

HOMEWORK ASSIGNMENT 6		Assigned: November 3, 2005
		Due: November 10, 2005
		20 Points – 1 point each day late

All data is in "<http://www.stat.lsu.edu/EXSTWeb/statlab/datasets/KNNLdata/>".

Complete the following questions from your textbook. There are two problems. The first problem is analysis of covariance example based on the "Copier maintenance" example (8.15 & 8.19). The original datasets are in CH01PR20.txt and CH08PR15.txt. As a special treat I have prepared a csv dataset called CH08PR15.csv!

The dataset 8.15 has variables Y and X1, but not X2. Dataset 8.19 has X1, but not Y or X1. You may fix this any way you wish. If you prefer you may use CH08PR15.csv as an external file. This external file has all 3 variables separated by commas. To use this dataset you must tell SAS that the separator is a comma. The first line has the values "minutes serviced model", so you will need to start reading data on the second line. My input segment is given below.

```
filename copier 'C:\Geaghan\Current\EXST7034\Fall2005\Datasets (KNNL)\CH01PR20.csv';
ODS HTML style=minimal rs=none
      body='C:\Geaghan\Current\EXST7034\Fall2005\SAS\CH01PR20A6.html' ;

DATA copier; infile copier missover DSD dlm="," firstobs=2;
      INPUT minutes machines model;
```

The second problem is a variable selection example based on the patient satisfaction example (9.17). The dataset is in CH06PR15.

A) 8.15 a - e	<p>Note that for this first part no <u>interaction</u> is fitted. The first model is the textbooks 8.33 ($Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$) where X_2 is an indicator variable.</p> <p>1) Fit the model and complete questions a, b and c. Answer the question in d.</p> <p>2) Plot the residuals on X1X2 (interaction) as suggested in the book. Is there an indication of an interaction? Try a second plot using the statement below. Note I placed the residuals, which I called "E", in a data set that I called "RESIDS".</p> <pre>PROC PLOT DATA=RESIDS; PLOT E*X1=X2 / VREF=0; RUN;</pre>
B) 8.19 a & b	<p>Note that for this second part an interaction is fitted.</p> <p>1) Complete both parts of problem 8.19.</p> <p>2) Can you figure out how to estimate both slopes and both intercepts with standard errors in a single model with pooled variance?</p>
C) 9.17 a, d & e	<p>Do only parts a, d and e with the following alterations.</p> <p>a) Use the stepwise selection options and a value of "SLENTY = 0.05" and "SLSTAY = 0.05".</p> <p>b and c) You do not need to answer these question.</p> <p>d) Use the backward selection options and a value of "SLSTAY = 0.05".</p> <p>e) Compare the two techniques used in parts a and d above.</p>

- 8.15. Refer to **Copier maintenance** Problem 1.20. The users of the copiers are either training institutions that use a small model, or business firms that use a large, commercial model. An analyst at Tri-City wishes to fit a regression model including both number of copiers serviced (X_1) and type of copier (X_2) as predictor variables and estimate the effect of copier model (S—small, L—large) on number of minutes spent on the service call. Records show that the models serviced in the 45 calls were:

$i:$	1	2	3	...	43	44	45
$X_{i2}:$	S	L	L	...	L	L	L

Assume that regression model (8.33) is appropriate, and let $X_2 = 1$ if small model and 0 if large, commercial model.

- a. Explain the meaning of all regression coefficients in the model.
 - b. Fit the regression model and state the estimated regression function.
 - c. Estimate the effect of copier model on mean service time with a 95 percent confidence interval. Interpret your interval estimate.
 - d. Why would the analyst wish to include X_1 , number of copiers, in the regression model when interest is in estimating the effect of type of copier model on service time?
 - e. Obtain the residuals and plot them against X_1X_2 . Is there any indication that an interaction term in the regression model would be helpful?
- *8.19. Refer to **Copier maintenance** Problems 1.20 and 8.15.
- a. Fit regression model (8.49) and state the estimated regression function.
 - b. Test whether the interaction term can be dropped from the model; control the α risk at .10. State the alternatives, decision rule, and conclusion. What is the P -value of the test? If the interaction term cannot be dropped from the model, describe the nature of the interaction effect.
- *9.17. Refer to **Patient satisfaction** Problems 6.15 and 9.9. The hospital administrator was interested to learn how the forward stepwise selection procedure and some of its variations would perform here.
- a. Determine the subset of variables that is selected as best by the forward stepwise regression procedure, using F limits of 3.0 and 2.9 to add or delete a variable, respectively. Show your steps.
 - b. To what level of significance in any individual test is the F limit of 3.0 for adding a variable approximately equivalent here?
 - c. Determine the subset of variables that is selected as best by the forward selection procedure, using an F limit of 3.0 to add a variable. Show your steps.
 - d. Determine the subset of variables that is selected as best by the backward elimination procedure, using an F limit of 2.9 to delete a variable. Show your steps.
 - e. Compare the results of the three selection procedures. How consistent are these results? How do the results compare with those for all possible regressions in Problem 9.9?