189.523B Assignment #2 (corrected 20001.02.06)
Due Friday 9 Febuary 2001 in class

Notes:

- This Assignment comprises 5 questions on 3 pages.
- The data are available in R format with the extension `.dput` or in text format with the extension `.dat` with the first row containing the variable names.
- Data files in R format can be read in using the `dget` command and `attach`'ed as in Assignment 1.
- If you use the R format, be aware that the **factors are not coded as such**. You will have to modify them appropriately.

1. The data in `malepigs.*` and `fempigs.*` are from an experimental piggery arranged for individual feeding of six pigs in each of five pens (covariate `pen`). From each of five litter, six young pigs, three males and three females, were selected and allotted to one of the pens. Three feeding treatments denoted by 1, 2, 3 (covariate `food`), containing increasing proportions ($p_1 < p_2 < p_3$) of protein, were used and each given to one male and one female in each pen. The pigs were individually weighed each week for 16 weeks. For each pig, the growth rate in pounds per week (covariate `growth`) was calculated as the slope of a line fitted by least-squares. The weight of each pig at the beginning of the experiment is also included (covariate `weight`). The variables in each of the files are summarized in the following table:

Column	Var. name	Factor/ Continous	Description	Values
Col. 1	`food`:	Factor	Type of food	1,2,3
Col. 2	`pen`:	Factor	Pen number	1,2,3,4,5
Col. 3	`growth`:	Continuous	Average growth rate	(in lbs/wk)
Col. 4	`weight`:	Continuous	Original weight	(in lbs)

The male students should use `malepigs.*` and the female students `fempigs.*`. Use the data to show that

(a) `pen` and `food` are orthogonal for the outcome `growth` with respect to the intercept.

(b) `pen` and `food` are not orthogonal for the outcome `growth` with respect to both the intercept and `weight`.

2. We wish to choose between the two models

$$\begin{aligned} M_1 : \mathbb{E}[Y_i] &= x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p \\ M_2 : \mathbb{E}[Y_i] &= x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + x_{i,p+1}\beta_{p+1} \end{aligned}$$

for $i = 1, \ldots, n$, where the $x_{i,j}$ and β_j, $j = 1, \ldots, p+1$, are respectively covariates and unknown scalars.
Show that the model chosen by the criterion $\widehat{\text{MSEP}}_3$ is approximately the same as that which is chosen by a t-test of the hypothesis

$$H_0 : \beta_{p+1} = 0$$

at the $\sqrt{2}$ critical value (i.e., we choose M_1 if $|t| \leq \sqrt{2}$ and M_2 if $|t| > \sqrt{2}$, for $t = \hat{\beta}_{p+1}/\widehat{\text{s.e.}}(\hat{\beta}_{p+1})$).

3. Show that $\widehat{\text{MSEP}}_2$ (also known as PRESS) is a slightly upwardly biased estimator of MSEP.

4. The file `cars.*` contains data on cars that were advertised for sale in a French newspaper in September 1985. There are 164 observations, each containing 4 variables:

Column	Var. name	Factor/ Continous	Description	Values
Col. 1	`type`:	Factor	Type of car	1=Peugeot 104
				2= Citroën 2CV
				3= Peugeot 304/305
				4= Renault 4
				5= Renault 5
				6= Peugeot 504/505
				7= Renault 18
Col. 2	`year`:	Continuous	Year of car	66, 67,. . . ,85
Col. 3	`kilo`:	Continuous	Kilometrage	(in 1000km)
Col. 4	`price`:	Continuous	Asking price	(in 1000F)

We are interested in the relation between price and the other variables.

(a) Make a plot of price against year. Does there seem to be a strong linear relationship?

(b) Calculate the variable `logpr<-log(price)` and produce a plot of `logpr` against `year`. Notice the improvement in linearity. Test that there is a significant linear relationship between `logpr` and `year`.

(c) Is it possible that there are different basic prices for each type of car? Fit a model with different intercepts according to type, and test the hypothesis that these intercepts are really different, assuming (for the moment) that the linear relationship with year is the same for each type.

(d) Do the different types of car depreciate in value at the same rate? Test the hypothesis that the linear effect of year is the same for all types of car, or, if you like, that there is an interaction between type and year. Which type of car depreciates the most? the least?

(e) Is there an effect of `kilo` on `logpr` allowing for `type`, `year` and their interaction?

(f) Is there an interaction between `kilo` and `type`, allowing for the variables fitted in part **(e)**?

5. **(a)** Among the models that you have fitted in Question **4.** (and any others that you think are reasonable), which is the best from the point of view of prediction? Use the MSEP criterion.

(b) A particular Renault 5, having logged 45,000 km, was purchased in France in 1979 for 16,000F. Use the model chosen in part **(a)** to predict the asking price for such a car.

End of Assignment 2.