

SCHOOL COSTS DATA SET

REGRESSION EXERCISES

c.dougherty@lse.ac.uk
April 2002

TABLE OF CONTENTS

1. Description of the data set
2. Exercises
3. Instructions for doing the exercises using Stata or EViews.

DOWNLOADING THE DATA SET

You will find the School Costs data set in three different formats at the Introduction to Econometrics web site, URL <http://econ.lse.ac.uk/ie/>. Stata format for Stata users, EViews format for EViews users, and comma-delimited ASCII (text format) for everybody else. It is also available in Excel format by email. To download, click on the filename and follow the instructions in the dialogue box. Note that you may wish to download to a directory other than that specified as the default in the dialogue box. If you are using the Stata file, the file will be called SCHOOLS.dta and it will be ready for use. If you are using the EViews file, it may download as SCHOOLS_wf1.bin. It should have extension wf1, not bin, so rename it as SCHOOLS.wf1. To do this, go to My Computer, browse until you find the downloaded file, click on File and then Rename. You will now be able to delete the .bin extension and to replace _wf1 with .wf1.

1. DESCRIPTION OF THE DATA SET

The data set is an extract taken with permission from a series of annual surveys of 105 Shanghai secondary schools undertaken undertaken by Fujian University staff during the period 1981-1985 with the support of the World Bank. The data in this data set relate to 1985. It was a time of rapid expansion of secondary education, both in terms of the number of schools and in the enrollments of some. To guard against the possibility that enrolments and budgets might be in disequilibrium, the 24 schools with enrolments that had increased (or decreased) by more than one third over the previous year were excluded from the sample. Likewise seven schools with incomplete data were excluded, leaving 74 in the sample. The data on capital expenditure were very volatile and not susceptible to meaningful analysis. The recurrent cost data comprised staff costs, non-staff administrative expenditure, non-staff instructional expenditure, expenditure on books and expenditure on utilities. Maintenance expenditure was excluded because, like capital expenditure, it was very volatile and seemingly determined more by the availability of a budget rather than actual year-to-year need.

Variables

<i>COST</i>	C	annual recurrent cost, in yuan (worth about US\$0.25 at the time)
<i>N</i>	C	number of students enrolled
<i>OCC</i>	D	occupational school (1 if technical or skilled worker's school, 0 if general or vocational school)
<i>REGULAR</i>	D	occupational school (1 if general or vocational school, 0 if technical or skilled worker's school)
<i>TECH</i>	D	technical school (1 if technical school, 0 otherwise)
<i>WORKER</i>	D	skilled workers' school (1 if skilled workers' school, 0 otherwise)
<i>VOC</i>	D	vocational school (1 if vocational school, 0 otherwise)
<i>GEN</i>	D	general school (1 if general school, 0 otherwise)
<i>RES</i>	D	residential school (1 if residential school, 0 otherwise)

C indicates a continuous variable, D a dummy variable.

2. EXERCISES

Note: These exercises merely replicate the regressions in the text.

Exercise 1 A simple school costs function

How do school costs vary with the number of students enrolled?

This is a baseline regression that will be used in comparisons with the dummy variable regressions. Regress *COST* on *N* and perform appropriate statistical tests. [Section 6.1.]

Exercise 2 Regression with a dummy variable

Are the costs of occupational schools different from those of regular schools?

Regress *COST* on *N* and *OCC*, and perform appropriate statistical tests. [Section 6.1.]

Exercise 3 Regression with more than two categories in the dummy variable classification

Are the costs of technical, skilled workers', and vocational schools different from those of general schools?

Regress *COST* on *N*, *TECH*, *WORKER*, and *VOC*, and perform appropriate statistical tests. [Section 6.2.]

Exercise 4 Regression with more than two categories in the dummy variable classification

How does the choice of reference (omitted) category affect the regression results?

Regress *COST* on *N*, *TECH*, *VOC*, and *GEN*, perform appropriate statistical tests, and compare the results with those of Exercise 3. [Section 6.2.]

Exercise 5 Test of the joint explanatory power of a set of dummy variables.

Do the technical, skilled workers', and vocational schools dummy variables jointly have significant explanatory power?

Perform an F test comparing the residual sum of squares (RSS) in Exercises 1 and 4. [Section 6.2.]

Exercise 6 Regression with two sets of dummy variables

Are the costs of occupational and/or residential schools different from those of non-residential general schools?

Regress $COST$ on N , OCC , and RES , and perform appropriate statistical tests. [Section 6.2.]

Exercise 7 Regression with a slope dummy variable

Are the marginal costs of occupational schools different from those of general schools?

Define a slope dummy variable $NOCC$ as the product of N and OCC , regress $COST$ on N , OCC , and $NOCC$, and perform appropriate statistical tests, including an F test on the joint explanatory power of the intercept and slope dummy variables. For the latter, you will need to compare RSS for this regression with that for the regression in Exercise 1. [Section 6.3.]

Exercise 8 Chow test

Are the cost functions of occupational schools different from those of general schools?

Run simple regressions of $COST$ on N , (1) for the whole sample, (2) for the regular schools only, and (3) for the occupational schools only, and perform a Chow test. Check that the F statistic is identical to that for the test of the joint explanatory power of the dummy variables in Exercise 5. [Section 6.4.]

3. INSTRUCTIONS FOR DOING THE EXERCISES USING STATA OR EViews**Stata**

```

Exercise 1:  reg COST N
Exercise 2:  reg COST N OCC
Exercise 3:  reg COST N TECH WORKER VOC
Exercise 4:  reg COST N TECH VOC GEN
Exercise 6:  reg COST N OCC RES
Exercise 7:  gen NOCC = N*OCC
             reg COST N OCC NOCC
Exercise 8:  reg COST N
             reg COST N if OCC==1
             reg COST N if OCC==0

```

EViews (from the command line)

Exercise 1: LS COST C N
Exercise 2: LS COST C N OCC
Exercise 3: LS COST C N TECH WORKER VOC
Exercise 4: LS COST C N TECH VOC GEN
Exercise 6: LS COST C N OCC RES
Exercise 7: GENR NOCC = N*OCC
LS COST C N OCC NOCC
Exercise 8: LS COST C N
SMPL OCC=1
LS COST C N
SMPL OCC=0
LS COST C N