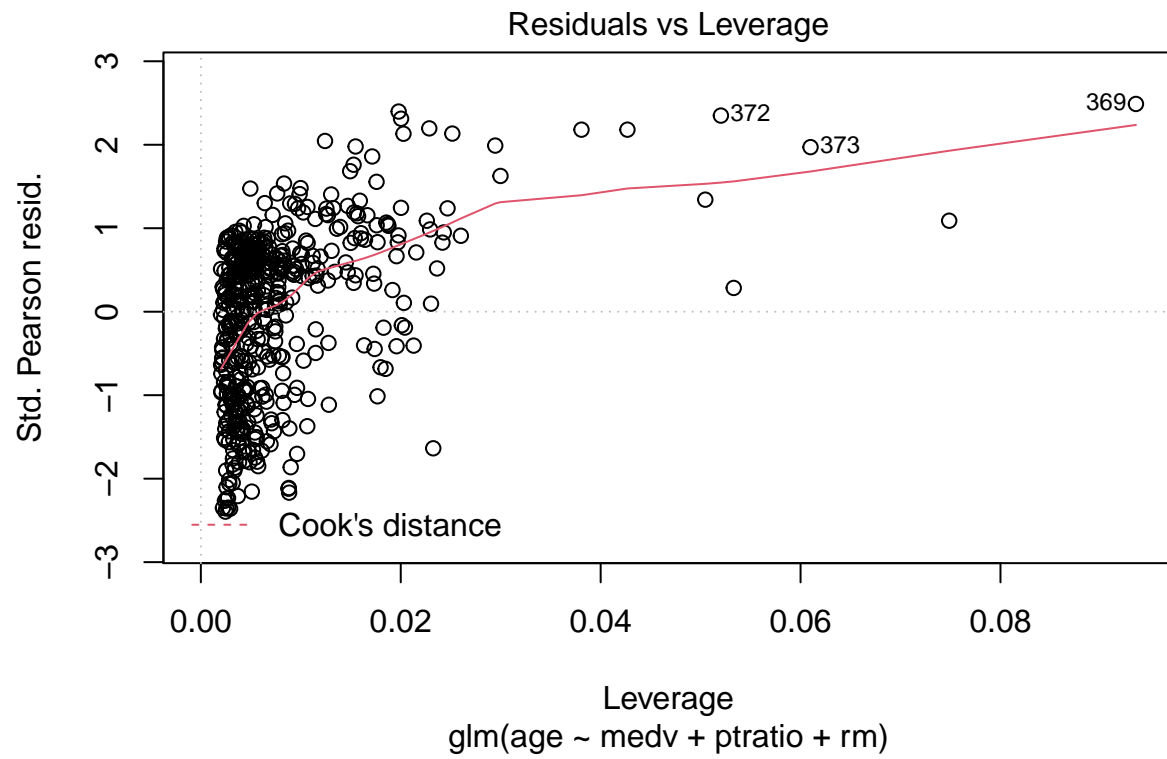
Plot

```
plot(model)
```

Normal Q–Q

Theoretical Quantiles
glm(age ~ medv + ptratio + rm)

Scale−Location

glm(age ~ medv + ptratio + rm)

Predicted values

Residuals vs Leverage

glm(age ~ medv + ptratio + rm)

When you plot the model you get different graphs based on the residuals and a Q-Q plot.