GENERAL LINEAR HYPOTHESES

So far we have considered comparison of nested models where the smaller model simply dropped variables from the model equation. For example:

$$E(Y_i) = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} \tag{1}$$

compared with

$$E(Y_i) = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,4}\,. \tag{2}$$

The first model is a special case of the second, where $\beta_3 = \beta_4 = 0$.

Let X_i denote the i-th column of the design matrix X, and suppose the full model is (2). In this case, the full model says that $E(Y)$ (as a column vector) lies in V, the space spanned by the columns of X, while the small model (1) says that $E(Y)$ lies in V_0, the space spanned by the first two columns of X.

The space V_0 is a subspace of V. We could consider other kind of subspaces. Note that if we assume $\beta_1 = \beta_2 = \beta_3$, then if J is the column of all 1's,

$$E(Y) = \beta_0 J + \beta_1 (X_1 + X_2 + X_3) + \beta_4 X_4\,.$$

This is the subspace spanned by the vectors $J, X_1 + X_2 + X_3$ and X_4. This forms a subspace of W_0 of V.

We can test the hypothesis

$$H_0 : E(Y) \in W_0 \text{ vs } H_1 : E(Y) \in V \setminus W_0$$

as before by the F-statistic

$$F = \frac{RSS_1 - RSS_2/(\Delta df)}{RSS_{FULL}/df(FULL)}$$

Here we need to determined the change in degrees of freedom between the two models. Note that the dull model had dimension 5, while the sub model has dimension 3. Thus there is a loss of 2 degrees of freedom.

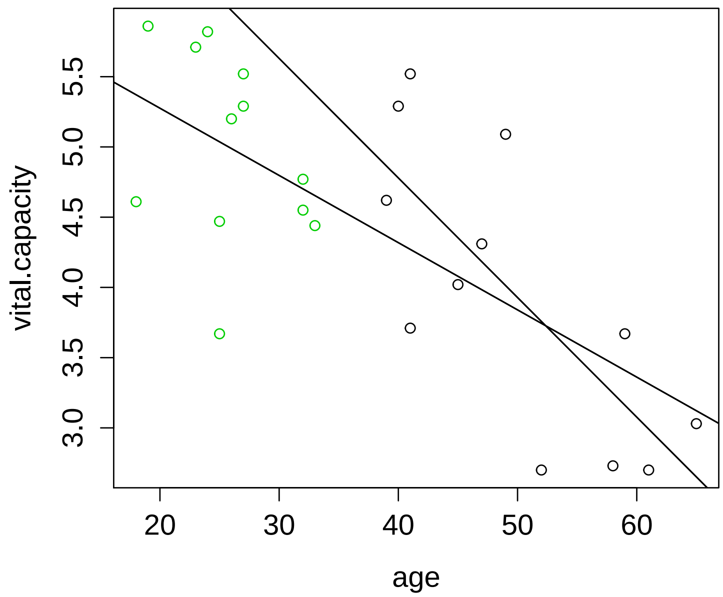
1. ANCOVA on Vital Capacity Data

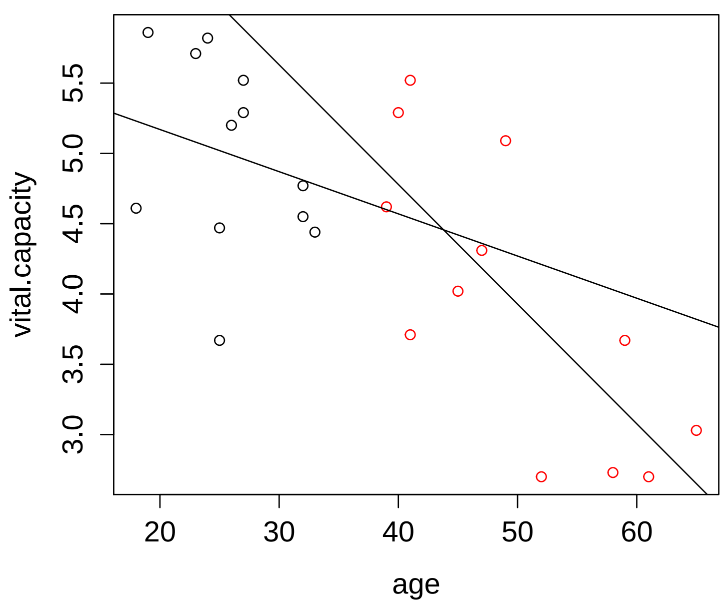
Consider the dataset `vitcap.txt`. This gives vital capacity (a measure of lung function) and age of individuals. Additionally, individuals have either been exposed to a toxin (group 1) or not (group 3).

First, perform an analysis of covariance comparing how lung function varies with age for the two groups. Be sure to include a plot displaying the relationship for the two groups. Is there a significant difference?

Now, suppose it is known that the coefficient of age in normal populations is -0.03. Consider the model that forces this coefficient to be -0.03 for group 3, but not for group 1.

Test the hypothesis that the slope for group 1 is different from -0.03. Then do an F-test comparing the two models.


vital.capacity
age
3.0
3.5
4.0
4.5
5.0
5.5
20
30
40
50
60


vital.capacity
age
3.0
3.5
4.0
4.5
5.0
5.5
20
30
40
50
60