1 Diagnostic Plots for Linear Models with `R`

There are six plots (selectable by `which`) are currently available:

1. a plot of residuals against fitted values,

2. a Scale-Location plot of *sqrt(| residuals |)* against fitted values,

3. a Normal Q-Q plot,

4. a plot of Cook's distances versus row labels,

5. a plot of residuals against leverages,

6. a plot of Cook's distances against *leverage/(1-leverage)*.

By default, the first three and 5 are provided, if you just type something like `plot(fit)`.

```
plot(lm(mpg~wt+cyl),which=c(1),pch=18,col="red")
plot(lm(mpg~wt+cyl),which=c(2),pch=18,col="red")
plot(lm(mpg~wt+cyl),which=c(3),pch=18,col="red")
plot(lm(mpg~wt+cyl),which=c(4),pch=18,col="red")
plot(lm(mpg~wt+cyl),which=c(5),pch=18,col="red")
plot(lm(mpg~wt+cyl),which=c(6),pch=18,col="red")
```

- The first one displays the residuals vs. the fitted values we use this to evlauate the mean, variance and correlation of residuals. If our assumptions of constant variance and uncorrelated residuals are violated we may be able to correct this with a variance-stabilizing transformation.

- The second plot helps us check the normality of the residuals. If the residuals are indeed normal, they should fall along the dashed line. Remember that the normality assumption for our errors allows us to determine the standard errors of our coefficients and predictions.

- The final plot will display our residuals vs. their leverage. The dashed red lines are level curves that denote a particular value of Cook's distance. We will pay attention to points lying beyond the distance of 1. Notice that when we have data with row labels, the points will be labeled with their names. Otherwise, the row number will be shown.

- **Plot 2** - The **Scale-Location** plot, also called Spread-Location (or S-L plot), takes the square root of the absolute residuals in order to diminish skewness (sqrt($|E|$)) is much less skewed than $|E|$ for Gaussian zero-mean E).

- **Plot 5** - The **Residual-Leverage** plot shows contours of equal Cook's distance, for values of `cook.levels` (by default 0.5 and 1) and omits cases with leverage one with a warning. If the leverages are constant (as is typically the case in a balanced aov situation) the plot uses factor level combinations instead of the leverages for the x-axis.
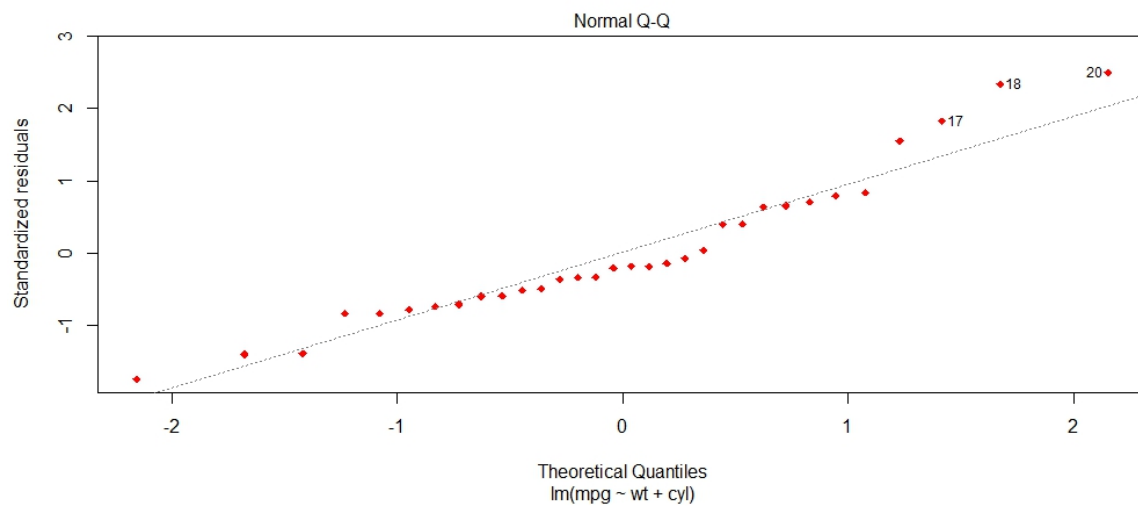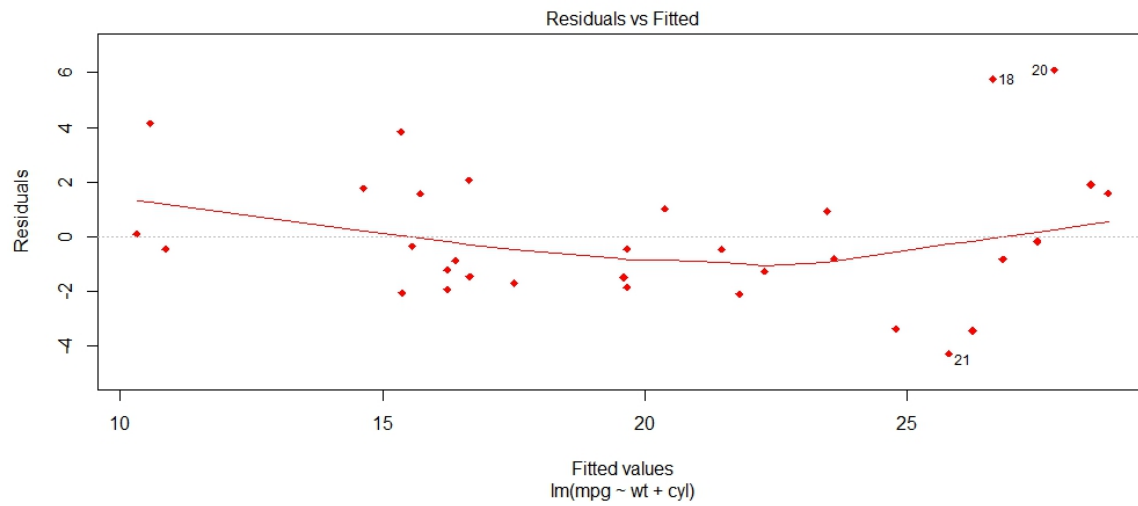 *(The factor levels are ordered by mean fitted value.)*

EDIT NOTE - FOLLOWING IN WRONG ORDER

Plot 1 : Residual Plot

Test for Constant Variance
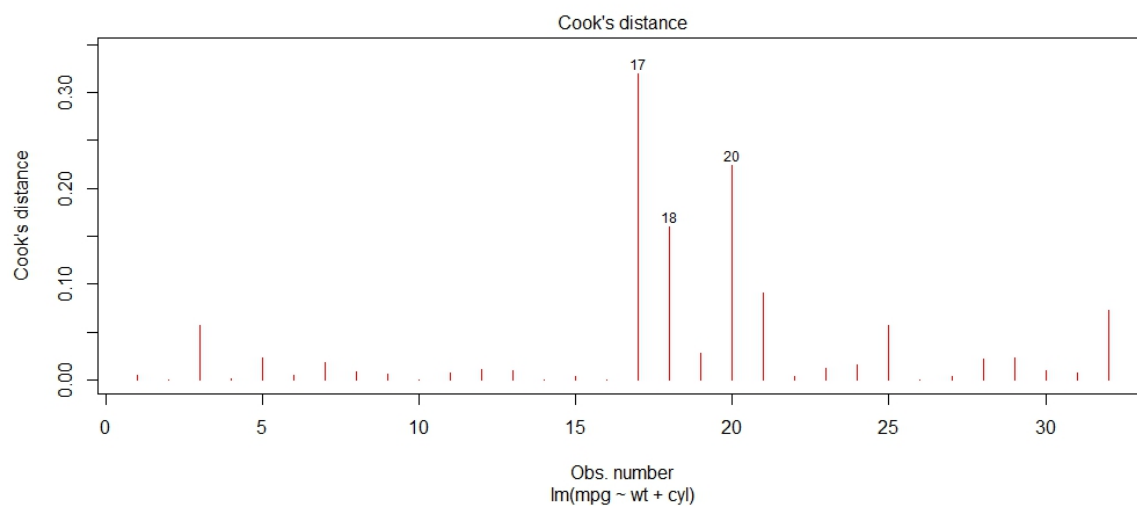

Plot 3 : Normal Probability Plot

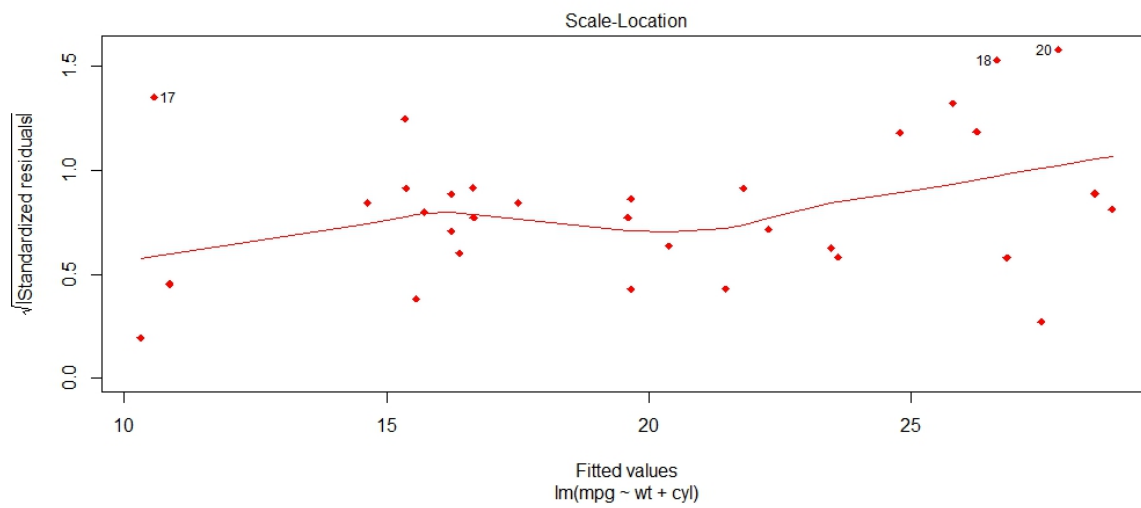This plot is used to assess the validity of the normality of the residuals.

Residuals vs Fitted

Normal Q-Q

Plot 5 : Cook's Distance

Plot 6 : Cook's Distance vs Leverage

Plot the four default plots together:

3

Scale-Location

Fitted values
lm(mpg ~ wt + cyl)

Cook's distance

Obs. number
lm(mpg ~ wt + cyl)

```
par(mfrow=c(4,1))
plot(fittedmodel)
par(opar)
```

Residuals vs Leverage

lm(mpg ~ wt + cyl)

Cook's dist vs Leverage $h_{ii}/(1-h_{ii})$

lm(mpg ~ wt + cyl)