Pak. J. Statist.
2005 Vol. 21(2) pp 159-165

WEIGHTED SCORE TEST FOR COMPARING THE SURVIVAL DISTRIBUTION IN CASE OF HEAVY CENSORING

Bahrawar Jan, S. Waqar Ali Shah, Mohammad Iqbal
Department of Statistics, University of Peshawar
Peshawar, Pakistan
and
Iftikhar-ud-Din
Govt. Post Graduate College Mansehra
Mansehra, Pakistan

ABSTRACT

To compare the survival distribution, the class of weighted mental Haenszal test is used in the data having censored observations. In this article a test statistic is proposed to compare the survival distributions of two groups in case of censoring times. The test is based on weights introduced for the censoring times. A real data set is used to compare the proposed test statistics with the existing log-rank family tests. The proposed test performs better to detect the difference in the groups survival distribution as compared to Breslow, Torane-Ware and the Peto's and Printice tests.

1. INTRODUCTION

In medical research, there is often a need to compare survival distributions. For instance researchers may wish to compare the survival times of patients in one age group to those in another, or to compare two groups exposed to different risk factors.

We are then interested in testing;

$$H_0:S_1(t)=S_2(t) \quad \text{against} \quad H_1:S_1(t)>S_2(t)$$
$$\text{or} \quad H_1:S_1(t)\neq S_2(t)$$

The non-parametric tests in the class of Weighted Mental Haenszal can be used to compare survival distribution for censored data.

Suppose there are two groups with m_1 and m_2 observations respectively. Let $x_1, \ldots\ldots, x_j$ failure observations and $x^+_{j+1},\ldots\ldots, x^+_{m_1}$ be the $m_{1}-j_1$ censored observations in group 1 and let $y_1, \ldots\ldots\ldots, y_{j2}$ be the j_2 failure observations and $y^+_{j2+1},\ldots\ldots\ldots, y^+_{m2}$ be the $m_{2}-j_2$ censored observations in group 2.

In Gehan's generalized Wilcoxon (Gehan 1965) or Breslow test, every observation (failed or censored) of one group is compared with every observation of the other group and a score w_{ij} is computed for the result of every comparison, where

$$W_{ij} = \begin{cases} +1 & x_i > y_j \text{ or } x_i^+ \geq y_j \\ 0 \quad \text{if} & x_i = y_j \text{ or } x_i^+ < y_j \text{ or } y_j^+ < x_i \text{ or } x_i^+ < y_j^+ \\ -1 & x_i < y_j \text{ or } x_i \leq y_j^+ \end{cases} \tag{1.1}$$

The test statistic to be used is:

$$W = \sum_{i=1}^{m_1} \sum_{j=2}^{m_2} w_{ij} \tag{1.2}$$

Where the sum is over all $m_1 m_2$ comparisons. Hence there is a contribution to the test statistic W for every comparison, where both observations are failures (except for ties) and for every comparison where a censored observation is equal to or larger than a failure. When m_1 and m_2 are large, an alternative method based on the pooled sample is adopted for calculating W (Mantel, 1967). In the two samples are combined into a single pooled sample of m_1+m_2 observations, it is the same as comparing each observation with the remaining $m_1+m_2 - 1$. Then $W = \sum_{i=1}^{m_1} w_i$.

The variance has been shown by Mantel (1967) to be:

$$\text{var}(W) = \frac{m_1 m_2 \sum_{i=1}^{m_1+m_2} w_i^2}{(m_1 + m_2)(m_1 + m_2 - 1)} \tag{1.3}$$

For the Cox-Mantel test (Cox 1959, 1972, Mantel 1966), let $t_1 < t_2 < \ldots\ldots\ldots\ldots < t_k$ be the distinct failure times in the two groups together and n_{i1} be the number of failure times equal to t_i, and a_{i1} the number of deaths (failures) in group 1. Let m_{i1} and m_{i2} be the number of patients at the ith failure time that belong to group 1 and group 2 respectively. For each of the distinct failure times, the data can be represented by a 2×2 contingency table:

Groups	**No. of Deaths**	**No. of Survivals**	**Total**
Group 1	a_{i1}	$m_{i1} - a_{i1}$	$\mathbf{m_{i1}}$
Group 2	a_{i2}	$m_{i2} - a_{i2}$	$\mathbf{m_{i2}}$
Total	$\mathbf{n_{i1}}$	$\mathbf{n_i - n_{i1}}$	$\mathbf{n_i}$

The Cox-Mantel test statistic, also referred to as logrank test (Mahesh K.B. and David M. 1995), is given by:

$$C = \frac{\sum_{i=1}^{k} a_{i1} - \sum_{i=1}^{k} E(A_{i1})}{\left[\sum_{i=1}^{k} \text{Var}(A_{i1})\right]^{1/2}}, \tag{1.4}$$

where,

$$E(A_{1i}) = \frac{m_{i1} n_{i1}}{n_i}, \tag{1.5}$$

and

$$\mathrm{var}(A_{1i}) = \frac{m_{i1} m_{i2} n_{i1} (n_i - n_{i1})}{n_i^2 (n_i - 1)}. \tag{1.6}$$

It is worth noting that that C^2 corresponds to the Mantel-Haenszel test statistics (Mantel and Haenszel, 1959), however, the later requires independence between the individual contingency tables. The Cox-Mental Test does not require such independency.

The Tarone-Ware (1977) class of statistics has the following form:

$$TW = \frac{\sum_{i=1}^{k} w_i \left[a_{i1} - E_0(A_{i1)})\right]}{\left[\sum_{i=1}^{k} w_i^2 \,\mathrm{var}_0(A_{i1})\right]^{1/2}}, \tag{1.7}$$

where w_i are constants used to weight the respective tables. Note that TW statistic includes the Cox-Mantel Statistic as a special case for w_i=1.

A further set of weights has been suggested by both Peto and Prentice. Their test is;

$$PP = \frac{\sum_{i=1}^{k} w_{pp} \left[a_{i1} - E_0(A_{i1)})\right]}{\left[\sum_{i=1}^{k} w_{pp}^2 \,\mathrm{var}_0(A_{i1})\right]^{1/2}}, \tag{1.8}$$

where $W_{pp} = \prod_{j=1}^{k} \left\{ \frac{n_j - d_j - 1}{n_j + 1} \right\}$.

The weighting for this test is similar to the corresponding estimate of Kaplan Meier survival function at each point.

Mantel (1966) generalized the logrank test (Savage, 1956). This generalization is based on a set of scores, w_i, assigned to the observations. The scores are functions of the logarithm of the survival function. To estimates the log survival function at t_i, using

$$e(t_i) = \sum_{t_j \le t_i} \frac{n_{j1}}{n_j}. \tag{1.9}$$

The scores suggested by Peto and Peto are $w_i = 1 - e(t_i)$ for an uncensored observation t_i and $-e(T)$ for an observation censored at T. In practice, for a censored observation t_i^+, $w_i = -e(t_j)$, where t_j is the largest uncensored observation such that $t_j \le t_i^+$. Thus the largest the uncensored observation, the smaller its score. Censored observations receive negative scores. The w scores sum identically to zero for the two groups together. The logrank test is based on the sum S of the w scores in one of two groups i-e

$$S = \sum_{i=1}^{k} a_{i1} - \sum_{i=1}^{k} \frac{m_{i1} n_{i1}}{n_i}$$

The Cox-Mantel test statistic can be written as:

$$C = \frac{S}{\left[\sum_{i=1}^{k} \operatorname{var}(A_{i1})\right]^{1/2}} . \tag{1.10}$$

A generalization of the Wilcoxon two-sample rank test is described by Peto and Peto (1972). Similar to the logrank test, this test assigns a score to every observation. For an uncensored observation t_i, the score is $u_i = \hat{S}(t+) + \hat{S}(t-) - 1$, , and for an observation censored at T, the score is $u_i = \hat{S}(T) - 1$, where $\hat{S}$ is the Kaplan-Meier estimate of the survival function. Using the same notation as used for Cox-Mantel test, the score for an uncensored observation t_i is $u_i = \hat{S}(t_i) + \hat{S}(t_{i-1}) - 1$ and $\hat{S}(0) = 1$ and that for a censored observation t_j^+ is $u_j = \hat{S}(t_i) - 1$, where $t_i < t_j^+$. These generalized Wilcoxon scores sum to zero. The test procedure after the scores are assigned is the same as for log-rank test.

2. PROPOSED TEST STATISTIC

The tests outlined in the introduction are developed only for the failure time comparison. We propose a new weighted score test to compare the survival times at both censored or failure times. Suppose there are two groups with having m_1 and m_2 observations. Let x_1, ……., x_j failure observations and $x_{j+1}^+, \ldots, x_{m_1}^+$ be the $m_1 - j_1$ censored observations in group 1 and let y_1, ……, y_{j2} be the j_2 failure observations and $y_{j2+1}^+, \ldots, y_{m2}^+$ be the $m_2 - j_2$ censored observations in group 2.

Let let $t_1 < t_2 < \ldots < t_k$ be the distinct survival times (failure or censored times) in the two groups together and D_i be the number of failure as a whole at time t_i, and d_{i1} the number of deaths (failures) in group 1. Let c_{i1} be the number censored at time t_i in group 1 and $\mathbf{C_i}$ be the number censored as a whole. Let n_{i1} and n_{i2} be the number of patients at the ith survival time that belong to group 1 and group 2 respectively.

For each of the distinct failure times, the data can be represented by a 2×3 contingency table:

Groups	No. of Deaths	No. of censored	No. of Survivals	Total
Group 1	d_{i1}	c_{i1}	l_{i1}	$\mathbf{n_{i1}}$
Group 2	d_{i2}	c_{i2}	l_{i2}	$\mathbf{n_{i2}}$
Total	$\mathbf{d_i}$	$\mathbf{c_i}$	$\mathbf{l_i}$	$\mathbf{n_i}$

The new test statistics is based on the weights developed as a function of the number censored as $W_i = \left\{\frac{n_{i1} - c_{i1}}{n_{i1}}\right\}$ known as non-censored rate. Where $W_i = 1$ for no censoring and $W_i < 1$ in case of censored observation at time t_i

Then proposed weighted scores test is then based on the score given as;

$$S_w = \sum_{i=1}^{k} W_i d_{i1} - \sum_{i=1}^{k} W_i \frac{d_i n_{i1}}{n_i} \tag{2.1}$$

The score value will be negative if there are no deaths at time t_i, that is

$$S_w = -\sum_{i=1}^{k} W_i \frac{d_i n_{i1}}{n_i} \tag{2.2}$$

The test statistic can be written as;

$$Cw = \frac{S_w}{\left[\sum_{i=1}^{k} W_i^2 \operatorname{var}(d_{i1})\right]^{1/2}}, \tag{2.3}$$

where

$$\operatorname{var}(d_{1i}) = \frac{n_{i1} n_{i2} d_i (n_i - d_i)}{n_i^2 (n_i - 1)}. \tag{2.4}$$

This test statistic also comes in the family of Tarone-Ware class of statistics. i.e.

$$Sw = \frac{\sum_{i=1}^{k} W_i \left[d_{i1} - E_0(d_{i1)})\right]}{\left[\sum_{i=1}^{k} W_i^2 \operatorname{var}_0(d_{i1})\right]^{1/2}} \tag{2.5}$$

where W_i are constants used as weight for censored time at $t_{i.}$. Note that if the weights are fixed to 1 then all test statistics are similar to Cox-Mantel Statistic. It happens only when there is no censoring.

It is assumed that this test statistics has a chi square distribution with 1 d.f.

3. COMPARISON OF THE TESTS BY APPLICATION TO REAL DATA

The Stanford Heart Transplant data are a classic survival data set with time dependent covariates. The well known data set has been considered and analyzed several times. The data set contain 103 patients 69 of them received transplants. The censoring rate is 27%. The response variable is Survival time. The observed survival time is indicated to be censored or uncensored by the survival status.

We make use of the proposed weighted score test on this data and compare it with the traditional tests in the class of weighted Mantel-Haenzsal test. Mostly the researchers considered age 40 years as the cutting point for the survival time of the patients. We considered the patients of age below 40 as group-I and above as group-II for the patients under investigation. Table-1 contain the comparison of these two groups, with the help of various tests i.e. Breslow, Torane-Ware, Peto's and Printice (Mahesh and David, 1995) and in comparison with the proposed one.

Table-1:
Comparison of Test Statistics

Test-Statistics	D.f	Chi-sq	p-values	Cum. Prob.
Breslow	1	0.06840	0.793681	0.206319
Torane-Ware	1	0.00126	0.971684	0.028316
Peto & Printice	1	1.95700	0.161835	0.838165
Proposed Weighted Scores-Test	1	2.13000	0.144441	0.855559

The Breslow uses the number at risk as weights for the difference of observed and estimated deaths, gives the chi square value as 0.0684 with p-value 0.794 fails to detect the difference. The Torane-Ware test uses the square root of the number at risk as weights for the difference of observed and estimated deaths, gives the chi square value as 0.001264 with p-value 0.972 also fails to detect the difference. The Peto's and Printice uses the Kaplan-Meier survival function as weights for the difference of observed and estimated deaths, gives the chi square value as 1.957 with p-value 0.162 has an improved value than the other two test to detect the difference. The proposed weighted score test uses the proposed Weighted Kaplan-Meier survival function as weights for the difference of observed and estimated deaths, gives the chi square value as 2.13 with p-value 0.144 is the most sensitive in comparison to the others to detect the difference between the two groups.

In contrast, both the Peto's and Printice and the proposed methods are more accurate than others in case of heavy censoring. Of these two, the proposed method is still performing slightly better, although the difference is perhaps less.

4. CONCLUSIONS

The proposed test may be considered in the class of weighted Mental Haenzsal tests. This test uses weights introduced for the weighted Kaplan Meier survival curves (Jan, B. 2004) as the weights for the difference of observed and estimated events occurrence.

The Stanford Heart Transplant data is used to compare the proposed test statistics with the class of weighted Mental Haenzsal score tests. The data is divided in to two groups by the cutting age point of 40 years of age recommended by many authors. The Breslow, Torane-Ware and the Peto's and Printice tests were failed to detect the difference in the survival distributions of the two groups. The proposed weighted score test is more closer to detect the difference having a p-value smaller than all others.

ACKNOWLEDGEMENT

We are thankful to the referee for reviewing the paper and suggested good points.

REFERENCES

1. Cox, D.R (1972). Regression Models with Life Tables (with discussion). *J. Roy. Statist. Soc.*, Series B, 34: 187-220.
2. Cox, D.R. (1959). The Analysis of Exponential Distribution Life-times with Two Types of Failures. *J. Roy. Statist. Soc.*, Series B, 21: 411-421.
3. Gehan, E.A. (1965). A Generalized Wilcoxon Test for Computing Arbitrarily Single-censored Samples. *Biometrika*, 52: 203-223.
4. Mahesh K.B. Parmar and David M. (1995). *Survival Analysis: A Practical Approach*. John Wiley & sons, England.
5. Mantel, N. (1966). Evaluation of Survival Data and Two New Rank Order Statistics Arising in its Consideration. *Cancer Chemotherapy Reports*, 50: 163-170.
6. Mantel, N. (1967). Ranking Procedures for Arbitrarily Restricted Observations. *Biometrics*, 23: 65-78.
7. Peto, R. and Peto, J. (1972). Asymptotically Efficient Rank Invariant Test Procedures (with discussion). *J. Roy. Statist. Soc.*, Series A, 135: 186-206.
8. Savage, I.R. (1956). Contribution to the Theory of Rank Order Statistics the Two Sample Case. *Ann. Math. Statist.*, 27: 590-615.
9. Tarone, R.E. and Ware, J. (1977). On distribution-free tests for equality of survival distributions. *Biometrika* 64.