Instructions:
1. You may use your formula sheet, a calculator, and the tables in the back of your text.
2. In order to get full credit, show as much of your work as possible.
3. Point values are given in parentheses.
4. Good luck!

1. (30) An experiment was conducted to study the performance of three different detergents (A, B and C) at two different water temperatures (warm and cold). Twenty-four loads of clothes were washed with 4 loads randomly assigned to each combination of detergent and water temperature. The response variable is a "whiteness" reading ascribed to each load of clothes (the higher the reading the whiter the clothes). Use the accompanying graph of treatment means and SPSS output to do a complete analysis of these data. Use $\alpha = 0.05$ for each hypothesis test you do. **Note:** Should you need it, Fisher's LSD value ($\alpha = 0.05$) for comparing two treatment means is 1.623.

The interaction test is significant at the .05 level. The easiest way to explain the interaction is to first note that washing with detergent B in warm water yields significantly whiter clothes, on average, than any of the other 5 treatments (using Fisher's LSD). However, we cannot reject the hypothesis that average whiteness is the same for all 5 treatments besides detergent B in warm water.
3. (12) A researcher wishes to estimate the relationship between a response $y$ and two independent variables $x_1$ and $x_2$ using a model of the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon.$$ 

Based on a data set of size $n = 200$, the researcher fits the above model and finds that $R^2 = 0.95$. He is somewhat surprised to find that the two $t$-tests of $H_0 : \beta_1 = 0$ and $H_0 : \beta_2 = 0$ have respective $P$-values of 0.52 and 0.67. He feels that such large $P$-values contradict the fact that the $R^2$ value is large. Provide a plausible explanation for this phenomenon. Be specific.

The two variables are undoubtedly highly correlated. The fact that both $P$-values are large means that both variables are not needed in the model, not that neither variable is needed.

4. Twenty-four middle level managers at a certain company were placed in groups of size 3 according to level of experience. The three managers within each group were randomly assigned to the three training conditions: no training, computer-assisted training, or computer-assisted training plus a behavior modeling workshop. After formal training, the managers were administered a 25-question multiple-choice test of managerial knowledge and the number of correct answers were recorded for each. The data from this randomized block design (RBD) yielded a block mean square of 35.09 and an error mean square of 3.64.

(a) (12) Given the same 24 managers, describe how this experiment could have been done as a completely randomized design (CRD).

Randomly select 8 managers and assign them to "no training." Randomly select 8 of the 16 remaining managers and assign them to "computer assisted training." Assign the remaining 8 to the third treatment.

(b) (16) Estimate the relative efficiency of the RBD relative to the CRD. What does this number actually tell you?

$$RE = \frac{7(35.09) + 8(2)(3.64)}{23(3.64)} = 3.63$$

About 4 times as many managers would have been needed for a CRD to yield as much info about treatment differences as did the RBD.
Practice Problem for Test 2

A petroleum company was interested in comparing the miles per gallon obtained by four different gasoline blends: A, B, C and D. Because there can be considerable variability due to differences in driving characteristics and car models, these two extraneous sources of variation were included as "blocking" variables in the study. The researcher selected four different types of cars and four different drivers. The drivers and car types were assigned to the blends in the way displayed in the table below. The mileage (in miles per gallon) obtained over each test run are also given in the table.

<table>
<thead>
<tr>
<th>Car Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Driver</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

Using the accompanying output, conduct an analysis to determine if there are differences in average miles per gallon between blends, and if so, determine which blends are best.

The ANOVA table shows that there are significant differences between blends in terms of average mpg. Tukey's procedure does not allow us to distinguish between blends A, B and D, but blends B and D yield significantly higher average mpg than blend C.