

# Trabajo Práctico

## Análisis de Datos de Duración - 2024

### Introducción

El cáncer de mama, conocido también como carcinoma mamario, es un tipo de cáncer que se origina en las células del tejido mamario. Se clasifica como un cáncer primario, lo que significa que se inicia en la mama y puede propagarse a otros tejidos u órganos del cuerpo.

El conjunto de datos **rotterdam** del paquete **survival** contiene información sobre el tiempo de supervivencia de 2982 pacientes femeninos con cáncer de mama primario, enrolados entre 1978 y 1993 en la ciudad de Rotterdam. Además, se cuenta con variables basales asociadas a cada paciente. El origen del estudio fue definido al momento de finalización de la cirugía primaria.

### Descripción de la base de datos rotterdam

Las variables que se considerarán de la base de datos original son:

- **pid**: Identificador del paciente;
- **year**: Año de la cirugía;
- **age**: Edad al momento de la cirugía (en años);
- **meno**: Estado menopáusico, donde 0 = premenopáusico y 1 = postmenopáusico;
- **size**: Tamaño del tumor, con niveles 20mm., entre 20 y 50mm., 50mm.;
- **grade**: Grado de diferenciación del tumor, con valores de 1 a 3;
- **pgr**: Receptores de progesterona (en fmol/l)
- **er**: Receptores de estrógeno (en fmol/l)
- **hormon**: Variable indicadora de haber recibido un tratamiento hormonal (1 = sí, 0 = no);
- **chemo**: Variable indicadora de haber recibido quimioterapia (1 = sí, 0 = no);
- **dtime**: Tiempo hasta la muerte o pérdida de seguimiento (en días);
- **death**: Variable indicadora de muerte (1 = muerte, 0 = censura).

## Base de datos a utilizar

Cada grupo contará con un subconjunto de la base original, considerando seis cohortes distintas según el año en que fue realizada la cirugía.

	Período
Grupo 1	1978 - 1985
Grupo 2	1986 - 1987
Grupo 3	1988
Grupo 4	1989 - 1990
Grupo 5	1991
Grupo 6	1992 - 1993

Para obtener la base de trabajo se presenta como ayuda el siguiente código, el cual no solo filtra según el período asignado sino también excluye las variables de la base original que no son consideradas en este trabajo:

```
library(survival)
library(tidyverse)

data <- rotterdam %>%

  ## Filtrar según el período asignado (ejemplo Grupo 1)
  filter(year <= 1985) %>% # Para filtrar entre años: year > 1985 & year <= 1987

  ## Eliminar las variables que no se utilizan
  select(-c(nodes, rtime, recur))
```

## Consigna

Aplicar técnicas de análisis de datos de duración para analizar y describir el conjunto de datos, para luego redactar un informe con los resultados obtenidos.

## Entrega

El trabajo práctico cuenta con tres instancias:

1. Presentación oral de un **análisis inicial** del problema. Deberán describir el conjunto de datos mediante técnicas gráficas y medidas resumen, y evaluar posibles diferencias entre grupos a través de tests de hipótesis. Fecha tentativa: fines de abril.

2. Presentación oral de un análisis utilizando **modelos de regresión de Cox**. Deberán realizar una evaluación completa: aplicar una selección de variables, evaluar la validez del modelo e interpretar los resultados. Fecha tentativa: principios de junio.
3. Entrega del **informe final**. Deberán plasmar en un informe todo el trabajo realizado, tanto la parte descriptiva como la del modelado. Fecha tentativa: a definir luego de la segunda presentación.

Las dos presentaciones son instancias de corrección y sugerencias. En ellas se marcarán cosas a mejorar que se espera que corrijan para la entrega final.

El formato de cómo presentar los resultados es una elección de cada grupo. Pueden optar por una presentación con diapositivas o bien, si ya comenzaron con la redacción del informe, mostrar directamente un archivo de tipo **.pdf**.

La entrega final del trabajo deberá contener un archivo en formato **.pdf** o similar con el informe final y un archivo que se pueda abrir mediante RStudio (**.R**, **.rmd**, **qmd**, etc.) donde se pueda replicar el análisis y los resultados que se encuentren en el informe.

El informe final puede contar como máximo con un total de **10 de páginas**, sin incluir portada ni índice (si se incluyera) pero sí considerando el anexo. Sean circiosos a la hora de definir qué es realmente necesario incluir en el informe. Hagan uso de las herramientas del software para la presentación de gráficos y sean concisos a la hora de interpretar los resultados y redactar sus conclusiones.

## Consideraciones finales

Algunas recomendaciones a tener en cuenta a la hora de realizar el trabajo:

- Los resultados deberán ser presentados no solo desde el punto de vista estadístico sino, principalmente, en **términos del problema**. Es decir, debe redactarse de forma tal que pueda ser entregado y comprendido por un equipo de investigación del área correspondiente, más que leído por docentes de una cátedra.
- En caso de querer presentar resultados puramente metodológicos, como fórmulas o gráficos específicos, utilice un **anexo**, donde puede volcar todo lo que considere que no debe formar parte del cuerpo principal del informe.
- Haga uso de las sugerencias que se encuentran en el campus con el nombre de *Recomendaciones TP*.