UNIVERSITY OF NAIROBI

COLLEGE OF BIOLOGICAL AND PHYSICAL SCIENCES

SCHOOL OF MATHEMATICS

ERASTUS KIMANI NDEKELE

I46/68601/2011

MORTALITY PROJECTION: CURVE FITTING AND LINEAR REGRESSION

As a partial requirement for the fulfillment of a postgraduate diploma in actuarial science

DECLARATION

This project as presented in this document is my original work and has not been replicated; extracted or copied from any other published or unpublished sources.

Erastus Kimani Ndekele I46/68601/2011 Signature ……………… Date …………….

This project was supervised and approved as original work by:

Prof. Patrick G.O. Weke Signature ………………. Date ……………..

ACKNOWLEDGEMENTS

I would first and foremost like to thank God Almighty for giving me the strength, will and grace to see this project to completion within the shortest possible time. I would also like to thank my parents and more so my father Mr. Erastus G.M. Ndekele for encouraging me all through the duration of this work and through many more endeavours. He has been patient and understanding and has assisted greatly in provision of resources instrumental to the completion of this work. I would also like to acknowledge the rest of my family, loved ones and friends who continued to provide suggestions, advice and encouragement to me as I pursued and continue to pursue this and many more such works.

Finally, I would like to give special thanks to my supervisor Prof. Patrick G.O. Weke, the patron of Actuarial Science at the University of Nairobi, for providing valuable reading material and insights as to the best way of approaching this work.

TABLE OF CONTENTS

SYMBOLS AND ABBREVIATIONS

Abbreviations:

ACF	Area Compatibility Factors
AIDS	Acquired Immune Deficiency Syndrome
CBD	Cairns-Blake-Dowd
CMI	Continuous Mortality Investigation
HMD	Human Mortality Database
L-C	Lee-Carter
SMR	Standard Mortality Ratio
UK	United Kingdom

Symbols:

d_x	Number of deaths between ages x and x+1
E_x	Exposed to risk at age x
μ_x	Force of mortality for (x)
$ln\mu_x$	Natural log of the force of mortality for (x)
q_x	Graduated q type mortality rate/probability of (x) dying within a year
$1 - q_x$	Probability that (x) survives for a year using graduated mortality rates. (Also p_x)
q^*x	Crude death rate of (x)
$1 - q^*x$	Crude survival rate of (x) for a year. (Can also be written as p^*x)
lnq^*x	Natural log of the crude death rate of (x)
$\ln(1 - q^*x)$	Natural log of the crude survival rate of (x)
(x)	A life aged x years exact
x	An independent variable used in linear regression.
R^2	Coefficient of determination.

ABSTRACT

This study was aimed at predicting future mortality rates given two main scenarios. The first is in the case where there is a body of mortality tables from the past and the second is in the case where such a database does not exist.

In the first case, linear regression was used to estimate mortality of specific age groups for a specific future year. In this case the specified future year was from 2000 to 2009 although similar computations can be carried out for age specific and single year life tables. To perform these linear regression two transformations were considered to the existing q type mortality rates: the log linear transformation and the logit transformation

For the second case, it was assumed that the only data that existed was that of actual deaths and exposed to risk. Thus, for adequate mortality projections to be done a Makeham curve was fitted in one case and a cubic spline graduation was done for the second case.

Both the Makeham and the cubic spline methods were observed to have a good fit to the data, the latter providing a better fit than the former. However the linear regression methods were observed to give an almost constant difference at every age and would probably not be the best methods of forecasting especially for very long time periods. The advantage of the linear regression method was that it was seen to keep the original shape of the graduation.

CHAPTER 1: INTRODUCTION

This project looks into the area of mortality studies and in specific life tables. It is well known that life tables are an important aspect of actuarial calculations especially in the life insurance and pension sectors. The main aim is to look at some methods of mortality projection and more so those that make use of particular mathematical laws or models.

1.1 BACKGROUND OF THE PROBLEM

A life table is used in actuarial calculations mainly to value life assurance contracts and pension agreements. The process of construction of life tables is however a tedious exercise and it usually takes quite a bit of time to come up with a complete life table. It is also a fact that, due to this time lag between the time that a life table is constructed and the time it is available for use, the conditions that existed during the time of constructing the life table may probably have changed. This change can be either drastic such as the discovery of a cure for a known ailment (for instance AIDS) or a slight change such as improved working conditions/increase in deplorable working conditions. In any case, the life table may be rendered obsolete before it is ready for use and thus the danger that there may never be a relevant life table for use.

It is due to this problem that the need for mortality projection has grown over the years. More specifically the need to model human mortality according to given laws[1] or mathematical formulae stretches back to 1825 when Gompertz came up with the first law for the force of mortality:

[1] A law of mortality is based on other biological factors other than just a best fitting curve. (Life Assurance mathematics by W.F. Scott)

$$\mu_X = BC^X$$

Where B and C are constants and x is the age of the life.

It was observed that Gompertz law was only accurate over middle ages and this led to another famous law of mortality known as Makeham's law, 1860, which involved adding a constant term to Gompertz law.

$$\mu_X = A + BC^X$$

The addition of the constant term by Makeham was an indication of an autonomous addition to the risk of mortality attributed to accidental deaths.

It should probably be noted at this point that the force of mortality is a random variable and can be modeled to fit any distribution. A good example of such is the Weibull's law where the force of mortality has been modeled to fit the Weibull distribution. Mixtures of various distributions may also be used to obtain a force of mortality that best describes the rate of decrement.

Several more complex formulas have evolved from the initial Gompertz and Makeham formulae to incorporate more factors such as mortality at all ages and other physiological factors.

Formulas such as:

$$\mu_X = \frac{\left(A+BC^X\right)}{1+DC^X} \quad \text{; Perks' formula 1932}$$

Where A, B, C and D are all parameters that should be estimated using various parametric methods to determine the force of mortality at the specified age x. This is an example of an evolution from the original Makeham equation where the terms BC^X and DC^X represent the decrease in intensity of a human life to resist death over time as advocated by Gompertz. The term A represents the superimposed increase in mortality due to accidents, which is especially high in young adults of around ages 16 years to 25 years.

L. Heligman and J.H. Pollard’s formula is another such formula. The authors claim that it is applicable to the whole human life span, taking into account the accident hump at young adult ages, the general decline in mortality due to age and, the mortality of a child adapting to its new environment. This was done in their representation of the Australian national mortality. (Benjamin and Pollard 1980) and the formula is given as:

$$\frac{q_x}{p_x} = A^{(x+B)^C} + Dexp\{-E(logx - logF)^2 + GH^x$$

J. H. Pollard and L. Heligman 1960-62

However, as much as there has been a lot of work on graduation by mathematical formulas such as those stated above (among others), they do not account for the general improvement of mortality over the recent years due to medical and technological advances. This has led to the adoption of even more sophisticated formulae taking into account not only individual/group ages, but also time or duration since joining the scheme in the case of a pension fund or since acceptance of a proposal in the case of life insurance contracts. Some of the renowned formulas include:

a) The Lee-Carter model (1992)

$$log\mu_{xt} = \alpha_x + \beta_x \kappa_t$$

This model also stems from the Gompertz model of 1825 but includes the time parameter represented by the Kappa variable κ_t. The similarity can be seen when the logarithm is removed. When this is done we obtain the equation:

$$\mu_{xt} = \alpha_x \beta_x^{\kappa_t}$$

It can be clearly seen that α_x and β_x correspond to the parameters A and B respectively in the Gompertz model and that they are dependent only on age, whereas the parameter κ_t represents time taken and replaces x in the original Gompertz model. The x (age) is incorporated in β_x.

b) The Cairns-Blake-Dowd model (CBD 2006)

$$q_{xt} = \frac{\exp\left(\kappa_t^{(0)} + \kappa_t^{(1)} x\right)}{1 + \exp\left(\kappa_t^{(0)} + \kappa_t^{(1)}\right)}$$

(Oliver Lockwood March 2009)

These and other such formulas, make an allowance for the general improvement in human mortality in the recent past together with the initial problem of forecasting mortality and mortality trends.

1.2 STATEMENT OF THE PROBLEM

As has been seen earlier, there is the need to be able to forecast mortality to avoid possible obsoleteness of life tables. The most ideal situation is to come up with flexible forecasting techniques that will be able to incorporate any future changes in mortality and the best chance of this is the use of mathematical graduation formulae. There is also a need to incorporate into the same forecasting methods the general improvement of mortality over time known as longevity as this has serious financial implications on actuarial calculations especially those of defined benefit pension funds.

The problem can therefore be summarized as: “How can mortality be modeled to include past, current and possible future changes in mortality as accurately as possible.”

1.3 OBJECTIVES OF THE STUDY

This study attempts to give a solution to some of the problems stated above. It mainly addresses the question of mortality projection given two case scenarios:

i. In the presence of a large database of past data and more specifically the q type mortality rates.
ii. When only the actual deaths and the exposed to risk are known for specified ages in a given year.

In the first case, linear regression analysis will be carried out to obtain the specific q type mortality rates at the desired point in the future (between the years 2000-2009). Two different linear transformation methods will be used in this case: the log transformation and the logit transformation.

For the second case, a curve fitting exercise will be done. First a Makeham curve will be fitted to the crude mortality rate (q^*x) to obtain a graduated curve with which mortality can be projected for future years. Second, a cubic spline graduation will be done to give another graduated curve with which mortality can be projected at any future period of time.

With regard to longevity, no mathematical formula will be fitted with this explicit purpose. However, the fitted curves can be used to compare the crude mortality rate and the graduated mortality at specific ages to account for either mortality risk[2] or longevity risk which are important to life assurance companies and defined benefit pension funds respectively.

1.3 SIGNIFICANCE OF THE STUDY

As already previously mentioned, the process of creating life tables is a tedious one. It starts by data collection, mainly from national registries, of information such as the total number of births and deaths in a given year, the total population in a given country by age, gender, region and such other factors. This is usually obtained through national census documents or for instance The Kenya Demographic and Health Survey document in Kenya. Other similar documents may be available as a result of the central governments' need to plan for their populations.

As such, if the process of creating a life table begins at the start of a given year it may take several months to complete and subsequently make the new tables available for use in that year. Further more, the conditions at the start of the year when the life table was being created may not persist throughout the year and as such the life table may be meaningless. Also new inventions and innovations may drastically change the mortality of populations in a given country or area making the life table obsolete again.

It is for these and many other reasons that mortality projection is significant to the planning authorities in a country, life assurance companies and pension fund companies.

In the case of planning authorities, they may use mortality projections as one of the tools they can use to check for estimates in population size of their country. This will aid governments to better plan for

[2] Mortality risk is the risk associated with a life dying earlier than possible. This is critical to life assurance companies for computation of reserves and premiums

provision of social amenities for their populations and other services such as in determination of adequate police to citizen ratios among others. It should be pointed out that other population projection methods are usually used and mortality projection is not frequently used for this purpose.

Life assurance companies benefit greatly from the use of mortality projections. First and foremost, a life table is instrumental in the computation of premiums and reserves. Mortality projections can be used to give the life office an idea of the pattern of mortality to expect in the coming years that they can then incorporate into their premium and reserving bases. An important aspect for life assurance companies is mortality risk. If a life assurance company over estimates the mortality of a group of lives in a certain policy (say a whole life policy or an endowment assurance policy), it runs the risk of not having accumulated enough reserves to payout claims as they arise and may end up becoming insolvent. Insurance companies offering life insurance products will therefore be very keen to check on mortality trends and to incorporate mortality projection results into their premium and reserving bases.

Pension fund companies are also very keen on mortality projections. If a pension scheme and more specifically a defined benefit pension scheme under estimates mortality, it runs the risk of paying too much unplanned pension benefits (especially those in the form of annuities) and may eventually become unable to meet these obligation and become insolvent. This is the risk associated with general improvement in mortality rates and is known as longevity risk. Thus, managers of defined benefit pension schemes and other pension schemes will be interested in modeling mortality to hedge against longevity risk.

Therefore the likely interested parties in mortality projections are: planning authorities, life assurance companies and pension fund managers. Individuals should also pay attention to mortality studies so as to increase their life expectancy by eliminating any factors that may increase their mortality.

CHAPTER 2: LITERATURE REVIEW

Graduation has been observed as a method/principal of adjusting a set of observed rates to provide a suitable basis for actuarial and demographic calculations of a practical nature (The Analysis of Mortality and Other Actuarial Statistics by B. Benjamin an J. H. Pollard). Thus it may be seen as a smoothing technique where observed rates are smoothed so that they can flow naturally from one data point to another in a Cartesian plane. The purpose for this kind of smoothing is to enable proper planning and projections about the future by introducing some kind of predictability in the observed values for decision-making purposes. Indeed Pollard and Benjamin observe that graduation is important for sound judgment in statistics for dealing with practical problems.

There are various ways to graduate data. One may use a graphical approach whereby hand polishing is used until sufficient smoothness is attained or graduation by reference to a standard table. These methods are largely subjective as it depends on the person carrying out the graduation, the degree of hand polishing required and probably to the choice of standard table to adopt. The focus of this project is graduations carried out by means of adherence to given laws of mortality or by adherence to certain mathematical formulae. The method is fairly objective in the sense that once a particular mathematical formula is fitted to a similar data set with similar assumptions, the results should be the same.

Over the years, mathematical formulae have developed to try and explain mortality over given time periods. The first of these was the Gompertz law for the force of mortality (1825) who postulated that the force of mortality μ_x satisfies the simple differential equation:

$$\frac{d\mu_x}{dx} = kx \qquad ; x > \alpha$$

Where x is the age of the life, α is the youngest age in the life table and k is a constant of proportionality.

This was solved to give:

$\mu_X = BC^X$ Where B and C are the constants of proportionality.

Makeham Later added a constant term to Gompertz formula in 1860, to represent the accident hump for young adult ages. He came up with the formula:

$$\mu_X = A + BC^X$$

Where A represents the autonomous accident hump for young adult ages which usually is from the ages of 16 to 21.

J. H. Pollard and B. Benjamin used a sample set of a mortality table from the ages 35 to 46 to fit a Makeham curve (Benjamin and Pollard 1980) and noticed that there was evidence of lack of fit to the data. This may have developed due to changes in mortality trends or simply because that curve did not adequately explain mortality for that data. Other formulae have developed to improve on the initial Gompertz and Makeham equations and Pollard and Heligman later developed their own law of mortality that they claim is applicable to the whole life span:

$$\frac{q_x}{p_x} = A^{(x+B)^C} + Dexp\{-E(logx - logF)^2 + GH^x$$

They claim that although the number of parameters seems excessive, they represent three distinct features: the mortality of a child adapting to its new environment, the mortality associated with aging of the body and the superimposed accident mortality (Benjamin and Pollard 1980)

The Pollard and Heligman model for mortality has not been used in this project although it would be interesting to further this project towards other mathematical formulas such as these and test their applicability in developing countries as well. This law for mortality together with others such as the Perks' formula mentioned above represents developments and curves allied to the Makeham curve. Another modern method of graduation, that has been preferred in this project, is the spline graduation technique and in particular the cubic spline graduation.

A spline has been described as a sufficiently smooth polynomial, which is piece wise defined and possesses a high degree of smoothness at the places where the polynomial pieces connect known as knots. (Wikipedia.org)

According to Pollard and Benjamin the name 'spline function' is given to a function obtained by joining together a sequence of polynomial arcs. The polynomials are chosen in such a way that derivatives up to and including the order n-1 of the polynomial (of order n) used is continuous everywhere. The cubic spline graduation method produces very smooth results that generally adhere to the data better than other graduation formula. This will be shown to be true in this project.

It may also be useful to mention a little bit on longevity risk as this may provide an appreciation of the use of mortality and the differences that mortality projections may develop with the introduction of time as a factor although this has not been done in this project.

Longevity risk has been defined as any potential risk attached to the increasing life expectancy of pensioners and policy holders, which can eventually translate in higher than expected pay-out-ratios for many pension funds and insurance companies (Wikipedia.org). A literal example of longevity risk is explained in the prologue of "Financial aspects of longevity risk" by Stephen Richards and Garvin Jones which, is quoted here below:

> *'In 1965, André-François Raffray, a lawyer in the southern French city of Arles, made a deal with a ninety-year-old local woman. In a contract relatively common in France, he agreed to pay her an income for the rest of her life in exchange for inheriting her house upon her death.*
>
> *Unfortunately for M. Raffray, the woman was Mme. Jeanne Calment, who went on to be the longest-lived person in the world at 122 years. She outlived the luckless M. Raffray, who paid more than the value of the house before pre-deceasing her.*
>
> *"In life, one sometimes makes bad deals", said Mme. Calment of M. Raffray.'*

Given the seriousness of longevity risk it is important to point out that mortality projections should evolve to take into account this type of risk and by extension mortality risk.

The L-C model is probably the most known model of longevity. It is stated as:

$$log\mu_{xt} = \alpha_x + \beta_x \kappa_t$$

Where the α_x, β_x and K_t are parameters to be estimated. This model is considered by the CMI in CMI (2005) report, and, further in the CMI (2007) report where it is fitted to the CMI's male assured lives data set and to England and Wales population data for both males and females. A major advantage of this method over the use of spline graduation is that the future projections can be easily computed by simply projecting the time series of the Kappa parameter (Oliver Lockwood March 2009). The other parameters can also be computed but only relate to age.

Another well-known longevity model is the CBD model mentioned earlier as:

$$q_{xt} = \frac{\exp\left(\kappa_t^{(0)} + \kappa_t^{(1)} x\right)}{1 + \exp\left(\kappa_t^{(0)} + \kappa_t^{(1)}\right)}$$

It can be noted that the CBD model is fitted directly to the q type mortality rate instead of the force of mortality. The model is also more flexible regarding the way mortality can evolve over time as evidenced by the introduction of the extra kappa parameter. The authors of this model fitted it to ages above 60 years (Oliver Lockwood March 2009). The CBD model has also been extended to account for the cohort effect but this will also not be mentioned here.

Also, linear regression has been carried out in this project as done by B. Benjamin and A. S. Soliman in their book ‘Mortality on the Move’. They used simple linear regression at particular points or for specific ages using the log transform and logit transform. The also used Benjamin and Pollards formula for all ages but as mentioned before this was substituted by the cubic spline graduation formula.

CHAPTER 3: METHODOLOGY

3.1 ASSUMPTIONS AND LIMITATIONS

3.1.1 ASSUMPTIONS.

The main assumptions are:

i. The data collected in the HMD and subsequently used in this project is correct and free of any bias.
ii. Grouping the data into five-year age groups has eliminated any misrepresentations of age.
iii. Any error terms as used in the models are assumed to be independent
iv. The number of deaths and the exposed to risk are assumed to be randomly sampled from the general population of the U.K.
v. The number of deaths is influenced by factors other than the exposed to risk.

All assumptions that underlie the methods used to compute the death rates also apply especially for the linear regression aspects. (http://www.mortality.org/Public/Docs/MethodsProtocol.pdf).

3.1.2 LIMITATIONS.

i. The first limitation of this project is that it does not account for the aspect of longevity. Thus once a curve is fitted, it assumed to hold in any time period. This is not true as the Makeham law of mortality is largely not applicable in this day and age due to medical advances and improved living standards.
ii. The project does not also take into account qualitative methods in trying to address the issue of mortality. All the exposed to risk are assumed to have 'normal mortality' and factors that can increase (or decrease) the force of mortality are not incorporated. A good example is conditions such as stress, depression, risky working conditions and so on.

iii. The choice of graduation technique is subjective and there are other methods of graduation that may be used. These are mainly: graduation with reference to a standard table and graphical graduation. These techniques could offer alternative results. Other mathematical formulas can also be used other than those applied in this project such as the Perks' and Pollard and Heligman formulas.

iv. The methods only apply to those countries that have complete and accurate records of the deaths and exposed to risk. Perhaps A.C.F's and S.M.R's can be used to extend these results to countries without proper records.

3.2 DATA COLLECTION

All the data was obtained in the human mortality database collection in the website: http://www.mortality.org. The reason for this is because all the data available can be assumed to be accurate and complete. However, most of the data available is for developed nations as they invest heavily in proper registration of births, deaths and exposed to risk in specific regions and nation as a whole. Methods of how the individual datasets are constructed are explained in the same website under the methods protocol. Another advantage of using this data is that it has been checked for errors. However, no attempt has been made to correct age misstatements or over/under enumeration of people or events. Any subsequent calculations in this project can be assumed to have the same errors. Also, since all the data is from developed nations, data for the United Kingdom (UK) was arbitrarily chosen, the results cannot be said to apply to developing countries as mortality trends are usually significantly different. The data were the abridged life tables from 1922 to 2009 grouped into 10-year intervals and five-year age groups.

3.3 DESCRIPTION OF METHODS

3.3.1 LINEAR REGRESSION: LOG LINEAR TRANSFORMATION

The steps are as follows:

i. Graduated UK rates q_x are transformed to logs i.e. y=ln qx

ii. x is taken as each of the 10 year intervals for specific age groups

iii. We obtain mean of x, mean of y and mean of xy

iv. We also obtain the values n*mean x*mean y and n*mean x

v. We calculate the parameters a and b using linear regression as

$$b = \frac{\Sigma xy - n\bar{x}\bar{y}}{\Sigma x^2 - n\bar{x}}$$

$$a = \bar{y} - n\bar{x}$$

Where: $\bar{x} = mean\ of\ x$

$\bar{y} = mean\ of\ y$

$n = number\ of\ observations$

$\Sigma xy = sumation\ over\ all\ values\ of\ x * y$

$\Sigma x^2 = summation\ over\ all\ values\ of\ x^2$

vi. The equation of a straight line ($y = a + bx$) is then used to estimate the new value of y.[3]

vii. The exponent of y is then taken to find q_x

[3] The same equation of a straight line will be used for the logit transformation and the same procedure for calculating a and b will be used.

It is important to note that this procedure is repeated for all age groups and hence it is equivalent to finding specific data points on the Cartesian plane for a specific future time period which in our case is the years 2000-2009.

The difference between the computed qx for 2000-2009 is then compared with those given in the HMD

3.3.2 LINEAR REGRESSION: LOGIT TRANSFORMATION

i. The given graduated UK rates are transformed: $y = 0.5ln(\frac{q_x}{1-q_{x)}})$

ii. The same steps 2-6 above as done for the log linear transformation are repeated and the parameters a, b are computed as before.

iii. The graduated rates qx are then computed as

$$q_x = e^{2y}/(1 + e^{2y})$$

Where: $q_x = graduated\ UK\ mortality\ rate$

$e^{2y} = exponent\ of\ 2 * y$

iv. The new q_x is then compared with that given by the HMD.

3.3.3 CURVE FITTING: MAKEHAM EQUATION

The Makeham law for the force of mortality as stated before is given as:

$$\mu_X = A + BC^X$$

where the second term of the equation is a Gompertz term representing mortality of human ages in the middle ages and A represents the autonomous risk of mortality especially in the young ages.

This model was only fitted for the year 2009 and the procedure used for fitting the Makeham equation involved estimating the parameters A, B and C. The steps followed were:

i. The crude death rate q*x is computed by dividing the actual deaths from the exposed to risk i.e.

$$q^*x = \frac{dx}{Ex}$$

Where: $q^*x = crude\ death\ rate$

$dx \ = Actual\ deaths\ during\ the\ year\ 2009$

$Ex \ = Exposed\ to\ risk$

ii. The survival probability and the force of interest were then obtained from:

$$(1 - q_x) = \exp - (\int_0^1 \mu_{x+t}\, dt)$$

$$(1 - q_x) = e^{-\mu_x}$$

$\mu_x = -\ln(1 - q_x)$

Where: $(1 - q_x) = The\ probability\ (x) survives\ for\ one\ year$

$$\mu_x = The\ force\ of\ mortality\ at\ age\ x$$

iii. Thus the log linear transformation of the force of mortality would be given by:

$ln\ \mu_x = \ln(-\ln(1 - q_x))$

Keeping the consistency of the notations used in this project and hence those defined we have:

$\mu_x = -\ln(1 - q^* x)$

And

$ln\ \mu_x = \ln(-\ln(1 - q^* x))$

Where the symbols remain as previously defined.

iv. A trend line is then fitted using excel between the ages 25 and 100. This is because on investigation of the log of the force of mortality between those ages it was identified that a straight line can be used to estimate the data points. This is consistent with the Gompertz model, which is accurate for middle ages.

v. The parameters B and C were estimated as follows:

$$y = a + bx$$

$\ln(-\ln(1 - q^*x)) = lnB + xlnC$ (from: $\mu_x = BC^x$)

Therefore by equating the terms we see that:

$$y = \ln(-\ln(1 - q^*x)$$

$$B = e^a$$

$$C = e^b$$

vi. The autonomous term A was obtained by taking the representative younger age 10 (any age may be taken), and applying the formula;[4]

$$A = (exp - (height\ of\ graduating\ curve)) - (\exp - (height\ of\ straight\ line))$$

vii. The survival probability $(1 - q^*x)$ is then obtained as:

$$(1 - q^*x) = \exp\left(-A - \left(\frac{B(C-1)}{\ln c}\right)C^x\right)$$

viii. The estimated q type mortality rate (q^*x) is then obtained as $1 - (1 - q^*x)$

ix. This estimated mortality rate is then compared with the actual mortality rate for the year 2009

[4] This formula is as applied in The Analysis of Mortality and Other Actuarial Statistics by B. Benjamin and J. H. Pollard

3.3.4 CURVE FITTING: CUBIC SPLINES

The steps taken in the fitting of the cubic spline equations are:

i. The observation of an almost linear relationship is assumed as before from the observation of $ln\mu_x$ and the ages 25 – 100 chosen for graduation using this method.

ii. Seven knots each of 10 years were arbitrarily chosen. They were at the ages of 25, 35, 45, 55, 65, 75 and 85 years. Knots can be chosen in any manner and they do not have to be of equal intervals.

iii. The log of the crude death rate (i.e. $y = lnq^{*}x$) was obtained and the natural cubic splines passing through the 7 knots computed using the formula:

$$y = a_0 + a_1 x + \sum_{j=1}^{7} b_j \phi_j$$

Where: $y = lnq^{*}x$

$$a_0, a_1 \text{ } and \text{ } b_j \text{ } are \text{ } constants$$

$$\phi_j = \begin{cases} 0 & ; x < x_j \\ (x - x_j)^3 & ; x \geq x_j \end{cases}$$

iv. $\Phi_{j's}$ are computed and the new formula that caters for the y intercept and slope is used:

$$y = a_0 + a_1 x + \sum_{j=1}^{n-2} b_j \, \Phi_j$$

Where: $\Phi_j = \phi_j - \left\{\frac{x_n - x_j}{x_n - x_{n-1}}\right\} \phi_{n-1}(x) + \left\{\frac{x_{n-1} - x_j}{x_n - x_{n-1}}\right\} \Phi_j$

The rest of the factors remain unchanged.

v. A multiple regression analysis is then carried out to determine the seven coefficients: $a_0, a_1, b_1, b_2, b_3, b_4$ and b_5

vi. The graduated mortality rate using cubic splines is then obtained as $q^*x = e^y = e^{lnq^*x}$

vii. The estimated mortality rate is then compared with that given by the HMD.

In the curve fitting methods of graduation it is important to remember that the mortality will be assumed to follow the new graduated rates as opposed to the crude rates. Thus, mortality projection in any given year can be estimated using the underlying mathematical formula.

CHAPTER 4: DATA ANALYSIS AND FINDINGS

4.1 LINEAR REGRESSION: LOG LINEAR TRANSFORMATION

The table that was used for the log linear transformation is in the Figures and Tables section as Table 1.

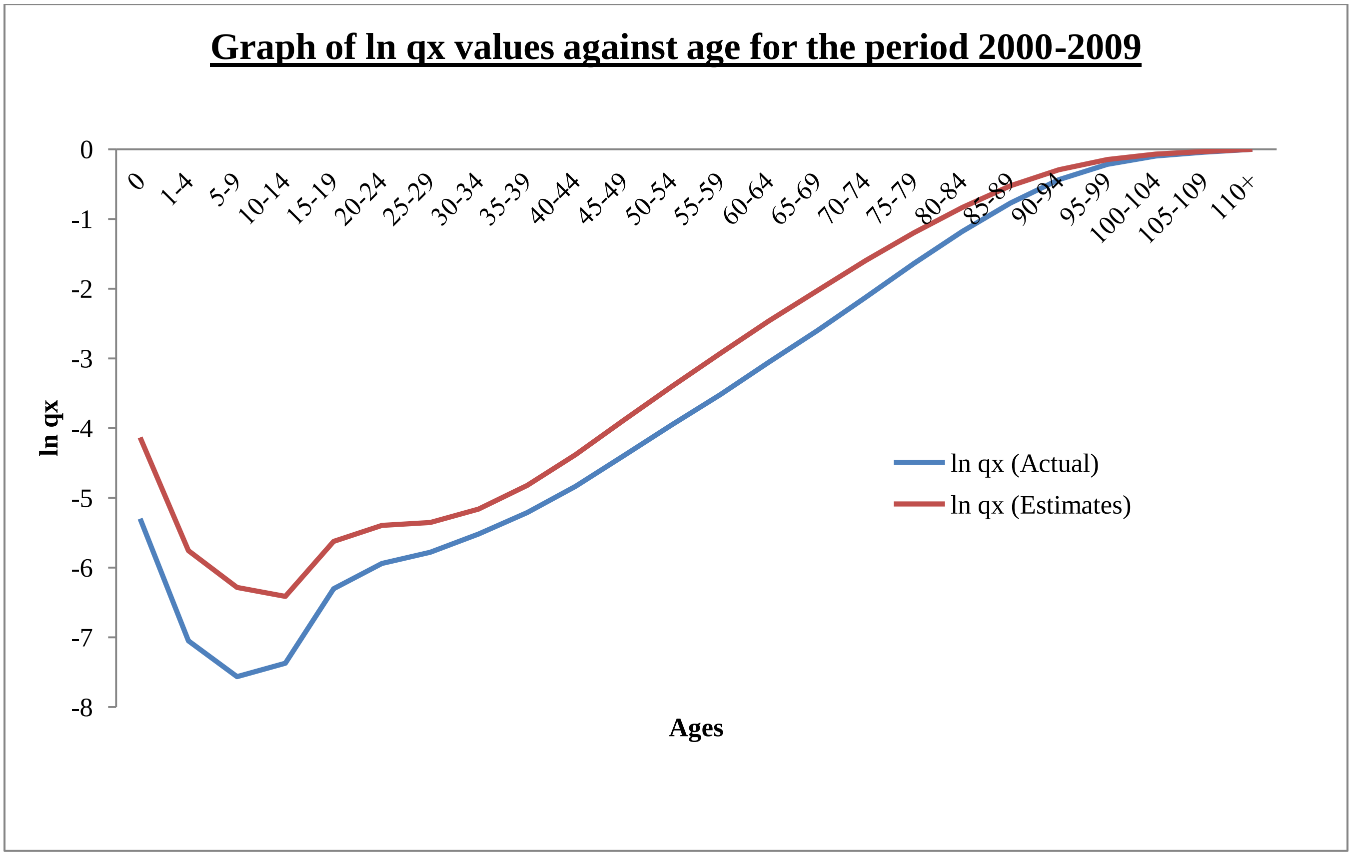
The results for the various age groups are shown below:

TABLE A: ESTIMATED AND ACTUAL MORTALITY:

Age	ln q_x (Actual)	ln q_x (Estimates)	q_x (Actual)	q_x (Estimates)
0	-5.298317367	-4.13408099	0.005	0.016017379
1-4	-7.047017346	-5.754537639	0.00087	0.003168371
5-9	-7.561681746	-6.282483969	0.00052	0.001868753
10-14	-7.369790739	-6.411860274	0.00063	0.001641967
15-19	-6.303439312	-5.624088113	0.00183	0.003609853
20-24	-5.940771433	-5.395798291	0.00263	0.004535598
25-29	-5.779584188	-5.35205971	0.00309	0.004738381
30-34	-5.516473376	-5.161595927	0.00402	0.005732544
35-39	-5.21213967	-4.822348177	0.00545	0.008047867
40-44	-4.832070786	-4.378254593	0.00797	0.01254724
45-49	-4.394099216	-3.885356867	0.01235	0.020540497
50-54	-3.950244218	-3.398221473	0.01925	0.033432678
55-59	-3.51694499	-2.92743418	0.02969	0.053534221
60-64	-3.056331896	-2.464727838	0.04706	0.085031982
65-69	-2.600991129	-2.026178421	0.0742	0.13183839
70-74	-2.124689985	-1.597977757	0.11947	0.202305215
75-79	-1.639021215	-1.197659839	0.19417	0.30189988
80-84	-1.17950786	-0.832813705	0.30743	0.434824097
85-89	-0.772861609	-0.5208336	0.46169	0.594025162
90-94	-0.43568723	-0.29311005	0.64682	0.745940046
95-99	-0.215175382	-0.145188364	0.8064	0.86485937
100-104	-0.094112897	-0.067483813	0.91018	0.934742851
105-109	-0.039884934	-0.03165395	0.9609	0.968841792
110+	0	0	1	1

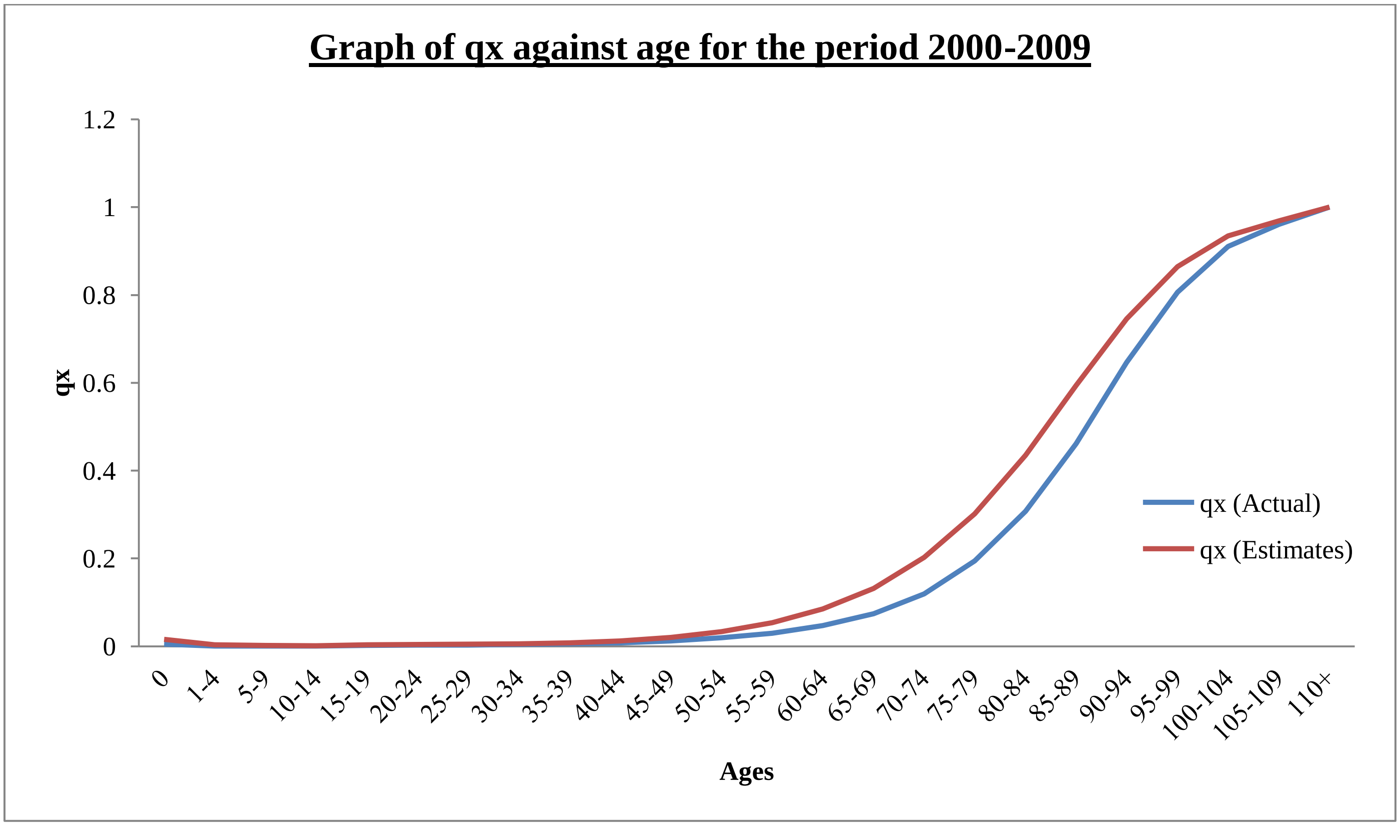
When the values are graphed, we get the following graphs:

Graph 1.1



It can be seen that the estimated log of the mortality rates differs at every age to the actual log of mortality rates. This is also the case with the actual mortality rates as shown in the graph below. This indicates that mortality does not follow a linear pattern with time. The mortality is affected by other factors apart from time alone and hence the discrepancy at every age. The major advantage of this method is that it keeps the initial shape of the curve as the assumptions and models that were used to come up with the graduated rates have not been changed and the only varying factor is time.

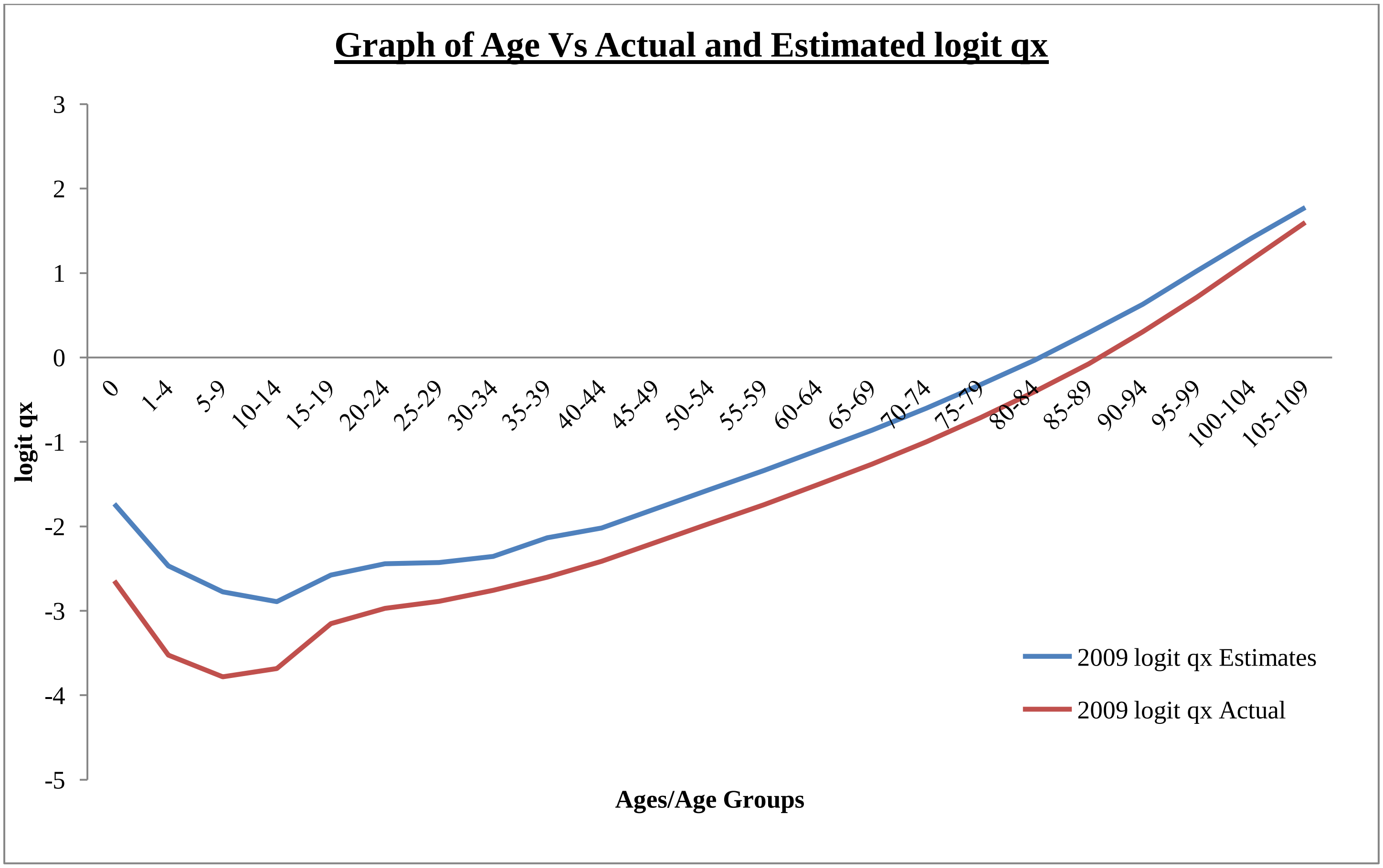
Graph 1.2



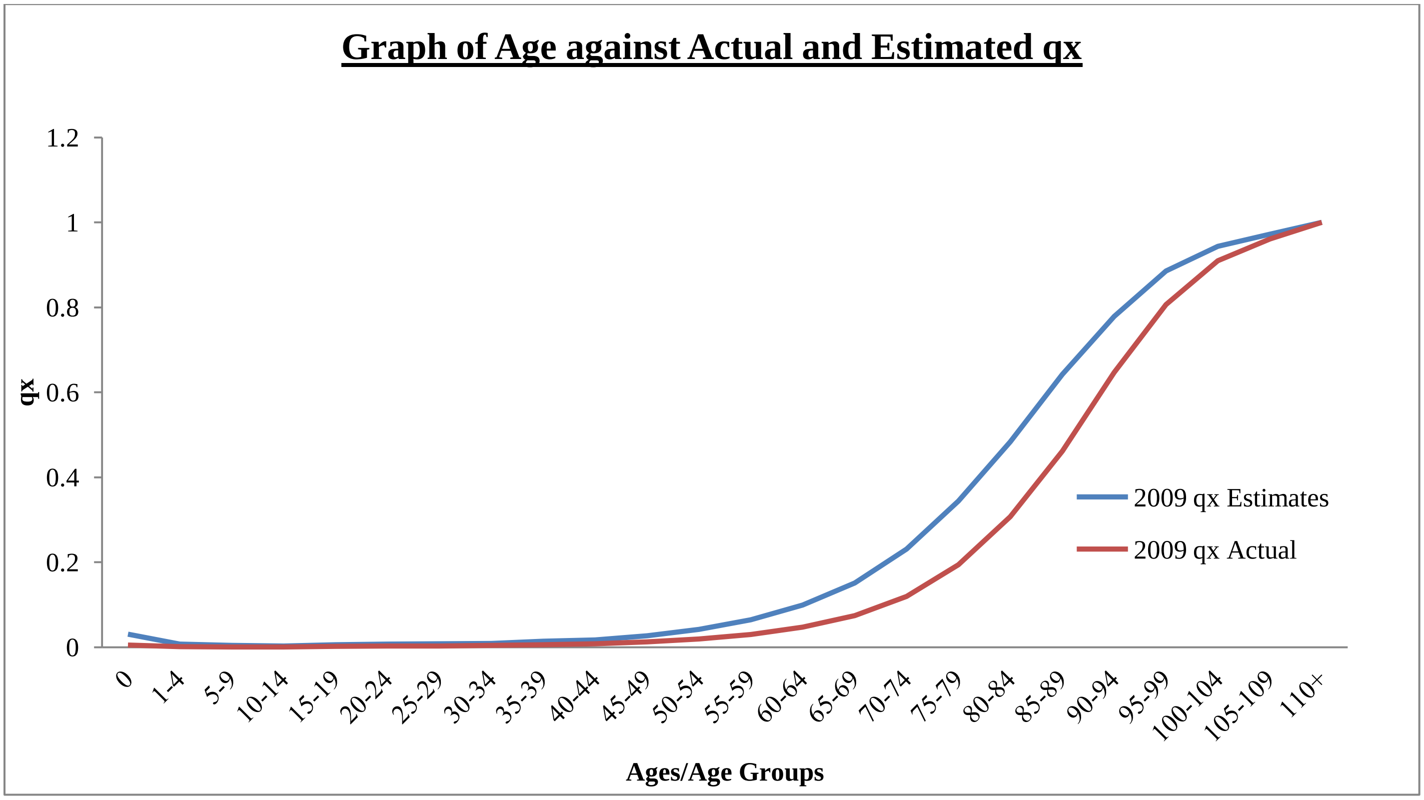
4.2 LINEAR REGRESSION: LOGIT TRANSFORMATION

The graphs of the actual and estimated logs of the mortality rates, and, the actual and estimated mortality rates differ at every age almost similarly to those obtained using the log linear transformation. This can again be attributed to the fact that the only thing being varied is time and that the underlying assumptions and formulae used to calculate the graduated mortality rates remain unchanged. It should also be noted that the mortality rates seem to be almost similar towards the older ages say from around 100 years and above. This may be attributed to the fact that there are fewer people at this age who are affected by other physiological factors and that the similarity may due to the few exposed to risk who survive to the elderly ages. For instance, at the age of 110 years and above, there is only one exposed to risk and the linear regression formula for both the log linear transformation and the logit method is unlikely to change this. These observations can be shown in the two graphs below:

Graph 2.1



Graph 2.2



4.3 CURVE FITTING: MAKEHAM EQUATION

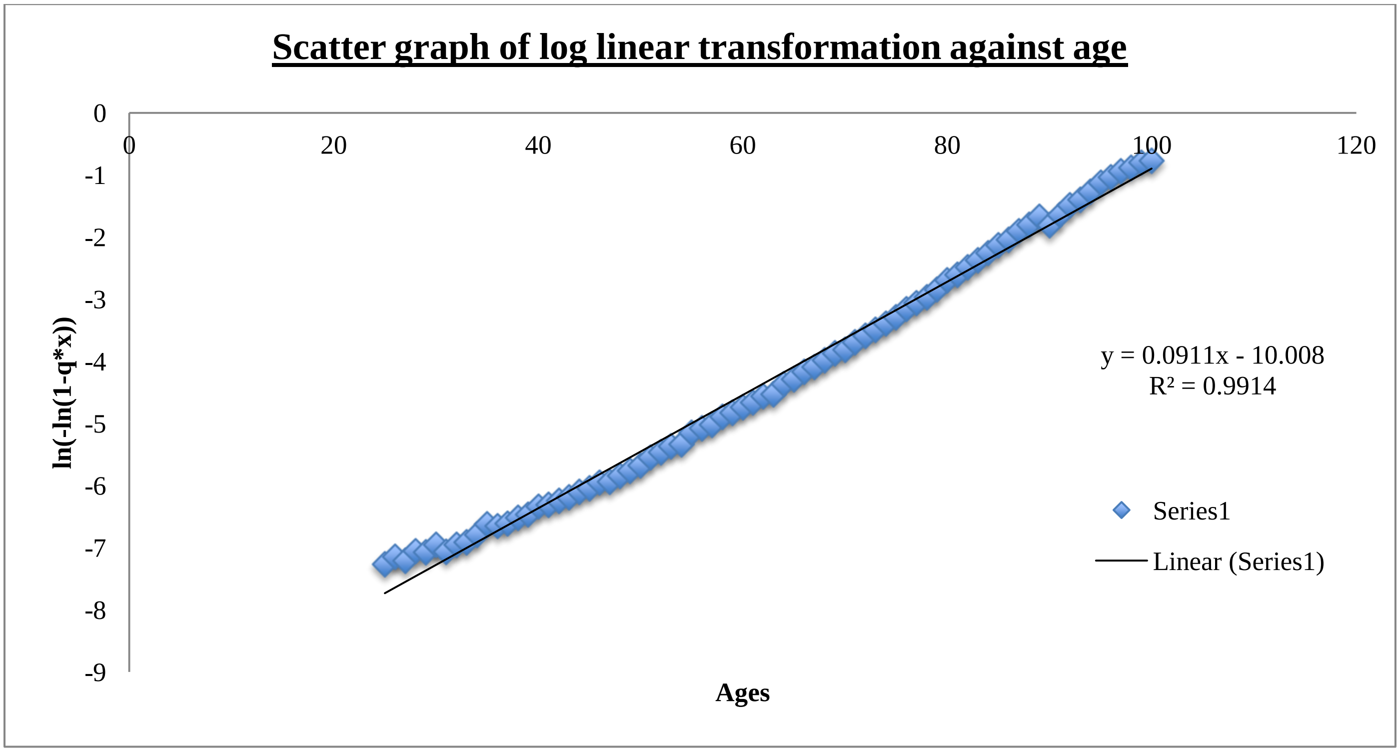
As mentioned earlier the Makeham equation is fitted to the data by first fitting the Gompertz term to the middle ages - which were taken to be from the ages of 25 years to 100 years - using linear regression and then the autonomous term is added. The trend line fitted using excel is:

$y = -10.008 + 0.0911x$

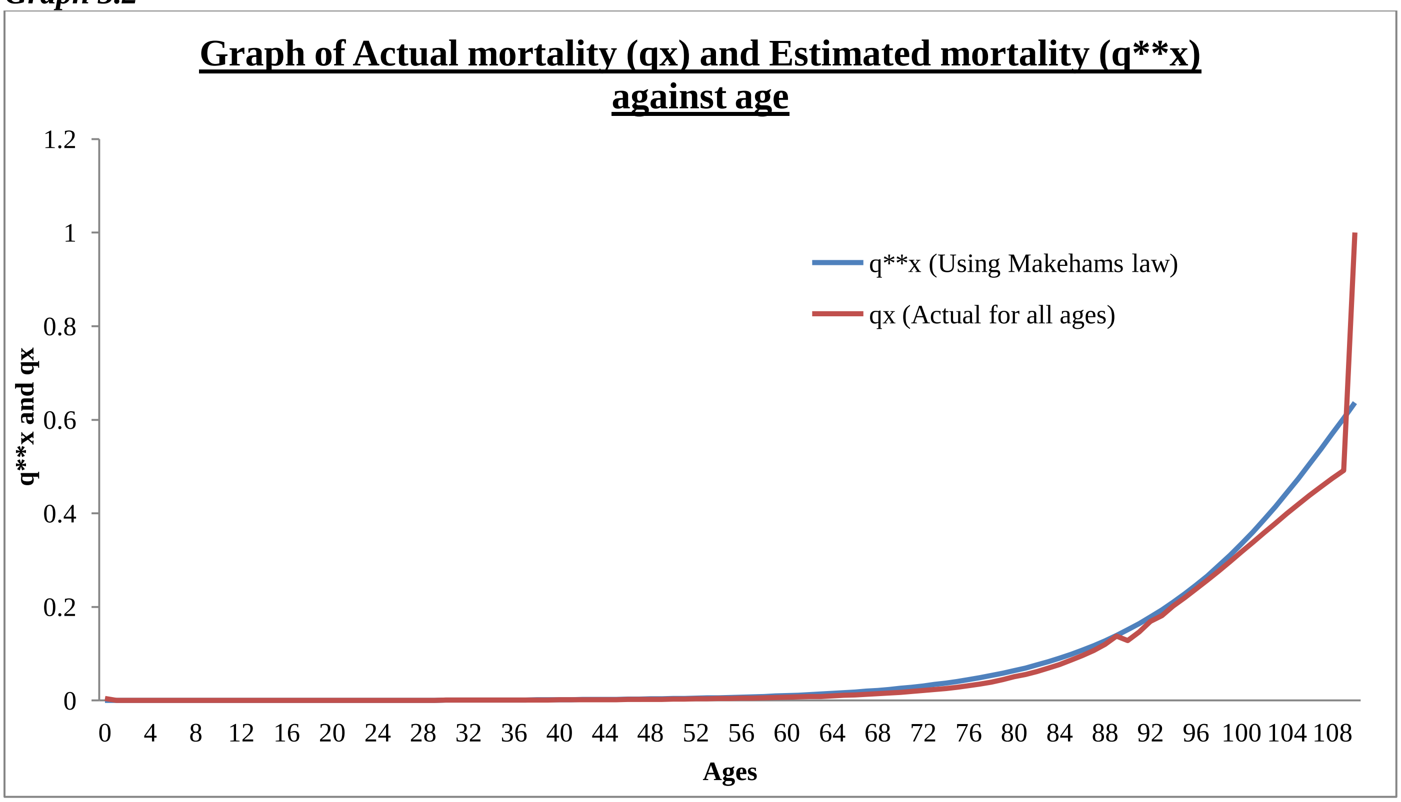
A coefficient of determination of 0.99143 is given by the trend equation. This shows that age highly affects mortality.

The graph that shows the fitted trend line is graph 3.1 below where series 1 is ln (-ln (1-q*x)) and linear series 1 is the fitted trend line for this log linear transformation: (The table that shows the Gompertz component of the Makeham equation is Table 3 under figures and tables)

Graph 3.1



Graph 3.2



Graph 3.2 shows the fitted Makeham curve, which shows some slight variations in the older ages.

The p value given by the Pearson chi square test done in r is 1 (Appendix 2). This means that there is a 100% chance that the null hypothesis is not false and so we do not reject it. However, due to the very small values involved there is the likelihood that the test is incorrect and this is indicated as a warning message in r. Other tests are then used to check the goodness of fit of the model.

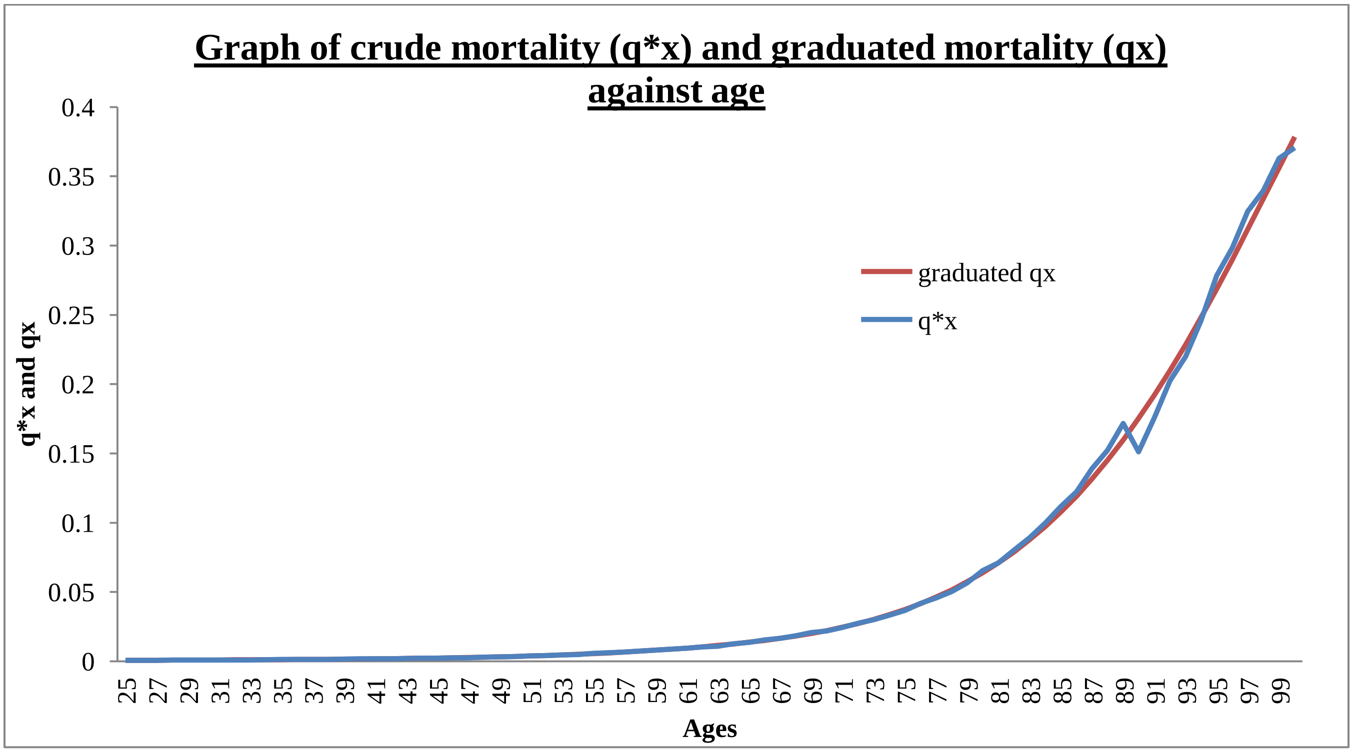
The Kolmogorov Smirnov test conducted in r shows that the Makeham model is a good fit (Appendix 2). The one sided Kolmogorov Smirnov test shows that the D value is 0.0541 and the p value is 0.723. Thus there is a 72.3% chance that the highest difference between this fitted model and the actual death rates is 0.0541. This indicates a good fit. The two sided Kolmogorov Smirnov test shows that there is an 85.89% chance that the highest difference between the graduated mortality rates and the actual mortality rates is 0.0811. Thus the graduated rates and the actual rates can be considered to come from the same distribution.

The Anderson Darling test shows p values of 0.42916 when there is no adjustment for ties and 0.41647 when there is an adjustment for ties. This means that at a critical values of 5% and 10%, the null hypothesis is not rejected and the curve is thus a good fit of the data.

The Shapiro Wilk test checks for normality. The p value given by r implies that the population from which the data comes from is not normally distributed.

4.4 CURVE FITTING: CUBIC SPLINES

Graph 4.1



Graph 4.1 above shows a cubic spline curve fitted to the crude mortality rates. From a simple first observation, the cubic spline method looks like it gives a better fit than the Makeham equation. The Multiple linear regression model was done using the program R and the results were:

$$ln\ q^{*}x = -9.498 + 0.08552x - 0.0001323\Phi_1 + 0.0001867\Phi_2 - 0.00001205\Phi_3 - 0.000103\Phi_4 + 0.0001706\Phi_5$$

The R code used to carry out the multiple regression analysis can be found in Appendix 1.

The Multiple R^2 value and adjusted R^2 values are given as 0.9994 in r which means that 99.94% of the variation in ln q*x can be explained by the independent variables chosen. The p value also shows that there is only a $2.2x10^{-14}$% chance that the observations were explained by chance. Thus the linear model is a very good fit.

Similarly, as in the Makeham model, the p value given by the chi square test done in r is 1. This implies a very good fit although other tests should then be used to check the goodness of fit of the model.

The Kolmogorov Smirnov test for goodness of fit (one sided test) shows that D=0.0395 and p=0.8883. This shows that the highest difference between the cubic splines graduated death rates and the actual death rates, is 0.0395 with a probability of 88.83%. This shows that the model is a good fit. The two-sided Kolmogorov Smirnov test shows that D=0.0395 and p=1. This implies that the graduated death rates are obtained from the same population or distribution as the actual death rates. (Appendix 2)

As in the Makeham model, the Shapiro Wilk test is very small and it implies that the population from which the data is obtained is not normally distributed.

CHAPTER 5: SUMMARY AND CONCLUSION

5.1 SUMMARY

The project has found that various methods can be used to model and project mortality. The linear transformation methods give a similar shape as that of the actual mortality when compared together but with an almost constant variation at all ages. This can be corrected by subtracting from the linear variation the constant value or some other constant (such as the mean of the variations). This may give a more accurate forecast of mortality.

The curve fitting methods used give a close approximation of mortality in the year 2009. In particular however, the graduation by cubic splines gives a more accurate prediction from the ages of about 60 to 100 years. This shows that the cubic spline graduation method is more accurate than the Gompertz Makeham model in this case although it is only applied for ages 25 to 100 years.

5.2 CONCLUSION

The log and logit transformations are good when there is a database of past mortality figures that are assumed to be accurate and checked for data entry errors. However, in the absence of such a database, then curve fitting can be used to try and model expected mortality from at least a recent periods' observed mortality. The formulae and laws used to fit the model may range from simple ones such as the Gompertz Makeham Laws to more complex ones allowing for the passage of time as a factor in addition to age and other physiological factors.

The major advantage of the linear regression models is that it keeps the original assumptions intact and the only assumption is that changes in observed mortality would be linear with time. However, since mortality does not follow a linear trend then these methods may be more successful if they are used for short time periods. Curve fitting has the advantage of trying use the most suitable model to mimic the observed mortality and hence project future mortality from the most appropriate model. The problem with this however is that such models may become out dated by the time they are ready for use or worse still, a major scientific break through may lead to discoveries that could improve mortality. Policy measures may also be put in place to improve the living conditions of people and hence improve living conditions and mortality rates.

There can be more done to improve on the results obtained here. For starters, area compatibility factors can be included to compare the results of this study to mortality in other areas, which may not have mortality records but may have similar demographic conditions. More sophisticated models can be used to try and reflect mortality patterns and other physiological factors can be included in such models other than age and time. In Kenya specifically, a database can be maintained by a reputable institution such as The University of Nairobi to facilitate further research into mortality trends and hopefully yield useful results for planning purposes, pension purposes and life offices.

The methods used to project mortality for given populations need to be reviewed regularly to ensure that they still represent mortality experience fairly accurately.

5.3 IMPLICATIONS OF THE STUDY

It can be seen that if it is assumed that there will be no major changes in mortality or other physiological factors for several years, then mathematical formulae may be used to estimate future mortality for a given number of years. This is useful in the sense that it keeps parameters constant and hence useful in planning

The study also implies that there are numerous formulae that can be developed to explain mortality and the most suitable of these formulae depends on factors such as age, working conditions, living conditions and other physiological factors. The choice of formulae to use may thus differ across countries, regions and even continents.

There is also room to accommodate the aspect of time in addition to that of age. Thus, separate life tables may be developed to take into account the aspect of populations living longer due to changing times and advancements in science. The addition of time as a factor could bring about major financial implications to actuaries and users of life tables.

The use of comparable factors is also important and could be adopted as another approach of developing life tables. This could especially be used in areas where the recording of births and deaths is incomplete or inaccurate. Factors such as geographical conditions could be used to develop life tables in such areas. Such comparisons could bring out further relationships between regions and interest groups.

REFERENCES

1. Arnold T. B. and Emerson J. W., 2011. Non-Parametric Goodness-of-Fit Tests For Discrete Null Distributions. The R Journal. Vol. 3/2. Available at http://journal.r-project.org/archive/2011-2/RJournal_2011-2_Arnold+Emerson.pdf. [Accessed on 28th May 2013].
2. Benjamin B. and Pollard J. H., 1980. The Analysis of Mortality and Other Actuarial Statistics. 2nd ed. London. William Heinemann Ltd.
3. Benjamin B. and Soliman A. S., . Mortality on The Move. Oxford. Actuarial Education Service.
4. NIST/SEMATECH. Engineering Statistics [e-handbook of statistical methods]. Available at http://www.itl.nist.gov/div898/handbook/eda/section3/eda35e.htm. [Accessed 27th June 2013].
5. Longhai Li, 2008? How To Install an R Package. Saskatchewan. s.n.
6. Oliver R., 2009. Time Series Modeling of Gompertz Makeham Mortality Curves: Historical Analysis, Forecasting and Life Insurance Applications [PDF]. Available at http://www.actuaries.org.uk/.../timeseriesmodellingofgompertz-makeham.pdf [Accessed on 1st April 2013].
7. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical. Computing, Vienna, Austria. URL http://www.R-project.org/
8. Ricci V. 2005. Fitting Distributions With R [PDF]. Free Software Foundation. Available at http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf [Accessed on 8th May 2013].
9. Scott W.F., 2000. Mortality Studies. Aberdeen. s.n
10. Stephen J. R. and Gavin L. J. 2004. Financial Aspects of Longevity Risk. Available at http://www.sias.org.uk/data/papers/LongevityRisk/DownloadPDF [Accessed on 22nd April 2013].

APPENDIX

1) R multiple linear regression Code.

```
>  setwd("~/Documents")
> Mydata<-read.table("~/Documents/MY DOCUMENTS/Graduation    Data.csv",header=T,sep=",")
> str(Mydata) #Checks how the data is structured or looks like
'data.frame': 76 obs. of  19 variables:
 $ Age                          : int  25 26 27 28 29 30 31 32 33 34 ...
 $ dx                           : int  305 334 307 362 351 380 327 362 380 439 ...
 $ Ex                           : num  435640 423859 416489 417669 414066 ...
 $ q.x                          : num  0.0007 0.000788 0.000737 0.000867 0.000848 ...
 $ ln.q.x                       : num  -7.26 -7.15 -7.21 -7.05 -7.07 ...
 $ X_1                          : int  0 1 8 27 64 125 216 343 512 729 ...
 $ X_2                          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ X_3                          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ X_4                          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ X_5                          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ X_6                          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ X_7                          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ X_1.1                        : num  0 5.9 40.6 108.3 207.2 ...
 $ X_2.1                        : num  0 0 0 0 0 0 0 0 0 0 ...
 $ X_3.1                        : num  0 0 0 0 0 0 0 0 0 0 ...
 $ X_4.1                        : num  0 0 0 0 0 0 0 0 0 0 ...
 $ X_5.1                        : num  0 0 0 0 0 0 0 0 0 0 ...
 $ Standard.2008..mortality.rates.qs: num  0.00052 0.0006 0.00059 0.00062 0.00067 0.00069 0.00074 0.00076
0.00088 0.00097 ...
 $ Weights..Ex.qs               : num  8.38e+08 7.06e+08 7.06e+08 6.74e+08 6.18e+08 ...
>  Mydata.lm<-lm(ln.q.x~ Age+X_1.1+X_2.1+X_3.1+X_4.1+X_5.1, data=Mydata) # fits a multiple linear
regression model of the given data
```

2) R code and output: Goodness of fit tests for Makeham Model

```
> x<-0:110
> A<--0.0000309191
> B<-0.0000450382
> C<-1.095378538
> y<-A+B*C^x
> Y<-1-exp(-y)
> X<-
c(0.00458,0.00028,0.00019,0.00014,0.00011,0.0001,0.00009,0.00007,0.00009,0.0001,0.00008,0.0001,0.00011,0.00013,0.00014,0.00021,0.00025,0.00036,0.00041,0.00045,0.00045,0.00047,0.0004,0.00046,0.00046,0.00052,0.00057,0.00055,0.00064,0.00065,0.00072,0.00067,0.00073,0.00076,0.00089,0.00102,0.001,0.00106,0.00116,0.00123,0.00143,0.00147,0.00156,0.00166,0.00186,0.002,0.00214,0.00219,0.00242,0.00262,0.00294,0.0031,0.00355,0.00391,0.00414,0.00473,0.0051,0.00541,0.00602,0.00655,0.00705,0.00764,0.00853,0.00876,0.0101,0.01096,0.01225,0.013,0.01441,0.01601,0.01756,0.01921,0.02124,0.0234,0.02594,0.02845,0.03195,0.03535,0.03941,0.04472,0.05068,0.05537,0.06201,0.0696,0.07718,0.08611,0.09569,0.10704,0.11959,0.13792,0.12827,0.14641,0.16915,0.18127,0.20168,0.21948,0.23787,0.25697,0.27668,0.29686,0.31737,0.33805,0.35875,0.37931,0.39955,0.41934,0.43853,0.45702,0.47468,0.49146,1)
> Makeham<-read.table("~/Documents/MY DOCUMENTS/Makeham Graduation for r.csv",header=TRUE,sep=",")
> chisq.test(Makeham) #performs the chi-square test

        Pearson's Chi-squared test

data:  Makeham
X-squared = 28.8616, df = 220, p-value = 1

Warning message:
In chisq.test(Makeham) : Chi-squared approximation may be incorrect
> ks.test(X,Y,A,B,C,alternative="two.side",exact=NULL) # We conduct a two sided K-S test

        Two-sample Kolmogorov-Smirnov test

data:  X and Y
D = 0.0811, p-value = 0.8589
alternative hypothesis: two-sided
```

```
ks.test(X,Y,A,B,C,alternative="g",exact=NULL) # We conduct a one sided K-S test

        Two-sample Kolmogorov-Smirnov test

data:  X and Y
D^+ = 0.0541, p-value = 0.723
alternative hypothesis: the CDF of x lies above that of y

Warning message:
In ks.test(X, Y, A, B, C, alternative = "g", exact = NULL) :
  p-value will be approximate in the presence of ties

> X<-
c(0.00007,0.00008,0.00009,0.00009,0.0001,0.0001,0.00011,0.00011,0.00011,0.00013,0.00014,0.00014,0.00019,0
.00021,0.00025,0.00028,0.00036,0.0004,0.00041,0.00045,0.00045,0.00046,0.00046,0.00047,0.00052,0.00055,0.0
0057,0.00064,0.00065,0.00067,0.00072,0.00073,0.00076,0.00089,0.001,0.00102,0.00106,0.00116,0.00123,0.001
43,0.00147,0.00156,0.00166,0.00186,0.002,0.00214,0.00219,0.00242,0.00262,0.00294,0.0031,0.00355,0.00391,0
.00414,0.00458,0.00473,0.0051,0.00541,0.00602,0.00655,0.00705,0.00764,0.00853,0.00876,0.0101,0.01096,0.01
225,0.013,0.01441,0.01601,0.01756,0.01921,0.02124,0.0234,0.02594,0.02845,0.03195,0.03535,0.03941,0.04472,
0.05068,0.05537,0.06201,0.0696,0.07718,0.08611,0.09569,0.10704,0.11959,0.12827,0.13792,0.14641,0.16915,0.
18127,0.20168,0.21948,0.23787,0.25697,0.27668,0.29686,0.31737,0.33805,0.35875,0.37931,0.39955,0.41934,0.
43853,0.45702,0.47468,0.49146,1) # These are the sorted values(from the highest to the lowest) of qx
> Y<-1-exp(-y)
> install.packages("adk",lib="~/Documents/")# installs packages for Anderson Darling tests
> library("adk",lib.loc="~/Documents/")
> adk.test(X,Y)#performs the Anderson Darling test
Anderson-Darling k-sample test.

Number of samples:  2
Sample sizes: 111 111
Total number of values: 222
Number of unique values: 215

Mean of Anderson-Darling Criterion: 1
```

Standard deviation of Anderson-Darling Criterion: 0.75489

T.AD = (Anderson-Darling Criterion - mean)/sigma

Null Hypothesis: All samples come from a common population.

```
                    t.obs P-value extrapolation
not adj. for ties -0.33989 0.42916             1
adj. for ties     -0.29737 0.41647             1
> shapiro.test(Y)#performs the Shapiro Wilk Test

        Shapiro-Wilk normality test

data:  Y
W = 0.6112, p-value = 1.061e-15
```

3) <u>R code and output: Goodness of fit tests for Cubic Spline.</u>

```
# This is a continuation of the code in Appendix 1
>  Spline<-read.table("~/Documents/MY DOCUMENTS/Spline graduation for r.csv",header=T,sep=",")# reads the table generated by spline graduation
> chisq.test(Spline)

	Pearson's Chi-squared test

data:  Spline
X-squared = 12.2857, df = 150, p-value = 1

Warning message:
In chisq.test(Spline) : Chi-squared approximation may be incorrect
> summary(Mydata.lm)

Call:
lm(formula = ln.q.x ~ Age + X_1.1 + X_2.1 + X_3.1 + X_4.1 + X_5.1,
    data = Mydata)

Residuals:
      Min        1Q    Median        3Q       Max
-0.152091 -0.019625  0.003336  0.025846  0.129065

Coefficients:
              Estimate Std. Error  t value Pr(>|t|)
(Intercept) -9.498e+00  2.787e-02 -340.804  < 2e-16 ***
Age          8.552e-02  3.895e-04  219.590  < 2e-16 ***
X_1.1       -1.323e-04  1.225e-05  -10.805  < 2e-16 ***
X_2.1        1.868e-04  3.990e-05    4.681 1.38e-05 ***
X_3.1       -1.205e-05  6.469e-05   -0.186  0.85278
X_4.1       -1.034e-04  8.088e-05   -1.278  0.20537
X_5.1        1.706e-04  6.363e-05    2.680  0.00919 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04894 on 69 degrees of freedom
```

Multiple R-squared: 0.9994,	Adjusted R-squared: 0.9994

F-statistic: 2.034e+04 on 6 and 69 DF, p-value: < 2.2e-16

The two sided Kolmogorov Smirnov test is performed as below

> Y<-

c(0.000636198,0.00069246,0.00075083,0.000810574,0.000871467,0.000933297,0.000995872,0.00105902,0.001
122603,0.001186513,0.001250682,0.001316284,0.001388408,0.001468966,0.001558803,0.001658873,0.001770
251,0.001894148,0.002031925,0.002185111,0.002355424,0.00254467,0.002754484,0.00298704,0.003244898,0.
003530913,0.003848273,0.004200535,0.004591663,0.005026074,0.00550869,0.006043178,0.006627648,0.0072
65134,0.007960959,0.008721118,0.009552374,0.010462376,0.011459787,0.012554442,0.013757525,0.0150866
73,0.016577366,0.018254446,0.020140347,0.022259963,0.024640822,0.027313242,0.030310468,0.033668785,0
.037427575,0.041629331,0.046319591,0.051546786,0.057361977,0.063818461,0.070971221,0.078876207,0.087
589409,0.097165717,0.107657542,0.119113186,0.131574963,0.145077062,0.159643183,0.175283963,0.191994
247,0.209750268,0.228506816,0.248194508,0.268717282,0.289950258,0.311738127,0.333894229,0.356200498,
0.378408426) # The graduated mortality rates using cubic splines

X<-

c(0.000700119,0.000787998,0.000737113,0.000866715,0.000847691,0.000953483,0.000855496,0.000960993,0.
000994443,0.001123714,0.001338015,0.00129736,0.001351145,0.001477006,0.00155608,0.001771173,0.00182
854,0.001946113,0.002055257,0.002244635,0.002377787,0.002611377,0.002630462,0.002905696,0.003149202,
0.003430682,0.003893916,0.004239612,0.004663989,0.004831139,0.005792477,0.006199807,0.00659252,0.007
454661,0.007970783,0.008715752,0.009400416,0.01044081,0.010777701,0.01264235,0.013665276,0.01538256,
0.016622711,0.018446463,0.020722495,0.021919395,0.024539459,0.027397977,0.029962712,0.033137592,0.03
6572933,0.041678059,0.045744148,0.050192112,0.056625696,0.065557955,0.071178079,0.080266548,0.08905
0454,0.099668346,0.111650165,0.122228836,0.138847462,0.152324829,0.171681969,0.151259896,0.17559846
4,0.202502773,0.219830854,0.245743111,0.278305418,0.298524782,0.324731214,0.339793488,0.363048362,0.
370662461)

ks.test(X,Y,alternative="two.sided",exact=NULL)

Two-sample Kolmogorov-Smirnov test

data: X and Y

D = 0.0395, p-value = 1

alternative hypothesis: two-sided

> ks.test(X,Y,alternative="g",exact=NULL)

Two-sample Kolmogorov-Smirnov test

data: X and Y

D^+ = 0.0395, p-value = 0.8883

alternative hypothesis: the CDF of x lies above that of y

The Shapiro Wilk test for the cubic Spline fit gives

```
> YSpline<-
c(0.00063619,0.00069246,0.00075083,0.000810574,0.000871467,0.000933297,0.000995872,0.00105902,0.0011
22603,0.001186513,0.001250682,0.001316284,0.001388408,0.001468966,0.001558803,0.001658873,0.0017702
51,0.001894148,0.002031925,0.002185111,0.002355424,0.00254467,0.002754484,0.00298704,0.003244898,0.0
03530913,0.003848273,0.004200535,0.004591663,0.005026074,0.00550869,0.006043178,0.006627648,0.00726
5134,0.007960959,0.008721118,0.009552374,0.010462376,0.011459787,0.012554442,0.013757525,0.01508667
3,0.016577366,0.018254446,0.020140347,0.022259963,0.024640822,0.027313242,0.030310468,0.033668785,0.
037427575,0.041629331,0.046319591,0.051546786,0.057361977,0.063818461,0.070971221,0.078876207,0.087
589409,0.097165717,0.107657542,0.119113186,0.131574963,0.145077062,0.159643183,0.175283963,0.191994
247,0.209750268,0.228506816,0.248194508,0.268717282,0.289950258,0.311738127,0.333894229,0.356200498,
0.378408426)
>  shapiro.test(YSpline)
```

Shapiro-Wilk normality test

data: YSpline

W = 0.6698, p-value = 8.7e-12

TABLES AND FIGURES

<u>Logit Transformation Estimates.</u>

Ages	a	b	2009 logit qx Estimates	2009 logit qx Actual	2009 qx/1-qx Estimates	2009 qx Estimates	2009 qx Actual
0	-1.774265577	-0.031830845	-1.733575174	-2.646652412	0.03120583	0.030261495	0.005
1-4	-2.52988423	-0.039996712	-2.466050591	-3.523073484	0.007211335	0.007159704	0.00087
5-9	-2.854179164	-0.035881177	-2.775077486	-3.780580806	0.003886855	0.003871806	0.00052
10-14	-2.957943289	-0.028323829	-2.890238883	-3.68458027	0.00308724	0.003077738	0.00063
15-19	-2.62249572	-0.02157559	-2.575703185	-3.150803818	0.005791254	0.005757909	0.00183
20-24	-2.490925359	-0.023318182	-2.443033696	-2.969068984	0.00755106	0.007494468	0.00263
25-29	-2.473143684	-0.022961056	-2.427020857	-2.888244702	0.007796801	0.007736482	0.00309
30-34	-2.396428971	-0.020869898	-2.355743233	-2.756222637	0.008991402	0.008911277	0.00402
35-39	-2.17011657	-0.018694907	-2.135879264	-2.603337382	0.013957217	0.013765095	0.00545
40-44	-2.046562456	-0.015664667	-2.01982267	-2.412034428	0.017603715	0.017299185	0.00797
45-49	-1.812540017	-0.013605027	-1.791343038	-2.19083616	0.027800922	0.027048937	0.01235
50-54	-1.579307723	-0.01157483	-1.563271679	-1.965403262	0.043869175	0.042025549	0.01925
55-59	-1.349333201	-0.009724903	-1.337713938	-1.74340266	0.068877351	0.064438966	0.02969
60-64	-1.109918809	-0.008696846	-1.101469084	-1.50406428	0.110478078	0.099486951	0.04706
65-69	-0.868622197	-0.008206794	-0.862639353	-1.261947039	0.178123401	0.151192482	0.0742
70-74	-0.605065245	-0.008900196	-0.601044543	-0.998729353	0.300565648	0.231103788	0.11947
75-79	-0.324775072	-0.010127254	-0.323106351	-0.711569369	0.524026668	0.343843503	0.19417
80-84	-0.031990114	-0.010345016	-0.033198078	-0.406080948	0.935760083	0.483407056	0.30743
85-89	0.298742437	-0.010974256	0.293777548	-0.076770466	1.799583223	0.642803975	0.46169
90-94	0.636649191	-0.009750213	0.629435407	0.302545103	3.521442879	0.778831664	0.64682
95-99	1.035819157	-0.010483844	1.023710435	0.713392861	7.7478925	0.885686753	0.8064
100-104	1.42504657	-0.009190317	1.411096115	1.157917357	16.81366973	0.943863336	0.91018
105-109	1.789111342	-0.007028136	1.776112719	1.600873939	34.89087861	0.972137768	0.9609
110+						1	1

Log Transformation Estimates

Age	ln qx (Actual)	ln qx (Estimates)	qx (Actual)	qx (Estimates)
0	-5.298317367	-4.13408099	0.005	0.016017379
1-4	-7.047017346	-5.754537639	0.00087	0.003168371
5-9	-7.561681746	-6.282483969	0.00052	0.001868753
10-14	-7.369790739	-6.411860274	0.00063	0.001641967
15-19	-6.303439312	-5.624088113	0.00183	0.003609853
20-24	-5.940771433	-5.395798291	0.00263	0.004535598
25-29	-5.779584188	-5.35205971	0.00309	0.004738381
30-34	-5.516473376	-5.161595927	0.00402	0.005732544
35-39	-5.21213967	-4.822348177	0.00545	0.008047867
40-44	-4.832070786	-4.378254593	0.00797	0.01254724
45-49	-4.394099216	-3.885356867	0.01235	0.020540497
50-54	-3.950244218	-3.398221473	0.01925	0.033432678
55-59	-3.51694499	-2.92743418	0.02969	0.053534221
60-64	-3.056331896	-2.464727838	0.04706	0.085031982
65-69	-2.600991129	-2.026178421	0.0742	0.13183839
70-74	-2.124689985	-1.597977757	0.11947	0.202305215
75-79	-1.639021215	-1.197659839	0.19417	0.30189988
80-84	-1.17950786	-0.832813705	0.30743	0.434824097
85-89	-0.772861609	-0.5208336	0.46169	0.594025162
90-94	-0.43568723	-0.29311005	0.64682	0.745940046
95-99	-0.215175382	-0.145188364	0.8064	0.86485937
100-104	-0.094112897	-0.067483813	0.91018	0.934742851
105-109	-0.039884934	-0.03165395	0.9609	0.968841792
110+	0	0	1	1

Gompertz Law for middle ages

Age	Force of mortality	1-q*x	q*x (Estimated using Gompertz)	qx (Actual)
25	0.000439225	0.999560872	0.000439128	0.00052
26	0.000481117	0.999518998	0.000481002	0.00057
27	0.000527006	0.999473133	0.000526867	0.00055
28	0.000577271	0.999422896	0.000577104	0.00064
29	0.00063233	0.99936787	0.00063213	0.00065
30	0.00069264	0.999307599	0.000692401	0.00072
31	0.000758703	0.999241584	0.000758416	0.00067
32	0.000831068	0.999169278	0.000830722	0.00073
33	0.000910334	0.999090081	0.000909919	0.00076
34	0.00099716	0.999003337	0.000996663	0.00089
35	0.001092267	0.998908329	0.001091671	0.00102
36	0.001196446	0.998804269	0.001195731	0.001
37	0.001310562	0.998690297	0.001309703	0.00106
38	0.001435561	0.998565469	0.001434531	0.00116
39	0.001572483	0.998428753	0.001571247	0.00123
40	0.001722464	0.998279019	0.001720981	0.00143
41	0.00188675	0.998115029	0.001884971	0.00147
42	0.002066705	0.997935429	0.002064571	0.00156
43	0.002263825	0.997738736	0.002261264	0.00166
44	0.002479745	0.997523327	0.002476673	0.00186
45	0.00271626	0.997287426	0.002712574	0.002
46	0.002975332	0.99702909	0.00297091	0.00214
47	0.003259115	0.99674619	0.00325381	0.00219
48	0.003569965	0.9964364	0.0035636	0.00242
49	0.003910463	0.996097173	0.003902827	0.00262
50	0.004283437	0.995725724	0.004274276	0.00294
51	0.004691985	0.995319005	0.004680995	0.0031
52	0.0051395	0.994873685	0.005126315	0.00355
53	0.005629698	0.994386119	0.005613881	0.00391
54	0.00616665	0.993852325	0.006147675	0.00414
55	0.006754816	0.993267946	0.006732054	0.00473
56	0.007399081	0.992628225	0.007371775	0.0051
57	0.008104794	0.991927961	0.008072039	0.00541
58	0.008877818	0.991161474	0.008838526	0.00602
59	0.009724571	0.99032256	0.00967744	0.00655
60	0.010652086	0.989404446	0.010595554	0.00705
61	0.011668067	0.988399741	0.011600259	0.00764
62	0.01278095	0.98730038	0.01269962	0.00853
63	0.013999978	0.986097566	0.013902434	0.00876

64	0.015335275	0.984781711	0.015218289	0.0101
65	0.016797932	0.983342367	0.016657633	0.01096
66	0.018400094	0.981768154	0.018231846	0.01225
67	0.020155068	0.980046688	0.019953312	0.013
68	0.022077429	0.978164494	0.021835506	0.01441
69	0.024183142	0.976106928	0.023893072	0.01601
70	0.026489694	0.97385808	0.02614192	0.01756
71	0.029016243	0.971400686	0.028599314	0.01921
72	0.031783769	0.968716025	0.031283975	0.02124
73	0.034815259	0.96578382	0.03421618	0.0234
74	0.038135887	0.962582129	0.037417871	0.02594
75	0.041773233	0.959087246	0.040912754	0.02845
76	0.045757502	0.955273586	0.044726414	0.03195
77	0.050121786	0.951113585	0.048886415	0.03535
78	0.054902329	0.946577597	0.053422403	0.03941
79	0.060138833	0.941633795	0.058366205	0.04472
80	0.065874787	0.936248088	0.063751912	0.05068
81	0.072157827	0.930384044	0.069615956	0.05537
82	0.079040135	0.924002838	0.075997162	0.06201
83	0.086578868	0.917063219	0.082936781	0.0696
84	0.094836634	0.909521507	0.090478493	0.07718
85	0.103882013	0.901331636	0.098668364	0.08611
86	0.113790128	0.892445236	0.107554764	0.09569
87	0.124643264	0.882811777	0.117188223	0.10704
88	0.136531556	0.872378791	0.127621209	0.11959
89	0.149553736	0.861092165	0.138907835	0.13792
90	0.163817953	0.848896547	0.151103453	0.12827
91	0.17944267	0.835735862	0.164264138	0.14641
92	0.19655765	0.821553968	0.178446032	0.16915
93	0.215305031	0.806295458	0.193704542	0.18127
94	0.23584051	0.789906646	0.210093354	0.20168
95	0.258334633	0.77233674	0.22766326	0.21948
96	0.282974212	0.75353922	0.24646078	0.23787
97	0.309963879	0.733473449	0.266526551	0.25697
98	0.339527781	0.712106514	0.287893486	0.27668
99	0.371911444	0.689415292	0.310584708	0.29686
100	0.407383814	0.665388756	0.334611244	0.31737

Makeham Law for all ages

Age	**Force of mortality using Makeham model**	**1-q**x**	**q**x (Using Makehams law)**	**qx (Actual for all ages)**
0	1.41191E-05	0.999985881	1.4119E-05	0.00458
1	1.84148E-05	0.999981585	1.84146E-05	0.00028
2	2.31202E-05	0.99997688	2.31199E-05	0.00019
3	2.82744E-05	0.999971726	2.8274E-05	0.00014
4	3.39201E-05	0.99996608	3.39196E-05	0.00011
5	4.01044E-05	0.999959896	4.01036E-05	0.00011
6	4.68785E-05	0.999953123	4.68774E-05	0.00009
7	5.42988E-05	0.999945703	5.42973E-05	0.00007
8	6.24267E-05	0.999937575	6.24248E-05	0.00009
9	7.13299E-05	0.999928673	7.13274E-05	0.0001
10	8.10823E-05	0.999918921	8.1079E-05	0.00008
11	9.17648E-05	0.999908239	9.17606E-05	0.0001
12	0.000103466	0.999896539	0.000103461	0.00011
13	0.000116284	0.999883723	0.000116277	0.00013
14	0.000130324	0.999869685	0.000130315	0.00014
15	0.000145703	0.999854308	0.000145692	0.00021
16	0.000162549	0.999837465	0.000162535	0.00025
17	0.000181001	0.999819015	0.000180985	0.00036
18	0.000201214	0.999798806	0.000201194	0.00041
19	0.000223355	0.99977667	0.00022333	0.00045
20	0.000247607	0.999752424	0.000247576	0.00045
21	0.000274172	0.999725865	0.000274135	0.00047
22	0.000303271	0.999696775	0.000303225	0.0004
23	0.000335146	0.99966491	0.00033509	0.00046
24	0.000370061	0.999630008	0.000369992	0.00046
25	0.000408306	0.999591778	0.000408222	0.00052
26	0.000450198	0.999549903	0.000450097	0.00057
27	0.000496086	0.999504037	0.000495963	0.00055
28	0.000546351	0.999453798	0.000546202	0.00064
29	0.000601411	0.99939877	0.00060123	0.00065
30	0.000661721	0.999338498	0.000661502	0.00072
31	0.000727784	0.99927248	0.00072752	0.00067
32	0.000800148	0.999200172	0.000799828	0.00073
33	0.000879414	0.999120972	0.000879028	0.00076
34	0.000966241	0.999034226	0.000965774	0.00089
35	0.001061348	0.998939215	0.001060785	0.00102
36	0.001165527	0.998835152	0.001164848	0.001
37	0.001279643	0.998721176	0.001278824	0.00106
38	0.001404642	0.998596344	0.001403656	0.00116

39	0.001541564	0.998459624	0.001540376	0.00123
40	0.001691545	0.998309885	0.001690115	0.00143
41	0.001855831	0.99814589	0.00185411	0.00147
42	0.002035786	0.997966284	0.002033716	0.00156
43	0.002232906	0.997769585	0.002230415	0.00166
44	0.002448826	0.99755417	0.00244583	0.00186
45	0.00268534	0.997318262	0.002681738	0.002
46	0.002944413	0.997059917	0.002940083	0.00214
47	0.003228196	0.996777009	0.003222991	0.00219
48	0.003539046	0.996467209	0.003532791	0.00242
49	0.003879544	0.996127972	0.003872028	0.00262
50	0.004252518	0.995756511	0.004243489	0.00294
51	0.004661066	0.99534978	0.00465022	0.0031
52	0.005108581	0.994904446	0.005095554	0.00355
53	0.005598779	0.994416865	0.005583135	0.00391
54	0.006135731	0.993883054	0.006116946	0.00414
55	0.006723897	0.993298658	0.006701342	0.00473
56	0.007368162	0.992658917	0.007341083	0.0051
57	0.008073875	0.991958631	0.008041369	0.00541
58	0.008846898	0.99119212	0.00880788	0.00602
59	0.009693652	0.99035318	0.00964682	0.00655
60	0.010621167	0.989435038	0.010564962	0.00705
61	0.011637147	0.988430302	0.011569698	0.00764
62	0.012750031	0.987330907	0.012669093	0.00853
63	0.013969059	0.986128056	0.013871944	0.00876
64	0.015304356	0.98481216	0.01518784	0.0101
65	0.016767013	0.983372771	0.016627229	0.01096
66	0.018369175	0.98179851	0.01820149	0.01225
67	0.020124149	0.98007699	0.01992301	0.013
68	0.02204651	0.978194739	0.021805261	0.01441
69	0.024152223	0.976137108	0.023862892	0.01601
70	0.026458775	0.973888191	0.026111809	0.01756
71	0.028985324	0.971430722	0.028569278	0.01921
72	0.03175285	0.968745978	0.031254022	0.02124
73	0.03478434	0.965813681	0.034186319	0.0234
74	0.038104968	0.962611892	0.037388108	0.02594
75	0.041742313	0.9591169	0.0408831	0.02845
76	0.045726583	0.955303122	0.044696878	0.03195
77	0.050090867	0.951142993	0.048857007	0.03535
78	0.05487141	0.946606865	0.053393135	0.03941
79	0.060107913	0.94166291	0.05833709	0.04472
80	0.065843867	0.936277036	0.063722964	0.05068

81	0.072126908	0.930412811	0.069587189	0.05537
82	0.079009216	0.924031408	0.075968592	0.06201
83	0.086547949	0.917091574	0.082908426	0.0696
84	0.094805715	0.909549629	0.090450371	0.07718
85	0.103851094	0.901359505	0.098640495	0.08611
86	0.113759209	0.89247283	0.10752717	0.09569
87	0.124612345	0.882839073	0.117160927	0.10704
88	0.136500637	0.872405764	0.127594236	0.11959
89	0.149522817	0.861118789	0.138881211	0.13792
90	0.163787034	0.848922794	0.151077206	0.12827
91	0.179411751	0.835761703	0.164238297	0.14641
92	0.19652673	0.82157937	0.17842063	0.16915
93	0.215274112	0.806320388	0.193679612	0.18127
94	0.235809591	0.78993107	0.21006893	0.20168
95	0.258303714	0.77236062	0.22763938	0.21948
96	0.282943293	0.753562519	0.246437481	0.23787
97	0.30993296	0.733496128	0.266503872	0.25697
98	0.339496862	0.712128532	0.287871468	0.27668
99	0.371880525	0.689436608	0.310563392	0.29686
100	0.407352895	0.665409329	0.334590671	0.31737
101	0.446208567	0.640050265	0.359949735	0.33805
102	0.488770237	0.613380243	0.386619757	0.35875
103	0.535391377	0.585440118	0.414559882	0.37931
104	0.586459173	0.556293541	0.443706459	0.39955
105	0.64239774	0.526029628	0.473970372	0.41934
106	0.703671647	0.494765361	0.505234639	0.43853
107	0.770789768	0.46264754	0.53735246	0.45702
108	0.844309519	0.429854062	0.570145938	0.47468
109	0.924841475	0.396594284	0.603405716	0.49146
110	1.013054452	0.363108187	0.636891813	1

Cubic Spline Table

lnq*x = -9.498+0.08552(Age)-0.0001323(ϕ1)+0.0001867(ϕ2)-0.00001205(ϕ3)-0.000103(ϕ4)+0.0001706(ϕ5)

Year	**Age**	**fitted ln q*x**	**graduated qx**	**q*x**
2009	25	-7.36	0.000636198	0.000700119
	26	-7.27526057	0.00069246	0.000787998
	27	-7.19433138	0.00075083	0.000737113
	28	-7.11776809	0.000810574	0.000866715
	29	-7.04533256	0.000871467	0.000847691
	30	-6.97678665	0.000933297	0.000953483
	31	-6.91189222	0.000995872	0.000855496
	32	-6.85041113	0.00105902	0.000960993
	33	-6.79210524	0.001122603	0.000994443
	34	-6.73673641	0.001186513	0.001123714
	35	-6.6840665	0.001250682	0.001338015
	36	-6.63294254	0.001316284	0.00129736
	37	-6.57959776	0.001388408	0.001351145
	38	-6.52319658	0.001468966	0.001477006
	39	-6.46383692	0.001558803	0.00155608
	40	-6.4016167	0.001658873	0.001771173
	41	-6.33663384	0.001770251	0.00182854
	42	-6.26898626	0.001894148	0.001946113
	43	-6.19877188	0.002031925	0.002055257
	44	-6.12608862	0.002185111	0.002244635
	45	-6.0510344	0.002355424	0.002377787
	46	-5.973754135	0.00254467	0.002611377
	47	-5.89452529	0.002754484	0.002630462
	48	-5.813472295	0.00298704	0.002905696
	49	-5.73067138	0.003244898	0.003149202
	50	-5.646198775	0.003530913	0.003430682
	51	-5.56013071	0.003848273	0.003893916
	52	-5.472543415	0.004200535	0.004239612
	53	-5.38351312	0.004591663	0.004663989
	54	-5.293116055	0.005026074	0.004831139
	55	-5.20142845	0.00550869	0.005792477
	56	-5.108825235	0.006043178	0.006199807
	57	-5.01650534	0.006627648	0.00659252
	58	-4.924668595	0.007265134	0.007454661
	59	-4.83320583	0.007960959	0.007970783
	60	-4.742007875	0.008721118	0.008715752
	61	-4.65096556	0.009552374	0.009400416
	62	-4.559969715	0.010462376	0.01044081

	63	-4.46891117	0.011459787	0.010777701
	64	-4.377680755	0.012554442	0.01264235
	65	-4.2861693	0.013757525	0.013665276
	66	-4.193943495	0.015086673	0.01538256
	67	-4.09971703	0.016577366	0.016622711
	68	-4.003346615	0.018254446	0.018446463
	69	-3.90503016	0.020140347	0.020722495
	70	-3.804965575	0.022259963	0.021919395
	71	-3.70335077	0.024640822	0.024539459
	72	-3.600383655	0.027313242	0.027397977
	73	-3.49626214	0.030310468	0.029962712
	74	-3.391184135	0.033668785	0.033137592
	75	-3.28534755	0.037427575	0.036572933
	76	-3.178950295	0.041629331	0.041678059
	77	-3.07219028	0.046319591	0.045744148
	78	-2.965265415	0.051546786	0.050192112
	79	-2.85837361	0.057361977	0.056625696
	80	-2.751712775	0.063818461	0.065557955
	81	-2.64548082	0.070971221	0.071178079
	82	-2.539875655	0.078876207	0.080266548
	83	-2.43509519	0.087589409	0.089050454
	84	-2.331337335	0.097165717	0.099668346
	85	-2.2288	0.107657542	0.111650165
	86	-2.127681095	0.119113186	0.122228836
	87	-2.02817853	0.131574963	0.138847462
	88	-1.930490215	0.145077062	0.152324829
	89	-1.83481406	0.159643183	0.171681969
	90	-1.741347975	0.175283963	0.151259896
	91	-1.65028987	0.191994247	0.175598464
	92	-1.561837655	0.209750268	0.202502773
	93	-1.47618924	0.228506816	0.219830854
	94	-1.393542535	0.248194508	0.245743111
	95	-1.31409545	0.268717282	0.278305418
	96	-1.238045895	0.289950258	0.298524782
	97	-1.16559178	0.311738127	0.324731214
	98	-1.096931015	0.333894229	0.339793488
	99	-1.03226151	0.356200498	0.363048362
	100	-0.971781175	0.378408426	0.370662461