DEPARTMENT OF STATISTICS
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Your title Here

Your Name Here

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfillment of the requirements for
the degree of M.Sc. in Statistics

Athens
2024

Acknowledgements

I would like to thank ....

YOUR TITLE HERE

Abstract

Survival analysis in clinical trials has an important role in biomedical research. ....

Ο ΤΙΤΛΟΣ ΣΤΑ ΕΛΛΗΝΙΚΑ

Περίληψη

Η ανάλυση δεδομένων επιβίωσης σε κλινικές δοκιμές έχει σημαντικό ρόλο στη βιοϊατρική έρευνα. ....

Contents

List of Tables

List of Figures

Chapter 1

Introduction

At the early stages of biomedical research, basic methods of survival analysis are used to elaborate data. In data from clinical trials (CT), the most common interest is the time till an event occurs, such as death or disease recurrence. Concerning this time, two or more groups of patients are compared. Groups of patients received different treatments, and the purpose is to investigate which treatment is better than the other. The first approach for this analysis is the non-parametric, Kaplan-Meier estimator and the logrank test that can evaluate if two survival curves are equal or not. However, this estimator can not explore how survival curves differ and, moreover, does not take into consideration possible confounders or characteristics of patients. The next step of the analysis, that can transcend these obstacles, is the Cox Proportional Hazards (PH) Model. Cox models are based on the baseline hazard function and can perform a multivariate analysis. These two methods, Kaplan-Meier estimator and Cox models are very simple, easy to implement, and achieve high power under proportional hazard assumption. The majority of clinical trials are designed based on these methods.

Nevertheless, there are some clinical trials where the treatment has early or late effect. In these occasions, the proportionality assumption is not valid. One representative example of clinical trials with potentially delayed treatment effects is cancer immunotherapies. This leads to considerate new methods for comparing the treatments, that deal with the proportionality assumption and are not limited to this. As a solution, there have been proposed some additional changes to the aforementioned approaches. These are the weighted logrank test, the stratification, or the time-varying coefficients, however, these methods have disadvantages. Restricted Mean Survival Time (RMST) is an alternative method for estimating the treatment effect.

Furthermore, treatments can be compared with the estimated RMST. The equality of RMST can be tested. Equivalently, the difference of RMST, that is the area between the two survival curves up to time τ, can be tested, if it is statistically significant different from the value of zero. In publications such as Royston and Parmar (2013), the power of RMST test outperforms the logrank and Cox test, when there are non-proportional hazards, and presents a little bit lower or almost equal power when the proportionality occurs. Finally, RMST is a promising method in the designing of CT. In some occasions, research of Weir and Trinquart (2018) has indicated by simulations and redesigning of processed CT, that the sample size, that is required to achieve the specific value of power, is remarkably lower than the sample size proposed under the hazard ratio design.

Chapter 2

Survival analysis

2.1 Definitions

Survival analysis is a statistical approach to analyze the expected duration of time until an event of interest occurs, often called failure. The purpose of survival analysis is to approach the portion of a population that will survive past a certain time, the rate of failure, or even more the characteristics that affect the probability of survival. It is widely used in medicine, physics, economics, sociology and engineering known us reliability theory. Examples of failure times could be the lifetimes of machine components in industrial reliability, the duration of strikes or periods of unemployment in economics or the lengths of tracks on a photographic plate in particle physics. In medicine, survival compares the failure times in two or more groups, to decide which one is longer and which covariates contribute to these results. The most common events are death, death from a specific cause, occurrence of a disease, recurrence of a disease after treatment,or infection. In order to define a failure time, there are three requirements: a time origin must be unambiguously defined, a scale for measuring the passage of time must be agreed and finally the meaning of failure must be entirely clear.

2.2 Distributions of failure time

There are several equivalent ways to characterize the probability distribution of a survival random variable. For the following a homogeneous of individuals is considered, with each having a failure time. This would be a single non-negative continuous random variable, T.

- density function:
$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t}.$$
- cumulative density function:
$$F(t) = P(T \leq t) = \int_0^t f(t)$$
- survival function:
$$S(t) = P(T > t) = 1 - F(t).$$

- hazard function:
$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t | T \leq t)}{\Delta t}.$$
- cumulative hazard function:
$$\Lambda(t) = \int_0^t \lambda(u) du.$$

The hazard function is the rate of failure given that the failure has not occurred before time t. It is strongly related to survival and density function.

$$\begin{aligned}\lambda(t) &= \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t | T \leq t)}{\Delta t} = \\ &= \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t | T \leq t)}{\Delta t \;\; P(T \leq t)} = \\ &= \frac{f(t)}{S(t)}.\end{aligned}$$

Relationship between distributions:

It is obvious that $\frac{dS(t)}{dt} = -f(t)$ and

$$-\frac{d}{dt}[\log S(t)] = -(\frac{1}{S(t)}) \; S'(t) = -\frac{-f(t)}{S(t)} = \frac{f(t)}{S(t)} = \lambda(t).$$

This concludes in the following relation:

$$\log S(t) = -\Lambda(t) \Rightarrow S(t) = \exp(-\Lambda(t)).$$

Important parametric survival distributions

On some occasions, the pattern of survival time for the study subjects follows a predictable pattern. Some distributions are widely used to model or simulate data. These distributions describe the event till an event occurs and should be over non-negative values. However, even distributions that have support that includes negative values, can be used to describe a transformation of time such as $\log(T)$.

Exponential distribution

Exponential distribution with parameter λ and mean $\frac{1}{\lambda}$ has the property of lack of memory. According to this property, the probability of an event occurring after a period s is equal to the probability of occurring unconditionally to period s. This reflects the constant hazard over time.

- density function: $f(t) = \lambda \exp(-\lambda t)$
- survival function: $S(t) = 1 - F(t) = \exp(-\lambda t)$
- hazard function: $\lambda(t) = \frac{f(t)}{S(t)} = \lambda$

- cumulative hazard function: $\Lambda(t) = \int_0^t \lambda(t) = \lambda t$

Weibull distribution

Assuming a Weibull distribution with scale λ and shape p.

- density function: $f(t) = p\lambda^p t^{p-1} \exp(-\lambda t)^p$
- survival function: $S(t) = 1 - F(t) = \exp(-\lambda t)^p$
- hazard function: $\lambda(t) = \frac{f(t)}{S(t)} = p\lambda(\lambda t)^{p-1}$
- cumulative hazard function: $\Lambda(t) = \int_0^t \lambda(t) = (\lambda t)^p$

For $p = 1$ Weibull distribution is equal to exponential. For $0 < p < 1$, the hazard is decreasing, while for $p > 1$ is increasing.

Piecewise exponential

Piecewise exponential distribution assumes hazard rate is constant within each interval but can vary between intervals. This approach is useful when the hazard rate is not constant over time, and the survival can be better captured by allowing for different rates in different time segments. This means that there are some specific time points where the hazard changes. Defining one change point τ_1 where the hazard rate differs, then there are two time intervals and two ratios noted:

- hazard function:
$$\lambda_t = \begin{cases} \lambda_0, & \text{for } 0 < t < \tau_1 \\ \lambda_1, & \text{for } t > \tau_1 \end{cases}$$
- cumulative hazard function:
$$\Lambda(t) = \int_0^t \lambda(t) = \int_0^{\tau_1} \lambda_0 dt + \int_{\tau_1}^t \lambda_1 dt =$$
$$= \begin{cases} \lambda_0 \tau_1, & \text{for } 0 < t < \tau_1 \\ \lambda_0 \tau_1 + \lambda_1 t - \lambda_1 \tau_1 = \tau_1(\lambda_0 - \lambda_1) + \lambda_1 t & \text{for } t > \tau_1 \end{cases}$$
- survival function:
$$S(t) = \exp(-\Lambda(t)) = \begin{cases} \exp(-\lambda_0 t), & \text{for } 0 < t < \tau_1 \\ \exp[-\tau_1(\lambda_0 - \lambda_1) - \lambda_1 t] & \text{for } t > \tau_1 \end{cases}$$

Chapter 3

The theory of everyhting

3.1 Definition of something

Another chapter here. Note that you can have (As we did for chapter 1) as a separate tex file.

3.2 The answer is 42

3.2.1 And one subsection

3.2.2 A second one

Chapter 4

Simulations

4.1 Simulation for estimating something

4.1.1 Scenarios of Simulations

This is an example for a Table. See Table 4.1.

scenario	80% N	N	120% N
sc:1 - type I error	0.056	0.054	0.054
sc:2 - type I error	0.06	0.062	0.054
sc:3 - type I error	0.054	0.06	0.052
sc:4 - type I error	0.062	0.064	0.062
sc:5 - power	0.756	0.842	0.902
sc:6 - power	0.822	0.894	0.94

Table 4.1: A caption here

And an example of a figure. Figure **??** shows ...

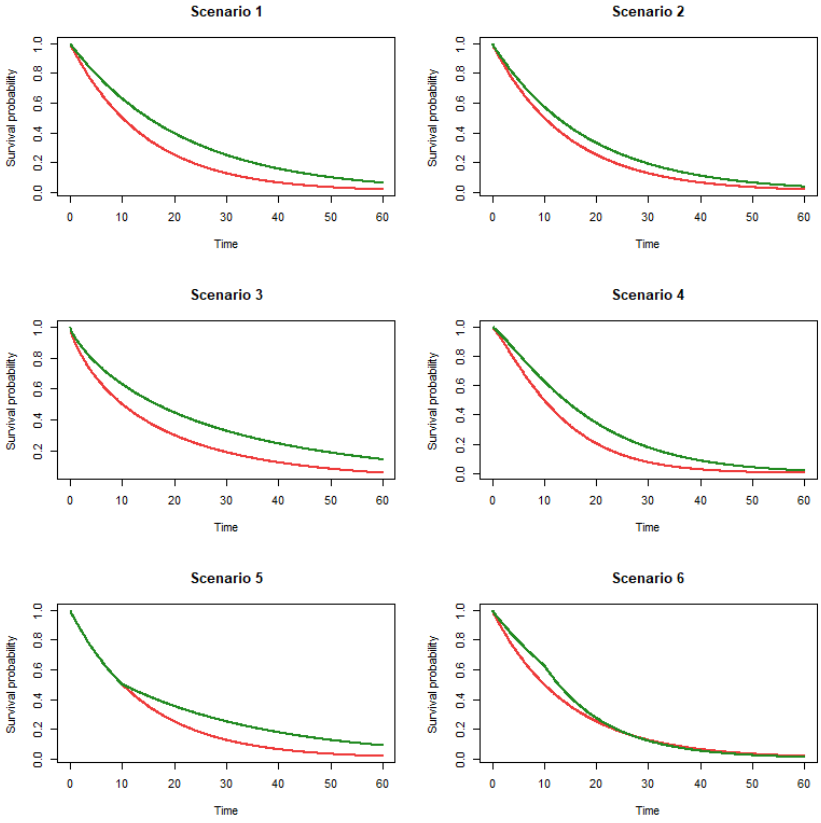

Figure 4.1: Survival curves of simulated scenarios comparing treatment arm (green line) with the control(red line)

Bibliography

Royston, P. and Parmar, M. K. (2013). Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Medical Research Methodology*, 13(1):1–15.

Weir, I. R. and Trinquart, L. (2018). Design of non-inferiority randomized trials using the difference in restricted mean survival times. *Clinical Trials*, 15(5):499–508.