1. The data in exercise 10.26 from Devore’s 6th edition (data set `ex10.26` in the `Devore6` package) are the results of the analyses of the amount of physiologically active polyunsaturated fatty acids (PAPFUA, in percentages) for six different brands of diet/imitation margarine. Of these margarines, Mazola and Fleishman’s are corn-based whereas the others are soybean-based.

   (a) Add a factor, say `base`, for corn-based versus soy-based to the data frame. Fit a linear model with `PAPFUA` as the response and `base` as the only term. Check the model matrix for this fitted model and describe how the columns in the model matrix are determined.

   (b) Fit a linear model with `PAPFUA` as the response and `brand` as a term. Check the model matrix for this fitted model and describe how the columns in the model matrix are determined.

   (c) Describe what the coefficient labelled "`brandChiffon`" in the second model represents. How would the model be changed if the true value of this coefficient was zero?

2. Consider the data on the yield of paddy (a grain grown in India) versus date of harvesting (measured as number of days after flowering) in the data set `xmp13.07` in the `Devore6` package. I showed a scatterplot of these data in class and suggested that a quadratic model could be reasonable.

   (a) Fit the model with `yield` of the form
   ```r
   > yield ~ days + I(days^2)
   ```
   Examine the model matrix. Describe how the model matrix is generated.

   (b) Fit the same model but omitting the `I()` from the second term. Describe the result.

   (c) Fit a model with `yield` as the response and a single term `poly(days, 2)`. Compare a summary of this model with a summary of the model in part (a). What is the same in the two fits and what is different?

   (d) Consider the model matrix from the `poly` fit. What is the matrix of crossproducts, `X'X`? (You can evaluate this matrix with the `crossprod` function. You may want to use the `zapsmall` function on the result to see the structure more clearly.) Explain why we say that the `poly` function generates “orthogonal” polynomials.

3. The data for exercise 1.19 in the book by Sen and Srivastava (package `SenSrivastava`; data set `E1.19`) provide the price of books versus the number of pages and a characterization of whether the book is a paperback or a hardcover book.

   (a) Provide separate plots of price versus number of pages by book type with an overlay of a fitted least squares line on each panel. (See the panel function `panel.lmline`. You probably want to combine it with `panel.xyplot`.) Comment on the suitability of the fits.

   (b) Repeat part (a) omitting the two points that are abnormally low prices in the "c" panel. You can detect these points on the basis of their price being less than 23 and the type being "c".

   (c) Fit a model with `Price` as the response and the expression `P*B` for the linear predictor. Examine the model matrix. How many terms are there? How are they calculated? What do the corresponding coefficients represent?

   (d) Repeat the model fit in part (c) omitting the two unusual data points. Consider the summary of each fit and comment on the ways in which those two points have influenced the fit. In particular, does removing them change predictions for the type (B) of "p"? Should it?