

Trabajo 4

Francisco Herrera Barajas

2025-05-23

Ejercicio 1

Importa la base de datos Reincidencia.csv a R studio. Obtén los estadísticos descriptivos (la media, mediana, mínimo y máximo) de los meses transcurridos hasta la reincidencia o hasta el final del estudio. (En el documento Word, incluya un pantallazo de estos resultados, y su interpretación).

Primero cargamos las librerías necesarias

```
library(survival)
library(prodlim)
library(ggplot2)
library(lubridate)
```

```
##
## Adjuntando el paquete: 'lubridate'
## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
```

```
library(ggsurvfit)
library(gtsummary)
library(tidycmprsk)
```

```
##
## Adjuntando el paquete: 'tidycmprsk'
## The following object is masked from 'package:gtsummary':
##
##   trial
```

```
library(condSURV)
library(ggfortify)
library(dplyr)
```

```
##
## Adjuntando el paquete: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v forcats 1.0.0      v stringr 1.5.1
## v purrr  1.0.4       v tibble  3.2.1
## v readr   2.1.5      v tidyr   1.3.1
```

```
## -- 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(psych)
```

```
##
## Adjuntando el paquete: 'psych'
##
## The following object is masked from 'package:ggsurvfit':
##
##      %+%
##
## The following objects are masked from 'package:ggplot2':
##
##      %+%, alpha
```

```
library(ggplot2)
```

```
library(survminer)
```

```
## Cargando paquete requerido: ggpubr
##
## Adjuntando el paquete: 'survminer'
##
## The following object is masked from 'package:survival':
##
##      myeloma
```

Cargamos la base de datos y la exploramos.

```
datos <- read.csv2("Reincidencia.csv")
head(datos)
```

```
##   Edad_Media Trabajo NivelIngresosAnual Alcoholismo Sexismo_hostil RESUMEN.1
## 1    -6.32      1          1           7          44           0
## 2    24.68      0          4           4          42           1
## 3    10.68      0          2           1          47           0
## 4   -16.32      0          4           0          13           0
## 5    -4.32      1          1          12          48           0
## 6    -7.32      0          0           6          23           0
##   Reincid Meses
## 1      0     53
## 2      0     40
## 3      0     30
## 4      0     17
## 5      0     67
## 6      0     67
```

Verificamos si la base de datos contiene NAs.

```
na <- sum(is.na(datos))
print(paste0("Número de Nas: ", na))
```

```
## [1] "Número de Nas: 10"
```

Verificamos el número de filas

```
num_row <- nrow(datos)
print(paste0("Número de filas: ", num_row))
```

```
## [1] "Número de filas: 335"
```

Borramos las filas que contengan NA

```
datos <- datos %>% drop_na()
```

Analizamos los datos sacando su descriptivos

```
describe.by(datos$Meses, group = datos$Reincid)
```

```
## Warning in describe.by(datos$Meses, group = datos$Reincid): describe.by is
## deprecated. Please use the describeBy function
```

```
##
## Descriptive statistics by group
## group: 0
##   vars   n mean    sd median trimmed  mad min max range  skew kurtosis   se
## X1     1 280 35.51 20.44     36  35.54 28.17   4  67   63 -0.04   -1.37 1.22
## -----
## group: 1
##   vars   n mean    sd median trimmed  mad min max range  skew kurtosis   se
## X1     1  45 24.29 19.14     20  22.78 22.24   1  69   68  0.51   -0.98 2.85
```

Primero analizamos para el grupo 0, es decir los participantes del programa que no reincidieron en agredir a su pareja.

Estos participantes forman el grupo más numeroso, $n = 280$. Su media y mediana son de 35,51 y 36 meses respectivamente transcurridos después de la terapia psicológica. Su mínimo es 4 y su máximo es 67.

Por otro lado encontramos el grupo 1, es decir aquellos participantes que reincidieron y volvieron agredir a su pareja.

Son un grupo de menor tamaño, $n = 45$. Su media es de 24.29 y su mediana de 20. Su máximo 69 y su mínimo 1. Me llama la atención que el máximo sea mayor que el grupo anterior. La desviación típica es elevada entre los participantes al igual que el grupo anterior.

Convertimos reincid en factor

Vemos que en el grupo reincidentes hubo un porcentaje importante que volvieron a cometer agresiones en los primeros meses.

Ejercicio 2

Utilizando la función `survfit(Surv())` (consultar línea 73 del script de la tutoría), obtener una tabla de mortalidad, agrupando los meses en intervalos de 10 meses. Interpreta, sólo de las dos primeras líneas, el número de eventos, el número de expuestos al evento, y la probabilidad de supervivencia. (En el documento Word, incluya un pantallazo de estos resultados, y su interpretación)

Utilizamos la función `surv()`

```
tiempo_supervivencia <- Surv(time = datos$Meses, event = datos$Reincid)
summary(tiempo_supervivencia)
```

```
##           time           status
##  Min.    : 1.00   Min.    :0.0000
##  1st Qu.:16.00   1st Qu.:0.0000
##  Median :34.00   Median :0.0000
##  Mean   :33.96   Mean    :0.1385
##  3rd Qu.:55.00   3rd Qu.:0.0000
##  Max.    :69.00   Max.    :1.0000
```

```
table(is.na(datos$Meses))
```

```
##
## FALSE
##    325
```

```
table(is.na(datos$Reincid))
```

```
##
## FALSE
##    325
```

```
summary(tiempo_supervivencia)
```

```
##           time           status
##  Min.    : 1.00   Min.    :0.0000
##  1st Qu.:16.00   1st Qu.:0.0000
##  Median :34.00   Median :0.0000
##  Mean   :33.96   Mean    :0.1385
##  3rd Qu.:55.00   3rd Qu.:0.0000
##  Max.    :69.00   Max.    :1.0000
```

```
modelo <- survfit(tiempo_supervivencia ~ 1)
```

```
# Crear tabla secuenciada cada 10 meses
summary(modelo, times = seq(1, 71, by = 10))
```

```
## Call: survfit(formula = tiempo_supervivencia ~ 1)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    1     325      1    0.997 0.00307    0.991    1.000
##   11     266     16    0.945 0.01298    0.920    0.971
##   21     217      6    0.921 0.01587    0.891    0.953
##   31     167      7    0.888 0.01970    0.850    0.928
##   41     136      4    0.866 0.02219    0.823    0.910
##   51      94      7    0.811 0.02883    0.757    0.870
##   61      36      3    0.772 0.03570    0.706    0.846
```

Esta es la tabla de mortalidad. El mayor número de reincidentes se encuentra entre los 10 primeros meses, 15 en total. Luego desciende y se mantiene estable. Por otro lado, el programa tiene bastante miembros censurados, a simple vista se ve que hay más gente que se va de la que reincide. La probabilidad de supervivencia al final del intervalo 60-70 es de aproximadamente 70%, lo que significa que el 70% de los individuos no habían reincidido hasta ese momento.

Viendo los datos por encima no entiendo como hacer el conteo la función `surv` porque si hago un conteo de reincidentes hay algunas variaciones. Además no parece que cuente un último caso en el mes 68. Pero no he sabido hacerlo de mejor manera.

3 Ejercicio

Utilizando el método Kaplan-Meier, obtén la representación gráfica de la función de supervivencia. Mirando esta gráfica, dar una estimación de la mediana de los tiempos de espera e interpreta este valor. (En el documento Word, incluya la gráfica y la explicación de cómo ha encontrado el valor de la mediana)

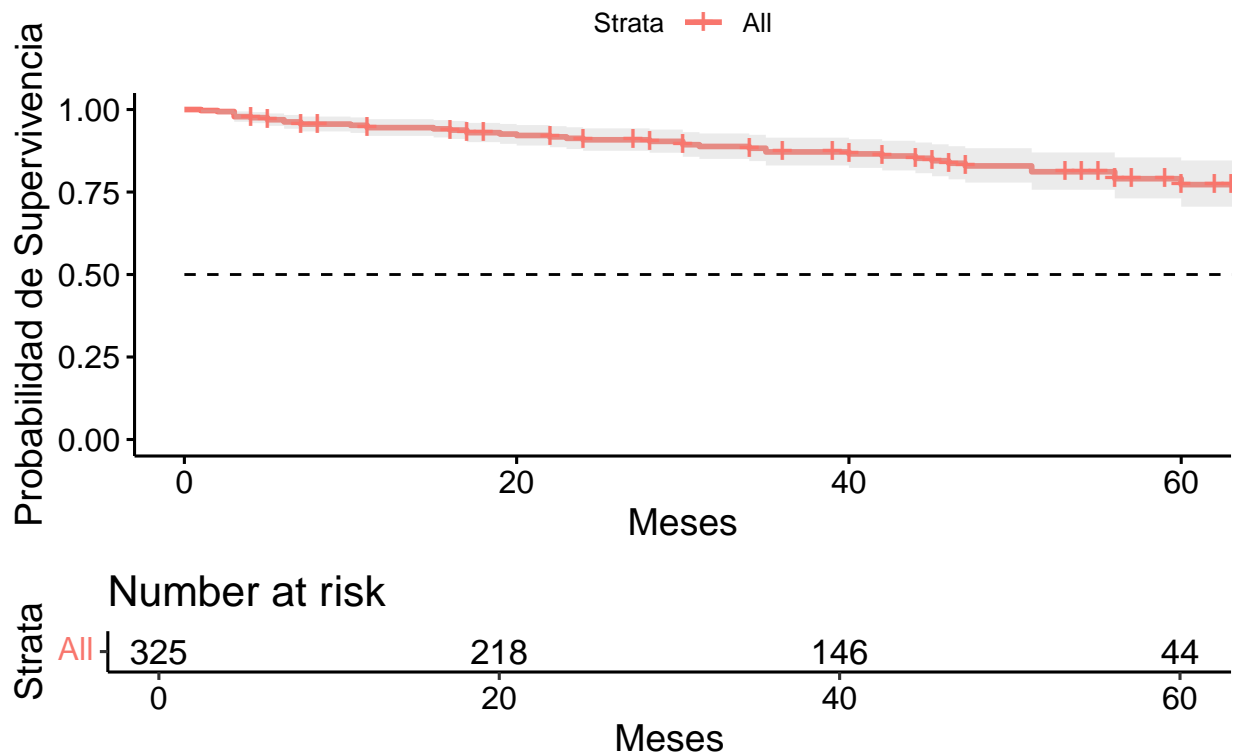
```
summary(modelo)$table["median"]
```

```
## median  
##      69
```

La media de los tiempos de espera es igual al tiempo máximo observado en el programa. Esto surge que la mayoría de los sujetos no reincidió.

```
# Ajustar el modelo  
modelo <- survfit(Surv(Meses, Reincid) ~ 1, data = datos)  
  
# Graficar  
print(ggsurvplot(modelo,  
  data = datos,  
  conf.int = TRUE,  
  xlab = "Meses",  
  ylab = "Probabilidad de Supervivencia",  
  title = "Función de Supervivencia - Kaplan-Meier",  
  risk.table = TRUE,  
  surv.median.line = "hv"))
```

Función de Supervivencia – Kaplan–Meier



La curva de Kaplan-Meier confirma que la probabilidad de no reincidir se mantuvo por encima del 50% durante todo el seguimiento. Lo cual es un valor bastante alto

Ejercicio 4

Por último, aplicar una regresión de Cox, metiendo como variables predictoras la edad media centrada, el nivel de ingresos, y la valoración del riesgo de reincidencia que aporta el psicólogo (esta variable es ordinal, pero hay que utilizarla como si fuera continua). Interpretar los coeficientes (transformados) de las variables que resulten estadísticamente significativas.

```
modelo1 <- coxph(Surv(time = Meses, event = Reincid) ~ Edad_Media + NivelIngresosAnual + RESUMEN.1, data = datos)
summary(modelo1)
```

```
## Call:
## coxph(formula = Surv(time = Meses, event = Reincid) ~ Edad_Media +
##       NivelIngresosAnual + RESUMEN.1, data = datos)
##
##   n= 325, number of events= 45
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## Edad_Media    -0.05092   0.95035  0.01566 -3.252  0.00115 **
## NivelIngresosAnual -0.07411   0.92857  0.07481 -0.991  0.32180
## RESUMEN.1       0.52275   1.68667  0.19456  2.687  0.00721 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## Edad_Media          0.9504      1.0522    0.9216    0.980
## NivelIngresosAnual   0.9286      1.0769    0.8019    1.075
## RESUMEN.1           1.6867      0.5929    1.1519    2.470
##
## Concordance= 0.655  (se = 0.047 )
## Likelihood ratio test= 19.1  on 3 df,  p=3e-04
## Wald test              = 17.98  on 3 df,  p=4e-04
## Score (logrank) test = 18.04  on 3 df,  p=4e-04
```

Interpretando estos resultados se puede ver que las variables edad (Edad_Media) y valoración del riesgo de reincidencia (RESUMEN.1) tienen un efecto significativo sobre el riesgo de reincidencia. Sin embargo, la variable nivel de ingreso (NivelIngresosAnual) no es significativa. La interpretación de como variará el coeficiente de la variable edad es que a medida que aumenta la edad, el riesgo de reincidencia disminuye, estimándose una reducción del 5% por cada año adicional. Por otro lado, un aumento en la variable valoración de riesgo se asocia con un incremento del 68,7% en el riesgo de reincidencia, lo que indica que la valoración de los psicólogos es importante.