STAT 210
Applied Statistics and Data Analysis
Multiple Linear Regression 2
Anova and Hypothesis Tests

Joaquin Ortega

Anova with Multiple Regression

Anova with Multiple Regression

The Anova calculations for multiple regression are almost identical to those of simple linear regression, except that now the degrees of freedom have to be adjusted, taking into account the number of parameters.

The table below shows the Anova values for multiple regression

Table 1: Anova table for multiple regression.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F_{obs}
Regression	$SSR = \hat{\beta}'\mathbf{X}'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{J}\mathbf{Y}$	$p-1$	$MSR = \frac{SSR}{p-1}$	$F = \frac{MSR}{MSE}$
Error	$SSE = \mathbf{Y}'\mathbf{Y} - \hat{\beta}\mathbf{X}'\mathbf{Y}$	$n-p$	$MSE = \frac{SSE}{n-p}$	
Total	SST	$n-1$		

Here, $\mathbf{J}$ is a matrix of 1's.

Anova with Multiple Regression

We have the following formulas for the sums of squares,

$$SSE = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$$
$$SST = \mathbf{Y}'(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{Y}$$
$$SSR = \mathbf{Y}'(\mathbf{H} - \frac{1}{n}\mathbf{J})\mathbf{Y}$$

These three sums of squares can be expressed as quadratic forms involving the matrices $\mathbf{H}, \mathbf{J}$, and $\mathbf{I}$.

Anova with Multiple Regression

If we call for the anova table associated with the first regression in the previous video using the `anova` command in `R`, we get

```r
(model1.anova <- anova(model1))
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## hp         1 678.37  678.37 100.1767 1.393e-10 ***
## wt         1 252.63  252.63  37.3059 1.593e-06 ***
## disp       1   0.06    0.06   0.0084    0.9275
## drat       1  12.15   12.15   1.7947    0.1915
## Residuals 27 182.84    6.77
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

and this does not look like table 1.

Anova with Multiple Regression

In this example there are

- $n = 32$ observations,
- 4 independent variables or regressors,
- $p = 5$ coefficients (add one for the intercept),

so $n - p = 32 - 5 = 27$, which are the degrees of freedom corresponding to `Residuals` in the table.

The sum of squares for the residuals is $SSE = 182.84$, and the mean square error is

$$MSE = \frac{SSE}{p-1} = \frac{182.84}{4} = 6.77.$$

This corresponds to the second line in table 1.

Anova with Multiple Regression

However, first line values, which correspond to the regression sum of squares, are split up into the individual regressors.

We will explain how this is done later on, but for the moment, we need to add degrees of freedom and sums of squares for the first four rows in our table to get the first row in table 1:

```
(SSE = model1.anova$`Sum Sq`[5])
```

```
## [1] 182.8376
```

```
(SSR = sum(model1.anova$`Sum Sq`[1:4]))
```

```
## [1] 943.2096
```

Anova with Multiple Regression

and since $p - 1 = 4$,

```
(MSE = SSE/27)
```

```
## [1] 6.771762
```

```
(MSR = SSR/4)
```

```
## [1] 235.8024
```

$$MSR = \frac{SSR}{p-1} = \frac{943.21}{4} = 235.8$$

Anova with Multiple Regression

The total sum of squares is

```
(SST = SSE + SSR)
```

```
## [1] 1126.047
```

Finally, F_{obs} is given by

```
(Fobs <- MSR/MSE)
```

```
## [1] 34.82143
```

Observe that this is the same value we get at the bottom line of the summary for the regression:

Anova with Multiple Regression

```r
summary(model1)
```

```
##
## Call:
## lm(formula = mpg ~ hp + wt + disp + drat, data = data.cars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5077 -1.9052 -0.5057  0.9821  5.6883
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 29.148738   6.293588   4.631  8.2e-05 ***
## hp          -0.034784   0.011597  -2.999  0.00576 **
## wt          -3.479668   1.078371  -3.227  0.00327 **
## disp         0.003815   0.010805   0.353  0.72675
## drat         1.768049   1.319779   1.340  0.19153
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.602 on 27 degrees of freedom
## Multiple R-squared:  0.8376, Adjusted R-squared:  0.8136
## F-statistic: 34.82 on 4 and 27 DF,  p-value: 2.704e-10
```

We can obtain the *p*-value with

```r
1-pf(Fobs,4,27)
```

```
## [1] 2.70431e-10
```

Coefficient of Determination

Coefficient of Determination

As in the case of simple linear regression, the coefficient of determination R^2 is defined by

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

and is interpreted as the proportion of the variability that is explained by the regression.

However, adding more variables to the model always increases R^2, even if the new variables do not add anything significant to the model.

Coefficient of Determination

Therefore, we need to adjust this coefficient to consider the number of variables in the model.

The new coefficient is known as the **adjusted coefficient of determination** or **adjusted R squared** and is defined as

$$R_a^2 = 1 - \frac{SSE/(n-p)}{SST/(n-1)} = 1 - \frac{MSE}{MST}.$$

As we stated when we introduced R^2 in the case of simple linear regression, caution must be exercised when using this coefficient in the context of model fitting.

Extra Sums of Squares

Extra Sums of Squares

If we add regressors to a model, the sums of squares will change. To assess the contribution of a new variable on an existing model, we look at the change in the sum of squares.

For example, the marginal increase when variable x_2 is added to a model based on variable x_1 is given by

$$SSR(\beta_2|\beta_1, \beta_0) = SSR(\beta_1, \beta_2|\beta_0) - SSR(\beta_1|\beta_0) \tag{1}$$

where $SSR(\beta_1, \beta_2|\beta_0)$ is the regression sum of squares for the (complete) model including β_1 and β_2 (β_0 is always included) and $SSR(\beta_1|\beta_0)$ is the sum for the model with only x_1 as a regressor. This is equivalent to

$$SSR(\beta_2|\beta_1, \beta_0) = SSE(\beta_1|\beta_0) - SSE(\beta_1, \beta_2|\beta_0).$$

Extra Sum of Squares

We can turn (1) around and get

$$SSR(\beta_1, \beta_2|\beta_0) = SSR(\beta_1|\beta_0) + SSR(\beta_2|\beta_1, \beta_0) \tag{2}$$

which can be seen as a decomposition of the regression sum of squares into two parts, the contribution of the simple model with only x_1 as a regressor plus the increase when a second regressor x_2 is added to the model.

Observe, however, that we could also have done this decomposition in a different order

$$SSR(\beta_1, \beta_2|\beta_0) = SSR(\beta_2|\beta_0) + SSR(\beta_1|\beta_2, \beta_0) \tag{3}$$

and the results may be different.

These are known as **incremental** or **type I** sums of squares.

Extra Sum of Squares

The sums of squares that appear in an anova table obtained with the command `anova` acting on an `lm` object are incremental sums of squares that follow the order set in the model's defining equation.

Thus, if the formula for the model is

$$\texttt{y} \sim \texttt{x1 + x2},$$

the first sum in the table corresponds to $SSR(\beta_1|\beta_0)$ and the second to $SSR(\beta_2|\beta_1, \beta_0)$.

Extra Sum of Squares

As an example, let us look at the third model fitted to the `mtcars` data. There we had `mpg ~ hp + wt` and we will produce the anova tables for the two possible orders for the regressors:

```r
anova(model3)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## hp         1 678.37  678.37 100.862 5.987e-11 ***
## wt         1 252.63  252.63  37.561 1.120e-06 ***
## Residuals 29 195.05    6.73
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Extra Sum of Squares

```r
model3a <- lm(mpg ~ wt + hp, data=data.cars)
anova(model3a)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## wt         1 847.73  847.73 126.041 4.488e-12 ***
## hp         1  83.27   83.27  12.381  0.001451 **
## Residuals 29 195.05    6.73
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Extra Sum of Squares

```r
anova(model3)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## hp         1 678.37  678.37 100.862 5.987e-11 ***
## wt         1 252.63  252.63  37.561 1.120e-06 ***
## Residuals 29 195.05    6.73
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(model3a)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## wt         1 847.73  847.73 126.041 4.488e-12 ***
## hp         1  83.27   83.27  12.381  0.001451 **
## Residuals 29 195.05    6.73
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Extra Sum of Squares

The sum of squares for the residuals are the same, as are the degrees of freedom appearing in the table. However, the sums of squares for the regressors are different. From these tables we have

$$SSR(hp|\beta_0) = 678.37 \qquad SSR(wt|\beta_0) = 847.73$$
$$SSR(wt|hp, \beta_0) = 252.63 \qquad SSR(hp|wt, \beta_0) = 83.27$$

Observe that the column sums are equal:

$$678.37 + 252.63 = 847.73 + 83.27 = 931$$

which corresponds to the sum of squares for the regression SSR.

Confidence Intervals

Regression Coefficients

We have assumed the errors to have a standard normal distribution, therefore $\hat{\beta}_j \sim N(\beta_j, \sigma^2 v_{jj})$ and

$$\frac{\hat{\beta}_j - \beta_j}{\sigma\sqrt{v_{jj}}} \sim N(0, 1).$$

σ is unknown but $(n - p - 1)MSE/\sigma^2 \sim \chi^2_{n-p-1}$, and therefore

$$\frac{\hat{\beta}_j - \beta_j}{MSE\sqrt{v_{jj}}} \sim t_{n-p-1}.$$

Hence, the α-level confidence interval is

$$\hat{\beta}_j \pm t_{n-p-1,1-\frac{\alpha}{2}}(MSEv_{jj})^{1/2}. \tag{4}$$

Mean value at a point

The average response value at $\mathbf{Z}' = (Z_1, Z_2, \ldots, Z_p)$ is $\hat{Y}(\mathbf{Z}) = \mathbf{Z}'\hat{\beta}$. Thus

$$E(\hat{Y}(\mathbf{Z})) = E(\mathbf{Z}'\hat{\beta}) = \mathbf{Z}'E(\hat{\beta}) = \mathbf{Z}'\beta = E(Y(\mathbf{Z}))$$

and

$$Var(\hat{Y}(\mathbf{Z})) = \mathbf{Z}'Var(\hat{\beta})\mathbf{Z} = \mathbf{Z}'\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z} = \sigma^2(\mathbf{Z}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z}).$$

Therefore $\hat{Y}(\mathbf{Z}) \sim N(E(Y(\mathbf{Z})), \sigma^2(\mathbf{Z}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z}))$ and

$$\frac{\hat{Y}(\mathbf{Z}) - E(Y(\mathbf{Z}))}{(MSE(\mathbf{Z}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z}))^{1/2}} \sim t_{n-p-1}.$$

The confidence interval at level α is

$$\hat{y}(\mathbf{z}) \pm t_{n-p-1,1-\frac{\alpha}{2}}(MSE(\mathbf{z}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{z})^{1/2}. \tag{5}$$

Prediction Interval

The prediction interval concerns the response $Y(\mathbf{Z})$ corresponding to a specific value of the regressors (and not the mean response which is $E(Y(\mathbf{Z})))$.

Let $Y(\mathbf{Z})$ be the observation at $\mathbf{Z}$ and let $\hat{Y}(\mathbf{Z}) = \mathbf{Z}'\hat{\beta}$.

Since $\hat{Y}(\mathbf{Z})$ is a linear combination of normal rvs, it is normal. Also, $Y(\mathbf{Z})$ is normal and therefore, the difference $Y(\mathbf{Z}) - \hat{Y}(\mathbf{Z})$ is a normal random variable. Moreover,

$$Y(\mathbf{Z}) - \hat{Y}(\mathbf{Z}) \sim N(0, \sigma^2(1 + \mathbf{Z}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z})$$

and the prediction interval is

$$\hat{y}(\mathbf{z}) \pm t_{n-p-1,1-\frac{\alpha}{2}}\left(MSE(1 + \mathbf{z}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{z})\right). \tag{6}$$

Hypotheses Tests

Significance Test

This test is done to determine if the regression is significant, i.e., if the regressor variables contribute at all to explain Y:

$$H_0: \ \beta_1 = \beta_2 = \cdots = \beta_p = 0 \quad vs. \quad H_A: \ \text{at least one } \beta_i \neq 0$$

For this test, we divide the total variability of the problem

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

into the part that is explained by the model and the part that remains unexplained:

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = SSE + SSR.$$

Significance Test

$$SST = SSE + SSR.$$

The first sum is the variability that the model has not been able to explain, $SSE = \sum \hat{\epsilon}_i$, and the second represents what has been explained, denoted by SSR.

We know

$$\frac{SSR}{\sigma^2} \sim \chi^2_{p-1} \quad \text{under } H_0.$$

It is also possible to prove that

$$\frac{SSE}{\sigma^2} \sim \chi^2_{n-p} \quad \text{under } H_0.$$

Significance Test

Since SSE and SSR turn out to be independent, the test statistic

$$F_0 = \frac{SSR/p-1}{SSE/n-p} \sim F_{p-1,n-p} \qquad \text{under } H_0.$$

When this statistic is large, the part explained by the regression is large with respect to the residual part, and the opposite is true when it is small.

We reject H_0 when the test statistic has a large value. The critical region is $\{F_0 > F_{p-1,n-p;1-\alpha}\}$.

Test on Individual Parameters

Sometimes it is useful to contrast individual parameters to determine the value of the variables within the model.

It may be that the model is more effective by deleting one of the variables.

Adding a variable always increases SSR and decreases SSE, but it may not be enough to lower MSE, so the contribution of the variable to the model may not be significant.

Test on Individual Parameters

The hypotheses for the test of significance of the parameter β_j is

$$H_0: \ \beta_j = 0 \qquad vs \qquad H_1: \ \beta_j \neq 0.$$

We know that $\hat{\beta}_j \sim N(\beta_j, \sigma^2 v_{jj})$. Under H_0

$$\frac{\hat{\beta}_j}{MSEv_{jj}} \sim t_{n-p-1}.$$

Since this is a bilateral test, the rejection region is

$$\Big\{ \frac{|\hat{\beta}_j|}{MSEv_{jj}} > t_{n-p-1,1-\frac{\alpha}{2}} \Big\}.$$

Tests on Subsets of Parameters

These contrasts look at the contribution of variables $x_{r+1}, \ldots, x_p$ to the model based on $x_1, \ldots, x_r$.

β_0 is not included in the hypothesis since it will always be in the model:

$H_0: \ \beta_{r+1} = \cdots = \beta_p = 0 \quad vs. \quad H_1: \text{ at least one } \beta_j \neq 0, \ r+1 \leq j \leq p.$

Fit the model $Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r + \epsilon$ with which we measure the contribution of $x_1, \ldots, x_r$ when they are alone, i.e,

$$SST = SSR(\beta_1, \ldots, \beta_r | \beta_0) + SSE(\beta_1, \ldots, \beta_r | \beta_0)$$

and then fit the complete model and compare to see the contribution of all the variables.

Tests on Subsets of Parameters

In the first model $SSR(\beta_1, \ldots, \beta_r|\beta_0)$ has $r-1$ degrees of freedom and in the second $SSR(\beta_1, \ldots, \beta_p|\beta_0)$ has $p-1$ degrees of freedom.

The difference between these two quantities is the contribution of $x_{r+1}, \ldots, x_p$ when $x_1, \ldots, x_r$, are in the model, which is written as

$$SSR(\beta_{r+1}, \ldots, \beta_p|\beta_0) = SSR(\beta_1, \ldots, \beta_p|\beta_0) - SSR(\beta_1, \ldots, \beta_r|\beta_0)$$

and has $p-r$ degrees of freedom. We use the statistic

$$F_0 = \frac{SSR(\beta_{r+1}, \ldots, \beta_p|\beta_0)/p-r}{SSR(\beta_1, \ldots, \beta_p|\beta_0)/n-p-1} \sim F_{p-r,n-p-1}.$$

for this test.

Hypotheses Tests: Example

As an example, let us consider the `mtcars` data again.

Our previous analysis saw that variables `disp` and `drat` did not seem to contribute much to the model. Let us test whether these two variables can be dropped from the model.

`model1` is the full model, and variables `hp` and `wt` were introduced as the first and second term in the regression equation.

The anova table for this model was calculated previously and is in `model1.anova`.

We calculate now the anova table for the reduced model

Hypotheses Tests: Example

```
model1.anova
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## hp         1 678.37  678.37 100.1767 1.393e-10 ***
## wt         1 252.63  252.63  37.3059 1.593e-06 ***
## disp       1   0.06    0.06   0.0084    0.9275
## drat       1  12.15   12.15   1.7947    0.1915
## Residuals 27 182.84    6.77
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hypotheses Tests: Example

```
(model4.anova <- anova(model4))
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## disp       1 808.89  808.89 95.0929 1.164e-10 ***
## wt         1  70.48   70.48  8.2852  0.007431 **
## Residuals 29 246.68    8.51
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hypotheses Tests: Example

```r
(SSRf = sum(model1.anova$`Sum Sq`[1:4])) # Complete model
```

```
## [1] 943.2096
```

```r
(SSRr = sum(model4.anova$`Sum Sq`[1:2])) # Reduced model
```

```
## [1] 879.3647
```

```r
(SSRnv = SSRf-SSRr)
```

```
## [1] 63.84494
```

```r
(SSEf = model1.anova$`Sum Sq`[5]) # Error sum of squares
```

```
## [1] 182.8376
```

Hypotheses Tests: Example

The test statistic is

```
(Fcomp <- (SSRnv/2)/(SSEf/27))
```

```
## [1] 4.714057
```

The p-value is given by

```
pf(Fcomp, 2, 27, lower.tail=FALSE)
```

```
## [1] 0.01753944
```

Hypotheses Tests: Example

This is what we get when we do an anova to compare the two models:

```r
anova(model4,model1)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ disp + wt
## Model 2: mpg ~ hp + wt + disp + drat
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     29 246.68
## 2     27 182.84  2    63.845 4.7141 0.01754 *
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual Analysis

Residual Analysis

1.- To verify the hypothesis of normality of the residuals, quantile plots of normality are made. If normality does not hold, transformations can be made to achieve it, but the linearity or homogeneity of the variance can be lost. The hypothesis of normality may not be too critical.

2.- If time is a regressor or measurements are taken periodically, a graph of the residuals against time can be made. In this graph, we can see if there is any tendency that can make us reject the hypothesis of independence.

3.-It is also useful to graph the residuals against the regressors or against the fitted values, to discover possible trends that indicate the need to introduce higher-order polynomial terms or make transformations of the variables. Usually, graphs are made of the residuals divided by their (sample) standard deviation to normalize them.

Properties of Residuals

1. $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.
2. $\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y}$ with $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.
3. $\mathbf{H}$ is the projection matrix on the subspace generated by the columns of the design matrix $\mathbf{X}$: $\mathbf{1}, \mathbf{X}_1, \ldots, \mathbf{X}_p$.
4. $\hat{\epsilon} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$.
5. $\mathbf{I} - \mathbf{H}$ is the projection matrix on the subspace orthogonal to $\langle \mathbf{1}, \mathbf{X}_1, \ldots, \mathbf{X}_p \rangle$.

Properties of Residuals

6. $\mathbf{H}$ and $\mathbf{I}-\mathbf{H}$ are idempotent and symmetrical: $\mathbf{H}^2=\mathbf{H}$, $(\mathbf{I}-\mathbf{H})^2=\mathbf{I}-\mathbf{H}$.
7. $E(\hat{\epsilon})=(\mathbf{I}-\mathbf{H})E(\mathbf{Y})=(\mathbf{I}-\mathbf{H})\mathbf{X}\beta=\mathbf{X}\beta-\mathbf{X}\beta=\mathbf{0}$.
8. $Cov(\hat{\boldsymbol{\epsilon}})=(\mathbf{I}-\mathbf{H})\sigma^2\mathbf{I}(\mathbf{I}-\mathbf{H})=\sigma^2(\mathbf{I}-\mathbf{H})$. Therefore

$$E(\hat{\epsilon}_i)=0$$
$$Var(\hat{\epsilon}_i)=\sigma^2(1-h_{ii}),$$
$$Cov(\hat{\epsilon}_i,\hat{\epsilon}_j)=-\sigma^2 h_{ij}$$

Properties of Residuals

Unlike the ϵ_i, the $\hat{\epsilon}_i$ do not have constant variance and are correlated.

We have

$$E\left(\frac{\hat{\epsilon}_i}{\sigma\sqrt{1-h_{ii}}}\right)=0, \qquad Var\left(\frac{\hat{\epsilon}_i}{\sigma\sqrt{1-h_{ii}}}\right)=1$$

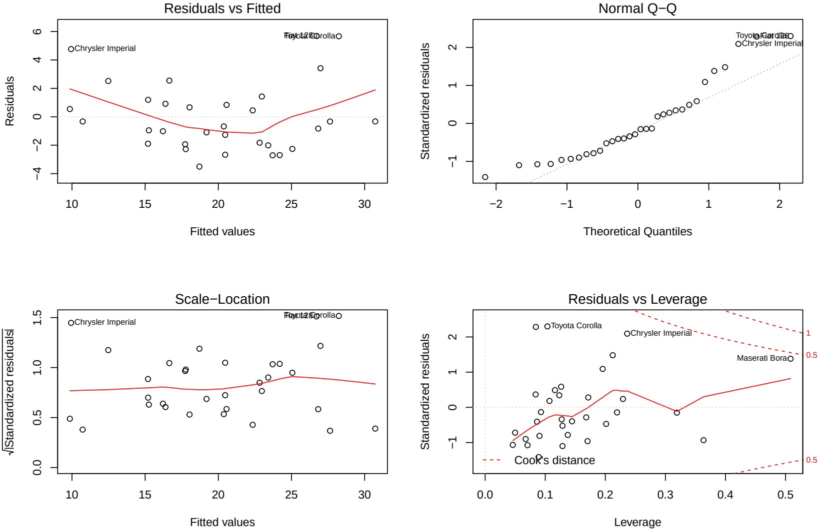
Since σ^2 is unknown, to standardize the residuals they are divided by the empirical standard deviation

$$r_i=\frac{\hat{\epsilon}_i}{MSE\sqrt{1-h_{ii}}}. \tag{7}$$

Standardized residuals are also known as internally studentized residuals.

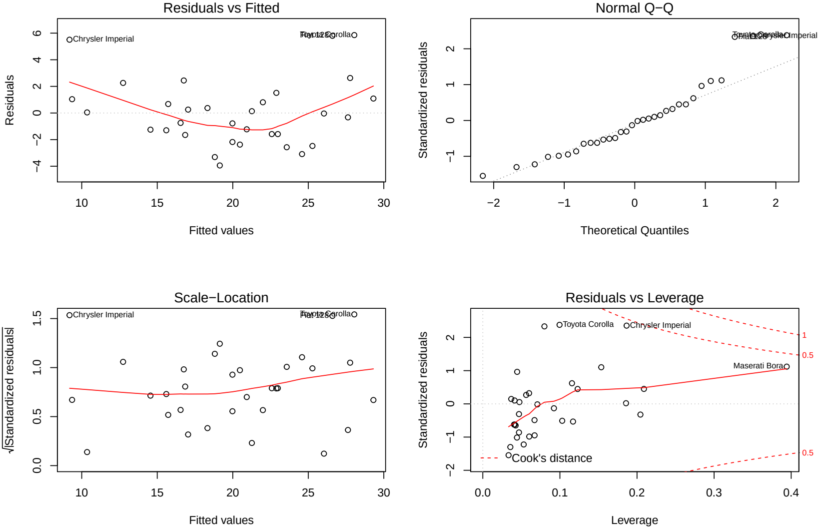
Properties of Residuals

Residual graphs for some of the models fitted for the `mtcars` data set. We start with the complete model



Properties of Residuals

and then plot the reduced model with only two variables: `model3`.



Properties of Residuals

We see that all graphs are very similar and reasonably good.

The graphs of residuals vs. fitted values have some curvature that suggests exploring a higher-order model. Still, the graphs of standardized residuals vs. fitted values do not show important tendencies, so that the curvature may be due to differences in variance.

The right tail in the normal quantile plot shows some points distant from the reference line.

The `R` function `rstandard()` computes standardized residuals according to (7). The function `stdres()` in the MASS package also computes standardized residuals.