1. Faraway (1st edition) problem 13.1

2. This problem will work with the `barley` dataset which you can download from the `lattice` library. This data set has $n = 120$ data points, each giving the crop yield of barley with covariates `variety` (10 types), `site` (6 types), and `year` (2 different years).

   (a) How many degrees of freedom would be used by the model will all interactions, (i.e. the regression `yield~variety*site*year`)? Would we be able to do significance testing on this model?

   (b) How many d.f. would be used by the model with all factors and two-way interactions, but not three-way interactions, (i.e. the regression `yield~(variety+site+year)**2`)? For both this part and the part above, show your d.f. calculation by hand, not by running the model in R.

   (c) From this point on, we will use the data set with data points 23 and 83 removed (these are identified as outliers, perhaps mistaken data entry, by your textbook), so your sample size is now 118. Run the model with all two-way interactions, `yield~(variety+site+year)**2`. Use ANOVA to check each of the two-way interactions for significance at the 0.05 level, removing them one at a time if appropriate. Show your R code/output as you perform each step of this procedure.

3. Pairwise comparisons. The `pulp` data set contains 20 data points, 5 each in groups A, B, C, D. The response `bright` is paper brightness and the covariate `operator` is some treatment applied during paper production, either A or B or C or D. We will investigate whether there are significant differences between any pair, e.g. brightness is significantly higher for production method A than for D, or statements of this type.

   (a) Begin by calculating the sample mean of `bright` in each group, $\hat{\alpha}_A$, $\hat{\alpha}_B$, $\hat{\alpha}_C$, and $\hat{\alpha}_D$.

   (b) Next calculate $\hat{\sigma}$, assuming that each observation is normally distributed as $Y_i \sim N(\alpha_{\dots}, \sigma^2)$ (where ... denotes the group, A or B or C or D, that data point i is assigned to)

   (c) Supposing that σ were known, what's the square root of the variance of $\hat{\alpha}_A - \hat{\alpha}_B$ in terms of σ? Now plug in $\hat{\sigma}$ in place of σ, this is now the standard error, $\mathrm{SE}(\hat{\alpha}_A - \hat{\alpha}_B)$. Repeat for every possible pairwise comparison.

   (d) Finally, calculate the Tukey honest significant difference (Tukey HSD) confidence interval for each possible pairwise comparison. At the 0.05 level, what conclusions can you draw about the four production methods?

4. Suppose that combinations of three drugs, called A and B and C, are being examined for their ability to lower blood pressure. Suppose that, without any medication, expected systolic blood pressure (the response Y) in the population being studied is 150. Any one drug on its own has no effect on blood pressure. However, drug A in combination with B or C will reduce blood pressure to 140. Drugs B and C are chemically very similar and it doesn't matter which one is used in combination with drug A. There's no benefit to using both—it's equivalent to just using one.

   Now suppose we want to write down a linear model, using treatment coding, to describe this scenario. What are the values of all the coefficients in the model:

   - The intercept term β_0
   - The one-way terms β_{A1}, β_{B1}, β_{C1}
   - If needed, the two-way interaction terms $\beta_{A1:B1}$, $\beta_{A1:C1}$, $\beta_{B1:C1}$
   - If needed, the three-way interaction term $\beta_{A1:B1:C1}$