

## Part 1: 20 points



Refer to the **CDI** data set

You have been asked to evaluate two alternative models for predicting the number of active physicians ( $Y$ ) in a CDI. Proposed model I includes as predictor variables total population ( $X_1$ ), land area ( $X_2$ ), and total personal income ( $X_3$ ). Proposed model II includes as predictor variables population density ( $X_1$ , total population divided by land area), percent of population greater than 64 years old ( $X_2$ ), and total personal income ( $X_3$ ).

- Prepare a histogram for each of the predictor variables. What noteworthy information is provided by your plots?
- Obtain the scatter plot matrix and the correlation matrix for each proposed model. Summarize the information provided.
- For each proposed model, fit the first-order regression model (6.5) with three predictor variables.
- Calculate  $R^2$  for each model. Is one model clearly preferable in terms of this measure?
- For each model, obtain the residuals and plot them against  $\hat{Y}$ , each of the three predictor variables, and each of the two-factor interaction terms. Also prepare a normal probability plot for each of the two fitted models. Interpret your plots and state your findings. Is one model clearly preferable in terms of appropriateness?

## Part 2: 20 points



Refer to the **CDI** data set

- For each geographic region, regress the number of serious crimes in a CDI ( $Y$ ) against population density ( $X_1$ , total population divided by land area), per capita personal income ( $X_2$ ), and percent high school graduates ( $X_3$ ). Use first-order regression model (6.5) with three predictor variables. State the estimated regression functions.
- Are the estimated regression functions similar for the four regions? Discuss.
- Calculate  $MSE$  and  $R^2$  for each region. Are these measures similar for the four regions? Discuss.
- Obtain the residuals for each fitted model and prepare a box plot of the residuals for each fitted model. Interpret your plots and state your findings.

## Part 3: 10 points



For each of the following regression models, indicate whether it is a general linear regression model. If it is not, state whether it can be expressed in the form of (6.7) by a suitable transformation:

- a.  $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 \log_{10} X_{i2} + \beta_3 X_{i1}^2 + \varepsilon_i$
- b.  $Y_i = \varepsilon_i \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}^2)$
- c.  $Y_i = \log_{10}(\beta_1 X_{i1}) + \beta_2 X_{i2} + \varepsilon_i$
- d.  $Y_i = \beta_0 \exp(\beta_1 X_{i1}) + \varepsilon_i$
- e.  $Y_i = [1 + \exp(\beta_0 + \beta_1 X_{i1} + \varepsilon_i)]^{-1}$

## Part 4: 10 points



Consider the multiple regression model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1}^2 + \beta_3 X_{i2} + \varepsilon_i \quad i = 1, \dots, n$$

where the  $\varepsilon_i$  are independent  $N(0, \sigma^2)$ .

- State the least squares criterion and derive the least squares normal equations.
- State the likelihood function and explain why the maximum likelihood estimators will be the same as the least squares estimators.

## Part 5: 20 points



**Patient satisfaction.** A hospital administrator wished to study the relation between patient satisfaction ( $Y$ ) and patient's age ( $X_1$ , in years), severity of illness ( $X_2$ , an index), and anxiety level ( $X_3$ , an index). The administrator randomly selected 46 patients and collected the data presented below, where larger values of  $Y$ ,  $X_2$ , and  $X_3$  are, respectively, associated with more satisfaction, increased severity of illness, and more anxiety.

| $i:$      | 1   | 2   | 3   | ... | 44  | 45  | 46  |
|-----------|-----|-----|-----|-----|-----|-----|-----|
| $X_{i1}:$ | 50  | 36  | 40  | ... | 45  | 37  | 28  |
| $X_{i2}:$ | 51  | 46  | 48  | ... | 51  | 53  | 46  |
| $X_{i3}:$ | 2.3 | 2.3 | 2.2 | ... | 2.2 | 2.1 | 1.8 |
| $Y_i:$    | 48  | 57  | 66  | ... | 68  | 59  | 92  |

## Part 5: 20 points

- Prepare a histogram for each of the predictor variables. Are any noteworthy features revealed by these plots?
- Obtain the scatter plot matrix and the correlation matrix. Interpret these and state your principal findings.
- Fit regression model (6.5) for three predictor variables to the data and state the estimated regression function. How is  $b_2$  interpreted here?
- Obtain the residuals and prepare a box plot of the residuals. Do there appear to be any outliers?
- Plot the residuals against  $\hat{Y}$ , each of the predictor variables, and each two-factor interaction term on separate graphs. Also prepare a normal probability plot. Interpret your plots and summarize your findings.
- Can you conduct a formal test for lack of fit here?
- Conduct the Breusch-Pagan test for constancy of the error variance, assuming  $\log \sigma_i^2 = \gamma_0 + \gamma_1 X_{i1} + \gamma_2 X_{i2} + \gamma_3 X_{i3}$ ; use  $\alpha = .01$ . State the alternatives, decision rule, and conclusion.

## Part 6: 20 points



Refer to **Patient satisfaction** Problem Assume that regression model (6.5) for three predictor variables with independent normal error terms is appropriate.

Test whether there is a regression relation; use  $\alpha = .10$ . State the alternatives, decision rule, and conclusion. What does your test imply about  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$ ? What is the  $P$ -value of the test?

Obtain joint interval estimates of  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$ , using a 90 percent family confidence coefficient. Interpret your results.

Calculate the coefficient of multiple determination. What does it indicate here?

Obtain an interval estimate of the mean satisfaction when  $X_{h1} = 35$ ,  $X_{h2} = 45$ , and  $X_{h3} = 2.2$ . Use a 90 percent confidence coefficient. Interpret your confidence interval.

Obtain a prediction interval for a new patient's satisfaction when  $X_{h1} = 35$ ,  $X_{h2} = 45$ , and  $X_{h3} = 2.2$ . Use a 90 percent confidence coefficient. Interpret your prediction interval.

## Assignment 4

- Submit your responses in Blackboard (in a **single** pdf file) by midnight, Friday, November 10.
- Use the following file name:  
LASTNAME\_FIRSTNAME\_ASUID\_ASSIGNMENTNUMBER
- Prepare your pdfs carefully; each week some of you will present their work. You are expected to work alone.
- Include the R commands you used.