DEPARTMENT OF POLITICAL SCIENCE
AND
INTERNATIONAL RELATIONS

Posc/Uapp 816

Assignment 8
MULTIVARIATE RELATIONSHIPS

Name_____
(Printed)
Student Number_____
(Social Security Number)
E-mail_____

The purpose of the questions in this assignment is to make sure everyone understands standardized variables and the concept of "statistical controls."

1. Suppose you wanted to explain cross-national variation in test scores, a topic on everyone's mind. Consider a couple of hypotheses. Scores are related to how much stress a society places on education, to its level of well-being, and/or to student work habits. Here are some data that might allow us to investigate these ideas, however superficially.[1] The data are for 13 countries that participated in the "International Assessment of Educational Progress."
   A. The variables are:
      i. Average percent correct on mathematics test for 9 year old students, 1991.
      ii. Average days in a school year.
      iii. Percent of students with two or more hours of homework daily.
      iv. Percent of students who report watching TV 5 hours or more a day.
      v. GNP per capita (dollars), 1991
      vi. Infant mortality rate, number of deaths of children under 1 year of age per 1,000 live births, 1994.

[1]Source: National Center for Educational Statistics, ***Digest of Educational Statistics 1994***; and Bureau of the Census, ***Statistical Abstract of the United States***.

Country	Percent correct	School days	Homework	Percent watching TV	GNP per capita	Infant mortality
Korea	**74.8**	**222**	**22**	**9**	**8639**	**21.7**
Hungary	**68.2**	**177**	**25**	**16**	**5727**	**12.5**
Taiwan	**68.1**	**222**	**31**	**8**	**9068**	**5.7**
Russia	**65.9**	**198**	**31**	**18**	**8639**	**27.0**
Israel	**64.4**	**215**	**35**	**24**	**13950**	**8.6**
Spain	**61.9**	**188**	**29**	**17**	**13370**	**6.9**
Ireland	**60.0**	**173**	**18**	**23**	**NA**	**NA**
Canada	**59.9**	**188**	**13**	**22**	**20840**	**6.9**
U.S.	**58.4**	**178**	**20**	**26**	**22550**	**8.1**
Slovenia	**55.8**	**190**	**15**	**8**	**NA**	**10.4**
Italy	**67.8**	**204**	**17**	**9**	**19630**	**7.6**
England	**59.5**	**192**	**9**	**23**	**17400**	**7.2**
Portugal	**55.5**	**172**	**20**	**20**	**6251**	**9.5**

B. Try building a model for these data. First obtain a correlation matrix so that you can anticipate results and problems.

i. Enter the correlation coefficients here.

	Percent correct	School days	Homework	TV	GNP	Infant mortality
School days						
Homework						
TV						
GNP						
Infant mortality						

C. Now, regress percent correct on number of school days.

i. The estimated model is:_____

ii. What is TSS? _____ R^2 ?_____

iii. Standard deviation about regression is _____

iv. Observed F for the model is _____ with _____ degrees of freedom.

v. Would you accept or reject the null hypothesis that days of schooling affects test scores? Why? Report what ever statistics are necessary and verbally state the substantive meaning of the significance test.

_____

_____

_____

_____

D. Try adding the student behavior variable, "percent reporting watching more than 5 hours of TV." That is, obtain the multiple regression of scores on school days and percent watching more than 5 hours.

i. What is the estimated model? _____

ii. Explain the meaning of the coefficient linking "percent watching" to scores.

_____

_____

_____

_____

iii. What is the TSS? _____ Now what is R^2 ?_____

iv. Test the model as a whole for significance. That is, test they hypothesis that $\beta_1 = \beta_2 = 0$. Use the F test results. What is your conclusion? Is the model as a whole significant? Report both the relevant test statistics and explain them **briefly**.

E. Assuming that television doesn't add much to the model, drop it and add the well being variable, infant mortality.

i. Now what is the estimated model? _____

ii. What is the TSS? _____ Now what is R^2 ?_____

iii. Test the model as a whole for significance. That is, test they hypothesis that $\beta_1 = \beta_2 = 0$. Use the F test results. What is your conclusion? Is the model as a whole significant? Report both the relevant test statistics and explain them **briefly**.

iv. Does the infant mortality "add" anything "significant" to the model? That is, is the increase in explained variation (or decrease in unexplained variation) statistically significant?

2. Try answering this question from Agresti and Finlay, ***Statistical Methods for Social Sciences,*** 3rd edition, page 380. "The percentage of women who get breast cancer is higher now than at the beginning of this century. Suppose that cancer incidence tends to increase with age, and suppose that women tend to live longer lives now than earlier in this century. Explain why a comparison of breast cancer rates now with the beginning of this century could show different results from these if we controlled for the age of the woman." Use diagrams of the sort shown in class to help you explain.

Go to Assignment page
Go to Statistics page