NON-PARAMETRIC METHODS FOR COMPARING TWO SURVIVAL DISTRIBUTIONS

M.Ramakrishnan
Department of Mathematics
RKM Vivekananda College, Chennai, India

R.Ravanan
Department of Statistics,
Presidency College, Chennai, India

ABSTRACT

Survival curves for each group gives comparison at some arbitrary point(s) but it does not provide a comparison of the total survival experience of two groups. So some Non-parametric methods like log rank test, Cox-Mantel test are popular methods used for comparing the survival distribution and also these methods takes whole follow-up period. In this paper non-parametric methods are used for data from WHAS.

Keywords: Survival Probabilities, Non-Parametric test, Kaplan-Meier estimate, Log-rank test, Cox-Mantel test..

INTRODUCTION:

Survival analysis is the study about survival data's. Survival data's include Survival time, response to a given treatment and patient characteristics related to response, survival and the development of a disease. Survival Curves are generated by Kaplan-Meier method. Traditional Kaplan-Meier method used for finding survival probabilities for censored and non-censored observations. Survival to any time point is calculated as the product of the conditional probabilities of surviving each time interval. The calculations are simplified by ignoring censoring times. Survival curves give rough idea of the difference between the distributions. It does not reveal whether the differences are significant or not. So the Non parametric tests are necessary.
Non parametric test like log rank test is used to test the null hypothesis that there is no difference between the populations in the probability of an event at any time point. Non parametric test like Cox-Mantel test is used to test the better survival distribution between two survival distributions.

FUNCTION OF SURVIVAL TIME AND NON-PARAMETRIC TESTS:

SURVIVORSHIP FUNCTION:

Let T be the survival time and S (t) is the probability that an individual survives longer than t. i. e. $S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t),$

$Where\ F(t) is\ the\ distribution\ function\ of\ the\ time\ t$,

S (t) is non increasing function of time t with the properties
S (t) = 1 for t = 0.
S (t) = 0 for t = ∞.
S (t) is known as the cumulative survival rate. The graph of S (t) is called the survival curve.

KAPLAN-MEIER METHOD OF ESTIMATION:

Let n be the total number of individuals whose survival times, censored or not, are available. Relabeling the survival times in order of increasing magnitude such that $t_1 \leq t_2 \leq \cdots .. t_n$ and the values of r are consecutive integers 1,2,….,n if there are no censored observation . If there are censored observations, they are not. Then the survival probabilities are calculated using $S(t) = \prod_{t_r \leq t} \frac{(n-r)}{(n-r+1)}$, where r runs through those positive integers for which $t_r \leq t$ and t_r is uncensored. The variance of S(t) is approximated by

$$var(S(t)) = [S(t)]^2 \sum_r \frac{1}{(n-r)(n-r+1)}$$

Where r includes those positive integers for which $t_{(r)} \leq t$ and $t_{(r)}$ corresponds to a death. Estimated Standard error is $\sqrt{var(S(t))}$. A 95% confidence interval for S(t) is S(t) $\mp$1.96 S.E $[S(t)]$.

THE LOG RANK TEST:

Let d_t be the number of deaths at time t and n_{1t} and n_{2t} be the numbers of patients still exposed to risk of dying at time up to t in the two treatment groups. The expected deaths for groups 1 and 2 at time t are

$$e_{1t} = \frac{n_{1t}}{n_{1t}+n_{2t}} * d_t, \quad e_{2t} = \frac{n_{2t}}{n_{1t}+n_{2t}} * d_t$$

Then the total numbers of expected deaths in the two groups
$E_1 = \sum e_{1t}$ $E_2 = \sum e_{2t}$. Let O_1 and O_2 be the observed numbers and E_1 and E_2 the expected numbers of death in two treatment groups.

$The\ Test\ statistic\ \chi^2 = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2}$ has approximately the chi-square distribution with one degree of freedom. A large χ^2 value
(e.g., $\geq \chi^2_{1,.05}$) would lead to the rejection of the null hypothesis in favor of the alternative that the two treatments are not equally effective at
α= 0.05.

COX-MANTEL TEST:

Let $t_1 \leq t_2 \leq \cdots .. t_k$ be the distinct failure times in the two groups together and m_i the number of failure times equal to t_i, or the multiplicity of t_i. Let R(t) be the set of individuals still exposed to risk of failure at time t, whose failure or censoring times are at least t. Let n_{1t} and n_{2t} be the number of patients in R (t) belongs to group 1 and 2, respectively. The total number of observations, failure or censored in R (t_i) is $r_i = n_{1t} + n_{2t}$ *and* A_i is the proportion of r_i that belong to group 2.

$$Define\ U = r_{(2)} - \sum_{i=1}^{k} m_i A_i$$

$$I = \sum_{i=1}^{k} \frac{m_i(r_i - m_i)}{r_i - 1} A_i(1 - A_i),$$

Now Test Statistic

$$C = \frac{U}{\sqrt{I}} \text{ as a standard normal variate under the null hypothesis}$$

COMPUTATION AND CALCULATION:

To analysis these methods let us take Survival time for 6 males and 6 females from the Worcester Heart Attack Study (WHAS) given below.
Males 1, 3, 4+, 10, 12, 18
Females: 1, 3+, 6, 10, 11, 12+.
Right censored times are denoted by a "+" sign.

Table 1 : KM estimate for 12 Patients with 95% confidence limits

Time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	12	2	0.833	0.108	0.6470	1.000
3	10	1	0.750	0.125	0.5410	1.000
6	7	1	0.643	0.146	0.4119	1.000
10	6	2	0.429	0.157	0.2086	0.880
11	4	1	0.321	0.150	0.1287	0.803
12	3	1	0.214	0.133	0.0635	0.723
18	1	1	0.000	NaN	NA	NA

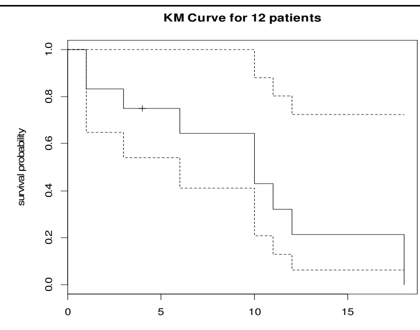


Table 1 : KM estimate for Male and Female Patients

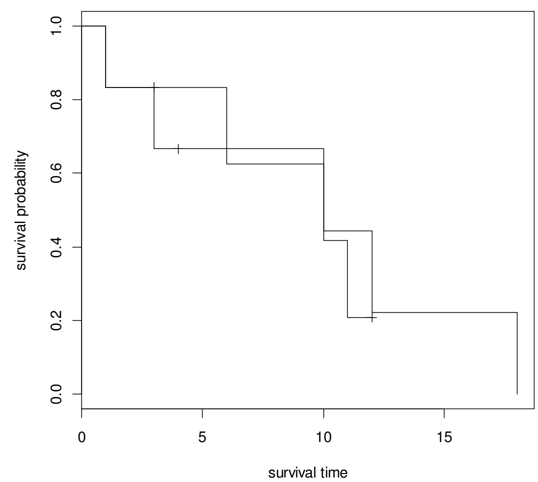
sex=Female

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	6	1	0.833	0.152	0.5827	1
6	4	1	0.625	0.213	0.3200	1
10	3	1	0.417	0.222	0.1468	1
11	2	1	0.208	0.184	0.0368	1

sex=Male

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	6	1	0.833	0.152	0.5827	1
3	5	1	0.667	0.192	0.3786	1
10	3	1	0.444	0.222	0.1668	1
12	2	1	0.222	0.192	0.0407	1
18	1	1	0.000	NaN	NA	NA

KM Curve for Male & Female patients



The survival Probability for the male survival time at t= 3 is obtained by S(3)= S(1)*4/5=0.6666 and S(10)= S(1)*S(3)*2/3=0.44444.

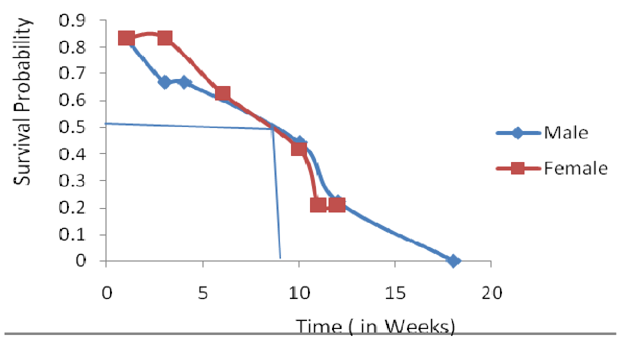


Fig. 2.1 survival curves for male and females patients

Computing the value for testing the two statistical distribution use Log-rank test. First we set null hypothesis as insignificant difference between male and female patient survival.

At survival time t = 3, number of death at t, $d_t = 1,$ Number of patients expected to death at time t for males is $n_{1t} = 5,$ Number of patients expected to death at time t for males is $n_{1t} = 5, e_{1t} = 0.5, e_{2t} = 0.5$

The same procedure of calculations is performed each time an event occurs. From the calculations for each time of death, the total number of expected death were 5.095238 for male patients and 3.904762 for females patients, and the observed number of death were 5 and 4. Now the value of χ^2 is 0.004103. The degrees of freedom are the number of groups minus one, i.e. 2 – 1 = 1. From a table of the χ^2 distribution the value of p is 0.9489 > 0.01. So that the difference between male and female is statistically insignificant.

When using Cox-Mantel test, In this problem, There are, k=7 distinct failure times in the two groups male and female, $r_{(1)}$=5, $r_{(2)}$=4. At failure time

t = 3, number of times death occurred at t is m(t)= 1, Number of risk set in male patients is n_{1t}=5, Number of risk set in female patients is n_{2t}=5, Total number of risk patients in both male and female is r(t) = 10, A(t)=5/10=0.5.

The same procedure of calculations is performed each time an event occurs. For comparison here $r_{(2)}$=4, the value of U is obtained as 0.0953 and I = 1.8216. Now Test Statistic C = 0.0706. The Corresponding p value is 0.4719 > 0.01. So in this method also accept null hypothesis. So that the difference between male and female is statistically insignificant.

CONCLUSION:

From the survival curves in fig 2.1, in early period, the probability of survival of male patients is greater than female patients. But in the late period probability of survival of both are same and Median survival time for both groups are observed as 8.5 weeks. But this approach does not provide a comparison of the total survival experience of male and female groups. When using log rank test and Cox-Mantel test, we observed that there is no significant difference between male and female patients survival. These tests are most likely to detect a difference between groups when the risk of an event is consistently greater for one group than another. Survival curves only gives pictorial representation but using non-parametric tests like log rank test and Cox-Mantel test, we can easily identify the relation of survival distribution.

REFERENCES:

[1] Altman DG, Bland JM.(1998),” Time to event (survival) data”. British Medical Journal (BMJ); vol. 317: pp. 468-469.

[2] Altman DG, Bland JM.(1998), “Survival probabilities (The Kaplan-Meier method)”. BMJ; vol. 317; pp. 1572 - 1580.

[3] Altman DG, Bland JM. (2004),” The log rank Test”, BMJ; vol. 328 ; pp. 1073.

[4] Collett D. (2003), “Modeling of Survival Data in Medical Research”. Chapman Hall, London, U.K.

[5] Elisa T. Lee. (1992), “Statistical methods for Survival Data Analysis”. Second Edition. A Wiley-Inter science publication, United States of America.

[6] Kaplan, E. L., and Paul Meier (Jun. 1958),” Non parametric Estimation from Incomplete Observations”, Journal of the American Statistical Association. Vol. 53, pp457 – 481.

[7] Rupert G. Miller J R (1981), “Survival Analysis”, John Wiley & Sons, United States of America.
