

Exploring Cross Sectional Data: The Comparative Health and Health Spending of Nations

1. Introduction

In this unit we explore the nature of cross-sectional data, which means observations of different “units” at the same point in time, in this case the set of 30 OECD countries in a given year. An alternative type of data, time-series data, consists of observations of the same unit over several time periods, e.g. the level of health spending in the U.K., annually from 1960 to 2004. (These two dimensions are sometimes combined in what’s called a “panel dataset.” Such a dataset might contain, for example, observations on the health expenditures of all European countries for each year from 1960 to 2001.)

An excellent resource for international comparative health data is the Organization of Economic Cooperative Development. The OECD is a public body that collects and analyzes primarily economic data for 30 relatively advanced economies.

It is possible to access the OECD health database directly online, via the University library homepage: choose the “Databases” tab; select (on the left) “Business and Economics” then “Economics and Statistics”; and the “OECD Health Data” should appear as an option on the right. However, for the purposes of the assignment on this topic we will access the OECD data using the program gretl. Reminder: there is a link to the program’s installer on the ECN 201 homepage.

2. The assignment

The final version of your report should be done as a Word document. Tables and graphs should be clearly labeled. The specifics of organization of the material, textual answers to the questions asked and placement of the graphs in the final document are left to your discretion. Grading will be based on completeness, clarity, neatness and presentation value.

The final version will be strictly due on **Sept. 18**. Late papers without a medical excuse will be docked a letter grade per day as of Sept. 19 as described on the syllabus. This strictness is necessary due to our tight schedule.

1. Select an OECD data file: if your family name begins with A–H, you should use OECD1998.gdt, otherwise you should use OECD2001.gdt. These file are available from the OECD folder inside the “Other docs” folder on the class website. They are password-protected; you will need to give your WFU username and password.

Note that the specific names for the variables will depend on the data file you use. For example, national expenditure on health care is called hexp98 in the 1998

data file and hexp01 in the 2001 file. I will refer to the variables generically below, leaving off the two final numbers. For example, health expenditure will be called hexp.

2. **Make a table** containing the variables total health expenditure per capita (hexp) and GDP per capita (gdppc). Sort the data by gdppc. Find the mean and median for each series. Display this as Table 1 in your report, and answer the following questions in relation to the table:

- (a) Which countries have the highest and the lowest GDP per capita?
- (b) Which countries have the highest and the lowest health expenditure per capita?
- (c) Where does the U.S. stand in relation to the mean and median values of these variables?

3. **Make an X-Y graph** with GDP per capita on the X axis and health expenditures per capita on the Y axis. Display this as Figure 1 in your report and answer the following questions about it:

- (a) Interpret the scatter plot and the estimated regression line. What does it say about the estimated relationship and the explanatory value of GDP in predicting a country's level of spending on health?
- (b) What countries are the most obvious "outliers?" Try to explain this.
- (c) What other variables would you want to include in an investigation of a country's total health spending, besides per-capita GDP?

4. **Make a table** including the following variables

- § Total Expenditures on Health per capita (hexp)
- § life expectancy of the total population at birth (lexp)
- § Practicing Physician Density/1000 pop (phys)

Sort the table in order of life expectancy. Find the mean and median for each series. Display this as Table 2 in your report and answer the following:

- (a) Which countries have the highest and lowest life spans?
- (b) Who has the highest and lowest number of physicians?
- (c) Who has the highest and lowest health spending?
- (d) Where does the U.S. fall in relation the means of these series?

5. **Make an X-Y graph** with health spending (hexp) as the X variable and life expectancy (lexp) as Y. Display this as Figure 2 in your report and answer the following:

- (a) Taken at face value, what does the line of best fit suggest? For instance, how many extra years of life expectancy are gained per \$1000 per capita spent on health? (The unit is dollars per capita, so you'll need to multiply the slope by 1000.)
- (b) What life expectancy does the equation predict for a country that spends nothing on health?
- (c) Any comment on Turkey?
- (d) Does this relationship seem to be linear? If not, what sort of shape might fit better?

6. **Make an X-Y graph** with health spending (hexp) as the X variable and infant mortality (infm) as Y. Display this as Figure 3 in your report and answer the following:

- (a) Describe the relationship shown in the graph.
- (b) The relationship does not seem to be linear. Try to explain why one would not really expect it to be linear.
- (c) Comment on the position of the U.S. in the distribution (here it might help to create a table too).

7. **Model a Single Variable Linear Regression.** Select “model” from the menu at top. Run an ordinary least squares regression where health expenditure per capita (hexp) as the dependent variable and % of population over 80 (o80) as the independent variable.

- (a) Write down the equation of the line of best fit.
- (b) Describe exactly, per percentage unit of population over 80, what it is saying.
- (c) Copy the regression to the clipboard and paste it into your WORD document. Shorten the output list of diagnostic statistics to what we described in class and what we shortened it to in the document “**The Basics of Single Variable Linear Regression in GRET**L.”

(d) Interpret each line of this shortened list. Generally is this relationship between health expenditures and % population over 80 a strong or a weak one?

(e) In the GRETl regression output window, click on “Graphs/Fitted, Actual Plot/ against o8o.” Pass your mouse over the outliers to label what country they represent. Copy the graph to the clipboard and paste it into your document. Write a few lines describing the outliers. What do they say about the relationship?

8. Make two more X-Y graphs This time you have a free choice: select any two pairs of variables from the data file that you think might reveal interesting relationships. Think in terms of cause and effect relationships: select for the X axis a variable that you think may have a causal influence on your Y-axis variable. In each case, display the graph and give an account of what it shows. Are there any surprises? Do any particular countries stand out as exceptional?

Just in case you're stumped, here are some possible questions that could guide your selection: Do more doctors seem to translate into longer life or lower infant mortality? More nurses? What about medical technology? How closely does health spending relate to the prevalence of high-tech medical equipment? Is there a clear relationship between high-tech equipment and health outcomes such as life expectancy?