

Homework 4

Linear Regression Models

Multiple Linear Regression Model

DUE: Wednesday, November 8, in class

- (1) Please sign your home work with your name and UNI number.
- (2) Homework is due **Wednesday, November 8, in class**.
- (3) No late homework, under any circumstances, will be accepted.
- (4) At the end of semester, one of your lowest homework scores will be dropped before the final grade is calculated.

Problem 1 (50p) (Problems 6.18 (b-f), 6.21, & 7.7 in ALRM book)

A commercial real estate company evaluates vacancy rates, square footage, rental rates, and operating expenses for commercial properties in a large metropolitan area in order to provide clients with quantitative information upon which to make rental decisions. The data are taken from 81 suburban commercial properties that are the newest, best located, most attractive, and expensive for five specific geographic areas. The variables are: rental rates Y , the age X_1 , operating expanses and taxes X_2 , vacancy rates X_3 , and total square footage X_4 .

Y	X_1	X_2	X_3	X_4
13.500	1	5.02	0.14	123,000
12.000	14	8.19	0.27	104,079
10.500	16	3.00	0.00	39,998
15.000	4	10.70	0.05	57,112
\vdots	\vdots	\vdots	\vdots	\vdots

(see file “Homework 4 data Problem1.txt” for a complete set of data)

(a)(5p) Obtain the scatter plot matrix and the correlation matrix. Interpret these and state your principal findings.

(b)(5p) Fit regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i,$$

for four predictor variables to the data and state the estimated regression function.

(c)(5p) Obtain the residuals and prepare a box plot of the residuals. Does the distribution appear to be fairly symmetrical?

(d)(5p) Plot the residuals against \hat{Y} , each predictor variable, and each two-factor interaction term on separate graphs. Also prepare a normal probability plot. Analyse your plots and summarize your findings.

(e)(5p) Can you conduct a formal test for lack of fit here?

(f)(10p) The commercial real estate company obtained information about additional three properties.

	1	2	3
$X_1 :$	4.0	6.0	12.0
$X_2 :$	10.0	11.5	12.5
$X_3 :$	0.10	0	0.32
$X_4 :$	80,000	120,000	340,000

Find separate prediction intervals for the rental rates for each of the new properties. Use 95% confidence coefficient in each case. Can the rental rates of these three properties be predicted fairly precisely? What is the family confidence level for the set of three predictions?

(g)(10p) Obtain the analysis of variance table that decomposes the regression sum of squares into extra sums of squares associated with X_4 ; with X_1 given X_4 ; with X_2 given X_1 and X_4 ; and with X_3 , given X_1 , X_2 and X_4 .

(h)(5p) Test whether X_3 can be dropped from the regression model given that X_1 , X_2 , and X_4 are retained. Use the F^* test statistic and level of significance 0.01. State the alternatives decision rule, and conclusion. What is the p -value of the test?

Problem 2 (50p) (Problems 8.15 & 8.19 in ALRM book)

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. Assume that the regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

is appropriate, and let $X_2 = 1$ if small model and 0 if large, commercial model.

Y	X_1	X_2
20	2	1
60	4	0
46	3	0
41	2	0
12	1	0
137	10	0
\vdots	\vdots	\vdots

(see file “Homework 4 data Problem2.txt” for a complete set of data)

- (a)(5p) Explain the meaning of all regression coefficients in the model.
- (b)(5p) Fit the regression model and state the estimated regression function.
- (c)(5p) Estimate the effect of copier model on mean service time with a 95 percent confidence interval. Interpret your interval estimate.
- (d)(10p) 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)(10p) Obtain the residuals and plot them against $X_1 X_2$. Is there any indication that an interaction term in the regression model would be helpful?
- (f)(5p) Fit regression model with interaction term as an additional explanatory variable, i.e.,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$$

- (g)(10p) Test whether the interaction term can be dropped from the model; control the α risk at 0.10. State the alternatives, decision rule, and conclusion. What is the p -value

of the test? If the interaction cannot be dropped from the model, describe the nature of the interaction effect.