Data 621 Blog 4

Bryan Persaud

11/13/2020

Beta Regression

For my fourth blog I will be going over beta regression. Beta regression is mostly used when you have a dependent variable that fall in the (0,1) interval.

Load Dataset

To demonstrate beta regression I will be using the betareg package. Within the package also has the gasoline yield dataset.

```r
library(betareg)
```

```r
data("GasolineYield")
head(GasolineYield, 32)
```

```
##    yield gravity pressure temp10 temp batch
## 1  0.122    50.8      8.6    190  205     1
## 2  0.223    50.8      8.6    190  275     1
## 3  0.347    50.8      8.6    190  345     1
## 4  0.457    50.8      8.6    190  407     1
## 5  0.080    40.8      3.5    210  218     2
## 6  0.131    40.8      3.5    210  273     2
## 7  0.266    40.8      3.5    210  347     2
## 8  0.074    40.0      6.1    217  212     3
## 9  0.182    40.0      6.1    217  272     3
## 10 0.304    40.0      6.1    217  340     3
## 11 0.069    38.4      6.1    220  235     4
## 12 0.152    38.4      6.1    220  300     4
## 13 0.260    38.4      6.1    220  365     4
## 14 0.336    38.4      6.1    220  410     4
## 15 0.144    40.3      4.8    231  307     5
## 16 0.268    40.3      4.8    231  367     5
## 17 0.349    40.3      4.8    231  395     5
## 18 0.100    32.2      5.2    236  267     6
## 19 0.248    32.2      5.2    236  360     6
## 20 0.317    32.2      5.2    236  402     6
## 21 0.028    41.3      1.8    267  235     7
## 22 0.064    41.3      1.8    267  275     7
```

```
## 23 0.161    41.3    1.8    267  358    7
## 24 0.278    41.3    1.8    267  416    7
## 25 0.050    38.1    1.2    274  285    8
## 26 0.176    38.1    1.2    274  365    8
## 27 0.321    38.1    1.2    274  444    8
## 28 0.140    32.2    2.4    284  351    9
## 29 0.232    32.2    2.4    284  424    9
## 30 0.085    31.8    0.2    316  365   10
## 31 0.147    31.8    0.2    316  379   10
## 32 0.180    31.8    0.2    316  428   10
```

Model

To create the model we will be using the betareg function. The variable we will be looking at is yield with two explanatory variables, temp and pressure.
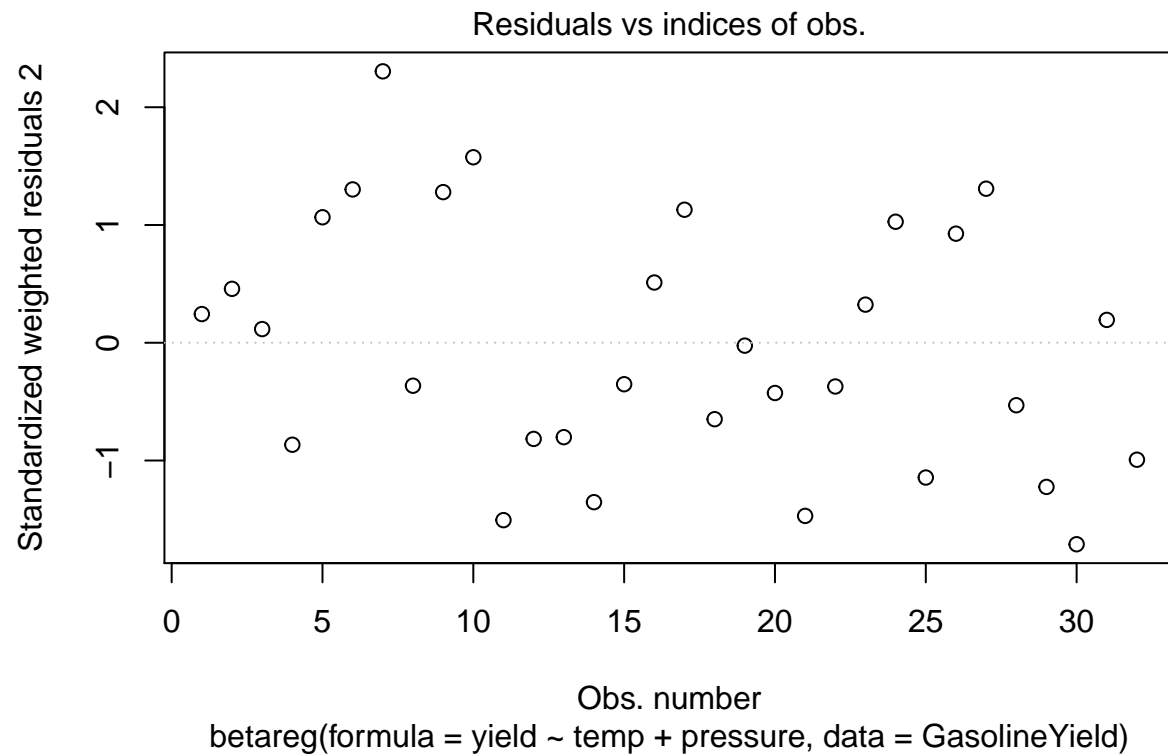
```
model <- betareg(yield ~ temp + pressure, data = GasolineYield)
summary(model)
```
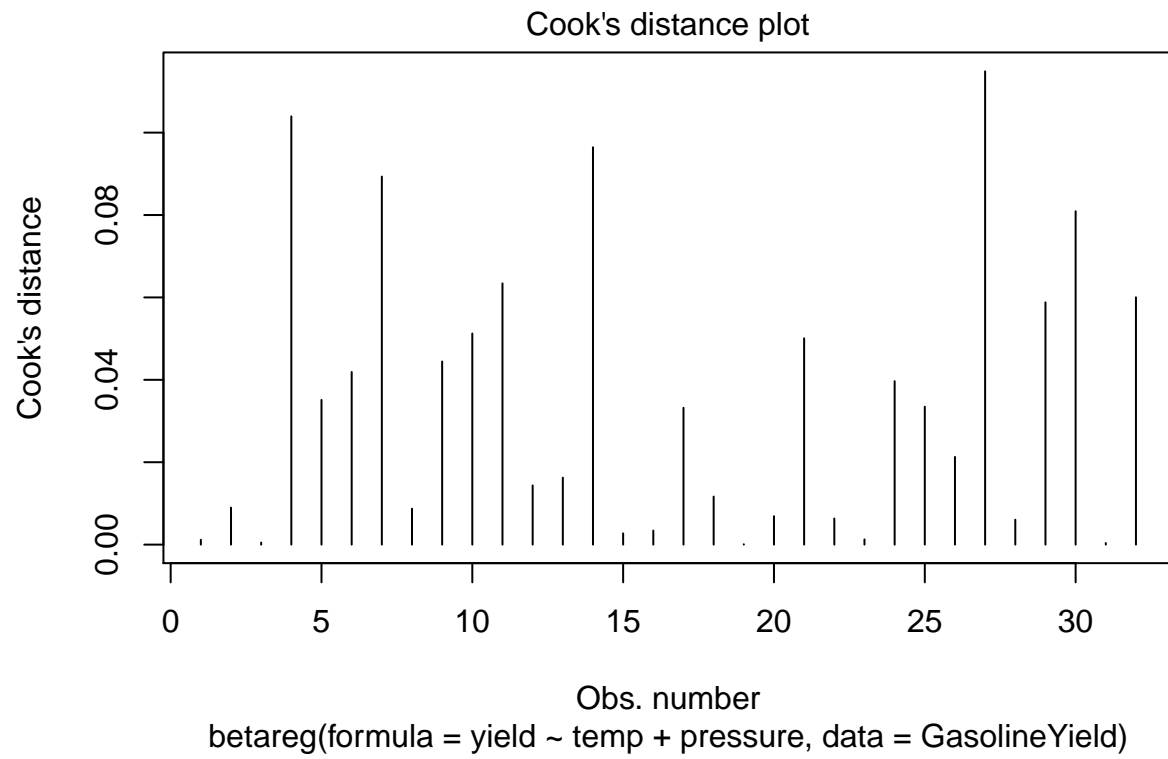
```
##
## Call:
## betareg(formula = yield ~ temp + pressure, data = GasolineYield)
##
## Standardized weighted residuals 2:
##     Min      1Q  Median      3Q     Max
## -1.7109 -0.8289 -0.1883  0.9519  2.3047
##
## Coefficients (mean model with logit link):
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.4993819  0.2717802  -20.23   <2e-16 ***
## temp         0.0097150  0.0006717   14.46   <2e-16 ***
## pressure     0.1745610  0.0160964   10.85   <2e-16 ***
##
## Phi coefficients (precision model with identity link):
##       Estimate Std. Error z value Pr(>|z|)
## (phi)   131.06      32.72   4.005 6.19e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Type of estimator: ML (maximum likelihood)
## Log-likelihood: 65.43 on 4 Df
## Pseudo R-squared: 0.8921
## Number of iterations: 28 (BFGS) + 5 (Fisher scoring)
```
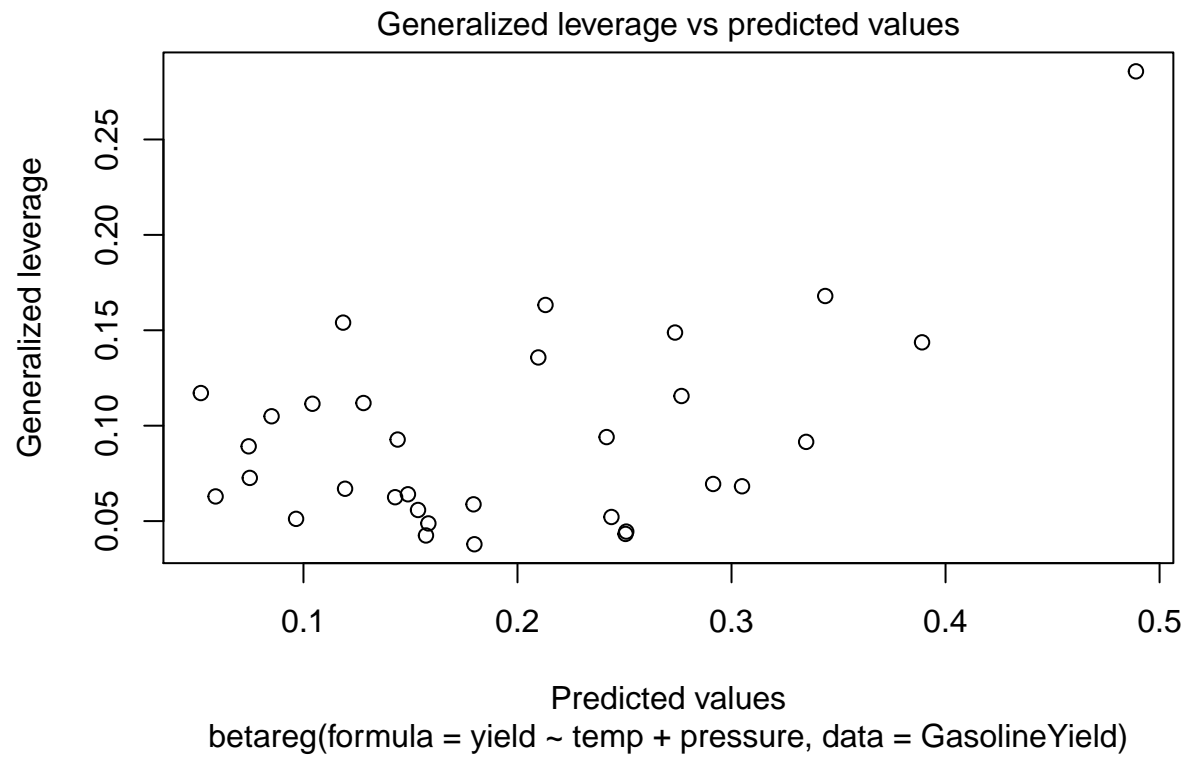
Here we see the information of the model created with the variable yield with temp and pressure. We see that there is a precision model in the data and a pseudo R-squared of 0.8921.

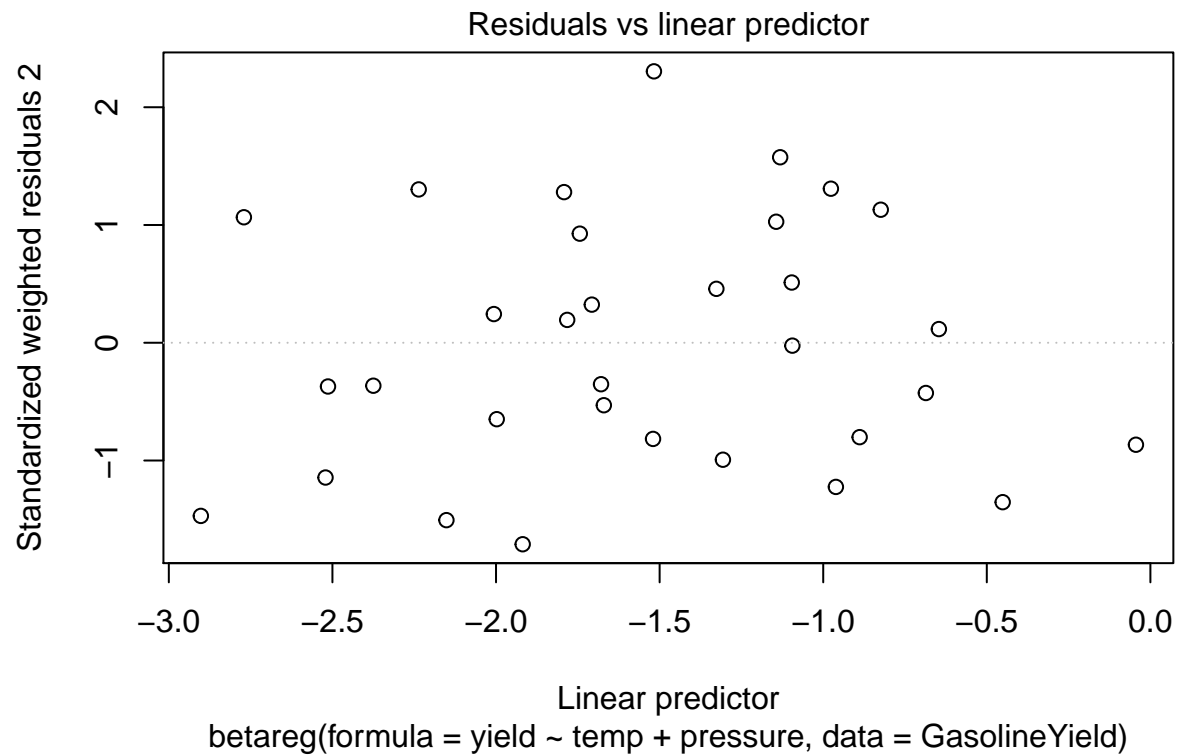Let's plot the model

```
plot(model)
```

Residuals vs indices of obs.

Obs. number
betareg(formula = yield ~ temp + pressure, data = GasolineYield)

3

Cook's distance plot

Obs. number
betareg(formula = yield ~ temp + pressure, data = GasolineYield)

Generalized leverage vs predicted values

Generalized leverage

Predicted values
betareg(formula = yield ~ temp + pressure, data = GasolineYield)

Residuals vs linear predictor

betareg(formula = yield ~ temp + pressure, data = GasolineYield)

When you plot the model you get different graphs, including a graph on Cook's distance and graphs on the residuals and leverage and predicted values.