Linear_Regression.R

Fiona

Fri Oct 13 15:40:34 2017

```
# --- Simple Linear Regression ---
# Import data
data <- read.csv("brewdog.csv", header=TRUE)

# Plot data
plot(data$ABV, data$Price)
text(data$ABV, data$Price, labels=data$Name, cex=0.7, pos=2)

# Use linear regression
fit <- lm(Price ~ ABV, data=data)

# Add regression line
abline(fit)

# Extract coefficient
coef(fit)
```

```
## (Intercept)         ABV
##  -2.7706730   0.9236609
```

```
# Show regression output
summary(fit)
```

```
##
## Call:
## lm(formula = Price ~ ABV, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3633 -1.4644 -0.6018  1.3776  5.8715
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.7707     1.2212  -2.269   0.0315 *
## ABV           0.9237     0.1215   7.601 3.55e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.52 on 27 degrees of freedom
## Multiple R-squared:  0.6815, Adjusted R-squared:  0.6697
## F-statistic: 57.78 on 1 and 27 DF,  p-value: 3.551e-08
```

```r
# show fitted values for each data point
fitted(fit)
```

```
##          1          2          3          4          5          6
##  5.5422752  1.8476316  6.6506683  7.0201326  8.1285257  7.4819631
##          7          8          9         10         11         12
## 11.2689728  1.8476316  8.3132579  6.4659361  0.7392385 11.1766067
##         13         14         15         16         17         18
## 11.1766067  4.0644177  3.2331229  6.7430344  1.5705333  5.7270074
##         19         20         21         22         23         24
##  6.0041056  3.8796856  3.8796856  1.8476316 10.6224102 11.0842406
##         25         26         27         28         29
##  2.4018281 12.4697320  4.9880786  0.3697741  1.7552655
```

```r
# show residuals for each data point
residuals(fit)
```

```
##           1           2           3           4           5           6
## -2.84227519 -0.04763155  3.34933172  2.97986735  5.87147426  4.51803690
##           7           8           9          10          11          12
##  1.73102717  0.10236845 -4.36325792 -2.96593610  0.96076154 -0.67660674
##          13          14          15          16          17          18
## -0.67660674 -1.46441774 -0.63312292  4.25696563  0.12946672 -2.87700737
##          19          20          21          22          23          24
## -2.75410565 -1.27968555 -1.37968555  0.75236845  1.37758981 -1.13424065
##          25          26          27          28          29
## -0.60182810 -2.51973201 -1.48807865  1.73022590 -0.05526546
```

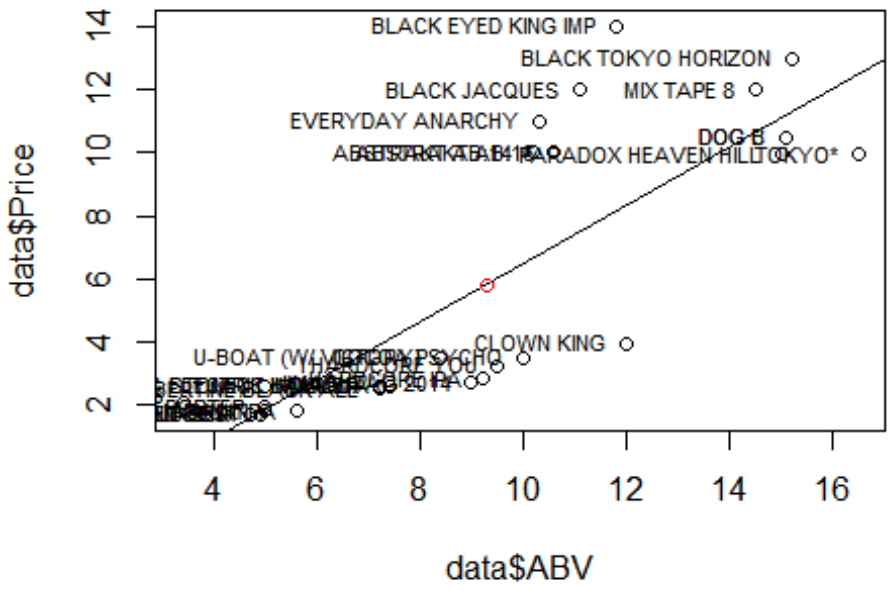
```r
# Work out the centroid value (mean of both)
mean(data$ABV)
```

```
## [1] 9.282759
```

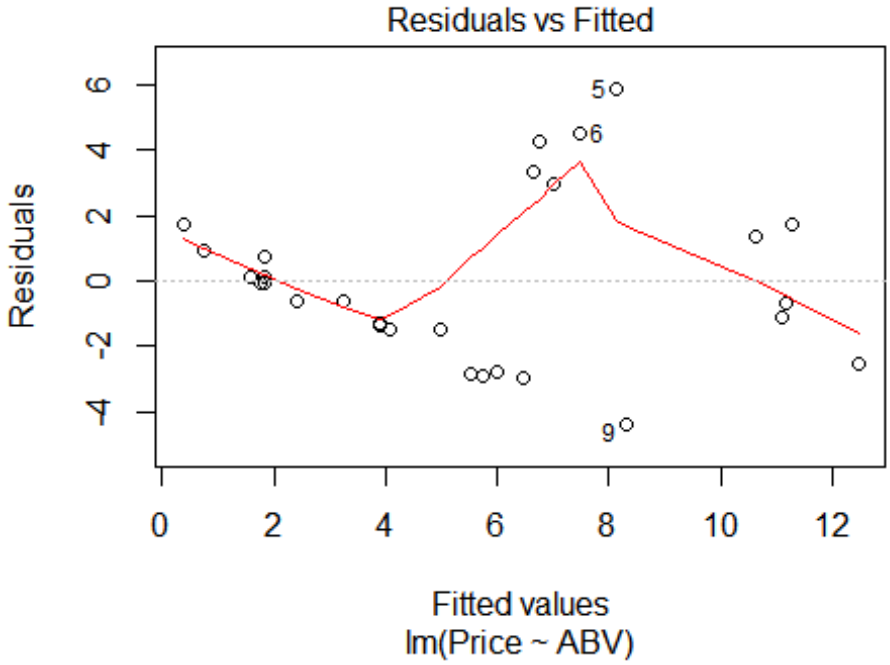
```r
mean(data$Price)
```

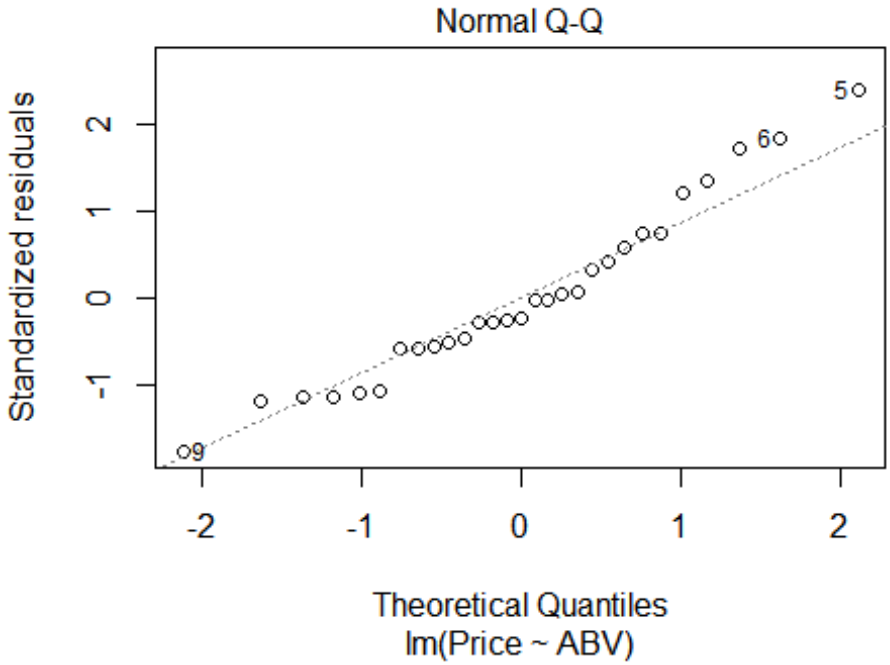
```
## [1] 5.803448
```

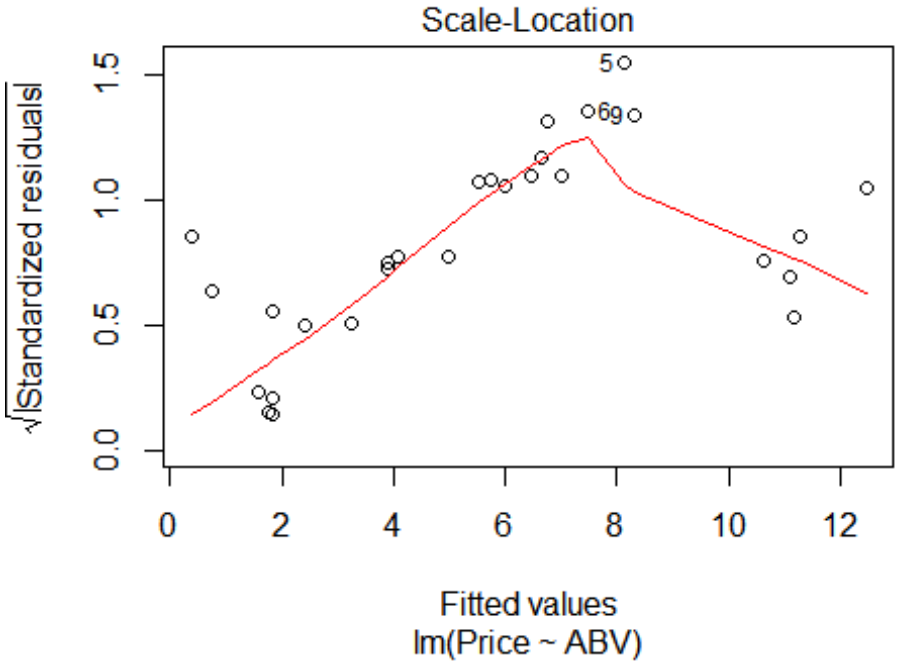
```r
# Plot centroid
points(mean(data$ABV), mean(data$Price), col = 'red')
```

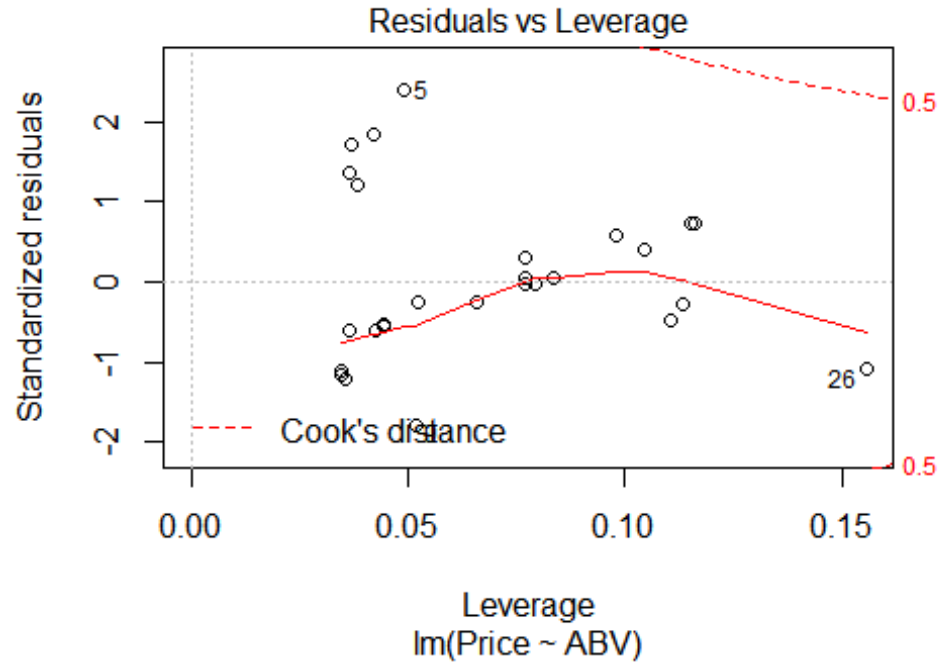

BLACK EYED KING IMP
BLACK TOKYO HORIZON
BLACK JACQUES
MIX TAPE 8
EVERYDAY ANARCHY
DOG B
CLOWN KING
U-BOAT
data$Price
data$ABV
2
4
6
8
10
12
14
16

```
plot(fit)
```


Residuals vs Fitted
Residuals
Fitted values
lm(Price ~ ABV)
5
6
9


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(Price ~ ABV)
5
6
9






```
anova(fit)

## Analysis of Variance Table
##
```

```
## Response: Price
##           Df Sum Sq Mean Sq F value    Pr(>F)
## ABV        1 367.01  367.01  57.778 3.551e-08 ***
## Residuals 27 171.51    6.35
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Histogram of residuals
hist(fit$residuals)
```