Bonferroni’s method to compare Survival Curves with Recurrent Events

Método de Bonferroni para comparar curvas de supervivencia con eventos recurrentes

Carlos M. Martínez M.

Key words: Survival Analysis, Recurrent Events, Statistical Tests, Bonferroni’s Method

Palabras Clave: Análisis de supervivencia, Eventos Recurrentes, Pruebas Estadísticas, Método de Bonferroni

ABSTRACT

Survival analysis provides useful statistical tools to study the occurrence of events such as epileptic seizures, viral diseases, malignant tumors, re-hospitalization of patients among others. This paper offers statistical tests of interest to epidemiologists, biostatisticians and other researchers interested in new methodologies to be applied in medicine and other related health research areas. The main objective is to propose a new procedure based on Bonferroni’s method to compare survival curves of population groups with recurrent events. Nonparametric comparison tests with recurrent events data are illustrated. The idea is applying this methodology with a sequential procedure controlling the error type I on multiple contrast tests. The hypothesis test is:

\[ H_0 : S_1(z) = S_2(z) = \ldots = S_k(z), \]

\[ H_1 : \text{At least one } S_r(z) \text{ is different with } r = 1, 2, \ldots, k. \]

Where, \( S_r(z) \) is the survival curve of the \( r \)-th group and its estimation is calculated by the Generalized Product Limit Estimator (GPLE or Kaplan-Meier estimator). The survival functions are estimated using R-language programs and counting processes.

INTRODUCTION

Recent developments on Survival Analysis (SA) have extended to include recurrent events. SA has evolved into particular applications that include many areas of science: economy, social science, astronomy, sociology, psychology, demography, medicine, engineering and others. Engineering and Medicine are the areas where SA has been traditionally applied. Its applications are known as Reliability Studies in Engineering and Survival Studies in Medicine. These applications basically are a model in time (T) from a given starting point until the occurrence of an event, being its main objectives the modeling functions such as density, cumulative distribution, survival, instantaneous hazard and cumulative risk functions, and the comparison of the mentioned functions taking into account the variations in population groups. Counting processes have benefited the modeling of the survival analysis functions. It is relatively recent the development of new procedures and tools of the survival analysis with the use of
counting processes. Highly specialized literature on this research area is not fully known; see Cook & Lawless (2007), Kalbfleisch & Prentice (2002) and Hougaard (2000). Also, martingales and stochastic integration were introduced in this area. Aalen (1975), Gill (1981), Therneau et al. (1990), Fleming & Harrington (1991), Andersen et al. (1975), Therneau & Hamilton (1997), Therneau & Grambsch (2000) and more recent Peña et al. (2001) have introduced and used these concepts in the modeling of the survival analysis. STATA, NCSS, SAS, S-Plus and R are programs where have been designed and adapted some of their routines to this topic.

Prentice et al. (1981), Andersen & Gill (1982) and Wei et al. (1989) are authors that have made some of the first studies in the modeling of survival analysis functions with recurrent event. Later, other authors like Wang & Chang (1999), Peña et al. (2001) and Peña & Slate (2005) have also done relevant investigations. In relation to the comparison of groups on the survival analysis with recurrent events Pepe & Cai (1993), Doganaksoy & Nelson (1998), Nelson (2003), Martínez (2009) and Martínez et al. (2009, 2011 and 2012) have made important contributions. In this research, issues of recurrent events data modeling and statistical tests are illustrated. A new methodology for comparing k survival curves with recurrent events using classical Bonferroni’s method is developed.

**METHODS**

GPLE model to estimate survival curves with recurrent event

Peña et al. (2001) developed a survival function estimator to recurrent events under the assumption of independent interoccurrence times and identically distributed, (IID). This estimator generalized the classical estimator of Kaplan & Meier (1958) called Generalized Product Limit Estimator (GPLE). They used two counting processes N and Y. The original idea of use counting process in Survival Analysis comes from Gill (1981) and it has been extended by Peña et al. (2001). Below, the survival function estimator is showed,

\[
\hat{S}(t) = \prod_{t \leq t_i} \left[ 1 - \frac{\Delta N(s, z)}{Y(s, z)} \right] 
\]

The authors considered two time scales: one related to calendar time (s) and other related to inter-occurrences time (z). N(s,z) represents the observed events number over the calendar period [0,s] whose inter-occurrence times were at most T (T ≤ z), and Y(s,z) represents the observed events number in the period [0,s] with T ≥ z. In the definition, S(z) is the survival function estimator, n is the units number, Tij is the jth inter-occurrence time in the ith unit, Sij is the jth calendar time in the ith unit, index i is the observation time of the ith unit, ki(s) is the occurrence total number of the event in the ith unit over the calendar period [0,s], see González & Peña (2003). The Tij’s are independent and identically distributed (IID) from an unknown distribution function F and the τis are IID from a distribution function G unknown too.

Table 1 shows the variables observations vector for each units of the study.

The authors defined N and Y as,

\[
N(s, z) = \sum_{i=1}^{n} \sum_{j=1}^{k_i(s)} I\{T_{ij} \leq z\} 
\]

Where, \( I\{\cdot\} \) is an indicator function that takes value one if the condition is true and value zero if it is false.

\[
N(s, z + \Delta z) = \sum_{i=1}^{n} \sum_{j=1}^{k_i(s)} I\{T_{ij} \leq z + \Delta z\} 
\]

\[
\Delta N(s, z) = N(s, z + \Delta z) - N(s, z) 
\]

Now, if \( z \to 0 \)

\[
\Delta N(s, z) = \sum_{i=1}^{n} \sum_{j=1}^{k_i(s)} I\{T_{ij} = z\} 
\]

\[
k_i(s) = \sum_{j=1}^{\infty} I\{S_{ij} \leq s\} \forall i = 1, 2, ..., n 
\]

Martínez. Bonferroni’s method to compare Survival Curves with Recurrent Events, p. 105-114
Furthermore,\n\[
Y(s, z) = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{k_i(s)} I(T_{ij} \geq z) + I(\min(s, \tau_i) - S_{ik_i(s)} \geq z) \right\}
\]
\[(7)\]

González et al. (2005) developed a package called survrec. This package is available at CRAN, \url{http://www.r-project.org/}, it is used to compute the survival function of Peña et al. (2001).

Figure 1 shows a hypothetical case of a patient with three recurrences at months 110, 185 and 280 from the beginning of the study and the last observation is a censored data by the right. For a calendar time scale \( s = 200 \) and an inter-occurrence time \( z = 100 \), notice that \( N(s = 200, z = 100) = 1 \) and \( Y(s = 200, z = 100) = 1 \).

Table 1. Observations vector of the variables in the units

<table>
<thead>
<tr>
<th>Unit</th>
<th>Observation period</th>
<th>Number of occurrence</th>
<th>Inter-occurrence times</th>
<th>Calendar times</th>
<th>Censored time</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>( \tau_i )</td>
<td>( k_i(s) )</td>
<td>( T_{ij} ), ( j = 1, 2, \ldots, i_k(s) )</td>
<td>( S_{ij} ), ( j = 1, 2, \ldots, i_k(s) )</td>
<td>( \tau_i - S_{ik_i(s)} )</td>
</tr>
<tr>
<td>1</td>
<td>( \tau_1 )</td>
<td>( k_1(s) )</td>
<td>( T_{11}, T_{12}, \ldots, T_{1k_1(s)}(s) )</td>
<td>( S_{11}, S_{12}, \ldots, S_{1k_1(s)}(s) )</td>
<td>( \tau_1 - S_{1k_1(s)} )</td>
</tr>
<tr>
<td>2</td>
<td>( \tau_2 )</td>
<td>( k_2(s) )</td>
<td>( T_{21}, T_{22}, \ldots, T_{2k_2(s)}(s) )</td>
<td>( S_{21}, S_{22}, \ldots, S_{2k_2(s)}(s) )</td>
<td>( \tau_2 - S_{2k_2(s)} )</td>
</tr>
<tr>
<td>\vdots</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
</tr>
<tr>
<td>n</td>
<td>( \tau_n )</td>
<td>( k_n(s) )</td>
<td>( T_{n1}, T_{n2}, \ldots, T_{nk_n(s)}(s) )</td>
<td>( S_{n1}, S_{n2}, \ldots, S_{nk_n(s)}(s) )</td>
<td>( \tau_n - S_{nk_n(s)} )</td>
</tr>
</tbody>
</table>

Figure 1. Counting process illustration for a hypothetical case. Source: González (2005)

GPLE model to estimate survival curves with recurrent event in the groups

The GPLE estimator in the groups is defined with a fundamental assumption of this approach, the units have been previously and properly classified in the groups according to a stratification variable denote by \( r \). In this way, we can estimate the survival function of each group shows:

\[
\hat{S}_r(z) = \prod_{s \leq z} \left[ 1 - \frac{\Delta N(s, z; r)}{Y(s, z; r)} \right] \quad \forall \ r = 1, 2
\]
\[(8)\]

Notice that all definitions eqs.: from (1) to (7) made in the investigation of Peña et al. (2001) are extensive to the estimations of survival functions in the groups. So, \( N(s, z; r) \) represents the observed

Martínez. Bonferroni’s method to compare Survival Curves with Recurrent Events, p. 105-114
events number for the rth group in the calendar period [0,s], whose inter-occurrence times were at most T and Y(s,z;r) represents the observed events number for the rth group in the period [0,s] with T ≥ z.

**Test to compare two survival curves with recurrent events**

Martinez (2009) and Martinez et al. (2009, 2011 and 2012) published hypothesis tests to compare survival curves of groups with recurrent events. The hypothesis of the test for two groups is:

\[ H_0 : S_1(z) = S_2(z) \]
\[ H_1 : S_1(z) \neq S_2(z) \]

Where, \( S_1(t) \) and \( S_2(t) \) are the group survival curves, respectively. The statistic test is:

\[
Z = \frac{\sum_{s \leq t} w_z \{ \Delta N(s, z; 1) - E[\Delta N(s, z; 1)] \}}{\sqrt{\sum_{s \leq t} w_z^2 Var[\Delta N(s, z; 1)]}}
\]

(9)

The statistic Z is asymptotic normally distributed and its square has a chi-square distribution with one freedom degree. Where, \( \Delta N(s, z; r) = N(s, z + \Delta z; r) - N(s, z; r) \) and \( N(s, z; r) \) has an hypergeometric behavior with expected value equal to \( Y(s, z; r) \times N(s, z) / Y(s, z) \) and variance equal to:

\[
Var[\Delta N(s, z; 1)] = \frac{Y(s, z) \times Y(s, z + 1)}{Y(s, z) - 1} Y(s, z; 1)^{\Delta N(s, z; 1)} \frac{N(s, z)}{N(s, z + 1)} \]

(10)

It is proposed the weights function \( w_z \), where,

\[
w_z = [S_c(z)]^\alpha [1 - S_c(z)]^\eta \frac{[Y(s, z)]^\alpha}{[Y(s, z + 1)]^\eta} \]

(11)

Table 2 shows the weights to SA.

<table>
<thead>
<tr>
<th>Test type</th>
<th>Test</th>
<th>Weight ( w_z )</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mantel-Haenszel</td>
<td>LRrec</td>
<td>1</td>
<td>Constant</td>
</tr>
<tr>
<td>Gehan</td>
<td>Grec</td>
<td>( Y(s, z) )</td>
<td>Decreasing</td>
</tr>
<tr>
<td>Tarone-Ware</td>
<td>TWrec</td>
<td>( [Y(s, z)]^{1/2} )</td>
<td>Decreasing</td>
</tr>
<tr>
<td>Peto-Peto</td>
<td>PPrec</td>
<td>( S(z) )</td>
<td>Decreasing</td>
</tr>
<tr>
<td>Fleming-Harrington</td>
<td>FHrec</td>
<td>( [S(z)]^\gamma [1 - S(z)]^\eta )</td>
<td>...</td>
</tr>
<tr>
<td>Martinez</td>
<td>CMrec</td>
<td>( [S(z)]^\gamma [1 - S(z)]^\eta )</td>
<td>...</td>
</tr>
</tbody>
</table>

Martinez (2012) developed a package called TestSurvRec (available at CRAN, http://www.r-project.org/) to compute the groups survival function. With the package TestSurvRec we can compute the p-values of these tests. The appropriate choice of weights depends on the curves behavior and the stable of its estimations. One appropriate selection of weights will allow put more importance on certain parts of the curves. Notice that, depending on the values of the parameters (\( \alpha, \beta, \gamma \) and \( \eta \)) the CMrec test is able to generate other tests to survival analysis with recurrent events, tests type: LRrec, Grec, TWrec, PPrec, FHrec and others. If the observation time of the unit is a large period time and the event occur only once in the unit; we are in presence of the traditional survival analysis and CMrec test is able to generate classical tests for this analysis. In CMrec test, if all parameters are zero, it implies that \( wz = 1 \), CMrec test generates the test type logrank. Now, if \( \alpha = 1 \) and the other parameters are zero, it implies that \( wz = Y(s, z) \), CMrec test generates the test type Gehan. If \( \gamma = 1 \) and the other parameters are zero, it implies that \( wz = S(z) \), CMrec test generates the test of Peto & Peto. If \( \gamma = 1 \) and \( \eta = 1 \) and the rest of the parameters are zero, CMrec test generates Fleming & Harrington test. So, we can conclude that these latter are particular cases of the proposal of Martinez (2009). All these tests are useful diverse fields such as medicine, engineering, social science and others.

**Martinez. Bonferroni’s method to compare Survival Curves with Recurrent Events, p. 105-114**
Problem
Suppose that you want to compare k survival curves of groups with recurrent events, where the error type I is equal to $\alpha$. The overall hypothesis test is:

$H_0: S_{i}(z) = S_{j}(z) = ... = S_{k}(z)$

$H_1: \text{At least one } S_{r}(z) \text{ is different with } r = 1, 2, ..., k.$

You can solve this common problem in two ways: One is using CMrec tests to compare k survival curves with recurrent events, see Martinez et al. (2009, 2011 and 2012) and other is with Bonferroni’s method. In this paper we propose and use Bonferroni’s method. Although, we should clarify that for groups with highly correlated data, this methods are inappropriate. In this paper, we assumed independents data.

Bonferroni’s algorithm. A proposal
In this investigation, we propose to apply the sequential procedure of Bonferroni to compare survival curves. The idea consists in use the procedure to make multiple contrasts controlling the error type I. Where, the null hypothesis of the tests is,

$H_{0i}: S_{r}(z) = S_{r'}(z)$

$H_{1i}: S_{r}(z) \neq S_{r'}(z)$

$H_0$ represents each one of the multiple hypothesis tests in the equality comparison of the curves in pairs, with $i = 1, 2, ..., q$ and $q = k \times (k-1)/2$. Notice that, $r \neq r'$; with $r = 1, r = 2, ..., r = k - 1$ and $r' = r + 1$, $r + 2, ..., k$. The hypothesis $H_0$ is true if all $H_0$ are true or if the pair of groups with more difference are equals. So, all k survival curves are equal. Bonferroni’s method is a procedure with control type I error. Multiple contrasts are made to compare survival curves in pairs. $H_0$ is true ($H_0$ is rejected) if at least $H_0$ is false with $i = 1, 2, ..., q$. Each hypothesis $H_0$ is $Sr(z) = Sr'(z)$ and $p-value$ is $p_{0i} = \text{Prob}(X_{i}^{2} > \chi_{1,1-\alpha}^{2})$. The statistic contrast tests for comparison of the hypothesis were proposed by Martinez et al.. Where, the statistic contrast $X_{0i}^{2}$ is the observed value. Now, if $X_{0i}^{2}$ is greater than the critical value if $X_{1,1-\alpha}^{2}$ it is because the $p-value$ is less than $\alpha$, and consequently rejects the null hypothesis $H_0$. Bonferroni proposed as critical values in each test must be equal to alpha divided by the number of tests, he suggests $\alpha/q$ for all multiple contrasts. Holm (1976) proposed as critical values: $\alpha/q, \alpha/(q-1), ..., \alpha/(q-i+1), ..., \alpha$. This author also showed that the overall significance level is $\alpha$, and that the power is higher than the classical Bonferroni procedure. Figure 2 shows Bonferroni’s procedure to compare survival curves.

Bonferroni’s procedure
1. Start by comparing the groups in pairwise. This method have a total number of tests ($q$) equals to $k \times (k -1)/2$. The total number of tests increases if the total number of groups ($k$) increases. To maintain order comparison could be made of a sequence as indicated: (1,2), (1,3), ..., (1, k), (2, 3), (2, 4), ..., (2,k), (3, 4), ..., (3,k), ..., (k -1, k). The pair $(r, r')$ indicates that $r$ group is contrasted with $r'$ group. In this method, we should use the tests proposed by Martinez et al. (2011)
2. Ordering the $p-values$ obtained: $p_{01} \geq p_{02} \geq ... \geq p_{0q}$.
3. Do, $r = q$.
4. Do, $i = r$.
5. Compare $p_{0i}$ with the Bonferroni critical value, $\alpha/q$.
6. If $p_{0i}$ is higher than $\alpha/q$ and the hypotheses: $H_{0i}, H_{0i-1},..., H_{01}$ are accepted. Finish it.
7. If $p_{0i}$ is less than $\alpha/q$ and the hypotheses $H_{0i}$ is rejected.
8. Do, $r = r - 1$.
9. If $r \neq 0$, go step 4.
10. If $r = 0$, finish the procedure.

APPLICATION
In the experiment Byar (1980) were measured the tumor recurrence time (months) of one hundred sixteen (116) sick patients with superficial bladder cancer. These patients underwent a process of randomization on assignment in the treatments: placebo (47 patients), pyridoxine (31 patients) and
thiotepa (38 patients). [See Andrews & Herzberg, (1985)]. For this data, the estimations of the survivor functions are made. Figures 3 and 4 show the graphical representation of the survival time data and the survival curves for the groups: placebo, pyridoxine and thiotepa. We used Bonferroni’s method to determine if significant differences exist among the survival curves of the groups.

Figure 2. Algorithm Bonferroni’s Method
1. Step one. Comparison of groups in pairwise.

1.1. Comparing survival curves of the groups: Placebo - Pyridoxine.

Table 3 shows the p-value of the comparison test of survival curves for the groups: placebo and pyridoxine. The results indicate that the equality hypothesis of survival curves cannot be rejected.

Placebo - Pyridoxine:

\[ H_0 : S_1(z) = S_2(z) \]
\[ H_1 : S_1(z) \neq S_2(z) \]

This test indicates that treatment with pyridoxine in patients with bladder cancer patients has not significant effect at the time of recurrence of tumors with respect to the placebo group.

1.2. Comparing survival curves of the groups: placebo vs. thiotepa.

Table 4 shows the comparison test p-value of the survival curves for the groups: placebo and thiotepa. Also, it is shows the pooled group curves.
1.3. Comparison of survival curves of the groups: pyridoxine vs. thiotepa.

Table 5 shows the p-value of the comparison test of the survival curves for the groups: pyridoxine and thiotepa. The result indicates that the equality hypothesis of survival curves cannot be rejected.

Pyridoxine-Thiotepa:

\[ H_0 : S_2(z) = S_3(z) \]
\[ H_1 : S_2(z) \neq S_3(z) \]

This test indicates that treatment with pyridoxine in patients with bladder cancer patients has not significant effect at the time of recurrence of tumors with the group treated with thiotepa.

2 Step two: Ordering of the p-values.

Table 6 indicates that the pair of group with more statistical difference is the pair placebo vs. thiotepa. The second pair with more statistical difference is the pair pyridoxine vs. thiotepa and with less difference is the pair of placebo vs. pyridoxine.

On our example, the global null hypothesis test is,

\[ H_0 : S_1(t) = S_2(t) = S_3(t) \]
\[ H_1 : \text{At least one } S_r(t) \text{ is different with } r = 1, 2, 3. \]
Where, $S_1(t)$, $S_2(t)$ and $S_3(t)$ are the survival curves of the groups: placebo, pyridoxine and thiotepa respectively. The hypothesis tests in pairs and ordered for the significations are:

**Placebo-Thiotepa**
- $H_{03}: S_1(t) = S_3(t)$
- $H_{13}: S_1(t) \neq S_3(t)$

**Pyridoxine-Thiotepa**
- $H_{02}: S_2(t) = S_3(t)$
- $H_{12}: S_2(t) \neq S_3(t)$

**Placebo-Pyridoxine**
- $H_{01}: S_1(t) = S_2(t)$
- $H_{11}: S_1(t) \neq S_2(t)$

3. **Step three**: $r = 3$
4. **Step four**: $i = 3$
5. **Step five: Comparing of groups**

$p_{03} = 0.223$ is greater than $\alpha/3 = 0.0167$. Therefore, the hypotheses: $H_{03}$ is accepted and the null hypotheses of the rest of the tests are accepted too.

4.6. **Step six: Decision**

The hypotheses: $H_{03}: S_1(z) = S_2(z) = S_3(z)$ is accepted.

Finish the procedure. **Byar’s experiment** shows:

For confidence level of 95%, there is no significant difference among the three treatments in tumor recurrence. There is no significant difference among three different treatments in tumor recurrence if the confidence level is 95%. For a confidence level of 90%, $H_{03}$ and $H_{02}$ are rejected and $H_{01}$ is not. So global null hypothesis is rejected, it indicates there is significant statistical difference in the application of the three treatments. Therefore, patients treated with thiotepa delays significantly tumor recurrence compared with patients treated with placebo and pyridoxine.

If, on the other hand, the confidence level is 90%, then $H_{05}$ and $H_{02}$ are rejected but $H_{01}$ is not. Therefore the global null hypothesis is rejected, resulting into a significant statistical difference when the three treatments are applied. Hence, the patients treated with thiotepa have a significant delay in tumor recurrence if compared with patients treated with placebo and pyridoxine.

**DISCUSSION AND CONCLUSIONS**

On this paper it proposes a methodology to compare $k$ survival curves of groups with recurrent events based on Bonferroni’s method, which is an alternative method to compare $k$ survival curves of population groups with such events. In the literature there are other works that propose methods to compare survival curves with these events. Martinez et al. were proposed these types of tests. The comparison Bonferroni’s procedure is an iterative method for comparing multiple groups whose contrast is made by comparing two by two, that part of the initial assumption of independence between the groups. In this paper has been worked with classic Bonferroni procedure and with a critical value equal to $\alpha/k$, where $k$ is the number of groups and $\alpha$ is the error type I. One conclusion in this paper is that this method of comparison is conservative and the probability of rejecting at least one hypothesis when all are true is not higher. However, in the methodology proposed, you can use alternative critical values to classical method of Bonferroni.

You can use the proposals of Holm or Simes. Other important conclusion is the following: If applying Bonferroni’s method to compare the survival curves of various groups, the curves pair with the greater statistical difference are equals, all the curves are equal. So, the global null hypothesis is not rejected. And if the curves pair with the greater statistical difference are different, at least there is a curve that it is different from the rest of them. Therefore, the global null hypothesis is rejected.

**REFERENCES**


Martinez. Bonferroni’s method to compare Survival Curves with Recurrent Events, p. 105-114


Martínez, C., Ramírez GJ, Vásquez ML. (2011). Statistical tests to compare k survival analysis functions involving recurrent events, Proceeding of the 58th World Statistics Congress 2011; Dublin Ireland.


Martínez. Bonferroni’s method to compare Survival Curves with Recurrent Events, p. 105-114

**Autor**

Carlos M. Martínez M. Profesor Asociado a Dedicación Exclusiva, adscrito al Departamento de Investigación Operativa, Escuela de Ingeniería Industrial, Facultad de Ingeniería, Universidad de Carabobo. Ingeniero Industrial, Magister y Doctor en Estadística por la Universidad Central de Venezuela

**Recibido:** 15/03/2013  **Aceptado:** 27/05/2013

*Martínez*. Bonferroni’s method to compare Survival Curves with Recurrent Events, *p. 105-114*