



Universidade de Vigo

---

MODELOS DE PREDICCIÓN CLÍNICA EN ESTUDIOS DE  
SUPERVIVENCIA

COMPARACIÓN DE DOS CLASIFICACIONES EN EL  
PRONÓSTICO DE PACIENTES CON SÍNDROME  
MIELODISPLÁSICO

---

Autora: Vanessa Balboa Barreiro

Directores: Francisco Gude Sampedro y César Sánchez Sellero

MÁSTER EN TÉCNICAS ESTADÍSTICAS

Julio 2013

---

# Índice general

<b>1. Introducción</b>	<b>3</b>
1.1. Descripción de la enfermedad . . . . .	3
1.2. Índices de pronóstico . . . . .	5
1.3. Datos disponibles en la unidad epidemiológica del Hospital Clínico Universitario de Santiago de Compostela . . . . .	8
1.4. Objetivos del estudio . . . . .	8
1.5. Estructura del trabajo . . . . .	9
<b>2. Análisis de supervivencia</b>	<b>11</b>
2.1. Censura . . . . .	12
2.2. Funciones elementales asociadas al tiempo de vida . . . . .	13
2.3. Estimador de Kaplan-Meier . . . . .	15
<b>3. Curvas ROC</b>	<b>19</b>
3.1. Sensibilidad y especificidad . . . . .	19
3.2. Descripción de una curva ROC . . . . .	21
3.2.1. Construcción de la curva ROC . . . . .	22

3.2.2. Curva ROC dependiente del tiempo . . . . .	22
3.2.3. Área bajo la curva . . . . .	23
3.3. Estimación de las curvas ROC . . . . .	23
3.4. Resultados para los datos de SMD . . . . .	24
<b>4. Evaluación de los modelos de pronóstico</b>	<b>29</b>
4.1. Ajuste general: Puntuación de Brier . . . . .	29
4.2. Discriminación . . . . .	31
4.2.1. Índice de concordancia C . . . . .	32
4.3. Reclasificación . . . . .	34
4.3.1. Mejora neta de reclasificación . . . . .	34
<b>5. Nueva propuesta con el modelo de regresión de Cox</b>	<b>41</b>
5.1. Contrastes de hipótesis . . . . .	42
5.1.1. Test de razón de verosimilitud . . . . .	42
5.1.2. Test de Wald . . . . .	43
5.1.3. Test de las puntuaciones ( <i>score test</i> ) . . . . .	43
5.2. Análisis de residuos . . . . .	44
5.3. Resultados . . . . .	44
<b>6. Otras aplicaciones realizadas en la Unidad de Epidemiología</b>	<b>51</b>

---

# Capítulo 1

## Introducción

Los modelos diagnósticos y pronósticos son muy comunes en Medicina, sobre todo en Medicina Clínica, y forman una parte importante de la investigación que se lleva a cabo en los hospitales y centros de atención primaria. Entre sus usos se pueden destacar:

- Diferenciar estados de la enfermedad.
- Clasificar la severidad de la enfermedad.
- Evaluar el riesgo de presentar futuras enfermedades.
- Estratificar el riesgo con el objetivo de tomar decisiones diagnósticas o terapéuticas.

Los modelos diagnósticos y pronósticos tienen propósitos diferentes. La diagnosis se ocupa de indagar acerca del estado actual del paciente tratando de identificar la existencia de una enfermedad hasta ese momento desconocida, por ejemplo un cáncer. Los modelos pronósticos añaden el elemento temporal. Aunque típicamente se refiere al curso clínico que posiblemente desarrollará un individuo enfermo, también se aplica a la predicción del riesgo de enfermedad en población sana. Pronóstico y predicción implican estimación del riesgo o la probabilidad de desarrollar una determinada enfermedad en el futuro.

### 1.1. Descripción de la enfermedad

Los Síndromes Mielodisplásicos (SMD) constituyen un grupo de trastornos en los que la médula ósea no funciona correctamente y las células de la misma no pueden producir suficientes células sanguíneas saludables. Para más información sobre esta enfermedad se pueden consultar las guías españolas de diagnóstico y tratamiento de los síndromes mieloplásticos y la leucemia mielomonocítica crónica, disponibles en [www.gesmd.es/es/](http://www.gesmd.es/es/).

Las células madre de la sangre, constituidas por una pequeña fracción de las células productoras de sangre, son necesarias para producir nuevas células. Estas se dividen originando dos células: una linfoide, la cual permanecerá como célula madre y, otra mieloide, que puede seguir cambiando y dividiéndose para producir células sanguíneas. La médula ósea sana produce células sanguíneas inmaduras, denominadas blastos, que luego se desarrollan dando lugar a:

- *Glóbulos rojos*: encargados de transportar oxígeno y otras sustancias a todos los tejidos del cuerpo.
- *Glóbulos blancos*: también conocidos como leucocitos, importantes en la defensa del organismo contra las infecciones y enfermedades. Los dos tipos principales de glóbulos blancos son: linfocitos, encargados de producir los anticuerpos que ayudan al organismo a combatir los gérmenes; y granulocitos, encargados de destruir las bacterias.
- *Plaquetas*: necesarias para la coagulación de la sangre, bloqueando las áreas de vasos sanguíneos dañadas por cortes o hematomas.

Por el contrario, en pacientes con el síndrome mielodisplásico (SMD), este proceso normal es interrumpido. La médula ósea produce células inmaduras, es decir, los blastos no se desarrollan plenamente. El organismo destruye muchas de estas células sanguíneas anormales, dejando al paciente con bajos recuentos sanguíneos, ya que no existen suficientes células sanguíneas normales. Como consecuencia de ello, los pacientes con SMD tienen menos células sanguíneas maduras y las que tienen pueden ser anormales y no funcionar adecuadamente. Cuando hay menos células sanguíneas sanas, se pueden presentar infecciones, anemia o sangrado fácil. El signo más común es la anemia, provocada por la insuficiencia de glóbulos rojos en la sangre. Esta afección puede causar que las personas se sientan cansadas, débiles y pueden lucir pálidas. Un estado grave podría causar dificultad para respirar. El resultado de una disminución de plaquetas, conocida como trombocitopenia, puede causar sangrado anormal o aparición de hematomas con mucha facilidad. Usualmente, los linfocitos no son anormales en los síndromes mielodisplásicos, en cambio, los neutrófilos, células cruciales para combatir las bacterias, pueden aparecer de forma anormal. Un déficit de neutrófilos, conocido como neutropenia, puede tener como consecuencia graves infecciones. En la Figura 1.1 se aprecia gráficamente la evolución de una célula sanguínea. Los SMD son por naturaleza heterogéneos y diversos con tendencia a evolucionar a leucemia mieloblástica aguda (LMA). En la LMA, las células madre mieloides habitualmente pasan a ser un tipo de glóbulos blancos inmaduros llamados mieloblastos (o blastocitos mieloides). Los mieloblastos en la LMA son anormales y no se convierten en glóbulos blancos sanos. Dando lugar, en ocasiones, a demasiados glóbulos rojos o plaquetas anormales, también denominados células o blastocitos leucémicos. Las células leucémicas se acumulan en la médula ósea y la sangre dejando menos lugar para las células sanas. Cuando esto ocurre, se puede presentar infección, anemia o sangrado fácil. Las células leucémicas se pueden diseminar fuera de la sangre hasta otras partes del cuerpo, como el sistema nervioso central (cerebro y médula espinal), la piel y las encías.

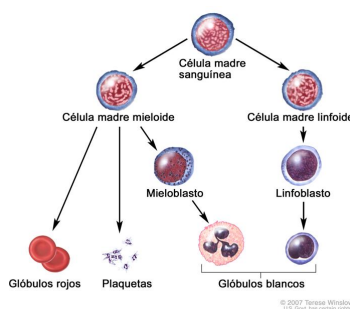


Figura 1.1: Una célula madre sanguínea pasa por varias etapas para convertirse en un glóbulo rojo, una plaqueta o un glóbulo blanco.

## 1.2. Índices de pronóstico

Como corresponde a un grupo heterogéneo de enfermedades, los SMD presentan gran variabilidad pronóstica, tanto en términos de supervivencia global como de riesgo de evolución a LMA. De ahí la necesidad de establecer un sistema de clasificación diagnóstica y pronóstica de este complejo grupo de neoplasias hematológicas.

La primera clasificación diagnóstica de los SMD fue establecida en 1982 por el grupo Franco-Americano-Británico (FAB) y se ha mantenido en vigor hasta la revisión realizada bajo el amparo de la Organización Mundial de la Salud (OMS) en el año 2008. En cualquier caso, la categorización pronóstica más extendida actualmente es el Índice Pronóstico Internacional (International Prognostic Scoring System, IPSS).

El índice pronóstico internacional, IPSS, divide a los pacientes en cuatro categorías pronósticas según el número de citopenias (disminución de un tipo celular de la sangre), ciertas características citogenéticas (cariotipo) y el porcentaje de blastos en la médula ósea, según se indica en el cuadro 1.1. Como habían mostrado estudios previos, las tres variables demostraron presentar un peso pronóstico independiente, tanto para la supervivencia global como para la evolución a LMA.

Variable Pronóstica	Grado				
	0	0.5	1	1.5	2
Blastos (%)	< 5	5 - 10	-	11 - 20	21 - 30
Cariotipo	bueno	intermedio	pobre	-	-
Citopenias	0 / 1	2 - 3	-	-	-

Tabla 1.1: Clasificación pronóstica (IPSS).

En la Tabla 1.1 se representa la clasificación pronóstica obtenida al usar el índice IPSS.

A cada intervalo de valores de las variables pronósticas, se le asigna una puntuación, a la que denominamos grado. Esta puntuación puede tomar valores entre 0, 0.5, 1, 1.5 y 2. En el caso del porcentaje de blastos, a un porcentaje inferior al 5 % se le adjudica el grado 0, si un individuo presenta un porcentaje entre el 5 y el 10 %, se le asigna un grado 0.5, a un porcentaje entre el 11 y el 20 % se le asigna el grado 1.5 mientras que a un porcentaje superior se le asignará el grado 2. La variable cariotipo representa las anomalías genéticas presentes en el individuo, obteniendo un mayor grado cuanto mayor sea el número de anomalías. Se denomina como *bueno* al cariotipo normal, es decir, cuando no existen anomalías, al que se le asigna un grado 0. El término *intermedio* recoge cualquier otra anomalía sin afectar al cromosoma 7 (grado 0.5) y, *pobre* se usa para referirse a un cariotipo complejo (grado 1), es decir, con un número de anomalías superior a 3, en el cual sí se ve afectado el cromosoma 7. En cuanto a las citopenias se otorga el grado cero a la no existencia de citopenias ó uni-citopenia (disminución de un tipo celular de la sangre). Se clasifica con grado 0.5 a la bicitopenia ó pancitopenia, es decir, la disminución de dos ó todos los tipos celulares de la sangre circulante. Se usa (–) para hacer referencia a un caso no aplicable.

El sistema de pronóstico IPSS es capaz de estratificar los pacientes en cuatro grupos con diferencias estadísticamente significativas en supervivencia global y tasa de progresión a LMA. Esta estratificación se realiza teniendo en cuenta la puntuación total del índice IPSS y definiendo cuatro intervalos. La puntuación total se obtiene como la suma de las puntuaciones de las tres variables.

La situación clínica del paciente, según el sistema de pronóstico IPSS sería la siguiente:

- *Riesgo bajo*: se corresponde con una puntuación igual a 0 y se denominará grupo 0.
- *Riesgo intermedio 1*: se corresponde con una puntuación en el intervalo 0.5 - 1 y se denominará grupo 1.
- *Riesgo intermedio 2*: se corresponde con una puntuación entre 1.5 y 2, se denominará grupo 2.
- *Riesgo alto*: se corresponde con una puntuación superior a 2. Este riesgo se denomina como grupo 3.

Por ejemplo, un paciente con un porcentaje de blastos inferior al 5 % (grado 0), un cariotipo pobre (grado 1) y pancitopenia (grado 0.5) sería clasificado como:

$$0 + 1 + 0,5 = 1,5$$

Es decir, sería un individuo con riesgo intermedio 2, pertenecería al grupo 2.

Desde su publicación ha sido considerado el sistema pronóstico de referencia para establecer el pronóstico individual y planificar el tratamiento en dicha enfermedad. Sin embargo,

el IPSS presenta importantes limitaciones, entre las que se encuentra la asignación inadecuada del riesgo para muchas alteraciones citogenéticas, especialmente en las de riesgo intermedio 1, o no tener en cuenta la importancia clínica de las severas neutropenias o plaquetopenias (disminución de granulocitos o plaquetas en la sangre, respectivamente) que determinan la necesidad de tratamiento terapéutico. En los últimos años se han reconocido nuevos factores de la enfermedad y del paciente con peso pronóstico independiente y se ha modificado el IPSS, dando lugar al IPSS revisado (IPSS-R).

El índice revisado sigue basándose en las variables presentes en el IPSS, pero incluye cinco en lugar de tres subgrupos citogenéticos con nuevas y específicas clasificaciones de un número de subconjuntos citogenéticos menos comunes. Esta revisión recoge la influencia peyorativa de la trombocitopenia y la neutropenia graves (polimorfonucleares [PMN]). Un trabajo muy reciente del Grupo Español de Síndromes Mielodisplásicos (GESMD) ha evidenciado que ambas anomalías citogenéticas tienen un peso pronóstico independiente tanto en supervivencia global como en progresión a LMA en pacientes con SMD de bajo riesgo.

La información disponible, aún no publicada, sobre el núcleo fundamental del sistema de pronóstico IPSS revisado, se muestra en la Tabla 1.2.

V. Pronóstico	Grado						
	0	0.5	1	1.5	2	3	4
Cariotipo	Muy bueno	-	Bueno	-	Intermedio	Pobre	Muy pobre
Blastos (%)	$\leq 2$	-	$> 2 - 5$	-	5 - 10	$> 10$	-
Hemoglobina	$\geq 10$	-	$8 - < 10$	$< 8$	-	-	-
Plaquetas	$\geq 100$	$50 - < 100$	$< 50$	-	-	-	-
PMN	$\geq 0,8$	$< 0,8$	-	-	-	-	-

Tabla 1.2: IPSS-R provisional.

El sistema pronóstico tradicional, el IPSS, hacía cuatro grupos, en cambio, el sistema IPSS-R define cinco categorías principales de pronóstico, o de riesgo.

- *Riesgo muy bajo*: se corresponde con una puntuación entre 0 y 1.5 puntos.
- *Riesgo bajo*: se corresponde con una puntuación en el intervalo (1.5, 3].
- *Riesgo intermedio*: se corresponde con una puntuación perteneciente al intervalo (3, 4.5].
- *Riesgo alto*: se corresponde con una puntuación perteneciente al intervalo (4.5, 6].
- *Riesgo muy alto*: se corresponde con una puntuación superior a 6 puntos.



El tratamiento de los pacientes con SMD de alto riesgo se enfoca a modificar la historia natural de la enfermedad y prolongar la esperanza de vida. Por el contrario, en los pacientes con SMD de bajo riesgo, el objetivo terapéutico se enfoca más hacia la mejora de la calidad de vida, con el control de los síntomas y el tratamiento de las citopenias, fundamentalmente de la anemia.

### 1.3. Datos disponibles en la unidad epidemiológica del Hospital Clínico Universitario de Santiago de Compostela

Con el propósito de describir la epidemiología de los Síndromes Mielodisplásicos en la población gallega, se realizó un estudio de 340 pacientes procedentes de la base de datos de la Fundación Pública Gallega de Medicina Xenómica. Para su inclusión debían haber sido diagnosticados de SMD según los criterios de la clasificación de la Organización Mundial de la Salud de 2008 en el período comprendido entre el 1 de enero de 2007 y el 18 de diciembre de 2011. Para ello se revisaron los estudios citogenéticos remitidos a dicho laboratorio con sospecha de SMD entre tales fechas. El período de seguimiento mínimo fue de un año, finalizando por lo tanto, el 17 de diciembre de 2012.

En la Tabla 1.3 se muestra un resumen de la base de datos disponible.

	IPSS-R					
IPSS	1	2	3	4	5	Total
0	62	77	4	0	0	143
1	7	68	44	10	0	129
2	0	0	15	24	15	54
3	0	1	0	1	12	14
Total	69	146	63	35	27	340

Tabla 1.3: Clasificación de los individuos según los dos sistemas pronósticos.

### 1.4. Objetivos del estudio

Debido a la gravedad de la enfermedad y a la necesidad de dar un tratamiento adecuado a la situación clínica de cada paciente, es fundamental tener un sistema de diagnóstico y pronóstico lo más eficaz posible. Es aquí donde radica la importancia de evaluar si el nuevo sistema de pronóstico mejora la eficacia del usado hasta el momento.

Finalmente, tras evaluar cuál de los dos sistemas de pronóstico es más eficiente, se procederá a la comprobación de si al añadir una covariable al mejor de ellos, se podría incluso mejorar los resultados obtenidos con ese sistema.

### 1.5. Estructura del trabajo

En el capítulo dos se pretende dar un marco de referencia básico sobre el análisis de supervivencia. Comenzando con una breve descripción de los tipos de censura existentes en los datos de supervivencia, así como las posibles soluciones a adoptar. Finalmente se presenta un compendio de las funciones básicas del análisis de supervivencia.

En el capítulo tres nos acercamos a los estudios de clasificación. En este capítulo se expone y describe la capacidad de discriminación de las curvas ROC y las curvas ROC dependientes del tiempo.

En el capítulo cuatro se consideran otras medidas de evaluación de un modelo. Además trataremos brevemente el tema de reclasificación.

En el capítulo cinco se estudia la información que aporta la inclusión de una nueva variable al marcador IPSS-R. El método usado es el modelo de Cox.

En el capítulo seis se presenta un breve resumen de las prácticas realizadas en la Unidad de Epidemiología del Complejo Hospitalario de Santiago de Compostela.



---

## Capítulo 2

# Análisis de supervivencia

En ciencias de la salud los estudios longitudinales son sin duda, los que aportan datos más precisos sobre la evolución de los sujetos pero, a la vez, son los más costosos y los de mayor complejidad en cuanto a su tratamiento estadístico (véase Stevenson (2009)). Se conoce como análisis de supervivencia a la colección de técnicas estadísticas apropiadas para este tipo de estudios, en los cuales cada paciente es seguido durante un determinado período de tiempo. Con el uso de estas técnicas se pretende describir y cuantificar el tiempo hasta un evento de interés.

En el análisis de supervivencia se emplea el término “fallo” para definir que sucede el evento de interés. Dicho evento, en el contexto médico, puede ser en realidad un “éxito”, como por ejemplo la recuperación de una terapia o la desaparición de un síntoma; o un fallo, como la recaída en una enfermedad o la muerte. El término “tiempo de supervivencia” se define como el tiempo transcurrido hasta que ocurre el evento a estudiar, es decir, es el intervalo de tiempo desde el acontecimiento o estado inicial hasta el estado final. Por ese motivo, es importante tener conocimiento de las fechas implicadas en el estudio.

En el análisis de supervivencia, el análisis de los datos puede ser realizado utilizando técnicas paramétricas, semi-paramétricas o no paramétricas. Entre los métodos paramétricos destaca el uso de las distribuciones Exponencial, Weibull y Lognormal para modelar la supervivencia. Entre los métodos no paramétricos o semi-paramétricos, los más usados son el método de Kaplan-Meier (véase Kaplan and Meier (1958)) y la regresión de Cox (véase Cox (1972)).

En epidemiología se utilizan con más frecuencia los métodos no paramétricos (o semi-paramétricos) que los métodos paramétricos. Este hecho se debe a que normalmente no se tiene una distribución teórica ajustada a los datos.

## 2.1. Censura

En estudios con datos longitudinales solo se conoce el tiempo exacto de supervivencia para aquellos individuos que realizaron el evento de interés durante el período de seguimiento. Sin embargo, aquellos individuos que no presentaron el evento una vez finalizado el período de observación o aquellos que son perdidos, el observador puede pensar que dichos sujetos no mostraron el evento de interés durante el período de seguimiento. Este sería un caso de censura. La respuesta, el tiempo hasta el evento, es una variable continua para la cual existen dificultades para llevar a cabo su estudio. En primer lugar, en la mayor parte de los estudios, la variable no sigue una distribución normal sino que tiene una distribución asimétrica. Una posible solución sería utilizar transformaciones que normalicen la variable. Una segunda dificultad se encuentra en el hecho de necesitar un período de observación suficientemente largo, en el cual se suelen “perder” observaciones, es decir, existen observaciones censuradas. Por ejemplo, en el supuesto de un seguimiento de 5 años tras un infarto de miocardio, donde el paciente  $i$  continua vivo. El tiempo de supervivencia de dicho paciente está censurado por la derecha tras esos 5 años, esto es, su tiempo de supervivencia es conocido únicamente hasta el momento de finalización del estudio, no se conoce lo que pasará tras ese período. La censura también puede ocurrir si se pierde el sujeto durante el seguimiento. Estas razones justifican el uso de métodos específicos para el estudio de este tipo de variable respuesta, el método más usado es el análisis de supervivencia, llegando a ser una técnica de vital importancia cuando se tienen observaciones censuradas.

Existen distintos tipos de censura: censura por la derecha, por la izquierda e intervalos de censura (véase Andersen y otros (1993)).

Se tienen *intervalos censurados* cuando, por ejemplo, el dispositivo de medición sólo funciona para un rango determinado de la respuesta; las medidas fuera de ese rango serán censuradas. Es decir, el tiempo de supervivencia es conocido únicamente si es medido en un intervalo concreto. Se puede encontrar este tipo de censura en estudios clínicos con seguimientos periódicos.

*Censura por la izquierda* significa que sólo se tiene conocimiento de los tiempos de supervivencia para los eventos que han ocurrido antes de un cierto tiempo. Un ejemplo sería observar un evento que ha ocurrido antes de que el individuo se incluya en el estudio. Por tanto, se desconoce el valor exacto de la observación.

Se tiene *censura por la derecha* cuando solo es conocida una cota inferior para el tiempo del evento de interés. Ante este tipo de censura se dice que una observación  $Z$  (tiempo de vida del individuo) es censurada por la derecha si se desconoce el valor de  $Z$  y, si sólo se sabe que es mayor que  $C$ .

Existen diferentes tipos de respuesta censurada por la derecha:

*Censura tipo I:* se fija una cota inferior común a todos los individuos. Sea  $T$  el tiempo de supervivencia y  $C$  el tiempo de censura, siendo esta última una constante. Suponiendo que  $T$  y  $C$  son independientes a lo largo del tiempo, los datos observados son:

$$Z = \min(T, C) = \begin{cases} T, & \text{para observaciones censuradas} \\ C, & \text{para observaciones no censuradas} \end{cases}$$

$$\delta = \mathbb{I}(T \leq C) = \begin{cases} 1, & \text{para observaciones censuradas} \\ 0, & \text{para observaciones no censuradas} \end{cases}$$

Siendo  $\delta$  la función indicadora de censura. Aquí las respuestas son observadas sólo si se encuentran por debajo de un valor fijo  $C$ .

*Censura tipo II:* el estudio continua hasta el fallo de los primeros  $r$  individuos, donde  $r$  es un número entero predeterminado ( $r < n$ ).

La determinación del tiempo para la censura de tipo I y el número de fallos para la censura de tipo II deben ser establecidos antes de iniciar el análisis.

*Censura aleatoria:* en este caso el tiempo de censura es una variable aleatoria, es decir, durante el transcurso de un estudio, existe un suceso independiente del de interés que provoca la salida de  $r$  individuos del estudio.

El modelo aleatorio de censura a la derecha se presenta con el par  $(Z, \delta)$  donde:

$$Z = \min(T, C) = \begin{cases} T, & \text{para observaciones censuradas} \\ C, & \text{para observaciones no censuradas} \end{cases}$$

$$\delta = \mathbb{I}(T \leq C) = \begin{cases} 1, & \text{para observaciones censuradas} \\ 0, & \text{para observaciones no censuradas} \end{cases}$$

Cada valor de la muestra será un par  $(Z_i, \delta_i)$ , con  $\delta_i = 1$  cuando el valor observado de  $Z$  corresponde con el tiempo de fallo del individuo y,  $\delta_i = 0$  cuando el valor registrado de  $Z$  es un tiempo de censura, es decir, no se ha observado el tiempo de fallo porque un suceso aleatorio e independiente del fallo ha ocurrido en el tiempo  $T_i$  interrumpiendo la observación de la vida hasta su fallo.

## 2.2. Funciones elementales asociadas al tiempo de vida

Sea  $T$  una variable aleatoria no negativa que representa el tiempo hasta la ocurrencia de un evento. En este capítulo, adoptaremos la terminología del análisis de supervivencia

refiriéndonos al evento de interés como “muerte” y al tiempo de espera como “tiempo de supervivencia”.

$T$  es una variable aleatoria continua en  $[0, \infty)$ . Sea  $f(t)$  la función de densidad de probabilidad de  $T$  y sea  $F(t)$  su correspondiente función de distribución acumulada:

$$F(t) = P\{T \leq t\} = \int_0^t f(x)dx$$

La *función de supervivencia* es:

$$S(t) = 1 - F(t) = P\{T > t\} = \int_t^{\infty} f(x)dx$$

Su valor en el tiempo  $t$  es igual a la probabilidad de que un individuo sobreviva por lo menos hasta el tiempo  $t$ . De forma más general, es igual a la probabilidad de que el individuo experimente el fallo con posterioridad al tiempo  $t$ . Es útil para comparar la progresión temporal hacia el fallo entre diferentes grupos de individuos. Se trata de una función continua decreciente, siendo  $S(0) = 1$  y  $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$ .

Otra función de gran utilidad en el análisis de supervivencia es la *función de riesgo* (o *Hazard rate*), denominada como  $\lambda(t)$ . La cual especifica la tasa instantánea de ocurrencia del evento en un intervalo de tiempo, dado que el individuo sobrevivió hasta el instante  $t$ . La función de riesgo se define como:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P\{t < T \leq t + \Delta t \mid T > t\}}{\Delta t}$$

El numerador de la expresión representa la probabilidad condicional de que el evento ocurra en el intervalo  $(t, t + \Delta t)$  dado que no ha ocurrido con anterioridad. El denominador es la longitud del intervalo. Tomando el límite cuando la longitud del intervalo tiende a 0, se obtiene la tasa instantánea de ocurrencia. La probabilidad condicionada en el numerador es lo mismo que la probabilidad conjunta de que  $T$  está en el intervalo  $(t, t + \Delta t)$  y  $T > t$ .

Es posible definir  $\lambda(t)$  como una tasa entre la función de densidad y la función de supervivencia:

$$\lambda(t) = \frac{f(t)}{S(t)}$$

la cual, algunos autores dan como la definición de función de riesgo.

Al tratarse  $f(t)$  de la derivada de la función  $S(t)$ , se puede reescribir la ecuación anterior como:

$$\lambda(t) = -\frac{d}{dt} \log S(t)$$

Si se integra la ecuación anterior entre 0 y  $t$  y se introduce la condición de contorno  $S(0) = 1$  (ya que es seguro que el evento no ha ocurrido en un tiempo anterior al instante 0), se obtiene una nueva fórmula para expresar la probabilidad de sobrevivir al menos hasta el instante  $t$  como la función de riesgo para todos los tiempos hasta  $t$ :

$$S(t) = \exp\left\{-\int_0^t \lambda(x)dx\right\} \quad (2.1)$$

La integral ubicada entre corchetes en la ecuación 2.1 se conoce como el *riesgo acumulado* y, se denota como:

$$\Lambda(t) = \int_0^t \lambda(x)dx = -\log S(t)$$

Esta función representa la suma de los riesgos acumulados desde el instante 0 hasta el tiempo  $t$ . Se trata de una función creciente en el intervalo  $[0, \infty)$

### 2.3. Estimador de Kaplan-Meier

El método de Kaplan-Meier se basa en los tiempos de supervivencia individual, asumiendo que la censura es independiente del tiempo de supervivencia. La probabilidad de sobrevivir en un período de tiempo puede ser calculada considerando pequeñas particiones de este tiempo. Estas particiones (o intervalos) se determinan por la ocurrencia del evento, es decir, la probabilidad de supervivencia es calculada cada vez que ocurre un evento. Por esa razón, entre un evento y otro, la probabilidad de supervivencia permanece constante.

Se trata del estimador no paramétrico de máxima verosimilitud de la función de supervivencia  $S(t)$ . Se define como el siguiente producto:

$$\widehat{S}(t) = \prod_{Z_{i:n} \leq t} \left(1 - \frac{1}{n - i + 1}\right)^{\delta_{[i:n]}}$$

donde  $Z_{[1:n]} \leq \dots \leq Z_{[n:n]}$  es la muestra ordenada de los  $Z_1, \dots, Z_n$  y  $\delta_{[1:n]}, \dots, \delta_{[n:n]}$  son los concomitantes, siendo  $n$  el tamaño de la muestra. El exponente es el indicador de censura. El denominador  $(n - i + 1)$  describe el número de pacientes en riesgo antes del instante  $Z_{[i:n]}$ , es decir, el número de individuos de la muestra para los cuales puede ocurrir el evento. En el caso de no existir censura, se trataría de la cantidad de supervivientes.

Los cálculos de supervivencia para los pacientes con SMD serán ilustrados en la Tabla 2.1, donde se recoge un extracto de los datos propuestos en este documento. Dicha tabla muestra los intervalos de tiempo (en días), el número de individuos en riesgo, el número de eventos sucedidos en cada instante, la probabilidad de supervivencia en los tiempos dados, su error estándar y los correspondientes intervalos de confianza al 95 %. El \* representa las observaciones censuradas. La probabilidad de supervivencia permanece en 1 hasta el tiempo del primer evento (2 días).



time	n <sup>o</sup> risk	n <sup>o</sup> event	survival	std. err	lower 95 % IC	upper 95 % IC
2	340	1	0.997	0.00294	0.991	1.000
3	339	1	0.994	0.00415	0.986	1.000
8	338	1	0.991	0.00507	0.981	1.000
16	337	2	0.985	0.00653	0.973	0.998
19	335	1*	0.982	0.00714	0.968	0.996
24	333	1	0.979	0.00770	0.964	0.995
27	332	1	0.967	0.00823	0.960	0.993
28	331	1*	0.973	0.00871	0.957	0.991
30	329	2	0.967	0.00961	0.949	0.987
34	327	1	0.965	0.01003	0.945	0.984

Tabla 2.1: Tabla de vida para pacientes diagnosticados con SMD.

Este estimador admite una representación gráfica por medio de una función escalonada a la que se denomina *curva de Kaplan-Meier*, donde la proporción de supervivencia permanece constante entre eventos. La información aportada por la tabla anterior se ve de forma más clara al representar la curva de K-M, Figura 2.1.

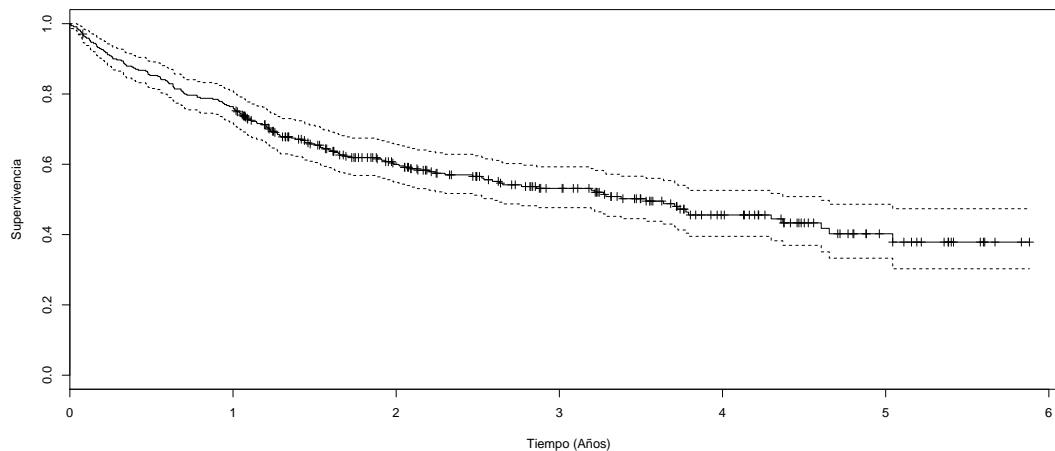


Figura 2.1: Curva de supervivencia Kaplan-Meier correspondiente a los pacientes diagnosticados con SMD.

En la Figura 2.1 se representa la curva de supervivencia correspondiente a los pacientes con Síndrome Mielodisplásico, junto con la banda de confianza al 95%. Los tiempos censurados son indicados con un (+) sobre la curva de supervivencia, los cuales facilitan la búsqueda de los tiempos de supervivencia de los sujetos a los cuales no les ha ocurrido el evento. La presencia de líneas horizontales largas, que indican que no existen cambios en la probabilidad de supervivencia estimada por un período, no siempre implican una evidencia de buenos logros. En ocasiones, se deben a una carencia de observaciones.

Observando la gráfica se ve como evoluciona la mortalidad a lo largo del período de estudio. Al año del estudio se puede apreciar una supervivencia del 76,4%, esa supervivencia desciende gradualmente hasta llegar al final del estudio, donde se tiene una supervivencia del 37,9%, es decir, durante el estudio se acumula una mortalidad del 62,1%.

En investigaciones médicas, con frecuencia, se quiere comparar la supervivencia de dos o más grupos de individuos. Dichos grupos pueden diferir respecto a cierto factor, como puede ser un marcador, la edad, el sexo, el tratamiento, etc. El efecto que ejerce este factor sobre la supervivencia es lo que interesa conocer. En base a los datos usados como ejemplo ilustrativo en este documento, interesa representar las curvas de K-M correspondientes a los distintos grupos de cada sistema de puntuación.

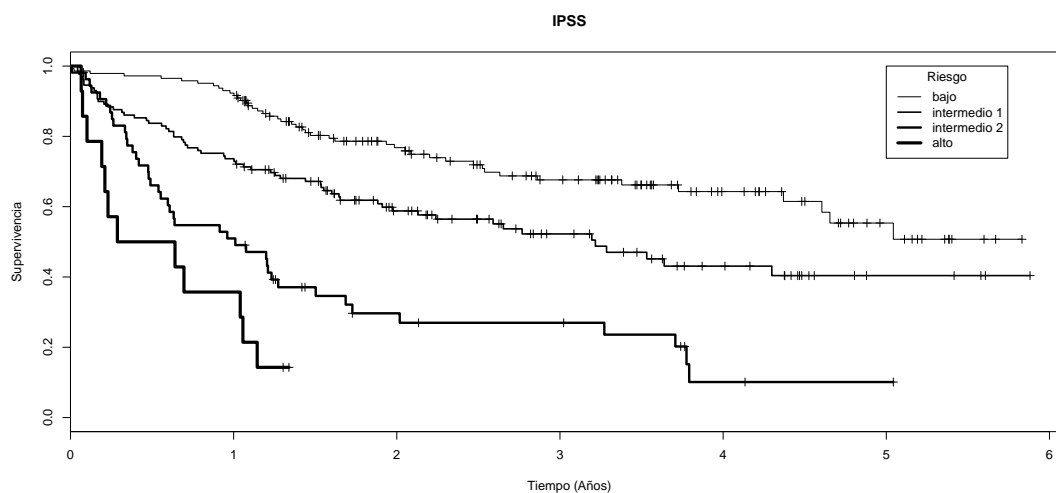


Figura 2.2: Curva Kaplan-Meier correspondiente a los pacientes clasificados mediante el IPSS.

En la Figura 2.2 se observan las distintas curvas de supervivencia descritas por los cuatro grupos pertenecientes a la clasificación mediante el sistema tradicional IPSS. La primera impresión que ofrece la gráfica es que existe una mayor censura en los grupos de menor riesgo y la existencia de una leve tendencia de agrupación entre las curvas de bajo y las de alto riesgo, llegando a diferenciarse a simple vista. Se puede apreciar como las curvas de K-M de los dos grupos de menor riesgo siguen trayectorias similares, sin presentar grandes cambios en la supervivencia. En cambio, las curvas correspondientes a los riesgos más altos presentan una caída brusca durante el primer año de seguimiento, llegando a alcanzar una probabilidad de supervivencia del 20% aproximadamente, en el caso del riesgo alto.

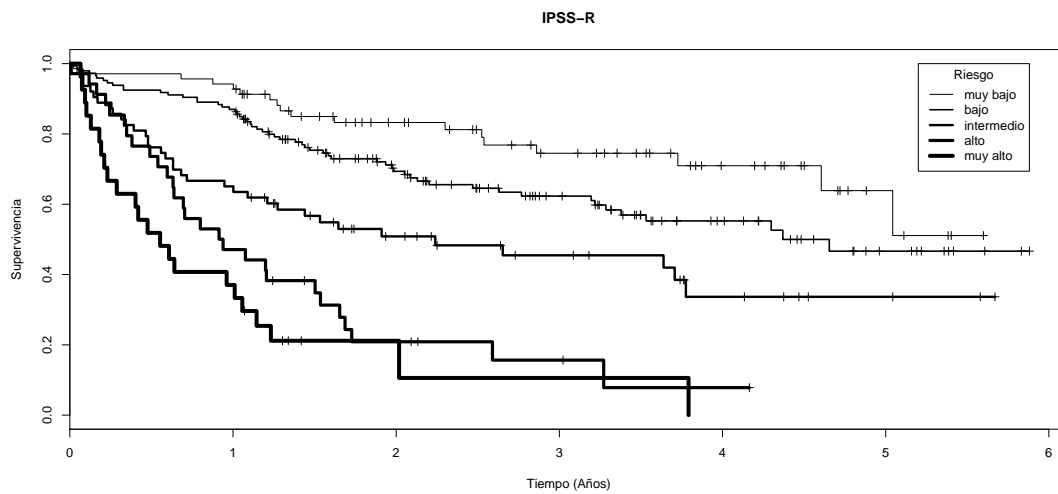


Figura 2.3: Curva Kaplan-Meier correspondiente a los pacientes clasificados mediante el IPSS-R.

En la Figura 2.3 sigue evidenciándose la agrupación entre las curvas de riesgos similares, formando dos grandes grupos. En este caso, como sucedía en el caso del sistema IPSS, las curvas de los riesgos más bajos presentan una caída más gradual en la supervivencia. En cambio, en los riesgos más altos, la caída es más agresiva, sobre todo en el primer año de seguimiento.

---

## Capítulo 3

# Curvas ROC

Uno de los problemas más comunes en las ciencias de la salud es el de clasificación o toma de decisiones clínicas. La clasificación se lleva a cabo a través de pruebas diagnósticas y/o pronósticas, para las cuales es fundamental conocer su exactitud y precisión. En caso contrario, por ejemplo, un paciente podría ser clasificado como sano cuando en realidad presenta la enfermedad.

En este caso disponemos de dos sistemas pronósticos, IPSS e IPSS-R, capaces de determinar el riesgo de SMD y de predecir la supervivencia y la progresión a LMA en pacientes con esta enfermedad. Nuestro objetivo es comprobar si la clasificación propuesta, IPSS-R, es más eficiente que el sistema usado hasta el momento, es decir, si la capacidad discriminatoria del sistema IPSS se ve mejorada por el sistema IPSS-R.

La curva ROC (*Receiver Operating Characteristic*) es una herramienta de uso muy extendido para evaluar la precisión de estos tests. A continuación, procedemos a la descripción de una curva ROC y nos centraremos en el caso de la curva ROC dependiente del tiempo, explicando cómo se estima y qué interpretación se deriva de ella.

Para ello debemos comenzar dando unas notaciones básicas, como la definición de los términos sensibilidad y especificidad.

### 3.1. Sensibilidad y especificidad

Las curvas ROC son muy usadas en la predicción y clasificación de marcadores cuando la variable respuesta es binaria, en este caso la precisión se resume a través de las tasas de clasificación correctas, definidas como *sensibilidad* =  $P\{X > c \mid Y_i = 1\}$ , y *especificidad* =  $P\{X \leq c \mid Y_i = 0\}$ . Donde  $Y_i = 1$  si el individuo  $i$  presenta la enfermedad y,  $Y_i = 0$

si está sano.  $X$  es el valor del marcador y  $c$  es el punto de corte o umbral de decisión, el cual clasifica como predicción positiva ( $X > c$ ) o predicción negativa ( $X \leq c$ ).

La sensibilidad es definida como la fracción de verdaderos positivos ( $FVP$ ) entre el número total de pacientes que presentan la enfermedad. Debemos especificar que un *verdadero positivo* es aquel paciente que presenta la enfermedad y tiene un resultado positivo en la prueba clínica. Por tanto, la ( $FVP$ ) es la probabilidad de clasificar correctamente a un individuo enfermo o, lo que es lo mismo, de clasificar a un verdadero positivo ( $VP$ ).

Este parámetro es especialmente importante cuando la enfermedad no debe pasar desapercibida y/o cuando el pronóstico mejora mucho con un tratamiento precoz.

$$\text{Sensibilidad} = \frac{VP}{VP + FN} = FVP$$

La especificidad es la fracción de verdaderos negativos ( $FVN$ ) entre el número total de pacientes que no presentan la enfermedad. Es la probabilidad de clasificar correctamente a un individuo sano, es decir, la probabilidad de que para un sujeto sano se obtenga un resultado negativo, verdadero negativo ( $VN$ ).

Este parámetro es importante cuando la enfermedad cambia la calidad de vida del paciente o las consecuencias del tratamiento pueden suponer un riesgo para el paciente, como una amputación por ejemplo. También es de gran importancia cuando no se quiere preocupar inútilmente a alguien con un falso positivo.

$$\text{Especificidad} = \frac{VN}{VN + FP} = FVN$$

En la Figura 3.1 se recoge la relación entre el resultado de una prueba diagnóstica y la presencia o ausencia de una enfermedad. La cantidad de sujetos realmente enfermos es la suma de los individuos colocados en las celdas VP y FN. Los realmente sanos se obtienen sumando los individuos ubicados en las celdas VN y FP.

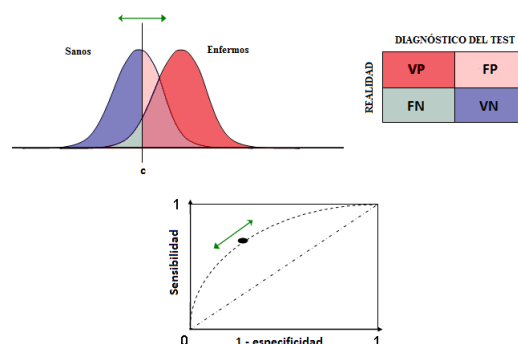


Figura 3.1: Referencias: VP: verdaderos positivos, VN: verdaderos negativos, FP: falsos positivos y FN: falsos negativos;  $c$  es el punto de corte.

Para clasificar un paciente como positivo o negativo es necesario establecer un punto de corte ( $c$ ) para la probabilidad predictiva. Si la predicción es más alta que el punto de corte, el paciente será clasificado como positivo, en caso contrario, como negativo. Se pueden tomar diferentes puntos de corte y representar gráficamente los resultados en la curva ROC. Existe una zona de posibles resultados de la prueba para la que la distribución de sujetos sanos y de sujetos enfermos se solapan, serían las zonas FN y FP representadas en la Figura 3.1. Al mover el punto de corte hacia la izquierda se aumenta la probabilidad de detectar pacientes enfermos, es decir, estamos aumentando el número de falsos positivos. Si se desplaza el punto de corte hacia la derecha, disminuirán los falsos positivos pero, en cambio, aumentará la cantidad de falsos negativos. Es decir, un aumento de sensibilidad implica una disminución de la especificidad, y viceversa.

### 3.2. Descripción de una curva ROC

Esta curva aporta información completa sobre el conjunto de todas las posibles combinaciones de las tasas de verdaderos positivos y falsos positivos. Una curva ROC se suele dibujar para visualizar y seleccionar clasificadores basados en su rendimiento (véase Fawcett (2006)). Se trata de gráficos bidimensionales en los cuales se dibuja la tasa, o fracción, de verdaderos positivos (*sensibilidad*) en el eje vertical, y la tasa de falsos positivos ( $1 - \text{especificidad}$ ) en el eje horizontal, Figura 3.2. Así mismo, un gráfico ROC representa el equilibrio relativo entre verdaderos positivos y falsos negativos. Siendo el vértice  $(0, 1)$ , vértice superior izquierdo, el representante de la clasificación perfecta. Por el contrario, la recta con ángulo de 45 grados representa la *no información*, es decir, el modelo no aporta información sobre la clasificación del marcador.

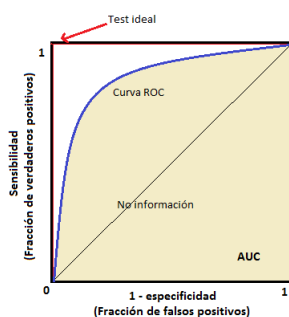


Figura 3.2: Curva ROC.

### 3.2.1. Construcción de la curva ROC

Existen diversos métodos para la construcción de las curvas ROC, los cuales dependen del tipo de datos que se manejen. Los datos clínicos pueden ser discretos o continuos. La mayoría de los datos de laboratorio son continuos pero, también es muy frecuente agrupar los datos continuos en categorías ordenadas para resumir la información y simplificar los cálculos. O puede suceder, como en este caso, que la variable respuesta sea discreta. En el caso del SMD, la clasificación de los individuos se lleva a cabo a través de los sistemas pronósticos IPSS e IPSS-R, variables respuestas, los cuales son variables discretas.

Para construir la curva ROC se empieza por calcular los puntos de *sensibilidad* =  $FVP$  y  $1 - \textit{especificidad}$  =  $FFP$  para todos los posibles puntos de corte,  $c$ , de la prueba. Si se tratara de una escala continua, estos puntos serían todos los valores observados, en cambio, si los resultados han sido agrupados en intervalos, los puntos de corte son los límites de estos intervalos. Por último, si la escala es discreta, los puntos de corte son todos los posibles valores de la escala. Hay que tener en cuenta que se debe considerar como otro punto de corte el punto  $-\infty$ . En este caso, al tratarse de marcadores discretos, la unión de los puntos no dará lugar a una curva diferenciable como en el caso de un marcador continuo, sino que dará lugar a un polígono. Los puntos  $(0, 0)$  y  $(1, 1)$  siempre pertenecen a la curva ROC.

Cada punto de corte  $c$  origina un punto diferente en el espacio ROC, que representa un par  $(FFP, FVP) = (1 - \textit{especificidad}(c), \textit{sensibilidad}(c))$ , los cuales definen la curva ROC.

### 3.2.2. Curva ROC dependiente del tiempo

Nuestro propósito es dar una predicción a cierto tiempo. Por lo tanto es necesario tener en cuenta la variable temporal involucrada en el estudio. La curva ROC es insuficiente, por ese motivo se consideran las curvas ROC dependientes del tiempo, denotadas como  $ROC(t)$  (véase Heagerty et al. (2000)). De esta forma son necesarias nuevas definiciones para la sensibilidad y especificidad dependientes del tiempo, fuertemente ligadas al concepto de máxima verosimilitud. Las medidas de precisión pueden ser usadas para calcular las curvas ROC para un tiempo específico. Heagerty propone dos tipos de estimadores para las curvas ROC, los cuales son válidos en presencia de datos censurados.

Cuando la variable respuesta depende del tiempo, se puede representar la curva  $ROC(t)$  como *sensibilidad*( $t$ ) vs. *1-especificidad*( $t$ ). Las definiciones propuestas por Heagerty para el caso dependiente del tiempo son:

$$\textit{Sensibilidad}(c, t) = P\{X > c \mid D(t) = 1\} = s_1(c, t)$$

$$\textit{Especificidad}(c, t) = P\{X \leq c \mid D(t) = 0\} = 1 - s_0(c, t)$$

Siendo  $X$  el marcador y  $D_i(t) = 1$  el indicador de que el individuo  $i$  ha sufrido un evento previo al instante  $t$  y  $c$  el punto de corte.

### 3.2.3. Área bajo la curva

El área bajo la curva, denotada como AUC, es conocida como un representante de la medida de concordancia entre el marcador y el indicador de la enfermedad. En curvas dependientes del tiempo, el  $AUC(t)$  estima la capacidad de distinguir o de discriminar entre enfermos y no enfermos, en este caso, la capacidad discriminatoria de cada sistema pronóstico en un cierto tiempo  $t$ . Se define como:

$$AUC(t) = \int_0^1 ROC_t(p) dp$$

donde  $p \in [0, 1]$  denota la fracción de falsos positivos.

Si el área bajo la curva valiese 1, es decir, valiese el 100 %, la prueba sería perfecta, ya que clasificaría al 100 % de los enfermos como enfermos ( $VP$ ) y al 100 % de los sanos como sanos ( $VN$ ). En este caso no existen FP ni FN. En cambio, si el área bajo la curva valiese 0.5 estaría discriminando un 50 %, es decir, existiría la misma probabilidad de clasificar a un enfermo como enfermo que como sano. Un área de 0.5 bajo la curva equivale a no discriminar y, se interpreta como una prueba “no informativa”.

## 3.3. Estimación de las curvas ROC

La estimación de las curvas ROC correspondientes a cada sistema pronóstico se lleva a cabo, de forma muy sencilla, usando las curvas Kaplan-Meier disponibles para cada grupo (ver capítulo 2), junto con las definiciones dadas por Heagerty y otros (2000) para la construcción de los pares ( $FFP, FVP$ ).

Se empieza definiendo la función bivalente  $S(c, t) = P(X > c, T > t)$ . Consideramos  $\widehat{S}(c, t)$  el estimador propuesto por Akitas (1994) para la supervivencia, definido de la siguiente manera:

$$\widehat{S}(c, t) = \frac{1}{n} \sum_i \widehat{S}(t | X = X_i) \mathbb{I}\{X_i > c\}$$

donde  $\widehat{S}(t | X = X_i)$  es el estimador adecuado de la función de supervivencia condicional y,  $c$  es el punto de corte correspondiente al grupo del marcador a estudio. En nuestro caso, donde el marcador toma un conjunto discreto de valores, el estimador se puede obtener



de manera sencilla mediante la siguiente expresión:

$$\widehat{S}(c, t) = \frac{1}{n} \sum_{j=1}^k n_j \mathbb{I}\{j > c\} \widehat{S}_j(t)$$

siendo  $k$  el número de valores posibles del marcador (cuatro para el IPSS y cinco para el IPSS-R),  $n_j$  el número de individuos que toman el valor  $j$  del marcador y  $\widehat{S}_j(t)$  el estimador Kaplan-Meier de la supervivencia para el grupo de individuos con valor  $j$  en el marcador.

Siguiendo esta notación podemos definir  $\widehat{S}(t) = \widehat{S}(-\infty, t)$ , que sería una estimación de la función de supervivencia como promedio de los estimadores Kaplan-Meier para cada valor del marcador.

La sensibilidad a un tiempo  $t$  se puede expresar así:

$$P\{X > c \mid D(t) = 1\} = \frac{P\{X > c, D(t) = 1\}}{P\{D(t) = 1\}} \quad (3.1)$$

El denominador de la expresión anterior se puede reescribir como:

$$P\{D(t) = 1\} = P\{T \leq t\} = 1 - S(t) \quad (3.2)$$

mientras que el numerador admite la expresión

$$P\{X > c, D(t) = 1\} = P\{X > c, T \leq t\} = P\{X > c\} - P\{X > c, T > t\} \quad (3.3)$$

De esta manera es posible estimar la sensibilidad a un tiempo  $t$  sustituyendo el numerador y el denominador de la expresión (3.1) por versiones empíricas de los términos a la derecha en (3.2) y (3.3). Como resultado, se obtiene el siguiente estimador de la sensibilidad a un tiempo  $t$ :

$$\widehat{P}\{X > c \mid D(t) = 1\} = \frac{\{1 - \widehat{F}_x(c)\} - \widehat{S}(c, t)}{1 - \widehat{S}(t)}$$

donde  $\widehat{F}_x$  representa la distribución empírica de los valores observados del marcador,  $X_1, \dots, X_n$ .

De manera similar la especificidad se puede estimar así:

$$Especificidad = \widehat{P}\{X \leq c \mid D(t) = 0\} = 1 - \frac{\widehat{S}(c, t)}{\widehat{S}(t)}$$

### 3.4. Resultados para los datos de SMD

El objetivo de este capítulo es analizar la capacidad discriminadora de ambos sistemas pronósticos mediante curvas ROC, es decir, ver como de buenos son los marcadores distinguiendo entre aquellos individuos que han sufrido el evento en un instante previo al

tiempo  $t$  y, los que no han sufrido el evento en ese período. La discriminación de cada sistema pronóstico es fácil de visualizar mediante los gráficos de las correspondientes curvas ROC. Al tratarse de datos de supervivencia, debemos usar curvas ROC dependientes del tiempo, prefijando los tiempos  $t$  de interés. En este caso se hacen tres cortes en el tiempo (1, 3 y 5 años ó lo que es lo mismo, 365, 1095 y 1825 días), para resumir la evolución del pronóstico de supervivencia a cierto tiempo.

Una vez fijados los tiempos a interés, es necesario determinar los puntos de corte. En el caso de los datos disponibles, los puntos de corte para el marcador IPSS toman los valores  $\{0, 1, 2, 3\}$ , correspondientes a los valores que toma el marcador. En el caso del marcador IPSS-R, los puntos de corte toman los valores  $\{1, 2, 3, 4, 5\}$ . A estos valores hay que añadirles el valor  $-\infty$ , punto de corte común a todas las curvas ROC que da lugar al par (1,1) en el espacio ROC.

El siguiente paso es calcular el conjunto de verdaderos positivos y falsos positivos para la construcción de la curva  $ROC(t)$  en cada instante  $t$ .

Los resultados obtenidos para el sistema pronóstico IPSS se muestran en la Tabla 3.1.

		$c = 0$	$c = 1$	$c = 2$	$c = 3$
$t = 365$	sens (c,t)	0.86	0.44	0.11	0
	espec (c,t)	0.51	0.87	0.98	1
	1- espec (c,t)	0.49	0.13	0.02	0
	(sens, 1- espec)	(0.86, 0.49)	(0.44, 0.13)	(0.11, 0.02)	(0, 0)
$t = 1095$	sens (c,t)	0.71	0.33	0.08	0
	espec (c,t)	0.54	0.92	0.99	1
	1- espec (c,t)	0.46	0.08	0.01	0
	(sens, 1- espec)	(0.71, 0.46)	(0.33, 0.08)	(0.08, 0.01)	(0, 0)
$t = 1825$	sens (c,t)	0.71	0.31	0.07	0
	espec (c,t)	0.58	0.96	0.99	1
	1- espec (c,t)	0.42	0.04	0.01	0
	(sens, 1- espec)	(0.71, 0.42)	(0.31, 0.04)	(0.07, 0.01)	(0, 0)

Tabla 3.1: Conjunto de verdaderos positivos y falsos positivos para el sistema pronóstico IPSS, siendo  $t = 365, 1095$  y  $1825$  días.

Las curvas ROC en el tiempo  $t$  serán el resultado de la unión de los conjuntos de puntos obtenidos, dando lugar a las curvas representadas en la Figura 3.3.

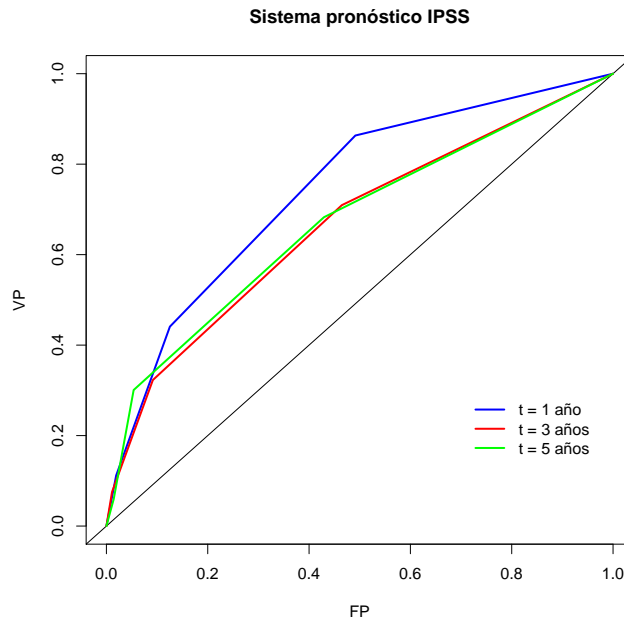


Figura 3.3: Curvas ROC dependientes del tiempo correspondientes al sistema pronóstico IPSS en los tiempos  $t = 1, 3$  y  $5$  años.

Los resultados para el sistema pronóstico IPSS-R se muestran en la Tabla 3.2.

		$c = 1$	$c = 2$	$c = 3$	$c = 4$	$c = 5$
$t = 365$	sens ( $c,t$ )	0.95	0.74	0.44	0.21	0
	espec ( $c,t$ )	0.25	0.75	0.90	0.95	1
	1- espec ( $c,t$ )	0.75	0.25	0.10	0.04	0
	(sens, 1- espec)	(0.95, 0.75)	(0.74, 0.25)	(0.44, 0.10)	(0.21, 0.04)	(0, 0)
$t = 1095$	sens ( $c,t$ )	0.89	0.55	0.33	0.15	0
	espec ( $c,t$ )	0.29	0.79	0.95	0.98	1
	1- espec ( $c,t$ )	0.71	0.21	0.05	0.02	0
	(sens, 1- espec)	(0.89, 0.71)	(0.55, 0.21)	(0.33, 0.05)	(0.15, 0.02)	(0, 0)
$t = 1825$	sens ( $c,t$ )	0.88	0.50	0.30	0.13	0
	espec ( $c,t$ )	0.35	0.84	0.98	1	1
	1- espec ( $c,t$ )	0.65	0.16	0.02	0.00	0
	(sens, 1- espec)	(0.88, 0.65)	(0.50, 0.16)	(0.30, 0.02)	(0.13, 0.00)	(0, 0)

Tabla 3.2: Conjunto de verdaderos positivos y falsos positivos para el sistema pronóstico IPSS-R, siendo  $t = 365, 1095$  y  $1825$  días.

Las curvas resultantes para el sistema pronóstico IPSS-R se ven representadas en la Figura 3.4. Se puede ver que las curvas no son líneas diferenciables. Se observa que las curvas

dan lugar a polígonos, por tratarse de marcadores discretos.

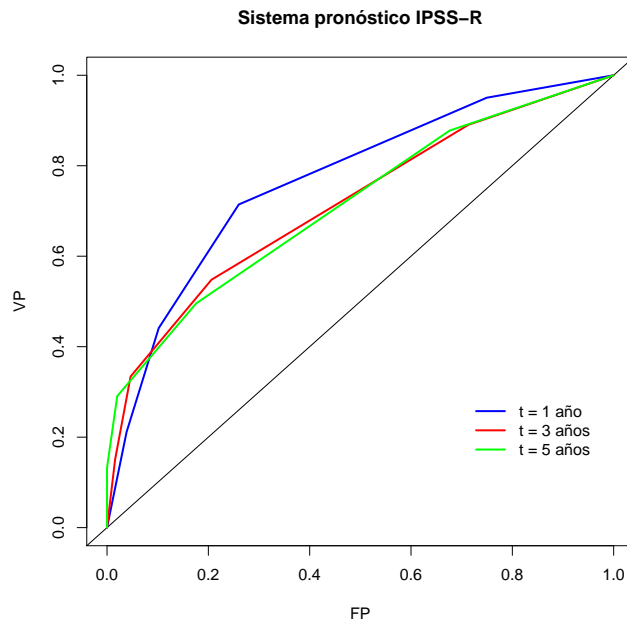


Figura 3.4: Curvas ROC dependientes del tiempo correspondientes al sistema pronóstico IPSS-R en los tiempos  $t = 1, 3$  y  $5$  años.

En la Tabla 3.3 se muestra el AUC correspondiente a cada sistema pronóstico en los tiempos  $t = 1, 3$  y  $5$  años. Se tiene que el AUC en cada instante  $t$ , sufre un ligero aumento en el sistema pronóstico IPSS-R frente al tradicional.

	IPSS	IPSS-R
$t = 1$ año	74.29	76.77
$t = 3$ años	66.64	71.47
$t = 5$ años	67.23	71.30

Tabla 3.3: Resumen  $AUC(t)$  para cada sistema pronóstico.

Una vez calculada la AUC dependiente del tiempo  $t$ , otra medida discriminatoria muy útil es la conocida como área integrada bajo la curva, (*iauc*), en el instante  $t$ . El test de Wilcoxon para muestras dependientes (véase Kains et al. (2008)) permite el contraste entre dos áreas bajo la curva para un tiempo  $t$  determinado. Como hipótesis nula se considera que el AUC del nuevo marcador no supera el AUC del tradicional. El contraste es el siguiente:

$$\begin{cases} H_0 : AUC_2(t) \leq AUC_1(t) \\ H_1 : AUC_2(t) > AUC_1(t) \end{cases}$$

Los resultados otorgados por este test se ven reflejados en la Tabla 3.4. El test concluye que para tiempos pequeños, no existen diferencias estadísticas entre ambos sistemas pronósticos, es decir, la nueva propuesta no mejora al marcador tradicional. En cambio, según avanza el tiempo, se prueba que existen diferencias estadísticas entre el sistema IPSS y el sistema IPSS-R, evidenciando una mayor área bajo la curva para el sistema IPSS-R.

	IPSS	IPSS-R	p-valor
t = 1 año	72.51	72.21	0.96
t = 3 años	70.69	72.84	4.37e-117
t = 5 años	69.92	72.74	3.63e-258

Tabla 3.4: Resultados del test de Wilcoxon para ambos sistemas pronósticos.

Hasta el momento se han mostrado las curvas ROC para tiempos concretos pero, también es interesante ver la evolución de dichas curvas para cada sistema pronóstico a lo largo del período de seguimiento, verificando los resultados mostrados anteriormente. Esta visualización es posible dibujando el valor de área bajo cada curva para cada tiempo  $t$  durante el período de estudio, que se muestra en la Figura 3.5.

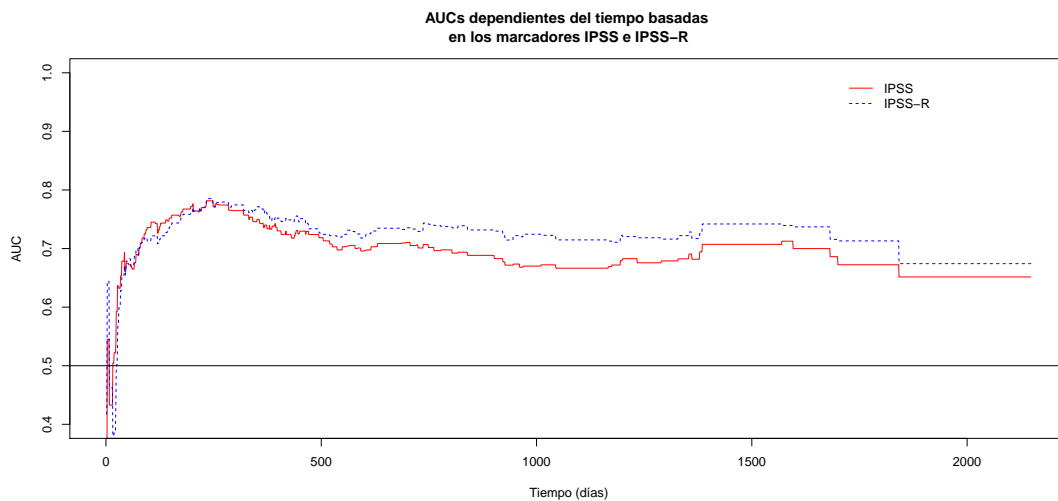


Figura 3.5: AUCs correspondientes a cada sistema pronóstico en cada instante durante el período de seguimiento.

Así, en la Figura 3.5 se observa que al comienzo del seguimiento existe solapamiento entre los dos sistemas. En cambio, a partir del primer año de seguimiento, aproximadamente, comienzan a existir diferencias entre ambos sistemas de puntuaciones, posicionándose el sistema de pronóstico IPSS-R sobre el sistema tradicional.

---

# Capítulo 4

## Evaluación de los modelos de pronóstico

Es fundamental evaluar el rendimiento de los marcadores. Con este propósito, se desarrollaron diversas medidas de evaluación que se pueden agrupar en tres grandes categorías: medidas de *ajuste general*, medidas de *calibración* y medidas de *discriminación*. Para una revisión detallada de este tema, se puede consultar Steyerberg (2009).

### 4.1. Ajuste general: Puntuación de Brier

Entre los parámetros más usados para evaluar el modelo están las medidas de rendimiento global, donde destacan el estadístico  $R^2$  y el sistema de puntuaciones de Brier (véase Brier (1950)). Nos centraremos en la puntuación de Brier.

La puntuación de Brier fue desarrollada para la predicción de la inexactitud de las predicciones meteorológicas. Más tarde se extendió para ser utilizada como una medida de precisión en problemas de supervivencia con presencia de censura (véase Graf y otros (1999)). Esta medida presenta diferente definición dependiendo de la existencia o no de censura en los datos.

Dado un tiempo  $t$ , la puntuación de Brier para un solo sujeto se define como el cuadrado de la diferencia entre el estado de supervivencia observada y la predicción basada en el modelo de supervivencia planteado en el tiempo  $t$ ,

$$BS(t, \hat{S}) = E(Y_i(t) - \hat{S}(t | X_i))^2 \quad (4.1)$$

donde  $Y_i = \mathbb{I}\{T_i > t\}$  indica el verdadero estado del sujeto  $i$ ,  $\hat{S}(t | X_i)$  es la estimación de la probabilidad de supervivencia en el tiempo  $t$  para el sujeto  $i$  y  $X_i$  es el marcador

correspondiente. En la expresión (4.1) se define la puntuación de Brier como el error de predicción esperado en un cierto tiempo.

Al considerar el caso en el que se observa censura en los datos, la expresión para el cálculo de la puntuación de Brier pasa a ser una ponderación de la expresión para el caso sin censura, expresión (4.1).

Sea  $G(t) = P(C > t)$  la función de supervivencia para la censura, es decir, la probabilidad de estar libre de censura hasta el momento  $t$ . Se denota  $\hat{G}(t)$  como el estimador Kaplan-Meier de  $G(t)$ . En estas condiciones, se define la puntuación de Brier de la siguiente forma:

$$BS(t | \hat{S}) = \frac{1}{n} \sum_{i=1, \dots, n} \begin{cases} \frac{(0 - \hat{S}(t|X_i))^2}{\hat{G}(Z_i)}, & Z_i \leq t, \delta_i = 1 \\ \frac{(1 - \hat{S}(t|X_i))^2}{\hat{G}(t)}, & Z_i > t \\ 0, & Z_i \leq t, \delta_i = 0 \end{cases}$$

donde  $1/\hat{G}(Z_i)$  es el peso considerado si ocurre el evento antes del tiempo  $t$  y,  $1/\hat{G}(t)$  si el evento ocurre después del tiempo  $t$ .

La puntuación de Brier integrada (IBS) es una medida general que resume el error de predicción hasta el instante  $t$ . Su definición viene dada por:

$$IBS = \frac{1}{T} \int_0^T BS(t | \hat{S}) dt$$

Se trata de una puntuación para evaluar la bondad de las funciones de supervivencia previstas de todas las observaciones en cada tiempo  $t$  entre 0 y  $T$  siendo  $T = \max(Z_i)$ , ( $i = 1, \dots, n$ ) para cada intervalo de tiempo considerado. Se basa en la predicción de las curvas de supervivencia. Toma valores entre 0, para un modelo perfecto, y 0.25, indicando que el modelo no aporta información. Es decir, cuanto menor sea la puntuación de Brier, mejores serán las predicciones, del mismo modo, cuanto menor sea el valor del IBS, menor será el error de predicción cometido a lo largo del período de tiempo definido, por lo tanto, los resultados obtenidos serán más precisos.

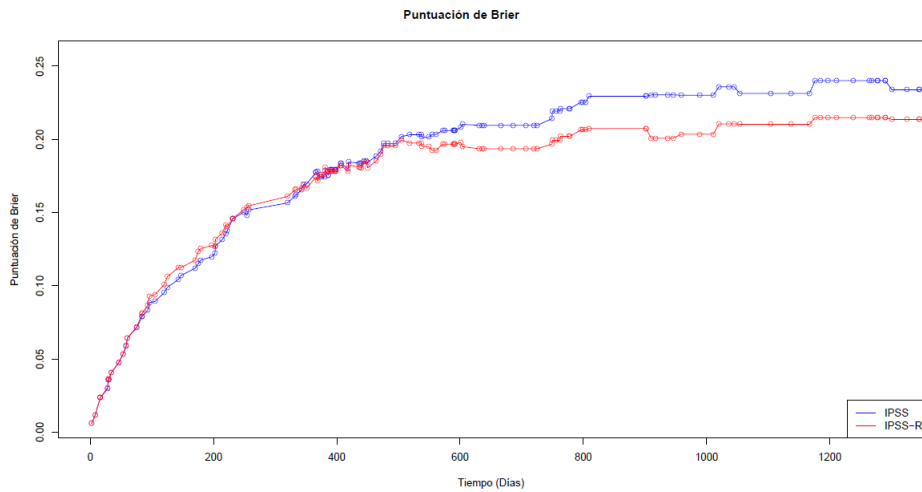


Figura 4.1: Representación de las puntuaciones de Brier para cada sistema pronóstico.

En la Figura 4.1 se representa la puntuación de Brier para cada tiempo  $t$ . No se representa el período de seguimiento por completo ya que, a partir del instante  $t = 1825$  días, se tienen menos observaciones. Al observar la Figura 4.1 se concluye que las puntuaciones de Brier son muy cercanas para ambos sistemas en tiempos pequeños. En el instante  $t = 400$  aproximadamente, se aprecia un cruce entre las dos puntuaciones, pasando a estar por encima la puntuación correspondiente al sistema pronóstico IPSS. Por lo tanto, se tienen mejores predicciones con el nuevo sistema propuesto. En la Tabla 4.1 se resume esta información de forma más detallada.

	$t = 365$ días	$t = 1095$ días	$t = 1825$ días
IPSS	0.103	0.168	0.188
IPSS-R	0.103	0.161	0.179

Tabla 4.1: Puntuación de Brier integrada para ambos sistemas pronósticos.

Lo que se observa en esta tabla es el valor de la puntuación de Brier integrada tomando como tiempos máximos tres puntos de corte en la escala temporal (365, 1095 y 1825 días).

## 4.2. Discriminación

La discriminación es la capacidad del modelo de asignar el resultado correcto a un par de sujetos seleccionados al azar; en otras palabras, permite distinguir entre los que desarrollarán el evento y los que no lo harán.



Este concepto indica cómo de eficaz es el modelo de predicción discriminando a los individuos que realizan el evento frente a los individuos que no realizan el evento.

La medida de discriminación más frecuentemente utilizada para modelos de error normal y resultado binario es el área bajo la curva (AUC). Su equivalente para datos con observaciones censuradas o modelos de regresión no logística es el índice de concordancia, índice C, (véase Harrell y otros (2001)). En el caso logístico, coinciden ambas medidas de discriminación.

En presencia de datos de supervivencia, el índice C es una generalización de las AUC(t), vistas en el capítulo 3. En este capítulo nos centraremos en el índice de concordancia.

### 4.2.1. Índice de concordancia C

El índice C calcula la probabilidad de que, para un par de pacientes comparables elegidos al azar, el paciente con probabilidad de riesgo más alta experimente un evento antes que el paciente con menor riesgo. Hasta el momento habíamos hablado de medidas a cierto tiempo pero, el índice C se trata de una medida global muy relacionada con el AUC.

Para entender la relación existente entre esta medida de discriminación global y el AUC cuando las curvas ROC no son dependientes del tiempo, se toman dos observaciones independientes  $(X_{i,i})$  y  $(X_{j,j})$  donde,  $X_i, X_j$  son los valores del marcador para los pacientes  $i, j$ , con  $i, j \in [1, n]$  siendo  $i \neq j$ .  $i$  es el indicador de que un paciente presente la enfermedad ( $i = 1$ ) o no la presente ( $i = 0$ ).

Por definición,  $C = P\{X_i > X_j \mid i=1, j=0\}$ .

Los supuestos anteriores implican que la medida de concordancia C es una medida ponderada del área bajo la curva ROC para plazos de tiempo fijados. Entonces se puede expresar de la siguiente forma:

$$C = \int_t AUC(t)w(t)dt$$

siendo  $w(t) = 2 \cdot f(t) \cdot S(t)$  la función de ponderación.

Llegando así a verificar la relación entre ambas medidas.

A continuación presentamos los resultados para el índice C en el caso del SMD. El software libre R nos permite calcular este índice usando el paquete `survcomp`, el cual toma como referencias Harrell (2001) y Pencina y D'Agostino (2004).

Para el cálculo del índice C definimos la siguiente ecuación usada en el paquete citado

anteriormente:

$$C = \frac{\sum_{i,j \in \varpi} \mathbb{I}\{x_i > x_j\}}{|\varpi|}$$

donde  $\varpi$  es el conjunto de todos los pares de pacientes  $i, j$  que cumplen una de las siguientes condiciones:

1. ambos pacientes  $i, j$  experimentaron un evento y,  $Z_i < Z_j$  con  $i, j \in [1, n]$ , siendo  $i \neq j$  y  $n$  el tamaño muestral.
2. sólo el paciente  $i$  experimentó un evento y,  $Z_i < c_j$ .

En el caso de la predicción del grupo de riesgo, se debe incluir una condición adicional: las predicciones de riesgo deben ser diferentes para los pacientes  $i, j$ .

Los errores estándar, los intervalos de confianza y los p-valores para el índice C se calculan asumiendo normalidad asintótica (véase Pencina y D'Agostino (2004)), Tabla 4.2.

	$\hat{C}$	<i>se</i>	IC(95 %)	<i>p</i> - valor
IPSS	0.745	0.025	(0.696, 0.794)	9.18e-23
IPSS-R	0.748	0.024	(0.701, 0.795)	4.25e-25

Tabla 4.2: Índice de concordancia, error estándar, intervalo de confianza al 95 % y p-valor para cada sistema pronóstico.

Como se observa en la Tabla 4.2, ambos sistemas pronósticos presentan índices de concordancia similares, con intervalos de confianza al 95 % estadísticamente significativos. El índice C del sistema pronóstico IPSS-R supera levemente al índice C para el sistema tradicional. El siguiente paso es verificar si esta diferencia es significativa. Para la comparación entre ambos índices pertenecientes a los mismos datos de supervivencia se emplea el test de Student para muestras dependientes. Como hipótesis a contrastar se considera que el índice de concordancia para el IPSS-R es superior al índice C para el IPSS. El resultado es un  $p$  - valor = 0,419 > 0,05, por lo tanto, no existen diferencias significativas entre los índices C.

Tras considerar las medidas discriminatorias adoptadas hasta el momento, AUC(t) e índice C, se puede decir que la nueva propuesta, IPSS-R, mejora al sistema pronóstico tradicional, aunque el índice C no presente una mejoría estadísticamente significativa.

### 4.3. Reclasificación

Para apreciar el rendimiento de un modelo de supervivencia es importante evaluar su capacidad predictiva. Actualmente existen nuevas medidas que aportan información adicional a lo visto hasta el momento, Cook (2008) propuso el uso de *tablas de reclasificación* y finalmente, Pencina y D’Agostino y otros (2011), amplian esta idea de reclasificación proponiendo dos nuevas medidas para cuantificar el grado de reclasificación correcta: *mejora neta de reclasificación* (NRI) y *mejoramiento integral de la discriminación* (IDI).

En este caso nos centraremos en la medida de mejora neta de reclasificación.

#### 4.3.1. Mejora neta de reclasificación

El NRI, o mejora neta de reclasificación, ofrece de forma sencilla e intuitiva la cuantificación de la mejora ofrecida por el nuevo marcador. Es decir, cuantifica la cantidad de pacientes que están correctamente reclasificados al usar el nuevo marcador frente al uso del marcador tradicional. Esta medida para cuantificar la mejora neta ha ido ganando popularidad entre los investigadores en los últimos tiempos.

Un movimiento ascendente (*up*) en las categorías para los sujetos que sufren el evento (*event*), implica una mejora de la reclasificación, en cambio, un movimiento descendente (*down*) implica un empeoramiento en la reclasificación. La interpretación es opuesta para los pacientes que no han sufrido el evento (*non-event*). En este caso el evento considerado es la muerte del paciente.

El NRI se define como:

$$NRI = P(up | event) - P(down | event) + P(down | non - event) - P(up | non - event) \quad (4.2)$$

La mejora en la reclasificación es cuantificada como la suma de las dos componentes de la expresión (4.2), una componente para los eventos y otra para los no eventos con ponderación proporcional a la incidencia de eventos.

Usando la regla de Bayes podemos reescribir la fórmula anterior de la siguiente forma:

$$NRI = \frac{P(event | up) \cdot P(up) - P(event | down) \cdot P(down)}{P(event)} + \frac{P(non - event | down) \cdot P(down) - P(non - event | up) \cdot P(up)}{P(non - event)} \quad (4.3)$$

Haciendo el cambio  $P(\text{non-event}) = 1 - P(\text{event})$  y  $P(\text{non-event} | *) = 1 - P(\text{event} | *)$ , donde  $*$  denota “up” ó “down”, se obtiene una nueva definición para la expresión (4.3):

$$NRI = \frac{P(\text{event} | \text{up}) \cdot P(\text{up}) - P(\text{event} | \text{down}) \cdot P(\text{down})}{P(\text{event})} + \frac{1 - P(\text{event} | \text{down}) \cdot P(\text{down}) - (1 - P(\text{event} | \text{up})) \cdot P(\text{up})}{1 - P(\text{event})}$$

Las cantidades en los numeradores representan el número esperados de movimientos ascendentes y descendentes entre los pacientes que experimentan el evento, es decir, se mueren (en el primer numerador) y, el número esperado de movimientos descendentes y ascendentes entre los pacientes que no experimentan el evento (en el segundo numerador). Los denominadores son los casos esperados de eventos y no eventos, respectivamente. Este enfoque supone la existencia de categorías de riesgo y el cálculo de los estimadores propuestos por Kaplan-Meier dentro de cada celda de la tabla de reclasificación.

Simplificando la ecuación anterior podemos reducirla a la siguiente expresión:

$$NRI = \frac{(P(\text{event} | \text{up}) - P(\text{event})) \cdot P(\text{up}) + (P(\text{event}) - P(\text{event} | \text{down})) \cdot P(\text{down})}{P(\text{event}) \cdot (1 - P(\text{event}))}$$

La extensión de estas fórmulas al análisis de supervivencia es inmediata, usando las estimaciones de Kaplan-Meier.

Teniendo en cuenta que  $n$  es el número total de sujetos, en este caso  $n = 340$ , se define  $n_U$  como la cantidad de reclasificados con movimiento ascendente y  $n_D$  los reclasificados con movimiento descendente. De este modo, la anterior fórmula puede expresarse de la siguiente forma:

$$NRI = \frac{P(\text{event} | \text{up}) \cdot n_U - P(\text{event} | \text{down}) \cdot n_D}{n \cdot P(\text{event})} + \frac{(1 - P(\text{event} | \text{down})) \cdot n_D - (1 - P(\text{event} | \text{up})) \cdot n_U}{n \cdot (1 - P(\text{event}))}$$

Esta fórmula, al contrario que las anteriores, no depende del número o incluso de la existencia de categorías de riesgo, ya que se supone que las probabilidades de evento entre los reclasificados hacia arriba o hacia abajo se puede conseguir con una puesta en común de todos los individuos con la misma reclasificación.

Ambos enfoques son equivalentes para muestras grandes pero, cuando se trabaja con muestras pequeñas es preferible este último planteamiento puesto que se eliminan celdas

con un número pequeño de eventos. En nuestro caso tenemos un número de pacientes suficiente, con lo cual, se pueden considerar las categorías de riesgo.

Además se pueden considerar dos subcomponentes muy útiles del NRI total: “*event NRI*” y “*non-event NRI*”:

$$\begin{aligned} eventNRI &= P(up | event) - P(down | event) \\ &= \frac{P(event | up) \cdot P(up) - P(event | down) \cdot P(down)}{P(event)} \end{aligned}$$

$$\begin{aligned} non - eventNRI &= P(down | non - event) - P(up | non - event) \\ &= \frac{P(non - event | down) \cdot P(down) - P(non - event | up) \cdot P(up)}{P(non - event)} \end{aligned}$$

Por supuesto,  $NRI = event\ NRI + non-event\ NRI$ .

Siguiendo la idea de Cook, se clasifica la predicción de riesgo en diferentes categorías para ambos marcadores. En este caso tomamos cuatro categorías de riesgo en cada marcador usado: 0 – 25, 25 – 50, 50 – 75, > 75 %. Tanto para los pacientes que sufren el evento como para los pacientes que no lo sufre.

Al tratarse de un estudio de supervivencia a cierto tiempo, hacemos una tabla de re-clasificación para los tiempos  $t = 1, 3$  y  $5$  años, intentando resumir así el período de seguimiento.

Tendremos, por lo tanto, seis tablas, tres correspondientes a cada tiempo para los eventos y otras tres para los no eventos.

	IPSS-R			
IPSS	[0, 0.25]	(0.25, 0.50]	(0.50, 0.75]	(0.75, 1]
[0, 0.25]	23.00		0.00	
(0.25, 0.50]	0.00	17.43	15.00	
(0.50, 0.75]		0.60	8.00	
(0.75, 1]				

Tabla 4.3: Tabla de reclasificación entre los eventos en el tiempo  $t = 365$  días.

IPSS	IPSS-R			
	[0, 0.25]	(0.25, 0.50]	(0.50, 0.75]	(0.75, 1]
[0, 0.25]	191.00		0.00	
(0.25, 0.50]	0.00	21.56	0.00	
(0.50, 0.75]		0.40	4.00	
(0.75, 1]				

Tabla 4.4: Tabla de reclasificación entre los pacientes que no sufren el evento para  $t = 365$  días.

En las Tablas 4.3 y 4.4 se muestran las reclasificaciones entre los pacientes que sufren el evento y entre los que no lo sufren en el primer año de seguimiento. Se observa que los pacientes se encuentran entre las tres primeras categorías de riesgo, es decir, al principio de seguimiento el riesgo de muerte es más bajo. Entre los que sufren el evento, Tabla 4.3, la mayor reclasificación se corresponde al cambio de categoría (0,25, 0,50] del sistema IPSS a la categoría (0,50, 0,75] del sistema IPSS-R entre los eventos. Es decir, el nuevo sistema clasifica con un riesgo más alto a estos pacientes.

Los resultados para  $t = 365$  días son:

$$eventNRI = 0,1795$$

$$non - eventNRI = 0,0015$$

Por lo tanto:

$$NRI = 0,1795 + 0,0015 = 0,1810$$

Se obtiene una reclasificación del 18% de los pacientes en el primer año de seguimiento al hacer el cambio al nuevo sistema pronóstico IPSS-R.

En las Tablas 4.5 y 4.6 se observan los datos de reclasificación entre eventos y no eventos siendo  $t = 1095$  días, es decir, al tercer año de seguimiento.

IPSS	IPSS-R			
	[0, 0.25]	(0.25, 0.50]	(0.50, 0.75]	(0.75, 1]
[0, 0.25]		0.00	0.00	
(0.25, 0.50]		29.74	1.42	0.00
(0.50, 0.75]		17.00	26.45	4.75
(0.75, 1]			12.75	44.99

Tabla 4.5: Tabla de reclasificación entre los eventos en el tiempo  $t = 1095$  días.

	IPSS-R			
IPSS	[0, 0.25]	(0.25, 0.50]	(0.50, 0.75]	(0.75, 1]
[0, 0.25]		0.00	0.00	
(0.25, 0.50]		47.26	2.57	0.00
(0.50, 0.75]		51.00	17.54	5.25
(0.75, 1]			2.25	7.00

Tabla 4.6: Tabla de reclasificación entre los pacientes que no sufren el evento para  $t = 1095$  días.

En este caso, los pacientes se sitúan, en ambas tablas, en las categorías de riesgo más alto. El valor de mejora neta es:

$$\begin{aligned} eventNRI &= -0,1479 \\ non - eventNRI &= 0,2514 \end{aligned}$$

Por lo tanto:

$$NRI = -0,1479 + 0,2514 = 0,1045$$

En la Tabla 4.5 se aprecia que el movimiento protagonizado por los pacientes es descendente, este hecho indica un empeoramiento de la reclasificación. A pesar de esto, se obtiene una buena reclasificación entre los pacientes que no sufren el evento, con lo cual la mejora neta es del 10 %.

Evaluando el tiempo  $t = 1825$  días, se obtiene la siguiente reclasificación:

	IPSS-R			
IPSS	[0, 0.25]	(0.25, 0.50]	(0.50, 0.75]	(0.75, 1]
[0, 0.25]				
(0.25, 0.50]		22.28	25.45	
(0.50, 0.75]		3.61	65.95	7.26
(0.75, 1]			13.60	52.00

Tabla 4.7: Tabla de reclasificación entre los eventos en el tiempo  $t = 1825$  días.

	IPSS-R			
IPSS	[0, 0.25]	(0.25, 0.50]	(0.50, 0.75]	(0.75, 1]
[0, 0.25]				
(0.25, 0.50]		39.71	55.54	
(0.50, 0.75]		3.38	46.04	2.73
(0.75, 1]			2.40	0.00

Tabla 4.8: Tabla de reclasificación entre los pacientes que no sufren el evento para  $t = 1825$  días.

En este caso, existe más movimiento, tanto entre los pacientes con evento, Tabla 4.7, como entre los pacientes sin evento, Tabla 4.8. Los valores de reclasificación para  $t = 1825$  días son los siguientes:

$$\begin{aligned} eventNRI &= -0,3074 \\ non - eventNRI &= 0,0763 \end{aligned}$$

Por lo tanto:

$$NRI = -0,3074 + 0,0763 = -0,2311$$

En este caso, se tiene un empeoramiento de la reclasificación. Este empeoramiento se puede ver afectado por la falta de observaciones.

Es fácil ver que si cambiamos las categorías de riesgo, por ejemplo, tomando dos categorías en vez de cuatro:  $< 50\%$  y  $> 50\%$ , Tabla 4.9. El resultado para la mejora neta de reclasificación variaría. En este caso el resultado sería:

	t = 365	t = 1095	t = 1825
event NRI	0.0785	0.3601	0.1074
non-event NRI	-0.0257	0.0021	-0.3813
NRI	0.0527	0.3622	-0.2738

Tabla 4.9: Valores para el NRI y sus subcomponentes en los tiempos  $t = 365$ ,  $1095$  y  $1825$  días.

En este caso se aprecia un ligero aumento del valor para el NRI en el tiempo  $t = 1095$  días, en cambio, disminuye en los demás tiempos considerados. Viendo el ejemplo se puede decir que no siempre es cierto que aumentando el número de categorías se consiga un mayor NRI.

Pencina y D'Agostino y otros sugieren un NRI continuo o sin categorías como medida de la mejora en la predicción de riesgo más objetiva y versátil. Su definición sigue siendo coherente con las fórmulas propuestas hasta el momento, encontrándose la única diferencia en la interpretación de los movimientos ascendentes y descendentes.

Los resultados para el NRI continuo se encuentran en la Tabla 4.10:

	t = 365	t = 1095	t = 1825
NRI	0.167	0.192	0.101

Tabla 4.10: Valores para el NRI continuo en los tiempos  $t = 365$ ,  $1095$  y  $1825$  días.

En esta última tabla observamos que la mejor reclasificación se tiene en el período intermedio de seguimiento. Este resultado es coherente con lo obtenido hasta el momento pues,



hemos visto a lo largo del estudio que el nuevo sistema pronóstico discrimina mejor en los años intermedios del período de seguimiento. Sin embargo, a tiempos pequeños, existe solapamiento entre ambos sistemas, impidiendo una buena diferenciación entre los dos y, a tiempos grandes, se tienen menos observaciones lo que provoca una menor precisión en la clasificación.

---

## Capítulo 5

# Nueva propuesta con el modelo de regresión de Cox

El modelo de regresión de Cox (1972) es el más utilizado para datos de supervivencia en el área médica. Es una extensión natural del modelo logístico para datos de supervivencia. Se utiliza cuando la variable dicotómica que se quiere estudiar (aparición del evento), depende del tiempo. Es decir, para el estudio no sólo es importante que los pacientes presenten el evento, sino también en que momento del estudio aparece dicho evento.

En este modelo el riesgo se define mediante:

$$\lambda(t | X) = \lambda_0(t)e^{\beta X} = \lambda_0(t)\exp(\beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p)$$

donde  $X$  es el vector de covariables. En el modelo logístico se usa el predictor lineal de intercepto mientras que, en el modelo de Cox, se usa una función de riesgo basal. Se dice que es un modelo semiparamétrico debido a que incluye una parte paramétrica y otra no paramétrica:

- La parte paramétrica:  $e^{\beta X}$  conocida como puntaje de riesgo (*risk score*).  $\beta$  es el vector de parámetros de la regresión.  $\beta$  es el predictor lineal  $\beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p$ , similar a la Odds Ratio del modelo logístico.
- La parte no paramétrica es  $\lambda_0(t)$ , llamada función de riesgo basal, es una función arbitraria no específica. Se trata de un factor de riesgo común a todos los individuos.

El modelo de regresión de Cox es también conocido como modelo de riesgos proporcionales (*proportional hazard model*) debido a que el cociente entre el riesgo para dos sujetos con el mismo vector de covariables es constante sobre el tiempo, es decir:

$$\frac{\lambda(t | X)}{\lambda(t | X^*)} = \frac{\lambda_0(t)e^{\beta X}}{\lambda_0(t)e^{\beta X^*}} = \frac{e^{\beta X}}{e^{\beta X^*}} = \exp[\beta(X - X^*)]$$

Este es el denominado supuesto de proporcionalidad según el cual la razón de riesgo es constante para todo par de observaciones, en cualquier instante  $t$ . Asimismo, presupone un efecto log-lineal de las covariables sobre la función hazard de forma multiplicativa. El modelo especifica para cada covariable un riesgo relativo constante en el tiempo; si  $X = X^* + 1_j, j \in \{1, \dots, p\}$ , entonces:

$$\frac{\lambda(t | X)}{\lambda(t | X^*)} = \exp[\beta_j]$$

luego,  $\exp[\beta_j]$  se interpreta como el efecto (constante) en el riesgo que tiene un incremento unitario de la  $j$ -ésima covariable (en tiempo  $t$ ).

Puede suceder que las covariables en un modelo de Cox también dependan del tiempo, en este caso, la especificación del modelo sería:

$$\lambda(t | X(t)) = \lambda_0(t)\exp[\beta^t X(t)] = \lambda_0(t)\exp[\beta_1 X_1(t) + \dots + \beta_p X_p(t)]$$

Si ha ocurrido una muerte en el tiempo  $t^*$ , entonces la verosimilitud de que le ocurra el evento de interés al individuo  $i$ -ésimo y no a otro es:

$$L_i(\beta) = \frac{\lambda_0(t^*)r_i(t^*)}{\sum_j Y_j(t^*)\lambda_0(t^*)r_j(t^*)} = \frac{r_i(t^*)}{\sum_j Y_j(t^*)r_j(t^*)}$$

siendo  $Y(t)$  una función indicadora tal que  $Y(t) = 1$  si un individuo está en riesgo en el tiempo  $t$  y, en el otro caso,  $Y(t) = 0$ . Se define el índice de riesgo para un paciente en el tiempo  $t$  como  $r(t) = \exp(\beta X(t))$  (véase Therneau y grambsch (2000)).

El producto  $L(\beta) = \prod L_i(\beta), i \in \{1, \dots, p\}$  se denomina verosimilitud parcial y fue introducida por Cox (1972). La maximización de  $\log(L(\beta))$  da una estimación para  $\beta$  sin necesidad de estimar el parámetro  $\lambda_0(t)$ .

## 5.1. Contrastes de hipótesis

Una vez que se ha ajustado un modelo de Cox, es necesario verificar su significación. Para ello existen tres contrastes de hipótesis, todos ellos asintóticamente equivalentes.

### 5.1.1. Test de razón de verosimilitud

Este test se define como:

$$LR = 2\log(L(\beta_0)) - \log(L(\hat{\beta})) \approx \chi_p^2$$

donde  $\beta_0$  son los valores iniciales de los coeficientes y  $\widehat{\beta}$  es la solución luego de ajustar el modelo.

Este es el test que presentan por defecto los paquetes computacionales.

### 5.1.2. Test de Wald

Este es, quizás, el test más natural debido a que proporciona un contraste para las variables en vez de una medida de significación global.

Sea  $b = (b_1, \dots, b_p)^t$  el vector que tiene como componentes los estimadores máximo verosímiles de  $\beta$ .  $I(\beta)$  es la matriz de información de Fisher evaluada en  $\beta$ .

El vector  $b$  es un vector aleatorio de distribución normal con media  $\beta$  y matriz de varianzas y covarianzas estimada por  $I^{-1}(b)$ .

El estadístico de contraste se define mediante:

$$W = (b - \beta_0)^t I(b) (b - \beta_0) \approx \chi_p^2$$

Si la hipótesis nula ( $H_0 : \beta = \beta_0$ ) se acepta se tiene que  $W \approx \chi_p^2$ .

### 5.1.3. Test de las puntuaciones (*score test*)

Esta prueba se basa en el vector  $U(\beta) = (U_1(\beta), \dots, U_p(\beta))^t$  donde  $U_i(\beta)$  es la derivada parcial de la función de log verosimilitud parcial con respecto a  $\beta_i, i \in \{1, \dots, p\}$ . Para muestras grandes  $U(\beta)$  tiene distribución asintótica normal con media el vector cero y matriz de varianzas y covarianzas  $I(\beta)$  cuando la hipótesis nula es cierta, siendo  $H_0 : \beta = \beta_0$ .

Este test está definido como:

$$S = U(\beta_0)^t I^{-1}(\beta_0) U(\beta_0)$$

Cuando se acepta la hipótesis nula se tiene  $S \approx \chi_p^2$ .

En los tres casos se rechaza la hipótesis nula,  $H_0 : \beta = \beta_0$ , si el estadístico es  $> \chi_{p,\alpha}^2$ .

## 5.2. Análisis de residuos

Los residuos de Schoenfeld (1980) se definen como la matriz:

$$s_i(\beta) = X_i(t_i) - \bar{X}(\beta, t_i)$$

donde  $i$  y  $t_i, i \in \{1, \dots, p\}$  son los individuos y el tiempo de ocurrencia del evento respectivamente,  $X_i$  es el vector de variables para el paciente  $i$ -ésimo y  $\bar{X}(\beta, t_i)$  es la media de las variables para aquellos individuos todavía en riesgo en el tiempo  $t$ .

Siguiendo con la notación de Thernau y Grambsch (2000), definimos  $\bar{X}(\beta, t_i)$  como:

$$\bar{X}(\beta, t_i) = \frac{\sum Y_j(t_i)r_j(t_i)X_j(t_i)}{\sum Y_j(t_i)r_j(t_i)}$$

La matriz  $s_{ij}$  tiene una fila por evento (muerte) y una columna por variable.

Bajo el supuesto de riesgos proporcionales, los residuos de Schoenfeld son útiles en la evaluación de la tendencia en el tiempo o de la falta de proporcionalidad.

## 5.3. Resultados

Un estudio de la GESMD mostró que una sobrecarga de hierro, definida como ferritina sérica superior a 1000 ng/ml, tiene un efecto peyorativo no sólo en la supervivencia global sino también en el riesgo de evolución a Leucemia Mieloide Aguda, tanto en pacientes de bajo como de alto riesgo.

Esta variable continua fue recogida en los pacientes incluidos en la base de datos usada. A su vez, también fue recogida de forma dicotómica la variable *transfusiones previas*. La importancia de esta variable radica en los niveles de ferritina en el momento del diagnóstico, los cuales se incrementan en estos pacientes. Esta variable toma valor 1 para indicar que la transfusión previa es conocida y, toma valor 2 como indicador de la ausencia de transfusión previa en la historia clínica o que no es conocida.

Con intención de averiguar si estas variables pueden aportar información adicional a la clasificación de los pacientes de forma más precisa y eficaz, hemos analizado el modelo de Cox con dichas variables. En capítulos anteriores hemos visto como el marcador propuesto, IPSS-R, mejoraba ligeramente al marcador tradicional. Veamos si el aporte de una nueva variable puede aportar más información.

A continuación se presentan las salidas aportadas por R para el modelo de Cox correspondiente a la incorporación de cada variable.

## CAPÍTULO 5. NUEVA PROPUESTA CON EL MODELO DE REGRESIÓN DE COX

---

En primer lugar se muestran los resultados para la incorporación de la variable continua Ferritina al marcador IPSS-R.

```

                coef    exp(coef)  se(coef)    z    Pr(>|z|)
factor(IPSS-R)2 0.4871477 1.6276670 0.2715737 1.794 0.072846 .
factor(IPSS-R)3 1.0657252 2.9029435 0.2922761 3.646 0.000266 ***
factor(IPSS-R)4 1.8261278 6.2097945 0.3077544 5.934 2.96e-09 ***
factor(IPSS-R)5 2.1464477 8.5544165 0.3274541 6.555 5.57e-11 ***
FERRITINA      0.0003458 1.0003458 0.0001248 2.771 0.005597 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

                exp(coef) exp(-coef) lower .95 upper .95
factor(IPSS-R)2      1.628      0.6144      0.9559      2.772
factor(IPSS-R)3      2.903      0.3445      1.6370      5.148
factor(IPSS-R)4      6.210      0.1610      3.3972     11.351
factor(IPSS-R)5      8.554      0.1169      4.5026     16.252
FERRITINA            1.000      0.9997      1.0001      1.001

```

```

Concordance= 0.698 (se = 0.024 )
Rsquare= 0.216 (max possible= 0.994 )
Likelihood ratio test= 82.53 on 5 df, p=2.22e-16
Wald test              = 93.38 on 5 df, p=0
Score (logrank) test = 112.2 on 5 df, p=0

```

Se observa que la variable Ferritina es significativa, con un p-valor inferior al 0.01. Este modelo resulta significativo para cualquiera de los tres criterios debido a que los p-valores son inferiores al 0.01.

La siguiente salida se corresponde al ajuste mediante el modelo de Cox para la incorporación de la variable Transfusión Previa.

```

                coef    exp(coef)  se(coef)    z    Pr(>|z|)
factor(IPSS-R)2 0.4466      1.5630      0.2724      1.640 0.101097
factor(IPSS-R)3 0.9553      2.5995      0.2967      3.220 0.001284 **
factor(IPSS-R)4 1.7293      5.6365      0.3114      5.553 2.82e-08 ***
factor(IPSS-R)5 2.1498      8.5835      0.3265      6.585 4.55e-11 ***
factor(tp)2     -0.6870      0.5031      0.1866     -3.681 0.000232 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

                exp(coef) exp(-coef) lower .95 upper .95
factor(IPSS-R)2      1.5630      0.6398      0.9164      2.6657

```

## CAPÍTULO 5. NUEVA PROPUESTA CON EL MODELO DE REGRESIÓN DE COX

---

```

factor(IPSS-R)3      2.5995      0.3847      1.4532      4.6500
factor(IPSS-R)4      5.6365      0.1774      3.0614     10.3778
factor(IPSS-R)5      8.5835      0.1165      4.5265     16.2765
factor(tp)2          0.5031      1.9878      0.3489      0.7253

```

```

Concordance= 0.711 (se = 0.023 )
Rsquare= 0.228 (max possible= 0.994 )
Likelihood ratio test= 88.2 on 5 df, p=0
Wald test           = 99.19 on 5 df, p=0
Score (logrank) test = 120 on 5 df, p=0

```

El p-valor para esta variable discreta es mucho menor que en el caso anterior. Es decir, la incorporación de esta variable al modelo es mucho más significativa. En cuanto a los p-valores de los criterios usados, todos ellos son significativos.

Por último, veamos si es más significativa la aportación de las dos variables en conjunto:

```

           coef      exp(coef)   se(coef)    z      Pr(>|z|)
factor(IPSS-R)2  0.4204180  1.5225979  0.2734575  1.537  0.12419
factor(IPSS-R)3  0.9344180  2.5457314  0.2972751  3.143  0.00167 **
factor(IPSS-R)4  1.7202707  5.5860403  0.3109414  5.532  3.16e-08 ***
factor(IPSS-R)5  2.0933694  8.1122023  0.3273060  6.396  1.60e-10 ***
FERRITINA       0.0002196  1.0002196  0.0001359  1.615  0.10625
factor(tp)2     -0.5844961  0.5573866  0.1985788 -2.943  0.00325 **

```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```

           exp(coef)  exp(-coef)  lower .95  upper .95
factor(IPSS-R)2     1.5226     0.6568     0.8909     2.6023
factor(IPSS-R)3     2.5457     0.3928     1.4216     4.5588
factor(IPSS-R)4     5.5860     0.1790     3.0369    10.2749
factor(IPSS-R)5     8.1122     0.1233     4.2711    15.4079
FERRITINA           1.0002     0.9998     1.0000     1.0005
factor(tp)2         0.5574     1.7941     0.3777     0.8226

```

```

Concordance= 0.7 (se = 0.024 )
Rsquare= 0.234 (max possible= 0.994 )
Likelihood ratio test= 90.57 on 6 df, p=0
Wald test           = 100.9 on 6 df, p=0
Score (logrank) test = 122.2 on 6 df, p=0

```

Se observa que al incorporar ambas variables, la Ferritina deja de ser significativa, con lo cual, este modelo no nos aporta más información que los anteriores. Nos centramos entonces en los dos primeros ajustes.

La interpretación de los coeficientes de un modelo de ajuste de Cox deben hacerse a través de la evaluación de la función exponencial evaluada sobre los coeficientes estimados para cada variable. Debe tenerse en cuenta que esta interpretación difiere para variables continuas y discretas.

- Interpretación para la Ferritina:

El exponencial del coeficiente estimado para la Ferritina es 1.000. Así por cada 1000 ng/ml que aumenta el nivel de Ferritina en el paciente, su riesgo de muerte se mantiene constante.

Por ser una variable continua, la interpretación puede hacerse para un rango de Ferritin sérica, utilizando  $exp(c\hat{\beta})$  en vez de  $exp(\hat{\beta})$ , siendo  $c$  el número de unidades de Ferritina sérica contenidas en ese rango.

- Interpretación para la variable Transfusión Previa:

Para el caso de la variable Transfusión Previa, la interpretación es distinta. En este caso tenemos una variable discreta, para la cual se toma como referencia el valor 1, indicador de transfusión previa conocida. Por lo tanto, vamos a comparar el riesgo de muerte si un paciente no realizó transfusión previa frente a uno que si la realizó. El exponencial del coeficiente estimado para los individuos sin transfusión previa es 0.503, con lo cual, la ausencia de transfusiones disminuye el riesgo de muerte hasta un 50 % aproximadamente. Es decir, en vez de tratarse de un factor de riesgo, el hecho de no recibir transfusiones previas se considera un factor protector.

Una condición indispensable para la aceptación del modelo de Cox es la verificación del supuesto de riesgos proporcionales.

El método que usaremos para comprobar si se verifica el supuesto de riesgos proporcionales en cada uno de los ajustes del modelo de Cox será analizando los residuos de Schoenfeld.

Las pruebas para el supuesto de riesgos proporcionales basadas en los residuos de Schoenfeld se obtienen usando la función `cox.zph`, disponible en el paquete `survival` de R. Esta función calcula un test de proporcionalidad para cada variable junto con un test global para el modelo.

En primer lugar se presentan los residuos de Schoenfeld para el modelo con la Ferritina y, en segundo lugar, para la incorporación de la Transfusión Previa.

	rho	chisq	p
factor(IPSS-R)2	-0.041325	2.71e-01	0.602
factor(IPSS-R)3	-0.152890	3.68e+00	0.055
factor(IPSS-R)4	-0.033172	1.76e-01	0.675
factor(IPSS-R)5	-0.094516	1.43e+00	0.231
FERRITINA	0.000286	1.56e-05	0.997
GLOBAL	NA	6.01e+00	0.305



## CAPÍTULO 5. NUEVA PROPUESTA CON EL MODELO DE REGRESIÓN DE COX

	rho	chisq	p
factor(IPSS-R)2	-0.0403	0.255	0.6136
factor(IPSS-R)3	-0.1479	3.499	0.0614
factor(IPSS-R)4	-0.0275	0.121	0.7284
factor(IPSS-R)5	-0.0617	0.590	0.4423
factor(tp)2	0.0311	0.161	0.6884
GLOBAL	NA	6.037	0.3026

Por lo que, no existen evidencias significativas al 5 % de que se viole el supuesto de riesgos proporcionales para ninguna de las variables de cada modelo, ni globalmente.

A continuación, se muestran los resultados del análisis de predicción clínica para ambos modelos.

En primer lugar se presentan las curvas ROC dependientes del tiempo como medida de discriminación entre el modelo basado en el marcador IPSS-R junto a la variable Ferritina y el modelo basado en el marcador IPSS-R junto a la variable Transfusión Previa, Figura 5.1.

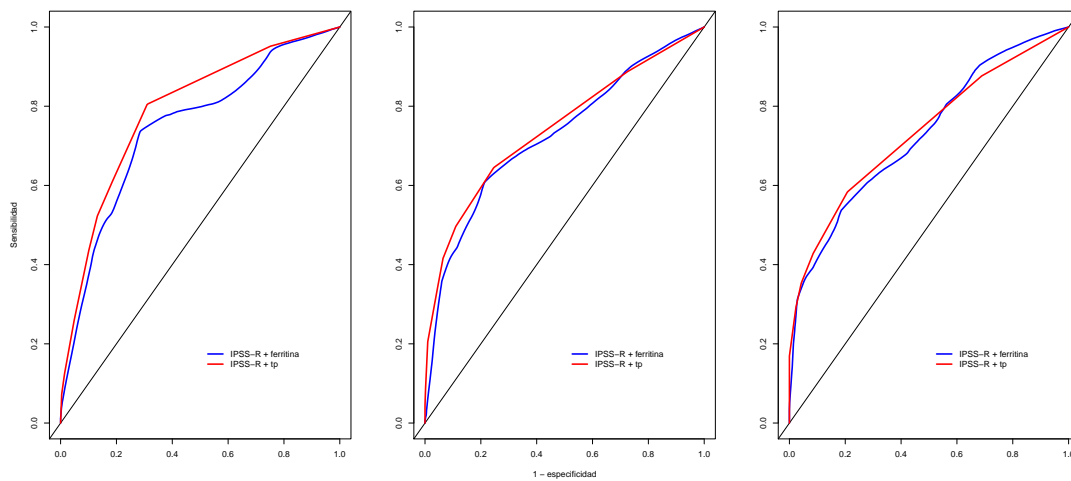


Figura 5.1: Curvas ROC dependientes del tiempo para ambos modelos en los tiempos  $t = 365, 1095$  y  $1825$  días.

En la Tabla 5.1 se muestra el AUC correspondiente a cada modelo para los tiempos  $t = 365, 1095$  y  $1825$  días, respectivamente.

	IPSS-R + Ferritina	IPSS-R + tp
t = 365 días	74.50	78.94
t = 1095 días	72.58	74.30
t = 1825 días	72.75	73.34

Tabla 5.1: Resumen AUC(t) para cada modelo.

En este caso se observa que la variable Transfusión Previa tiene una mayor aportación en la discriminación que la variable Ferritina. Es necesario comprobar que estas diferencias entre los AUCs son estadísticamente significativas, para ello realizamos el test de Wilcoxon para muestras dependientes. Los resultados se pueden apreciar en la Tabla 5.2.

	IPSS-R + Ferritina	IPSS-R + tp	p-valor
t = 365 días	69.68	74.35	1.00e-61
t = 1095 días	71.53	75.09	7.32e-181
t = 1825 días	72.09	74.95	5.61e-300

Tabla 5.2: Resultados del test de Wilcoxon para ambos modelos.

Tras ver los resultados del test de Wilcoxon podemos afirmar que las diferencias entre ambos modelos son significativas, teniendo una mayor capacidad discriminatoria el modelo basado en el marcador IPSS-R junto a la variable Transfusión Previa.

Si comparamos los AUCs dependientes del tiempo entre el marcador IPSS-R y el mismo marcador con la variable  $tp$ , se tiene que la información que aporta esta variable es significativa, Tabla 5.3.

	IPSS-R	IPSS-R + tp	p-valor
t = 365 días	72.51	74.35	1.69e-60
t = 1095 días	70.69	75.09	1.38e-179
t = 1825 días	69.92	74.95	5.61e-300

Tabla 5.3: Resultados del test de Wilcoxon para comparar ambos modelos.

Al contrario que el AUC, que resultó significativamente mayor para la incorporación de una variable al sistema de pronóstico, la puntuación de Brier no presenta diferencias notables entre los modelos propuestos como se refleja en la Figura 5.2.

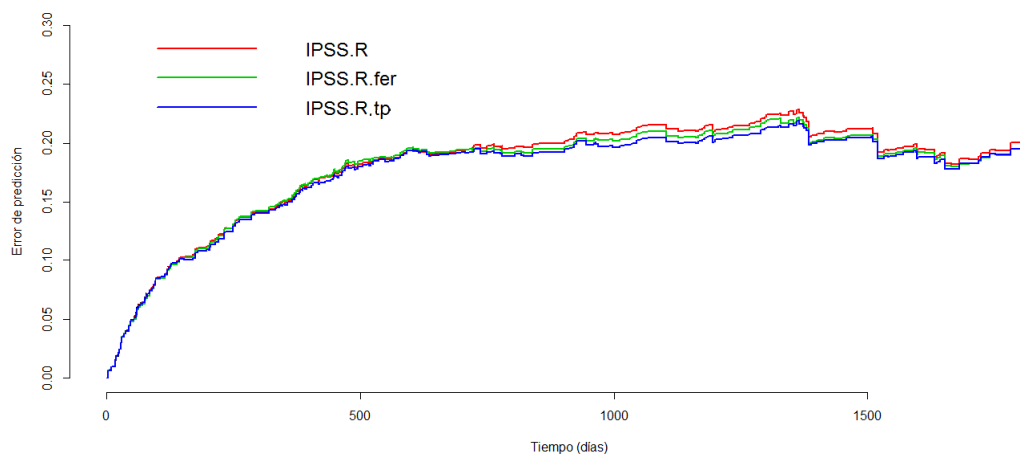


Figura 5.2: Puntuación de Brier para los tres modelos propuestos.

Como última medida de discriminación presentamos el índice de concordancia  $C$ , para el que se aporta su estimación, el error estándar, los intervalos de confianza y los  $p$ -valores, Tabla 5.4.

	$\hat{C}$	se	IC (95 %)	p-valor
IPSS-R + Ferritina	0.698	0.022	(0.655, 0.741)	1.04e-19
IPSS-R + tp	0.750	0.022	(0.708, 0.793)	4.84e-31

Tabla 5.4: Índice de concordancia para ambos modelos.

Presenta un mayor índice de concordancia el modelo con el aporte de la Transfusión Previa. El test de Student para muestras dependientes da  $p - valor = 1,56e - 12 < 0,05$  con lo cual, se verifica que las diferencias entre los índices  $C$  son significativas. Es decir, el aporte de esta nueva variable al marcador IPSS-R es mucho mejor que el aporte que puede añadir la variable Ferritina.

---

## Capítulo 6

# Otras aplicaciones realizadas en la Unidad de Epidemiología

Por último, se presenta un breve resumen de lo que ha sido mi paso por la Unidad de Epidemiología Clínica del Hospital Clínico Universitario de Santiago de Compostela. Hasta el momento se llevó a cabo la descripción de un caso concreto realizado en esta Unidad de forma más individualizada.

Además del tema expuesto, se han llevado a cabo otras tareas como consultas clínicas y la asistencia a sesiones de carácter formativo.

En dichas sesiones fueron expuestos diversos temas, tanto de índole clínica como estadística. Entre los temas tratados podemos destacar:

- La elaboración de proyectos de investigación.
- El ensayo clínico.
- Una introducción a los modelos mixtos.
- La importancia del tamaño muestral.
- Una breve introducción al tratamiento de datos espacio-temporales usando los programas R y BayesX.

Entre las tareas realizadas se encuentran las siguientes consultas a las que se le dio respuesta durante el período de la prácticas:

- Análisis de supervivencia en pacientes con tumor de páncreas. La base de datos constaba de 460 registros de pacientes diagnosticados con dicha enfermedad. El

objetivo del estudio era comprobar la efectividad del uso de un marcador a la hora de predecir el comportamiento del tumor. Para ello se usaron las curvas Kaplan-Meier, disponibles en el paquete `survival` de R.

- Determinación de la utilidad de tres marcadores en el diagnóstico del Hiperaldosterismo Primario (HAP). Para ello se estudiaron 220 individuos diagnosticados de hipertensión arterial.

En primer lugar se representaron las curvas de densidad para los marcadores usados, en relación a la presencia o no de HAP. A continuación, se determinaron las curvas ROC y las áreas bajo la curva ROC, hallando sus correspondientes puntos de corte óptimos mediante el índice de Youden. Para dichos puntos se calcula la sensibilidad, especificidad, valores predictivos, razones de probabilidad y precisión diagnóstica de los diferentes marcadores. Para la realización de estos análisis se ha utilizado el paquete `pROC` disponible en R.

- Evaluación de la incidencia y los predictores para cada transición experimentada por los sujetos a estudio. Se dispone de 408 individuos con Ablación de Flutter (AFI) los cuales pueden experimentar diferentes estados: desarrollar Fibrilación Auricular (AF), un Accidente Cerebrovascular (STROKE), ambas arritmias, o ninguna y finalmente, pueden morir. Se trata de un modelo de multiestado con 4 estados posibles.

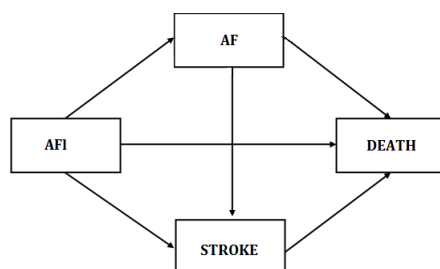


Figura 6.1: Gráfico con los cuatros estados y las posibles transiciones, representadas con flechas.

Se evaluaron las distintas transiciones durante el curso de la enfermedad y se estudiaron los factores de pronóstico para cada transición. Todos los análisis estadísticos se llevaron a cabo en el software R, utilizando los paquetes: `survival`, `splines`, `tdc.msm` y `Epi`.

- Evaluación de la influencia de la glucosa en la evolución de pacientes con trasplante hepático. Se dispone de 630 individuos a los que se le realizó un trasplante hepático. El seguimiento mínimo en estos pacientes es de un año, comenzando desde el momento previo a la cirugía, donde se recoge la medida de la glucosa basal. El análisis se realiza usando los programas estadísticos R y BayesX. El análisis de supervivencia se llevó a cabo usando el paquete `survival` y `splines`.

---

## Referencias

- AKRITAS M. G., 1994, Nearest neighbor estimation of a bivariate distribution under random censoring. *Annals of Statistics* 22, 1299-1327.
- ANDERSON P. K., BORGAN O., GILL R. D., KEIDING N., 1993, Statistical Models Based on Counting Processes. *Ed. Springer*.
- BRIER G. W., 1950, Verification of forecasts expressed in terms of probabilities. *Monthly Weather Review* 78, 1-3.
- COX D. R., 1972, Regression models and life-tables. *Journal of the Royal Statistical Society. Series B* 34, 187-220.
- COOK N. R., 2008, Statistical evaluation of prognostic versus diagnostic models: Beyond the ROC curve. *Clinical Chemistry* 56, 17-23.
- FAWCETT T., 2006, An Introduction to ROC Analysis. *Pattern Recognition Letters* 27, 861-874.
- GRAF E., SCHMOOR C., SAUERBREI W., SCHUMACHER M., 1999, Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine* 18, 2529-2545.
- HARRELL F. E. JR., 2001, Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis. *Ed. Springer*.
- HEAGERTY P. J., LUMLEY T., PEPE M. S., 2000, Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. *Biometrics* 56, 337-344.
- HEAGERTY P. J., ZHENG Y., 2005, Survival Model Predictive Accuracy and ROC Curves. *Biometrics*; 61, 92-105.
- KAINS H., DESMEDT C., SOTIRIOU C., BONTEMPI G., 2008, A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all?. *Bioinformatics* 24, 2200-2208.

## REFERENCIAS

---

KAPLAN E. L., MEIER P., 1958, Non parametric estimation from incomplete observations. *Journal of the American Statistical Association* 53, 457-481.

PENCINA M. J., D'AGOSTINO R. B. SR., 2004, Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in Medicine* 23, 2109-2123.

PENCINA M. J., D'AGOSTINO R. B. SR., D'AGOSTINO R. B. JR., 2008, Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine* 27, 157-172.

PENCINA M. J., D'AGOSTINO R. B. SR., STEYERBERG E. W., 2011, Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in Medicine* 30, 11-21.

SCHOENFELD D., 1980, Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika* 67, 145-153.

STEVENSON M., 2009, An Introduction to Survival Analysis.

STEYERBERG E. W., 2009, Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. *Ed. Springer*.

THERNEAU T. M., GRAMBSCH P. M., 2000, Modeling Survival Data: Extending the Cox Model. *Ed. Springer*.

WILCOXON F., 1945, Individual Comparisons by Ranking Methods. *Biometrics Bulletin* 1, 80-83.