14.75J 9/23/2003
Prof. Victor Lavy
Problem Set 1 Solutions

Feel free to work in groups, but please turn in your own individual results.

General comments: Please pay close attention to what the question is asking and answer it as directly as possible. Also, for questions where there is a lot of writing, please turn it in typewritten.

1. World Bank Development Report Data

The point of this exercise is to help you understand the faces of development on the one hand, and acquaint you with Stata on the other. You will need to download the Stata data set (wbdr.dta; this data set contains 1997 variables, unless otherwise noted) from the course website. There you will also find instructions for using Stata. Hand in your do-file with the problem set. No need to hand in the log file, although you can hand in the parts that are *directly* relevant for answering questions below.

> a. What is the mean and standard deviation of GNP per capita, illiteracy rate, infant mortality rate and under 5 mortality rate in 1997 across countries?

Variable	Obs	Mean	Std. Dev.	Min	Max
gnppc	160	5340.887	9307.99	112	46448
illit_t	138	20.94928	21.7864	0	86
mort_inf	195	41.16923	38.40462	4	170
mort_5	177	62.48588	64.37586	5	286

> b. Compute the correlation coefficient among GNP per capita, illiteracy rate, infant mortality rate and under 5 mortality rate in 1997. Are the directions of the correlations with GNP per capita what you expected?

	gnppc	illit_t	mort_inf	mort_5
gnppc	1.0000			
illit_t	-0.4820	1.0000		
mort_inf	-0.5531	0.8170	1.0000	
mort_5	-0.5051	0.8043	0.9908	1.0000

The correlation coefficient describes the relationship between two variables. It is bounded between –1 and 1. A high absolute value of the correlation coefficient means that the two variables move closely together. For example, people's heights and weights are strongly positively correlated, so their correlation is positive and close to 1. When the correlation coefficient is close to zero, then the two variables do not move together. Remember correlation does not imply causality. Just because two variables are highly correlated does not mean one causes the other.

The correlation coefficients in the above table are what we expected. We expect higher GNP countries to have better education and health because they can spend more on schools and health care. Conversely, with a more educated and healthier labor force, countries can produce more. Hence, the negative correlation between gnppc and both mortality and illiteracy rates. The positive correlation among the three measurements of health care and education is also as expected. They are all indicators of poverty, and low education countries are usually countries with poor health care as well.

    c. Compare the male illiteracy rate to the female illiteracy rate. Why might they be different?

```
                                   ---------- difference ----------
                        count     minimum      average     maximum
-----------------------------------------------------------------
illit_m<illit_f            98         -43    -13.10204          -1
illit_m=illit_f            30
illit_m>illit_f            10           1            5          22
                    ----------
jointly defined           138         -43    -8.942029          22
jointly missing            72
                    ----------
total                     210
```

correlation:

```
             |  illit_m  illit_f
-------------+------------------
     illit_m |   1.0000
     illit_f |   0.9433   1.0000
```

```
    Variable |        Obs        Mean    Std. Dev.        Min         Max
-------------+--------------------------------------------------------
     illit_m |        138    16.43478     17.89261          0          78
     illit_f |        138    25.37681     26.20977          0          93
```

Countries with high male illiteracy tend to have high female illiteracy also-we observe this from the very strong positive correlation. However, male illiteracy is usually less than female illiteracy, as shown in the first table above which comes from the "compare" command. This can also be seen from the fact that in the third table, the mean is much lower on male illiteracy. You can also see from this table that the standard deviation of female illiteracy is much higher—this may be because in poor countries with high illiteracy the women tend to have much higher illiteracy than men, while the richer countries with low illiteracy women also tend to have more gender equality in education so that the illiteracy rate for women is just as low as it is for men. As a result, women's illiteracy varies more widely than men's does.

d. The richest and poorest countries:

    i. Restrict your data set to countries for which we have GNP per capita   data for both 1997 and 1977. What were the ten poorest countries (in terms of GNP per capita) in 1997? In 1977? Are there countries that are in both lists?

1997:

```
     +--------------------------+
     |          country   gnppc |
     |--------------------------|
  1. | Congo, Dem. Rep.     114 |
  2. |          Burundi     141 |
  3. |     Sierra Leone     150 |
  4. |          Malawi      163 |
  5. |           Niger      202 |
     |--------------------------|
  6. |          Rwanda      207 |
  7. |           Nepal      216 |
  8. |            Chad      218 |
  9. |       Madagascar     229 |
 10. |     Guinea-Bissau    232 |
     +--------------------------+
```

1977:

```
     +------------------------+
     |        country  gnppc77 |
     |------------------------|
  1. |          China     129 |
  2. |          Nepal     156 |
  3. |          Malawi    156 |
  4. |          Burundi   180 |
  5. |    Guinea-Bissau   187 |
     |------------------------|
  6. |       Bangladesh   208 |
  7. |     Burkina Faso   209 |
  8. |           India    214 |
  9. |            Chad    242 |
 10. |          Rwanda    265 |
     +------------------------+
```

Burundi, Malawi, Rwanda, Nepal, Chad, and Guinea Bissau are on both lists. The most common mistake in this and the next question was to forget to restrict the data set to countries which have data for both years.

    ii. Continue using the restricted data set. What were the five richest countries (in terms of GNP per capita) in 1997? In 1977? Are there countries that are in both lists?

```
     +---------------------+
     |       country  gnppc |
     |---------------------|
107. |    Singapore   32486 |
108. |      Norway    35947 |
```

```
109. |      Denmark    36418 |
110. |       Japan     43574 |
111. | Switzerland     46448 |
     +------------------------+
```

```
     +------------------------+
     |        country  gnppc77 |
     |------------------------|
107. |         Sweden    20857 |
108. | United States    21050 |
109. |          Japan    24845 |
110. |        Denmark    24987 |
111. |    Switzerland    38314 |
```

Switzerland, Japan and Denmark are in both lists.

iii. Return to using the original data set. What is the correlation between 1997 and 1977 GNP per capita? Between the 1997 and 1977 infant mortality rates? How do you interpret these correlations?

```
             |    gnppc   gnppc77
-------------+------------------
      gnppc  |   1.0000
    gnppc77  |   0.9616    1.0000
```

```
             | mort_inf    mort77
-------------+------------------
   mort_inf  |   1.0000
     mort77  |   0.9102    1.0000
```

These high and positive correlations say that countries that were relatively poor in 1977 – either measured by GNP or infant mortality – were still relatively poor in 1997. There is great persistence to poverty, as already suggested by parts (i) when we saw that 6 out of 10 countries that were poorest in 1977 were still poorest in 1997, and (ii) when we saw that 3 out of 5 countries that were richest in 1997 were still richest in 1997.

Note that the level of GNP could have risen, and the level of infant mortality could have fallen, for every country. That is, over time, every country could be doing somewhat better in terms of income and health. However, on a relative basis, poor countries stay poor and rich countries stay rich.

2. This problem is intended to help you gain insight into cross country growth comparisons and understand better the pros and cons of the different ways to measure welfare. You will use WDI data (http://devdata.worldbank.org/dataonline/) for 5 countries over a period spanning at least 40 years from 2002. Please email the TA the countries you choose by 4pm Thursday (9/11), so that there is not too much overlap in the class over countries.

Much of this question is pulled directly from Chapter 2 of the textbook. It discusses measures of welfare, measurement issues, and the relationship between growth and inequality.

a. Choose 4 measurements of welfare. For each measurement, use Excel to graph the changes over a period (as long as possible) for the 5 countries of your choice. (Please put welfare on the Y-axis and years on the X-axis.) You should produce 4 graphs.

A comment: when graphing series that do not have an observation for every year, do not graph a zero for years when there is no data, leave those years blank. The graph will be much easier to read that way, it is very hard to interpret when it is continually bouncing down to zero.

I didn't grade your choice of measurement, but some are clearly better measures than others. GDP per capita in constant dollars is more informative than GDP per capita in current dollars which is in turn better than simple GDP.

b. Compare the 4 graphs. Discuss why they might look different (or the same) and discuss the pros and cons of each measure of wealth. (Hint: How is each measurement constructed? What information does each measurement reveal? How might each measurement mislead a researcher?)

The answer to this question depends a lot on which four measurements you chose to use. Note that the main thing the question asks about is issues about each of the measurements. The measurements are the focus of the question, so discussing the differences among the countries that seem to be revealed through these measurements is not enough.

For example, suppose I chose GDP per capita in constant 1995 dollars. This is a measure of the total amount of goods and services produced in the country, divided by the population of the country and transformed into 1995 dollars using the exchange rate and CPI. This is a measure of income per person because production and sale of goods and services is how people generate income. Since the amount of income people have is a very important determinant of their welfare, this is quite a good welfare measure. It is also good that it is in contant dollars, that means it takes into account inflation, so a country will not appear to have increasing welfare just because of inflation. It has several drawbacks, however. It is only a measure of average income, so if there are a few very rich people and many very poor people it is possible for most of the country to be much worse off than the per capita GDP would imply. It also may not be the best way to make cross-country comparisons for reasons discussed in parts (c) and (d).

So for each measurement, you should explain why you want to use it as a measure of welfare and why it is imperfect. Then you should compare the different graphs and how the comparison of the different countries differs with the different measures—a country with low per capita income could have high life expectancy, and you might say that implies the country probably has a good health care system, maybe provided by the government.

c. Recall the in class discussion about exchange rates and non-tradeable goods such as services. What other measurement issues do you have to take into account when making cross-country comparisons?

Although there were many different possible answers to this question which I gave full credit to, referring to different cultures, geography, government and economic systems and so on, the intended answer to this question is on pages 11-16 of the textbook. There are two main reasons mentioned her:e developing countries tend to have less efficient tax collection systems and so a greater tendency for income to be underreported. There may also be a lot of subsistence farming, and since that food is never sold it is very difficult to count in income measures.

    d. What measurements do you think should be used to get the most accurate picture of a countries' growth in wealth? Cross-country comparisons? (You can use more than one).

There are many possible answers here as well, but the best one would be that to measure a country's growth, you will want to measure everything in real values, rather than nominal values, because that corrects for inflation so you know how much their income is changing in real terms. For cross-country comparisons, the purchasing power parity (PPP) measure should be used to compare countries because it was designed for that purpose, and although it isn't perfect it is the best we have. (see pp.12-15 of the book)

    e. Why might a researcher want to look at the income share of different segments of income distribution (e.g. top 10%) over time? For one country, graph the income share held by the lowest 20%, second lowest 20%, the highest 10% and the highest 20%. What does this graph say about changes income inequality?

skipped

    f. Hypothesize about the relationship between growth and inequality: Do you think that countries which grow faster will experience more inequality? Do you think rich countries or poor countries generally experience more inequality? Please give a brief – 4 sentence – justification.

Once again, there are many possible answers to this question, and there is truly no right answer because these questions are not settled, but there is a very good overview of the issues in the textbook, pages 21-25.