Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.4, No.7, 2014

www.iiste.org

Comparison Of Survival Models And Estimation Of Their Parameters With Respect To Mortality in a Given Population .

Kiche J. , Aduda J. , Athiany K.H.O, Apudo. B.O, Wanjoya A.

Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture

& Technology, P. O. Box 62000-00200, Nairobi, Kenya

* E-mail of the corresponding author:kicheopiyo07@gmail.com

ABSTRACT

In this research, we consider three different survival models under the assumption of Gompertz model as the baseline distribution. We compare the fitting results of the Exponential distribution , the Gompertz distribution and the Gompertz - Makeham model in a given population. As the human lifespan decreases, more and more people are becoming interested in mortality rates at higher ages. The aim of this study is to estimate the robust and reliable estimates of level and trend in mortality in Homa-Bay county , Kenya . The purpose of this study is to find out if the population of Homa-Bay area in Nyanza Province fits the Gompertz model and also to compare different survival models parametrically in a population. And also to determine the relationship between death rate and age in the area. . Akaike Information Criterion (AIC) is used to identify the best model. The expected output is that the preferred model is the one which satisfies the characteristics of the given population.

Keywords: Gompertz distribution, Gompertz-Makeham Model , Expected life time, Force of mortality, AIC, Stochastic process.

1.0 INTRODUCTION

The main objective of providing demographic data is to provide up-to-date information for policy makers, planners, researchers and program managers. This information guides the planning, implementation , monitoring and evaluation of population and health programs in Kenya.

Survival analysis examines and models the time it takes for events to occur. The prototypical such event is death, from which the name 'survival analysis' and much of its terminology derives, survival analysis focuses on the distribution of survival times.

Mortality rate is a measure of the number of deaths (in general or due to a specific cause) in a population.

1.1 Mortality Statistics.

Mortality statistics provide a valuable measure for assessing community health status. The importance of mortality statistics derives both from the significance of death in an individual's life as well as their potential to improve the public's health when used to systematically assess and monitor the health status of a whole community. Within the realm of public health, mortality statistics are often used as a cornerstone in formulating health plans and policies to prevent or reduce premature mortality and improve our quality of life.

The Gompertz model is widely used in demography and in various branches of science. Initially, the model was developed by Gompertz (1825) to describe age patterns of mortality. Over time, researchers have started using it as a growth model (Winsor, 1932).

The Gompertz distribution plays an important role in survival analysis especially in modeling human mortality. The Gompertz distribution is one of the most important growth models. The British actuary ,Benjamin Gompertz, proposed a simple formula in 1825 for describing the mortality rates of the elderly. This famous law Gompertz distribution states that death rates increase exponentially with age. The Gompertz–Makeham law states that the death rate is the sum of an age-independent component (the Makeham term, named after William

Makeham) and an age-dependent component (the Gompertz function, named after Benjamin Gompertz), which increases exponentially with age.

2. SURVIVAL MODELS USED IN THIS STUDY.

2.1 Survival Function.

The survival function is used to represent the probability that an individual survives from sometimes origin to sometime beyond t.

$$s(x) = Pr(X > x) \tag{1}$$

Also the integral of the probability density function $f(x)$:

$$s(x) = Pr(X > x) = \int_0^x f(t)dt \tag{2}$$

Thus, given a survival function, we can calculate the probability density function

$$f(x) = -\frac{ds(x)}{dx} \tag{3}$$

2.2 The Hazard Function

Hazard function is widely used to express the risk or hazard of death (failure) at some time x, and is obtained from the probability that an individual dies at time t, conditional on he/she having survived to time x. This is sometimes called instantaneous failure rate or force of mortality.

It is defined as

$$h(x) = \mu(x) = \lim_{\Delta x \to 0} \frac{\Pr\left[x \le X < x + \frac{\Delta x}{X} \ge x\right]}{\Delta x} \tag{4}$$

$$= \lim_{\Delta x \to 0} \frac{\Pr[x \le X < x + \Delta x]}{\Delta x Pr(X \ge x)}$$

$$= \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x\, s(x)}$$

$$= \frac{1}{s(x)} \left\{ \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} \right\} \tag{5}$$

Since X is a continuous random variable

$$h(x) = \frac{f(x)}{s(x)} = -\frac{d}{dx} log[s(x)] \tag{6}$$

The cumulative hazard is

$$H(x) = \int_0^x h(u)du = -\log[s(x)] \tag{7}$$

Thus for a continuous lifetimes

$$s(x) = \exp[-H(x)] = \exp[-\int_0^x h(u)du] \tag{8}$$

2.3 The Exponential Distribution

The probability density function (pdf) of an exponential distribution is

$$f(x;\beta) = \begin{cases} \beta e^{-\beta x}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

Here $\beta > 0$ is the parameter of the distribution, often called the *rate parameter.*

We also note that if the force of mortality, that is if the hazard rate is constant in a given population, then it follows an exponential distribution. By this we mean that regardless of the age group, the rate will be the same.

2.4 The Gompertz model

The force of mortality in the Gompertz model is given as

$$\mu(x) = \alpha e^{\beta x} \tag{9}$$

The corresponding cohort p.d.f is

$$f_0(x) = \alpha exp\left[\beta x + \frac{\alpha}{\beta}(1 - e^{\beta x})\right] \tag{10}$$

And the survival function is

$$s(x) = exp\left[\frac{\alpha}{\beta}(1 - e^{\beta x})\right] \tag{11}$$

Also note that log-mortality is a linear function of age

$$log\ \mu(x) = log(\alpha) + \beta x$$

This suggests a regression approach may be useful

2.4.1 Gompertz Life Expectancy and Its Approximation

Without the Makeham term, for the proportionally changing Gompertz force of mortality used by Vaupel (1986) is

$$\mu_G(x,y) = \alpha(y)e^{\beta x}$$

Missov and Lenart (2011) showed that the corresponding life expectancy of age x in year y can be solved as

$$e_G(x,y) = \frac{1}{\beta} e^{\frac{\alpha(y)}{\beta}} E_1\left(\frac{\alpha(y)}{\beta} e^{\beta x}\right) \tag{12}$$

where

$$E_1(z) = \int_z^\infty \frac{e^{-t}}{t} dt$$

denotes the exponential integral.

Note that

$$E_1(z) = \lim_{s\to 0} \int_z^\infty t^{s-1} e^{-t} dt = \lim_{s\to 0} \Gamma(s,z) \tag{13}$$

2.5 The Gompertz –Makeham model

The Gompertz –Makeham model specifies the force of mortality as

$$\mu(x) = \alpha_1 + \alpha_2 e^{\beta x} \tag{14}$$

The α_1 parameter in this expression represents the constant, age –independent component of mortality; the $\alpha_2 e^{\beta x}$ term is just a Gompertz function describing the senescent component.
The cohort PDF for the Gompertz –Makeham model is

$$f(x) = \left(\alpha_1 + \alpha_2 e^{\beta x}\right) exp\left[-\alpha_1 x + \frac{\alpha_2}{\beta}\left(1 - e^{\beta x}\right)\right] \tag{15}$$

And the Gompertz-Makeham survival function is

$$s(x) = exp\left[-\alpha_1 x + \frac{\alpha_2}{\beta}\left(1 - e^{\beta x}\right)\right] \tag{16}$$

The following table gives the summary of the quantities used in survival analysis for a particular given model:

Distribution	$f(x)$	$s(x)$	μ_x
Exponential	$\beta e^{-\beta x}$	$e^{-\beta x}$	β
Gompertz	$\alpha exp\left[\beta x + \frac{\alpha}{\beta}\left(1 - e^{\beta x}\right)\right]$	$exp\left[\frac{\alpha}{\beta}\left(1 - e^{\beta x}\right)\right]$	$\alpha e^{\beta x}$
Gompertz-Makeham	$\left(\alpha_1 + \alpha_2 e^{\beta x}\right) exp\left[-\alpha_1 x + \frac{\alpha_2}{\beta}\left(1 - e^{\beta x}\right)\right]$	$exp\left[-\alpha_1 x + \frac{\alpha_2}{\beta}\left(1 - e^{\beta x}\right)\right]$	$\alpha_1 + \alpha_2 e^{\beta x}$

2.5.1 The Gompertz-Makeham Life Expectancy and Its Approximation

Life expectancy at age x in year y can be given explicitly by

$$e_{GM}(x, y) = \frac{1}{\beta} e^{\frac{\alpha_2(y)}{\beta}} \left(\frac{\alpha_2(y)}{\beta}\right)^{\frac{\alpha_1(y)}{\beta}} \Gamma\left(\frac{-\alpha_1(y)}{\beta}, \frac{\alpha_2(y)}{\beta} e^{\beta x}\right) \tag{17}$$

where

$$\Gamma(s, z) = \int_z^{\infty} t^{s-1} e^{-t} dt$$

denotes the upper incomplete gamma function

If α_2 is close to 0, $e_{GM}(0, y)$, can be approximated by

$$e_{GM}(0, y) = \frac{1}{\alpha_1} - \frac{\left(\frac{\alpha_2}{\beta} e^{\gamma - 1}\right)^{\frac{\alpha_1}{\beta}}}{\alpha_1\left(1 - \frac{\alpha_1}{\beta}\right)} \tag{18}$$

where $\gamma \approx 0.57722$ is the Euler-Mascheroni constant.
For credible extreme values of human mortality, $0 < \frac{\alpha_2(y)}{\beta} e^{\beta x} \leq 1$

and $0 < \frac{\alpha_1(y)}{\beta} \leq 0.1$, $\Gamma\left(\frac{-\alpha_1(y)}{\beta}, \frac{\alpha_2(y)}{\beta} e^{\beta x}\right)$

can be approximated by

$$\Gamma(s, z) = \frac{1}{s + s^2} exp\{(1 - \gamma)s + 0.3225 s^2\} - \sum_{k=0}^{\infty} (-1)^k \frac{z^{s+k}}{k!\,(s + k)} \tag{19}$$

where $\xi(n) = \sum_{k=1}^{\infty} k^{-n}$ is the Riemann zeta function and $0.3225 \approx \frac{\xi(2) - 1}{2}$.

2.6 Stochastic Process

The word stochastic has a Greek origin meaning randomness.

Basically we can define stochastic process as a sequence of random variables in a given system Mathematically, a stochastic process is a collection of random variables $\{X_t\}_{t\epsilon T}$

Where $T\epsilon\Re$ is called the index set.

If T is discrete ie $T = \{0,1,2,...\}$, then the stochastic process $\{X_t\}_{t\epsilon T}$ is called a stochastic process in discrete time

Again if T is in an interval in $\Re$ i.e. $T = [0,\infty)$, then the process $\{X_t\}_{t\epsilon T}$ is called a stochastic process in continuous time.

Note that in this research work, we are talking about stochastic process which means randomness to clarify and explain more on mortality as a risk measure since it can occur anytime thus it is unexpected and hence the term randomness.

3.0 METHODOLOGY

3.1 Model Comparison:

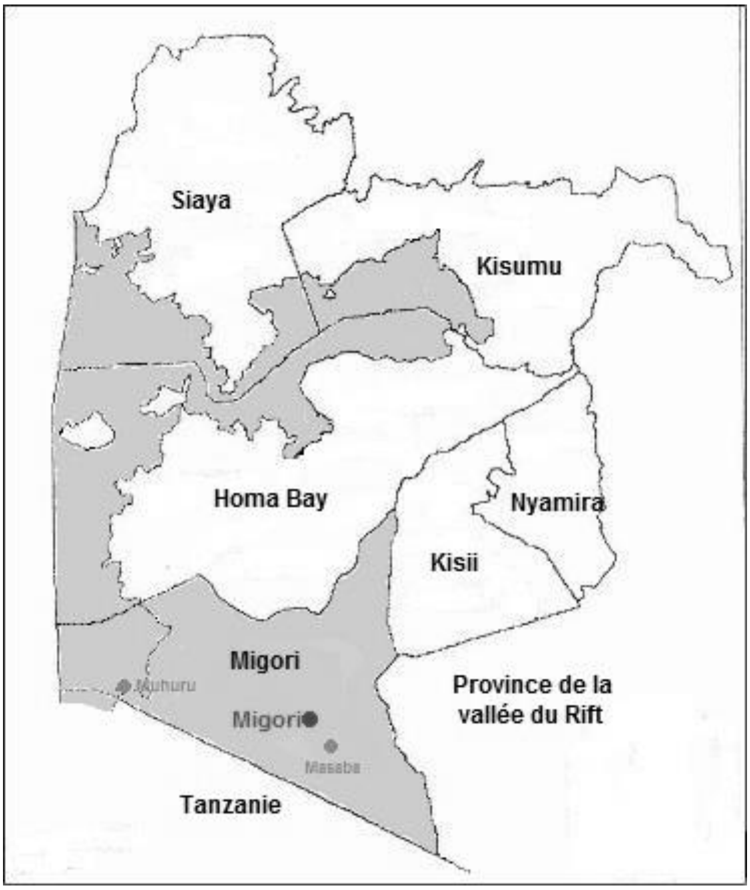
The different survival models in this case were compared using the AIC technique .
Model comparison and selection are among the most common problems of statistical practice, with numerous procedures for choosing among a set of models proposed in the literature.
The AIC is a measure of the relative quality of a statistical model, for a given set of data. As such, AIC provides a means for model selection.

AIC deals with the trade- off between the goodness of fit of the model and the complexity of the model.
In the general case, the AIC is

$$AIC = 2k - 2\ln(L) \qquad (20)$$

where k is the number of parameters in the statistical model, and L is the maximized value of the likelihood function for the estimated model.
Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. Hence AIC not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters. This penalty discourages over-fitting (increasing the number of free parameters in the model improves the goodness of the fit, regardless of the number of free parameters in the data-generating process).

3.2 Research Study Area
The study site considered in this research is the Homa-Bay county in Nyanza Province.
The site has resources which can increase the likelihood of the success of the study.



Geography of Homa Bay County

Homa Bay County is located in the now defunct Nyanza Province, it borders Lake Victoria to the West and North, and the following counties; Kisumu and Kericho to the North East, Nyamira and Kisii to the East, and Migori to the South.

4.0 RESULTS AND DATA ANALYSIS

Empirical results

Table 1: Presents age-specific mortality rates for women age 15-49.

Age	Deaths	Exposure	Mortality rates
15-19	35	21051	1.7
20-24	82	24580	3.3
25-29	134	22762	5.9
30-34	134	19218	7.0
35-39	131	14593	9.0
40-44	105	9444	11.1
45-49	56	5478	10.3
Overall 15-49	677	117126	5.8

Table 2 : Presents age-specific mortality rates for men age 15-49.

Age	Deaths	Exposure	Mortality rates
15-19	52	20943	2.5
20-24	76	24702	3.1
25-29	95	22915	4.1
30-34	136	19030	7.2
35-39	118	14090	8.4
40-44	114	9298	12.3
45-49	84	5627	14.9
Overall 15-49	675	116604	6.0

Table 3 : Long Term Population and Household Projection (2014 to 2030)

Year	Population size	Number of Households
2014	83,708	16,742
2018	93,122	18,624
2022	103,594	20,719
2026	115,243	23,048
2030	128,203	25,640

Table 4: Shows the AIC values of the given distributions.

	Exponential	Gompertz	Gompertz-Makeham
Log Likelihood Value	-41.70	-36.09	-34.56
Number of Parameters	1	2	3
AIC	85.47	73.16	61.93
Rank	3	2	1

Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.4, No.7, 2014

Table 5: Shows the Correlation value between age and the death rate.

		age	Death rate
Age	Pearson Correlation	1	.727**
	Sig. (2-tailed)		.005

We find that from the above given analysis , there is a strong positive relationship between the death rate and age, that is death rate increases exponentially with age.

Table 6: Model Summary and Parameter Estimates

Dependent variable: population size

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	.999	6.500E3	1	4	.000	-4.030E6	2.042E3
Growth	1.000	1.105E5	1	4	.000	-42.604	.027
Exponential	1.000	1.105E5	1	4	.000	3.144E-19	.027

The independent variable is year.

Figure 1: Depicts mortality rate and Age For Women

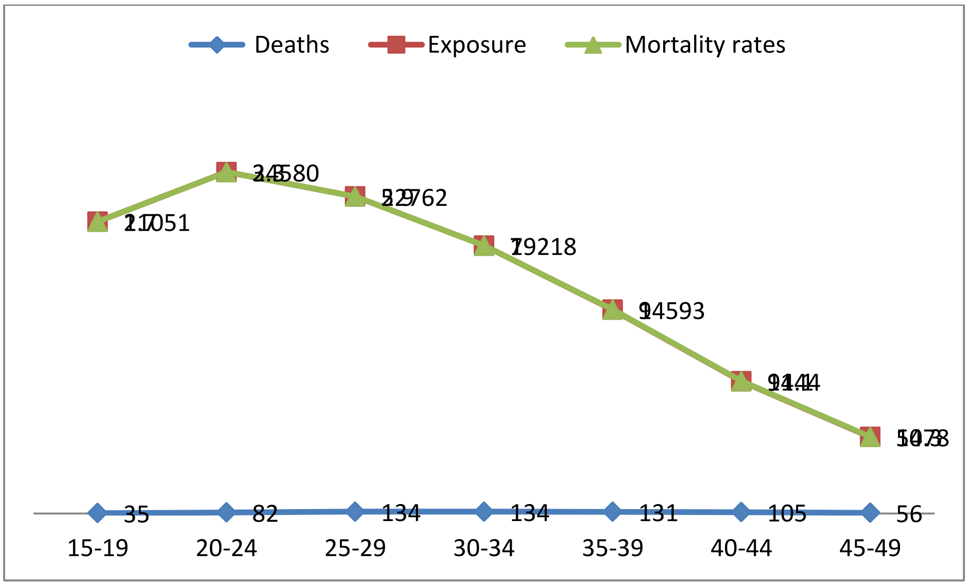


Figure 2: Shows mortality rate and age for men.

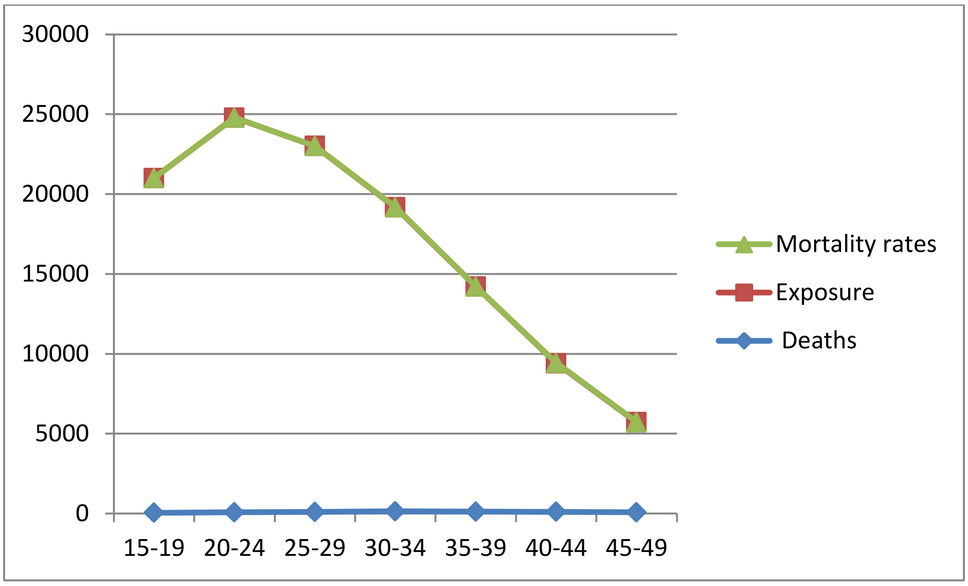


Table 7: Proves that death rate is directly proportional to the age.

Age interval	Probability of death	Number of persons surviving the exact age	Number of deaths during age interval
0-1	0.0250	80000	2500
2-5	0.0051	77500	500
6-10	0.00393	77000	300
11-15	0.00259	76700	250
16-20	0.00622	76100	600
21-25	0.00835	75300	800
26-30	0.00789	74500	750
31-35	0.00875	73750	825
36-40	0.0128	72925	1200
41-45	0.0217	71725	2000
46-50	0.0332	69725	3000
51-55	0.0575	66725	5050
56 and above	0.0857	61675	7050

Figure 3: Depicts the population size with the respective year.

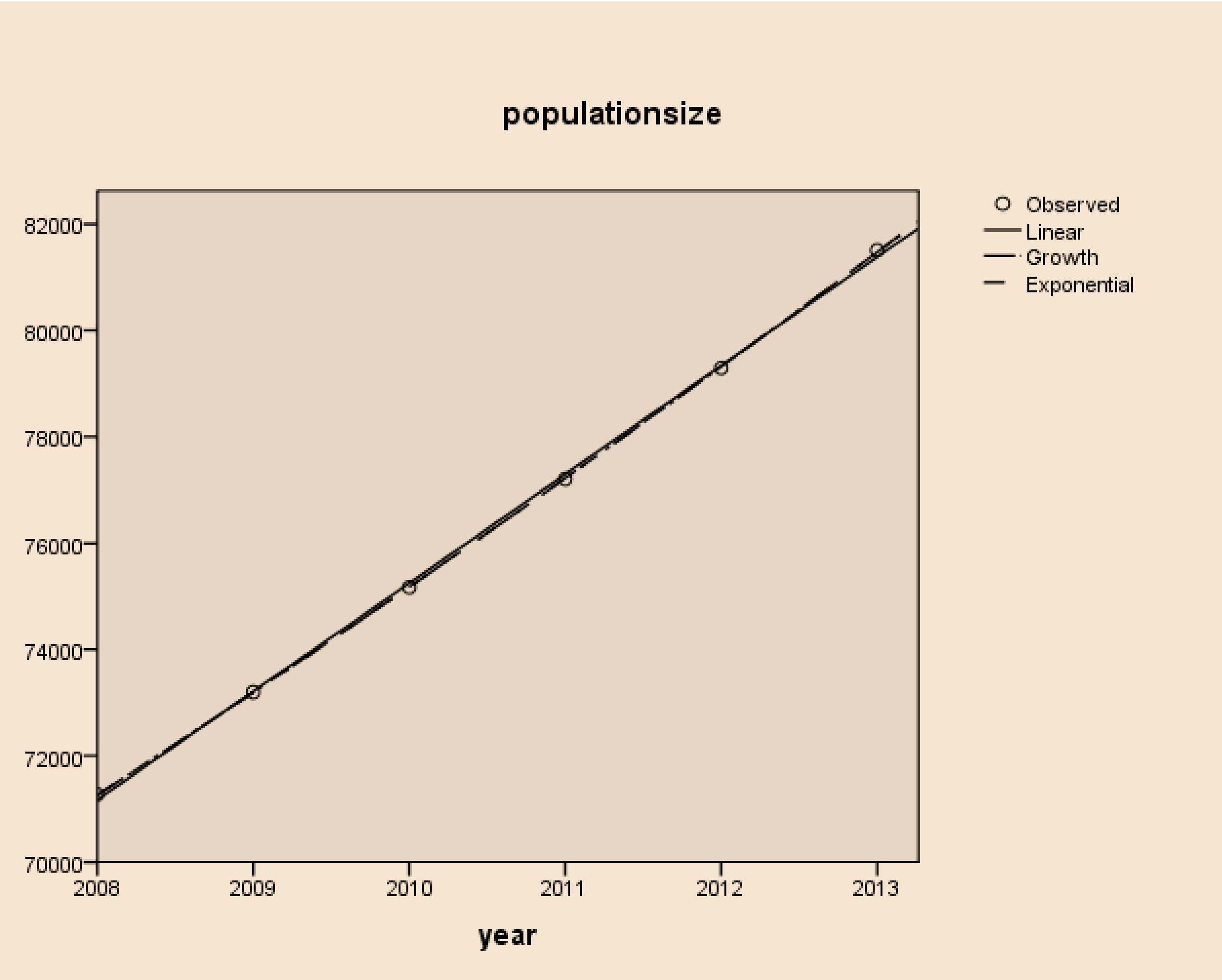


Figure 4: Shows the number of households and the population size in a corresponding year

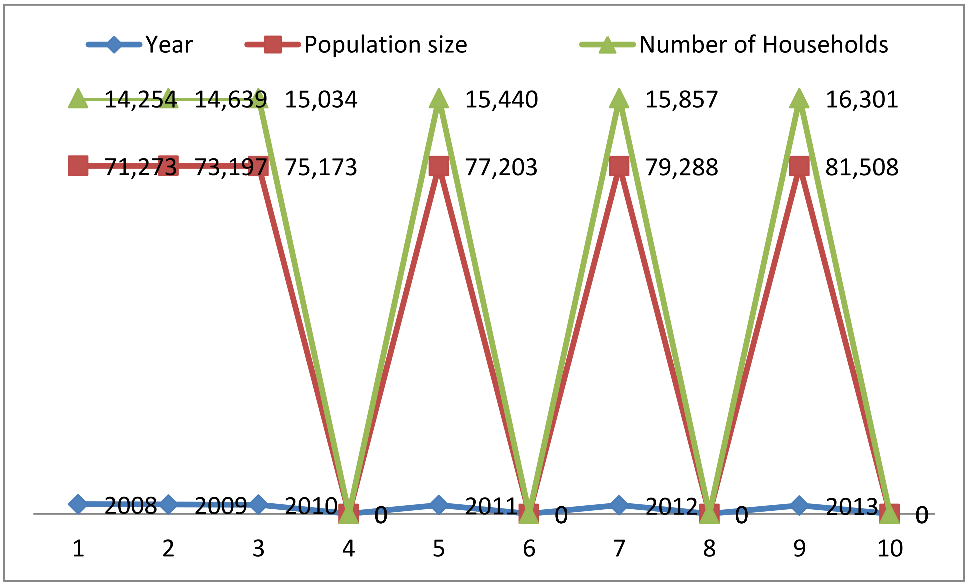
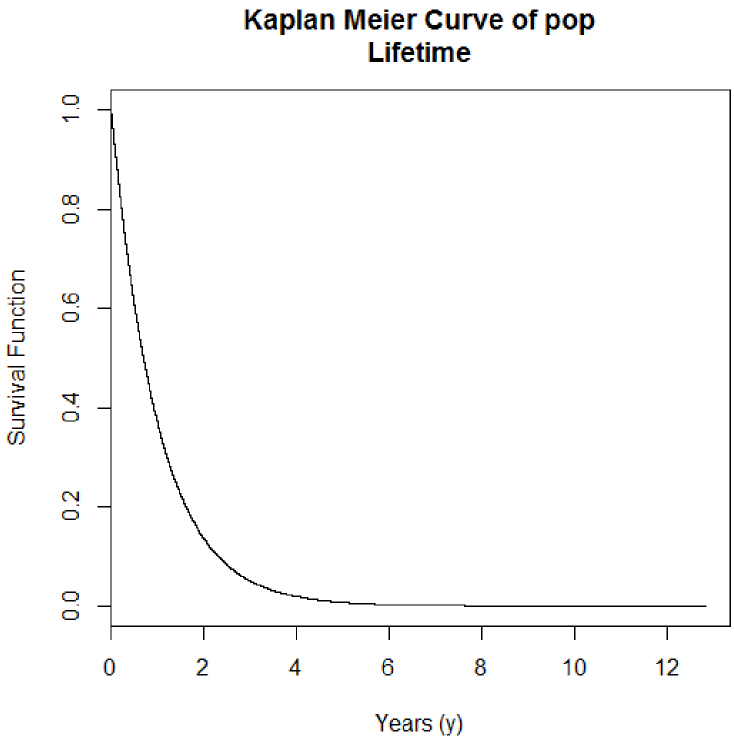


Table 8: Showing the values of the parameters estimates.

Distribution	Number of parameters	Parameter estimates using MLE		
Exponential	1	$\beta = 1.000$		
Gompertz	2	$\alpha = 0.021$	$\beta = 0.437$	
Gompertz - Makeham	3	$\alpha_1 = 0.042$	$\alpha_2 = 0.0011$	$\beta = 1.329$

Figure 5: Shows the Kaplan Meier curve of population lifetime



Demographic Challenges.

- High poverty levels
- Low employment opportunities
- High mortality /infant mortality rates
- High dependency levels
- Low life expectancy
- High rate of in-migration

The noted demographic challenges should be addressed by the policy makers and economic planners in order to help in the reduction of mortality cases.

Discussion

The life expectancy of Homa Bay is 48 years for males and 52 for females. This is a great deal lower than the national average, which is 57 and 58 for males and females respectively.

At ages 20-29 and 35-39, female mortality exceeds male mortality, with a wider difference at age group 25-29, while the rates are nearly the same at age group 30-34. Above age 40, male mortality exceeds female mortality by wider margins as age advances.

The rates are stable, showing expected increase in rate of mortality for both sexes as their age increases . The rise is steeper for men at older ages and steeper for women at younger ages.
The overall mortality rates are lower among women than men(5.8 and 6.0 deaths per 1000)
.

Conclusion

According to the findings of this research , mortality rate in Homa –Bay county, Kenya is not only influenced by age but also by other age independent factors including accidents, social factors and economic factors and many others.
Therefore, the best model that would fit well in this population is the Gompertz-Makeham distribution since it has the minimum AIC value . We also note that if the force of mortality, that is if the hazard rate is constant in a given population, then it follows an exponential distribution. By this we mean that regardless of the age group, the rate will always be the same.
This study has also shown that mortality declines as economic performance improves. From this work , one can also find the corresponding life expectancy of age x in year y using the given life expected models.
Thus mortality is not only related to age but also to other factors which include social factors and economic factors which are summarized as age –independent factors.
This study was meant to provide future survivorship for people in this region in order to help in the economic planning putting to consideration the age factor and mortality rate.

Recommended further area of research.
The different survival models should now be compared and their parameters estimated and applied in a larger area including now the entire country in order to determine the model that best fits the population of Kenya. This will assist the economic planners and demographers to make the required policies concerning the population in the country and help in the allocation and distribution of the resources hence solves the problem of the demographic challenges . Since even the burial sites nowadays have become a problem, once it is known in advance that a given population conforms to or fits a given survival model ,then the issue can be comfortably be addressed by the planners in the country.

Acknowledgement .
Much appreciation goes to the KNBS who through their general director and ICT director allowed and provided me with the required data that really assisted me in my research studies.

REFERENCES

1) Abramowitz, M. and I. Stegun. 1965. *Handbook of Mathematical Functions*. Washington, DC: US Government Printing Office.
2) Manton, K.G., Stallard, E. and Vaupel, J.W(1981) methods for comparing the mortality experience of heterogeneous populations, *Demography* **18**,389-410
3) Pakin, Yu. V. , Hrisanov, S.M.(1984) Critical Analysis of the Applicability of the Gompertz-Makeham Law in Human Populations. *Gerontology* 30: 8-12.
4) Hougaard,P.(1984) Life Table methods for heterogeneous populations: Distributions describing heterogeneity, *Biometrika* **71**, 75-83.

5) Vaupel, J. 1986. "How change in age-specific mortality affects life expectancy." *Population Studies* 40(1):147-157.
6) Pollard, J. and E. Valkovics. 1992. "The Gompertz distribution and its applications." *Genus* 48(3-4):15-29.
7) Benjamin, B. and Pollard, J.H.(1993) The analysis of mortality and other actuarial statistics, The Institute of Actuaries and the Faculty of Actuaries, London.
8) Carriere, JF (1994): An investigation of the Gompertz law of mortality, Actuarial Research Clearing House , 2, 161-177.
9) Kunimura, D. 1998. "The Gompertz Distribution-Estimation of Parameters." Actuarial Research Clearing House 2:65-76.
10) Willemse, W. and H. Koppelaar. 2000. "Knowledge Elicitation of Gompertz' Law of Morality." *Scandinavian Actuarial Journal* pp. 168-180.
11) Preston, S., P. Heuveline and M. Guillot. 2001. *Demography: measuring and modeling population processes*. Wiley-Blackwell.