Instruction:

(A) Questions in this paper should be answered by students whose **surnames** fall within the range **N-Z**.

(B) Use the Excel file '**Dataset3_part3b** to answer the questions asked.

(**C**) A heavy penalty will be applied if your answers are not based on dataset assigned to you.

Instructions for Dataset3_part3b: Multiple Regression Analysis

A random sample of 1800 women working in manufacturing industries in country S were interviewed and the following information was collected (and saved in **Dataset3_part3b**): hourly wage in dollars; current grade completed by the employee, number of hours worked per week; and place of residence.

The variables saved in **Dataset3_part3b** are:

- hw (Y, hourly wage in dollars)
- cgc (X1, current grade completed by the employee)
- nhw (X2, number of hours worked per week)
- resi (X3, place of residence, coded 1 if the employee lives in south and 0 otherwise)

The dependent variable for your analysis is **hw**.

Answer the following questions using Dataset3_part3b

(a) Estimate a regression model using X1 and X2 to predict Y (state the multiple regression equation).

(b) Interpret the meaning of the slopes.

(c) Predict Y when X1 = 14 and X2 = 40.

(d) Compute a 95% confidence interval estimate of the mean Y for all women working in manufacturing industries in country S when X1 = 14 and X2 = 40 and interpret its meaning.

(e) Compute a 95% prediction interval of Y for a woman working in a manufacturing industry in country S when X1 = 14 and X2 = 40 and interpret its meaning.

(f) Plot the residuals to test the assumptions of the regression model. Is there any evidence of violation of the regression assumptions? Explain.

(g) Determine the variance inflation factor (VIF) for each independent variable (X1 and X2) in the model. Is there reason to suspect the existence of collinearity? Why?

(h) At the 0.05 level of significance, determine whether each independent variable (X1 and X2) makes a significant contribution to the regression model (use t tests and follow all the necessary steps). On the basis of these results, indicate the independent variables to include in the model.

(i) Test for the significance of the overall multiple regression model (with two independent variables, X1 and X2) at 5% level of significance.

(j) Determine whether there is a significant relationship between Y and each independent variable (X1 and X2) at the 5% level of significance (hint: testing portions of the multiple regression model using the partial F test).

(k) Compute the coefficients of partial determination for a multiple regression model containing X1 and X2 and interpret their meaning.

(l) Estimate a regression model using X1, X2 and X3 to predict Y (state the multiple regression equation, the regression equation for women (working in manufacturing industries) living in south, the regression equation for women (working in manufacturing industries) not living in south) and interpret the coefficient for X3.

(m) Estimate a regression model using X1, X2, X3, an interaction between X1 and X2, an interaction between X1 and X3, and an interaction between X2 and X3 to predict Y.

(n) Test whether the three interactions significantly improve the regression model. Assume 5% level of significance (hint: test the joint significance of the three interaction terms using the partial F test. If you reject the null hypothesis, test the contribution of each interaction separately (using the partial F test) in order to determine which interaction terms to include in the model).