STAT 441/541 Statistical Methods II

Homework Assignment 3 Multiple Regression Part 1

Submit a single pdf document to the Dropbox folder *Homework Assignment 3 Multiple Regression Part 1*.

Start solution for each dataset on a new page.

Dataset 1

Scenario

Mercury contamination in freshwater fish has been a recognized problem in North America for over four decades. High concentrations of mercury in fish can pose a serious health threat to humans and birds. The purpose of this assignment is to investigate the relationships between mercury concentrations and selected physical and chemical lake characteristics. Data has been collected to determine if these characteristics strongly influenced the bioaccumulation of mercury in largemouth bass. The study included 53 lakes of various sizes which are hydrologically diverse.

The data set contains the following variables:

Lake ID numbers of the 53 lakes

EHg Estimated mercury concentration () for a three-year-old fish

Alk Alkalinity level of lake ()

pH Degree of acidity () or alkalinity ()

Ca Calcium level ()

Chlo Chlorophyll ()

The R code file: Mercury in Fish Part 1 R Code.R

The dataset is Excel file: Mercury in Fish.xlsx

Multiple regression will be used with EHg as the dependent variable and Alk, pH, Ca, and Chlo as independent variables. For consistency, we will give the model in terms of the mnemonic variable names. That is, we will use the variable names EHg, Alk, pH, Ca, and Chlo in the model equation.

**(a)** Paste the scatterplot matrix from R output. Looking at the scatterplot matrix:

(i) Describe relationships between the dependent variable EHg and independent variables Alk, pH, Ca, and Chlo.

(ii) Describe any interesting relationships among the independent variables.

**(b)** Paste the R output that gives the correlation matrix, number of observations, and p-values for the corresponding correlation coefficients. Looking at the correlation matrix:

> rcorr(mat, type="pearson")

EHg Alk pH Ca Chlo

EHg 1.00 -0.63 -0.61 -0.46 -0.51

Alk -0.63 1.00 0.72 0.83 0.48

pH -0.61 0.72 1.00 0.58 0.61

Ca -0.46 0.83 0.58 1.00 0.41

Chlo -0.51 0.48 0.61 0.41 1.00

n= 53

P

EHg Alk pH Ca Chlo

EHg 0.0000 0.0000 0.0005 0.0001

Alk 0.0000 0.0000 0.0000 0.0003

pH 0.0000 0.0000 0.0000 0.0000

Ca 0.0005 0.0000 0.0000 0.0023

Chlo 0.0001 0.0003 0.0000 0.0023

(i) Are there any significant correlations between the dependent variable and independent variables Alk, pH, Ca, and Chlo at a significance level of ? If so, which ones? Justify your answer by giving the sample correlation coefficients and their p-values.

(ii) Are there significant correlations among independent variables at a significance level of ? If so, which ones? Justify your answer by giving the sample correlation coefficients and their p-values.

(iii) Does the scatter plot of EHg and Alk reflect the size of the sample correlation coefficient for EHg and Alk? Justify your answer.

**(c)** Paste the Coefficient Table from R output. Give the estimated regression model and interpret each of the estimated slope parameters for Alk, pH, Ca, and Chlo.

**(d)** Paste the regression summary information that is below the Coefficient Table from R output and perform an overall F test for the proposed multiple regression model using the five-step method and a significance level of :

Hypotheses:

Test Statistic:

P-value:

Decision about the null hypothesis:

Conclusion:

**(e)** Paste the Coefficient Table from R output and test the partial slope parameter for Alk using the five-step method and a significance level of :

Test for :

Hypotheses:

Test Statistic:

P-value:

Decision about the null hypothesis:

Conclusion:

**(f)** Use the Coefficient Table from Part (e) and test the partial slope parameter for Ca using the five-step method and a significance level of :

(ii) Test for :

Hypotheses:

Test Statistic:

P-value:

Decision about the null hypothesis:

Conclusion: