Bayesian Estimation of parameters for Survival models using the Cox Proportional model
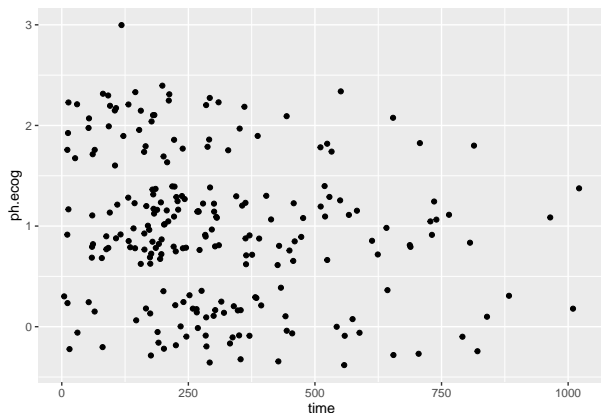
Daniela Pico Julian Usuga Deivid Zhang

2022-06-19

Introduction

The primary aim of our study is examine the mortality and survival of the people from African countries.

Specifically, to estimate survival time and probability, to identify predictive factors, and to describe causes and circumstances of deaths.

Methods

Study Population

Many countries don't have good registries management on government institutions and thus they don't have a good record of important events such as population deaths, the Demographic and Health Surveys (DHS) facilitates multiple datasets containing data collected from questionnaires performed in households from a large list of countries.

In our study we will consider the Individual Women's data, which consist of questionnaires performed to womens. they were asked about wherever they had siblings, their survival status, age and death date (in case they have died).

Data sources

The primary data source was the Demographic and Health Surveys (DHS) Individual Recode (IR) where each row consist of a woman and their responses on multiple questions.

Data sources

The primary data source was the Demographic and Health Surveys (DHS) Individual Recode (IR) where each row consist of a woman and their responses on multiple questions.

Basic concepts

Survival analysis

Survival analysis is a collection of statistical methods to study the time that elapses until the occurence of an event(also known as event analysis failure times or survival analysis).The name is survival is because these methods are generally used to study times to death, sime fields of application of survival analysis include: health sciences, engineering, social sciences, economics, epidemiology, among others. Survival data must face a problem (often unavoidable) called censorship, censorship occurs when you have some information about the survival times of a patient but the time to failure is not exactly known.

There are three essential quantities of interest for survival analysis: the hazard rate, the cumulative haz-

ard function, and the survival probability. It is these quantities that are used to form the likelihood the survival models described in later sections.

Let T be a random variable denoting the survival time of a unit or person. The survival function, which is denoted $S(t)$ gives the probability that a unit or person will survives beyond some specific time t. i.e,

$$S(t) = P(T > t)$$

The survival function $S(t)$ is fundamental in AS since for different values of t provide crucial information of survival data In some situations it may be more interest to quantify the risk of failure at a given instant than to estimate survival; a function of interest in survival analysis that allows do this is the danger function

The hazard function usually denoted as $h(t)$ or $\lambda(t)$ is given by:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T + \Delta t | T \geq t)}{\Delta t}$$

The numerator represents the conditional probability of the event occurring in an infinitesimal interval $[t, t + \Delta t]$ (as $\Delta t \to 0$) given that the unit has survived to t $(T > t)$. survived until t (T>t).

The cumulative hazard is defined as:

$$H(t) = \int_0^t h(s) ds$$

Where $h(s)$ is the hazard function.There is a one-to-one relationship between the hazard function, the cumulative hazard and the probability of survival, as follows

$$S(t) = exp(-H(t)) = exp(-\int_0^t h(s) ds)$$

Censorship

An observation in a random variable T is said to be right-censored if all that is known about of T is that it is greater than some value c. In AS, T refers at the time of occurrence of some particular event and a case is considered right-censored if it stops observing before for the event to occur. Survival probability can be estimated nonparametrically over time observations (censored and uncensored) using the Kaplan-Meyer method. Others two common approaches to

modelling survival data. The first is to model the instantaneous rate of the event (known as the hazard) as a function of time. This includes the class of models known as proportional and non-proportional hazards regression models.The second is to model the event time itself. This includes the class of models known as accelerated failure time (AFT) models.

Models

Under this modeling framework, it is proposed to implement a of Bayesian Cox proportional hazard to later make a comparison versus the frequentest approach and the kaplan Meier curve.

Kaplan–Meier estimator

The Kaplan-Meier estimator, also known as the limit product estimator, is a nonparametric method for estimating the survival function. survival function by maximizing the sample likelihood function. Suppose one has k different failure times $t_1 < t_2 < ... < t_k$, in each time $t_j (j = 1, 2, ...k)$ there are n_j subjects that are under observation and at risk of an event of interest.

The K-M estimator is defined as

$$\hat{S}_{KM}(t) = \prod_{j:t_j \leq t} [1 - \frac{d_j}{n_j}] \quad para \quad t_1 \leq t \leq t_k \quad y \quad d_j = \#faults$$

Cox proportional hazards model

Under a hazard scale formulation, we model the hazard of the event for individual i at time t using the regression model:

$$h(t|X_i) = h_0(t) exp(X_i \beta)$$

Where h_0 is called the hazard base, that is, $h0(t)$ is the risk when all X_i variables are 0. $h_0(t)$ characterizes the way in which hazard changes as a function of survival time, while the second term characterizes the way the hazard changes as function of the covariates at the same time guarantees that this is positive, it is also called the linear predictor.

The formulation of the Cox model in terms of the hazard and the survival function are given by: From (2) we note that

$$S(t; x) = exp(-H(t; x))$$

Where $H(t; \vec{x})$ is the cumulative hazard for a subject with covariates $\vec{x} = (x_1, x_2, ..x_k)$ Assuming survival time is continuous

$$H(t; \vec{x}) = \int_0^t h(s; \vec{x})ds = \int_0^t h_0(s)exp(\beta^T \vec{x})ds$$

$$= exp(\beta^T \vec{x}) \int_0^t h_0(s)ds = exp(\beta^T \vec{x})H_0(T)$$

Cox model in terms of cumulative hazard. In this expression $H_0(t)$ is the cumulative baseline hazard. This relationship can be thought of as a baseline cumulative risk measure which is modified according to the function

From the above relationship, the Cox model can be formulated in terms of survival:

$$S(t; \vec{x}) = exp(-H(t; \vec{x})) = exp(e^{\beta^T \vec{x}})H_0(t) =$$

$$[exp(-H_0(t)]^{e^{\beta^T \vec{x}}} = [S_0(t)]^{e^{\beta^T \vec{x}}}$$

Cox model in terms of survival. In this expression $S_0(t)$ is the survival baseline.