Measuring socio-economic status in LMIC
How to construct wealth index

Fuyu Guo

September 13, 2020

Table of Contents

Table of Contents

What is common to health research and policy intervention is that there is differential impact with respect to health health outcome or health service utilization based on socio-economic status (SES) or socio-economic position (SEP), which is always a synonym to SES.

SES Wikipedia

SES is an economic and sociological combined total measure of a person's work experience and of an individual's or family's economic and social position in relation to others. The household income, earners' education, and occupation are examined, as well as combined income, whereas for an individual's SES only their own attributes are assessed. However, SES is more commonly used to depict an economic difference in society as a whole.

Table of Contents

Income is the sum of all the wages, salaries, profits, interest payments, rents, and other forms of earnings received in a given period of time (wikipedia). Income could be used to measure the material aspect of SEP.

HIC

- Formal income measuring system (e.g. through tax).
- Limited sources of income, mostly through monetary system.
- Less likely to vary across seasons.
- Different interpretation in communist society or in countries with rapid development (China).

LMIC

- Similar to HIC in MICs like countries in South America.
- Usually no official data in LICs.
- Multiple sources of income. Income could be paid in the form of goods.
- Self-employed, casual and seasonal labour problem.
- Bidirectional relationship between income and health.

Table of Contents

- Consumption expenditure measures how income is used by a household what goods and services are purchased.
- According to "permanent income hypothesis", consumption expenditure is more stable than the income.
- Compared to the income, it is a long-term measurement of SEP, and thus is preferred.

HIC

- Macro data: Caluclated in the GDP.
- Mico data: Household surveys and expenditure diaries.
- Not a measurement of SEP in studies examing inequlities in health outcomes.

LMIC

- Difficult to use expenditure diaries.
- Long list of potential expenditure items are given to respondent, very time-consuming.
- Including home-produced products.
- Misreport or recall bias.
- Seasonality for rural households.
- What sorts of expenditure should be included. Debates on health consumption.

Table of Contents

- Because of all the difficulties to use income or consumption expenditure in LMICs. An asset-based wealth-index was firsly introduced by DHS, and then by the World Bank
- The main idea is that instead of measuring economic indicators, ownership of a range of durable assets and access to basic services (e.g. electricty supply, sanitation supply) are used to assess a household or a individual SEP
- The arising problem is how to determine one's SEP through a amount of different variables

- The main aim of PCA is to reduce the number of dimensions
- suppose we have n correlated variables from X_1 to X_n, through PCA we will get

$$PC_1 = a_{11}X_1 + A_{12}X_2 + \cdots + a_{1n}X_n$$

$$PC_m = a_{m1}X_1 + A_{m2}X_2 + \cdots + a_{mn}X_n$$

where PC includes m uncorrelated componets in which PC_1 explains the largest possible amount of variation in the orginial data. a_{mn} represents the weight for the m th principal component and the n th variable.

- The higher the degree of correlation among the original data, the fewer components are required to capture the common information
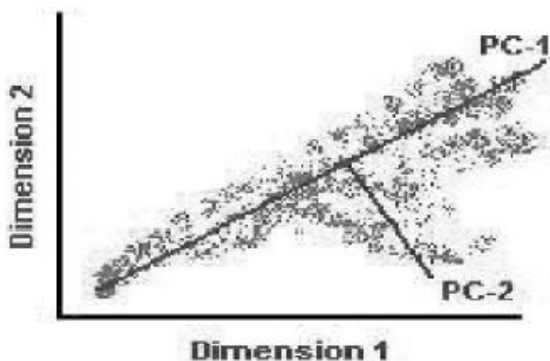
Figure 1. Representation of two sequential components in PCA. *Source*: [http://www.ucl.ac.uk/oncology/MicroCore/HTML_resource/PCA_1.htm], accessed 7 September 2005

Selecting variables into PCA is the major concern.

- Ownership of real estate (e.g.farmland, house).
- Ownership of durable assets (e.g. car, refrigerator, television).
- Infrastructure (e.g. sanitation facility and source of water).
- Housing characteristics (e.g. number of rooms and building material).

Generally speaking,

- PCA works better on correlated variables.
- The distribution of variables are expected to vary across households.
- Variables with low standard deviation carry low weights from the PCA.
- Unequally distributed variables are given large weights.
- Taking outcome variables into the PCA always leads to larger inequality among households (e.g. for the water saniation study).

In PCA, we aim to avoid **Clumping** and **Truncation**

- Clumping or clustering is described as households being grouped together in a small number of distinct clusters.
- Truncation implies a more even distribution of SES, but spread over a narrow range, making differentiating between socio-economic groups difficult.

Solutions to those two potential problems

- Including more variables.
- Using continuous variables instead of binary variables.

Example from DHS in Brazil and Ethiopia

Table 1. Results from principal components analysis

Variable description Brazil/Ethiopia	Brazil urban			Brazil rural			Ethiopia urban			Ethiopia rural		
	Mean	Std. dev.	Factor score	Mean	Std. dev.	Factor score	Mean	Std. dev.	Factor score	Mean	Std. dev.	Factor score
Electricity	0.987	0.114	0.158	0.694	0.461	0.347	0.829	0.376	0.297	0.012	0.107	0.171
Radio	0.881	0.323	0.216	0.765	0.423	0.171	0.689	0.463	0.294	0.139	0.345	0.210
Television	0.716	0.450	0.372	0.314	0.463	0.345	0.215	0.411	0.327	0.000	0.010	0.024
Refrigerator	0.821	0.383	0.363	0.425	0.493	0.397						
Car	0.296	0.456	0.295	0.135	0.341	0.256	0.035	0.184	0.176			
Bicycle										0.003	0.058	0.106
Telephone							0.136	0.343	0.291			
No. of rooms for sleeping	2.150	0.899	0.143	2.175	0.916	0.105						
Source of water supply												
Piped into residence/dwelling	0.760	0.427	0.243	0.200	0.400	0.179	0.007	0.086	0.033			
Piped into yard, plot/compound	0.044	0.204	−0.182	0.051	0.219	−0.033	0.414	0.493	0.367	0.001	0.024	0.105
/Piped outside compound							0.441	0.497	−0.221	0.061	0.239	0.092
Well, spring inside/covered well	0.075	0.264	−0.126	0.381	0.485	0.096	0.012	0.108	−0.077	0.065	0.247	0.103
Well or spring outside/open well	0.054	0.227	0.122	0.276	0.447	−0.154	0.049	0.217	−0.060	0.072	0.259	−0.106
Bottled water/	0.047	0.212	0.062									
/Covered, open spring							0.023	0.151	−0.103	0.427	0.495	0.071
/River							0.048	0.214	−0.152	0.333	0.471	−0.108
Other/and pond, lake, dam, rain	0.019	0.138	−0.135	0.092	0.289	−0.143	0.005	0.070	−0.011	0.041	0.199	−0.033
Sanitation facility												
Toilet to sewer/flush toilet	0.410	0.491	0.277	0.059	0.235	0.109	0.035	0.184	0.147			
Toilet to open space or river/	0.054	0.225	−0.089	0.093	0.290	0.057						
Latrine to sewer/	0.128	0.334	0.062	0.035	0.182	0.098						
Latrine no connection/	0.217	0.411	−0.049	0.176	0.380	0.210						
Traditional latrine/pit	0.138	0.345	−0.184	0.218	0.412	0.133	0.714	0.452	0.218	0.096	0.294	0.580
/Ventilated improved pit latrine							0.033	0.177	0.056	0.001	0.026	0.077
No facility/and bush or field	0.053	0.224	−0.238	0.420	0.493	−0.395	0.218	0.413	−0.328	0.904	0.295	−0.586
Type of floor material												
Earth or sand	0.032	0.175	−0.175	0.191	0.393	−0.294	0.345	0.475	−0.312	0.696	0.460	−0.263
/Dung							0.108	0.310	−0.121	0.283	0.451	0.228
Wood planks/and reed or bamboo	0.070	0.255	0.004	0.059	0.236	0.096	0.028	0.166	0.037	0.002	0.048	0.065

- Clumping or truncation is a problem in rural Ethiopia and urban Brazil.
- Factor score is the weight for the orginal variable.

technical points

- Categorical varaibles are coverted into binary variables.
- Similar variables with low frequencies are combined.
- Excluding or imputating missing values, for low SEP households are more likely to respond with missing values.
- Standardizing or scaling the orginal variables to remove the units.

- After summing all the variables multipling by the factor score, we will get the first *PC* which is the *"wealth-index"*.
- In this study the first *PC* only accounted for 12% to 27% variation.
- Higher order of *PC* was found not associated with the SEP.

Since, the absolute value of the wealth-index is hard to interprete, researchers usually use different cut-offs to differentiate households into catergories based on their wealth-index.

Poor, middle and rich

- 1%-40% Poor
- 41%-80% Middle
- 81%-100% Rich

5 quantiles

- 1%-20% Poorest
- 21%-40% Second
- 41%-60% Middle
- 61%-80% Fourth
- 81%-100% Richest

Table 2. Mean socio-economic score by quintile

Site	N	Poorest	Second	Middle	Fourth	Richest
Urban Brazil	10 527	−2.96	−0.82	0.35	1.33	2.14
Rural Brazil	2756	−2.68	−1.44	−0.01	1.40	2.80
Urban Ethiopia	3629	−2.82	−1.17	0.02	1.22	2.83
Rural Ethiopia	10 443	−1.08	−0.72	−0.43	0.20	2.85

For the wealth-index distribution, the wealth-index for Urban Brazil is skewed to left. For the Rural Ethiopia, the wealth-index is skewed to the right.
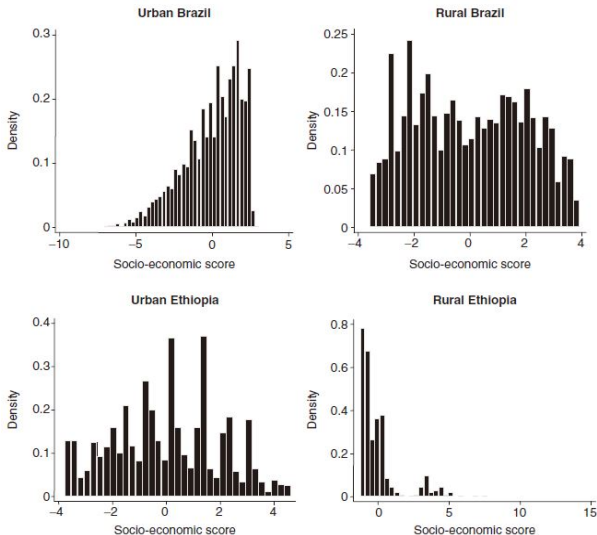

Figure 2. Distribution of socio-economic scores

Using the cluster analysis on the wealth-index to group the households into 3 groups The ideal shares of household in poor, middle and rich groups are 40%, 40%, 20%

Table 4. Proportion of households in low, medium and high socio-economic group for entire sample

Site	N	Low (%)	Medium (%)	High (%)
Urban Brazil	10 527	17.77	36.28	45.95
Rural Brazil	2 756	35.92	29.75	34.33
Urban Ethiopia	3629	38.58	40.20	21.22
Rural Ethiopia	10 443	59.26	30.73	10.01

limitations

- Only using binary variables will violate the assumption behand the PCA.
- The finer catergory of assets are usually ignored (e.g. color television VS black and white television)
- A measure of relative rather than absolute SEP, lacking comparability between countries or time.
- Ignoring Different types of assets between rural and urban (e.g. more access to electricty supply in the urban; more farmland in the rural).
- Usually a measuremnt on the household or family level.

- Wealth-index is a method trying to differentiate households in different SEP. Alternative methods includes cluster analysis and factor analysis.
- Wealth-index was used to measure early-life SEP when other indicators were not available. The assets used in the PCA are very different between HIC and LMIC.

Table of Contents

- Howe, L. D., Galobardes, B., Matijasevich, A., Gordon, D., Johnston, D., Onwujekwe, O., ... & Hargreaves, J. R. (2012). Measuring socio-economic position for epidemiological studies in low-and middle-income countries: a methods of measurement in epidemiology paper. International journal of epidemiology, 41(3), 871-886.
- Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: how to use principal components analysis. Health policy and planning, 21(6), 459-468.