

Exercici d'events recorrents en R (part II)

Jose Calatayud Mateu

2025-04-21

```
# Llibreries
library(devtools)
```

```
## Loading required package: usethis
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr    1.5.1
## v ggplot2    3.5.1      v tibble     3.2.1
## v lubridate  1.9.3      v tidyr      1.3.1
## v purrr      1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
```

```
devtools::install_github("isglobal-brge/TestSurvRec")
```

```
## Skipping install of 'TestSurvRec' from a github remote, the SHA1 (db98fc68) has not changed since last
## Use 'force = TRUE' to force installation
```

```
library("TestSurvRec")
```

```
## Loading required package: survrec
## Loading required package: boot
```

```
devtools::install_github("isglobal-brge/survrec", build = FALSE)
```

```
## Skipping install of 'survrec' from a github remote, the SHA1 (91402d0f) has not changed since last i
## Use 'force = TRUE' to force installation
```

```
library(survrec)
```

Procedim a descarregar manualment l'arxiu associats als paquet “gcmrec” que està dedicat a funcions de modelatge de models generals per events recurrents i el data sobre el qual treballarem “lymphoma”:

```
source(file="lymphoma.R")
source(file="survrec.R")
source(file="gcmrec.R")
```

El data “lymphoma” contenen els temps de recaiguda del càncer després del primer tractament en pacients diagnosticats amb limfoma de grau baix.

```
head(lymphoma)
```

##	id	time	event	enum	delay	age	sex	distrib	effage
## 1	6	3.900826	0	1	17	79	1	1	CR
## 2	7	63.173554	0	1	33	25	1	1	CR
## 3	8	41.289256	0	1	26	37	1	2	CR
## 4	11	29.421488	1	1	31	43	2	2	CR
## 5	11	20.826446	1	2	31	43	2	2	CR
## 6	11	17.950413	0	3	31	43	2	2	CR

Exercici 2: Comparación de Curvas de Supervivencia

- *Investiga com el paquet TestSurvRec compara completament dues corbes de supervivència (mira Secció de Referències d'ambdues paquets i el manuscrit describint com funciona)*

El paquet **TestSurvRec** proporciona mètodes para comparar dues corbes de supervivència en situacions de events recurrents.

El paquet consta de les següents funcions entre altres bases de datos:

- Dif.Surv.Rec
- Plot.Data.Events
- Plot.Data.Rec
- Plot.Surv.Rec
- Print-Summary

La función principal **Dif.Surv.Rec** implementa un procedimiento de prueba robusto que tiene en cuenta la correlación entre eventos recurrentes del mismo individuo. Utiliza un enfoque de proceso de conteo y calcula una estadística de prueba basada en la diferencia entre las dos curvas de supervivencia estimadas, considerando todo el período de seguimiento.

Todas las funciones trabajan con una estructura de data type de eventos recurrentes que incorpora una variable **group** indicadora del tratamiento al que pertenecen. Así, modificamos la base de lymphoma para poder trabajar correctamente con el grupo de interés que se desea comparar:

```
lymphoma_sex<- lymphoma %>%
  mutate("group" = sex %>% as.factor)

Plot.Surv.Rec(lymphoma_sex)
```

```

##
##      Group= 1
##      time n.event n.risk  surv std.error
##      1.85      1      69 0.986   0.0143
##      3.34      1      65 0.970   0.0205
##      4.40      1      62 0.955   0.0254
##      5.22      1      59 0.939   0.0296
##      6.25      1      57 0.922   0.0333
##      6.48      1      56 0.906   0.0365
##      7.14      1      53 0.888   0.0395
##      7.44      1      52 0.871   0.0422
##      8.07      1      51 0.854   0.0446
##      8.33      1      50 0.837   0.0468
##      8.36      1      49 0.820   0.0488
##      8.96      1      48 0.803   0.0507
##      9.19      1      47 0.786   0.0523
##      9.42      1      46 0.769   0.0539
##      9.88      1      45 0.752   0.0552
##     10.05      1      44 0.735   0.0565
##     10.48      1      42 0.717   0.0578
##     11.64      1      39 0.699   0.0591
##     12.26      1      38 0.680   0.0602
##     12.43      1      37 0.662   0.0613
##     13.29      1      35 0.643   0.0623
##     13.95      1      33 0.624   0.0633
##     15.54      1      30 0.603   0.0644
##     21.12      1      26 0.580   0.0658
##     21.92      1      25 0.556   0.0670
##     26.71      1      22 0.531   0.0683
##     27.11      1      20 0.505   0.0697
##     32.03      1      17 0.475   0.0713
##     37.06      1      16 0.445   0.0724
##     62.94      1       8 0.390   0.0799
##     81.82      1       3 0.260   0.1016
##    120.76      1       1 0.000   0.0000
##
##
##      Group= 2
##      time n.event n.risk  surv std.error
##      5.52      1      33 0.970   0.0294
##      6.15      1      32 0.939   0.0409
##      7.67      1      30 0.908   0.0498
##      7.77      1      29 0.877   0.0568
##      8.30      1      27 0.844   0.0630
##      8.76      1      26 0.812   0.0681
##     10.05      1      24 0.778   0.0729
##     10.18      1      23 0.744   0.0769
##     11.14      1      21 0.709   0.0806
##     12.46      1      20 0.673   0.0837
##     13.72      1      19 0.638   0.0861
##     16.99      1      18 0.602   0.0879
##     20.83      1      13 0.556   0.0917
##     23.11      1      12 0.510   0.0942
##     25.42      1      10 0.459   0.0964

```

```
## 26.35      1      8 0.401    0.0982
## 29.42      1      7 0.344    0.0974
##
##           Group   Median
## 1 Pooled Group 26.71074
## 2 1er Group   32.03306
## 3 2do Group   25.42149
```

```
plot(1:10, 2:11)
```

En la representació anterior, es mostren les gràfiques de les corbes de supervivència estimades de l'objecte de classe `TestSurvRec`. Les corbes de supervivència de ambdues grups son estimats mitjançant el mètode PHS del paquet `survrec` assumint que els temps inter-ocurrències son independents. Endemès, també plotetja la corba de supervivència pooled, sense distingir entre classes de sexe.

- *Utilitzant aquest mètode:*

NOTA: la variable `distrib` codifica les lesions relacionades amb el diagnostic en 4 categories (0=Single, 1=Localized, 2=More than one nodal site, 3=Generalized)

- *Compara el número de vegades en recaiguda de cancer entre homes i dones (variable `sex`)*

Per tant, es comparen els temps de recaiguda entre homes (`sex=1`) y mujeres (`sex=0`). Abans de res, plotejem

```
Dif.Surv.Rec(lymphoma_sex, "all")
```

```
## Nomb.Est   Chi.square   p.value
## 1 LRrec    0.1894371795 0.6633855
## 2 Grec     0.0007768251 0.9777646
## 3 TWrec    0.0528432059 0.8181877
## 4 PPrecc   0.0376726615 0.8461019
## 5 PMrec    0.0354733341 0.8506073
## 6 PPrrec   0.0403693094 0.8407602
## 7 HFrec    0.1894371795 0.6633855
## 8 CMrec    0.1894371795 0.6633855
## 9 Mrec     0.2454391078 0.6203050
```

Aquesta funció conté tests per comparar corbes de supervivència amb events recurrents utilitzant l'estimador PSH. La funció retorna diferents tests que depenen dels pesos que assignen així la selecció apropiada dels pesos en l'anàlisi de supervivència depèn del comportament de les seves corbes. Amb l'elecció dels valors dels paràmetres (α, β, γ ; i η) en la nostra proposta, l'estadístic és capaç d'ajustar-se a aquest comportament.

Amb aquesta proposta, es poden dur a terme estudis d'anàlisi de supervivència amb esdeveniments recurrents i generar proves per a altres tipus d'anàlisi, incloent-hi els tests clàssics com: **logrank**, **Gehan**, **Peto-Peto**, **Fleming-Harrington**, entre d'altres.

Observem que, si tots els paràmetres de l'estadístic són zero, aleshores $w_z = 1$, la qual cosa genera el test **logrank** per a l'anàlisi amb esdeveniments recurrents (és el que la funció retorna per defecte).

Si $\alpha = 1$ i la resta de paràmetres són zero, llavors $w_z = Y(s, z)$, el que correspon al test **Gehan**.

Si $\gamma = 1$ i els altres paràmetres són zero, llavors $w_z = S(z)$, el que dona lloc al test de **Peto-Peto**.

Finalment, si $\gamma = 1$, $\eta = 1$ i els altres paràmetres són zero, s'obté el test de **Fleming-Harrington**.

Per tal que l'estadístic pugui generar els tests de l'anàlisi de supervivència clàssica, totes les unitats d'estudi han d'experimentar només un esdeveniment o una censura.

En la proposta, l'estadístic de comparació per a esdeveniments recurrents depèn dels processos de recompte N i Y , tots dos indexats doblement. Amb l'índex s mesurem el temps de calendari, i amb l'índex z mesurem els temps entre esdeveniments (gap times).

Per tant, si el temps d'observació tendeix a infinit i l'esdeveniment només pot ocórrer una vegada per unitat, el test de comparació de la proposta esdevé el test clàssic ponderat de comparació de grups en l'anàlisi de supervivència.

Aleshores, com que la hipòtesis del model son:

$$H_0 : S_1(t) = S_2(t) \quad H_0 : S_1(t) \neq S_2(t)$$

Independement del test aplicat, tots ells presenten un p-value > 0.05 aleshores no hi ha evidència estadística per rebutjar la hipòtesi nul · la d'igualtat de corbes de supervivència.

- *Compara el número de vegades en recaiguda de cancer entre pacients tenint lesions úniques i lesions localitzades (variable distrib)*

De la mateixa forma que en l'exercici anterior, creem una nova variable grup per distrib però filtrem només els casos en que son lesions úniques i lesions localitzades

```
lymphoma_distrib1 <- lymphoma %>%
  filter(distrib==0 | distrib==1) %>%
  mutate(group=distrib)

Dif.Surv.Rec(lymphoma_distrib1, "all")
```

```
##   Nomb.Est Chi.square   p.value
## 1 LRrec      2.751177 0.0971828
## 2 Grec       3.203988 0.0734589
## 3 TWrec      3.034240 0.0815247
## 4 PPrecc     3.092627 0.0786478
## 5 PMrec      3.093417 0.0786096
## 6 PPrecc     3.094571 0.0785538
## 7 HFrec      2.751177 0.0971828
## 8 CMrec      2.751177 0.0971828
## 9 Mrec       2.751177 0.0971828
```

En cap dels test aplicat, no hi ha evidència estadística suficient per un nivell de significació del 95% per rebutjar la hipòtesi nul · la d'igualtat de corbes de supervivència entre el número de vegades en recaiguda de cancer entre pacients tenint lesions úniques i lesions localitzades

- *Compara el número de vegades en recaiguda de cancer entre pacients ab lesions úniques i lesions en més d'un lloc nodal (variable distrib)*

Anàleg, al apartat anterior:

```
lymphoma_distrib2 <- lymphoma %>%
  filter(distrib==0 | distrib==2) %>%
  mutate(group=distrib)

Dif.Surv.Rec(lymphoma_distrib2, "all")
```

##	Nomb.Est	Chi.square	p.value
## 1	LRrec	5.839279	0.0156723
## 2	Grec	5.125427	0.0235778
## 3	TWrec	5.607358	0.0178852
## 4	PPrec	5.480682	0.0192278
## 5	PMrec	5.437236	0.0197118
## 6	PPrrec	5.494695	0.0190743
## 7	HFreq	5.839279	0.0156723
## 8	CMrec	5.839279	0.0156723
## 9	Mrec	4.494656	0.0340010

En canvi, en aquest cas, tots els tests presentats indiquen que sí hi ha evidència estadística suficient per rebutjar l'igualtat d'estimacions de corbes de supervivència entre el número de vegades en recaiguda de cancer entre pacients ab lesions úniques i lesions en més d'un lloc nodal.