

# Análisis de Supervivencia - Máster en Técnicas Estadísticas

## Trabajos para la evaluación continua - Parte 2

2 de mayo de 2025

**Modo de entrega y plazo:** Cada alumno/a resolverá el ejercicio que le haya sido asignado, de manera individual. Enviará un informe en pdf con las respuestas a las cuestiones planteadas, así como un script ejecutable del código R utilizado, a la dirección **jacobo@uvigo.gal** antes del lunes 26 de mayo de 2025 (incluido). Los documentos deben nombrarse como ApellidosNombre.pdf y ApellidosNombre.R. Se valorará la corrección de las respuestas, y la claridad y la concisión en la presentación de resultados. El profesor podrá contactar con el/la alumno/a para resolver las dudas que surjan durante la corrección del ejercicio.

**Ejercicio 1.** Para los datos `bmt{KMsurv}` considérense los dos eventos competitivos 'recaída' y 'muerte en remisión' (individuos que mueren sin haber recaído en la enfermedad). Para el grupo 1 (ALL) se pide:

- 1.1 Estimar las intensidades de transición mediante un modelo de intensidades de transición constantes
- 1.2 Construir intervalos de confianza al 95 % para los parámetros del modelo
- 1.3 Representar gráficamente las intensidades de transición, las intensidades de transición acumuladas, y las funciones de incidencia acumulativa del modelo. ¿Qué proporción de individuos recaen a largo plazo? ¿Qué proporción muere en remisión a largo plazo?
- 1.4 Construir un gráfico de bondad de ajuste para el modelo utilizado

**Ejercicio 2.** Para los datos `bmt{KMsurv}` considérense los dos eventos competitivos 'recaída' y 'muerte en remisión' (individuos que mueren sin haber recaído en la enfermedad). Para el grupo 2 (AML Low Risk) se pide:

- 2.1 Estimar las intensidades de transición mediante un modelo de intensidades de transición Weibull

- 2.2 Contrastar al 5 % la hipótesis nula de que las intensidades de transición son homogéneas en el tiempo
- 2.3 Representar gráficamente las funciones de incidencia acumulativa del modelo. ¿Qué proporción de individuos recaen a largo plazo? ¿Qué proporción muere en remisión a largo plazo?
- 2.4 Construir un gráfico de bondad de ajuste para el modelo utilizado

**Ejercicio 3.** Para los datos `bmt{KMsurv}` considérense los dos eventos competitivos 'recaída' y 'muerte en remisión' (individuos que mueren sin haber recaído en la enfermedad). Para el grupo 3 (AML High Risk) se pide:

- 3.1 Estimar las intensidades de transición acumuladas mediante un método no paramétrico a 6, 12 y 18 meses. Aportar los correspondientes intervalos de confianza al 95 %
- 3.2 Estimar las funciones de incidencia acumulada a 6, 12 y 18 meses. Interpretar los valores obtenidos. Aportar los correspondientes intervalos de confianza al 95 %
- 3.3 Mediante métodos no paramétricos, dar respuesta a las siguientes preguntas, indicando las limitaciones encontradas en el análisis: (a) ¿Qué proporción de individuos recaen a largo plazo? (b) ¿Qué proporción muere en remisión a largo plazo?
- 3.4 ¿Crees que un modelo de intensidades de transición constantes sería adecuado para este estudio? Razonar la respuesta

**Ejercicio 4.** Para los datos `bmt{KMsurv}` considérense los dos eventos competitivos 'recaída' y 'muerte en remisión' (individuos que mueren sin haber recaído en la enfermedad). Se pide:

- 4.1 Mediante un modelo de regresión Weibull, estimar el efecto del grupo de riesgo sobre las intensidades de transición. ¿Es estadísticamente significativo al 5 %? Interpretar el efecto estimado tomando el grupo 1 (ALL) como grupo de referencia
- 4.2 Construir intervalos de confianza para los coeficientes de la regresión Weibull al 95 %. ¿Puede aceptarse al 5 % que las intensidades de transición son homogéneas en el tiempo?
- 4.3 Representar gráficamente las funciones de incidencia acumulada estimadas a partir del modelo Weibull para los tres grupos de riesgo. ¿Qué proporción de individuos sufren recaída en el largo plazo para cada grupo de riesgo?
- 4.4 Construir un gráfico de bondad de ajuste para el modelo utilizado

**Ejercicio 5.** Para los datos `bmt{KMsurv}` considérense los dos eventos competitivos 'recaída' y 'muerte en remisión' (individuos que mueren sin haber recaído en la enfermedad). Se pide:

- 5.1 Mediante un modelo de regresión de intensidades de transición proporcionales, estimar el efecto del grupo de riesgo. ¿Es estadísticamente significativo al 5 %? Obtener los factores multiplicativos de riesgo para las intensidades de transición tomando el grupo 1 (ALL) como grupo de referencia
- 5.2 Construir intervalos de confianza para los coeficientes de la regresión al 95 %
- 5.3 Representar gráficamente las intensidades de transición estimadas a partir del modelo para los tres grupos de riesgo
- 5.4 Construir un gráfico de bondad de ajuste para el modelo utilizado

**Ejercicio 6.** Para los datos `bmt{KMsurv}` considérense los dos eventos competitivos 'recaída' y 'muerte en remisión' (individuos que mueren sin haber recaído en la enfermedad). Se pide:

- 6.1 Mediante un modelo de regresión de riesgos de subdistribución proporcionales, estimar el efecto del grupo de riesgo. ¿Es estadísticamente significativo al 5 %? Obtener los factores multiplicativos de riesgo para los riesgos de subdistribución tomando el grupo 1 (ALL) como grupo de referencia
- 6.2 Discutir las limitaciones o posibles inconsistencias del modelo utilizado
- 6.3 Representar gráficamente las funciones de incidencia acumulada estimadas a partir del modelo para los tres grupos de riesgo
- 6.4 Construir un gráfico de bondad de ajuste para el modelo utilizado

**Ejercicio 7.** Considérense los  $n = 1835$  adultos (edad  $\geq 18$ ) con leucemia en el conjunto de datos `ebmt1{mstate}`, que pueden experimentar una recaída (evento 1) o una muerte en remisión (evento 2) a lo largo del seguimiento. Para el subconjunto de individuos con `score` de riesgo bajo, se pide:

- 7.1 Estimar las intensidades de transición mediante un modelo de intensidades de transición constantes
- 7.2 Construir intervalos de confianza al 95 % para los parámetros del modelo
- 7.3 Representar gráficamente las intensidades de transición, las intensidades de transición acumuladas, y las funciones de incidencia acumulativa del modelo. ¿Qué proporción de individuos recaen a largo plazo? ¿Qué proporción muere en remisión a largo plazo?
- 7.4 Construir un gráfico de bondad de ajuste para el modelo utilizado

**Ejercicio 8.** Considérense los  $n = 1835$  adultos (edad  $\geq 18$ ) con leucemia en el conjunto de datos `ebmt1{mstate}`, que pueden experimentar una recaída (evento 1) o una muerte en remisión (evento 2) a lo largo del seguimiento. Para el subconjunto de individuos con `score` de riesgo medio, se pide:

- 8.1 Estimar las intensidades de transición mediante un modelo de intensidades de transición Weibull
- 8.2 Contrastar al 5 % la hipótesis nula de que las intensidades de transición son homogéneas en el tiempo
- 8.3 Representar gráficamente las funciones de incidencia acumulativa del modelo. ¿Qué proporción de individuos recaen a largo plazo? ¿Qué proporción muere en remisión a largo plazo?
- 8.4 Construir un gráfico de bondad de ajuste para el modelo utilizado

**Ejercicio 9.** Considérense los  $n = 1835$  adultos (edad  $\geq 18$ ) con leucemia en el conjunto de datos `ebmt1{mstate}`, que pueden experimentar una recaída (evento 1) o una muerte en remisión (evento 2) a lo largo del seguimiento. Para el subconjunto de individuos con `score` de riesgo alto, se pide:

- 9.1 Estimar las intensidades de transición acumuladas mediante un método no paramétrico a 1, 2 y 3 años. Aportar los correspondientes intervalos de confianza al 95 %
- 9.2 Estimar las funciones de incidencia acumulada a 6, 12 y 18 meses. Interpretar los valores obtenidos. Aportar los correspondientes intervalos de confianza al 95 %
- 9.3 Mediante métodos no paramétricos, dar respuesta a las siguientes preguntas, indicando las limitaciones encontradas en el análisis: (a) ¿Qué proporción de individuos recaen a largo plazo? (b) ¿Qué proporción muere en remisión a largo plazo?
- 9.4 ¿Crees que un modelo de intensidades de transición constantes sería adecuado para este estudio? Razonar la respuesta

**Ejercicio 10.** Considérense los  $n = 1835$  adultos (edad  $\geq 18$ ) con leucemia en el conjunto de datos `ebmt1{mstate}`, que pueden experimentar una recaída (evento 1) o una muerte en remisión (evento 2) a lo largo del seguimiento. Se pide:

- 10.1 Mediante un modelo de regresión Weibull, estimar el efecto del `score` sobre las intensidades de transición. ¿Es estadísticamente significativo al 5 %? Interpretar el efecto estimado tomando el `score` de riesgo bajo como grupo de referencia
- 10.2 Construir intervalos de confianza para los coeficientes de la regresión Weibull al 95 %. ¿Puede aceptarse al 5 % que las intensidades de transición son homogéneas en el tiempo?
- 10.3 Representar gráficamente las funciones de incidencia acumulada estimadas a partir del modelo Weibull para los tres grupos de `score`. ¿Qué proporción de individuos sufren recaída en el largo plazo para cada grupo de riesgo?

10.4 Construir un gráfico de bondad de ajuste para el modelo utilizado

**Ejercicio 11.** Considérense los  $n = 1835$  adultos (edad  $\geq 18$ ) con leucemia en el conjunto de datos `ebmt1{mstate}`, que pueden experimentar una recaída (evento 1) o una muerte en remisión (evento 2) a lo largo del seguimiento. Se pide:

- 11.1 Mediante un modelo de regresión de intensidades de transición proporcionales, estimar el efecto del `score`. ¿Es estadísticamente significativo al 5 %? Obtener los factores multiplicativos de riesgo para las intensidades de transición tomando el `score` de riesgo bajo como grupo de referencia
- 11.2 Construir intervalos de confianza para los coeficientes de la regresión al 95 %
- 11.3 Representar gráficamente las intensidades de transición estimadas a partir del modelo para los tres grupos de `score`
- 11.4 Construir un gráfico de bondad de ajuste para el modelo utilizado

**Ejercicio 12.** Considérense los  $n = 1835$  adultos (edad  $\geq 18$ ) con leucemia en el conjunto de datos `ebmt1{mstate}`, que pueden experimentar una recaída (evento 1) o una muerte en remisión (evento 2) a lo largo del seguimiento. Se pide:

- 12.1 Mediante un modelo de regresión de riesgos de subdistribución proporcionales, estimar el efecto del `score`. ¿Es estadísticamente significativo al 5 %? Obtener los factores multiplicativos de riesgo para los riesgos de subdistribución tomando el `score` de riesgo bajo como grupo de referencia
- 12.2 Discutir las limitaciones o posibles inconsistencias del modelo utilizado
- 12.3 Representar gráficamente las funciones de incidencia acumulada estimadas a partir del modelo para los tres grupos de `score`
- 12.4 Construir un gráfico de bondad de ajuste para el modelo utilizado

**Ejercicio 13.** El archivo `electrical_appliance.txt` recoge el número de ciclos hasta el fallo y el tipo de fallo para 36 unidades de un electrodoméstico puestas a prueba. Se quiere estimar la intensidad de transición y la función de incidencia acumulativa para el fallo tipo 9. Se pide: (a) Realizar las citadas estimaciones bajo el supuesto de intensidades de transición constantes y bajo el supuesto de intensidades de transición tipo Weibull; (b) Realizar las citadas estimaciones por métodos no paramétricos, y construir gráficos de bondad de ajuste para los dos modelos utilizados en el apartado (a).

**Ejercicio 14.** Considérense los tiempos de desempleo (truncados por la izquierda y censurados por la derecha) de  $n = 1009$  mujeres gallegas casadas disponibles en el archivo `un_data.txt`. Se pide: (a) estimar mediante métodos

no paramétricos la función de supervivencia del tiempo total en desempleo; (b) estimar mediante métodos no paramétricos las funciones acumulativas de riesgo para los eventos 'encontrar empleo' y 'abandonar la búsqueda'; (c) discutir, a partir de las estimaciones realizadas en (b), la plausibilidad del modelo de intensidades de transición constantes; (d) representar gráficamente las funciones de incidencia acumulada de Aalen- Johansen para los dos eventos antedichos; (e) ¿podría asumirse que la variable de truncamiento está uniformemente distribuida? (f) estudiar el efecto que habría tenido en las respuestas a los apartados (a)–(d) ignorar la existencia de truncamiento

**Ejercicio 15.** Considérense los tiempos de desempleo (truncados por la izquierda y censurados por la derecha) de  $n = 1009$  mujeres gallegas casadas disponibles en el archivo `un_data.txt`. Se pide:

- 15.1 Estudiar el efecto de la edad al comenzar el período de desempleo (variable  $x$ ) sobre los eventos 'encontrar empleo' y 'abandonar la búsqueda' mediante modelos de intensidades de transición proporcionales. Comparar los resultados basados en el predictor lineal  $\beta x$  con los del predictor parabólico  $\beta_1 x + \beta_2 x^2$ . ¿Qué modelo es mejor? Discutir
- 15.2 Representar gráficamente las intensidades de transición acumulativas para las mujeres que entran en el desempleo a los 30, 35 y 40 años a partir del modelo con predictor parabólico del apartado anterior. Interpretar los resultados obtenidos
- 15.3 Estudiar el efecto que habría tenido en los resultados ignorar la existencia de truncamiento

**Ejercicio 16.** Considérense los tiempos de desempleo (truncados por la izquierda y censurados por la derecha) de  $n = 1009$  mujeres gallegas casadas disponibles en el archivo `un_data.txt`. Se pide:

- 16.1 Estudiar el efecto de la edad al comenzar el período de desempleo (variable  $x$ ) sobre los eventos 'encontrar empleo' y 'abandonar la búsqueda' mediante modelos de intensidades de transición proporcionales, incluyendo la edad categorizada de acuerdo con tres grupos: menores de 30 años, entre 30 y 40 años, y mayores de 40 años
- 16.2 Representar gráficamente las intensidades de transición acumulativas para los tres grupos de edad a partir del modelo del apartado anterior. Comentar
- 16.3 Construir un gráfico de bondad de ajuste para el modelo utilizado
- 16.4 Estudiar el efecto que habría tenido en los resultados ignorar la existencia de truncamiento

**Ejercicio 17.** Considérese un modelo con dos riesgos competitivos e intensidades de transición lineales  $\lambda_j(t) = \lambda_j t$ ,  $j = 1, 2$ . Se pide:

- 17.1 Obtener los estimadores de máxima verosimilitud de  $\lambda_1$  y  $\lambda_2$  a partir de una muestra aleatoria simple censurada del proceso
- 17.2 Calcular la función de supervivencia del tiempo de absorción  $T$
- 17.3 Demostrar que  $T$  y el indicador de evento  $\eta$  son variables aleatorias independientes
- 17.4 Calcular las funciones de incidencia acumulada y las probabilidades de absorción
- 17.5 Diseñar un algoritmo que simule el proceso. Para  $\lambda_1 = 1$  y  $\lambda_2 = 1/2$ , simular muestras aleatorias simples no censuradas del proceso con tamaños muestrales  $n = 100$  y  $n = 500$  y estudiar la precisión de los estimadores obtenidos en el primer apartado

**Ejercicio 18.** Si  $S(t_1, t_2) = (e^{\lambda_1 \theta t_1} + e^{\lambda_2 \theta t_2} - 1)^{-1/\theta}$ , con  $\lambda_1, \lambda_2, \theta > 0$ , es la función de supervivencia conjunta de dos tiempos de transición latentes  $T_1$  y  $T_2$ , se pide

- 18.1 Las supervivencias netas  $S_{T_j}(t)$
- 18.2 La supervivencia global  $S_T(t)$
- 18.3 Comprobar que las intensidades de transición son  $\lambda_j(t) = \lambda_j e^{\lambda_j \theta t} (e^{\lambda_1 \theta t} + e^{\lambda_2 \theta t} - 1)^{-1}$
- 18.4 Un script para simular los valores de  $(T, \eta)$
- 18.5 Ilustrar mediante simulaciones la posible inconsistencia del estimador de Kaplan-Meier de  $S_{T_1}(t)$  que interpreta el evento tipo 2 como censura independiente sobre  $T_1$

**Ejercicio 19.** Considérese un proceso con dos eventos competitivos y riesgos de subdistribución dados por  $\lambda_j^*(t) = e^{-\lambda_j t}$ ,  $\lambda_j > 0$ ,  $j = 1, 2$ . Se pide:

- 19.1 Calcular las funciones de incidencia acumulada  $F_j(t) = P(T \leq t, \eta = j)$  y las probabilidades de absorción  $p_j = P(\eta = j)$ ,  $j = 1, 2$
- 19.2 Utilizar el método de la inversión para diseñar un algoritmo que simule el proceso; en el primer paso se simula  $\eta$  y en el segundo se simula  $T$  condicionadamente al valor de  $\eta$  del primer paso
- 19.3 Para el caso  $\lambda_1 = \lambda_2 = 1$ , simular muestras aleatorias simples del proceso con tamaños muestrales  $n = 100$  y  $n = 500$  y estudiar la precisión de los estimadores de Aalen-Johansen de  $F_1(t)$  y  $F_2(t)$
- 19.4 Repetir el apartado anterior introduciendo censura por la derecha, de acuerdo con una variable  $C$  uniformemente distribuida en el intervalo  $[0, 5]$ . Discutir comparativamente los resultados de ambos apartados

**Ejercicio 20.** Para los datos `bmt{KMsurv}` considérense los dos eventos competitivos 'recaída' y 'muerte en remisión' (individuos que mueren sin haber recaído en la enfermedad). Para el grupo 1 (ALL) se pide:

- 20.1 Estimar las intensidades de transición mediante un modelo de intensidades de transición de tipo lognormal
- 20.2 Construir intervalos de confianza al 95 % para los parámetros del modelo
- 20.3 Representar gráficamente las intensidades de transición, las intensidades de transición acumuladas, y las funciones de incidencia acumulativa del modelo. ¿Qué proporción de individuos recaen a largo plazo? ¿Qué proporción muere en remisión a largo plazo?
- 20.4 Construir un gráfico de bondad de ajuste para el modelo utilizado