EXST7034 : Regression Techniques

HOMEWORK ASSIGNMENTS : General Information

PC/NT Workstations are available in room 11 and room 48, both in the basement of the Ag. Admin. Building. Class schedules will be posted and you can use either when regular classes are not scheduled. Room 50 should become available as an open lab in a few weeks. The use of SAS is strongly encouraged.

Most homework assignments can be done in two to four pages. **Do not pass in raw computer printouts unless requested**. If you need to reference computer output, cut out the parts you need and tape or paste them onto written homework, or write your answers on selected output pages. Do not pass in computer pages you do not need. Homework is generally due on Thursday of each week.

All of the data files for this book are available on disk. The disk is included with the textbook. I will also maintain a listing of the data files on the internet at

http://www.stat.lsu.edu/exstweb/statlab/datasets/NKNWData/index.html

HOMEWORK ASSIGNMENT 1

Assigned: September 8, 2005
Due: September 15, 2005
20 Points – 1 point each day late

Questions are from chapter one of your text, Kutner, Nachtsheim, Neter and Li (2005) or Neter, Kutner, Nachtsheim, Wasserman (1996).

A) Answer question 1.1 and 1.2 on page 33 (or 36 in NKNW).

B) Complete the following questions using problem 1.20. (e.g. the **Copier maintenance** problem on pages 38 & 39).

1) Answer the 4 questions (a, b, c and d) under problem 1.20 from your textbook.

2) In addition to interpreting b_0 in 1.20c, give an interpretation for b_1.

3) Give the complete ANOVA table.

4) Give the normal equations for the model (numerically).

5) Give the values of b_0 and b_1 and their corresponding standard errors.

6) Also answer the 2 questions under problem 1.24 (parts a & b).

B) Complete the following questions using problem 1.20. (e.g. the **Copier maintenance** problem on pages 38 & 39).

C) Complete Problem 2.5; all parts using the values from problem 1.20 (**Copier maintenance** problem). Some parts of this homework may have to be done by hand, but I would encourage you to use SAS (or other package) as much as possible.

D) 2) Problem 2.14; all parts.

1.1. Refer to the sales volume example on page 4. Suppose that the number of units sold is measured accurately, but clerical errors are frequently made in determining the dollar sales. Would the relation between the number of units sold and dollar sales still be a functional one? Discuss.

1.2. The members of a health spa pay annual membership dues of \$300 plus a charge of \$2 for each visit to the spa. Let Y denote the dollar cost for the year for a member and X the number of visits by the member during the year. Express the relation between X and Y mathematically. Is it a functional relation or a statistical relation?

1.20. **Calculator maintenance**. The Tri-City Office Equipment Corporation sells an imported desk calculator on a franchise basis and performs preventive maintenance and repair service on this calculator. The data below have been collected from 18 recent calls on users to perform routine preventive maintenance service; for each call, X is the number of machines serviced and Y is the total number of minutes spent by the service person. Assume that first-order regression model (1.1) is appropriate.

i:	1	2	3	. . .	16	17	18
X_i:	7	6	5	. . .	1	4	5
Y_i:	97	86	78	. . .	17	49	68

a. Obtain the estimated regression function.

b. Plot the estimated regression function and the data. How well does the estimated regression function fit the data?

c. Interpret b_0 in your estimated regression function. Does b_0 provide any relevant information here? Explain.

d. Obtain a point estimate of the mean service time when $X = 5$ machines are serviced.

1.24. Refer to **Calculator maintenance** Problem 1.20.

a. Obtain the residuals e_i and the sum of the squared residuals Σe_i^2. What is the relation between the sum of the squared residuals here and the quantity Q in (1.8)?

b. Obtain point estimates of σ^2 and σ. In what units is σ expressed?

Problem 2.5

2.5. Refer to **Calculator maintenance** Problem 1.20.

a. Estimate the change in the mean service time when the number of machines serviced increases by one. Use a 90 percent confidence interval. Interpret your confidence interval.

b. Conduct a t test to determine whether or not there is a linear association between X and Y here; control the α risk at .10. State the alternatives, decision rule, and conclusion. What is the P-value of your test?

c. Are your results in parts (a) and (b) consistent? Explain.

d. The manufacturer has suggested that the mean required time should not increase by more than 14 minutes for each additional machine that is serviced on a service call. Conduct a test to decide whether this standard is being satisfied by Tri-City. Control the risk of a Type I error at .05. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

e. Does b_0 give any relevant information here about the "start-up" time on calls—i.e., about the time required before service work is begun on the machines at a customer location?

2.14. Refer to **Calculator maintenance** Problem 1.20.

a. Obtain a 90 percent confidence interval for the mean service time on calls in which six machines are serviced. Interpret your confidence interval.

b. Obtain a 90 percent prediction interval for the service time on the next call in which six machines are serviced. Is your prediction interval wider than the corresponding confidence interval in part (a)? Should it be?

c. Management wishes to estimate the expected service time *per machine* on calls in which six machines are serviced. Obtain an appropriate 90 percent confidence interval by converting the interval obtained in part (a). Interpret the converted confidence interval.

d. Determine the boundary values of the 90 percent confidence band for the regression line when $X_h = 6$. Is your confidence band wider at this point than the confidence interval in part (a)? Should it be?