Linear statistical models
Inference for the less than full rank model

Notes by Yao-ban Chan and Owen Jones

Recap

We consider the less than full rank model:

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where the errors $\boldsymbol{\varepsilon}$ have mean $\mathbf{0}$, variance $\sigma^2 I$, and (sometimes) are assumed to be jointly normally distributed.

In a less than full rank model, $r(X) < p$, and this means that $X^T X$ is singular. In turn, this means that not all quantities associated with the model can be estimated.

The quantities that can be estimated are called estimable.

Inference for the less than full rank model

In this section, we develop procedures for testing hypotheses for the less than full rank model.

As you might expect, not all hypotheses can be tested. In general, if you cannot estimate the value of something, it is difficult to test any hypotheses about it!

Hypotheses which can be tested are called *testable*. This is defined as follows.

Testability

Definition 7.1

A hypothesis H_0 is testable if there exists a set of estimable functions $\mathbf{c}_1^T\boldsymbol{\beta}, \mathbf{c}_2^T\boldsymbol{\beta}, \ldots, \mathbf{c}_m^T\boldsymbol{\beta}$ such that H_0 is true if and only if

$$\mathbf{c}_1^T\boldsymbol{\beta} = \mathbf{c}_2^T\boldsymbol{\beta} = \ldots = \mathbf{c}_m^T\boldsymbol{\beta} = \mathbf{0},$$

and $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m$ are linearly independent.

Testability

That is, a testable hypothesis is of the form $H_0 : C\boldsymbol{\beta} = \mathbf{0}$, where

$$C = \begin{bmatrix} \mathbf{c}_1^T \\ \vdots \\ \mathbf{c}_m^T \end{bmatrix}$$

is $m \times p$ of rank m, and each $\mathbf{c}_i^T\boldsymbol{\beta}$ is estimable.

Testability

Now recall that a linear function $\mathbf{c}^T\boldsymbol{\beta}$ is estimable if and only if

$$\mathbf{c}^T(X^TX)^cX^TX = \mathbf{c}^T.$$

Therefore $H_0 : C\boldsymbol{\beta} = \mathbf{0}$ is testable if and only if C is of full rank and

$$C(X^TX)^cX^TX = C.$$

Note that since $r((X^TX)^cX^TX) = r(X) = r$, the maximum number of linearly independent estimable functions in a less than full rank model is r, so $m \leq r \leq p$.

Testability

Example. Consider the one-way classification model with fixed effects and $k = 3$. The linear model that we use is

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}.$$

Consider the hypothesis that the means of all three populations are equal. This is equivalent to $H_0 : \tau_1 = \tau_2 = \tau_3$.

This hypothesis is true if and only if

$$\tau_1 - \tau_2 = 0$$

and

$$\tau_2 - \tau_3 = 0.$$

Testability

So we can express this hypothesis as $H_0 : C\boldsymbol{\beta} = \mathbf{0}$, where

$$C = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}.$$

$\tau_1 - \tau_2$ is a contrast, so it is estimable. Similarly, $\tau_2 - \tau_3$ is estimable. The rows of C are obviously linearly independent, so H_0 is testable.

Testability

Once we have determined that a hypothesis is testable, how can we test it?

We look back at the full rank case. Here the hypothesis $H_0 : C\boldsymbol{\beta} = \mathbf{0}$ is a special case of the general linear hypothesis $C\boldsymbol{\beta} = \boldsymbol{\delta}^*$, with $\boldsymbol{\delta}^* = \mathbf{0}$.

The F statistic for this hypothesis (for the full rank case) is

$$\frac{(C\mathbf{b})^T[C(X^TX)^{-1}C^T]^{-1}C\mathbf{b}/m}{SS_{Res}/(n-p)},$$

which under the null hypothesis has an F distribution with m and $n-p$ degrees of freedom, where $m = r(C)$.

Testability

In the less than full rank case, X^TX does not have an inverse. However, we can use a conditional inverse.

The other change is that $\frac{SS_{Res}}{n-p}$ is no longer the estimator of the variance, s^2. We change this to the new estimator, $s^2 = \frac{SS_{Res}}{n-r}$ (where $r = r(X)$).

Testability

Therefore our proposed statistic for testing this hypothesis in the less than full rank model is

$$\frac{(C\mathbf{b})^T[C(X^TX)^cC^T]^{-1}C\mathbf{b}/m}{s^2},$$

which under the null hypothesis should follow an F distribution with m and $n-r$ degrees of freedom.

The following theorems justify this statistic.

Testability

Theorem 7.2

In the linear model $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, assume $\boldsymbol{\varepsilon} \sim MVN(\mathbf{0}, \sigma^2 I)$. Suppose that $C\boldsymbol{\beta} = \mathbf{0}$ is testable, where C is an $m \times p$ matrix with rank m. Then

$$\frac{(C\mathbf{b})^T[C(X^TX)^cC^T]^{-1}C\mathbf{b}}{\sigma^2}$$

has a noncentral χ^2 distribution with m degrees of freedom and noncentrality parameter

$$\lambda = \frac{(C\boldsymbol{\beta})^T[C(X^TX)^cC^T]^{-1}C\boldsymbol{\beta}}{2\sigma^2}.$$

Testability

Proof. Since the hypothesis is testable, $C\mathbf{b}$ is not dependent on the conditional inverse that we use to calculate $\mathbf{b}$.

Since

$$C\mathbf{b} = C(X^TX)^cX^T\mathbf{y},$$

we see that $C\mathbf{b}$ is a multivariate normal vector with mean

$$C(X^TX)^cX^TX\boldsymbol{\beta} = C\boldsymbol{\beta}$$

and variance

$$C(X^TX)^cX^T\sigma^2IX(X^TX)^cC^T = C(X^TX)^cC^T\sigma^2.$$

The result now follows from Corollary 3.10 provided that $C(X^TX)^cC^T$ is invertible, which is left as an exercise.

Testability

Theorem 7.3

In the linear model $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, assume $\boldsymbol{\varepsilon} \sim MVN(\mathbf{0}, \sigma^2 I)$. Suppose that $C\boldsymbol{\beta} = \mathbf{0}$ is testable. Then $C\mathbf{b}$ and s^2 are independent.

Proof.

We use Theorem 3.13. We know that

$$C\mathbf{b} = C(X^T X)^c X^T \mathbf{y}.$$

We can also write

$$SS_{Res} = \mathbf{y}^T [I - H]\mathbf{y}.$$

Testability

Theorem 7.3

In the linear model $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, assume $\boldsymbol{\varepsilon} \sim MVN(\mathbf{0}, \sigma^2 I)$. Suppose that $C\boldsymbol{\beta} = \mathbf{0}$ is testable. Then $C\mathbf{b}$ and s^2 are independent.

Proof.

$$
\begin{aligned}
BVA &= C(X^TX)^cX^T\sigma^2 I[I-H] \\
&= [C(X^TX)^cX^T - C(X^TX)^cX^TH]\sigma^2 \\
&= [C(X^TX)^cX^T - C(X^TX)^cX^T]\sigma^2 \\
&= 0,
\end{aligned}
$$

so $C\mathbf{b}$ is independent of SS_{Res}, and hence of s^2.

Testability

Remember our test statistic is

$$\frac{(C\mathbf{b})^T[C(X^TX)^cC^T]^{-1}C\mathbf{b}/m}{s^2}.$$

If the null hypothesis is true and $C\boldsymbol{\beta} = \mathbf{0}$, then it has an F distribution, with m and $n - r$ degrees of freedom.

If the null hypothesis is false, then since $[C(X^TX)^cC^T]^{-1}$ is positive definite (proof required), the mean of the numerator will increase. Thus we use a right-tailed test.

Carbon removal example

Example. Let us look at the carbon removal example from the previous section. We compare three methods of removing carbon from wastewater. The data is:

AF	FS	FCC
34.6	38.8	26.7
35.1	39.0	26.7
35.3	40.1	27.0

Carbon removal example

We test whether the populations have the same mean, i.e. $H_0 : \tau_1 = \tau_2 = \tau_3$. This can be written in matrix form as $H_0 : C\boldsymbol{\beta} = \mathbf{0}$, where

$$C = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}.$$

```
> library(MASS)
> n <- 9
> r <- 3
> y <- c(34.6,35.1,35.3,38.8,39.0,40.1,26.7,26.7,27.0)
> X <- matrix(c(rep(1,n),rep(0,27)),n,r+1)
> X[1:3,2] <- 1
> X[4:6,3] <- 1
> X[7:9,4] <- 1
```

Carbon removal example

```
> (b <- ginv(t(X)%*%X)%*%t(X)%*%y)
       [,1]
[1,] 25.275
[2,]  9.725
[3,] 14.025
[4,]  1.525
> (C <- matrix(c(0,0,1,1,-1,0,0,-1),2,4))
     [,1] [,2] [,3] [,4]
[1,]    0    1   -1    0
[2,]    0    1    0   -1
> (numer <- t(C%*%b) %*%
+        solve(C %*% ginv(t(X)%*%X)%*%t(C)) %*% C%*%b)
       [,1]
[1,] 241.98
```

Carbon removal example

```
> e <- y - X%*%b
> (s2 <- sum(e^2)/(n-r))

[1] 0.2166667

> (Fstat <- (numer/2)/s2)

         [,1]
[1,] 558.4154

> pf(Fstat, 2, n-r, lower=F)

             [,1]
[1,] 1.525846e-07
```

We can reject H_0 firmly, so the populations are not all the same. It is still possible that *some* of the populations are the same, but not all of them.

One-factor model

In a one-factor model (one-way classification), if we look closer at our F statistic, we can see that we do not need all of the individual data to test hypotheses. This is because we have simplified formulas for various quantities.

In particular, we can write $X^T X$ as

$$X^T X = \begin{bmatrix} n & n_1 & n_2 & \dots & n_k \\ n_1 & n_1 & 0 & \dots & 0 \\ n_2 & 0 & n_2 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ n_k & 0 & 0 & \dots & n_k \end{bmatrix}$$

One-factor model

It is easy to check that $r(X) = k$, and in particular

$$(X^TX)^c = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & \frac{1}{n_1} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{n_k} \end{bmatrix}.$$

Also we have

$$X^T\mathbf{y} = \begin{bmatrix} \sum_{ij} y_{ij} \\ \sum_j y_{1j} \\ \sum_j y_{2j} \\ \vdots \\ \sum_j y_{kj} \end{bmatrix}.$$

One-factor model

Multiplying out gives us

$$\mathbf{b} = (X^T X)^c X^T \mathbf{y} = \begin{bmatrix} 0 \\ \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_k \end{bmatrix}.$$

This means that if we know or are given s^2, the only other information that we need to test our hypotheses is the means, and number, of the samples from the various populations. We do not need the samples themselves.

Tennis ball example

Example. A tennis ball manufacturer is studying the life span of a newly developed tennis ball on five different surfaces. The response is the number of hours that the ball is used before it is judged to be dead. A study is conducted and the following data obtained:

Surface	Clay	Grass	Composition	Wood	Asphalt
Mean	6.2	6.8	6.4	5	4.4
Number	20	22	24	21	25

We are also given $s^2 = 8.87$.

Tennis ball example

We test if the lifespan of the ball is different on hard surfaces (wood and asphalt) vs. soft surfaces.

This gives the hypothesis

$$H_0 : \frac{1}{3}(\tau_1 + \tau_2 + \tau_3) - \frac{1}{2}(\tau_4 + \tau_5) = 0$$

with corresponding matrix

$$C = \left[\begin{array}{cccccc} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{2} & -\frac{1}{2} \end{array}\right].$$

Tennis ball example

```
> n <- 112
> r <- 5
> (XtXc <- diag(c(0,1/c(20,22,24,21,25))))
     [,1] [,2]       [,3]       [,4]       [,5] [,6]
[1,]    0 0.00 0.00000000 0.00000000 0.00000000 0.00
[2,]    0 0.05 0.00000000 0.00000000 0.00000000 0.00
[3,]    0 0.00 0.04545455 0.00000000 0.00000000 0.00
[4,]    0 0.00 0.00000000 0.04166667 0.00000000 0.00
[5,]    0 0.00 0.00000000 0.00000000 0.04761905 0.00
[6,]    0 0.00 0.00000000 0.00000000 0.00000000 0.04
> b <- c(0,6.2,6.8,6.4,5,4.4)
> s2 <- 8.87
```

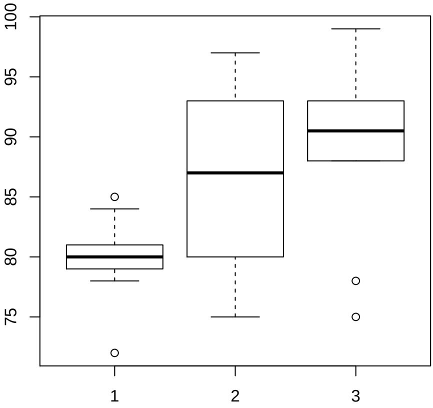
Tennis ball example

```
> C <- matrix(c(0,rep(1/3,3),rep(-1/2,2)),1,6)
> (Fstat <- ( t(C%*%b) %*% solve(C%*%XtXc%*%t(C))
+   %*% C%*%b )/s2)

        [,1]
[1,] 9.47411

> pf(Fstat, 1, n-r, lower=F)

            [,1]
[1,] 0.002647318
```

We can reject the null hypothesis that the ball lasts as long on hard and soft surfaces.

Maths marks example

We return to our running example of mathematics marks. We fit a one-factor model using the class of the students.



Maths marks example

Does the class have any effect?

This is equivalent to the hypothesis $H_0 : \tau_1 = \tau_2 = \tau_3$.

```
> (C <- matrix(c(0,1,-1,0,0,0,1,-1),2,4,byrow=TRUE))

     [,1] [,2] [,3] [,4]
[1,]    0    1   -1    0
[2,]    0    0    1   -1

> round(C %*% XtXc %*% t(X) %*% X,3) #testable?

     [,1] [,2] [,3] [,4]
[1,]    0    1   -1    0
[2,]    0    0    1   -1
```

Maths marks example — $H_0 : \tau_1 = \tau_2 = \tau_3$

```
> (SS <- t(C %*% b) %*% solve(C %*% XtXc %*% t(C)) %*% C %*% b)

         [,1]
[1,] 474.0667

> s2

[1] 42.14074

> (Fstat <- (SS/2)/s2)

         [,1]
[1,] 5.624802

> pf(Fstat, 2, n-r, lower.tail=FALSE)

            [,1]
[1,] 0.009077098
```

Maths marks example — $H_0 : \tau_1 = \tau_2 = \tau_3$

```
> contrasts(maths$class.f) <- contr.treatment(3)
> model <- lm(maths.y ~ class.f, data=maths)
> summary(model)

Call:
lm(formula = maths.y ~ class.f, data = maths)

Residuals:
   Min     1Q Median     3Q    Max
-14.40  -1.80   0.85   3.60  10.50

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   79.900      2.053  38.922  < 2e-16 ***
class.f2       6.600      2.903   2.273  0.03117 *
class.f3       9.500      2.903   3.272  0.00292 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.492 on 27 degrees of freedom
Multiple R-squared:  0.2941, Adjusted R-squared:  0.2418
F-statistic: 5.625 on 2 and 27 DF,  p-value: 0.009077
```

Maths marks example — $H_0 : \tau_1 = \tau_2 = \tau_3$

```
> anova(model)
Analysis of Variance Table

Response: maths.y
          Df  Sum Sq Mean Sq F value   Pr(>F)
class.f    2  474.07 237.033  5.6248 0.009077 **
Residuals 27 1137.80  42.141
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```

Maths marks example — $H_0 : \tau_1 = \tau_2 = \tau_3$

```
> basemodel <- lm(maths.y ~ 1, data=maths)
> anova(basemodel, model)

Analysis of Variance Table

Model 1: maths.y ~ 1
Model 2: maths.y ~ class.f
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1     29 1611.9
2     27 1137.8  2    474.07 5.6248 0.009077 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```

Maths marks example — $H_0 : 2\tau_2 = \tau_1 + \tau_3$

```
> C <- matrix(c(0,-1,2,-1),1,4)
> (SS <- t(C %*% b) %*% solve(C %*% XtXc %*% t(C)) %*% C %*% b)

         [,1]
[1,] 22.81667

> (Fstat <- (SS/1)/s2)

          [,1]
[1,] 0.5414396

> pf(Fstat, 1, n-r, lower=F)

          [,1]
[1,] 0.4681814
```

Maths marks example — $H_0 : 2\tau_2 = \tau_1 + \tau_3$

The hypothesis is equivalent to
$0(\mu + \tau_1) + 2(\tau_2 - \tau_1) - (\tau_3 - \tau_1) = 0$.

```
> library(car)
> linearHypothesis(model, c(0,2,-1), 0)

Linear hypothesis test

Hypothesis:
2 class.f2 - class.f3 = 0

Model 1: restricted model
Model 2: maths.y ~ class.f

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     28 1160.6
2     27 1137.8  1    22.817 0.5414 0.4682
```

Maths marks example — $H_0 : 2\tau_2 = \tau_1 + \tau_3$

The hypothesis is also equivalent to
$0\bar{\mu} + 0(\mu_1 - \bar{\mu}) + 3(\mu_2 - \bar{\mu}) = 0$.

```
> contrasts(maths$class.f) <- contr.sum(3)
> modelsum <- lm(maths.y ~ class.f, data=maths)
> linearHypothesis(modelsum, c(0,0,3), 0)

Linear hypothesis test

Hypothesis:
3 class.f2 = 0

Model 1: restricted model
Model 2: maths.y ~ class.f

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     28 1160.6
2     27 1137.8  1    22.817 0.5414 0.4682
```

Two-factor models

In this section, we look at two-factor models (two-way classification), but the ideas extend easily any number of factors.

In a basic two-factor model, we assume that each level of each factor affects the overall mean μ by a specific amount. We name these effects τ_i for the ith level of factor 1 and β_j for the jth level of factor 2. Then our model is

$$y_{ijk} = \mu+\tau_i+\beta_j+\varepsilon_{ijk}, \qquad i = 1, \ldots, a, j = 1, \ldots, b, k = 1, \ldots, n_{ij}.$$

In matrix form (with 1 sample from each combination of factor levels):

Two-factor models

$$
X = \left[\begin{array}{c|cccc|cccc}
1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\
1 & 1 & 0 & \dots & 0 & 0 & 1 & & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & & \ddots & \vdots \\
1 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 1 \\
\hline
1 & 0 & 1 & \dots & 0 & 1 & 0 & \dots & 0 \\
1 & 0 & 1 & \dots & 0 & 0 & 1 & & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & & \ddots & \vdots \\
1 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 1 \\
\hline
\vdots & & & & & & & & \\
\hline
1 & 0 & 0 & \dots & 1 & 1 & 0 & \dots & 0 \\
1 & 0 & 0 & \dots & 1 & 0 & 1 & & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & & \ddots & \vdots \\
1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1
\end{array}\right], \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_a \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_b \end{bmatrix}
$$

Two-factor models

This model is known as the *additive* model. It assumes that the effects from each factor can be added to produce the overall effect.

Any hypothesis that we can test in a one-factor model can be tested for each factor in an additive two-factor model.

Theorem 7.4

In an additive two-factor model, every contrast in the τ's is estimable. Similarly, every contrast in the β's is estimable.

Two-factor models

The most common hypotheses that we will want to test are

$$\tau_1 = \tau_2 = \ldots = \tau_a$$

and

$$\beta_1 = \beta_2 = \ldots = \beta_b.$$

Because they are all composed of treatment contrasts for one factor, they are testable. We can use the theory already developed to test them.

Two-factor models

Example. We model the time taken to dissolve a capsule in a biological fluid. A study is conducted with 1 sample from each combination of factor levels and the following data found:

Time		Fluid type	
		Gastric	Duodenal
Capsule	A	39.5	31.2
type	B	47.4	44

Two-factor models

The linear model is

$$\mathbf{y} = \begin{bmatrix} 39.5 \\ 47.4 \\ 31.2 \\ 44 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \beta_1 \\ \beta_2 \end{bmatrix}.$$

We test the hypotheses that there is no difference in the response for the levels of each of the two factors.

Two-factor models

The first factor (fluid type) gives the hypothesis

$$H_0 : \tau_1 = \tau_2 \text{ or } \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \end{bmatrix} \boldsymbol{\beta} = \mathbf{0}.$$

```
> y <- c(39.5,47.4,31.2,44)
> X <- matrix(c(1,1,1,1,1,1,0,0,0,0,1,1,1,0,1,0,0,1,0,1),
+     4,5)
> n <- 4
> r <- 3
> b <- ginv(t(X)%*%X) %*% t(X)%*%y
> s2 <- sum((y - X%*%b)^2)/(n-r)
> (C <- matrix(c(0,1,-1,0,0),1,5))

     [,1] [,2] [,3] [,4] [,5]
[1,]    0    1   -1    0    0
```

Two-factor models

```
> (Fstat <- t(C%*%b) %*% solve(C %*% ginv(t(X)%*%X)
+  %*% t(C)) %*% C%*%b / s2)

         [,1]
[1,] 5.701374

> pf(Fstat, 1, n-r, lower=F)

        [,1]
[1,] 0.25249
```

We cannot reject the null hypothesis.

Two-factor models

The second factor (capsule type) gives the hypothesis

$$H_0 : \beta_1 = \beta_2 \text{ or } \left[\begin{array}{ccccc} 0 & 0 & 0 & 1 & -1 \end{array}\right] \boldsymbol{\beta} = \mathbf{0}.$$

```
> C <- matrix(c(0,0,0,1,-1),1,5)
> (Fstat <- t(C%*%b) %*% solve(C %*% ginv(t(X)%*%X)
+  %*% t(C)) %*% C%*%b / s2)

         [,1]
[1,] 17.84631

> pf(Fstat, 1, n-r, lower=F)

          [,1]
[1,] 0.1479737
```

We cannot reject the null hypothesis either.

Interaction

In some cases, it is possible that *interaction* between factors may occur.

Interaction happens when one factor affects the effect of another factor.

For example, if the effect of factor 1 when factor 2 is at level 1 is different from the effect of factor 1 when factor 2 is at level 2, then there is interaction.

Interaction

Example. Suppose that we are studying the effect of pressure and temperature on viscosity, and the *actual* means of the response variable for each of the combinations are given by:

		Pressure			
		1	2	3	4
Temperature	1	4	6	4	3
	2	8	2	7	5

Interaction

When the pressure is at level 1, changing the temperature from level 1 to level 2 results in an increase of viscosity of 4.

However, when the pressure is at level 2, changing the temperature from level 1 to level 2 results in a *decrease* of viscosity of 4!

In this case, the factors interact.

Interaction

If, on the other hand, the actual means were:

		Pressure			
		1	2	3	4
Temperature	1	4	6	4	3
	2	8	10	8	7

then there would be no interaction between the factors. Even though the factors themselves are significant, the *combination* of factor levels has no effect apart from the individual factor effects.

Interaction

An additive model assumes that there is no interaction between the factors, so the effects of the factor levels can be measured in isolation from the other factor(s).

If we have interaction, or want to test whether there is interaction, we must use a different model:

$$y_{ijk} = \mu + \tau_i + \beta_j + \xi_{ij} + \varepsilon_{ijk},$$

where ξ_{ij} is an interaction term which quantifies the effect of factor 1 being at level i at the same time that factor 2 is at level j.

Interaction

Example. Consider the previous example (dissolving a capsule in fluid). If we allow an interaction term, $\mathbf{y}$ stays the same, but the linear model becomes

$$X = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \beta_1 \\ \beta_2 \\ \xi_{11} \\ \xi_{12} \\ \xi_{21} \\ \xi_{22} \end{bmatrix}.$$

Interaction

In a two-factor model with interaction, we are often interested in testing whether there is interaction or not.

However, testing the presence of interaction is not quite as straightforward as it may seem. It seems like we would want to test the hypothesis

$$H_0 : \xi_{11} = \xi_{12} = \ldots = \xi_{1b} = \xi_{21} = \ldots = \xi_{ab},$$

but it turns out that this is not correct. We illustrate with an example.

Interaction

Example. Suppose we have a two-factor model with the following actual means:

		Factor I	
		1	2
Factor II	1	6	5
	2	6	5

There is clearly no interaction between the factors.

Interaction

One possible parameter set is

$$\mu = 0, \tau_1 = 5, \tau_2 = 4, \beta_1 = \beta_2 = 1,$$

$$\xi_{ij} = 0 \qquad \forall i, j.$$

However, an equally valid parameter set is

$$\mu = 0, \tau_1 = 2, \tau_2 = 1, \beta_1 = 3, \beta_2 = 2,$$

$$\xi_{11} = 1, \xi_{12} = 2, \xi_{21} = 1, \xi_{22} = 2.$$

Interaction

Thus, while $\xi_{ij} = 0$ for all i, j implies no interaction, it is not actually necessary.

Moreover, the hypothesis $H_0 : \xi_{ij} = 0 \, \forall i, j$ is not even testable.

Nor is $H_0 : \xi_{ij}$ the same $\forall i, j$.

Example

Consider a two-way classification where each factor has two levels, with one observation from each combination of levels. We have

$$
\begin{aligned}
X &= \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \\
(X^TX)^c &= \left[\begin{array}{c|c} 0_{5\times 5} & 0_{5\times 4} \\ \hline 0_{4\times 5} & I_4 \end{array}\right] \\
(X^TX)^cX^TX &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.
\end{aligned}
$$

Example

We can write the hypothesis $H_0 : \xi_{ij}$ the same $\forall i, j$ as $C\boldsymbol{\beta} = \mathbf{0}$ where

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}.$$

This hypothesis is not testable since

$$C(X^TX)^cX^TX = \begin{bmatrix} 0 & 0 & 0 & 1 & -1 & 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 1 & -1 & 1 & 0 & 0 & -1 \end{bmatrix} \neq C.$$

Interaction

Theorem 7.5

For the linear model

$$y_{ijk} = \mu + \tau_i + \beta_j + \xi_{ij} + \varepsilon_{ijk},$$

there is no interaction if and only if

$$(\xi_{ij} - \xi_{ij'}) - (\xi_{i'j} - \xi_{i'j'}) = 0,$$

for all $i \neq i', j \neq j'$.

Moreover these quantities are all estimable.

Interaction

Proof. A sample which has level i of factor 1 and level j of factor 2 has mean

$$\mu_{ij} = \mu + \tau_i + \beta_j + \xi_{ij}.$$

Now take two levels of factor 1 (i and i') and two levels of factor 2 (j and j').

No interaction between these levels means the difference in means that results from switching factor 2 from j to j' is the same whether factor 1 is at level i or i'.

Interaction

In other words,

$$\mu_{ij} - \mu_{ij'} = \mu_{i'j} - \mu_{i'j'},$$

and expanding gives

$$\mu+\tau_i+\beta_j+\xi_{ij}-\mu-\tau_i-\beta_{j'}-\xi_{ij'} = \mu+\tau_{i'}+\beta_j+\xi_{i'j}-\mu-\tau_{i'}-\beta_{j'}-\xi_{i'j'}$$

which reduces to

$$(\xi_{ij} - \xi_{ij'}) - (\xi_{i'j} - \xi_{i'j'}) = 0.$$

In order for there to be no interaction at all, we need this condition to hold for all i, i', j, j'.

The group means μ_{ij} are all elements of $X\boldsymbol{\beta}$, and thus linear combinations of them are estimable.

Interaction

Theorem 7.5 generates $ab(a-1)(b-1)$ equations. However, it can be shown that all but $(a-1)(b-1)$ of them are redundant.

Example. In a two-factor design with two levels in each factor, Theorem 7.5 shows that there is no interaction if and only if

$$\begin{aligned}
(\xi_{11}-\xi_{12})-(\xi_{21}-\xi_{22}) &= 0\\
(\xi_{21}-\xi_{22})-(\xi_{11}-\xi_{12}) &= 0\\
(\xi_{12}-\xi_{11})-(\xi_{22}-\xi_{21}) &= 0\\
(\xi_{12}-\xi_{21})-(\xi_{12}-\xi_{11}) &= 0.
\end{aligned}$$

Interaction

It is easy to see that all of these equations are equivalent, so we need only test one.

This gives the hypothesis $H_0 : C\boldsymbol{\beta} = \mathbf{0}$, where

$$\begin{aligned} C &= \left[\begin{array}{ccccccccc} 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \end{array} \right], \\ \boldsymbol{\beta} &= \left[\begin{array}{ccccccccc} \mu & \tau_1 & \tau_2 & \beta_1 & \beta_2 & \xi_{11} & \xi_{12} & \xi_{21} & \xi_{22} \end{array} \right]^T. \end{aligned}$$

Interaction considerations

Some things to consider when testing for interaction:

1) If we have one sample per combination of factors, it is impossible to account for or test for interaction.

This is because $r(X) = n$ and therefore $n - r$, the residual degrees of freedom, is 0.

Essentially we treat each combination of factors as a separate population. If we have one sample from each population, then we have no way to estimate the variance!

Interaction considerations

2) If we test for interaction and find that there is none, we theoretically should still use the residual sum of squares from the full model with interaction, unless there is a convincing data-related reason to think that there is no interaction.

This follows from the same reasoning as using SS_{Res} for the full model in sequential tests: we cannot be sure that there is no interaction, we just haven't found any!

However, for practical purposes, this may take away too many degrees of freedom from SS_{Res}. So if you find no interaction, it's OK to use the SS_{Res} from an additive model.

Interaction considerations

3) It is possible to have interaction between three or more factors.

However, this is hard to test for and hard to interpret. In practice most people only look at two-factor interactions.

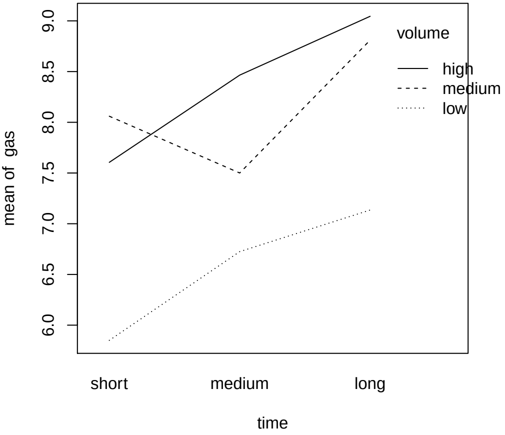
Engine example

We look at the effect of pre-chamber volume ratio and injection timing on the emission of noxious gas from an engine. The factors have 3 levels each.

```
> str(engine)

'data.frame': 18 obs. of  3 variables:
 $ gas   : num  6.27 8.08 7.34 5.43 8.04 7.87 6.94 7.48 8.61 6.5
 $ volume: Factor w/ 3 levels "low","medium",..: 1 2 3 1 2 3 1 2
 $ time  : Factor w/ 3 levels "short","medium",..: 1 1 1 1 1 1 2

> means

      [,1]  [,2]  [,3]
[1,] 5.850 6.725 7.135
[2,] 8.060 7.500 8.810
[3,] 7.605 8.465 9.045
```

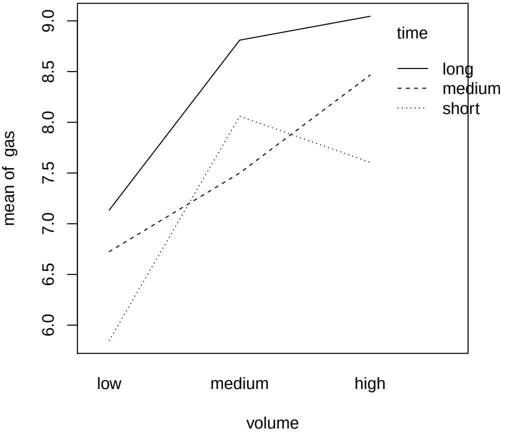
Engine example

```
> with(engine, interaction.plot(time, volume, gas))
```



Engine example

```
> with(engine, interaction.plot(volume, time, gas))
```



Engine example

Without interaction (additive model):

```
> y <- engine$gas
> n <- length(y)
> X <- matrix(c(rep(1,n),rep(0,n*6)),n,7)
> X[cbind(1:n,as.numeric(engine$volume)+1)] <- 1
> X[cbind(1:n,as.numeric(engine$time)+4)] <- 1
> X
      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
 [1,]    1    1    0    0    1    0    0
 [2,]    1    0    1    0    1    0    0
 [3,]    1    0    0    1    1    0    0
 [4,]    1    1    0    0    1    0    0
 [5,]    1    0    1    0    1    0    0
 [6,]    1    0    0    1    1    0    0
 [7,]    1    1    0    0    0    1    0
 [8,]    1    0    1    0    0    1    0
 [9,]    1    0    0    1    0    1    0
[10,]    1    1    0    0    0    1    0
[11,]    1    0    1    0    0    1    0
[12,]    1    0    0    1    0    1    0
[13,]    1    1    0    0    0    0    1
[14,]    1    0    1    0    0    0    1
```

Testing differences among populations

Testing $\tau_1 = \tau_2 = \tau_3$:

```
> r <- rankMatrix(X)[1]
> XtXc <- ginv(t(X)%*%X)
> b <- XtXc%*%t(X)%*%y
> s2 <- sum((y-X%*%b)^2)/(n-r)
> (C <- matrix(c(0,1,-1,rep(0,6),1,-1,rep(0,3)),2,7,byrow=T

     [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]    0    1   -1    0    0    0    0
[2,]    0    0    1   -1    0    0    0

> round(C %*% XtXc %*% t(X) %*% X,3) #testable?

     [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]    0    1   -1    0    0    0    0
[2,]    0    0    1   -1    0    0    0
```

Testing differences among populations

```
> (Fstat <- (t(C%*%b)%*%solve(C%*%XtXc%*%t(C))%*%
+    C%*%b/2)/s2)
         [,1]
[1,] 35.72603
> pf(Fstat,2,n-r,lower=F)
             [,1]
[1,] 5.219933e-06
```

Testing differences among populations

Testing $\beta_1 = \beta_2 = \beta_3$:

```
> (C <- matrix(c(rep(0,4),1,-1,rep(0,6),1,-1),2,7,byrow=T))
     [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]    0    0    0    0    1   -1    0
[2,]    0    0    0    0    0    1   -1
> (Fstat <- (t(C%*%b)%*%solve(C%*%XtXc%*%t(C))%*%
+    C%*%b/2)/s2)
         [,1]
[1,] 13.00833
> pf(Fstat,2,n-r,lower=F)
             [,1]
[1,] 0.0007897809
```

Testing for interaction

```
> X <- matrix(c(rep(1,n),rep(0,n*15)),n,7+9)
> X[cbind(1:n,as.numeric(engine$volume)+1)] <- 1
> X[cbind(1:n,as.numeric(engine$time)+4)] <- 1
> X[cbind(1:n,as.numeric(engine$time)*3+as.numeric(engine$volume)+4)] <- 1
> X[,-(1:7)]
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
 [1,]    1    0    0    0    0    0    0    0    0
 [2,]    0    1    0    0    0    0    0    0    0
 [3,]    0    0    1    0    0    0    0    0    0
 [4,]    1    0    0    0    0    0    0    0    0
 [5,]    0    1    0    0    0    0    0    0    0
 [6,]    0    0    1    0    0    0    0    0    0
 [7,]    0    0    0    1    0    0    0    0    0
 [8,]    0    0    0    0    1    0    0    0    0
 [9,]    0    0    0    0    0    1    0    0    0
[10,]    0    0    0    1    0    0    0    0    0
[11,]    0    0    0    0    1    0    0    0    0
[12,]    0    0    0    0    0    1    0    0    0
[13,]    0    0    0    0    0    0    1    0    0
[14,]    0    0    0    0    0    0    0    1    0
[15,]    0    0    0    0    0    0    0    0    1
[16,]    0    0    0    0    0    0    1    0    0
[17,]    0    0    0    0    0    0    0    1    0
```

Testing for interaction

```
> (r <- rankMatrix(X)[1])
[1] 9
> XtXc <- ginv(t(X)%*%X)
> b <- XtXc%*%t(X)%*%y
> s2 <- sum((y-X%*%b)^2)/(n-r)
```

Testing for interaction

```
> C <- matrix(0,4,16)
> C[1,c(8,9,11,12)] <- c(1,-1,-1,1)
> C[2,c(9,10,12,13)] <- c(1,-1,-1,1)
> C[3,c(11,12,14,15)] <- c(1,-1,-1,1)
> C[4,c(12,13,15,16)] <- c(1,-1,-1,1)
> C[,-(1:7)]
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,]    1   -1    0   -1    1    0    0    0    0
[2,]    0    1   -1    0   -1    1    0    0    0
[3,]    0    0    0    1   -1    0   -1    1    0
[4,]    0    0    0    0    1   -1    0   -1    1
```

Testing for interaction

```
> (Fstat <- (t(C%*%b)%*%solve(C%*%XtXc%*%t(C))%*%
+    C%*%b/4)/s2)
        [,1]
[1,] 4.47684
> pf(Fstat,4,n-r,lower=F)
           [,1]
[1,] 0.02891813
```

Engine example

For an additive model we have

$$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}.$$

Let $\mu_{ij} = \mu + \tau_i + \beta_j$ (estimable). Then R estimates the following (for `contr.treatment`):

Label	Estimated quantity
Intercept	$\mu_{11} = \mu + \tau_1 + \beta_1$
f2	$\mu_{21} - \mu_{11} = \tau_2 - \tau_1$
f3	$\mu_{31} - \mu_{11} = \tau_3 - \tau_1$
g2	$\mu_{12} - \mu_{11} = \beta_2 - \beta_1$
g3	$\mu_{13} - \mu_{11} = \beta_3 - \beta_1$

Engine example

```
> model <- lm(gas ~ volume + time, data=engine)
> summary(model)
Call:
lm(formula = gas ~ volume + time, data = engine)

Residuals:
     Min       1Q   Median       3Q      Max
-0.62333 -0.15000  0.03583  0.21375  0.49500

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.0533     0.2109  28.703 3.83e-13 ***
volumemedium  1.5533     0.2310   6.724 1.42e-05 ***
volumehigh    1.8017     0.2310   7.798 2.95e-06 ***
timemedium    0.3917     0.2310   1.695 0.113817
timelong      1.1583     0.2310   5.014 0.000237 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Engine example

```
> b[1]+b[2]+b[5]
[1] 6.053333
> b[3]-b[2]
[1] 1.553333
> b[4]-b[2]
[1] 1.801667
> b[6]-b[5]
[1] 0.3916667
> b[7]-b[5]
[1] 1.158333
```

Testing differences among populations

To test $\tau_1 = \tau_2 = \tau_3$ and $\beta_1 = \beta_2 = \beta_3$:

```
> anova(model)

Analysis of Variance Table

Response: gas
          Df  Sum Sq Mean Sq F value    Pr(>F)
volume     2 11.4410  5.7205  35.726  5.22e-06 ***
time       2  4.1658  2.0829  13.008 0.0007898 ***
Residuals 13  2.0816  0.1601
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```

Engine example

```
> basemodel <- lm(gas ~ 1, data=engine)
> model2 <- lm(gas ~ time, data=engine)
> anova(basemodel,model2, model)
```

```
Analysis of Variance Table

Model 1: gas ~ 1
Model 2: gas ~ time
Model 3: gas ~ volume + time
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     17 17.6885
2     15 13.5226  2    4.1658 13.008 0.0007898 ***
3     13  2.0816  2   11.4410 35.726  5.22e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

Engine example

```
> model3 <- lm(gas ~ volume, data=engine)
> anova(basemodel,model3, model)
```

```
Analysis of Variance Table

Model 1: gas ~ 1
Model 2: gas ~ volume
Model 3: gas ~ volume + time
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     17 17.6885
2     15  6.2474  2   11.4410 35.726  5.22e-06 ***
3     13  2.0816  2    4.1658 13.008 0.0007898 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```

Interaction

If we have interaction, we use the model

$$y_{ijk} = \mu + \tau_i + \beta_j + \xi_{ij} + \epsilon_{ijk}.$$

Let $\mu_{ij} = \mu + \tau_i + \beta_j + \xi_{ij}$. Then R estimates the following:

Intercept	$\mu_{11} = \mu + \tau_1 + \beta_1 + \xi_{11}$
f2	$\mu_{21} - \mu_{11} = \tau_2 - \tau_1 + \xi_{21} - \xi_{11}$
f3	$\mu_{31} - \mu_{11} = \tau_3 - \tau_1 + \xi_{31} - \xi_{11}$
g2	$\mu_{12} - \mu_{11} = \beta_2 - \beta_1 + \xi_{12} - \xi_{11}$
g3	$\mu_{13} - \mu_{11} = \beta_3 - \beta_1 + \xi_{13} - \xi_{11}$
f2:g2	$\mu_{22} - \mu_{21} - \mu_{12} + \mu_{11} = \xi_{22} - \xi_{21} - \xi_{12} + \xi_{11}$
f3:g2	$\mu_{32} - \mu_{31} - \mu_{12} + \mu_{11} = \xi_{32} - \xi_{31} - \xi_{12} + \xi_{11}$
f2:g3	$\mu_{23} - \mu_{21} - \mu_{13} + \mu_{11} = \xi_{23} - \xi_{21} - \xi_{13} + \xi_{11}$
f3:g3	$\mu_{33} - \mu_{31} - \mu_{13} + \mu_{11} = \xi_{33} - \xi_{31} - \xi_{13} + \xi_{11}$

Testing f2:g2 = f3:g2 = f2:g3 = f3:g3 = 0 is the test for no interaction.

Interaction

```
> imodel <- lm(gas ~ volume * time, data=engine)
> summary(imodel)
Call:
lm(formula = gas ~ volume * time, data = engine)

Residuals:
   Min     1Q Median     3Q    Max
 -0.42  -0.13   0.00   0.13   0.42

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)               5.8500     0.1967  29.745 2.68e-10 ***
volumemedium              2.2100     0.2781   7.946 2.34e-05 ***
volumehigh                1.7550     0.2781   6.310 0.000139 ***
timemedium                0.8750     0.2781   3.146 0.011815 *
timelong                  1.2850     0.2781   4.620 0.001254 **
volumemedium:timemedium  -1.4350     0.3933  -3.648 0.005333 **
volumehigh:timemedium    -0.0150     0.3933  -0.038 0.970413
volumemedium:timelong    -0.5350     0.3933  -1.360 0.206882
volumehigh:timelong       0.1550     0.3933   0.394 0.702715
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Interaction

```
> b[1]+b[2]+b[5]+b[8]
[1] 5.85
> b[c(3,4)] - b[2] + b[c(9,10)] - b[8]
[1] 2.210 1.755
> b[c(6,7)] - b[5] + b[c(11,14)] - b[8]
[1] 0.875 1.285
> b[c(12,13,15,16)] - b[c(9,10,9,10)] -
+  b[c(11,11,14,14)] + b[c(8,8,8,8)]
[1] -1.435 -0.015 -0.535  0.155
```

Testing for interaction

```
> anova(imodel)
Analysis of Variance Table

Response: gas
            Df  Sum Sq Mean Sq F value    Pr(>F)
volume       2 11.4410  5.7205 73.9456 2.594e-06 ***
time         2  4.1658  2.0829 26.9246 0.0001591 ***
volume:time  4  1.3853  0.3463  4.4768 0.0289181 *
Residuals    9  0.6962  0.0774
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```

Testing for interaction

```
> anova(model, imodel)

Analysis of Variance Table

Model 1: gas ~ volume + time
Model 2: gas ~ volume * time
  Res.Df     RSS Df Sum of Sq      F  Pr(>F)
1     13 2.08158
2      9 0.69625  4    1.3853 4.4768 0.02892 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```

ANCOVA

We can also do analysis of covariance (ANCOVA) using the linear model framework.

In this case we have one (or more) categorical predictors and one (or more) continuous predictors. For example:

$$y_{ij} = \mu + \tau_i + \beta x_{ij} + \xi_i x_{ij} + \varepsilon_{ij}.$$

We can think of this simple model as fitting several regression lines, one to each population (assuming equal variances across populations).

ANCOVA

Interaction in this case means that the slopes of the regression lines (effect of continuous predictor) are different for each population.

A model without interaction assumes that the slopes are the same (but the intercepts may be different):

$$y_{ij} = \mu + \tau_i + \beta x_{ij} + \varepsilon_{ij}.$$

We fit these models using the less than full rank model.

ANCOVA

Suppose we fit the lines $y = \alpha_i + \beta_i x$ to each subpopulation. For the interaction model, R estimates:

Intercept	$\alpha_1 = \mu + \tau_1$
f2	$\alpha_2 - \alpha_1 = \tau_2 - \tau_1$
f3	$\alpha_3 - \alpha_1 = \tau_3 - \tau_1$
x	$\beta_1 = \beta + \xi_1$
f2:x	$\beta_2 - \beta_1 = \xi_2 - \xi_1$
f3:x	$\beta_3 - \beta_1 = \xi_3 - \xi_1$

Exam marks example

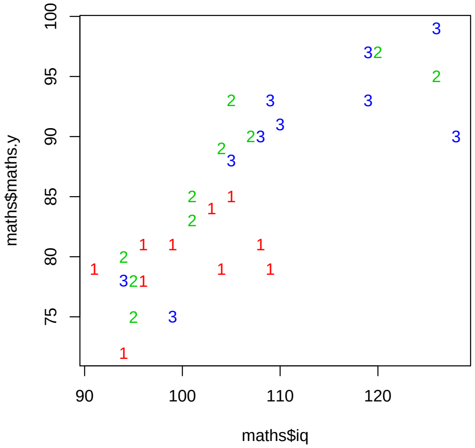
The `maths` dataset also has another component: the IQ of the student.

```
> str(maths)
```

```
'data.frame': 30 obs. of  5 variables:
 $ X      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ maths.y: int  81 84 81 79 78 79 81 85 72 79 ...
 $ iq     : int  99 103 108 109 96 104 96 105 94 91 ...
 $ class  : int  1 1 1 1 1 1 1 1 1 1 ...
 $ class.f: Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1
  ..- attr(*, "contrasts")= num [1:3, 1:2] 0 1 0 0 0 1
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr  "1" "2" "3"
  .. .. ..$ : chr  "2" "3"
```

Exam marks example

```
> plot(maths$iq, maths$maths.y, pch=array(maths$class.f),
+  col=maths$class+1)
```



Exam marks example

```
> model <- lm(maths.y ~ class.f * iq, data=maths)
> summary(model)

Call:
lm(formula = maths.y ~ class.f * iq, data = maths)

Residuals:
    Min      1Q  Median      3Q     Max
-8.2507 -1.8312  0.9807  2.4711  6.3765

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  52.7577    21.9941   2.399   0.0246 *
class.f2    -30.9642    25.7058  -1.205   0.2401
class.f3    -24.0093    25.8357  -0.929   0.3620
iq            0.2701     0.2185   1.236   0.2284
class.f2:iq   0.3474     0.2524   1.376   0.1815
class.f3:iq   0.2729     0.2497   1.093   0.2852
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Exam marks example

```
> amodel <- lm(maths.y ~ class.f + iq, data = maths)
> anova(amodel, model)

Analysis of Variance Table

Model 1: maths.y ~ class.f + iq
Model 2: maths.y ~ class.f * iq
  Res.Df    RSS Df Sum of Sq    F Pr(>F)
1     26 423.42
2     24 392.36  2    31.062 0.95 0.4008
```

Interaction is not significant, so we remove the interaction term and fit an additive model.

Exam marks example

```
> summary(amodel)

Call:
lm(formula = maths.y ~ class.f + iq, data = maths)

Residuals:
   Min     1Q Median     3Q    Max
-8.137 -2.842  1.220  2.662  6.393

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 26.02809    8.23338   3.161  0.00396 **
class.f2     4.29503    1.83799   2.337  0.02743 *
class.f3     3.49636    2.01959   1.731  0.09526 .
iq           0.53604    0.08093   6.623 5.03e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.036 on 26 degrees of freedom
Multiple R-squared:  0.7373, Adjusted R-squared:  0.707
```

Exam marks example

```
> basemodel <- lm(maths.y ~ class.f, data = maths)
> anova(basemodel,amodel)
```

```
Analysis of Variance Table

Model 1: maths.y ~ class.f
Model 2: maths.y ~ class.f + iq
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     27 1137.80
2     26  423.42  1    714.38 43.866 5.032e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```

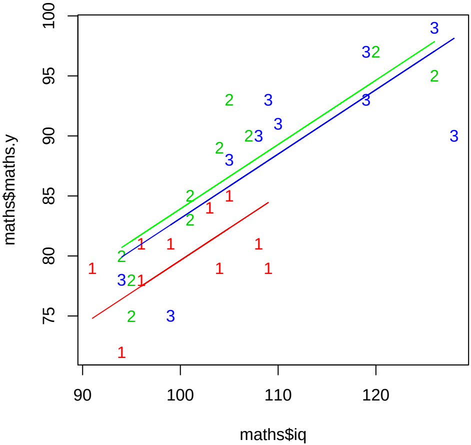
Clearly IQ is significant.

Exam marks example

```
> basemodel <- lm(maths.y ~ iq, data = maths)
> anova(basemodel, amodel)

Analysis of Variance Table

Model 1: maths.y ~ iq
Model 2: maths.y ~ class.f + iq
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     28 518.13
2     26 423.42  2    94.707 2.9077 0.0725 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The class is not very significant. However, since it is so close, we will retain it. Remember that it was significant in the one-factor model!

The fitted ANCOVA model



Exam marks example

We can also do the analysis from matrix theory.

```
> maths <- read.csv("../data/maths.csv")
> maths$class.f <- factor(maths$class)
> y <- maths$maths.y
> n <- 30
> X <- matrix(0, n, 8)
> X[,1] <- 1
> X[cbind(1:n,maths$class+1)] <- 1
> X[,5] <- maths$iq
> X[cbind(1:n,maths$class+5)] <- maths$iq
> r <- rankMatrix(X)[1]
```

Exam marks example

```
> X
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
 [1,]    1    1    0    0   99   99    0    0
 [2,]    1    1    0    0  103  103    0    0
 [3,]    1    1    0    0  108  108    0    0
 [4,]    1    1    0    0  109  109    0    0
 [5,]    1    1    0    0   96   96    0    0
 [6,]    1    1    0    0  104  104    0    0
 [7,]    1    1    0    0   96   96    0    0
 [8,]    1    1    0    0  105  105    0    0
 [9,]    1    1    0    0   94   94    0    0
[10,]    1    1    0    0   91   91    0    0
[11,]    1    0    1    0  101    0  101    0
[12,]    1    0    1    0   95    0   95    0
[13,]    1    0    1    0  105    0  105    0
[14,]    1    0    1    0   94    0   94    0
[15,]    1    0    1    0  101    0  101    0
[16,]    1    0    1    0  126    0  126    0
[17,]    1    0    1    0  107    0  107    0
[18,]    1    0    1    0  104    0  104    0
[19,]    1    0    1    0  120    0  120    0
[20,]    1    0    1    0   95    0   95    0
[21,]    1    0    0    1  108    0    0  108
[22,]    1    0    0    1   99    0    0   99
[23,]    1    0    0    1  126    0    0  126
[24,]    1    0    0    1  119    0    0  119
[25,]    1    0    0    1  109    0    0  109
[26,]    1    0    0    1  110    0    0  110
[27,]    1    0    0    1  105    0    0  105
[28,]    1    0    0    1  119    0    0  119
[29,]    1    0    0    1  128    0    0  128
```

Exam marks example

We check which parameters/contrasts can be estimated.

```
> XtX <- t(X) %*% X
> XtXc <- matrix(0, 8, 8)
> XtXc[c(2:4,6:8),c(2:4,6:8)] <- solve(XtX[c(2:4,6:8),c(2:4,6:8)])
> A <- XtXc %*% XtX
> round(c(1,0,0,0,0,0,0,0) %*% A,3)

     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    0    0    0    0    0    0    0    0

> round(c(1,1,0,0,0,0,0,0) %*% A,3)

     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    1    1    0    0    0    0    0    0

> round(c(1,0,1,0,0,0,0,0) %*% A,3)

     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    1    0    1    0    0    0    0    0
```

Exam marks example

```
> round(c(1,0,0,1,0,0,0,0) %*% A,3)
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    1    0    0    1    0    0    0    0
> round(c(0,0,0,0,1,0,0,0) %*% A,3)
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    0    0    0    0    0    0    0    0
> round(c(0,0,0,0,1,1,0,0) %*% A,3)
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    0    0    0    0    1    1    0    0
> round(c(0,0,0,0,1,0,1,0) %*% A,3)
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    0    0    0    0    1    0    1    0
> round(c(0,0,0,0,1,0,0,1) %*% A,3)
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    0    0    0    0    1    0    0    1
```

Exam marks example

Check the fit:

```
> b <- XtXc %*% t(X) %*% y
> s2 <- sum( (y - X%*%b)^2 )/(n-r)
> b[3]-b[2]

[1] -30.96419

> b[5]+b[6]

[1] 0.270073

> b[7]-b[6]

[1] 0.3473557

> sqrt(s2)

[1] 4.043299
```

Exam marks example

Test for interaction:

```
> (C <- matrix(c(0,0,0,0,0,1,-1,0,0,0,0,0,0,1,0,-1), 2, 8, byrow=T))

     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    0    0    0    0    0    1   -1    0
[2,]    0    0    0    0    0    1    0   -1

> (Fstat2 <- t(b) %*% t(C) %*%  solve( C %*% XtXc %*% t(C) ) %*%
+     C %*% b / 2 / s2)

          [,1]
[1,] 0.9500218

> pf(Fstat2,2,n-r,lower=F)

          [,1]
[1,] 0.4008011
```

Exam marks example

Model without interaction:

```
> X <- X[,1:5]
> r <- rankMatrix(X)[1]
> XtX <- t(X) %*% X
> XtXc <- matrix(0, 5, 5)
> XtXc[2:5,2:5] <- solve(XtX[2:5,2:5])
```

The iq coefficient is now estimable:

```
> A <- XtXc %*% XtX
> round(c(0,0,0,0,1) %*% A, 3)
```

```
     [,1] [,2] [,3] [,4] [,5]
[1,]    0    0    0    0    1
```

Exam marks example

Check model without interaction:

```
> b <- XtXc %*% t(X) %*% y
> s2 <- sum( (y - X%*%b)^2 )/(n-r)
> b[3]-b[2]

[1] 4.295033

> b[5]

[1] 0.5360389

> sqrt(s2)

[1] 4.03552
```

Exam marks example

Test significance of class:

```
> (C <- matrix(c(0,1,-1,0,0,0,1,0,-1,0), 2, 5, byrow=T))

     [,1] [,2] [,3] [,4] [,5]
[1,]    0    1   -1    0    0
[2,]    0    1    0   -1    0

> (Fstat <- t(b) %*% t(C) %*% solve( C %*% XtXc %*% t(C) ) %*%
+     C %*% b / 2 / s2)

         [,1]
[1,] 2.907729

> pf(Fstat,2,n-r,lower=F)

           [,1]
[1,] 0.07250318
```

Exam marks example

Test significance of IQ:

```
> (C <- matrix(c(0,0,0,0,1), 1, 5, byrow=T))

     [,1] [,2] [,3] [,4] [,5]
[1,]    0    0    0    0    1

> (Fstat <- t(b) %*% t(C) %*% solve( C %*% XtXc %*% t(C) ) %*%
+     C %*% b / 1 / s2)

         [,1]
[1,] 43.86618

> pf(Fstat,1,n-r,lower=F)

             [,1]
[1,] 5.032089e-07
```

Motor Trends car tests

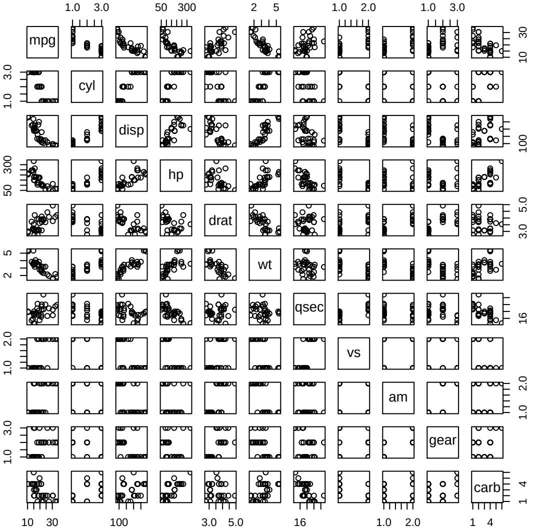
The US magazine *Motor Trends* published a dataset in 1974 on fuel consumption of cars for 32 different models. The variables are:

- `mpg`: Miles/(US) gallon
- `cyl`: Number of cylinders
- `disp`: Displacement (cu.in.)
- `hp`: Gross horsepower
- `drat`: Rear axle ratio
- `wt`: Weight (1000 lbs)
- `qsec`: 1/4 mile time
- `vs`: V/S (engine type)
- `am`: Transmission (0 = automatic, 1 = manual)
- `gear`: Number of forward gears
- `carb`: Number of carburetors

Motor Trends car tests

```
> data(mtcars)
> str(mtcars)
'data.frame': 32 obs. of  11 variables:
 $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
 $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
 $ disp: num  160 160 108 258 360 ...
 $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
 $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
 $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
 $ qsec: num  16.5 17 18.6 19.4 17 ...
 $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
 $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
 $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
 $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
> mtcars$cyl <- factor(mtcars$cyl)
> mtcars$vs <- factor(mtcars$vs)
> mtcars$am <- factor(mtcars$am)
> mtcars$gear <- factor(mtcars$gear)
> mtcars$carb <- factor(mtcars$carb)
```

Motor Trends car tests



Motor Trends car tests

```
> model <- lm(mpg ~ ., data=mtcars)
> summary(model)
Call:
lm(formula = mpg ~ ., data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-3.5087 -1.3584 -0.0948  0.7745  4.6251

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 23.87913   20.06582   1.190   0.2525
cyl6        -2.64870    3.04089  -0.871   0.3975
cyl8        -0.33616    7.15954  -0.047   0.9632
disp         0.03555    0.03190   1.114   0.2827
hp          -0.07051    0.03943  -1.788   0.0939 .
drat         1.18283    2.48348   0.476   0.6407
wt          -4.52978    2.53875  -1.784   0.0946 .
qsec         0.36784    0.93540   0.393   0.6997
vs1          1.93085    2.87126   0.672   0.5115
am1          1.21212    3.21355   0.377   0.7113
gear4        1.11435    3.79952   0.293   0.7733
gear5        2.52840    3.73636   0.677   0.5089
carb2       -0.97935    2.31797  -0.423   0.6787
carb3        2.99964    4.29355   0.699   0.4955
carb4        1.09142    4.44962   0.245   0.8096
carb6        4.47757    6.38406   0.701   0.4938
carb8        7.25041    8.36057   0.867   0.3995
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Motor Trends car tests

```
> model2 <- step(model)
Start:  AIC=76.4
mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb

       Df Sum of Sq    RSS    AIC
- carb  5   13.5989 134.00 69.828
- gear  2    3.9729 124.38 73.442
- am    1    1.1420 121.55 74.705
- qsec  1    1.2413 121.64 74.732
- drat  1    1.8208 122.22 74.884
- cyl   2   10.9314 131.33 75.184
- vs    1    3.6299 124.03 75.354
<none>              120.40 76.403
- disp  1    9.9672 130.37 76.948
- wt    1   25.5541 145.96 80.562
- hp    1   25.6715 146.07 80.588

Step:  AIC=69.83
mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear

       Df Sum of Sq    RSS    AIC
- gear  2    5.0215 139.02 67.005
- disp  1    0.9934 135.00 68.064
- drat  1    1.1854 135.19 68.110
- vs    1    3.6763 137.68 68.694
- cyl   2   12.5642 146.57 68.696
- qsec  1    5.2634 139.26 69.061
<none>              134.00 69.828
- am    1   11.9255 145.93 70.556
- wt    1   19.7963 153.80 72.237
- hp    1   22.7935 156.79 72.855
```

Motor Trends car tests

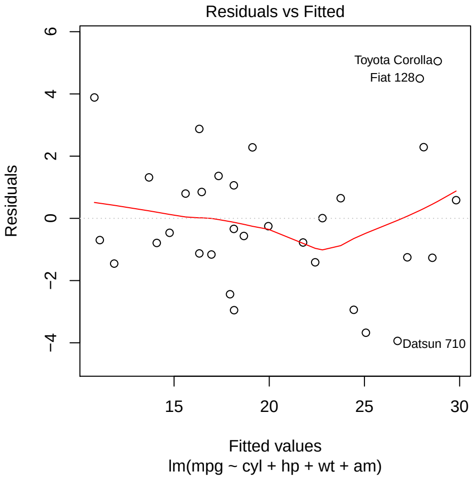
```
> summary(model2)

Call:
lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9387 -1.2560 -0.4013  1.1253  5.0513

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
cyl6        -3.03134    1.40728  -2.154  0.04068 *
cyl8        -2.16368    2.28425  -0.947  0.35225
hp          -0.03211    0.01369  -2.345  0.02693 *
wt          -2.49683    0.88559  -2.819  0.00908 **
am1          1.80921    1.39630   1.296  0.20646
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Motor Trends car tests

```
> model3 <- lm(mpg ~ (cyl + hp + wt + am)^2, data=mtcars)
> anova(model2, model3)

Analysis of Variance Table

Model 1: mpg ~ cyl + hp + wt + am
Model 2: mpg ~ (cyl + hp + wt + am)^2
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     26 151.03
2     17 102.47  9     48.56 0.8952 0.5496
```

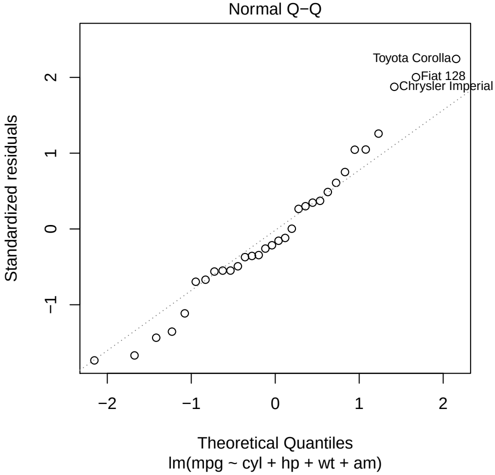
Motor Trends car tests

```
> plot(model2, which=1)
```



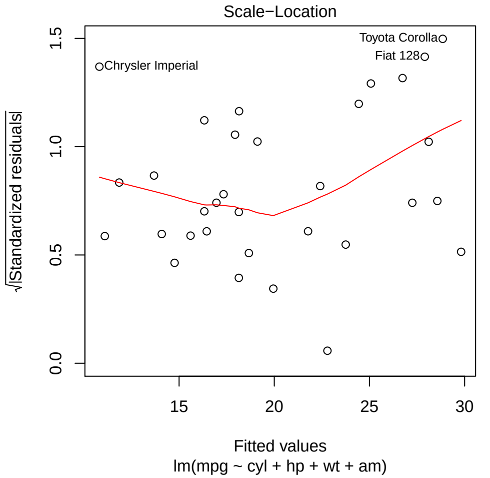
Motor Trends car tests

```
> plot(model2, which=2)
```



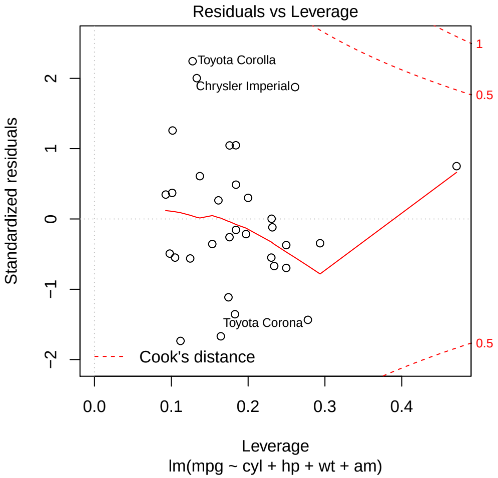
Motor Trends car tests

```
> plot(model2, which=3)
```



Motor Trends car tests

```
> plot(model2, which=5)
```



Motor Trends car tests

```
> model4 <- step(model3)
Start:  AIC=67.24
mpg ~ (cyl + hp + wt + am)^2

         Df Sum of Sq    RSS    AIC
- cyl:wt  2    0.6926 103.16 63.457
- cyl:am  2    1.8176 104.28 63.804
- cyl:hp  2    5.8623 108.33 65.022
- hp:wt   1    0.0052 102.47 65.243
- hp:am   1    3.4193 105.89 66.292
- wt:am   1    6.1169 108.58 67.097
<none>                102.47 67.241

Step:  AIC=63.46
mpg ~ cyl + hp + wt + am + cyl:hp + cyl:am + hp:wt + hp:am +
    wt:am

         Df Sum of Sq    RSS    AIC
- cyl:am  2    1.3945 104.55 59.886
- hp:wt   1    0.0841 103.24 61.483
- hp:am   1    3.6818 106.84 62.579
- cyl:hp  2   13.1830 116.34 63.305
<none>                103.16 63.457
- wt:am   1    9.9355 113.09 64.399

Step:  AIC=59.89
mpg ~ cyl + hp + wt + am + cyl:hp + hp:wt + hp:am + wt:am

         Df Sum of Sq    RSS    AIC
- hp:wt   1    0.0663 104.62 57.907
- hp:am   1    3.5035 108.06 58.941
```

Motor Trends car tests

```
> summary(model4)
Call:
lm(formula = mpg ~ cyl + hp + wt + am + cyl:hp + wt:am, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-2.8777 -1.4603 -0.5024  1.2795  3.8468

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  36.65881    3.98711   9.194 3.64e-09 ***
cyl6         -7.19711    5.60784  -1.283  0.21213
cyl8        -10.82118    4.22762  -2.560  0.01752 *
hp           -0.08268    0.03401  -2.431  0.02326 *
wt           -2.31293    0.81181  -2.849  0.00908 **
am1           9.14282    4.12170   2.218  0.03669 *
cyl6:hp       0.05954    0.05035   1.182  0.24913
cyl8:hp       0.07634    0.03565   2.142  0.04305 *
wt:am1       -3.04685    1.51646  -2.009  0.05639 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.172 on 23 degrees of freedom
```