The Impact of Socio-economic Factors in the Global Decline of Mortality levels

Dr. Amin I. Adam
Associate Professor in Applied Statistics, Dean Faculty of Economics & Political Sciences, Omdurman Islamic University, Omdurman,
P. O. Box 382, Sudan

Abstract:

In this paper we studied the effect of the economic and social factors in the global decline of mortality levels. These factors include: the real per capita income, the daily calorie supply, the adult literacy rate for males, the adult literacy rate for females, the number of population per physician, the percentage of population living in urban areas, the percentage of population with access to health services, and the percentage of population with access to safe water; are the available variables which we used for their possible influence in mortality levels. On the other hand, and among many mortality measures we used Infant Mortality Rate because it is more likely be affected by these socio-economic variables than other mortality measures. Based on our data, which were collected form the periodicals of the United Nations, the empirical findings showed that all the investigated variables are highly correlated with IMR, with correlation-coefficients greater than and equal 0.7 for all. The number of population per doctor is the sole variable that turned out to be positively related with IMR. For the multiple regression of mortality on these variables we found that: the log of the real per capita income, the adult literacy rate for females, and the percentage of population having access to safe water are the only variables which turned out to be statistically significant. The diagnosis we made did not confirm the absence of 'mutlicollineariy' problem. Accordingly, a remedy for 'multicollinearity' was suggested using the Principal

Component Analysis to extract independent factors to represent the data. The first extracted factor is shown to be interpreted as a contrast between the log of the number of population per doctor and the rest of the variables. It explained about 87% of the variations in mortality. The study showed a significant difference between the developed and the developing countries concerning mortality levels and concerning the extracted factors using the PCA. The factor analysis technique showed that all the socio-economic variables turned out to be represented by one and only one factor. This factor is identified, without any doubt, to be the infant mortality. This means that mortality reductions are more likely to happen with the economic development and consequently with improving health and medical technology, along with the better diet and greater educational attainment and awareness of health protection as an auxiliary factor.

Key Words: Mortality levels, Infant Mortality Rate, Socio-economic factors.

الملخص:

تبحث هذه الورقة في أثر العوامل الاجتماعية والاقتصادية في الانخفاض العالمي في معدلات الوفيات. وتم استخدام متغيرات اقتصادية واجتماعية شملت: متوسط دخل الفرد الحقيقي ومعدل السعرات اليومية ومعدل التعليم بالنسبة للرجال ومعدل التعليم بالنسبة للنساء وعدد الأشخاص لكل طبيب عام ونسبة السكان الذين يقطنون المناطق الحضرية ونسبة السكان الذين يتمتعون بخدمات صحية ونسبة السكان الذين يتمتعون بماء صحي؛ وهي متغيرات لها بيانات متاحة ويُعتقد أن لها تأثيراً ما في معدلات الوفيات. وفي المقابل ومن بين المعدلات المختلفة للوفيات تم استخدام معدل وفيات الأطفال حديثي الولادة لارتباطه أكثر من غيره بهذه العوامل الاقتصادية والاجتماعية. وبناءً على المعلومات التي الحصول عليها من دوريات الأمم المتحدة فإن نتائج الدراسة أوضحت أن كل المتغيرات قيد الدراسة لها ارتباط عالٍ مع معدل وفيات الأطفال حديثي الولادة، بدرجة ارتباط أعلى من 0.7. وباستثناء متغير عدد الأشخاص لكل طبيب عمومي الذي له علاقة موجبة فإن كل المتغيرات الأخرى لها علاقة سالبة مع معدل وفيات الأطفال حديثي الولادة،

بمعنى أن التحسين والزيادة في مؤشرات هذه المتغيرات له علاقة قوية في انخفاض معدل وفيات الأطفال حديثي الولادة. وبتطبيق نموذج الانحدار المتعدد للوفيات وجد أن لوغريثم متوسط الدخل الحقيقي ومعدل التعلم بالنسبة للنساء ونسبة السكان الذين يتمتعون بماء صحي هي المتغيرات التي لها دلالة إحصائية في وفيات الأطفال حديثي الولادة. وعند الفحص لمشكلة الارتباط الخطي بين المتغيرات تبين أن هذه المشكلة بالفعل واقعة بين هذه المتغيرات. وبناءً عليه تم استخدام طريقة تحليل المكونات الأساسية لاشتقاق عوامل جديدة مستقلة تمثل المتغيرات الاقتصادية والاجتماعية قيد الدراسة. وقد وُجِد أن العامل المشتق الأول بطريقة المكونات الأساسية يمكن تفسيره على أنه مقارنة بين (لوغريثم) عدد الأشخاص لكل طبيب عام وبين بقية المتغيرات في الدراسة . وقد فسر هذا العامل حوالي ٧٨% من جملة الفروقات في مصير وفيات الأطفال .كما أبانت الدراسة وجود فروق جوهرية بين الدول المتقدمة والدول والنامية من حيث معدلات وفيات الأطفال حديثي الولادة ومن حيث العوامل المستخلصة بواسطة طريقة المكونات الأساسية .كما أوضحة طريقة تحليل العامل أن كل العوامل الاقتصادية والاجتماعية جاءت ممثلة بعامل واحد فقط ، وهذا العامل يمكن التعرف عليه بدون شك على أنه تمثل وفيات الأطفال . وهذا يعنى أن التخفيض في معدلات وفيات الأطفال يرتبط حدوثه بالتحسين في الاقتصاد (الدخل) وبالتالي بالتحسين في الظروف الصحية والطبية ، بجانب التغذية الجيدة والملكات التعليمية الواسعة والوعي بالوقاية الصحية كعوامل مساعدة .

1- Introduction and Objectives

The mortality rates, all over the World, are noticed to be decreasing through the past decades. These reductions were varied from one continent to another and from one country to another. Accordingly, there are many socio-economic variables which could be suggested to be responsible from these mortality reductions which were experienced by the World countries. These variables represent: rising living standards, modernization, urbanization, education, modern medical technology, family planning programmes, etc. The real per capita income, the daily calorie supply, the adult literacy rate for males, the adult literacy rate for females, the number of population per physician, the percentage of population living in urban areas, the percentage of population with access to health services, and the percentage of population with access to safe water; are the variables which are used to explain mortality reductions. Among so many measures that could be used to represent mortality behaviours, we chosed infant mortality rate (IMR). Infant mortality rate (the number of deaths under one year age in a year per 1000 live births) is chosen because its contribution to the total loss of years of human life is great, i.e., its level is relatively high. In addition, infant mortality seems to be more affected by the socio-economic variables than other mortality measures. Moreover, infant mortality is generally agreed to provide an index of the prevalent level of living in developing countries.

Our objective, then, is to assess the effectiveness of these variables in mortality levels, and to see the relative strength of each variable in terms of its impact on mortality reductions.

2- Data of the Study

The previous discussion indicates that a proper understanding of the effect of socio-economic variables in mortality decline requires a variety of detailed data on these variables. But the main problems concerning data are their availability and reliability. The data we managed to collect are the recent ones and collected from UNDP(2010), UNFPA(2010), UNICEF(2010) and The World Bank(2010). They are cross-sectional data (data collected on variables at one point in time) for 129 countries representing almost all the world countries, with the exception of some other very small countries and islands. The number of population of these excluded countries and islands is very small and for most of it, if not all, our variables under investigation are not available.

3- Methods and Techniques of Analysis

Putting our objective in mind, i.e., the effectiveness of socio-economic variables in mortality decline, and besides the use of the correlation and regression techniques we are going to use the 'Principle Components analysis' and 'factor analysis'.

4- Infant Mortality Rates for World Countries

Before assessing what socio-economic variables affecting mortality decline let us make a comparison between continents and groups of countries concerning levels of infant mortality.

According to table(1) and for the world as a whole, infant mortality rate was estimated at about 118 per 1000 population for 1960 compared with only 62 in 2010. But this decrease in infant mortality rate was by no means uniform for the major continent-groups and even, sometimes, inside each continent-group. In 1960, infant mortality rates were 167, 130, 106, 70, 43 and 27 for Africa, Asia, Latin America, Oceania, Europe and Northern America; respectively; compared with 104, 60, 47, 25, 12 and 9 for the same group countries in 2010. It is very clear now the absolute reduction in mortality for Japan, Europe, Northern America, USSR and Oceania (with the exception of Papua New Guinea); which are classified as developed countries by UN(2007a), UN(2009b), and UNICEF(2010); was much less pronounced than for the rest of the countries, i.e., the developing countries, and this is mainly due to the existence of greater room for reduction in comparison with the developed countries. Within each of these continent-groups there were also considerable variations in the levels of infant mortality for subgroup countries. In Africa, for example, infant mortality was very high in the Western part while in the Northern and Southern parts the situation seems to be better. In Asia, a very low infant mortality was found in Japan (about 4 in 2010) while in Southern Asia the rate was 102. Latin American countries all had mortality intermediate between the lowest and highest levels. For the USSR, infant mortality rate was slightly lower, but the lowest rates were found in Japan, Europe, Northern America, Australia and New Zealand.

Table(1): World Infant Mortality 'IMR' (per 1000 population) for 1960 and 2010

Continent	Region	IMR 1960	IMR 2010	Reduction	Reduction %
Africa		167	104	63	37.72
	Eastern Africa	153	105	48	31.37
	Middle Africa	174	111	63	36.21
	Northern Africa	167	75	92	55.09
	Southern Africa	140	83	57	40.71
	Western Africa	187	118	69	36.90
Asia		130	60	70	53.85
	Eastern Asia	98	31	67	68.37
	Southeastern Asia	116	60	56	48.28
	Southern Asia	164	102	62	37.80
	Western Asia	139	50	89	64.03
	Japan	31	4	27	87.10
Latin America		106	47	59	55.66
	Caribbean	97	40	57	58.76
	Central America	114	46	68	59.65
	South America	104	50	54	51.92
Ocienia		70	25	43	64.29
	Australia	21	8	13	61.90
	New Zealand	23	10	13	56.52
	Papua New Guinea	165	58	107	64.85
Europe		43	12	31	72.09
	Eastern Europe	49	15	34	69.39
	Northern Europe	22	8	14	63.64
	Southern Europe	72	15	57	79.17
	Western Europe	28	8	20	71.43
Northern America		27	9	18	66.67
	Canada	28	7	21	75.00
	USA	26	10	16	61.54
USSR		38	25	13	34.21
Whole World		**118**	**62**	**56**	47.46

* Source: ref. '28', '29', '30' and '31'.

5- Socio-economic Variables Affecting Mortality Decline

It is very clear from what we have said just earlier that the high mortality rates characterized the economically underdeveloped regions of

Asia and Africa, and, as suggested by Bogue(1969), these high levels are likely to be related to the low per capita income, high rates of illiteracy, a large proportion of males engaged in agriculture, and a large number of persons per physician. Similarly, low levels of mortality, which prevailed in Europe, Northern America and other parts, were accompanied by high income levels, low illiteracy rates, an economy that is primarily non-agricultural, social reforms, and a more adequate supply of physicians. This suggests a relationship between mortality decline which has been experienced by those affluent countries and the development which created these socio-economic conditions. The main socio-economic variables which are thought to be associated with mortality reductions are described individually below:

5.1- Economic development and rising income levels

It is found by Glass and Revelle(1976), Pichat (1973), Preston(1978), UN(2009) and others, that the reduction in mortality rate in Europe for the last two centuries has been due largely to the rising income levels and economic development. It is also recognized that this economic progress played the same role in reducing mortality rate in other parts of the world.

So income level is inversely related to mortality rate, since better income levels mean improved supply of food and diet, promotion of health services and medical research, purification of water supply, etc; which all help directly or indirectly in reducing mortality. The appropriate variable that we are going to use for this effect is the real per capita income(RPCI).

5. 2- Sanitary reforms and public health measures

Improvement in sanitation was considered by Glass and Revelle(1976) and UN(2009) as a direct cause for the decline in mortality from typhus, cholera, plague, dysentery, diarrhoea and enteritis. Purification of the water supply was one of great improvements in sanitation and public health all over the world. The percentage of population having access to safe water (PPSW) is the available variable that will be used for its own right for this effect. Personal hygiene also developed along with the improvement in sanitation and public health. The best examples of this development are: common use of soap and substitution of cotton for wool. These are believed by UN(2009) to have had a direct impact on the reduction of typhus, which is transmitted by the body louse. Part of what we said could be represented, also, by the percentage of population having access to health services (PPHS) which we will meet later.

5.3- Nutrition

It is suggested by Hauser(1980), Hutchinson(2003), Overbeek(1976), Preston(1978) and UN(2010), that a better diet was probably the main reason behind the substantial fall in tuberculosis mortality in many parts of the world during the second half of the last century, and was possibly a cause, although to a lesser extent, in the fall of mortality from typhus. The daily calorie supply as a percentage of the requirement (DCS), which we are going to use as a variable affecting mortality, is the most appropriate available measure for nutrition. The daily calorie supply, however, is obviously related to income level which we have mentioned earlier.

5.4- Educational attainment

Rising levels of education and health consciousness have played an important role in reducing mortality, especially infant mortality. As reported by UNICEF(2010), increased educational attainment of mothers and fathers makes them more aware of the importance of child care and health protection such as immunization. The role of mothers especially is very important, since with better education will be easier for them to follow publications, radio and television programmes about child rearing and health care which of course reflects on the health of their children especially during pregnancy. The adult literacy rates for both males (ALRM) and females (ALRF) are the available variables which will be used in this respect.

5.5- Advances in medicine, improved health services and availability of better health consultants

There is a belief to Gray and Tangri(1990), Habakkuk(1974) and UN(2010) that improved medical practices had a role - however modest - in reducing mortality. It is beyond the scope of this study to mention the tremendous achievements in medical therapy which have contributed so much to mortality reduction, and in some situations it is believed to abolish totally from the developed countries mortality from certain dread diseases. Promotion of health services in general and the availability of health consultants have played an important role in mortality decline. The percentage of population with access to health services (PPHS) and the number of population per doctor (NPPD), which we are going to use as factors affecting mortality, are the most appropriate available measures found representing health services and health consultants, respectively.

5.6- Urbanization

Urbanization is considered by UN(2010) as a determinant factor in mortality mainly because urban areas are in most cases accompanied by better medical facilities and public services, better education, higher income, more aware of health problems, etc. However, Pollard, Yusuf and Pollard (1990) pointed to the fact that rural mortality might generally be less than urban mortality, although they confessed that the difference is pretty small. The relative contaminated fresh air the urban areas compared with the rural areas and the direct effect of this on mortality could be a justification behind their views. The percentage of population living in urban areas (PPU) is the variable that will be seen as an indicator of the degree of urbanization.

5.7- Other variables

There are many other variables believed to have an effect on mortality decline, directly or indirectly. However, they weren't considered by UN(2010) as important as the previously mentioned ones. These include: the improvement in housing conditions, the increasing health consciousness of the people, the development of transportation facilities, etc. Unfortunately those variables won't be considered in this study because of the unavailability of the suitable data to represent them.

6- Mortality and Simple Correlations

Starting with the correlations of mortality and each of our socio-economic variables, Table (2) shows that all the prescribed correlations turned out to be very high. Most of the absolute values of correlation coefficients are - 0.7 and above. The Adult literacy rate for males (ALRM) and for females (ALRF), which we used to represent the educational attainment, have the largest correlation with mortality, with coefficients of -0.87 and -0.86, respectively. On the other hand, the number of people per doctor (NPPD) is the variable with the lowest correlation with mortality, with correlation of about 0.68. However, this relative lowest correlation is itself highly significant, statistically, as indicated by the t-value given by

$$t = 0.68(113-2)/(1-(0.68)^2) = 9.77$$

when compared with the critical t-value of 1.98 under 111 degrees of freedom and 5% level of significance. These strong correlations, in all, signify the strong association between mortality and the socio-economic variables.

Table (2): The Correlations Between Infant Mortality Rate 'IMR' and the Socio-economic Variables

Variable	IMR	Log IMR	Variable	IMR	Log IMR
ALRM	-0.81	-0.83	Log ALRM	-0.83	-0.75
ALRF	-0.86	-0.83	Log ALRF	-0.81	-0.74
PPSW	-0.81	-0.84	Log PPSW	-0.78	-0.76
PPHS	-0.81	-0.78	Log PPHS	-0.76	-0.70
DCS	-0.80	-0.76	Log DCS	-0.81	-0.75
PPU	-0.75	-0.76	Log PPU	-0.72	-0.69
RPCI	-0.73	-0.88	Log RPCI	-0.87	-0.91
NPPD	0.68	0.55	Log NPPD	0.87	0.83

Another important point from Table (2) is that the directions of the correlations between mortality and the socio-economic variables coincided with our prior expectations. They all suggest a close linear relationship between mortality and the socio-economic variables. Table (2) also tells us, in the third column, about the improvement in the correlation for some of the variables, especially the real per capita income (RPCI) and the percentage of population with access to safe water (PPSW), when log transformation was introduced for mortality. However, the last two columns of table (2), with log transformation introduced to the real per capita income (RPCI) and the number of people per doctor (NPPD), the correlations became far better. This will be needed in the next section for building the multiple relationship between mortality and all these variables together. It is also worth saying that neither the quadratic nor the reciprocal transformations of our socio-economic variables appeared to improve their relationship with mortality.

On the other hand, most if not all of our socio-economic variables turned out to be significantly correlated with each other. The adult literacy rate for males (ALRM) and for females (ALRF) are very strongly related according to their correlation coefficient of 0.96. These high correlations are good in situations where some variables can be considered as proxies of others and are very important, also, for forming simple and parsimonious factors in 'Factor Analysis', which we will meet later. However, for the multiple linear regression, which we will see shortly, these high correlations might create a multicollinearity problem.

7- Mortality and Multiple Regression

The analysis in the previous section suggests that it may be helpful to transform the infant mortality rate (IMR), the real per capita income (RPCI) and the number of people per doctor (NPPD) by taking logarithms. After some experimentation, the really for snggeafcd form of modest regressing mortality on all our socio-economic variables is

$$\text{Log IMR} = \underset{(8.97)}{4.42} + \underset{(1.79)}{0.0025}\,\text{PPU} - \underset{(-0.13)}{0.00017}\,\text{PPHS} - \underset{(-0.86)}{0.00132}\,\text{DCS}$$

$$- \underset{(-3.52)}{0.00737}\,\text{ALRF} - \underset{(-2.23)}{0.00271}\,\text{PPSW} - \underset{(-0.82)}{0.0534}\,\log \text{NPPD}$$

$$- \underset{(-6.60)}{0.644}\,\log\text{RPCI} + \underset{(1.06)}{0.00336}\,\text{ALRM}$$

$$R^2(\text{adjusted}) = 0.887\ ,\ \text{F-ratio} = 89.33$$

$$n = 91 \ , df = 82 \ , s = 0.1574$$

where inside the brackets are the t-values of the corresponding coefficients.

This estimated model is highly significant according to the value of F-ratio (89.33) which is greater than the tabulated value of F under 8 and 82 degrees of freedom and with any level of significance more than 0.0005. The adjusted R^2 (for the degrees of freedom) indicates that about 89% of the total variation in mortality is explained by the variation in all these variables together. However, the t-test values show that the real per capita income (RPCI), the adult literacy rate for females (ALRF) and the percentage of population having access to safe water (PPSW) are the only significant variables at the 5% level. Hence the insignificant appearance of most of our explanatory variables and the strange sign of the coefficients of the adult literacy rate for males (ALRM) and log of the number of people per doctor (NPPD), made this model seems distorted.

It is very clear now that we are facing the multicollinearity problem. We have already seen how strong are the correlations between our socio-economic variables themselves, so multicollinearity is not a surprise. However, it might not only be the multicollinearity that is to be blamed from this lack of significance of regression coefficient, but heteroscedasticity could be responsible as well, especially with this type of data, i.e., the cross-sectional data. To pursue the matter more, let us apply a formal test for heteroscedasticity. This test is the Breusch-Pagan test, due to Breusch and Pagan(1979), and is chosen among so many tests because it is general and simple at the same time. It considers the variance of the disturbance term to be a function of some known variables (which we consider in this case as all our explanatory variables). Specifically speaking, we assume that

$$\sigma_i^2 = \delta_0+\delta_1 PPU+\delta_2 PPHS+\delta_3 DCS+\delta_4 ALRF+\delta_5 PPSW+\delta_6 \log NPPD +\delta_7 \log RPCI+\delta_8 ALRM$$

that is, σ_i^2 is a linear function of all our socio-economic variables. If

$$\delta_1 = \delta_2 = ... = \delta_8 = 0$$

then $\sigma_i^2=\delta_0$, which is constant. Therefore, to test whether σ_i^2 is homoscedastic, we test the hypothesis that

$$H_o : \delta_1 = \delta_2 = ... = \delta_8 = 0$$

and the first step in performing this test is the use the residuals from the previous equation in order to get an estimate of the variance of the disturbance term, that is

$$\sigma^2 = e_i^2/N = 0.22326$$

The second step involves constructing a new variable 'P', which is simply each residual squared and divided by the by σ^2, that is

$$P = e_i^2/\sigma^2 = e_i^2/0.022326$$

In the last step we regress 'P', thus constructed, on our explanatory variables and obtain the explained sum of squares (ESS). The ESS appeared to be 14.461. Since ESS/2 (=7.2305) is less than the chi-squared value (15.507) with 8 degrees of freedom and 5% level of significance, we accept the hypothesis of homoscedasticity and hence there is no evidence to support the existence of heteroscedasticity. It is then the multicollinearity problem that we suffered.

Multicollinearity, however, does not destroy the unbiasedness and consistency properties of our OLS estimates, but these estimates are no longer minimum variance or efficient. To solve the problem, we are going to use the 'Principal Component' technique. This technique is used to build new variables (principal components) from the original ones (our explanatory variables), so that the new variables are independent of each other. According to Koutsoyiannis(1973), if these new variables can be given any specific meaning, then they can be used as variables in their own right and the transformation provides a defensible solution to the multicollinearity problem. Each of the principal components (PCs) is then formed as a linear combination as

$$PC = a_1PPU + a_2PPHS + a_3DCS + a_4ALRF + a_5PPSW + a_6\log NPPD + a_7\log RPCI + a_8ALRM$$

and the variance of any such linear combination is to be calculated. The first principal component (PC1) is that linear function which has the maximum possible variance, the second principal component (PC2) is the linear function with maximum possible variance subject to being uncorrelated with the PC1, the third principal component (PC3) is the linear function which maximises variance subject to being uncorrelated with the PC1 and PC2, and so on up to the number of the original variables. The aim of the PC technique, however, is to choose a few PCs that account for the maximum possible proportion of the original variation. The term 'eigenvalue' is simply the variance of the corresponding PC, and therefore a measure of its importance in explaining the variation. The result shows that the first PC explains about 80% of the total variation between the socio-economic variables. Almost 95% of the total variation is attributable to the first four PCs. There is no simple rule for determining the number of PCs to be extracted, but there are many 'rules-of-thumb' have been produced. one guideline suggests that only PCs that account for variances (eigenvalues) greater than 1.0 should be included, while another one considers an eigenvalue of 0.7 as satisfactory. A third criterion proposes to choose the PC where the cumulative variance exceeds 80% of the total variation. In this study, however, we are going to retain the first four PCs, although the first PC may be adequate to represent the data.

In regressing mortality on the four PCs extracted, the regression equation we came up with is

$$\text{Log IMR} = \underset{(21.14)}{2.95} + \underset{(14.54)}{0.00921}\ PC1 + \underset{(0.55)}{0.00066}\ PC2 - \underset{(-0.20)}{0.0003}\ PC3 + \underset{(0.17)}{0.00033}\ PC4$$

$$R^2(\text{adjusted}) = 0.818\ ,\ \text{F-ratio} = 102.31$$

$$n = 91\ ,\ df = 86\ ,\ s = 0.1996$$

where inside the brackets are t-values of the corresponding coefficients.

The model, however, is still not convincing. Apart from PC1, all the other three PCs are far from being important in the model. So it seems reasonable to rely on the first PC only, which could be considered as a general measure of underdevelopment, loaded on all the socio-economic variables, and having the appropriate sign for such a measure, or equivalently as a contrast between the log of the number of people per doctor and the rest of the explanatory variables. Hence, the regression of mortality on the first PC is

$$\text{Log IMR} = \underset{(41.6)}{2.98} + \underset{(20.53)}{0.00907}\ PC1$$

$$R^2 = 0.824\ ,\ \text{F-ratio} = 421.34$$

$$n = 91\ ,\ df = 89\ ,\ s = 0.1967$$

where inside the brackets are the t-values of the corresponding coefficients.

The model is extremely significant overall as indicated by the high value of the F-ratio (421.34) over the tabulated F under the 1 and 89 degrees of freedom with any level of significance more than 0.0005. The coefficient of determination, R^2, is very high indicating that the first PC explained about 82% of the mortality variation, leaving less than 18% of the variation to be explained by other PCs and other variables (including ones which are outside our investigation). The t-values in the brackets confirmed that the coefficient of the PC1 and the constant term are all individually significant.

The coefficient of PC1 is positive and hence, in an indirect way, our socio-economic variables are inversely related to mortality, since PC1 is negatively related to these variables.

As far as the outliers are concerned, there are five cases (countries) considered having an unusual observations. For Tanzania, Gabon and Saudi Arabia the IMR of the 39, 34 and 24, respectively, are appeared to be predicted instead of 103, 100 and 67. On the other hand, higher values of IMR are expected for Finland and Japan, i.e., 15 instead of 6 and 12 instead of 4, respectively. Although this seems a problem in terms of the fit of the regression model, however, we have no much objection in not rejecting these outlier observations.

Because of the sharp differences in mortality levels between areas in the world, the effect of the socio-economic variables might be different as well. We have already met the classification of the World into two groups: the developing and developed countries. So need we to test whether there is a difference or not in mortality regressions in the two groups. This will be done by applying the Chow test, due to Chow(1983). The test is based on the assumptions that the disturbances in mortality equations of the two groups are distributed normally with zero mean and constant or homoscedastic variance σ^2 and that the disturbances of the two regressions are independently distributed. We have already obtained the combined mortality regression of the whole sample with the residuals sum of squares (RSS) given by 16.296 with df = N-k (=91-2=98), where k is the number of parameters estimated and N is the number of active observations. We run two mortality regressions for the developing and developed countries, and obtain their residuals sum of squares (RSS), that is, RSS1 (=1.7508) and RSS2 (=0.70372), with df = N_1-k (=66-2=64) and N_2-k (=25-2=23), respectively. We then apply the F-test as

$$F = \{ (RSS - (RSS1+RSS2))/k \} / \{ (RSS1+RSS2)/(N-2k) \} = 245.304$$

and since it is exceeding the critical F under 2 and 87 degrees of freedom, we accept the hypothesis that the two mortality regressions of developing and developed countries are not the same. This is very clear from the result of the two regressions. The PC1 explains about 73% of the variation in the log of the infant mortality rate for the developing countries compared with only 12% for the developed countries. In other words, while for the first group the PC1 is very significant, for the developed group the probability value is only 8.4%. It might be the case that the countries in the second group have already reached a level in education, nutrition and economic situations where potential for further changes was more limited and hence their effect on mortality became latent, while for the first group the room for improvement in nutrition, education, and economic situations was very large and led directly to mortality decline, especially the infant mortality. The evidence for this is strengthened by the fact that the standard error of the residuals is similar for the two subgroups while the regression mean square is much greater for the underdeveloped group than for the developed group, showing much greater variation within the third world than in the other group.

The multi-step Chow test procedure can be substantially abridged by the use of two new variables: 'D' and 'DPC1'. While the variable 'D' is taking a value of '0' for the developing countries and '1' for the developed countries and hence will help us to distinguish between the intercepts of the two groups, the multiplicative variable 'DPC1' ('D' multiplied by PC1) will enable us to differentiate between slope coefficients. The modified mortality regression we reached is

$$\text{Log IMR} = \underset{(36.69)}{2.73} + \underset{(0.01)}{0.010\,D} + \underset{(13.04)}{0.00690\,PC1} + \underset{(0.36)}{0.00162\,DPC1}$$

$$R^2(\text{adjusted}) = 0.871\ ,\ \text{F-ratio} = 204.21$$

$$n = 91\ ,\ df = 87\ ,\ s = 0.1680$$

where inside the brackets are the t-values of the corresponding coefficients.

As this regression equation shows, both 'D' and 'DPC1' are statistically insignificant indicating no difference between the two groups. This result, however, is affected by the collinearity of 'D' and 'DPC1' with PC1 and hence might not contradict what we saw before.

8- Mortality and Factor Analysis

analysis is used here to identify a relatively small number of factors that can be used to represent relationships among the socio-economic variables. The aim is to see if we can identify separate economic, social and demographic dimensions from these variables. The basic assumption of factor analysis is that these underlying dimensions, or factors, can be used to explain complex phenomena. Observed correlations between the variables result from their sharing these factors. Each variable, then, is expressed as a linear combination of factors, called common factors, which are not actually observed.

To help us decide how many factors we need to represent the data, it is helpful to examine the percentage of the total variance explained by each. The total variance is the sum of the variance of each variable. For simplicity, all variables and factors are expressed in standardized form, with a zero mean and a unit variance.

Using the principal component method for extraction, and from the initial statistics for each factor along with the variance containing the variance explained by each factor, we found that the linear combination formed by the first factor has a variance (eigenvalue) of 9.67, which is about 74.4% of the total variance. Almost 85% of the total variance is attributable to the first two factors. The remaining eleven factors together account for 15% of the total variance. Thus a model with two factors may be adequate to represent the data. The two extracted factors can then be expressed from the observed variables as

$$\begin{aligned} FACTOR1 = {} & 0.94704 \log RPCI + 0.83885\, DCS + 0.86616\, ALRM \\ & + 0.86247\, ALRF - 0.93493 \log NPPD \\ & + 0.83990\, PPHS + 0.89545\, PPSW \\ & + 0.87071\, PPU - 0.93379 \log IMR \\ & - 0.89630\, PCP - 0.89247 \log TFR \\ & - 0.93663\, PLFA \end{aligned}$$

$$FACTOR2 = 0.95680\, FPLF$$

The 'communality' result shows that more than 70% of the variance of every variable is explained by either of the two factors. It remains for us, then, to say that : the first factor could be interpreted as a general index of development and the second factor is just a measure of the female participation in the labour force.

9- Summary

This study has traced the effect of the economic and social factors on mortality behaviour. It is believed that the reduction in mortality and fertility which have been experienced by countries in all over the world is due to the general improvement in these economic and social conditions. The real per capita income, the daily calorie supply, the adult literacy rate for males, the adult literacy rate for females, the number of population per physician, the percentage of population living in urban areas, the percentage of population with access to health services, and the percentage of population with access to safe water; are the variables which are suggested as determinants of mortality behaviour.

The Results showed that all the investigated variables are highly correlated with infant mortality rate, with correlation-coefficients greater than and equal 0.7 for all. As we expected, apart from the number of population per doctor all other variables turned to be negatively related with mortality. Because of the collinearity between our variables, the individual effect of many of them appeared to be negligible in the multiple regression of the mortality on these variables. This is actually what was happened when the log of the real per capita income, the adult literacy rate for females, and the percentage of population having access to safe water are the only variables which turned out to be significant in the linear multiple relationship with mortality. The check for 'muliticollinearity' problem suggested the existence of the problem. Accordingly, a remedy for multicollinearity was made using the principal component method where the first principal component was extracted to represent the data. This principal component could be interpreted as a contrast between the log of the number of population per doctor and the rest of the variables. It explained more than 87% of the variation in mortality.

In testing the difference or otherwise in mortality between the developing and developed countries, it was found that there is a considerable difference between the two groups. While the first principal component is statistically significant in the group of developing countries and explained more than 73% of the decline in mortality, it is even insignificant in the group of developed countries and explained only 12% of the variation in mortality. This does not mean that our socio-economic variables are no longer effective in mortality change in the developed countries. What happened is that neither the socio-economic variables nor mortality were varying in this group.

The factor analysis showed that all the socio-economic variables turned out to be represented by one and only one factor. This factor is identified, without any doubt, to be the infant mortality. So factor analysis confirmed the great association between the socio-economic variables and mortality behaviours.

It is therefore possible for the further reductions in mortality to occur with economic development and consequently improving health and medical technology, along with the better diet beside the greater educational attainment and awareness of health protection as an auxiliary factor. A concern must be given to control fertility levels. This is mainly to reach equilibrium and stability in the population growth, through the direct effect of mortality reduction of fertility decline, and hence better quality of life in the earth.

10- References

1. Benjamin, B.(1959). Demographic Analysis. George Allen and Unwin Ltd, London.
2. Bogue, D.J.(1969). Principles of Demography. John Wiley and Sons, Inc., New York.
3. Breusch, T.S. and A.R. Pagan(1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. Econometrica, vol. 47, pp. 1287-1294.
4. Cattell, R.B.(1978). The Scientific Use of Factor Analysis in Behavioural and Life Sciences. Plenum Press, New York.
5. Chow, G.C.(1983). Econometric Methods. McGraw-Hill Book Company, New York.
6. Dutta, M.(1975). Econometric Methods. South-Western Publishing company, Inc., Cincinnati.
7. Glass, D.V. and R.C. Revelle(1976). Population and Social Change. Edward Arnold Ltd, London.
8. Gray, H.P. and S.S. Tangri(1990). Economic Development and Population Growth: "Factors Influencing Population Growth and Fertility Reduction". Heath and Company, Massachusetts.
9. Habakkuk, H.J.(1974). Population Growth and Economic Development since 1750. Leicester University Press, Leicester.
10. Hauser, P.M.(1980). World Population and Development: Challenges and Prospects. Syracuse University Press, New York.

11. Hutchinson, J.(2003). Population and Food Supply. Cambridge University Press, Cambridge.

12. Koutsoyiannis, A.(1973). Theory of Econometrics. Haper & Row Publishers, Inc., New York.

13. Jolliffe, I.T.(2010). Principal Component Analysis. Springer Verlag, New York.

14. Maddala, G.S.(2009).Infrodncyir to Econometrics. McGraw-Hill Book Company, New York.

15. Mardia, K.V., J.T. Kent and J.M. Bibby(1980). Multivariate Analysis. Academic Press, London.

16. McIntosh, J.(1983). Reproductive Behaviour in Peasant Societies : A Theoretical and Empirical Analysis. Review of Economic Studies, pp. 133-142.

17. Morrison, D.F.(2004). Multivariate Statistical Methods. McGraw-Hill Book Company, New York.

18. Overbeek, J.(1976). The Population Challenge. Greenwood Press, London.

19. Pichat, J.B.(1973). Main Trends in Demography. George Allen and Unwin Ltd, London.

20. Pollard, A.H., F. Yusuf and G.N. Pollard(1990). Demographic Techniques. Pergamon Press Austraila, Rushcutters Bay.

21. Population Crisis Committee(1990). 1990 Report on Progress Towards Population Stablization. Population Crisis Committee, Washington.

22. Preston, S.H.(1978). The Effect of Infant and Child Mortality on Fertility. Academic Press, London.

23. Stewart, M.B. and K.F. Wallis(1981). Introductory Econometrics, 2nd ed. Basil Blackwell, Oxford.

24. Teitelbaum, M.S. and J.M. Winter(1986). The Fear of Population

Decline. Academic Press, London.

25. The World Bank(2010). World Development Report: "World Development Indicators". Oxford University Press.
26. UN(2005). World Population Prospects as Assessed in 1968. United Nations, New York.
27. UN(2007). Indirect Techniques for Demographic Estimation. United Nations, New York.
28. UN(2009). Modelling Economic and Demographic Development. United Nations, New York.
29. UN(2010). The Determinants and Consequences of Population Trends, No. 50. United Nations, New York.
30. UNDP(2010). Human Development Report. Oxford University Press.
31. UNFPA(2010). The State of World Population. UNFPA, New York.
32. UNICEF(2010). The State of World's Children. Oxford University Press.