Measurement of Policy Outcomes: Homework 6

Please send your answers before December 2nd, 2021 11:59 A.M. to measurement.homework@gmail.com

Date of the tutorial session: December 3rd, 2021

Please try to keep your document within a reasonable length.

Education in France (10 points)

The dataset **panel97.dta** is an excerpt from the 1997 panel of French primary school pupils. It contains 9641 observations, 1 for each pupil.

French primary school normally lasts 5 years, from age 6 to age 11. The first to fith grades are called, respectively, "**cp**", "**ce1**", "**ce2**", "**cm1**" and "**cm2**" (literally, "*preparatory course*", "*elementary course 1*", "*elementary course 2*", "*intermediate course 1*" and "*intermediate course 2*").

The sampled pupils were tested at their entry into primary school and were then followed for 5 to 6 years. They were tested again at the beginning of the 3rd grade of primary school, whether they reached that grade in 1999 or in 2000, i.e. whether they did not repeat any grade between the first and the third, or they repeated either the first or the second. Test 2 yielded two separate scores, one measuring numeracy (math) and the other measuring literacy (proficiency in the French language). Finally, pupils were tested again, spearately for math and literacy, soon after the end of primary school, i.e. when they entered middle school (in their "6th grade"). Pupils who did not repeat a grade in primary school were tested in 2002 whereas those who repeated once in primary school were tested in 2003.

The variables available in the dataset are detailed in the appendix.

Questions

(a) Compare the distributions of raw test scores for each test. Comment. (**1 point**)

(b) How many missing values are there for each test score? Is this a problem if you want to study pupils' trajectories or progress? (**1 point**)

(c) Plot the global score after 3 years against the entry score. Can you find a more meaningful way to present this information? (**1 point**)

(d) Analyzing "progress" (**2 points**)

- With the variables at hand, how would you measure a pupil's progress between the first and the third year of primary school? Between the third year and the final year of primary school?
- Implement the strategy you suggest and provide a summary of the distributions of the variables you computed.

(e) Mobility between two successive tests. Focus on test scores expressed in deciles (they are stored under variables prefixed by "`d_`" in the dataset) (**1 point**)

- The point of this question is to compute transition matrices between variables present in your dataset. Each cell $p_{i,j}$ of the 10x10 transition matrix between test A and test B represents the probability (as estimated by the realized share of students) that a student will be in decile j on test B, conditional on her score being in decile i on test A.
- Compute transition matrices across deciles of the global score distribution between test 1 and test 2, then between test 2 and test 3. Comment.

(f) In France, children are supposed to start primary school the year they turn six. Hence a school cohort typically includes pupils born between January 1st and December 31st of a given year.

(a) Compare the global test scores of pupils born in January vs. those born in December on each of the three tests ("`gscore_t1`", "`gscore_t2`", "`gscore_t3`") and provide the 95% confidence interval for the difference between test scores. Interpret your findings. (**1.5 points**)

(b) Compare the evolution of the January vs. December global test score gap in the different years with the evolution of the global test score gap between children of executives/professionals and children of blue-collars (use the "`parent_occ`" variable), with the 95% confidence interval for that gap. Interpret your results. (**1.5 points**)

(c) Why is the comparison of test scores of children born in January vs. December likely to underestimate the true effect of relative age differences on test scores? (**1 point**)

Human Development Index: Old vs. New Definition (10 points)

The Human Development Index (HDI) is a composite index that aims to provide a broader characterization of "development" than is possible by focusing on national income alone. For this purpose, the HDI aggregates country-level attainments in health and education, as well as income. The index has been published since 1990 in the United Nations' Human Development Reports.

Old definition. Until 2009, the HDI was computed as a simple arithmetic mean of three indexes for health, education, and education:

- The health index was based on life expectancy at birth (LE_i)
- The education index was itself a composite index combining the adult literacy rate[1] (AL_i), with a two-third weight, and the gross enrolment rate in primary, secondary and tertiary education[2] (GE_i), with a one-third weight.
- The income index was based on the logarithm of gross domestic product per capita at purchasing power parity (GDP_i)

All four elementary indicators were normalized by taking the proportional country's achievement over a prefixed scale. The HDI for country i was then computed as follows:

$$HDI_i^{old} = \frac{1}{3}\Bigg(\underbrace{\frac{LE_i - 20}{85-20}}_{\text{health index}}\Bigg) + \frac{1}{3}\Bigg[\underbrace{\frac{2}{3}\left(\frac{AL_i}{100}\right) + \frac{1}{3}\left(\frac{GE_i}{100}\right)}_{\text{education index}}\Bigg] + \frac{1}{3}\Bigg(\underbrace{\frac{\ln GDP_i - \ln 100}{\ln 40,000 - \ln 100}}_{\text{income index}}\Bigg), \tag{1}$$

where $\ln(\cdot)$ denotes the logarithm function.

[1] The adult literacy rate is defined as the proportion of people aged 15 years and older who declare that they are able to read and write.

[2] The gross enrolment rate in primary, secondary and tertiary education is computed as the total enrolment in primary, secondary and tertiary education (without taking age into consideration) divided by the population of the official age for education in the given country (e.g. 6 to 22, 7 to 21). For the tertiary level, the population used is that of the five-year age group following on from the secondary school leaving age.

New definition. In 2010, the formula defining the HDI was changed to:

$$HDI_i^{new} = \underbrace{\left(\frac{LE_i - 20}{85 - 20}\right)}_{\text{health index}}^{\frac{1}{3}} \left[\underbrace{\frac{1}{2}\left(\frac{MYS_i}{15}\right) + \frac{1}{2}\left(\frac{EYS_i}{18}\right)}_{\text{education index}}\right]^{\frac{1}{3}} \underbrace{\left(\frac{\ln GNI_i - \ln 100}{\ln 75,000 - \ln 100}\right)}_{\text{income index}}^{\frac{1}{3}}, \tag{2}$$

where MYS_i denotes the average years of schooling in country i (for individuals aged 25 or older), EYS_i denotes the expected years of schooling[3], and GNI_i denotes the gross national income per capita at purchasing power parity.

Dataset. The `hdi_2019.dta` file is a dataset that provides all the necessary information to compute the old and new version of the HDI index for a sample of 126 countries in 2019. The dataset dictionary is shown in the appendix.

1. Compare and contrast the old vs. new definition of the HDI index by focusing on the following dimensions:
   - **(a)** The choice of the functional form that is used to compute the index. [**0.5 point**]
   - **(b)** The choice of the education indicators. [**0.5 point**]
   - **(c)** The choice of the income indicator. [**0.5 point**]
2. Compute the old and new version of the HDI index for all countries in the dataset and produce a scatterplot showing the value of the new HDI (on the Y-axis) and the value of the old HDI (on the X-axis) for each country. Comment. [**1.5 points**]
3. Compared to the old version of the HDI, the new version has been claimed to penalize countries with uneven development across the three components of the index (health, education and income). Describe and implement an empirical test to assess the validity of this claim using the dataset provided. [**2 points**]
4. Suppose that policymakers are provided with the two following options: (1) Increase life expectancy by 3 years in the 50% poorest countries (in terms of GNI) in the dataset; (2) Increase life expectancy by 3 years in the 50 % richest countries.
   - **(a)** Using the data, assess empirically which policy option would be favored if the policymakers are seeking to maximize the average value of the HDI (across all countries) under the old version of the index? [**1.5 points**]
   - **(b)** Perform the same comparison using the new version of the HDI index. Which policy option would now be favored? [**1 point**]
   - **(c)** Using the equation of the old HDI, show the effect of an additional year of life expectancy on the HDI index of country i is equivalent to increasing its GDP by $\Delta \ln GDP_i^* = \frac{\ln 40,000 - \ln 100}{65}$. [**1 point**]
   - **(d)** Use a similar approach to compute the variation in GNI, denoted by $\Delta \ln GNI_i^*$, that would be equivalent to the effect of adding an extra year of life expectancy on the value of the HDI index of country i under the new definition. [**1 point**]
   - **(e)** Based on the results from questions 4a and 4b, what criticism could be addressed to the new version of the HDI index from the perspective of promoting human development? [**0.5 point**]

[3]Expected years of schooling is the number of years of schooling that a child of school entrance age can expect to receive if prevailing patterns of age-specific enrolment rates persist throughout the child's life.

Appendix

Table 1: Variables of the Dataset **panel97.dta**.

Variable	Definition
pupil_id	Pupil's identification number (identification variable)
birthm	Pupil's month of birth (1 to 12)
parent_occ	Occupation of the pupil's responsible parent: 1 (Farmer), 2 (Self-employed), 3 (Executive and professional), 4 (Technician and associate professional), 5 (Clerical, sales and service workers), 6 (Blue-collar workers), 8 (Unemployed or economically inactive), 9 (Unknown)
year_t1	Year in which the pupil took test 1 (beginning of cp)
gscore_t1	Raw global score on test 1 = percentage of items answered correctly on test 1 (beginning of cp)
year_t2	Year in which the pupil took test 2 (beginning of ce2)
fscore_t2	Raw score in literacy on test 2 = percentage of items answered correctly on literacy component of test 2 in ce2 ("f" stands for "French language")
mscore_t2	Raw score in math on test 2 = percentage of items answered correctly on math component of test 2 in ce2 ("m" stands for "Math")
gscore_t2	$\frac{\text{fscore\_t2}+\text{mscore\_t2}}{2}$: "Raw global score for test 2" ("g" stands for "global")
year_t3	Year in which the pupil took test 3 (beginning of middle-school)
fscore_t3	Percentage of items answered correctly on literacy component of test 3 (beginning of middle-school)
mscore_t3	Percentage of items answered correctly on math component of test 3 (beginning of middle-school)
gscore_t3	$\frac{\text{fscore\_t3}+\text{mscore\_t3}}{2}$: "Raw global score on test 3"
p_gscore_t1, **p_fscore_t1**,..., **p_mscore_t3**	Percentile rank of the pupil on the different test scores (ranges from 1 to 100)
d_gscore_t1, **d_fscore_t1**,..., **d_mscore_t3**	Decile to which the pupil belongs for each test score (ranges from 1 to 10)

Table 2: HDI Dataset Dictionary

Variable	Label
country	Country name
le	Life expectancy at birth in 2019 (in years)
al	Adult literacy rate (averaged over 2008-2018)
ge	Gross enrolment ratio in primary, secondary, and tertiary education (averaged over 2014–2019)
mys	Mean years of schooling in 2019 (in years)
eys	Expected years of schooling in 2019 (in years)
gdp	Gross domestic product per capita in 2019 (in 2017 PPP $)
gni	Gross national income (GNI) per capita in 2019 (in 2017 PPP $)