Estimation of Longevity using Survival Functions

Ajayi Moses Adedapo[a]*, Shangodoyin Dahud Kehinde[b],

Thaga K[c], Mokgathle Lucky[d]

[a, b,c,d] Statistics Department, University of Botswana, Gaborone, +267, Botswana

[a] Email:ttkdapo@gmail.com

Abstract

Estimation of life expectancy and longevity are frequently used in several fields basically on convectional model. The study considers Longevity estimation models and variances as alternative to conventional model incorporating non-parametric, semi-parametric and parametric with relevant survival functions to formulate longevity models that will accommodate expected variables in the estimation of life expectancy. These models can be used to analyze both grouped and individual data. We consider Kaplan Meier model as nonparametric methods, Cox proportional model as semi parametric methods. The parametric methods considered are exponential, weibull and gompertz proportional models which are commonly used for analyzing discrete data. The derived variances of KM and Cox proportional longevity estimator models are equivalent this confirm the result of major work done in this field. The other derived longevity models have difference variances due to their peculiarities. We shall consider these models empirically in another paper.

Keywords: longevity; parametric; expectancy; model.

1. Introduction

Estimates of longevity and life expectancy are frequently used in several fields including biostatistics, demography, economics, engineering and sociology.

----

* Corresponding author.

E-mail address: ttkdapo@gmail.com.

The expressions duration analysis, event-history analysis, failure-time analysis, reliability analysis, and transition analysis refer essentially to the same group of techniques although the emphases in certain modelling aspects could differ across disciplines [4]. The concepts of longevity and life expectancy are fairly easy to understand: the longest lived individual and how long individuals live on average, respectively. What is not easily understood is that there are many different methods of estimating life expectancy, and the different methods can give widely different answers. When selecting a method for estimating life expectancy, it is important to ensure that the method used is appropriate for the data available and for the life history of the subjects. Although there is rarely only one correct method of brief demographic information, it is often possible to show methods that are clearly incorrect or give misleading results [5].

According to [3], the methodological developments of life expectancy that have had the most profound impact are the Kaplan-Meier method for estimating the survival function, the log-rank test for comparing the equality of two or more survival distributions, and the Cox proportional hazards (PH) model for examining the covariate effects on the hazard function. The accelerated failure time (AFT) model was proposed but seldom used. In this study, we will develop longevity models, using non-parametric, semi-parametric and parametric with survival function. We shall compare nonparametric methods (the Kaplan-Meier), semi parametric methods (the Cox PH model) and parametric methods (Parametric PH model) which are commonly used for analyzing academic retiree's data.

2. Longevity Estimation model

Longevity estimation model could be useful in replacing of the conventional model in cohort life table for estimating life expectancy. The expected years of survival in cohort life table is given as

$$\ell(x) = \frac{T(x)}{l(x)} \tag{1}$$

Where T_x is the total number of per-years lived by the cohort after age x

$$\ell(x) = \frac{\sum_{max}^{min} nL(x)}{l(x)}$$

From life table, $l(x)$ is the number of survivors in the life table at exact age x out of an initial population call radix at age 0. $nL(x)$ or $L(x,n)$ is total number of person-years lived by the total population between ages x and x+n and where "n" is the age interval.

$$nL(x) = L(x,n) = 0.5n\big(l(x) + l(x+n)\big)$$

$$\ell(x) = \frac{\sum\limits^{x}_{\text{max}} o.5n\big[l(x)+l(x+n)\big]}{l(x)}$$

$$= \frac{0.5n\left(\sum\limits^{x}_{\text{max}}\big[l(x)+l(x+n)\big]\right)}{l(x)}$$

Using the life table formulae we have

$$l(x+n) = P(x)*l(x) \tag{2}$$

Where p_x is the probability of survival from age x to x+n.

If $p(x) = \dfrac{l(x+n)}{l(x)}$, then we have

$$\ell(x) = 0.5n\sum\limits^{x}_{\text{max}}(p(x)+1)$$

$$= 0.5n\sum\limits^{x}_{\text{max}} p(x) + 0.5n\sum\limits^{x}_{\text{max}} 1 \tag{3}$$

The Kaplan Meier estimate of survival function

$$s(x) = \prod\limits_{\leq x} p(x) \text{ Could be modified as}$$

$$s(x) = p(x) \times s(x-1) \tag{4}$$

Where $p(0) = 1$, Therefore, $\qquad p(x) = \dfrac{s(x)}{s(x-1)}$

$$\sum\limits^{x}_{\text{max}} p(x) = \sum\limits^{x}_{\text{max}} \frac{s(x)}{s(x-1)} \tag{5}$$

Therefore

$$LG(x) = 0.5n \sum_{\max}^{x} \frac{s(x)}{s(x-1)} + 0.5n \sum_{\max}^{x} 1$$

$$LG(x) = 0.5n \left[\sum_{\max}^{x} \left[s(x)(s(x-1)^{-1}) \right] + I_{\geq x} \right] \tag{6}$$

Where $I_{\geq x}$ = summing 1 up to the age x from the highest age. Using this expression the Kaplan Meier model is

$$LG(x)_{km} = 0.5n \sum_{\max}^{x} \frac{s(x)}{s(x-1)} + 0.5n \sum_{\max}^{x} 1$$

$$= (0.5n) \left[\sum_{\max}^{x} \frac{\prod [1-h(x)]}{\prod [1-h(x-1)]} + \sum_{\geq x}^{x} 1 \right] \tag{7}$$

From equation (6), we then deduce longevity estimator (LG) for selected non parametric, semi-parametric and parametric survival models.

From equation (6), we have the general variance as

$$Var(LG(x)) = (0.5n)^2 \left[\sum_{\max}^{x} Var\left(\frac{s(x)}{s(x-1)} \right) \right] \tag{8}$$

This involves ratio of variable $s(x)$ and $s(x-1)$. The general formula is

$$Var\left(\frac{s(x)}{s(x-1)} \right) = \frac{\mu_{s(x)}^2}{\mu_{s(x-1)}^4} Var(s(x-1)) + \frac{1}{\mu_{s(x)}^2} Var(s(x))$$

$$-2\rho\sigma_{s(x)}\sigma_{s(x-1)}\mu_{s(x)}\mu_{s(x-1)}^{-1} \tag{9}$$

Using (5.9) in (5.8) gives

$$VarLG(x) = (0.5n)^2 \left[\sum_{\max}^{x} \left(\frac{\mu_{s(x)}^2}{\mu_{s(x-1)}^4} \sigma_{s(x-1)}^2 + \frac{1}{\mu_{s(x)}^2} \sigma_{s(x)}^2 - 2\rho\sigma_{s(x)}\sigma_{s(x-1)}\mu_{s(x)}\mu_{s(x-1)}^{-1} \right) \right] \tag{10}$$

This latter equation shall be used to derive the model variance for other forms of longevity to be considered later in this chapter.

3.   Longevity Estimation using Kaplan Meier (Km)

Consider the longevity equation of KM

$$LG(x)_{km} = 0.5n\left[\sum_{max}^{x}\left[s(x)(s(x-1)^{-1})\right] + I_{\geq x}\right] \tag{11}$$

The variance is of the form

$$Var\left[LG(x)_{km}\right] = (0.5n)^2\left[\sum_{max}^{x}Var\left(\frac{s(x)}{s(x-1)}\right)\right] \tag{12}$$

Equation (11) involves ratio of estimators and is given as

$$Var\left(\frac{s(x)}{s(x-1)}\right) \cong \frac{\mu_{s(x)}^2}{\mu_{s(x-1)}^4}Var(s(x-1)) + \frac{1}{\mu_{s(x)}^2}Var(s(x))$$

$$-2\rho\sigma_{s(x)}\sigma_{s(x-1)}\mu_{s(x)}\mu_{s(x-1)}^{-1} \tag{13}$$

where $\sigma_{s(x)} = \sqrt{Var(s(x))}$, $\sigma_{s(x-1)} = \sqrt{Var(s(x-1))}$ and $\rho = corr\left[s(x), s(x-1)\right]$

The required variance of $LG(x)_{km}$ is

$$Var\left[LG(x)_{km}\right] = (0.5n)^2\left[\sum_{max}^{x}\left(\frac{\mu_{s(x)}^2}{\mu_{s(x-1)}^4}\sigma_{s(x-1)}^2 + \frac{1}{\mu_{s(x)}^2}\sigma_{s(x)}^2 - 2\rho\sigma_{s(x)}\sigma_{s(x-1)}\mu_{s(x)}\mu_{s(x-1)}^{-1}\right)\right] \tag{14}$$

From equation (14) it is required to find Var. ($s(x)$) and by mathematical indication we obtain Var($s(x-1)$).

To do this, we have $Var(\ln s(x)) = Var\sum_{max}^{x}ln(1-h(x))$ which can be obtained using Delta method [1,2].

But suppose that it is the number of individuals at risk in x_j with d_j as the number of deaths at x_j. Given that

r_j is the total number of individual surviving in the interval $(x_j x_{j+1})$, we can deduce that random number

$r_j - d_j$ has a binomial distribution with parameter r_j and $1-\left(\dfrac{d_j}{r_j}\right)$. Thus, the conditional variance of

$r_j - d_j$ is given by $\quad Var\left(r_j - d_j / r_j\right) = r_j h_j(1-h_j)$; Therefore

$$Var\left(\hat{h}_j / r_j\right) = Var(1-\hat{h}_j) = \frac{h_j(1-h_j)}{r_j} \tag{15}$$

By Delta method, we have

$$Var\left[lns(x / r_j)\right] = Var\sum_{max}^{x}ln(1-\hat{h}(x_j r_j))$$

$$= \sum_{j:x_j<x}^{r} \left(\frac{d}{dh(x_j)} \left(ln(1-h(x_j)) \right) \right)^2 Var(h(x_j / r_j))$$

$$= \sum_{1}^{r} \left(\frac{1}{1-h(x_j)} \right)^2 \left(\frac{h(x_j)(1-h(x_j))}{r_j} \right)$$

$$= \sum_{1}^{r} \frac{h(x_j)}{r_j(1-h(x_j))} \qquad \text{Where j = 1, 2... r}$$

Replacing $h(x_j)$ with $d(x_j)/r_j$ gives

$$Var(lns(x/r_j) = \sum_{j:x<x} \left[\frac{d_j}{r_j(r_j-d_j)} \right]$$

Finally, if $z = lns(x)$ then $\ell^z = \ell^{lns(x)}$, therefore $Var(s(x)) = Var(\ell^z) = \left[\frac{d}{dz} \ell^z \right]^2 Var(z)$

$$= (s(x))^2 Var(lns(x))$$

$$= (s(x))^2 \sum_{j:x_{j<x}} \frac{d_j}{r_j(r_j-d_j)} \tag{16}$$

Also

$$Var(s(x-1) = (s(x-1))^2 \sum_{j:x_{j<x-1}} \frac{d_j}{r_j(r_j-d_j)} \tag{17}$$

Let $\sigma_{s(x)} = \sqrt{Var(s(x))} = (s(x)) \left[\sum_{j:x_{j<x}} \frac{d_j}{r_j(r_j-d_j)} \right]^{1/2}$

$\sigma_{s(x-1)} = \sqrt{Var(s(x-1))} = (s(x-1) \left[\sum_{j:x_{j<x-1}} \frac{d_j}{r_j(r_j-d_j)} \right]^{1/2}$. Then

$$Var(LG(x)_{km}) = (0.5n)^2 \left\{ \sum_{\max}^{x} \left[\frac{\mu_{s(x)}^2}{\mu_{s(x-1)}^4} Var(s(x-1) + \frac{1}{\mu_{s(x)}^2} Var(s(x) \right] - 2\rho s(x)s(x-1)\mu_{s(x)}\mu_{s(x-1)}^{-1} \left[\sum_{j:x_{j<x}} \frac{d_j}{r_j(r_j-d_j)} \right]^{1/2} \left[\sum_{j:x_{j<x-1}} \frac{d_j}{r_j(r_j-d_j)} \right]^{1/2} \right\} \tag{18}$$

4. Longevity Estimation using Cox Proportional (CP) Model

The Cox Proportional Hazard (CPH) model is given by

$h(x/y) = h_0(x)\exp(\beta_i y_i + \ldots\ldots + \beta_p y_p) = h_0(x)\exp(\beta' y)$. $h_0(x)$ is the base line hazard function, this

is the hazard function for an individual for which all the variables included in the model are zero. Y =

$(y_1,\ldots\ldots y_q)$ is the values of the vector of explanatory variables for a particular individual and β' is the vector of

regression coefficient. The survival function is given by

$$s(x/y) = s_0(x)\exp\left(\sum_{j=1}^{p}\beta_j y_j\right) \qquad (19)$$

Also

$$s((x-1/y)) = s_0(x-1)\exp\left(\sum_{j=1}^{p}\beta_j y_j\right) \qquad (20)$$

Since $s_0(x)$ is the base line survival function at age x, then we can extend to point x-1 which in turn gives the

longevity model for CP model is

$$LG(x)_{cp} = 0.5n\left\{\sum_{max}^{x}\left[s_0(x)\exp\left(\sum_{j=1}^{p}\beta_j y_j\right)s_0(x-1)\exp\left(\sum_{j=1}^{p}\beta_j y_j\right)\right]^{-1} + I_{\geq x}\right\} \qquad (21)$$

$$LG(x)_{cp} = 0.5n\left\{\sum_{max}^{x}\left[s_0(x)s_0(x-1)^{-1}\right] + I_{\geq x}\right\} \qquad (22)$$

The equation (22) is the same as equation (11)

The variance of Cox proportional longevity model could be derived using equation (22)

$$Var(LG(x)_{cp}) = (0.5n)^2\left\{\sum_{max}^{x}\left[\frac{\mu_{s_0(x)}^2}{\mu_{s_0(x-1)}^4}Var(s_0(x-1)) + \frac{1}{\mu_{s(x)}^2}Var(s_0(x))\right] - 2\rho s_0(x)s_0(x-1)\mu_{s_0(x)}\mu_{s_0(x-1)}^{-1}\left[\sum_{j:x_j<x}\frac{d_j}{r_j(r_j-d_j)}\right]^{1/2}\left[\sum_{j:x_j<x-1}\frac{d_j}{r_j(r_j-d_j)}\right]^{1/2}\right\}$$

$$\ldots\ldots\ldots$$

(23)

5.   Longevity Estimation Using Exponential PH Model

In literature the exponential PH model is a special case of the Weibull model when $\gamma = 1$. The hazard function

under this model is to assume that it is constant over time. The survival and hazard function are written

respectively as $s_{(x)} = \exp(-\lambda x)$ and $h(x/y) = \lambda \exp(\beta_1 y_1 + \beta_2 y_2 + ... + \beta_p y_p) = \lambda \exp(\beta' y)$.

Under exponential PH model, with covariates $(y_1, y_2, y_3, ... y_p)$, the survival function is

$$s(x/y) = \exp\left[-\lambda x \exp(\beta' y)\right] \tag{24}$$

Also

$$s(x-1/y) = \exp\left[-\lambda(x-1)\exp(\beta' y)\right] \tag{25}$$

Therefore using (24 and 25) in (6) gives the exponential proportional estimation of longevity as

$$LG(x)_{EXP} = 0.5n\left\{\sum_{\max}^{x}\left(\exp\left[-\lambda(x)\exp(\beta' y)\right]\left[\exp\left(-\lambda(x-1)\exp(\beta' y)\right)\right]^{-1}\right) + I_{\geq x}\right\} \tag{26}$$

$$LG(x)_{EXP} = 0.5n\left\{\sum_{\max}^{x}\exp\left(\lambda \exp\beta' y\right) + I_{\geq x}\right\} \tag{27}$$

$$LG(x)_{EXP} = 0.5n\left\{\sum_{\max}^{x}\exp\left(h(x)\right) + I_{\geq x}\right\} \tag{28}$$

The variance of longevity exponential proportional model is derived from equation (28) as:

$$VarLG(x)_{\exp} = (0.5n)^2\left\{\sum_{\max}^{x}Var(\exp h(x))\right\} \tag{29}$$

But let $\exp h(x) = \varphi$, then $Var(\exp h(x)) = Var(\varphi)$ are $\ln\varphi = h(x)$, $Var(\ln\varphi) = Var(h(x))$.

Using Delta method $Var\ln\varphi = \dfrac{1}{\varphi^2}Var(\varphi) = Var(h(x))$ from the last expression, we have

$\varphi^2 Var(h(x)) = Var(\varphi) = Var(\exp h(x))$. Therefore

$$Var(\exp h(x)) = (\exp h(x))^2 Var(h(x)) \tag{30}$$

But $h(x) = \lambda \exp(\beta' y)$. If we use Taylor series approximation (Alex, 2009) we have $\exp\beta' y \square 1 + \beta' y$.

Thus $Var(h(x)) = \lambda^2 \sum \beta_i^2 Var(y_i)$ So we have;

$$Var\left[LG(x)_{\exp}\right] = \left(0.5n\exp h(x)\lambda\right)^2 \sum \beta_i^2 Var(y_i) \tag{31}$$

6. Longevity Estimation using Weibull PH Model

Suppose that survival times are assumed to have a Weibull distribution with scale parameter λ shape parameter γ , so the survival and hazard function of a W(λ , γ) distribution are given respectively by $s(x) = \exp(-\lambda x^{\gamma})$ and $h(x) = \lambda\gamma(x)^{\gamma-1}$ With λ , γ > 0. The hazard rate increases when γ > 1 and decreases when γ < 1 as time goes on. Whens γ = 1, the hazard rate remains constant, which is the special exponential case.

Under the Weibull PH model, the hazard function of a particular patient with covariates $(y_1, y_2, y_3, \ldots\ldots y_p)$ is given by:

$$h(x/y) = \lambda\gamma(x)^{\gamma-1}\exp(\beta_1 y_1 + \beta_2 y_2 + \ldots + \beta_p y_p) = \lambda\gamma(x)^{\gamma-1}\exp(\beta' y).$$

The effects of the explanatory variables in the model alter the scale parameter of the distribution, while the shape parameter remains constant. The corresponding survival function is given by:

$$s(x/y) = \exp\left[-\lambda x^{\gamma}\exp(\beta' y)\right] \qquad (32)$$

and

$$s(x-1/y) = \exp\left[-\lambda(x-1)^{\gamma}\exp(\beta' y)\right] \qquad (33)$$

Given equation (5.32) and (5.33) we can relate $s(x)$ and $h(x)$ as

$$s(x) = \exp\left(-h(x)\gamma^{-1}\right) \text{ and } s(x-1) = \exp\left(-h(x-1)\gamma^{-1}\right)$$

The longevity model using Weibull distribution and gives equation (6) and (33) is

$$LG(x)_{WB} = (0.5n)\left[\sum_{max}^{x}\left(\frac{\exp\left(-h(x)\gamma^{-1}\right)}{\exp\left(-h(x-1)\gamma^{-1}\right)}\right) + I_{\geq x}\right]$$

$$= (0.5n)\left[\sum_{max}^{x}\exp\left(\frac{h(x)}{\gamma}\right) + I_{\geq x}\right] \qquad (34)$$

To derive the variance of $LG(x)_{WB}$ we have

$$VarLG(x)_{WB} = (0.5n)^2\sum\gamma^2 Var(h(x)) \qquad (35)$$

By taking $h(x) = \lambda\gamma x^{\gamma-1}\exp(\beta' y) \square \lambda\gamma x^{\gamma-1}\left(1+\sum\beta_i y_i\right)$ then

$$Var(h(x)) = \left(\lambda \gamma x^{\gamma-1}\right)^2 \sum \beta_i^2 Var(y_i)$$

$$VarLG(x)_{WB} = (0.5n)^2 \left[\sum \gamma^2 \left(\lambda \gamma x^{\gamma-1}\right)^2 \sum \left(\beta_i^2 Var(y_i)\right) \right] \tag{36}$$

7.   Longevity Estimation using Gompertz PH Model

The survival and hazard function of the Gompertz distribution are given respectively by

$s_{(x)} = \exp\left(\dfrac{\lambda}{\theta}\left(1 - \exp(\theta x)\right)\right)$ and $h(x) = \lambda \exp(\theta x)$. For $0 \le x < \infty$ and $\lambda > 0$. The parameter θ

determines the shape of the hazard function and to be equals to zero. The survival time has an exponential

distribution. The hazard function of a particular person is given by $h(x / \underline{y}) = \lambda \exp\left(\sum \hat{\beta}_i' y_i\right) \exp(\theta x)$,

consequently we may write the survival function as:

$$s(x / y) = \exp\left(\frac{\lambda}{\varphi}(1 - \exp \varphi x \exp \beta' y)\right) \tag{37}$$

$$= \exp\left(\frac{\lambda}{\varphi}(1 - h(x))\right) \text{ by mathematical induction, we have}$$

$$s(x - 1 / y) = \exp\left(\frac{\lambda}{\varphi}(1 - h(x-1))\right) \tag{38}$$

By using equation (5.6), we have the longevity estimator for Gompertz model distribution as

$$LG(x)_{GP} = 0.5n \left\{ \sum_{\text{max}}^{x} \left(\frac{\exp\left(\dfrac{\lambda}{\varphi}(1 - h(x))\right)}{\exp\left(\dfrac{\lambda}{\varphi}(1 - h(x-1))\right)} \right) + I_{\ge x} \right\} \tag{39}$$

We derive the variance of $LG(x)_{GP}$ model by using equation (5.15) as

$$VarLG(x)_{GP} = (0.5n)^2 \left[\sum_{\text{max}}^{x} Var\left(\frac{\exp\left(\dfrac{\lambda}{\varphi}h(x)\right)}{\exp\left(\dfrac{\lambda}{\varphi}h(x-1)\right)} \right) \right].$$

So let $s_x = \exp\left(\dfrac{\lambda}{\varphi} h(x)\right)$, then we have to find $Var\left(\dfrac{s_x}{s_{x-1}}\right)$ as defined in equation (9).

But in order to find variance of s_x, we use delta method to get

$$Var\left[\exp\left(\frac{\lambda}{\varphi} h(x)\right)\right] = \left(\frac{\lambda}{\varphi}\right)^2 \rho^{\frac{\lambda}{\varphi} h(x)} Var(h(x)) = \sigma^2_{s(x)} \tag{40}$$

and

$$Var\left[\exp\left(\frac{\lambda}{\varphi} h(x-1)\right)\right] = \left(\frac{\lambda}{\varphi}\right)^2 \rho^{\frac{\lambda(h(x-1))}{\varphi}} Var\left[h(x-1)\right] = \sigma^2_{s(x-1)} \tag{41}$$

Thus using elements of (5.40) and (5.41), to get the required variance as:

$$VarLG(x)_{GP} = (0.5n)^2 \left[\sum_{\max}^{x}\left(\frac{\mu^2_{s(x)}}{\mu^4_{s(x-1)}}\sigma^2_{s(x-1)} + \frac{1}{\mu^2_{s(x)}}\sigma^2_{s(x)} - 2\rho\sigma_{s(x)}\sigma_{s(x-1)}\mu_{s(x)}\mu^{-1}_{s(x-1)}\right)\right] \tag{42}$$

8. Conclusion

We developed longevity models in replacing the conventional models for estimating life expectancy with the view of making useful comparison with their model variances as well as to allow data to conform with appropriate model. We incorporated non-parametric, semi-parametric and parametric with survival function. These models can be used to analyse both grouped and individual data. We consider Kaplan Meier model as nonparametric methods, Cox proportional model as semi parametric methods. The parametric methods considered are exponential, weibull and gompertz proportional models which are commonly used for analyzing discrete data. Overall, as expected the derived variances of KM and Cox proportional longevity estimator models are equivalent. The other derived longevity models have difference variances due to their peculiarities. We shall consider these models empirical in another paper.

References:

[1] Alex, Papanicolaou (2009): Taylor Approximation and the Delta Method, http://www.stanford.edu

[2] David W. Hosmer, Stanley Lemeshow and Susanne (2008): Applied Survival Analysis: Regression Modeling of Time-to-Event Data, Second Edition, John Wiley & Sons, Inc.

[3] Jiezhi Qi (2009): Comparison of Proportional Hazards and Accelerated Failure Time Models, University of Saskatchewan Saskatoon, Saskatchewan.

[4] Joseph C. Gardiner (2010): Survival Analysis: Overview of Parametric, Nonparametric and Semi parametric approaches and New Developments Division of Biostatistics, Department of Epidemiology, Michigan State University, East Lansing, MI 48824

[5] Robert J. Wiese and Kevin Willis (2004): Calculation of Longevity and Life Expectancy in Captive Elephants TECHNICAL REPORT, Minnesota Zoological Gardens, Apple Valley, Minnesota