Assignment 1 (Due next Friday (09/06/02))

1. A randomized experiment was conducted to test if factor $B$ has larger effect than factor $A$. The data are given in the following.
   a) Construct the randomization reference distribution.
   b) Calculate the $p$-value based on the randomization distribution.
   c) Calculate the $p$-value based on two-sample $t$ test.
   d) Comment on the procedures and the results in b and c.

   $\begin{array}{cccc}
   A & B & B & A & B \\
   3 & 5 & 5 & 1 & 8
   \end{array}$

2. 
   a) For the typing experiment discussed in class, use a statistical model to quantify the gains from using randomization (as illustrated in the second sequence in the lecture note) and from using balance in addition to randomization.
   b) Suppose that the following sequence is obtained from using balanced randomization:


   Would you use it for the study? If not, what would you do? What aspect of the sequence makes you concerned? Can you relate it to the possibility that the learning effect may decay over time? Express it in more rigorous terms. (Hint: the terms in the models should represent the effects you have identified as potentially influencing the comparison).

3. Data on sales last year ($Y$, in thousand squares) in 26 sales districts are given below for a maker of asphalt roofing shingles. (only part of the data shown here, you can go to the course website to download the complete data entitled roofing.dat). Shown also are
promotional expenditures \((X_1,\text{in thousand dollars})\), number of active accounts \((X_2)\), number of competing brands \((X_3)\), and district potential \((X_4, \text{coded})\) for each of the districts.

\[
\begin{array}{cccccc}
\text{y} & \text{x1} & \text{x2} & \text{x3} & \text{x4} \\
79.3 & 5.5 & 31.0 & 10.0 & 8.0 \\
200.1 & 2.5 & 55.0 & 8.0 & 6.0 \\
163.2 & 8.0 & 67.0 & 12.0 & 9.0 \\
200.1 & 3.0 & 50.0 & 7.0 & 16.0 \\
146.0 & 3.0 & 38.0 & 8.0 & 15.0 \\
177.7 & 2.9 & 71.0 & 12.0 & 17.0 \\
30.9 & 8.0 & 30.0 & 12.0 & 8.0 \\
291.9 & 9.0 & 56.0 & 5.0 & 10.0 \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
\end{array}
\]

a) Fit the multiple regression model containing all four explanatory variables. Does it appear that all predictors should be retained?

b) Among the possible models that contain only two predictors, which one you guess is the best? Justify your guess by comparing to the full model.