Question 6.1CC: Factorial experiments (cont'd)

This Question continues with the work on data set `missile` *from Question CC5.2 from the previous sheet. Complete Question CC5.2 if not done so yet. Then, continue with the following questions.*

(a) Display again the model `summary` of the fitted interaction model. Read and interpret from this the result of the overall F–test.

(b) Use function `anova()` to perform (sequential) analysis of variance to investigate the significance of the effects included in the model. At $\alpha = 0.05$, what is the simplest appropriate model? Do your results change when changing the order of terms?

(c) Test (via a partial F–test) the hypothesis $H_0 = E(\text{life}|\text{material}) = \mu + \tau_i^{\mathcal{A}}$ vs $H_{\text{A}} = E(\text{life}|\text{material, temperature}) = \mu + \tau_i^{\mathcal{A}} + \tau_j^{\mathcal{B}}$ at a 5% significance level. Write your conclusions below. [*Hint*: Fit both models and then use `anova( , )`.]

(d) Let's go back to the model with interaction. Using the values in the corresponding ANOVA table calculate R^2. Interpret the result and verify it by comparison with the R^2 given in the output from the `summary()` function.

Question 6.2CC: Mallows' $C_{\mathcal{I}}$

(a) Read in the 'Portland cement' data discussed in Example 3.20 in lectures (see the Lecture 25 code). Please use the code as follows to ensure that the ordering and names of variables correspond to those used in the lecture.

```
data(cement, package="MASS")
cement=cement[,c(5,1,2,3,4)]
names(cement)= c("heat", "aluminate", "tri.silicate", "ferite", "di.silicate")
```

(b) Fit a linear model using `heat` as response, and the four chemical compounds as predictors. Save the fitted model, which will serve as our 'full' model in what follows, into an object `cement.fit.full`. Extract, for later use, the residual standard error from this model, and save it into an object `s`.

(c) Produce a few sequential ANOVA tables for these data (using several orders of inclusion) to get impression on the relative importance of the variables. Summarize your conclusions.

(d) Produce a function `CI` which implements the quantity $C_{\mathcal{I}}$. Your function should have the shape

```
CI<-function(model, var){
  pI <- length(model$coef)
  n  <-...
  RSSI <- ...
  ci <- ...
  return(ci)
}
```

where `model` is a fitted model object, and `var` the error variance obtained from the *full* model.

(e) Apply your function `CI` to as many of the 16 submodels of the full model as you are able to do, in order to reproduce the results from Table 3.5 in the lecture notes. There are multiple ways of doing this, for instance the model '$1 + x_1 + x_3$' could be fitted via any of

```
cement.fit.1.3 = lm(heat~  aluminate+ferite,   data = cement)
cement.fit.1.3 = lm(heat~.,   data = cement[,c(1,2,4)])
```

(f) Contemplate more efficient ways of carrying out this analysis (you will probably not have time to put your thoughts into practice).