Statistica & Applicazioni
Vol. III, n. speciale 1, 2005
pp. 37-44

Multivariate permutation tests for evaluating effectiveness of universities through the analysis of student dropouts

Stefano Bonnini[§]

Luigi Salmaso[‡]

Aldo Solari[§]

Summary: *This paper deals with the problem of testing for dropping out (by censoring all graduate students) in analogy with survival analysis. A conditional test based on the probability of dropout and on the distribution of observed time to dropout is given. The application of the proposed solution to all students who began one of the three new laurea programs in the Civil Area of the Faculty of Engineering at the University of Padua is presented. A simulation study shows that the proposed solution presents a very good overall performance.*

Keywords: Competing Risks, Permutation test, Nonparametric combination, Survival analysis, Unequal censoring.

1. Introduction

The analysis of university student careers may be viewed in a competing risks framework, where the two risks that are faced by the students are completion and dropping out.

However, one can view competing risks data as a censored data problem in which is desired to estimate the distribution of time until failure from a particular cause when the other cause of failure is not in effect.

In the current paper, the event of interest is the student's dropping out from university, where the dropout is defined as withdrawal from the degree

[§] Dipartimento di Scienze Statistiche – Università degli Studi di Padova – via C. Battisti, 241, 35121 PADOVA (e-mail: bonnini@stat.unipd.it, solari@stat.unipd.it).

[‡] Dipartimento di Tecnica e Gestione dei Sistemi Industriali – Università degli Studi di Padova – stradella San Nicola, 3, 36100 VICENZA (e-mail: salmaso@gest.unipd.it).

program for whatever reason. For each student, if the dropout occurs, we observe the failure time, whereas if it does not occurs, we observe the length of time to graduation, recorded as a right-censored observation.

Our objective is the comparison of survival functions between two groups. The problem can be formalized by considering the two groups as the "agents" (i.e. universities, faculties, etc.) with different institutional policies and practices.

2. Formulation of the problem

In the general right-censored survival problem, the life process of a subject can be represented by random variables (r.v.)

- X , the true survival time under test;
- Z , the observed portion of survival time;
- O , the indicator of whether Z denotes actual survival time $(O = 1)$ or a censored record of it $(O = 0)$.

In our context

- the event of interest is the student's dropout from the degree program;
- the r.v. X denotes the length of time to dropout;
- the r.v. Z represents the time elapsed either because dropout occurs or because observation must cease, that is be censored, because we fail to observe all graduate students.

A full probabilistic description of the observable pair (O, Z) is given by the quantities p , $F^{Z|O=1}$ and $F^{Z|O=0}$, where $p = \Pr\{O = 1\}$ is the probability that an observation is uncensored, $F^{Z|O=1}(t) = \Pr\{Z \le t \mid O = 1\}$ is the conditional distribution of observation time, given that censoring does not occur, and $F^{Z|O=0}(t) = \Pr\{Z \le t \mid O = 0\}$ is the conditional distribution of observation time, given that censoring does occur.

2.1 *Random censorship model*

Nearly all statistical methods for censored survival data are based on the assumption that censoring effects are, in a very specific sense, noninformative with respect to the distribution of survival time. Noninformative censoring can be characterized by the random-censorship model, where the observable pair (O, Z) is structurally represented by $Z = \min(X, C)$ and $O = 1$ if $X \le C$ or $O = 0$ if $X > C$, where the random variable C with distribution function G , is independent of X .

It follows from the statistical independence of X and C that $\Pr\{O=1\}=\int_0^\infty F(t)dG(t)$, $\Pr\{Z\in(t-dt,t),O=1\}=[1-G(t)]dF(t)$ and $\Pr\{Z\in(t-dt,t),O=0\}=[1-F(t)]dG(t)$.

3. Tests for the comparison of survival time distributions

A fundamental problem is the comparison of the survival-time distributions from two groups; this is a common occurrence when two treatments are being evaluated. In our context, the *j*th treatment, $j=1,2$, corresponds to the institutional policy and practice of the *j*th "agent" (faculty, degree program, etc).

Clearly, there are other factors such age, sex, etc, which may also be related to survival time and, therefore, must be dealt with. A possible solution is to construct homogeneous strata with respect to possible confounding factors. For example, stratification by propensity score (Rosenbaum and Rubin, 1983) may be used.

In the following the discussion is restricted to the simpler situation in which the n_j observations (O_{ji},Z_{ji}), $i=1,\dots,n_j$ forming the *j*th group are homogeneous with respect both survival time and censoring.

We now turn attention to nonparametric tests for the null hypothesis of the equality of survival time distribution functions across two samples. The most popular test for comparing survival curves in the presence of right-censoring is the logrank test for testing the null hypothesis $H_0:F_1=F_2$ (Heimann and Neuhaus, 1998). If we assume that the two groups are defined according to increasing levels of a treatment (i.e. in the present context, institutional policy and practice of the agents), it may be of interest to test against the one-sided alternative $H_1:F_1<F_2$, because the expected treatment effects are assumed to decrease time to dropout.

As frequently occurs in such studies the event of interest (dropout) can only be observed in a relatively small number of students. Therefore, the asymptotic distribution under the null hypothesis of the above mentioned test statistics may not be appropriate. This has been observed by several authors, who then proposed so-called exact methods. This provides an exact small sample test when the censoring patterns in the samples being compared are equal, i.e. $G_1=G_2$.

Since censoring intensities are not likely to be equal in practice, the permutation test must, in general, be viewed as an approximate test. In fact, when the censoring patterns are unequal, the observation pairs from the first sample do not have the same distribution as those from the second sample, even when the null hypothesis $H_0 : F_1 = F_2$ is true.

3.1 *Analysis based on (O,Z)*

Censoring observations cannot be prevented since they represent failure from another cause (graduation), hence the complete observation of the survival time of interest is an unachievable event. Because of this ambiguity, a statistician faced with this type of data should first determine whether an inference about F is actually necessary. Often an analysis based only on the distribution of (O, Z) is sufficient (Lagakos, 1979). Therefore, the null hypothesis of interest is given by

$$\begin{aligned} H_0 &: \Pr\{O_1 = 1, Z_1 \leq t\} = \Pr\{O_2 = 1, Z_2 \leq t\} \\ &: \Pr\{O_1 = 1\}\Pr\{Z_1 \leq t \mid O_1 = 1\} = \Pr\{O_2 = 1\}\Pr\{Z_2 \leq t \mid O_2 = 1\} \end{aligned}$$

Hence, the null hypothesis requires the joint distributional equality of the missing data process in the two groups, and of response variable Z conditional on O, that is

$$\begin{aligned} H_0 &: H_0^O \cap H_0^{Z|O} \\ &: \left[\Pr\{O_1 = 1\} = \Pr\{O_2 = 1\}\right] \cap \left[\Pr\{Z_1 \leq t \mid O_1 = 1\} = \Pr\{Z_2 \leq t \mid O_2 = 1\}\right] \\ &: \left[p_1 = p_2\right] \cap \left[F_1^{Z|O=1} = F_2^{Z|O=1}\right] \end{aligned}$$

We may write the alternative hypothesis as

$$\begin{aligned} H_1 &: H_1^O \cup H_1^{Z|O} \\ &: \left[\Pr\{O_1 = 1\} \leq \Pr\{O_2 = 1\}\right] \cup \left[\Pr\{Z_1 \leq t \mid O_1 = 1\} \geq \Pr\{Z_2 \leq t \mid O_2 = 1\}\right] \\ &: \left[p_1 \leq p_2\right] \cup \left[F_1^{Z|O=1} \geq F_2^{Z|O=1}\right] \end{aligned}$$

where at least one inequality is strict because the expected treatment effects are:

- increasing probability of dropout;
- stochastically lower time to dropout.

In the permutation context, let us denote the observed data set by $(\mathbf{O}, \mathbf{Z})$ and the *b*th permutation of it by $(\mathbf{O}^{*b}, \mathbf{Z}^{*b})$, $b = 1, \ldots, n!$.

For testing the subhypothesis H_0^O against the subalternative H_1^O, we consider the Fisher's exact probability test

$$T_O = T_O(\mathbf{O},\mathbf{Z}) = \sum_{i=1}^{n_2} O_{2i} \ . \tag{1}$$

Following Pesarin (2001), for testing $H_0^{Z|O}$ against $H_1^{Z|O}$ we use a test statistic

$$T_{Z|O} = T_{Z|O}(\mathbf{O},\mathbf{Z}) = \left(\sum_{i=1}^{n_1} Z_{1i}O_{1i}\right)\cdot\left(\frac{\nu_2}{\nu_1}\right)^{1/2} - \left(\sum_{i=1}^{n_2} Z_{2i}O_{2i}\right)\cdot\left(\frac{\nu_1}{\nu_2}\right)^{1/2}, \tag{2}$$

that is invariant in mean and variance with respect to the effective sample size $\nu_j = \sum_{i=1}^{n_j} O_{ji}$, that vary according to the random attribution of units to the two groups, because units with missing data partecipate in the permutation mechanism as well as all other units.

Then, using the Fisher's combining function, the overall solution becomes

$$T_F'' = T_F''(\mathbf{O},\mathbf{Z}) = -2\cdot\left\{\log(\lambda_O) + \log(\lambda_{Z|O})\right\}, \tag{3}$$

where $\lambda_O = \sum_{b=1}^{n!} I\left\{T_O^{*b} \geq T_O\right\}/n!$ and $\lambda_{Z|O} = \sum_{b=1}^{n!} I\left\{T_{Z|O}^{*b} \geq T_{Z|O}\right\}/n!$ are the *p*-values associated with the permutation tests (1) and (2), respectively, where $I\{\cdot\}$ denotes the indicator function and $T_O^{*b} = T_O\left(\mathbf{O}^{*b},\mathbf{Z}^{*b}\right)$, $T_{Z|O}^{*b} = T_{Z|O}\left(\mathbf{O}^{*b},\mathbf{Z}^{*b}\right)$ are the values of the test statistics computed on $\left(\mathbf{O}^{*b},\mathbf{Z}^{*b}\right)$.

4. Application to the faculty of Engineering

Our data refers to students enrolling in the 2001/2002 academic year in one of the 3 new laurea programs in the Civil Area of the Faculty of Engineering at the University of Padua, Italy. The follow-up period for each freshman is 3 years, and the time unit for events (dropout and graduation) is the academic year, as shown in Table 1. Although data refer to the most recent cohort available, we acknowledge that information regarding time to dropout and graduation $\#\left[Z_j = t, O_j = o\right]/n_j = \hat{\Pr}\{Z_j = t, O_j = o\}$, $o = 0{,}1$, is not complete. In Table 2, we arbitrarily complete the information of $\hat{\Pr}\{Z_j = t, O_j = 1\}$ and $\hat{\Pr}\{Z_j = t, O_j = 0\}$ for t = 4, 5 and 6, assuming that by the end of the sixth year there is a forced termination: successful completion or dropout.

Table 1. *Observed number of dropouts and graduations; A: Environmental and Territorial Engineering; B: Civil Engineering; C: Building Engineering.*

Group	n_j		$\#\{Z_{ji}=t,O_{ji}=1\}$						$\#\{Z_{ji}=t,O_{ji}=0\}$					
		t	1	2	3	4	5	6	1	2	3	4	5	6
A	78		9	4	3	?	?	?	0	0	35	?	?	?
B	149		26	13	5	?	?	?	0	0	45	?	?	?
C	133		31	14	3	?	?	?	0	0	30	?	?	?

Table 2. *Probability of observing a dropout or a graduation.*

Group	n_j		$\hat{\Pr}\{Z_j=t,O_j=1\}$						$\hat{\Pr}\{Z_j=t,O_j=0\}$					
		t	1	2	3	4	5	6	1	2	3	4	5	6
A	78		.12	.05	.04	.03	.02	.01	0	0	.45	.18	.07	.04
B	149		.17	.09	.03	.03	.02	.01	0	0	.30	.21	.09	.04
C	133		.23	.11	.02	.02	.02	.01	0	0	.23	.22	.11	.04

Note that under the random censorship model, we have

$$f_j(t)=\frac{\Pr\{Z_j=t,O_j=1\}}{1-G_j(t-1)} \quad \text{and} \quad g_j(t)=\frac{\Pr\{Z_j=t,O_j=0\}}{1-F_j(t)},$$

thus we are able to estimate F and G by using $\hat{\Pr}\{Z_j=t,O_j=1\}$ and $\hat{\Pr}\{Z_j=t,O_j=0\}$.

4.1 *Illustrating example*

For purpose of illustration, generating two samples by using the estimated probabilities obtained above, we obtain the following data (Table 3):

Table 3. *Generated number of dropouts and graduations.*

Group	n_j		$\#\{Z_{ji}=t,O_{ji}=1\}$						$\#\{Z_{ji}=t,O_{ji}=0\}$					
		t	1	2	3	4	5	6	1	2	3	4	5	6
B	149		26	9	4	4	3	3	0	0	53	37	7	5
C	133		34	12	1	1	3	1	0	0	36	26	17	2

The resulting *p*-values of the unconditional and conditional logrank test are .1190 and .1160, respectively, whereas for the combined permutation test $T_F^{''}$ the associated *p*-value is .0338, significant at $\alpha = .05$. We can also examine the separate contribution of the frequency of dropout and of the time to dropout, by looking at the adjusted *p*-values of the partial tests T_O and $T_{Z|O}$. As a results, students of the two areas dropout substantially with the same frequency ($\lambda_O = .1691$) but with significantly early time to dropout for the group B with respect to the group A ($\lambda_{Z|O} = .0419$).

4.2 *Simulation study*

The effect of unequal censoring will, in general, depend in complicated ways on the censoring mechanisms involved, and the permutation test can be either conservative or nonconservative. A good deal of insight can be gained, however, by a Monte Carlo study of the size of the tests assuming unequal censoring patterns. In order to evaluate the size of the tests assuming unequal censoring, by considering under the null hypothesis the same F for the three groups as $(F_A + F_B + F_C)/3$. The size of the tests was estimated at $\alpha = .05$ using 1000 Monte Carlo simulations and 1000 permutations. The results are shown in Table 4, where L^a and L^c denote respectively the unconditional and conditional version of the logrank test.

Table 4. *Estimated size of the tests under unequal censoring.*

Configuration	Groups size	L^a	L^c	T''_F	T_O	$T_{Z\|O}$	% censoring
A vs B	(78,149)	.062	.056	.046	.065	.024	[65,64]
B vs C	(149,133)	.056	.040	.045	.042	.028	[64,63]
A vs C	(78,133)	.061	.059	.053	.065	.024	[65,63]

The size was also estimated under equal censoring, by considering the same G for the three groups as $(G_A + G_B + G_C)/3$ (see Table 5).

Table 5. *Estimated size of the tests under equal censoring.*

Configuration	Groups size	L^a	L^c	T''^*_F	T^*_O	$T^*_{Z\|O}$	% censoring
A vs B	(78,149)	.053	.051	.048	.043	.039	[64,64]
B vs C	(149,133)	.058	.055	.043	.047	.044	[64,64]
A vs C	(78,133)	.054	.055	.046	.046	.049	[64,64]

The estimated power of the considered tests are shown in Table 6, either under equal and unequal censoring.

Table 6. *Estimated power of the tests.*

Configuration	Groups size	L^a	L^c	T''_F	T_O	$T_{Z\|O}$	% censoring
Unequal Censoring							
A vs B	(78,149)	.425	.416	.338	.375	.140	[73,64]
B vs C	(149,133)	.319	.319	.361	.274	.180	[64,59]
A vs C	(78,133)	.778	.772	.789	.722	.293	[73,59]
Equal Censoring							
A vs B	(78,149)	.425	.412	.424	.334	.118	[73,64]
B vs C	(149,133)	.259	.264	.342	.202	.232	[64,59]
A vs C	(78,133)	.732	.727	.775	.631	.397	[73,59]

5. Conclusions

This paper mainly looks at the problem of testing for dropping out (by censoring all graduate students) in analogy with survival analysis. The application of the proposed solution to all students who began one of the 3 new laurea programs in the Civil Area of the Faculty of Engineering at the University of Padua allows us to examine the separate contribution of the frequency of dropout and of the time to dropout to the possible global significance, whereas this kind of analysis is not possible with the log-rank type tests. Finally, a comparative simulation study shows that the proposed solution presents a very good overall performance.

Riferimenti Bibliografici

Heimann G., Neuhaus G. (1998). Permutational Distribution of the Log Rank Statistic under Random Censorship with Applications to Carcinogenicity Assays. *Biometrics*, **54**, 168-184.

Lagakos S. W. (1979). General Right Censoring and Its Impact on the Analysis of Survival Data. *Biometrics*, **35**, 139-156.

Pesarin F. (2001). *Multivariate Permutation Tests with applications in Biostatistics*. Wiley, Chichester.