Statistical Methodology

Preamble

The Arab Knowledge Index (AKI) consists of six composite thematic indices, expressing six key developmental sectors, namely: Pre-University Education, Technical Vocational Education and Training (TVET), Higher Education, the Economy, Information and Communications Technology (ICT) and Research & Development (R&D) and Innovation. Each of these six indices was constructed in accordance with the standard international methodologies for the construction of composite indicators.[1] The following is a presentation of the steps used in constructing these indicators.

Variable Selection

Selecting the individual variables included in the construction of each of the six knowledge indicators is relying on a well-defined, clear and scientific methodology based on a thorough review of relevant international and local literature. In addition, the concepts and experience of international organisations and agencies are explored. In each of the six sectors, it also relied on consultations with a large number of specialists in various countries of the world such as Egypt, Jordan, United Arab Emirates, Canada, United Kingdom and United States through a special questionnaire. The participants were able to express their agreement with the selected list of variables and their various aggregations; reject or propose additions or amendments to it. In accordance with the feedback from the participants and the rest of the core team who prepared the Index, a final list of the variables was compiled.

Principal Components Analysis was used to confirm the consistency of the selected variables and the structure of their classification into the various sub-indicators. These results supported the consistency of the conceptual context in selecting the variables and their classification in the various sub groups, in which the explained variance ratio in most cases exceeded 50 per cent.[2]

The results of the in-depth correlation analysis and alpha Cronbach coefficient confirmed the validity of the selection and classification of the variables, in which the alpha Cronbach coefficient exceeded 0.70 in more than 80 per cent of cases. Weak consistency of certain variables could be attributed to lack of data and/or the nature of the correlation between these variables.[3]

Data Used

The 304 variables incorporated into the construction of the six Arab Knowledge indices can be classified into three types. The first consists of the hard data obtained from various resources such as the World Bank, the UNESCO, other Uniterd Nations agencies and others.[4] The second type includes composite indicators computed by some of the international institutions such as the International Telecommunication Union (ITU), the European Union (EU) and the Organisation for Economic Cooperation and Development (OECD). The third type consists of survey data which are used when variables have no data or data are incomplete.

For the sake of transparency, simplicity and the possibility of replicating the results, no attempts were made to impute missing values to the various variables. The use of the arithmetical mean formula in computing an indicator is equivalent of imputing each of the missing values of the indicator to the mean value of the variable. The missing values, indicated by the symbol "n/a", were not entered in the composite sub-indicators which were computed, as is customary in similar cases, by just using the available data for each country.[5]

Data-processing was performed on the assumption that the data were error-free, the team having reviewed it more than once to ensure there were no errors in the data entry. The variables that might lead to a biased indicator value were processed using suitable statistical methods.[6] It had been observed that some indicators were linked to other, size-dependent indicators, such as population or GDP; these indicators were therefore rescaled using the size.

Normalisation

The values of variables were normalised in the range of 1-100, in which the higher values indicated better results. The rescaling or the 'maximum–minimum' method was used, in which the maximum and minimum indicate the largest and smallest value of the available indicator values respectively. The normalisation criterion depends on whether the variable is good i.e. has a positive relation with the overall index, or bad i.e. has a negative relation with the overall index. The good indicators were normalised using the following formula:

$$\text{Normalised indicator value of the country} = 99 \times \left(\frac{\text{raw indicator value of the country - raw minimum value of the indicator across countries}}{\text{raw maximum value of the indicator across countries - raw minimum value of the indicator across countries}} \right) + 1$$

In the case of the bad indicators i.e. indicators with an inversely correlated relation, this formula should be adjusted as follows:

$$\text{Normalised indicator value of the country} = 99 \times \left(\frac{\text{raw maximum value of the indicator across countries - raw indicator value of the country}}{\text{raw maximum value of the indicator across countries - raw minimum value of the indicator across countries}} \right) + 1$$

Weighting

The methods of estimating weights used in the construction of the six knowledge indices range from equal weighting, budget allocation process and the factor analysis method. Equal weights are used in the absence of clear evidence of the diversity of significance of each Index, as well as in the absence of sound and complete information concerning the existence of causal relationships or a lack of consensus on a classical method for estimating weights.

The budget allocation process method was also used for weighting, in which a group of specialists, concerned and experienced experts were invited to attend a workshop for each of the six knowledge sectors. Each expert was given a budget consisting of 100 points for the variables or sub-indicators used. If the variable or sub-indicator was believed to have greater relative importance, it was allocated a greater number of points. Subsequently, the weights were estimated by calculating the average of the total points allocated to each variable or sub-indicator.[7]

The weights were also assessed using factor analysis, which is based on aggregating the linked sub-indicators to form a single factor containing as much information as possible that is shared between these linked indicators. The resulting estimated weights using both the budget allocation and factor analysis methods were extremely consistent with each other, as well as being consistent with the initial weights estimates, based on the intellectual and conceptual framework.

Index Computation

The Arab Knowledge Index was calculated for 22 Arab countries in its first version using the most recent and best available data for the various variables of each country, together with the choice of the year 2005 as a cut-off year. The values of the composite sub-indices for the Knowledge Index were calculated by applying a series of successive aggregations starting with the more detailed level of variables and ending by attaining the overall index. In the case of ICT for example, the sub-indicators relating to infrastructure and digital content were aggregated to form a single composite indicator expressing the infrastructure and digital content. Similarly, the sub-indicators relating to each of communications affordability, individual, company and government usage were aggregated to form a sub-composite indicator for each. By aggregating these three

sub-indicators, a single composite indicator was created expressing the direct ICT indicators. Using the same method, one composite indicator was created expressing the indirect indicators to ICT.

The overall ICT Index was created by aggregating the two composite indicators for the directed and indirect indicators.

Due to an inability to obtain data for all the sub-indicators of each country, some countries have missing values for certain sub-indicators. For example, in the case of infrastructure and digital content, no data was available for the Internet bandwidth (kb/s per user) indicator for Libya and Yemen. Accordingly, and with a desire to calculate the indicators for all the countries, a maximum number of sub-indicators for each country was calculated using the available data. For example, the infrastructure and digital content indicator for Libya and Yemen alone was calculated through the use of four sub-indicators – electricity production in terms of kWh per capita, the extent of mobile phone network coverage as a percentage of population, ratio of households with a computer and access to digital content. There was no sub-indicator concerning Internet bandwidth or kb/s per user due to the unavailability of data.

On the other hand, when the data for sub-indicators were available for a single country only or for two countries at the most, the decision was taken, however, not to calculate this indicator, because the data were insufficient for normalisation. Subsequently, this indicator was excluded from the calculation of the rest of the overall composite indicators, and the results of the calculation were not presented subsequently.

There are two well-known aggregation methods: arithmetical (or linear) and geometric aggregations and any of them can be chosen by the researcher. Sometimes, both methods are used together for comparing their outcomes to assess the sensitivity of the Index to the aggregation method.

The linear aggregation formula of the sub-indicators (SIj) to compute the composite indicator (CI) takes the following form:

$$CI = \sum_{j=1}^{n} w_j \times SI_j$$

The geometric aggregation formula of sub-indicators (SIj) to compute the composite indicator (CI) takes the following form:

$$CI = \left(\prod_{j=1}^{n} SI_j^{w_j} \right)^{1/\sum_{j=1}^{n} w_j} = exp\left(\frac{\sum_{j=1}^{n} w_j \times ln(SI_j)}{\sum_{j=1}^{n} w_j} \right)$$

CI is the proposed composite index to be computed, w_j is the relative weight of the sub-indicator SI_j, n is the number of sub-indicators aggregated to form the composite indicator and exp and ln are the exponential and logarithmic functions respectively. The arithmetical (or linear) aggregation method is employed to compute all knowledge sub-indicators in this Index.

Index Sensitivity

Like all composite indices, the construction of the Arab Knowledge Index and its six sub-indices depends on the choices of the researchers, reflecting the unavoidable elements of uncertainty. These elements include the methods for variables selection, the structure of the sub-indices, imputation of the missing values, normalisation, weighting and aggregations. The Index sensitivity study aims to assess the effect of uncertainty elements separately or jointly on the performance of the Index.

The results of the sensitivity analysis of the knowledge indices for weighting and normalisation methods showed the insensitivity of the indicators to these elements i.e. the indicator performance does not differ significantly due to variation of these elements. For example, to study the sensitivity of the knowledge indices to the aggregation method, each index was calculated by using both arithmetic and geometric means. The geometric mean was used because it offers a partial compensability between indicators,

contrary to the arithmetic mean that assumes full compensability between the indicators incorporated into the Index calculation, where both methods – arithmetic and geometric aggregations – were applied to the sub-indicator of infrastructure and digital content. Despite the geometric aggregation formula giving lower values for the indicator than the results of the arithmetic formula, the performances of the sub-indicator of infrastructure, digital content and the overall ICT indicator do not differ significantly. Thus, the ranking of the countries is still mostly unchanged, meaning that the highest ranked countries, according to the results of arithmetic aggregations, maintain the same rank according to the results of geometric aggregation.

It is noteworthy that the construction of the Arab Knowledge indices – despite their validity in reflecting the true knowledge situation in the Arab region – is open to future development, including the completion and improvement of the data quality and global sensitivity of the Index.[8]

Endnotes

1 OECD 2008b.
2 For more information about Principal Components Analysis refer to Hair et al. (2010).
3 For more information about Cronbach's coefficient alpha refer to Tavakol & Dennick 2011.
4 For more information about the data sources of the Arab Knowledge indices, refer to the annex.
5 Cornell University et al. 2015.
6 Groeneveld and Meeden 1984.
7 For more information about the Budget Allocation Process method, refer to OECD 2008b.
8 Refer to Saltelli et al. 2008.