Using Categorical Data Analysis to Estimate Population Change in EU25: Absolute Numbers and Crude Rates

Adanma Ngozi Ohia, Joel Chiadikobi Nwaubani, Uchechi Mgbafor Ezeji, Uzokwe Chinwe Aduago, Christiana Uzoma Ezechukwu, Opara Peace

Abstract — Statistics on population change and the structure of population are increasingly used to support policymaking and to provide the opportunity to monitor demographic behaviour within political, economic, social and cultural contexts. Specifically, this concerns demographic developments that focus on a likely reduction in the relative importance of the working age population and a corresponding increase in the number of older persons. These statistics may be used to support a range of different analyses, including studies relating to population ageing and its effects on the sustainability of public finance and welfare, the evaluation of fertility as a background for family policies, or the economic and social impact of demographic change. This research aims to highlight the population change in twenty-five countries of the European Union. We consider the use of categorical data analysis to estimate the population change in EU25: absolute numbers and crude rates from 2003-2017. The data used in this study are from the Eurostat/World population prospects and estimated on actual base year from. Since the main focus is to have a better understanding of the population change in EU25, the analysis of association table (ANOAS) is given in order to ascertain the percentage of the data which is covered by each model. We find and estimate the association model with the best fit and in conclusion we find out that the Row-Column Effects Association Model (RC) of the multivariate model (Ms=4) has the best fit among all - covering a total of 99.9% of the data observed.

Index Terms — Association models, Log-linear and non-linear models, population and EU25.

I. INTRODUCTION

Population change or population growth in a given year is the difference between the population size on 1 January of the given year and the corresponding level from 1 January of the previous year. It consists of two components: natural change and net migration plus statistical adjustment. Natural population change is the difference between the number of live births and the number of deaths. If natural change is positive, then it is often referred to as a natural increase. Net migration is the difference between the number of immigrants and the number of emigrants. The rate may be positive or negative. The growth rate is a factor in determining how great a burden would be imposed on a country by the changing needs of its people for infrastructure (e.g., schools, hospitals, housing, roads), resources (e.g., food, water, electricity), and jobs. Rapid population growth can be seen as a threat to a nation as well as threatening by neighboring countries.

Europe’s ageing society and its relatively static number of inhabitants may be contrasted against a rapid expansion in the world’s population, driven largely by population growth in developing countries. However, the demographic challenge that the EU-27 is confronted with is by no means unique. Most developed, and also some emerging economies, will undergo considerable changes in their demographic composition in the next five decades. Shrinking working age populations, a higher proportion of elderly persons, and increasing old-age dependency rates suggest that there will be a considerable burden to provide for social expenditure related to population ageing (pensions, healthcare, institutional care). The challenges associated with an ageing society are likely to be even more acute in countries such as Japan and South Korea where the old-age dependency ratio will rise rapidly and to a very high level, while the share of older persons in the total population is also expected to increase at a rapid pace in China [1].

A. Historical background

The European Union (EU) is a confederation of twenty-eight nations. It was established in 1992 by the Treaty of the European Union (Maastricht Treaty) and it is the de facto successor to the European Communities (ECSC, EEC, EURATOM) of the six member states established in 1951, 1957 and 1965 respectively. Since then, new enlargements have increased the number of its member States and its competencies have been extended. It is an open process of European integration and it is one of the largest economic and political entities in the world [2]. Luxembourg City, together with Brussels and Strasbourg, is one of the three official capitals of the European Union and the seat of the European Court of Justice – the highest judicial authority in the EU. The European Union is a single market with a common commercial, agricultural, fisheries and regional policy to help the poorer regions. It introduced a single currency, the Euro, adopted by 15 Member States.

The European Union was also formally established when the Maastricht Treaty came into force on 1 November 1993. In 1995 Austria, Sweden and Finland entered the newly established Union. The next treaty was signed in Amsterdam in 1997. In 1999 the euro replaced the national currencies in the 11 Member States, the so-called Eurozone. Greece joined

Published on September 11, 2020.
Adanma Ngozi Ohia, University of Portharcourt, Nigeria.
Joel Chiadikobi Nwaubani, Advance Higher Education London, UK.
(corresponding e-mail: joelnocool@yahoo.com)
Uchechi Mgbafor Ezeji, Amuzukwu Girls’ Secondary School, Nigeria.
Uzokwe Chinwe Aduago, Michael Okpara University of Agriculture, Nigeria.
Christiana Uzoma Ezechukwu, Agricultural Development Programme (ADP), Nigeria.
Opara Peace, Abia Secondary Education Management Board (SEMB), Nigeria.

DOI: http://dx.doi.org/10.24018/ejbmrm.2020.5.5.481
Vol 5 | Issue 5 | September 2020
the Euro in 2001. In 2004, ten new countries (eight of which were formerly communist countries) joined the European Union. In addition, Romania and Bulgaria joined the EU in 2007, while Croatia joined in 2013. Slovenia also joined the Eurozone in 2007. In 2008 Cyprus and Malta adopted the euro respectively [3].

B. Europe Population from 2014–2017 (in million inhabitants)

Europe which is about 11% of the world’s population is the third-most populous continent after Asia and Africa with a population of 739 million. The population of the EU27 was 507.24 million in 2014, 508.52 million in 2015, 510.18 million in 2016 and 511.37 million in 2017. Europe is bordered by the Arctic Ocean to the north, the Atlantic Ocean to the west, the Mediterranean Sea to the south, and the Black Sea and connected waterways to the southeast. Europe is a peninsula that makes up the western fifth of the Eurasian landmass. Its maritime borders are made up of the Arctic Ocean to the north, the Atlantic Ocean to the west, and the Mediterranean, Black, and Caspian Seas to the south. The southern regions are more mountainous, while moving north the terrain descends from the high Alps, Pyrenees, and Carpathians, through hilly uplands, into broad, low northern plains, which are vast in the east. This extended lowland is known as the Great European Plain, and at its heart lies the North German Plain. An arc of uplands also exists along the north-western seaboard, which begins in the western parts of the islands of Britain and Ireland, and then continues along the mountainous, fjord-cut spine of Norway.

C. European Population Growth

Europe’s population growth rate is approximately 0.212%. The average annual percent change in the population, resulting from a surplus or deficit of births over deaths and the balance of migrants entering and leaving a country. The rate may be positive or negative. The growth rate is a factor in determining how great a burden would be imposed on a country by the changing needs of its people for infrastructure, resources, and jobs. Rapid population growth can be seen as threatening by neighboring countries.

D. European Union Member States

The European Union is described as “a family of democratic European countries”. On 23 July 1957 six founding members formed the European Coal and Steel Community, which was the first stage of today’s EU. Twenty-one countries have since been introduced into successive waves of enlargement. The EU17 consists of Belgium, Germany, Estonia, Ireland, Greece, Spain, France, Italy, Cyprus, Luxembourg, Malta, the Netherlands, Austria, Portugal, Slovakia, Slovenia, Austria, Ireland, Latvia, Lithuania, Luxembourg, Hungary, Croatia, Malta, the Netherlands, Austria, Poland, Portugal, Romania, Slovenia, Slovakia, Finland, Sweden and the United Kingdom.

| Year | Enlargement History | Total |
|------|---------------------|-------|
| 1952/1958 | Belgium Germany1, France | 6 |
| 1961 | Ireland Luxembourg Holland | 9 |
| 1968 | Ireland United Kingdom | 10 |
| 1973 | Cyprus Greece | 12 |
| 1975 | Austria Finland Sweden | 15 |
| 1980 | Czech Republic Hungary Poland Estonia | 25 |
| 2004 | Malta Cyprus Lithuania Latvia Bulgaria Romania | 27 |
| 2007 | Croatia1 | 28 |

Notes:
1 On 3 October 1990, East Germany joined with West Germany and its territory automatically became part of the EU.
2 In 1973, as in 1995, Norway, despite the fact that it signed the respective Accession Treaties, did not endorse them because of the negative referendum of the Norwegian citizens.
3 Greenland, a self-governing territory under Danish sovereignty since 1979, left the European Communities in 1985 following a referendum.
4 The population of Croatia in 2013 was 4,475,611 inhabitants and became the 28th member of the EU on July 1, 2013. In April 2009 it became a member of NATO.

Any country willing to join the European Union must adhere to the economic and political terms, known as the Copenhagen Criteria (following the Copenhagen Summit in June 1993). This basically requires a secular, democratic government, the rule and respect of law and corresponding social and institutional freedoms. Under the EU Treaty, each Member State and the European Parliament must agree unanimously on any enlargement.

E. Candidate countries (En route to EU membership)

There are currently six candidate countries to join the EU. These are:
- Albania;
- Republic of North Macedonia;
- Iceland;
- Montenegro;
- Serbia;
- Turkey.

In addition, the Balkan countries Bosnia and Herzegovina and Kosovo are recognized as potential candidates.

F. Demographics Data Population

The most populous country in the EU is Germany with 82.2 million inhabitants. The top five are: The United Kingdom with (65.65 million), France (64.1 million), Italy (59.29 million) and Spain (46.4). Greece ranks 10th with a population of 11.1 million according to the official census of 2017. The country with the smallest population in the EU is Malta with only 0.4 (432089) million Residents followed by Luxembourg with 0.5 (590321) million of which 50.9% are Luxembourgers while 49.1% are foreign citizens. Cyprus is
followed after Luxembourg with 1.19 million. The largest city in the EU is London, with the population of the entire metropolitan complex according to 2011 census, 8 173 941. The metro population in 2018 estimated the population of London to be 8 787 892.

In 2011, natural increase (the positive difference between live births and deaths) added 0.4 million (32 %) to the population growth in the EU-27. Some 68 % of the EU-27’s population growth came from net migration plus statistical adjustment, which continued to be the main determinant of population growth, contributing 0.9 million in 2011.

The contribution of net migration plus statistical adjustment to the total population change in the EU-27 has been greater than that of natural change since 2002. The share of net migration plus statistical adjustment in total population growth peaked, in relative terms, in 2003 (95 % of total change). Since this date, the contribution of net migration plus statistical adjustment decreased somewhat. Thus, the share of natural change in total population growth followed an upward development up to 2009 (37%) and declined again over the most recent years (from 2010 onwards). In continuation, Table 2 below shows the 10 largest cities in the EU.

TABLE 2: Top 10 largest cities in the European Union – (estimate 2017) [3]

| A/A | Country       | Town Planning Complex       | Population |
|-----|---------------|-----------------------------|------------|
| 1   | France        | Paris                       | 122741346  |
| 2   | United Kingdom| London                      | 8 787892   |
| 3   | Germany       | Berlin                      | 7 245576   |
| 4   | Spain         | Madrid                      | 6514620    |
| 5   | Spain         | Barcelona                   | 64393081   |
| 6   | Germany       | Essen – Bochum- Dortmund    | 5 126013   |
| 7   | Italy         | Milan                       | 4851565    |
| 8   | Italy         | Rome                        | 4039813    |
| 9   | Greece        | Athens                      | 3750124    |
| 10  | Poland        | Warsaw                      | 1 781740   |

G. European population compared with world population

By 1 July 2017 the world’s population had reached 7 547 859 (Seven billion, five hundred and forty-seven million, eight hundred and fifty nine thousand) inhabitants, according to United Nations’ (UN’s) World Population Prospects, 2010 revision. Most population scenarios project that population numbers will continue to grow in the coming decades. Asia accounted for the majority of the world’s population (just over 60 % in 2012) with 4.250 billion inhabitants, while Africa was the next most populous continent with 1.070 billion inhabitants, or 15.2 % of the global total. In comparison, there were 504 million inhabitants in the EU-27 in 2012, accounting for just over 7 % of the world’s population.

Global population doubled between 1960 and 2000. From the turn of the millennium to 2012, the world’s population grew by a further 929 million persons. The overall increase in global population between 1960 and 2017 can be largely attributed to growth in Asia, Africa and Latin America.

The latest UN population projections that the pace at which the world’s population is expanding will slow somewhat in the coming decades; nevertheless, the total number of inhabitants is projected to reach more than 9 550 944 891 by 2050. According to these projections (the medium variant), the world’s population will also be relatively older (in other words, with a higher median age) in 2050. At the same time, according to Eurostat's population projections, the EU-27's population is predicted to grow at a slower rate before peaking in 2040–45, contracting thereafter.

H. World Population Prospects 2017

The 2017 world population of 7.6 billion is expected to reach 8.6 billion in 2030, 9.8 billion in 2050 and 11.2 billion in 2100, according to a new United Nations report being launched today. With roughly 83 million people being added to the world’s population every year, the upward trend in population size is expected to continue, even assuming that fertility levels will continue to decline [1].

TABLE 3: Continental Population

| Continent | Population 2050 |
|-----------|-----------------|
| Asia      | 4 140 000 000   |
| Africa    | 995 000 000     |
| Europe    | 739 000 000     |
| North America | 529 000 000 |
| South America | 386 000 000 |
| Australia | 36 000 000      |

I. Shifts in country population rankings / Population 2050 (estimate)

The new projections include some notable findings at the country level. China (with 1.4 billion inhabitants) and India (1.3 billion inhabitants) remain the two most populous countries, comprising 19% and 18% of the total global population. By 2024, or in roughly seven years, the population of India is expected to surpass that of China.

TABLE 4: Country Population Rankings 2050 (estimate)

| Country   | Population 2050 |
|-----------|-----------------|
| 1         | India           | 1 620 050 849   |
| 2         | China           | 1 384 976 976   |
| 3         | Nigeria         | 440 355 062     |
| 4         | USA             | 400 853 042     |
| 5         | Indonesia       | 321 377 092     |
| 6         | Pakistan        | 271 081 825     |
| 7         | Brazil          | 231 120 024     |
| 8         | Bangladesh      | 201 947 716     |
| 9         | Ethiopia        | 187 572 656     |
| 10        | Philippines     | 157 117 506     |
| Total     | 9 550 944 891   |

Among the ten largest countries worldwide, Nigeria is growing the most rapidly as shown in Table 4 above. Consequently, the population of Nigeria, currently the world’s 7th largest, is projected to surpass that of the United States and become the third largest country in the world shortly before 2050.

J. Large movements of refugees and other migrants

There continue to be large movements of migrants between regions, often from low- and middle-income
countries toward high-income countries. The volume of the net inflow of migrants to high-income countries in 2010-2015 was 3.2 million per year. Although international migration at or around current levels will be insufficient to compensate fully for the expected loss of population tied to low levels of fertility, especially in the European region, the movement of people between countries can help attenuate some of the adverse consequences of population ageing.

The Syrian refugee crisis has had a major impact on levels and patterns of international migration in recent years, affecting several countries. Generally, the net outflow from the Syrian Arab Republic was 4.2 million persons from 2010-2017. Most of these refugees migrated to Syria’s neighboring countries, contributing to a substantial increase in the net inflow of migrants especially to Turkey, Greece, Lebanon and Jordan.

II. STATISTICAL DATA AND PROCESSING OF THE RESULTS

The data used for this study was obtained from the Eurostat/World Population Prospects, the statistical office of the European Union, calculated on an annual basis. The processing of observations (data) consists of the population change in the EU25: absolute numbers and crude rates. Below in Table 5 are the lists of data under our study for the subsequent years from 2003-2017.

| TABLE 5: POPULATION CHANGE IN EU27: ABSOLUTE NUMBERS AND CRUDE RATES [5] |
|-----------------|------------------|-----------------|-----------------|
| geo time        | 2003             | 2004            | 2005            |
| Belgium         | 10 004 500       | 10 045 200      | 10 085 400      |
| Czech Republic  | 10 308 600       | 10 319 100      | 10 332 900      |
| Denmark         | 5 154 300        | 5 171 400       | 5 186 600       |
| Germany         | 80 013 900       | 80 624 600      | 81 156 400      |
| Estonia         | 1 561 300        | 1 533 100       | 1 492 400       |
| Ireland         | 3 534 300        | 3 558 500       | 3 586 400       |
| Greece          | 10 256 300       | 10 369 900      | 10 465 600      |
| Spain           | 389 391 000      | 390 677 000     | 391 894 000     |
| Italy           | 584 154 915      | 587 453 995     | 590 297 305     |
| Cyprus          | 595 110 000      | 611 200 000     | 626 100 000     |
| Latvia          | 26 506 000       | 26 644 000      | 25 633 000      |
| Lithuania       | 37 040 200       | 37 010 000      | 36 826 000      |
| Luxembourg      | 37 800 000       | 39 220 000      | 39 750 000      |
| Hungary         | 103 734 000      | 103 693 000     | 103 575 000     |
| Malta           | 35 770 000       | 36 130 000      | 36 640 000      |
| Netherlands     | 150 690 000      | 151 840 000     | 152 930 000     |
| Austria         | 77 549 000       | 78 407 000      | 79 050 000      |
| Poland          | 382 426 000      | 383 637 000     | 384 614 000     |
| Portugal        | 99 679 000       | 99 700 000      | 99 826 000      |
| Slovenia        | 1 999 400        | 1 996 500       | 1 991 800       |
| Slovakia        | 53 033 000       | 53 505 000      | 53 548 000      |
| Finland         | 50 138 000       | 50 420 000      | 50 665 000      |
| Sweden          | 861 740 000      | 868 610 000     | 871 860 000     |
| United Kingdom  | 596 440 000      | 599 504 000     | 601 684 000     |
With the help of the Categorical Data Analysis program (CDAS), we were able to ascertain the results of our data for the population change in the EU25: absolute numbers and crude rates from 2003-2017. Table 6 below shows the values (prices) of all the models to be estimated. The statistics used to compare and contrast the association (correlation) of the models are \( X^2 \) (Pearson Chi-Square) and the Likelihood-Ratio Chi-Square \( G^2 \)-probability ratio [6]. In continuation, the following results are shown below:

| Models | \( \chi^2 \) (Pearson/Chi-Square) | Likelihood-Ratio Chi-Square \( G^2 \) | Degrees of Freedom | Index of Dissimilarity | Maximum Deviation |
|--------|----------------------------------|----------------------------------------|--------------------|-----------------------|------------------|
| O      | 123379.93577                     | 123599.24741                          | 80                 | 0.00265               | 0.00000003       |
| U      | 103522.90731                     | 103376.02003                          | 79                 | 0.00236               | 0.00061689       |
| R      | 103394.71974                     | 103242.19437                          | 70                 | 0.00236               | 0.00070555       |
| C      | 20439.20893                      | 20437.38281                           | 72                 | 0.00104               | 0.00001520       |
| R+C    | 20306.16632                      | 20301.82262                           | 63                 | 0.00103               | 0.00008217       |
| RC     | 14672.60555                      | 14675.23959                           | 63                 | 0.00092               | 0.84482370       |

**A. Association Model**

In the categorical data analysis system, we apply the methodology by considering six of the most commonly used association models. These are:

1. The model of Independence or null association model which is also symbolized by (O) and holds that there is no relationship between the variables. The log-linear model is:
   \[
   \log(F_{ij}) = \lambda + \lambda_{A(i)} + \lambda_{B(j)},
   \]
   where \( \lambda \) denotes the natural logarithm, \( F_{ij} \) the expected frequencies under the independence model, \( \lambda_{A(i)} \) are the rows main effects, \( \lambda_{B(j)} \) are the columns main effects [7].

2. The Uniform association model, which is symbolized by (U) in log-linear form is:
   \[
   \log(F_{ij}) = \lambda + \lambda_{A(i)} + \lambda_{B(j)} + \varphi X_i Y_j
   \]
   where \( \varphi \) is a single parameter for interaction, \( X_i, Y_j \) are the scores for the row and column variables (\( i = 1, \ldots, I \), \( j = 1, \ldots, J \)) respectively.

3. The row effects model (R) where linear-by-linear interaction holds that [8],
   \[
   \log(F_{ij}) = \lambda + \lambda_{A(i)} + \lambda_{B(j)} + \varphi_{i} y_j
   \]
   where \( y_j \) are fixed scores for the column variable (\( j = 1, \ldots, J \)), \( \varphi_{i} \) are unknown scores for the row variable (\( i = 1, \ldots, I \)).

4. The column effects model (C) is the same as the R model with a change in subscripts:
   \[
   \log(F_{ij}) = \lambda + \lambda_{A(i)} + \lambda_{B(j)} + \varphi_1 x_i
   \]
   where \( x_i \) are fixed scores for the row variable (\( i = 1, \ldots, I \)), \( \varphi_{j} \) are unknown scores for the column variable (\( j = 1, \ldots, J \)).

5. The model that allows both row and column effects in additive form is called the R+C model [9]. The log-frequency version of the above model is
   \[
   \log(F_{ij}) = \lambda + \lambda_{A(i)} + \lambda_{B(j)} + \sum_{k=1}^{I-1} p_k x_i Z_{A(k)} + \sum_{k=1}^{J-1} q_k y_j Z_{B(k)}
   \]
   where \( f_{ij} \) are the scores (as defined earlier), \( Z_{A(i)}, Z_{B(j)} \) denote to variables (dummy variables) for the row and column levels respectively.

6. The model, instead of additive row and column effects on the local odds ratios, has multiplicative effects called the R*C model or model II [10]. The log-multiplicative model is:
   \[
   \log(F_{ij}) = \lambda + \lambda_{A(i)} + \lambda_{B(j)} + \varphi_{i} y_j
   \]
   where the row score parameters \( \mu_i \) and column score parameters \( y_j \) are not known but are estimated in the data.

We aim at finding out the model that has the best fit among the 6 models which we are examining, i.e., population change in the EU25: absolute numbers and crude rates from 2003-2017. For this reason, first, we are going to examine the Index of Dissimilarity (L2), which shows that, the lesser the number, the more our model will give the best fit to match the data under consideration.

We analyze the six association model describe above, with the help of the categorical data analysis statistical program [11]. We used the Pearson chi-squared \( (X^2) \) statistics, the likelihood-ratio chi-square \( (G^2) \) statistics and the index of dissimilarity which is equated by:

\[
D = \frac{\sum_{ij} \left| f_{ij} - F_{ij} \right|^2}{\sum_{ij} F_{ij}^2}
\]

where:
- \( f_{ij} \) are the observed frequencies,
- \( F_{ij} \) are the expected frequencies (under the model).

Additionally, we have the following results as shown in the table below:

| Models | Index of Dissimilarity (D) |
|--------|---------------------------|
| 1. Null Association-Independence Model (O) | 0.00265 |
| 2. Uniform Association Model (U) | 0.00236 |
| 3. Row-Effects Association Model (R) | 0.00236 |
| 4. Column-Effects Association Model (C) | 0.00104 |
| 5. Row+Column Effects Association Model (R*C) | 0.00103 |
| 6. Row Column Effects Association Model (R*C) | 0.00092 |

At first sight in Table 7 above, it seems that the Row Column Effects of the Association Model (R*C) adjusted
better to the percentage of the population change in the EU25 for the years under our study (having the lowest index of dissimilarity) with D=0.00092.

Since we have models with similar lower ratio, we justify the model with the best fit to match both countries and years by calculating the Index BIC (Bayesian Information criterion). In statistics, the Bayesian Information Criterion (BIC) or Schwarz criterion is a Criterion for model selection among a finite set of models; the model with the lowest BIC is preferred. It is based, in part, on the likelihood function. When fitting models, it is possible to increase the likelihood by adding parameters but doing so may result in over-fitting. BIC attempt to resolve this problem by introducing a penalty term for the number of parameters in the model and gives the best solution [12].

The formula for this calculation is:

\[ BIC = G^2 - (D.F.) \log(n) \]  
\[ \text{(2)} \]

Notations:
- \( G^2 \): the likelihood-ratio chi-square statistics;
- d.f.: degrees of freedom of the models;
- n: the size of the sample (2240630715.0000);
- \( \log(n) = \log(2240630715.0000) = 21.5300232 \).

When comparing a number of models, the model with the smallest index of BIC is assumed to be the best. So, we choose the models that have similar and high INDEX OF DISSIMILARITY out of the six models. More precisely, we will consider the 3rd, 5th and 6th models respectively.

Subsequently, the calculation is as follows:

- 3rd model: BIC = \( G^2 - (D.F.) \log(n) = 21197.91551 - (72*21.5300232 = 19647.7538 \)
- 5th model: BIC = \( G^2 - (D.F.) \log(n) = 21036.38760 - (63*21.5300232) = 19679.9961 \)
- 6th model: BIC = \( G^2 - (D.F.) \log(n) = 15327.73587 - (63*21.5300232) = 13971.3444 \)

As we can see from the above calculations, the 6th model (Row Column-effects of the Association Model (RC)) finally accounts for the best fit from the results since it has the smallest index of BIC.

B. Analysis of the Association Models

In continuation, the association model passes through several tests to ascertain the accuracy, quality, or satisfactory fit of each model. Tests are done through the use of the likelihood-ratio chi-square (\( G^2 \)) statistics and the Pearson chi-squared (\( \chi^2 \)) distribution. In the case of the \( \chi^2 \) distribution, the Statgraph programme will be of good help.

Initially, we observe that the likelihood-ratio chi-square statistic for the Independence model (O) which is \( G^2=123960.12346 \) with 80 degrees of freedom. The 95% reference point chi-square distribution is 102.023. It has an unacceptable fit because the \( \chi^2 \) distribution is much smaller than the likelihood-ratio chi-square statistic \( G^2 \).

Subsequently, the Uniform association model has \( G^2=104991.90963 \) with 79 degrees of freedom. The 95% reference point chi-square distribution is 100.887. As it could be noticed, this statistic is not accepted and does not have a satisfactory fit since the \( \chi^2 \) distribution is much smaller than the likelihood-ratio chi-square statistic \( G^2 \).

Moreover, the statistic \( G^2 \) for the Row model (R) is reduced dramatically for 103931.40198 with 70 degrees of freedom. The 95% reference point chi-square distribution is 90.6892. The row model is also not accepted because the \( \chi^2 \) distribution is smaller than the likelihood-ratio chi-square statistic \( G^2 \).

The Column model (C) has \( G^2=21197.91551 \) with 72 degrees of freedom. The 95% reference point chi-square distribution is 92.9779 which show even the worst fit as we could observe that the \( \chi^2 \) distribution is very much smaller than the likelihood-ratio chi-square statistic \( G^2 \).

The statistics of the model R+C, that takes into account the effects for both Countries and Years in additive form, is \( G^2=21036.38760 \) with 63 degrees of freedom. The 95% reference point chi-square distribution is 82.6494 has equally unacceptable fit since the \( \chi^2 \) distribution is smaller than the likelihood-ratio chi-square statistic \( G^2 \).

By virtue of the index of dissimilarity of the models, the model R+C has the best fit. However, upon clear test to ascertain the accuracy, quality, or satisfactory fit of the model, we find out that it has a very poor fitting, making it unacceptable [13, 14].

Consequently, we have to find out the degree or level of effects on each model. To verify this, we will construct the analysis of association (ANOAS) table.

C. Analysis of Association Table (ANOAS)

The ANOAS table in table 6 below was given by [15]. In this table, the \( \chi^2 \) is divided so that it can be used as two factor analysis of variance by making use of the \( G^2 \) (0) statistics for the base (zero) independence model which measures the total deviation of the variables. In other words, we can find the percentage of the baseline chi-squared \( \chi^2 \) distribution, which have effects on each of the models on the phenomenon being studied.

| TABLE 8: ASSOCIATION ANALYSIS |
|-----------------------------|
| Models | Likelihood-\( G^2 \) | Degrees of Freedom | Index of Dissimilarity |
|--------|----------------|-----------------|---------------------|
| O      | 123960.12346   | 80              | 0.00266             |
| U      | 104991.90963   | 79              | 0.00237             |
| R      | 103931.40198   | 70              | 0.00237             |
| C      | 21197.91551    | 72              | 0.00104             |
| R+C    | 21036.38760    | 63              | 0.00103             |
| RC     | 15327.73587    | 63              | 0.00093             |

The analysis of association table has the following differences of our models: O-U is the overall or total effects of the models, U-C are the column effects model, C-CR are the column effects model which gives the effects of the column, RC and the residuals of the models.
From the ANOAS table we created, the uniform association model (U) as shown in Table 7 above, covers 17.13% of the total baseline chi-squared $X^2$ distribution, the Row model (R) accounts for a large percentage of 68.17%, the Column-effects which gives the Row-effects (C) covers a small percentage of 1.23%. Finally, the row column effects (Residuals) are weak also because the RC model accounts for only 13.47%of the baseline chi-squared value.

We could therefore observe that at the rate of 13.47%, the variation which is attributed to the null-independence has been measured from the model of RC. This rate is quite satisfactory and we can say that the corresponding percentage of population change in 25 countries of the European Union (Belgium, Czech Republic, Denmark, Germany, Estonia, Ireland, Greece, Spain, France, Italy, Cyprus, Latvia, Lithuania, Luxembourg, Hungary, Malta, Netherlands, Austria, Poland, Portugal, Slovenia, Slovakia, Finland, Sweden and United Kingdom), as shown from the data depended slightly positive on the association of both the countries and also from the years under our study (2003-2017).

We could as well say that the percentage of the data is explained by the column effects model (on the local odds ratios in a multiplicative way), thus, giving it a satisfactory fit but not acceptable due to poor adjustment that was found earlier in our study, i.e. the value of the Pearson chi-squared $X^2$ distribution for the 95% reference point is unacceptable because it has a bad fit.

Furthermore, because our best model (RC) under our study is unacceptable because it has a bad fit based on the index BIC upon clear examination to test the accuracy, quality or satisfactory fit of the model, therefore, we shall proceed to examining the multivariate model to find the model with a satisfactory fit.

### III. THE MULTIVARIATE MODELS

In the RC (M) association model, M represents the dimension fit to be, which is utilized by the row-column dimension (RCDIM PROGRAM). As shown in Table 10 below the multivariate model RC (M=4) is the acceptable model with the best fit [13].

The results are as follows:

| Effects                  | Models | Df  | $G^2$ | Percentage |
|--------------------------|--------|-----|-------|------------|
| 1. General               | O-U    | 1   | 19868.2138 | 17.13%     |
| 2. Rows                  | U-C    | 7   | 82893.9941 | 68.17%     |
| 3. Column-effects which provides the Row-effects | C-RC | 9   | 1615.2791  | 1.23%      |
| 4. Residuals             | RC     | 63  | 153277.3587 | 13.47%     |
| Total                    | O      | 80  | 123960.12346 | 100.00%    |

Model RC (1) multivariate row, column, M=1
Model RC (2) multivariate row, column, M=2

### A. Examination of the multivariate model

In the RC (M) association model, M represents the dimension to be fit, which is utilized by PROG RCDIM. As it is shown above the multivariate model RC (M=4) is the acceptable model with the best fit.

The multivariate model RC (1) with M=1 has likelihood-ratio chi-square statistic $G^2 = 15327.85142$ with 63 degrees of freedom (d.f.). Thus, we find out that the 95% reference point of the $X^2$ distribution is 82.6494. It has a bad fit. We could conclude therefore that this model has a bad fit because the $X^2$ distribution is much smaller than the likelihood-ratio chi-square statistic $G^2$.

Moreover, the multivariate model RC (2) with M=2 has likelihood-ratio chi-square statistic $G^2 = 486.20303$ with 48 degrees of freedom (d.f.). The 95th percentile of the reference chi-square distribution is 65.2594. Similarly, this model has a bad fit since the $X^2$ distribution.

The multivariate model RC (3) with M=3, has likelihood-ratio chi-square statistic $G^2 = 22.35796$ with 35 degrees of freedom (d.f.). Thus, we find out that the 95% reference point of the $X^2$ distribution is 49.8718. It has a very bad fit, thus, making it unacceptable.

Finally, the multivariate model RC (4) with M=4, has likelihood-ratio chi-square statistic $G^2 = 22.35796$ with 24 degrees of freedom (d.f.). Thus, we find out that the 95% reference point of the $X^2$ distribution is 36.35796. Here we could see that the multivariable RC model with M = 4 has the best fit. We equally observed that the model M=4 covers $\{165674.62855-62.918999 \} / 165674.62855 = 99.9%$ of all observations.

When a model with a smaller M has is satisfactory fit, it is accepted because it gives a better explanation of the interaction between the rows and columns. Therefore, we’ll prefer the multivariate model M=4 which has a perfect fit.

### B. Evaluation of the multivariate model

The practical implementation of multivariate statistics to a particular problem involve several types of univariate and multivariate analysis in order to understand the relationships between variables and their relevance to the actual problem being studied. In addition, multivariate statistics is concerned with multivariate probability distributions, in terms of both: how these can be used to represent the distributions of observed data; and how they can be used as part of statistical inference, particularly where several different quantities are of interest to the same analysis [13].

The expected frequencies under the independent and column effects models for population change in EU25: absolute numbers and crude rates from 2003-2017 are given below:

Note: The multivariate model RC (M=4) seems to give much better fit, particularly at the end of nominal scale.
| Countries | Years | Data | Values of O Model (f_i) | Values of RC Model (F_i) |
|-----------|-------|------|-------------------------|-------------------------|
| ROW | COLUMN | DATA | OBSERVED | EXPECTED |
| 1 | 1 | 95 | 153 | 80 |
| 2 | 1 | 85 | 180 | 98 |
| 3 | 1 | 160 | 98 | 141 |
| 4 | 1 | 187 | 96 | 172 |
| 5 | 1 | 300 | 195 | 366 |
| 6 | 1 | 183 | 88 | 166 |
| 7 | 1 | 126 | 93 | 146 |
| 8 | 1 | 74 | 90 | 75 |
| 9 | 1 | 41 | 87 | 39 |
| 10 | 1 | 13 | 33 | 15 |
| 11 | 1 | 13 | 30 | 11 |
| 2 | 1 | 473 | 600 | 462 |
| 3 | 1 | 544 | 339 | 556 |
| 4 | 1 | 425 | 707 | 437 |
| 5 | 1 | 778 | 386 | 733 |
| 6 | 1 | 365 | 378 | 421 |
| 7 | 1 | 294 | 375 | 269 |
| 8 | 1 | 231 | 344 | 234 |
| 9 | 1 | 447 | 363 | 452 |
| 10 | 1 | 249 | 355 | 252 |
| 11 | 1 | 315 | 341 | 305 |
| 12 | 1 | 326 | 347 | 301 |
| 13 | 1 | 443 | 354 | 469 |
| 1 | 2 | 1820 | 2835 | 1804 |
| 2 | 2 | 1756 | 1602 | 1776 |
| 3 | 2 | 1723 | 3337 | 1739 |
| 4 | 2 | 1947 | 1786 | 1967 |
| 5 | 2 | 2099 | 1770 | 2004 |
| 6 | 2 | 1921 | 1625 | 1913 |
| 7 | 2 | 1967 | 1716 | 1982 |
| 8 | 2 | 1938 | 1675 | 1941 |
| 9 | 2 | 2045 | 1610 | 2037 |
| 10 | 2 | 2043 | 1639 | 2026 |
| 11 | 2 | 2045 | 1671 | 2063 |
| 12 | 2 | 4192 | 5678 | 4230 |
| 13 | 2 | 3970 | 3209 | 3923 |
| 14 | 2 | 7428 | 5419 | 7424 |
| 15 | 2 | 5723 | 5289 | 5723 |
| 16 | 2 | 4469 | 5085 | 4477 |
| 17 | 2 | 3768 | 5177 | 3768 |
| 18 | 2 | 4500 | 5279 | 4497 |
| 19 | 2 | 27 | 57 | 30 |
| 20 | 2 | 26 | 32 | 22 |
| 21 | 2 | 26 | 67 | 23 |
| 22 | 2 | 20 | 36 | 21 |
| 23 | 2 | 23 | 36 | 32 |
| 24 | 2 | 47 | 35 | 41 |
| 25 | 2 | 26 | 32 | 29 |
| 26 | 2 | 37 | 34 | 33 |
| 27 | 2 | 25 | 33 | 25 |
| 28 | 2 | 56 | 32 | 56 |
| 29 | 2 | 70 | 33 | 70 |
| 30 | 2 | 78 | 33 | 78 |
| 31 | 3 | 3493 | 7603 | 3572 |
| 32 | 3 | 3685 | 4297 | 3588 |
| 33 | 3 | 4238 | 8951 | 4174 |
| 34 | 3 | 4386 | 4895 | 4438 |
| 35 | 3 | 5125 | 4791 | 5333 |
| 36 | 3 | 6094 | 4746 | 5932 |
| 37 | 3 | 6257 | 4539 | 6267 |
| 38 | 3 | 6252 | 4602 | 6205 |
| 39 | 3 | 5836 | 4492 | 5816 |
| 40 | 3 | 6005 | 4318 | 6040 |
| 41 | 3 | 5312 | 4396 | 5463 |
| 42 | 3 | 5250 | 4483 | 5105 |
| 43 | 3 | 3985 | 12828 | 3902 |
| 44 | 3 | 4884 | 7250 | 4991 |
| 45 | 3 | 5993 | 15103 | 6064 |
| 46 | 3 | 8427 | 8258 | 8351 |
| 47 | 3 | 9542 | 8083 | 9376 |
| 48 | 3 | 9751 | 8008 | 9882 |
| 49 | 3 | 8793 | 7355 | 8751 |
| 50 | 3 | 10143 | 7764 | 10216 |

**TABLE 11: EVALUATION OF THE MULTIVARIATE MODEL**

DOI: http://dx.doi.org/10.24018/ejbmnr.2020.5.5.481
In 2017, total financial transactions, ‘ed important, equivalent to 19.2% of GDP in 2017. The next highest natural growth rates were recorded in Ireland, that the variation within the population from 2003–2017, it holds:

For example, we compared France and Greece and observed that $\hat{f}_2 = 1.296, \exp(1.296) = 3.61$. This means that the variation within the population in France is more than three times compared to that of Greece.

In the case of Italy and the Netherlands, we find that $\tau_{10} = 0.384, \exp(0.384) = 2.421$. That is, the population change in Italy outnumbered that of Holland by 2.42 proportionally.

The difference between Denmark and Ireland was slightly higher, where $\exp(-0.013152111) - 0.289098319 = 0.30225043, \exp(0.30225043) = 1.35$. In other words, the population change in Denmark increased slightly with 1.35 than Ireland.

As for the difference between Austria and Portugal, we observe that $\tau_{19} = 0.553021826 - 1.0926491 = -0.539627274, \exp(-0.539627274) = 0.58$. That is, the number of inhabitants (variation within the population change) is 0.583 lower than that of Portugal.

In the case of Mediterranean countries like Greece and Spain, we see that $\hat{f}_{21} = 1.289668683, \exp(1.289668683) = 8.37$. We could say that Greece has a lower population change by 0.27 than Spain.

By comparing Belgium and Greece, we observed that, $\tau_{12} = 1.107649788-1.289668683, \exp(1.124589) = -0.83$. It means that Belgium has a lower population change by 0.83 than Greece.

Finally, in the Scandinavian countries such as Finland and Sweden, the difference between them is, $\hat{f}_{23} = -1.2187 - 0.5622 = -1.7809$ with $\exp(-1.7809) = 0.168$, thus, the variation within the population in Finland is smaller than its neighboring country Sweden.

V. RESEARCH FINDINGS: COMPARISON OF THE POPULATION CHANGE IN EU25

By comparing some of the countries in EU25 to ascertain the variation within the population from 2003–2017, it holds:

From Table 11 above, it is evident that the prices (value) of the model RC (M = 4) fully fitted to the data.

IV. LOGARITHMS OF RC (M = 4) MODEL

The RC (M=4) from the RCDIM (ROW TAU-S) are as follows:

| Country     | $\ln(-0.330335) = -1.107647988$ | $\ln(0.99118) = -0.00904$ | $\ln(-0.986963) = -0.01312272759$ | $\ln(1.685072) = 0.52180$ | $\ln(-3.434861) = -1.2380$ | $\ln(-1.335225) = -0.289120381$ | $\ln(-0.275362) = -1.289668683$ | $\ln(0.03418891096) = 0.34188891096$ | $\ln(1.415027) = 0.341846122$ | $\ln(1.388701) = 0.3283687778$ | $\ln(-1.82367) = -0.601$ | $\ln(-3.130306) = -1.1410$ | $\ln(-1.700391) = -0.5309$ | $\ln(0.08966) = -2.4117$ | $\ln(-1.82028) = -0.5909$ | $\ln(1.950344) = 0.6680$ | $\ln(-0.575237) = -0.5529731492$ | $\ln(1.121523) = 0.115$ | $\ln(-0.335356) = -1.092562625$ | $\ln(-0.1247) = 2.0818$ | $\ln(-2.2101) = -0.7930$ | $\ln(-2.41695) = -0.8825$ | $\ln(0.2956) = -1.2187$ | $\ln(-0.56996) = 0.5622$ | $\ln(3.06409) = 1.1197$ |

\[
\hat{f}_0 \text{ Expected frequencies of the independence model.}
\]
\[
\hat{f}_0 \text{ Expected frequencies of the multivariate model RC (M = 4).}
\]

VI. SUMMARY

The highest proportion of government expenditure in the EU goes to social protection and health. In 2017, total government expenditure in the European Union (EU) amounted to 46.7% of gross domestic product (GDP). This share has steadily decreased since 2012, when it stood at 49.7% of GDP.

Among the main functions of general government expenditure in the EU, ‘social protection’ is by far the most important, equivalent to 19.2% of GDP in 2017. The next most important areas are ‘health’ (7.0%), ‘general public services’ (6.0%) such as external affairs and public debt transactions, ‘education’ (4.6%) and ‘economic affairs’ (4.4%). The functions ‘public order and safety’ (1.7%), ‘defence’ (1.2%), ‘recreation, culture and religion’ (1.1%), ‘environmental protection’ (0.8%) and ‘housing and community amenities’ (0.6%) have more limited weights.

Highest natural growth rates were recorded in Ireland, Cyprus, France, Luxembourg and the United Kingdom. In 2010, 5.4 million children were born in the EU27. The crude birth rate was 10.7 per 1000 inhabitants, the same as in 2009. The highest birth rates were recorded in Ireland.
of the United Kingdom (13.0%), France (12.8%), Cyprus (12.4%) and Sweden (12.3%), and the lowest rates in Germany (8.3%), Latvia (8.6%), Hungary (9.0%), Italy (9.3%), Austria (9.4%), Portugal (9.5%) and Malta (9.6%).

There were 4.8 million deaths registered in the EU27 in 2010. The crude death rate was 9.7 per 1000 inhabitants, unchanged compared with 2009. The highest death rates were observed in Bulgaria (14.6%), Latvia (13.4%), Hungary (13.0%), Lithuania (12.8%) and Romania (12.1%), and the lowest rates in Ireland (6.2%), Cyprus (6.7%), Malta (7.2%) and Luxembourg (7.4%). Consequently, the highest natural growth of the population (the difference between live births and deaths per 1000 inhabitants) was registered in Ireland (+10.3%), well ahead of Cyprus (+5.7%), France (+4.4%), Luxembourg (+4.2%) and the United Kingdom (+3.9%). Eight Member States had a negative natural growth, with the largest declines in Latvia (-4.8%), Bulgaria (-4.6%), Hungary (-4.0%), Germany and Romania (both -2.2%). Highest population growth was recorded in 2017 in Luxembourg, Sweden, Malta, Belgium and the United Kingdom.

In 2010, over 60% of the increase in the EU27 population came from migration. In relative terms, Luxembourg (+15.1%), Malta (+5.4%), Sweden (+5.3%), Italy (+5.2%) and Belgium (+5.1%) had the largest net inflows, while Lithuania (-23.7%) and Ireland (-7.5%) recorded the highest net outflows.

In conclusion, the population increased in twenty Member States and decreased in seven, with considerable variations between Member States. The largest relative increases were observed in Luxembourg (+19.3%), Sweden (+8.0%), Malta (+7.8%), Belgium (+7.2%) and the United Kingdom (+6.6%), and the largest decreases in Lithuania (-25.7%), Latvia (-8.4%) and Bulgaria (-7.8%).

In the context of the annual demographic balance, Eurostat produces net migration figures by taking the difference between total population change and natural change; this concept is referred to as net migration plus statistical adjustment.

Most of the global increase is attributable to a small number of countries. From 2017 to 2050, it is expected that half of the world’s population growth will be concentrated in just nine countries: India, Nigeria, the Democratic Republic of the Congo, Pakistan, Ethiopia, the United Republic of Tanzania, the United States of America, Uganda and Indonesia (ordered by their expected contribution to total growth).

The group of 47 least developed countries (LDCs) continues to have a relatively high level of fertility, which stood at 4.3 births per woman in 2010-2015. As a result, the population of these countries has been growing rapidly, at around 2.4% per year. Although this rate of increase is expected to slow significantly over the coming decades, the combined population of the LDCs, roughly one billion in 2017, is projected to increase by 33% between 2017 and 2030, and to reach 1.9 billion persons in 2050.

Similarly, Africa continues to experience high rates of population growth. Between 2017 and 2050, the populations of 26 African countries are projected to expand to at least double their current size.

The concentration of global population growth in the poorest countries presents a considerable challenge to governments in implementing the 2030 Agenda for Sustainable Development, which seeks to end poverty and hunger, expand and update health and education systems, achieve gender equality and women’s empowerment, reduce inequality and ensure that no one is left behind.

VII. Conclusion

All the six association models show bad fit. Therefore, we proceeded to the multivariate model in which the row column-effects RC multivariate model of RC ((M = 4) is better than all because it covers 99.98% of the data, thus, giving the best fit among all. However, to be more precise, the population of a country is affected or influenced by several factors; these could be as a result of:

- The standard of living of each country; every country can prosper if its citizens are productive and do not display behaviors that hinder development such as corruption and poor work ethic [16].
- The degree of utilization of the workforce, which is directly related to efficiency in the production of goods and services [17].
- A country’s mortality rate and
- Several other factors which are difficult to be determined in each country.

We should mention that Θ, which is 0.08906, i.e. approximately equal to 1, show that we have independence between the variables. To determine the degree of correlation (association) between these countries and years, we use the θ of the second model, (the uniform association model which is typified by (U) for calculation. More specifically, the innate correlation index in log-linear form is equal to:

$$\log(F_{ij}) = \lambda + \lambda_{A(i)} + \lambda_{B(j)} + \phi_{i}F_{i}$$

where ϕ is a single parameter for interaction.

The parameter of interaction

$$\phi = \log \theta = \log (0.08906)$$

$$\Phi = \ln (0.08906) = -2.41844498$$

$$|\varphi_{1/2}| = \sqrt{2.41844498} = 1.555135036$$

Since there is a correlation based on our study on population change in EU25: absolute numbers and crude rates, the general observation is that in all the 25 countries of the European Union, over the years from 2003–2017, there was a slight increase in each year of their population.

REFERENCES

[1] United Nations Population Division, World Population Prospects: The 2009 Revision Highlights 2017.
[2] Eurostat (2010): “European Data Agency, “Eurostat-Online-Data- Agency, (2020_20), OECD statistics at regional level”.
[3] Nwaubani, J., Zelka, M., & Tsianta, A. (2020). A Model for Comparing Government Expenditure on Civil Servants-Gross Wages

DOI: http://dx.doi.org/10.24018/ejbmr.2020.5.5.481

Vol 5 | Issue 5 | September 2020
and Salaries in the EU24. *European Journal of Business and Management Research, Vol 5 No 1* (2020).

[4] Office for Official Publications of the European Union, 2017.

[5] UN – Eurostat Population “European Commission, Department of Economic and Social Affairs.” 27 January 2017.

[6] Eliason P. Scott & Clifford Clogg, (1990), *Categorical Data Analysis (CDAS)* NBER working P., 510.

[7] Diewert, W. Erwin, (1995). Axiomatic and Economic Approaches to Elementary Price indexes. NBER working paper 5104.

[8] Goodman, L.A., (1979a). Multiple Models for the Analysis of Occupational Mobility Tables &Other Kinds of Cross-Classification Tables. “American Journal of Sociology”, 84:804-819.

[9] Goodman, L.A., (1979b). Multiple Models for the Analysis of Occupational Mobility Tables and Other Kinds of Cross-Classification Tables.”American Journal of Sociology”.

[10] Goodman, L.A., (1981a). Association models and the Bivariate Normal for Contingency Tables with Ordered Categories. *Biometrika*, 68:347-55.

[11] Clogg, C.C. (1990), *Analysis of Association (ANOAS) Program*.

[12] Schwarz, Gideon E. (1978), “Estimating the dimension of a model”, Annals of Statistics, 6 (2): 461–464, doi:10.1214/aos/1176344136, MR 0468014.

[13] Haritou A., Nwaubani J C. (2008). “Categorical Data Analysis” - *Working paper (University Press)*.

[14] Haritou A., Nwaubani J C. (2010). “Categorical Data Analysis” - *Working paper (University Press)*.

[15] Goodman, L.A., (1981b). Association Models and Canonical Correlation in the Analysis of Cross-Classifications Having Ordered Categories. “Journal of American Statistical Association”, 76.3,20-34.

[16] Fairbanks, Michael. (2005). Changing the Mind of a Nation: Elements in a Process for Creating Prosperity, “in Culture Matters”, Huntington, editors, “New York: Basic Books”, pp.270-281.

[17] Olson, M. (1982), The Rise and Decline of Nations: Economic Growth, Economic Rigidities and Stagflation, New Haven: Yale University Press.