Challenges and Chances of Classical Cox Regression

Mengying Xia
Emory University, USA

Leigh Wang
Northwestern University, USA

INTRODUCTION

A brand-new era of big data and with that the outburst of interest in prediction modeling has arrived. With this upsurge comes the challenge and chances of evaluating statistical prediction models. Kattan and Gerds (2020) provide guidance for the evaluation and critique of a statistical prediction model. Almost all of the prediction activities fall into one of three settings. The first is the prediction of a present state or condition that is not yet known, such as diagnosis or presence of disease. The second category of prediction models comprise those that predict the probability of an event happening in the future. The third category is similar to the second, except there is also the possibility of a competing event occurring.

The advance in computational algorithms plays an essential role in statistical inference and machine learning research. Survival analysis is a statistical method of analysis of the duration of time until an event has occurred, such as death or biological problems. Survival analysis is very important especially in the medical standpoint because while events are occurring, statistics are able to interpret all of the factors that go into the outcome to try to figure out why the event occurred or what could've been done differently in order to possibly change the outcome.

Cox regression usually provides better estimates of survival probabilities and cumulative hazards. It is mainly used to measure the survival time of patients and the variable(s) that affect this outcome which can become a huge breakthrough in the medical community. Cox regression is the only time where independent variables are appropriate and it has the advantage of preserving the variable in its original quantitative form while using the maximum amount of information. The Cox model can observe a whole data set instead of having to split it up into groups or sections based on time.

Cox regression models and methods are very important when it comes to the medical field. This method is essential to calculating survival based on different variables that can affect one's survival rate. Medical facilities can have accurate studies based on the attributes of their patients and the variables, getting survival rates that can lead them in the direction of curing diseases, making medicines, and so much more. Having this method within medical technology, different facilities can keep logs of data for their patients using the data a Cox regression model would give them, and can observe it easily. The best Cox models are those which include censored data— observations where the event *didn't* happen, as well as data from observations where the event actually occurred.

DOI: 10.4018/978-1-7998-9220-5.ch146

C

BACKGROUND

The Cox Proportional Hazards Regression Model, or the Cox regression model for short, was discovered by David Roxbee Cox in 1972. It is commonly used for survival analysis, as it is concerned with the amount of time that passes until a particular event occurs, such as a death of a patient. Hazard function is the term for the rate at which a patient's death is known. Being a semiparametric model, it is a regression model with both a finite- and an infinite-dimensional component. This means there are no assumptions on the shape of the baseline hazard function and the measure of effect is the hazard rate. The hazard represents the anticipated number of events per a single unit of time. Consequently, the hazard in a group can exceed 1 at times. The three main uses for Cox regression are independence of survival times between separate individuals of a sample, a multiplicative relationship between independent variable predictors and the hazard (as opposed to a linear relationship like of a multiple linear regression analysis can give), and a constant hazard ratio over time progression (LaMorte, 2016).

Cox regression is statistical in medical research for investigating the association between the survival time of patients and one or more predictor variables, meaning, it uses data collected in medical studies to explore possible outcomes and medical breakthroughs. This model coincides with quantitative predictor variables and categorical variables. Quantitative predictor variables provide information on an associated dependent variable regarding a particular outcome. The Cox model is beneficial because while others can only analyze one risk factor at a time, the Cox method can analyze multiple risk factors at the same time which is extremely beneficial being that medical research is constantly changing and can be very crucial at times. Cox proportional hazard regression is commonly used to model censored survival data. The purpose of the Cox proportional hazards regression model (CM) is to model the simultaneous effect of multiple factors on the survival. The CM aims to estimate hazard ratios over time. The model equation is written as follows:

$$h(t \mid z_1,\ldots,z_p) = h_0(t)\exp\left(\sum_i \beta_i z_i\right) \tag{1}$$

Where (z_i) i=1, …, p are the values of the covariates $Z_1, \ldots, Z_p$ on which the hazard may depend and $h_0(t)$ represents the baseline hazard. The baseline hazard is defined as the value of the hazard when $z_i=0$, for i in1, p. The predicted hazard (i.e., h(t)), or the rate of suffering the event of interest in the next instant, is the product of the baseline hazard ($h_0(t)$) and the exponential function of the linear combination of the predictors. Thus, the predictors have a multiplicative or proportional effect on the predicted hazard, as LaMorte has noted.

CURRENT RESEARCH

Addressing a Public Health Problem

Kim *et al*. (2021) examined the association between BMI and clinical outcomes among patients with COVID-19 infection. A total of 10,861 patients with COVID-19 infection who were admitted to the Northwell Health system hospitals between March 1, 2020, and April 27, 2020, were included in this study. BMI was classified as underweight, normal weight, overweight, and obesity classes I, II, and III. Primary outcomes were invasive mechanical ventilation (IMV) and death. This study showed that