A Regression Example using iNZightRegression

Simon Potter, David Banks, Tom Elliott
The University of Auckland

December 3, 2013

The iNZightRegression package was created to provide additional functions which provide modified output, as well as some additional functionality. This document will demonstrate the use of iNZightRegression for modelling some regression data.

1 Linear Models

Here, we will be using the course.df dataset from the s20x library, which contains data collected from STATS 20x students. The first function is iNZightSummary(), which is shown below:

```
> library(iNZightRegression)
> library(s20x)
> data(course.df)
> model_1 <- lm(Exam ~ Test + Assign + Attend, data = course.df)
> iNZightSummary(model_1)
```

```
Model for: Exam

Coefficients:
            Estimate Std. Error t value    p-value
(Intercept)  -0.4960     3.1331  -0.158    0.874
Test          2.6014     0.2903   8.962 1.65e-15   ***
Assign        1.4804     0.2647   5.592 1.11e-07   ***
Attend                                   0.048   *
  No               0          -       -        -
  Yes         4.1047     2.0576   1.995      0.048 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.48 on 142 degrees of freedom
Multiple R-squared: 0.6919,        Adjusted R-squared: 0.6854
```

If compared to the basic summary() function provided by base R, there are several differences. The main difference is the way in which factors are printed.

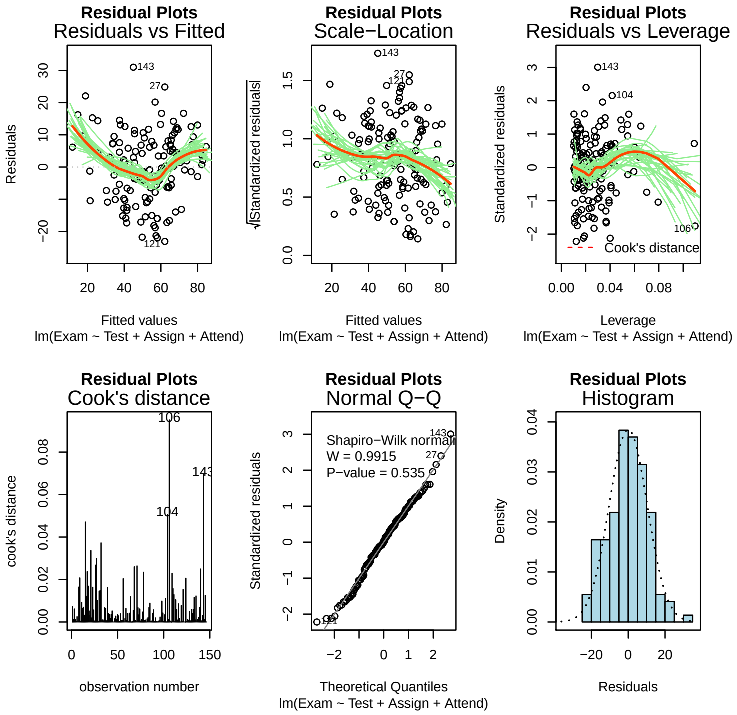
The baseline is included with an estimate of 0, and Type III Anova p-values are given for each factor.

Several additional arguments are demonstrated below, which allow bootstrap simulation of estimates, reordering of factors (use the most common level as the baseline), and excluding confounding variables from the output. If, for example, we are *only* insterested in the effect of not attending lectures on exam marks, we could use the following function call:

```
> iNZightSummary(model_1, method = 'bootstrap',
+                reorder.factors = TRUE, exclude = c('Test', 'Assign'))

Model for: Exam

Coefficients:
            Estimate Std. Error t value  p-value
(Intercept)    3.609      3.884   0.929 0.3544
Attend
  Yes              0          -       -        -
  No          -4.105      2.135  -1.923   0.0565 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.48 on 142 degrees of freedom
Multiple R-squared: 0.6919,        Adjusted R-squared: 0.6854
```

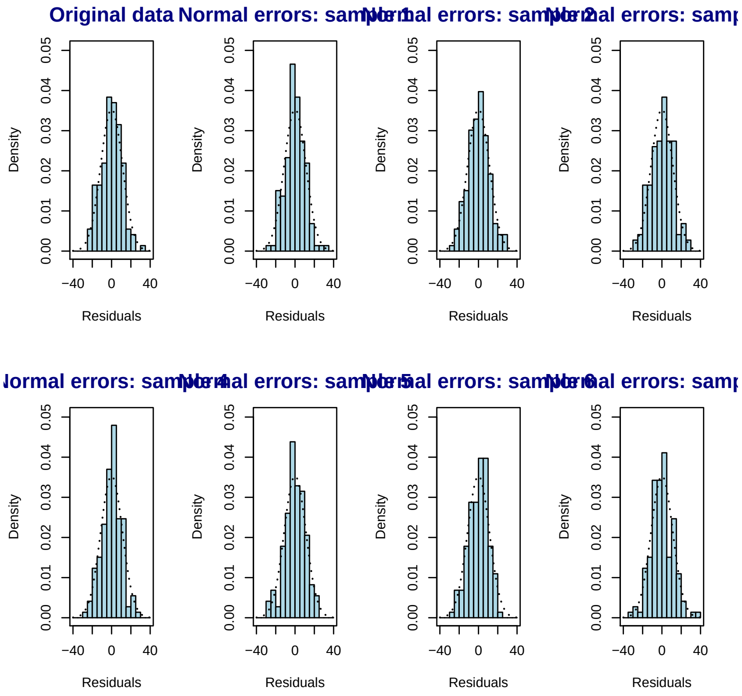
Summary and diagnostic plots can be produced simply by using the new `plotlm6` function, which takes several arguments. It provides 6 plots, which can be selected by the `which` argument (7 means all-in-one). Each residual plot shows a loess curve through the data, as well as a bootstrap sample of loess curves which are calculated from non-parametric bootstrap samples. The number of smoothers can be adjusted, or disabled completely as desired.

```
> plotlm6(model_1, which = 7, main = 'Residual Plots')
```

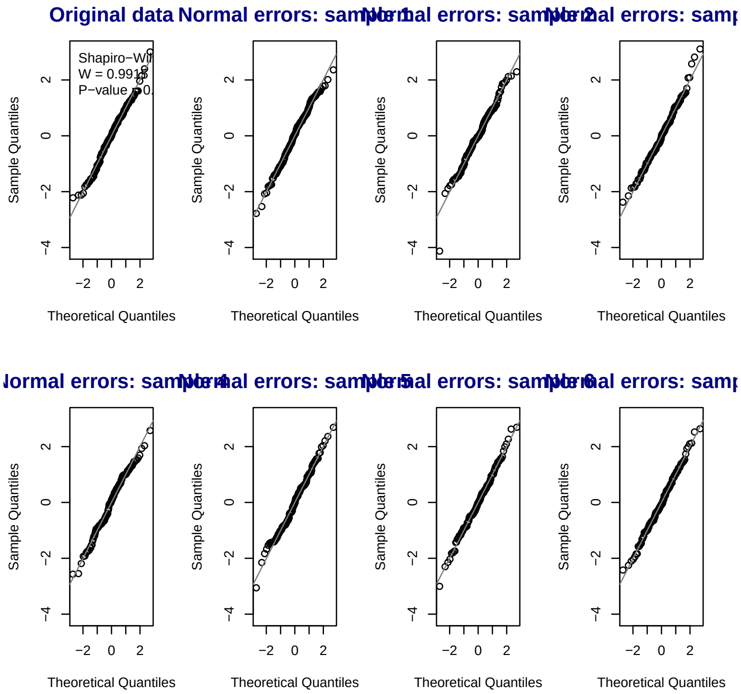


For normal data, there are also the `histogramArray` and `qqplotArray` functions, which display normally distributed (by simulation) datasets with the same mean and variance as those from the fitted model, which allow users to easy decide if the normality assumption is satisfied.

```
> histogramArray(model_1)
```



```
> qqplotArray(model_1)
```


Original data
Normal errors: sample 1
Normal errors: sample 2
Normal errors: samp
Shapiro−Wil
W = 0.9913
P−value = 0.
Sample Quantiles
−4
−2
0
2
−2
0
2
Theoretical Quantiles
lormal errors: sample 4
Normal errors: sample 5
Normal errors: sample 6
Normal errors: samp