El modelo de regresión de Cox

Eva Boj del Val

Departamento de Matemática Económica, Financiera y Actuarial

Facultad de Economía y Empresa

Universidad de Barcelona

Agosto de 2025

Índice

1. El modelo de Cox de riesgos proporcionales

Formulación del modelo

El modelo de Cox[1] expresa la función de riesgo $h(t)$ en función del tiempo t y de un conjunto de covariables / variables explicativas / predictores / factores de riesgo / variables de confusión, $\mathbf{X}=\left(X_1,\ldots,X_p\right)$, que definen al sujeto en estudio del siguiente modo:

$$h(t,\mathbf{X})=h\left(t,X_1,\ldots,X_p\right)=h_0(t)\exp\left(\sum_{j=1}^{p}\beta_j X_j\right). \tag{1}$$

Como hipótesis de partida supondremos que los tiempos de supervivencia tienen distribuciones continuas, que están tomados de forma exacta y que no existe posibilidad de empates. Para cada sujeto i para $i=1,\ldots,n$ conoceremos su tiempo de muerte/fallo t_i, su estado de fallo o censura d_i, variable codificada con 1 si el dato no está censurado y con 0 si el dato sí lo está, y las covariables fijas $\mathbf{X}_i=\left(X_{i1},\ldots,X_{ip}\right)$. Si incluimos el subíndice i para denotar a un sujeto determinado, el modelo (1) se podría re-escribir como:

$$h\left(t_i,\mathbf{X}_i\right)=h\left(t_i,X_{i1},\ldots,X_{ip}\right)=h_0\left(t_i\right)\exp\left(\sum_{j=1}^{p}\beta_j X_{ij}\right).$$

A la función $h_0(t)$ se la denomina "función de riesgo basal" y corresponde al riesgo de un individuo que tiene como valor en todos los predictores 0, el cual sería el "individuo de referencia" de cara a la interpretación posterior del análisis:

$$h\left(t,X_1=0,\ldots,X_p=0\right)=h_0(t)\exp\left(\sum_{j=1}^{p}\beta_j 0\right)=h_0(t)\exp(0)=h_0(t)1=h_0(t).$$

También se interpreta que la función de riesgo basal sería aquella función "básica" del modelo si éste no incorporara predictores.

La función de riesgo basal, $h_0(t)$, es la única parte de la expresión del modelo de Cox

[1] Cox, D. R. (1972). Regression Models and Life Tables. *Journal of the Royal Statistical Society, Series B*, 34, 187–220.

que depende del tiempo t. La otra parte, $\exp\left(\sum_{j=1}^{p}\beta_j X_j\right)$, sólo depende del vector de covariables $\mathbf{X}=\left(X_1,\ldots,X_p\right)$ de los sujetos, que en este apartado supondremos "independiente del tiempo".

Una variable independiente del tiempo se define como una variable cuyos valores no varían a lo largo del tiempo. Por ejemplo el sexo, la raza o el grupo de tratamiento son variables fijas, sólo toman un valor, el inicial. También podríamos considerar variables como el hecho de ser o no fumador (estado de fumador) como variable independiente del tiempo, ya que aunque el estado de fumador puede variar en el tiempo, se suele suponer que para el estudio no varía, se parte de un estado inicial y se supone que no cambia hasta el final, y por lo tanto que sólo toma un valor por individuo. Otro ejemplo de este tipo podría ser la variable "estado inicial de la enfermedad".

También cabe notar que hay variables cambiantes con seguridad, pero que también se suelen tratar como independientes del tiempo. Por ejemplo la edad y el peso de los sujetos sí varían con el tiempo, pero puede ser apropiado tratarlas como independientes del tiempo en análisis determinados. Esto es posible siempre que los valores de estos predictores no varíen en exceso a lo largo del tiempo, o bien si el efecto de dichas variables en el riesgo de supervivencia depende esencialmente de un único valor de medición.

Existe la posibilidad de considerar predictores dependientes del tiempo a los que denominaremos $\mathbf{X}(t)=\left(X_1(t),\ldots,X_p(t)\right)$. Por ejemplo, el estado corriente de la enfermedad o medidas de tensión arterial sucesivas. En tal caso es posible utilizar la modelización de Cox pero no se suele satisfacer la condición de "riesgos proporcionales" que más abajo definimos. En esta situación en que se tienen en cuenta predictores que dependen del tiempo la regresión se denomina "modelo de Cox ampliado". A éste modelo dedicaremos un apartado.

El modelo de Cox (1) se considera un modelo "semiparamétrico" debido a que incluye una parte paramétrica y otra parte no paramétrica:

a) La parte "paramétrica" se corresponde con $\exp\left(\sum_{j=1}^{p}\beta_j X_j\right)$, es decir, con la exponencial del predictor lineal $\eta = \sum_{j=1}^{p} \beta_j X_j$. En esta parte del modelo se estiman los parámetros $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)$ de la regresión mediante la maximización de la denominada "función de verosimiltud parcial" que estudiaremos con detalle.

b) La parte "no paramétrica" es la función de riesgo basal $h_0(t)$. Ésta es una función arbitraria y no especificada y se estima en un segundo estadio, condicionada a la estimación de los parámetros $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)$ de la regresión. Es por esta componente no paramétrica de la fórmula que el modelo de Cox se considera "semiparamétrico".

Una vez estimada la parte paramétrica, $\exp\left(\sum_{j=1}^{p} \hat{\beta}_j X_j\right)$, y posteriormente la no paramétrica, $\hat{h}_0(t)$, tendremos el modelo semiparamétrico completo:

$$\hat{h}(t, \mathbf{X}) = \hat{h}_0(t) \exp\left(\sum_{j=1}^{p} \hat{\beta}_j X_j\right).$$

Tal y como ya hemos dicho, en el modelo de Cox $h_0(t)$ no está especificada y por lo tanto la distribución del error tampoco. A diferencia, en un modelo paramétrico la forma funcional está completamente especificada, excepto por los parámetros de la distribución supuesta, los cuáles deben ser estimados. Por ejemplo, en el modelo paramétrico Weibull la función de riesgo queda definida como:

$$h(t, \mathbf{X}) = \lambda p t^{p-1},$$

donde $\lambda = \exp\left[\sum_{j=1}^{p} \beta_j X_j\right]$ y $h_0(t) = p t^{p-1}$. En este modelo debemos estimar los parámetros λ, p y $\boldsymbol{\beta}$.

El hecho de que el modelo de Cox sea un modelo "semiparamétrico" hace que sea bien recibido en análisis de supervivencia. Al no tener especificada la función de riesgo basal es posible estimar los coeficientes de la regresión, calcular las razones de riesgo y ajustar las curvas de supervivencia a una gran variedad de situaciones.

Podemos decir que el modelo de Cox es "robusto" en el sentido de que los resultados obtenidos en los ajustes tenderán a aproximarse a los del modelo paramétrico correcto. Por ejemplo, si el modelo paramétrico correcto para el estudio es el Exponencial (o lo mismo para el Weibull o log-logístico) las curvas de supervivencia obtenidas con el modelo de Cox serán similares a las obtenidas con el modelo Exponencial.

Si supiéramos ciertamente cuál es el modelo paramétrico correcto, preferiríamos dicho modelo antes que el modelo de Cox, pero deberíamos estar completamente seguros de que el modelo paramétrico es el apropiado. Aunque podemos utilizar tests de bondad de ajuste, es difícil asegurar que la distribución paramétrica es la correcta. Por lo que el modelo de regresión de Cox resulta una buena alternativa a los modelos paramétricos. Dicho de otro modo, con el modelo de Cox evitamos utilizar un modelo paramétrico incorrecto para el estudio.

Podríamos definir el modelo de regresión de Cox (1) de un modo más general en lo referente a la parte paramétrica:

$$h(t,\mathbf{X}) = h_0(t)\psi(\eta). \tag{2}$$

La parte $\psi(\eta)$ se interpreta como el "riesgo relativo" en el momento t de un individuo con perfil $\mathbf{X} = (X_1,\ldots,X_p)$ respecto de un individuo con $\mathbf{X} = (0,\ldots,0)$. El modelo (1) se correspondería con $\psi(\eta) = \exp\left(\sum_{j=1}^{p}\beta_j X_j\right)$. Esta parametrización del modelo de Cox se denomina forma "log lineal" y es la más popular. También se consideran la forma "lineal" con $\psi(\eta) = 1 + \sum_{j=1}^{p}\beta_j X_j$ y la "logística" con $\psi(\eta) = \log\left(1 + \exp\left(\sum_{j=1}^{p}\beta_j X_j\right)\right)$, a parte de alguna familia dependiente de otros parámetros adicionales que obtiene a las lineales (log lineal y lineal) como casos particulares.

Si analizamos la expresión paramétrica del modelo de Cox definida en (1), observaremos que al suponer que el predictor lineal está relacionado con las curvas de riesgo (o funciones de supervivencia) a través de la exponencial, esto nos asegura que nunca obtendremos valores negativos en la estimación del modelo. Esta es una propiedad deseable que la forma lineal no siempre cumple.

Otra propiedad atractiva del modelo de Cox es que, aunque la función de riesgo basal no esté especificada es posible estimar los parámetros del predictor lineal. Una vez estimados los coeficientes, podremos también estudiar el efecto de las variables explicativas de interés y calcular las denominadas ratios o razones de riesgo que más abajo definimos, todo ello de forma independiente a la estimación de las funciones de riesgo. En el modelo de Cox estimamos en un segundo estadio la función de riesgo $h(t,\mathbf{X})$ (y las correspondientes curvas de supervivencia $S(t,\mathbf{X})$) con muy pocos supuestos ya que la función de riesgo basal, $h_0(t)$ (o de supervivencia basal, $S_0(t)$) son funciones no especificadas.

Un último punto sobre la "popularidad" del modelo de Cox es que es preferido al modelo de regresión logístico cuando la variable objeto de estudio son tiempos de supervivencia que pueden estar o no censurados, pues el modelo de Cox utiliza más información sobre estos datos que el logístico. El logístico únicamente trata con respuestas del tipo 0 y 1 y no tiene en cuenta las censuras.

Comentar que para el modelo de Cox de riesgos proporcionales se procede de forma análoga que con los modelos lineales o lineales generalizados en lo referente al tratamiento y codificación de predictores, excepto que el término "intercept" o medio queda absorbido por la función de riesgo basal. Se pueden tener en cuenta los efectos principales y también interacciones. En general si incluimos interacciones entre variables, también incluiremos sus efectos principales aunque éstos se vuelvan menos significativos con la introducción de la interacción. Si la naturaleza del predictor lo permite se pueden hacer transformaciones de las variables para que entren de un modo más adecuado en el modelo, por ejemplo de forma cuadrática o cúbica, o bien estandarizadas por algún método. En el modelo de Cox, y debido a la interpretación que se realiza, interesará tener a predictores continuos discretizados en clases que sean de interés y susceptibles de interpretación en las denominadas razones de riesgo que en breve definiremos.

Hipótesis de riesgos proporcionales

En el modelo de Cox se busca como primer paso la relación entre los riesgos de muerte de dos individuos expuestos a factores de riesgo diferentes. Para ello, el modelo parte de una hipótesis fundamental, la de que los riesgos son proporcionales.

Para comprender esta noción definiremos previamente la denominada "razón de riesgos" (Hazard Ratio, *HR*) entre dos sujetos con diferente vector de covariables $\mathbf{X}=(X_1,\ldots,X_p)$ y $\mathbf{X}^*=(X_1^*,\ldots,X_p^*)$ como:

$$HR = \frac{h\left(t, \mathbf{X}^*\right)}{h\left(t, \mathbf{X}\right)}. \tag{3}$$

Al igual que se realiza con los *odds ratios*, típicamente se evalúa en el numerador el grupo de mayor riesgo definido por $\mathbf{X}^*$ y en el denominador el grupo de menor riesgo definido por $\mathbf{X}$. En tal caso se espera que el HR sea mayor que 1, ya que $h\left(t, \mathbf{X}^*\right) > h\left(t, \mathbf{X}\right)$ y cuantifica cuántas veces es mayor el riesgo de morir con perfil $\mathbf{X}^*$ que con $\mathbf{X}$. Es más fácil la interpretación si excede del valor base unidad que indica que tienen el mismo riesgo, que si disminuye de la unidad.

Si sustituimos la expresión del modelo (1) en (3) obtenemos lo siguiente:

$$HR = \frac{h_0\left(t\right)\exp\left(\sum_{j=1}^{p} \beta_j X_j^*\right)}{h_0\left(t\right)\exp\left(\sum_{j=1}^{p} \beta_j X_j\right)} = \frac{\exp\left(\sum_{j=1}^{p} \beta_j X_j^*\right)}{\exp\left(\sum_{j=1}^{p} \beta_j X_j\right)} = \exp\left(\sum_{j=1}^{p} \beta_j \left(X_j^* - X_j\right)\right),$$

con lo que

$$HR = \exp\left(\sum_{j=1}^{p} \beta_j \left(X_j^* - X_j\right)\right). \tag{4}$$

Observamos que el resultado de la razón de riesgo (4) no depende de la función de riesgo basal, tan sólo del valor de los predictores y de las betas estimadas, i.e., no depende del tiempo. Por lo tanto, en el modelo de Cox se supone la hipótesis de que los riesgos son proporcionales, ya que se suponen covariables no dependientes del tiempo.

La hipótesis de riesgos proporcionales significa explícitamente que la razón de riesgo (3) es constante en el tiempo: $h\left(t, \mathbf{X}^*\right) = \text{constante} \times h\left(t, \mathbf{X}\right)$. Si lo aplicamos a la expresión resultante en el modelo de Cox (4), denominando θ a la constante y una vez estimados los coeficientes de la regresión por máxima verosimilitud parcial, tenemos que la razón de proporcionalidad es constante en el tiempo e igual a:

$$\hat{\theta} = \exp\left(\sum_{j=1}^{p} \hat{\beta}_j \left(X_j^* - X_j\right)\right). \tag{5}$$

En el caso de dos individuos, i y j que sólo se diferencian en la k-ésima variable, supongamos que X_k vale 0 para i y 1 para j, entonces tenemos que para cualquier tiempo t la denominada razón de riesgo es:

$$HR = \frac{h\left(t, X_1, \ldots, X_{k-1}, 1, X_{k+1}, \ldots, X_p\right)}{h\left(t, X_1, \ldots, X_{k-1}, 0, X_{k+1}, \ldots, X_p\right)}, \tag{6}$$

y vale exactamente $HR = \exp\left(\beta_k\right)$, ya que:

$$\frac{h_0\left(t\right)\exp\left(\beta_1 X_1 + \cdots + \beta_{k-1} X_{k-1} + \beta_k 1 + \beta_{k+1} X_{k+1} + \cdots + \beta_p X_p\right)}{h_0\left(t\right)\exp\left(\beta_1 X_1 + \cdots + \beta_{k-1} X_{k-1} + \beta_k 0 + \beta_{k+1} X_{k+1} + \cdots + \beta_p X_p\right)} =$$

$$\frac{\exp\left(\beta_1 X_1 + \cdots + \beta_{k-1} X_{k-1} + \beta_k 1 + \beta_{k+1} X_{k+1} + \cdots + \beta_p X_p\right)}{\exp\left(\beta_1 X_1 + \cdots + \beta_{k-1} X_{k-1} + \beta_k 0 + \beta_{k+1} X_{k+1} + \cdots + \beta_p X_p\right)} = \exp\left(\beta_k\right).$$

Comentar que en caso de covariables continuas el $HR = \exp\left(\beta_j\right)$ representa la razón de riesgos al incrementar en una unidad la covariable continua X_j. Si nos resulta interesante estimar la razón de riesgos al incrementar la covariable X_j en c unidades, lo haremos mediante $\exp\left(c\beta_j\right)$.

Al utilizar la regresión de Cox para unos datos determinados será necesario verificar que se cumple esta hipótesis de proporcionalidad de riesgos. Para ello se suele comprobar que el efecto de cada variable es constante en el tiempo.

Existen varios métodos para estudiar el cumplimiento de la hipótesis de proporcionalidad. Por un lado puede utilizarse un método gráfico: si una variable, por ejemplo, toma únicamente los valores 0 y 1, pueden representarse las curvas de supervivencia para los dos grupos de sujetos definidos por dicha variable y estudiar si son paralelas. No obstante, existen métodos estadísticos algo más rigurosos como el estudio de los denominados residuos de Schoenfeld. A este tema le dedicaremos un apartado.

Ejemplo / ilustración en que no sería adecuado suponer la hipótesis de riesgos proporcionales:

Este es un ejemplo dónde no se cumple la hipótesis de riesgos proporcionales. Consideremos un estudio en el que los pacientes con cáncer se reparten aleatoriamente a cirugía o a radioterapia sin cirugía. Definimos una variable que toma los valores 0 y 1 que indica si ha habido cirugía o no respectivamente. Supongamos que ésta es la única variable predictora de interés en el modelo de Cox y por lo tanto el HR se calcularía como:

$$HR = \frac{\hat{h}(t, X=1)}{\hat{h}(t, X=0)}.$$

Supongamos que cuando un paciente recibe cirugía para eliminar un tumor canceroso hay, por lo general, un alto riesgo de complicaciones de la propia cirugía o quizás de muerte temprana en el proceso de recuperación. Una vez que el paciente pasa este período crítico, es cuando se pueden observar las ventajas de la cirugía.

Supongamos que dibujamos las curvas de riesgo para los dos grupos y que observamos que se cruzan aproximadamente en el día 3, que antes del día 3 el riesgo para el grupo de cirugía es más alto que el riesgo para el grupo sin cirugía, mientras que después de 3 días, el riesgo para el grupo de cirugía es inferior que el riesgo para el grupo sin cirugía.

Escrito en términos de HR tendríamos lo siguiente:

$$t = 2: \qquad HR = \frac{\hat{h}(2, X=1)}{\hat{h}(2, X=0)} < 1$$

$$t = 3: \qquad HR = \frac{\hat{h}(3, X=1)}{\hat{h}(3, X=0)} = 1$$

$$t = 5: \qquad HR = \frac{\hat{h}(5, X=1)}{\hat{h}(5, X=0)} > 1$$

Con lo que el HR que obtenemos no es constante a lo largo del tiempo t.

Cabe notar que en estudios que comparan cirugía ante no cirugía, podríamos esperar ver funciones de riesgo para cada grupo como las que se explican en este ejemplo.

Encontraremos alguna solución a este problema cuando estudiemos el modelo de Cox estratificado y el modelo de Cox ampliado, el cual permite la utilización de predictores dependientes del tiempo.

Algunas soluciones a este problema podrían ser:

- Estratificar en dos grupos sin ajustar ningún modelo y obtener las curvas Kaplan-Meier (KM) por separado para cada grupo;

- Comenzar el análisis en el día 3, aplicando el modelo de Cox únicamente a los supervivientes después de los tres días;
- Estimar dos modelos de Cox, uno para los datos referentes al tiempo menor que tres, y otro para los datos a partir del tiempo tres;
- Estimar el modelo de Cox ampliado incluyendo una variable dependiente del tiempo para medir la interacción de los grupos de cirugía y el tiempo.

Como en cualquier estudio estadístico las diferentes opciones pueden llevar a diferentes conclusiones. De modo que el investigador deberá sopesar las ventajas / inconvenientes de cada opción.

Regla general: Si los riesgos del numerador y denominador del HR se cruzan, la hipótesis de riesgos proporcionales no se cumple y por lo tanto el modelo (1) de riesgos proporcionales no es adecuado.

Función de verosimilitud parcial. Estimación de los coeficientes.

En el modelo de regresión de Cox los parámetros $\boldsymbol{\beta} = \left(\beta_1, \ldots, \beta_p\right)$ se estiman maximizando el logaritmo de la denominada "función de verosimilitud parcial". La maximización de dicha función se realiza mediante métodos numéricos, obteniendo de esta forma la estimación $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)$. Con la estimación de estos parámetros ya tendremos la componente paramétrica totalmente especificada en el modelo:

$$h\left(t, \mathbf{X}\right) = h_0\left(t\right) \exp\left(\sum_{j=1}^{p} \hat{\beta}_j X_j\right),$$

y consecuentemente podremos hacer inferencia sobre dicho vector de parámetros y calcular los HR de interés para el estudio.

La función de verosimilitud parcial que a continuación vamos a definir, se denomina parcial debido a que tiene en cuenta únicamente en la función de verosimilitud las probabilidades de los tiempos de muerte/fallo y no incluye las probabilidades de los tiempos de datos censurados. Sin embargo, en el cálculo de las probabilidades de los tiempos de muerte sí tiene en cuenta a todos los sujetos (censurados o no *a posteriori*) objeto de riesgo al inicio de los diferentes tiempos de muerte.

Denominamos $L \equiv L\left(\beta_1,\ldots,\beta_p\right)$ a la función de verosimilitud parcial. Supongamos que tenemos k tiempos de muerte y que no hay empates. Así, tendremos $n-k$ tiempos censurados. Los tiempos de muerte ordenados los denotamos por $t_{(1)},\ldots,t_{(k)}$, y denotamos por $R\left(t_{(i)}\right)$ para $i=1,\ldots,k$ al conjunto de los sujetos a riesgo en el tiempo $t_{(i)}$. Denominamos por $L_i \equiv L_{t_{(i)}}\left(\beta_1,\ldots,\beta_p\right)$ para $i=1,\ldots,k$ a las porciones de la verosimilitud total anterior debidas a la aportación de los diferentes tiempos de muerte $t_{(1)},\ldots,t_{(k)}$.

Construiremos la función de verosimilitud total como el producto de cada una de las aportaciones de los k tiempos de muerte:

$$L=\prod_{i=1}^{k} L_i .$$

Una vez tenemos la verosimilitud total construida hacemos el logaritmo y derivamos respecto de los parámetros:

$$\frac{\partial \log L}{\partial \beta_j}, \tag{7}$$

$$\frac{\partial^2 \log L}{\partial \beta_i \partial \beta_j}. \tag{8}$$

De (7) igualando a 0, $\frac{\partial \log L}{\partial \beta_j}=0$ para $j=1,\ldots,p$ obtendremos las ecuaciones que nos permitirán obtener las estimaciones de $\hat{\boldsymbol{\beta}}=\left(\hat{\beta}_1,\ldots,\hat{\beta}_p\right)$ mediante la utilización de algún método numérico.

De (8) comprobamos que realmente es un máximo y podemos obtener, como ocurre cuando se trabaja en general con una función de verosimilitud, la “matriz de información (observada)”, $\mathbf{I}(\boldsymbol{\beta})$, donde cada elemento se iguala a:

$$I_{ij}(\boldsymbol{\beta})=-\frac{\partial^2 \log L}{\partial \beta_i \partial \beta_j}.$$

Así, la matriz de varianzas y covarianzas $p\times p$ estimada es $\hat{\boldsymbol{\Sigma}}=\mathbf{I}^{-1}\left(\hat{\boldsymbol{\beta}}\right)$. Cabe notar que este estimador, obtenido a partir de la maximización de la función de verosimilitud parcial es asintóticamente no sesgado, eficiente y normal. Por un lado, el estimador $\hat{\boldsymbol{\beta}}$ estima consistentemente el vector de parámetros $\boldsymbol{\beta}$, pero no es completamente eficiente, es decir no alcanza la cota de Cramer-Rao. Finalmente, la distribución de $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)$ es aproximadamente normal de media $\left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)$ y matriz de varianzas y covarianzas $\hat{\boldsymbol{\Sigma}}$.

Veamos cuánto vale exactamente cada una de las $L_i \equiv L_{t_{(i)}}\left(\beta_1, \ldots, \beta_p\right)$ para $i = 1, \ldots, k$:

$$L_{t_{(i)}}\left(\beta_1, \ldots, \beta_p\right) = \frac{h\left(t_{(i)}, \mathbf{X}_{(i)}\right)}{\sum_{l \in R\left(t_{(i)}\right)} h\left(t_{(i)}, \mathbf{X}_l\right)} = \frac{h_0\left(t_{(i)}\right) \exp\left(\sum_{j=1}^{p} \beta_j X_{(i)j}\right)}{\sum_{l \in R\left(t_{(i)}\right)} h_0\left(t_{(i)}\right) \exp\left(\sum_{j=1}^{p} \beta_j X_{lj}\right)},$$

siendo $\mathbf{X}_{(i)}$ el vector de covariables para el sujeto con tiempo de muerte $t_{(i)}$ y $\mathbf{X}_l$ para $l \in R\left(t_{(i)}\right)$ el vector de covariables de cada uno de los sujetos de $R\left(t_{(i)}\right)$.

Como podemos ver la función de riesgo basal se anula en el numerador y en el denominador, con lo que nos queda la expresión:

$$L_{t_{(i)}}\left(\beta_1, \ldots, \beta_p\right) = \frac{\exp\left(\sum_{j=1}^{p} \beta_j X_{(i)j}\right)}{\sum_{l \in R\left(t_{(i)}\right)} \exp\left(\sum_{j=1}^{p} \beta_j X_{lj}\right)}. \tag{9}$$

Observamos que la función de verosimilitud parcial total así calculada no depende de las cuantías de los tiempos, tan sólo de su ordenación y de si el dato estaba o no censurado. Como consecuencia podríamos obtener las mismas estimaciones de $\hat{\boldsymbol{\beta}}$ para distintos datos, siempre que éstos tengan el mismo patrón de orden y censura en los tiempos de supervivencia.

<u>Función de verosimilitud en caso de empates</u>

En este apartado comentamos tres opciones para tener en cuenta los empates en los tiempos de supervivencia. Serán los tres métodos que podremos escoger en la función `coxph` del paquete survival[2] de `R`:

- Método de Breslow (ties "breslow")
- Método de Efron (ties "efron")
- Método "Exact partial likelihood" (ties "exact")

Por defecto en la función `coxph` la opción es "efron". Según el manual esta es la opción más eficiente computacionalmente para tiempos continuos. El método "exact" es equivalente a un modelo de regresión logística condicional y será apropiado cuando los tiempos sean un conjunto discreto y pequeño. En éste último, si los datos tienen tamaño "grande" el tiempo computacional puede ser excesivo. Por otro lado, en caso de tener un número elevado de empates, la aproximación discreta es, como veremos a continuación, complicada computacionalmente.

Estudiaremos las fórmulas de cada método teniendo en cuenta las siguientes referencias:

Lee, E. T. and Wang, J. W. (2003). *Satistical Methods for suvirval data analysis*. Wiley Series. (pp. 298-305).

Therneau, T. M. and Grambsch, P. M. (2000). *Modelling Survival Data: Extending the Cox model*. Springer-Verlag.

Cabe notar que, según el autor del paquete survival y siguendo los comentarios de su libro Therneau and Grambsch (2000) p. 53, cuando el número de empates es elevado la propuesta es utilizar la opción "efron", la cual empíricamente es la que suele dar estimaciones de los coeficientes "similares" a los de la aproximación discreta del método "exact". Es por ello que la opción por defecto de la función `coxph` es "efron", a diferencia de otros *softwares* que también estiman el modelo de Cox, en los que la opción por defecto suele ser "breslow" como por ejemplo SAS.

Para poder explicitar las fórmulas, vamos a añadir alguna nomenclatura adicional. Comenzamos por escribir de forma matricial la expresión (9) suponiendo vectores fila para los parámetros y para las covariables:

---

[2] Therneau, T. M. (2024). *A Package for Survival Analysis in R*. R package version 3.7-0, https://CRAN.R-project.org/package=survival.

$$L_{t_{(i)}}(\boldsymbol{\beta}) = \frac{\exp\left(\boldsymbol{\beta}\mathbf{X}'_{(i)}\right)}{\sum_{l \in R\left(t_{(i)}\right)} \exp\left(\boldsymbol{\beta}\mathbf{X}'_{l}\right)}. \tag{10}$$

Con lo que la función de verosimilitud parcial total será:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{k} L_{t_{(i)}}(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\exp\left(\boldsymbol{\beta}\mathbf{X}'_{(i)}\right)}{\sum_{l \in R\left(t_{(i)}\right)} \exp\left(\boldsymbol{\beta}\mathbf{X}'_{l}\right)} \tag{11}$$

Denotamos por $m_{(i)}$ a la multiplicidad de $t_{(i)}$ para $i = 1,\ldots,k$, es decir, el número de eventos que coinciden en su fallo en $t_{(i)}$. Tendremos que $m_{(i)} > 1$ si hay más de una observación con tiempo de muerte $t_{(i)}$ y $m_{(i)} = 1$ si sólo hay un sujeto con tiempo de muerte $t_{(i)}$. Sea $r_i = card\left[R\left(t_{(i)}\right)\right]$, i.e., el número de sujetos en riesgo en el tiempo $t_{(i)}$. Por ejemplo, para $\{15, 16+, 20, 20, 20, 21, 24, 24\}$ tenemos $n = 8$, $k = 4$, $t_{(1)} = 15$, $t_{(2)} = 20$, $t_{(3)} = 21$, $t_{(4)} = 24$, $m_{(1)} = 1$, $m_{(2)} = 3$, $m_{(3)} = 1$, $m_{(4)} = 2$, $r_1 = 8$, $r_2 = 6$, $r_3 = 3$ y $r_4 = 2$.

Supongamos que para cada $R\left(t_{(i)}\right)$ podemos seleccionar aleatoriamente $m_{(i)}$ sujetos. Denotamos cada una de estas selecciones por $\mathbf{u}_{(j)}$. Hay ${}_{r_i}C_{m_{(i)}} = r_i! \big/ \left[m_{(i)}!\left(r_i - m_{(i)}\right)!\right]$ posibles $\mathbf{u}_{(j)}$'s. Sea $\mathbf{U}_i$ el conjunto de todas ellas. Por ejemplo, para $R\left(t_{(2)}\right)$ podemos seleccionar aleatoriamente combinaciones de $m_{(2)} = 3$ de entre los $r_2 = 6$ sujetos $\{20, 20, 20, 21, 24, 24\}$. Hay un total de ${}_6C_3 = 20$ posibles selecciones o subconjuntos, y por lo tanto tendremos $\mathbf{U}_2 = \left\{\mathbf{u}_{(1)},\ldots,\mathbf{u}_{(20)}\right\}$. Ahora nos centraremos en los sujetos empatados. Sea $\mathbf{X}_k = \left(X_{k1},\ldots,X_{kp}\right)$ el vector de covariables del *k*-ésimo sujeto, $\mathbf{Z}_{\mathbf{u}_{(j)}} = \sum_{k \in \mathbf{u}_{(j)}} \mathbf{X}_k = \left(Z_{\mathbf{u}_{(j)}1},\ldots,Z_{\mathbf{u}_{(j)}p}\right)$, donde $Z_{\mathbf{u}_{(j)}l}$ es la suma de la *l*-ésima covariable de los $m_{(i)}$ sujetos que están en $\mathbf{u}_{(j)}$. Sea $\mathbf{u}^*_{(i)}$ el conjunto de $m_{(i)}$ sujetos que muere/falla en $t_{(i)}$ y $\mathbf{Z}_{\mathbf{u}^*_{(i)}} = \sum_{k \in \mathbf{u}^*_{(i)}} \mathbf{X}_k = \left(Z_{\mathbf{u}^*_{(i)}1},\ldots,Z_{\mathbf{u}^*_{(i)}p}\right)$, donde $Z_{\mathbf{u}^*_{(i)}l}$ es la suma de la *l*-ésima covariable de los $m_{(i)}$ sujetos que están en $\mathbf{u}^*_{(i)}$ (que fallan en $t_{(i)}$). Por ejemplo, $Z_{\mathbf{u}^*_{(2)}1}$ es igual a la suma de la primera componente de las covariables de los tres sujetos que tienen el tiempo de muerte en 20.

Tiempo continuo

Podemos utilizar las aproximaciones "breslow" y "efron" cuando $m_{(i)}$ es pequeño en comparación con r_i:

- Breslow [3]

$$L_B(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\exp\left(\boldsymbol{\beta}\mathbf{Z}'_{\mathbf{u}^*_{(i)}}\right)}{\left[\sum_{l \in R\left(t_{(i)}\right)} \exp\left(\boldsymbol{\beta}\mathbf{X}'_l\right)\right]^{m_{(i)}}}, \tag{12}$$

- Efron [4]

$$L_E(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\exp\left(\boldsymbol{\beta}\mathbf{Z}'_{\mathbf{u}^*_{(i)}}\right)}{\prod_{j=1}^{m_{(i)}}\left[\sum_{l \in R\left(t_{(i)}\right)} \exp\left(\boldsymbol{\beta}\mathbf{X}'_l\right) - \left[(j-1)/m_{(i)}\right]\sum_{l \in \mathbf{u}^*_{(i)}} \exp\left(\boldsymbol{\beta}\mathbf{X}'_l\right)\right]}. \tag{13}$$

Tiempo discreto

Cuando los tiempos se observan en tiempo discreto los empates son empates "verdaderos", es decir, los fallos pasan realmente al mismo tiempo. Cox[5] propone en este caso el siguiente modelo logístico:

$$\frac{h(t,\mathbf{X})dt}{1-h(t,\mathbf{X})dt} = \frac{h_0(t)dt}{1-h_0(t)dt}\exp(\boldsymbol{\beta}\mathbf{X}').$$

Este modelo se reduce al modelo (1) en escala continua para el tiempo. Con este modelo sustituyendo en (10) por el siguiente término con empates:

[3] Breslow, N. (1974). Covariance Analysis of Survival Data under the Proportional Hazards Model. *International Statistical Review*, 43, 43–54.

[4] Efron, B. (1977). The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association*, 72, 557–565.

[5] Cox, D. R. (1972). Regression Models and Life Tables. *Journal of the Royal Statistical Society, Series B*, 34, 187–220.

$$\frac{\exp\left(\boldsymbol{\beta}\mathbf{Z}'_{\mathbf{u}^*_{(i)}}\right)}{\sum_{\mathbf{u}_{(j)}\in\mathbf{U}_i}\exp\left(\boldsymbol{\beta}\mathbf{Z}'_{\mathbf{u}_{(j)}}\right)},$$

resulta la siguiente función de verosimilitud parcial para el caso de empates en tiempo discreto:

$$L_D(\boldsymbol{\beta})=\prod_{i=1}^{k}\frac{\exp\left(\boldsymbol{\beta}\mathbf{Z}'_{\mathbf{u}^*_{(i)}}\right)}{\sum_{\mathbf{u}_{(j)}\in\mathbf{U}_i}\exp\left(\boldsymbol{\beta}\mathbf{Z}'_{\mathbf{u}_{(j)}}\right)}. \tag{14}$$

En esta expresión, el término *i*-ésimo representa la probabilidad condicional de observar los $m_{(i)}$ fallos dado que hay $m_{(i)}$ fallos en el tiempo $t_{(i)}$ y el conjunto a riesgo $R\left(t_{(i)}\right)$. El número de términos del denominador es ${}_{r_i}C_{m_{(i)}}=r_i!\Big/\left[m_{(i)}!\left(r_i-m_{(i)}\right)!\right]$ y será grande si $m_{(i)}$ es grande. Para su cálculo se utiliza un algoritmo recurrente que hace que el tiempo computacional se rebaje y sea manejable. La ecuación (14) se puede también considerar una aproximación de la función de verosimilitud parcial para tiempos de supervivencia continuos con empates haciendo el supuesto de que los empates son verdaderos como si se hubieran observado en tiempo de escala discreta.

En la mayoría de casos prácticos las tres aproximaciones, (12), (13) y (14) son aproximaciones razonables de la función de verosimilitud exacta para tiempos de supervivencia en tiempo continuo con empates. Cuando no hay empates, las tres se reducen a (11).

Contrastes de hipótesis

Como ya hemos comentado, a partir de la función de verosimilitud parcial obtenemos una estimación de los coeficientes $\hat{\boldsymbol{\beta}}=\left(\hat{\beta}_1,\ldots,\hat{\beta}_p\right)$ cuya distribución es aproximadamente normal de media $\left(\hat{\beta}_1,\ldots,\hat{\beta}_p\right)$ y matriz de varianzas y covarianzas $\hat{\boldsymbol{\Sigma}}$. Lo que nos permite utilizar tests análogos a los utilizados en un modelo lineal o lineal generalizado.

Para resolver la hipótesis $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$ podemos utilizar el **estadístico de Wald** $z = \dfrac{\hat{\beta}_j}{\sqrt{V\hat{a}r\left(\hat{\beta}_j\right)}}$.

La fórmula para calcular un intervalo de confianza aproximado de nivel $(1-\alpha)$ para el coeficiente β_j es la siguiente:

$$\hat{\beta}_j \pm z_{1-\alpha/2}\sqrt{V\hat{a}r\left(\hat{\beta}_j\right)}.$$

Si lo que queremos es hacer el test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ se suelen utilizar tres contrastes:

1) **El contraste de Wald**. Este contraste se basa en que los coeficientes $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)$ siguen una distribución aproximadamente normal de media $\left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)$ y matriz de varianzas y covarianzas $\hat{\boldsymbol{\Sigma}} = \mathbf{I}^{-1}\left(\hat{\boldsymbol{\beta}}\right)$. El estadístico se define como:

$$\chi_W = \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)^T \mathbf{I}\left(\hat{\boldsymbol{\beta}}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right),$$

que bajo la hipótesis nula sigue una distribución χ^2 con p grados de libertad.

2) **El contraste de la razón de verosimilitud**. En este contraste se utiliza el valor de la función de verosimilitud parcial evaluada en $\hat{\boldsymbol{\beta}}$, $L\left(\hat{\boldsymbol{\beta}}\right)$, y evaluada en $\boldsymbol{\beta}_0$, $L\left(\boldsymbol{\beta}_0\right)$:

$$\chi_{LR} = 2\left(\log L\left(\hat{\boldsymbol{\beta}}\right) - \log L\left(\boldsymbol{\beta}_0\right)\right),$$

que bajo la hipótesis nula sigue una distribución χ^2 con p grados de libertad.

3) **El contraste del "score" (Log Rank)**. En este contraste se utiliza el gradiente (derivadas) del logaritmo de la verosimilitud parcial evaluada en la hipótesis nula y supone que bajo la hipótesis nula el vector scores:

$$\chi_S = \left(\frac{\partial L(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \right)^T \left(-\frac{\partial L^2(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right)^{-1} \frac{\partial L(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}},$$

es aproximadamente (multi)normal de medias 0 y matriz de varianzas y covarianzas $\mathbf{I}\left(\hat{\boldsymbol{\beta}}\right)$, con lo que el estadístico sigue también una distribución χ^2 con p grados de libertad.

Comentarios:

El contraste de Wald tiene una interpretación más directa que el contraste de verosimilitud y el del score, sin embargo no es invariante ante diferentes parametrizaciones y los otros dos sí. Con el contraste del score sólo hace falta maximizar bajo la hipótesis nula, con lo que si hay que realizar el test para varios parámetros es más rápido computacionalmente. Sin embargo, el test de la máxima verosimilitud converge más rápido hacia la distribución normal.

Ante la duda de cuál utilizar, es recomendable decantarse por el test de la máxima verosimilitud.

Test de verosimilitud en la comparación de modelos anidados

Supongamos que tenemos dos modelos de Cox anidados, el menor con p predictores con vector de parámetros $\left(\beta_1,\ldots,\beta_p\right)$, y el mayor que incluye al anterior con $p+q$ predictores con vector de parámetros $\left(\beta_1,\ldots,\beta_p,\beta_{p+1},\ldots,\beta_{p+q}\right)$. Supongamos que queremos hacer el siguiente contraste $H_0:\beta_{p+1}=\beta_{p+2}=\cdots=\beta_{p+q}=0$ frente a la alternativa de que alguno de los coeficientes $\beta_{p+1},\ldots,\beta_{p+q}$ no sea 0. Podemos trabajar con el **estadístico de deviance** (similar a los modelos lineales generalizados) definido como:

$$D=-2\log L\left(\hat{\boldsymbol{\beta}}\right). \tag{15}$$

Calculamos la deviance para los dos modelos a comparar: el menor $D_1=-2\log L_1\left(\hat{\boldsymbol{\beta}}\right)$ y el mayor $D_2=-2\log L_2\left(\hat{\boldsymbol{\beta}}\right)$. Con lo que podemos realizar el test de hipótesis mediante:

$$D_1 - D_2 = -2\log L_1\left(\hat{\boldsymbol{\beta}}\right) - \left(-2\log L_2\left(\hat{\boldsymbol{\beta}}\right)\right) = -2\log\frac{L_1\left(\hat{\boldsymbol{\beta}}\right)}{L_2\left(\hat{\boldsymbol{\beta}}\right)},$$

que bajo la hipótesis nula sigue una distribución χ^2 con q grados de libertad.

De este modo podemos realizar lo que sería un análisis "anova" similar al que se realiza para los modelos lineales o lineales generalizados.

También puede ser útil el **Criterio de Akaike** del mismo modo que en los modelos lineales o lineales generalizados para decidir cuántas variables incluir en el modelo. El criterio se puede definir como: $AIC\left(\hat{\boldsymbol{\beta}}\right) = -2\log L\left(\hat{\boldsymbol{\beta}}\right) + kp$ y se busca minimizar con la inclusión de parámetros. Una práctica extendida es escoger $k = 2$.

Comentario sobre discretización de predictores continuos:

Como ya hemos comentado en el modelo de Cox es usual discretizar predictores de naturaleza continua en clases que representen los diferentes riesgos a analizar mediante las denominadas razones de riesgo (o HR). Si deseamos hacer, por ejemplo, dos clases una será la de mayor riesgo y la otra la de menor. A la variable que resuma la discretización la codificaremos con un 1 para la clase de mayor riesgo y con un 0 para la de menor. Puede pasar que la discretización venga dada "ad-hoc" por un "experto" según el significado de la covariable. En otro caso hay que decidir cuál será el punto de corte que dividirá a las clases de riesgo de la covariable continua. Se puede proceder de diferentes formas. Si se tienen en cuenta características de la covariable se puede por ejemplo dividir en función de valores de la variable mayores o menores a la media o la mediana, o decidir clases con el mismo número de observaciones, etc. Sin embargo, parece adecuado en el caso de no disponer de experiencia previa utilizar algún estadístico que tenga en cuenta también la variable explicativa y el entorno. Una solución en esta dirección se puede encontrar en la siguiente referencia: Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis techniques for censored and truncated data*. Springer. En el apartado "8.6 Discretizing a Continuous Covariate" que comienza en la página 272. En dicha referencia se propone elegir el punto de corte que maximize el estadístico del Score (Log Rank). El punto de corte así obtenido será el que rechace más la hipótesis nula de que el coeficiente de la variable sea cero en un modelo con una única variable, la analizada. Con lo que nos quedaremos con el punto de corte que nos dé un mayor valor del estadístico. En la referencia está descrito el procedimiento y cómo tener en cuenta *a posteriori* en el test de Wald el hecho de haber discretizado después de realizar un número determinado de contrastes de hipótesis.

Intervalos de confianza para los *HR*

Un intervalo de confianza para un HR de un efecto principal, $\exp\left(\hat{\beta}\right)$, sin tener en cuenta interacciones de ningún orden, se calcula como:

$$\exp\left(\hat{\beta} \pm z_{1-\alpha/2}\sqrt{V\hat{a}r\left(\hat{\beta}\right)}\right) = \exp\left(\hat{\beta} \pm z_{1-\alpha/2} es\left(\hat{\beta}\right)\right). \qquad (16)$$

Si queremos por ejemplo que el intervalo sea de un nivel de confianza del 95% tendremos $z_{1-\alpha/2} = z_{1-0.05/2} = 1.96$:

$$\exp\left(\hat{\beta} \pm 1.96\sqrt{V\hat{a}r\left(\hat{\beta}\right)}\right) = \exp\left(\hat{\beta} \pm 1.96 es\left(\hat{\beta}\right)\right).$$

En el caso de covariables continuas si analizamos la razón de riesgos al incrementar la covariable X_i en c unidades resulta $HR = \exp\left(c\beta_i\right)$, siendo su intervalo de confianza del 95%:

$$\exp\left(c\hat{\beta}_i \pm 1.96|c| es\left(\hat{\beta}_i\right)\right).$$

Sin embargo, cuando intervienen términos de interacción en el modelo es más complicado el análisis de los intervalos de confianza y su interpretación. En tal caso hay que tener en cuenta la matriz de varianzas y covarianzas de los coeficientes, además de fijar valores específicos de algunas de las variables.

Pongamos una situación de interacción en la que explicitaremos la fórmula correspondiente a la razón de riesgo y al correspondiente intervalo de confianza. Supongamos X_1 una variable de exposición a un grupo que toma valores 0 y 1, y el término de interacción $X_1 \times W_1, \ldots, X_1 \times W_k$, con coeficiente β_1 para X_1 y $\delta_1, \ldots, \delta_k$ para $X_1 \times W_1, \ldots, X_1 \times W_k$. Estamos interesados en calcular un intervalo de confianza del 95% para $\exp\left(\hat{l}\right)$, donde $\hat{l} = \hat{\beta}_1 + \hat{\delta}_1 \times W_1 + \hat{\delta}_2 \times W_2 + \cdots + \hat{\delta}_k \times W_k$. El intervalo de confianza lo calcularemos como:

$$\exp\left(\hat{l} \pm 1.96\sqrt{Var\left(\hat{l}\right)}\right),$$

Y en este caso la fórmula de la varianza viene dada por:

$$Var\left(\hat{l}\right) = Var\left(\hat{\beta}_1 + \hat{\delta}_1 \times W_1 + \hat{\delta}_2 \times W_2 + \cdots + \hat{\delta}_k \times W_k\right) =$$
$$Var\left(\hat{\beta}_1\right) + \sum_{j=1}^{k} W_j^2 Var\left(\hat{\delta}_j\right) + 2\sum_{j=1}^{k} W_j Cov\left(\hat{\beta}_1, \hat{\delta}_j\right) + 2\sum_{j=1}^{k}\sum_{l=j}^{k} W_j W_l Cov\left(\hat{\delta}_j, \hat{\delta}_l\right),$$

para valores fijados de $W_1, \ldots, W_k$.

Ajuste de las curvas de riesgo / supervivencia en el modelo de Cox

El modelo de Cox (1) puede expresarse fácilmente en términos de funciones de supervivencia. El modelo en función de las funciones de riesgo es:

$$h(t, \mathbf{X}) = h_0(t)\exp\left(\sum_{j=1}^{p} \beta_j X_j\right).$$

Puesto que se cumple la relación $S(t) = \exp\left(-\int_0^t h(s)\,ds\right)$, es fácil comprobar que el modelo de Cox en términos de funciones de supervivencia es:

$$S(t, \mathbf{X}) = S_0(t)^{\exp\left(\sum_{j=1}^{p} \beta_j X_j\right)}, \tag{17}$$

dónde $S_0(t)$ se denomina función de "supervivencia basal" siguiendo la idea de la función de riesgo basal, $h_0(t)$, que se corresponde con la supervivencia de un individuo "base" al que le corresponderían covariables todas iguales a 0.

En el modelo de Cox ajustamos las curvas de supervivencia teniendo en cuenta la parte paramétrica ya estimada del modelo, $\exp\left(\sum_{j=1}^{p} \hat{\beta}_j X_j\right)$, con lo que el ajuste de las curvas tiene en cuenta los valores de las variables explicativas utilizadas como predictores en el modelo. A estas curvas las denominamos "curvas de supervivencia ajustadas" y al igual que las curvas KM se trazan como funciones escalonadas para cada uno de los tiempos de supervivencia. Las curvas estimadas serán:

$$\hat{S}(t, \mathbf{X}) = \hat{S}_0(t)^{\exp\left(\sum_{j=1}^{p} \hat{\beta}_j X_j\right)}. \tag{18}$$

En este paso del proceso de estimación se ajusta la parte no paramétrica relativa a la función de supervivencia basal. Sin embargo, observamos que el input a introducir para la estimación de la función de supervivencia depende de los tiempos de supervivencia y también de los valores de las variables explicativas del modelo.

Así, para la estimación de $\hat{S}(t, \mathbf{X})$ debemos indicar los valores de $\mathbf{X}$ para los cuáles queremos dibujar la curva de supervivencia. A no ser que estemos interesados en dibujar la de algún perfil en concreto, una práctica habitual es escoger el valor medio (o la mediana) de los predictores:

$$\hat{S}(t, \bar{\mathbf{X}}) = \hat{S}_0(t)^{\exp\left(\sum_{j=1}^{p} \hat{\beta}_j \bar{X}_j\right)}. \tag{19}$$

La curva resultante sería una curva "resumen" de todas las posibles curvas para los diferentes valores de las covariables.

Si por ejemplo, la variable X_1 es una variable de exposición que puede tomar los valores 0 y 1 para los dos grupos en estudio, y queremos dibujar en el mismo gráfico las curvas de supervivencia de los dos grupos teniendo en cuenta el resto de variables de confusión, podemos dibujar:

$$\hat{S}(t, 1, \bar{X}_2, \bar{X}_3, \dots, \bar{X}_p) = \hat{S}_0(t)^{exp(\hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3 + \cdots + \hat{\beta}_p \bar{X}_p)},$$

y

$$\hat{S}(t, 0, \bar{X}_2, \bar{X}_3, \dots, \bar{X}_p) = \hat{S}_0(t)^{exp(\hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3 + \cdots + \hat{\beta}_p \bar{X}_p)}.$$

Los métodos de estimación de esta función no paramétrica y no especificada que contempla el paquete `survival` de `R` son tres:

- Método Nelson-Aalen-Breslow (por defecto en ties "breslow")
- Método Efron, método por defecto (por defecto en ties "efron")
- Método Kalbfleish-Prentice (por defecto en ties "exact")

Ahora pasaremos a comentar las formulaciones de dichos métodos que encontramos por ejemplo en: Lee, E. T. and Wang, J. W. (2003). *Satistical Methods for suvirval data analysis*. Wiley Series, pp. 319-322.

Pero antes comentar que los métodos de estimación de esta parte no paramétrica se basan en metodologías que maximizan una función de verosimilitud definida con las aportaciones de las curvas de supervivencia, con lo que bajo condiciones de regularidad, las curvas siguen una distribución normal de la cual podemos calcular fácilmente la esperanza y la varianza. Es por ello que podremos calcular intervalos de confianza.

Método Nelson-Aalen-Breslow

$$\hat{H}_0(t) = \sum_{t_{(i)} \leq t} \left[\frac{m_{(i)}}{\sum_{l \in R(t_{(i)})} \exp\left(\hat{\boldsymbol{\beta}} \mathbf{X}_l^{'}\right)} \right]. \tag{20}$$

donde $\hat{S}_0(t) = \exp\left[-\hat{H}_0(t)\right]$ y las curvas estimadas serán (18).

Método Efron

$$\hat{H}_0(t) = \sum_{t_{(i)} \leq t} \left[\sum_{j=1}^{m_{(i)}} \frac{1}{\sum_{l \in R(t_{(i)})} \exp\left(\hat{\boldsymbol{\beta}} \mathbf{X}_l^{'}\right) - \left[(j-1)/m_{(i)}\right] \sum_{l \in \mathbf{u}_{(i)}^{*}} \exp\left(\hat{\boldsymbol{\beta}} \mathbf{X}_l^{'}\right)} \right]. \tag{21}$$

donde $\hat{S}_0(t) = \exp\left[-\hat{H}_0(t)\right]$ y las curvas estimadas serán (18).

La referencia para éste método la encontramos en el libro Therneau and Grambsch (2000) del autor del paquete `survival` de `R`, dónde éste es el método por defecto en la estimación de la función de supervivencia.

Método Kalbfleish-Prentice

$$\hat{S}_0(t) = \prod_{t_{(i)} \leq t} \alpha_i$$

dónde α_i para $i = 1, \ldots, k$ se obtienen de solucionar el sistema de k ecuaciones:

$$\sum_{j \in \mathbf{u}_{(i)}^{*}} \frac{\exp\left(\hat{\boldsymbol{\beta}} \mathbf{X}_j^{'}\right)}{1 - \hat{\alpha}_i^{\exp\left(\hat{\boldsymbol{\beta}} \mathbf{X}_j^{'}\right)}} = \sum_{l \in R(t_{(i)})} \exp\left(\boldsymbol{\beta} \mathbf{X}_l^{'}\right) \quad \text{para} \quad i = 1, \ldots, k, \tag{22}$$

y las curvas estimadas serán (18).

En particular, la solución cuando no hay empates viene dada por:

$$\hat{S}_0\left(t\right)=\prod_{t_{(i)}\leq t}\left[1-\frac{\exp\left(\boldsymbol{\beta}\mathbf{X}_{(i)}^{'}\right)}{\sum_{l\in R\left(t_{(i)}\right)}\exp\left(\boldsymbol{\beta}\mathbf{X}_{l}^{'}\right)}\right]^{\exp\left(-\boldsymbol{\beta}\mathbf{X}_{(i)}^{'}\right)}.$$

2. Evaluación de la hipótesis de riesgos proporcionales

En este apartado vamos a explicar tres formas de evaluar la hipótesis de riesgos proporcionales:

- Métodos gráficos. El más popular es el uso de los gráficos denominados “log-log”. Si se cumple la hipótesis de riesgos proporcionales estos gráficos deben ser visualmente paralelos para las diferentes clases de las covariables del modelo por separado. Otro método es el de comparar las curvas de supervivencia “observada” y “esperada” para las diferentes clases de las covariables por separado. Éstas curvas deberían ser parecidas visualmente.
- Estudio de la bondad de ajuste (goodness-of-fit, GOF) de la hipótesis de riesgos proporcionales mediante un test estadístico para cada variable del modelo. Mediante el test se obtienen *p*-valores con los que decidir si se cumple o no la hipótesis de riesgos proporcionales. De este modo tenemos una medida menos subjetiva que con el método gráfico y menos complicada computacionalmente que con la inclusión de variables dependientes del tiempo. Sin embargo, el método GOF es demasiado global como para detectar algunas desviaciones de la hipótesis de riesgos proporcionales que el método gráfico y el de inclusión de variables dependientes del tiempo sí pueden detectar y que pueden ser de ayuda en las conclusiones e interpretaciones de la hipótesis.
- Estudio mediante el uso de variables dependientes del tiempo. Este camino implica utilizar el modelo de Cox ampliado con los datos introducidos en formato “counting process” o “(start, stop)”. Aunque dedicaremos un apartado a este modelo ampliado, ahora explicaremos su uso en la validación de la hipótesis de riesgos proporcionales.

Recordemos que en el modelo de Cox (1),

$$h\left(t, X_1, \ldots, X_p\right) = h_0\left(t\right) \exp\left(\sum_{j=1}^{p} \beta_j X_j\right),$$

se introduce como input el tiempo y el valor de los predictores fijos (o independientes del tiempo). El tiempo interviene en la función de riesgo basal no paramétrica y los predictores en la parte paramétrica del predictor lineal.

Recordemos también la expresión (4) en que calculamos la razón de riesgos entre dos sujetos con diferente vector de covariables $\mathbf{X}=\left(X_1,\ldots,X_p\right)$ y $\mathbf{X}^*=\left(X_1^*,\ldots,X_p^*\right)$:

$$HR=\exp\left(\sum_{j=1}^{p}\beta_j\left(X_j^*-X_j\right)\right).$$

Y finalmente recordemos en qué consistía la hipótesis de riesgos proporcionales (5):

$$\hat{\theta}=\exp\left(\sum_{j=1}^{p}\hat{\beta}_j\left(X_j^*-X_j\right)\right),$$

en la que se asume que $\dfrac{\hat{h}\left(t,\mathbf{X}^*\right)}{\hat{h}\left(t,\mathbf{X}\right)}=\hat{\theta}$ es constante a lo largo del tiempo t, es decir, $\hat{h}\left(t,\mathbf{X}^*\right)=\hat{\theta}\times\hat{h}\left(t,\mathbf{X}\right)$.

Si calculamos los riesgos del numerador y denominador de un *HR*, los dibujamos y éstos se cruzan, significa que la hipótesis de riesgos proporcionales no se cumple. Sin embargo, si no se cruzan no significa que sean proporcionales de modo obligatorio, hay que investigar un poco más. Para ello haremos uso de alguno de los métodos que describimos a continuación.

Métodos gráficos

Gráfico log-log de la curva de supervievencia

El gráfico denominado log-log consiste en graficar la función:

$$-\ln\left(-\ln\left(\hat{S}(t)\right)\right). \tag{23}$$

La primera función de (23), $\ln\left(\hat{S}(t)\right)$, siempre es negativa, con lo que para aplicar de nuevo el logaritmo debemos cambiar el signo. Finalmente, los valores resultantes pueden ser o no negativos. Partimos de los valores de $\hat{S}(t)$ que toman valores entre 0 y 1 y obtenemos valores que oscilan entre $-\infty$ y $+\infty$.

Un ejemplo de valor positivo sería: $\hat{S}=0.54$, con lo que $\ln\left(\hat{S}(t)\right)=-0.6161861$ y finalmente $-\ln\left(-\ln\left(\hat{S}(t)\right)\right)=-\ln\left(0.6161861\right)=0.4842062$.

Un ejemplo de valor negativo sería: $\hat{S} = 0.25$, con lo que $\ln\left(\hat{S}(t)\right) = -1.386294$ y finalmente $-\ln\left(-\ln\left(\hat{S}(t)\right)\right) = -\ln(1.386294) = -0.3266343$.

Veamos de qué modo pueden utilizarse las curvas log-log para evaluar la hipótesis de riesgos proporcionales. La expresión que resulta en el modelo de Cox, teniendo en cuenta la definición de la función de supervivencia (18),

$$\hat{S}(t,\mathbf{X}) = \hat{S}_0(t)^{\exp\left(\sum_{j=1}^{p}\hat{\beta}_j X_j\right)},$$

es la siguiente: primero aplicamos logaritmo,

$$\ln\left(\hat{S}(t,\mathbf{X})\right) = \ln\left(\hat{S}_0(t)^{\exp\left(\sum_{j=1}^{p}\hat{\beta}_j X_j\right)}\right) = \exp\left(\sum_{j=1}^{p}\hat{\beta}_j X_j\right)\ln\left(\hat{S}_0(t)\right),$$

cambiamos el signo,

$$-\ln\left(\hat{S}(t,\mathbf{X})\right) = \exp\left(\sum_{j=1}^{p}\hat{\beta}_j X_j\right)\times\left(-\ln\left(\hat{S}_0(t)\right)\right),$$

volvemos a aplicar logaritmo,

$$\ln\left(-\ln\left(\hat{S}(t,\mathbf{X})\right)\right) = \ln\left(\exp\left(\sum_{j=1}^{p}\hat{\beta}_j X_j\right)\times\left(-\ln\left(\hat{S}_0(t)\right)\right)\right) = \sum_{j=1}^{p}\hat{\beta}_j X_j + \ln\left(-\ln\left(\hat{S}_0(t)\right)\right) \quad (24)$$

y, finalmente, cambiamos el signo:

$$-\ln\left(-\ln\left(\hat{S}(t,\mathbf{X})\right)\right) = -\sum_{j=1}^{p}\hat{\beta}_j X_j - \ln\left(-\ln\left(\hat{S}_0(t)\right)\right). \quad (25)$$

Algunos programas que realizan este tipo de gráficos se basan en (24) sin cambiar por segunda vez el signo. Este es el caso de la función `plot.survfit` del paquete `survival` de `R`. Pero en realidad da igual porque al final acabamos evaluando las diferencias de dicha función evaluada en dos valores diferentes de las covariables, con lo que no importa cuál sumamos o restamos, ya que lo importante es que las diferencias se mantenga paralelas en el gráfico log-log.

Supongamos dos sujetos con vectores de covariables $\mathbf{X}_1 = \left(X_{11}, \ldots, X_{1p} \right)$ y $\mathbf{X}_2 = \left(X_{21}, \ldots, X_{2p} \right)$. La diferencia $\ln\left(-\ln\left(\hat{S}\left(t, \mathbf{X}_1\right) \right) \right) - \ln\left(-\ln\left(\hat{S}\left(t, \mathbf{X}_2\right) \right) \right)$ para un momento dado del tiempo (continuo) t es la siguiente:

$$\sum_{j=1}^{p} \hat{\beta}_j X_{1j} + \ln\left(-\ln\left(\hat{S}_0\left(t\right) \right) \right) - \sum_{j=1}^{p} \hat{\beta}_j X_{2j} - \ln\left(-\ln\left(\hat{S}_0\left(t\right) \right) \right),$$

Los términos $\ln\left(-\ln\left(\hat{S}_0\left(t\right) \right) \right)$ quedan (hipotéticamente) eliminados y obtenemos que:

$$\ln\left(-\ln\left(\hat{S}\left(t, \mathbf{X}_1\right) \right) \right) - \ln\left(-\ln\left(\hat{S}\left(t, \mathbf{X}_2\right) \right) \right) = \sum_{j=1}^{p} \hat{\beta}_j \left(X_{1j} - X_{2j} \right).$$

Si despejamos una de las funciones obtenemos:

$$\ln\left(-\ln\left(\hat{S}\left(t, \mathbf{X}_1\right) \right) \right) = \ln\left(-\ln\left(\hat{S}\left(t, \mathbf{X}_2\right) \right) \right) + \sum_{j=1}^{p} \hat{\beta}_j \left(X_{1j} - X_{2j} \right). \qquad (26)$$

La interpretación respecto de la hipótesis de riesgos proporcionales es que si al dibujar el gráfico log-log obtenemos que las curvas son paralelas, entonces se cumplirá la hipótesis de riesgos proporcionales, ya que en este caso la diferencia o la distancia entre las dos curvas debería ser igual a $\sum_{j=1}^{p} \hat{\beta}_j \left(X_{1j} - X_{2j} \right)$ constante y no depender del tiempo.

Para llevar a la práctica el estudio se utilizan los plots "empíricos" que grafican $-\ln\left(-\ln\left(\hat{S} \right) \right)$ ó $\ln\left(-\ln\left(\hat{S} \right) \right)$ en dos situaciones diferentes:

1. En el caso en que $\hat{S}$ es una curva KM para cada covariable de forma individualizada. En este caso se estudia el gráfico para contrastar la hipótesis de riesgos proporcionales en las clases de la covariable. Se estudia si las gráficas resultantes son o no paralelas. Un problema de este caso es que si queremos testear la diferencia de más de una variable a la vez tendremos un número elevado de curvas (igual al producto del número de clases que tenga cada variable incluida), y será bastante difícil hacer una interpretación visual en el mismo gráfico. Además, el número de datos para cada curva será cada vez más pequeño, llegando a ser en ocasiones imposible dibujar las curvas.

2. En el caso en que $\hat{S}$ es una curva obtenida con el modelo de Cox de riesgos proporcionales utilizando covariables que ya sabemos que cumplen la hipótesis de riesgos proporcionales y sin tener en cuenta la covariable en estudio de la cuál se quiere testear la hipótesis de proporcionalidad en el tiempo. Si suponemos que la variable en estudio tiene dos clases, se trataría de dibujar en el mismo gráfico las curvas estimadas para cada subconjunto dado por dichas clases y utilizando la media global para el resto de variables que cumplen la hipótesis de proporcionalidad. De este modo obtenemos dos curvas y podemos realizar visualmente la inspección de si son o no paralelas.

Este es un método bastante subjetivo. A no ser que se vea claramente que las curvas no son paralelas se debe ser conservador y, ante la duda, no rechazar la hipótesis de proporcionalidad. Además, si la covariable era inicialmente continua y se ha discretizado, la forma del gráfico dependerá mucho de los puntos de corte utilizados. De cara a la interpretación de los resultados es recomendable no utilizar más de 3 clases para el estudio. También será importante que sean clases balanceadas, es decir con más ó menos el mismo número de sujetos en cada clase.

Gráfico de las curvas "observadas" versus "esperadas"

Este método gráfico es el paralelo al método de test GOF que veremos en el siguiente subapartado y, por lo tanto, es de forma natural una alternativa razonable a los gráficos log-log anteriores.

Al igual que con los gráficos log–log, con los gráficos observados versus los esperados podemos seguir dos estrategias:

1. Validar la hipótesis de riesgos proporcionales para las variables de una en una. En este caso obtendremos los plots "observados" mediante el ajuste de curvas KM para cada una de las clases de la variable en estudio. Los plots "esperados" los obtendremos mediante un modelo de Cox incluyendo la variable y dibujando las curvas resultantes para cada uno de los perfiles de las clases. Si las curvas "observadas" y "esperadas" son similares aceptaremos la hipótesis de riesgos proporcionales y si no lo son la rechazaremos.

2. Validar la hipótesis de riesgos proporcionales de una covariable específica incluidas en el modelo otras covariables. La estrategia en este caso para obtener

los plots "observados" pasa por utilizar un modelo de Cox estratificado, dónde el modelo contiene todas las variables y la que se está analizando es la variable de estrato. Al modelo de Cox estratificado le dedicaremos un apartado.

Al igual que ocurre con los gráficos log-log este método es visual, con lo que si tenemos duda en la aceptación o rechazo de la hipótesis de proporcionalidad es mejor ser conservador y no rechazar la hipótesis, a no ser que se vea muy claro que no se cumple, es decir, que los plots "observados" y los "esperados" no se parezcan casi.

Comentario para el caso 1 cuando el predictor es continuo

Discretizaremos en clases, supongamos k, y procederemos del siguiente modo. Para las curvas "observadas" simplemente estimaremos las k curvas KM para cada clase (o estrato) por separado. Pero para la estimación de las curvas "esperadas" tenemos dos posibilidades: a) tratar la variable como continua y estimar las curvas asociadas a cada clase utilizando el perfil medio correspondiente a cada una de las k clases, es decir la media de la variable en cada clase, o bien, b) tratar la variable como categórica codificando $k-1$ variables binarias y estimando las curvas para cada uno de los k perfiles como es usual.

Método GOF

Testearemos de un modo más objetivo la hipótesis nula de proporcionalidad de los riesgos. Puesto que la hipótesis nula será la de proporcionalidad, si el *p*-valor es grande aceptaremos la hipótesis y si es pequeño la rechazaremos. Lo que nos interesa es aceptar dicha hipótesis para que el modelo sea válido, y por lo tanto buscaremos *p*-valores elevados. El contraste lo realizaremos para cada variable del modelo por separado y obtendremos por tanto un *p*-valor por variable.

El método se basa en los denominados residuos de Schoenfeld, los cuales se calculan para cada predictor y para cada individuo. Si por ejemplo tenemos dos predictores en el modelo, para cada sujeto con tiempo de muerte tendremos dos residuos de Schoenfeld.

Los residuos de Schoenfeld se definen como:

$$R_{ij} = \delta_i \left(X_{ij} - \frac{\sum_{l \in R\left(t_{(i)}\right)} X_{lj} \exp\left(\hat{\beta}\mathbf{X}_l^{'}\right)}{\sum_{l \in R\left(t_{(i)}\right)} \exp\left(\hat{\beta}\mathbf{X}_l^{'}\right)} \right), \tag{27}$$

para $i = 1,\ldots,n$ y $j = 1,\ldots,p$, siendo $\delta_i = 1$ el indicador de si el sujeto no está censurado y $\delta_i = 0$ si el sujeto está censurado. Por lo tanto estos residuos sólo están definidos para los individuos no censurados. Observamos que los residuos así calculados son una media ponderada de la covariable con pesos los riesgos del resto de sujetos a riesgo en el momento de fallo.

La idea del test es la siguiente: para un predictor en particular se cumplirá la hipótesis de proporcionalidad de los riesgos si los residuos de Schoenfeld de dicho predictor no están correlacionados con los tiempos de supervivencia. Gráficamente, si dibujamos los residuos de Schoenfeld del predictor, éstos serán horizontales si se cumple la hipótesis de proporcionalidad ya que en tal caso los residuos son independientes del tiempo.

Para la realización debemos: calcular los residuos de Schoenfeld del modelo de Cox en estudio; ordenar los tiempos de fallo etiquetando el orden a cada tiempo; calcular la correlación entre los residuos y la variable de orden creada. Y realizamos el contraste de si $H_0 : \rho = 0$ para cada covariable por separado. En el caso de aceptar la hipótesis nula de que la correlación es cero, se cumplirá la hipótesis de riesgos proporcionales para el predictor correspondiente. Con lo que, tal y como ya habíamos comentado, nos interesará que los *p*-valores del contraste sean elevados para aceptar la hipótesis de proporcionalidad.

Comentario: Aunque el test nos da una medida más objetiva de la aceptación o no de la hipótesis de riesgos proporcionales que los métodos gráficos, lo único que nos indica con seguridad es que no hay suficientes pruebas como para rechazar la hipótesis nula. También puede influir el tamaño de la muestra en el *p*-valor: puede pasar que una violación grande de la hipótesis nula no se detecte si la muestra es pequeña, y a la inversa, una violación pequeña de la hipótesis nula puede ser sumamente significativa si la muestra es grande. Con lo que la recomendación es la siguiente: utilizar conjuntamente los métodos gráficos y el test, ya que por un lado el test es más objetivo y, por otro, los gráficos pueden mostrar desviaciones concretas e información acerca del no cumplimiento de la hipótesis de riesgos porporcionales.

Variables dependientes en el tiempo (modelo de Cox extendido)

Cuando utilizamos variables predictoras dependientes del tiempo con el objetivo de evaluar la hipótesis de riesgos proporcionales el modelo se “extiende” con la inclusión del producto (o sea, la interacción) de las covariables con funciones del tiempo, es decir, $X \times g(t)$, donde $g(t)$ es una función del tiempo t.

Por ejemplo, si queremos evaluar si el sexo cumple o no la hipótesis de riesgos proporcionales podemos hacer:

$$h(t, Sex) = h_0(t)\exp\left(\beta Sex + \delta\left(Sex \times t\right)\right),$$

y evaluar con un test si $\delta = 0$. En caso en que nos salga que $\delta \neq 0$ podremos afirmar que para la variable sexo no se cumple la hipótesis de riesgos proporcionales.

En general, si analizamos las **variables una a una** tendremos:

$$h(t, X) = h_0(t)\exp\left(\beta X + \delta\left(X \times g(t)\right)\right), \tag{28}$$

donde la función $g(t)$ puede tomar diferentes expresiones, por ejemplo:

- $g(t) = t$, $g(t) = t^2$, $g(t) = \sqrt{t}$;
- $g(t) = \ln(t)$ ó $g(t) = \log(t)$, $g(t) = t \cdot \ln(t)$;
- $g(t) = \begin{cases} 1 & si\ t \geq t_0 \\ 0 & si\ t < t_0 \end{cases}$ (Función de cambio de punto).

Bajo la hipótesis nula $H_0 : \delta = 0$ el modelo (28) queda como $h(t, X) = h_0(t)\exp(\beta X)$.

Para realizar el test podemos utilizar el *p*-valor del contraste de Wald o bien del de la razón de máxima verosimilitud.

A parte de utilizar este tipo de contrastes covariable a covariable también los podemos utilizar para un conjunto de covariables simultáneamente o para analizar un predictor ajustado por otros en el modelo.

Si **tenemos simultáneamente varios predictores**, la expresión (28) será:

$$h(t, \mathbf{X}) = h_0(t)\exp\left(\sum_{j=1}^{p} \beta_j X_j + \sum_{j=1}^{p} \delta_j\left(X_j \times g_j(t)\right)\right) \tag{29}$$

donde la función $g_j(t)$ será la función del tiempo t para el predictor *j*-ésimo para $j = 1, \ldots, p$.

La hipótesis nula ahora será $H_0 : \delta_1 = \delta_2 = \cdots = \delta_p = 0$ y el modelo bajo la hipótesis nula quedará $h(t, \mathbf{X}) = h_0(t) \exp\left(\sum_{j=1}^{p} \beta_j X_j \right)$. Podremos contrastar la hipótesis con el test de la razón de verosimilitud en el que $LR = -2\log L_{Cox} - (-2\log L_{CoxExt})$, y seguirá una chi-cuadrado con p grados de libertad.

Si ahora queremos analizar la hipótesis de proporcionalidad para **un predictor ajustado por otros** en el modelo, tendremos el siguiente caso:

$$h(t, \mathbf{X}) = h_0(t) \exp\left(\sum_{j=1}^{p-1} \beta_j X_j + \beta^* X^* + \delta^* \left(X^* \times g^*(t) \right) \right) \tag{30}$$

donde la función $g^*(t)$ será la función del tiempo t para el predictor X^* estudiado. La hipótesis nula será $H_0 : \delta^* = 0$ y la contrastaremos con el test de Wald o el de razón de verosimilitud, en este caso siguiendo una chi-cuadrado con un grado de libertad.

Comentario: Respecto de las funciones del tiempo $g_j(t)$ a escoger, puede pasar que en función de la expresión que incluyamos obtengamos diferentes conclusiones en la realización del test de riesgos proporcionales.

Diagnóstico del modelo: Otros residuos

Otros residuos que se calculan en el modelo de Cox y, concretamente, con la función/comando del paquete `survival` de `R`, `residuals`, son:

- Residuos de martingala
- Residuos de deviance
- Residuos de score
- Residuos de Schoenfeld (que ya hemos visto con detalle)

En dicha función podemos especificar type=c("martingale", "deviance", "score", "schoenfeld", "dfbeta", "dfbetas", "scaledsch", "partial").

Algunas orientaciones en la práctica son:

Respecto de los residuos de martingala: éstos son una transformación de los denominados residuos de Cox-Snell. Tienen una distribución asimétrica y su esperanza debería ser asintóticamente 0. Son útiles para indicar si con las covariables del modelo hemos predicho bien los tiempos de supervivencia. Sirven para analizar transformaciones de las covariables que mejoren dichos residuos. Su fórmula es:

$$R_{M_i} = \delta_i - R_i , \tag{31}$$

para $i = 1,\ldots,n$, siendo $\delta_i = 1$ el indicador de si el sujeto no está censurado y $\delta_i = 0$ si el sujeto está censurado. Dónde $R_i = -\log \hat{S}(t_i, \mathbf{X}_i)$ son los denominados residuos de Cox-Snell extendidos. Los residuos de Cox-Snell se utilizan del siguiente modo: si el modelo de riesgos proporcionales estimado es adecuado el plot de dichos residuos y su correspondiente curva KM, $\hat{S}(R)$, aparecerían como una recta de 45 grados.

Respecto de los residuos basados en deviances: son los análogos para datos censurados de los residuos de desviance de un modelo lineal generalizado de la familia exponencial de McCullagh y Nelder. Estos residuos sí se distribuyen de forma simétrica alrededor del cero y sirven para analizar el ajuste del modelo para cada sujeto no censurado y por lo tanto detectar observaciones "atípicas". Su fórmula es:

$$R_{D_i} = sign\left(R_{M_i}\right)\sqrt{2\left[-R_{M_i} - \delta_i \log\left(\delta_i - R_{M_i}\right)\right]}, \tag{32}$$

para $i = 1,\ldots,n$.

Respecto de los residuos de score: estos residuos se calculan para cada sujeto y cada covariable. Los gráficos de éstos versus las covariables (una a una) indican la "influencia" de los individuos en la estimación de los coeficientes. Los "dfbeta" y "dfbetas" (el "dfbeta" estandarizado) se calculan para cada sujeto e indican el cambio aproximado en el vector de coeficientes si el sujeto concreto no estuviese en el modelo. Su fórmula es:

$$R_{ij}(t) = \int_0^t \left\{X_{ij}(u) - \overline{X}_j(u)\right\} d\hat{M}_i(u), \tag{33}$$

para $i = 1,\ldots,n$. Dónde:

$$\overline{X}_j(u) = \frac{\sum_{i=1}^n J_i(t) X_{ij} \exp\left(\boldsymbol{\beta} X_i^{'}(t)\right)}{\sum_{i=1}^n J_i(t) \exp\left(\boldsymbol{\beta} X_i^{'}(t)\right)} \text{ y } \hat{M}_i(u) = N_i(u) - \int_0^t J_i(s) \exp\left(\boldsymbol{\beta} X_i^{'}(s)\right) d\hat{H}_0(s),$$

siendo $J_i(t)=1$ si el individuo i está a riesgo antes de t, y $N_i(t)$ el indicador del proceso de contaje (*counting process*) de si el individuo *i*-ésimo ha tenido fallo, es decir le ha ocurrido el suceso de interés. Observamos que estos residuos tienen en cuenta variables dependientes del tiempo. Su expresión se particulariza cuando las variables son independientes del tiempo.

Una referencia para más detalle sobre las formulaciones de los diferentes tipos de residuos y su utilización en el modelo de Cox puede ser por ejemplo: Lee, E. T. and Wang, J. W. (2003). *Satistical Methods for suvirval data analysis*. Wiley Series, pp. 330-337.

3. El modelo de Cox estratificado

El modelo de Cox estratificado es una modificación del modelo de Cox (1) de riesgos proporcionales que permite incluir un control mediante una estratificación dada por las clases de un predictor que no cumple la hipótesis de proporcionalidad de riesgos dados otros predictores que sí la cumplen. Consideraremos la posibilidad de estratificar por un único predictor o por varios a la vez. Distinguiremos entre el uso de una versión "sin interacción" de dicho modelo y una alternativa incluyendo "interacción".

Formulación general del modelo de Cox estratificado

Supongamos que tenemos p predictores que cumplen la hipótesis de proporcionalidad de riesgos, $\mathbf{X} = \left(X_1, \ldots, X_p\right)$, y que queremos estratificar por k que no cumplen dicha hipótesis, $\mathbf{Z} = \left(Z_1, \ldots, Z_k\right)$. El primer paso es generar (o codificar) una variable Z^* que recoja las diferentes categorías en que queremos estratificar, supongamos k^* en total. Para ello si alguna de las variables incluidas para estratificar es de naturaleza continua debemos discretizarla previamente. Si únicamente estratificamos por una variable, el número total de estratos diferentes será exactamente el número de clases de la variable.

Supongamos que estratificamos por k variables, mediante la nueva variable Z^* que recoge en su codificación las diferentes combinaciones, k^*, de todas las clases provenientes de las diferentes clases de las k variables iniciales. El modelo general de Cox estratificado "sin interacción" será:

$$h_g\left(t, \mathbf{X}\right) = h_{0g}\left(t\right) \exp\left(\sum_{j=1}^{p} \beta_j X_j\right), \tag{34}$$

con $g = 1, \ldots, k^*$ estratos definidos por Z^*.

Observamos que Z^* no se incluye en la parte paramétrica del predictor lineal, con lo que los HR que calculemos serán comunes para todos los estratos y no se podrán calcular para las variables que no cumplen la hipótesis de riesgos proporcionales y que nos sirven para estratificar el modelo. A este modelo se le denomina de "no-interacción".

Sin embargo, las variables $\mathbf{Z} = \left(Z_1, \ldots, Z_k\right)$ sí servirán para diferenciar las funciones de riesgo basal a través de Z^*, con lo que tendremos en total k^* funciones de riesgo basal y por lo tanto también k^* curvas de riesgo estimadas: $\hat{h}_{0g}(t)$ y $\hat{h}_g(t)$ para $g = 1,\ldots,k^*$, y sus correspondientes funciones de supervivencia:

$$\hat{S}_g(t, \mathbf{X}) = \hat{S}_{0g}(t)^{\exp\left(\sum_{j=1}^{p} \beta_j X_j\right)}. \qquad (35)$$

Para estimar el modelo se calcula la función de verosimilitud parcial del siguiente modo:

$$L = \prod_{g=1}^{k^*} L_g \qquad (36)$$

donde cada L_g para $g = 1,\ldots,k^*$ es la función de verosimilitud parcial de cada uno de los estratos por separado. Observamos pues, que las estimaciones de los parámetros son comunes para todos los estratos, pero que la función de verosimilitud parcial total sí distingue dichos estratos. Una vez estimados los parámetros $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1,\ldots,\hat{\beta}_p\right)$, la idea es que para cada función de verosimilitud parcial se estiman las diferentes funciones de riesgo asociadas, es decir para cada estrato.

Hipótesis de "no-interacción"

Vamos a ver qué significa la hipótesis de "no-interacción", cómo se evalúa y qué hacer cuando no se cumple.

Supongamos que tenemos k^* estratos que no cumplen la hipótesis de riesgos proporcionales. Estimaremos el modelo de Cox estratificado incluyendo Z^* como variable de estrato. En este modelo obtendremos una única estimación de $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1,\ldots,\hat{\beta}_p\right)$ común a todos los estratos y lo denominaremos de "no-interacción".

Otra posibilidad (aunque con otros resultados) es dividir previamente los datos en los k^* estratos y ajustar para cada uno por separado un modelo de Cox sin estratificar. De este modo obtendríamos k^* estimaciones de $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1,\ldots,\hat{\beta}_p\right)$, una para cada modelo. Se trataría de decidir si los k^* coeficientes diferentes que obtenemos para cada uno de los p predictores de estos k^* modelos son o no estadísticamente diferentes. En el caso de que consideremos que son iguales, el modelo de Cox estratificado será el adecuado. En caso contrario deberemos estimar un modelo de Cox estratificado que tenga en cuenta términos de interacción. Vamos a ver cómo contrastamos si los coeficientes son diferentes para el total de modelos individuales y cómo construimos para ello el denominado modelo de Cox estratificado "con interacción".

El modelo de Cox estratificado "sin interacción" viene dado por la fórmula (34):

$$h_g\left(t,\mathbf{X}\right)=h_{0g}\left(t\right)\exp\left(\sum_{j=1}^{p}\beta_j X_j\right),$$

con $g=1,\dots,k^*$ estratos definidos por Z^*.

Los modelos individuales de Cox para cada estrato los definiremos con la siguiente nomenclatura:

$$h_g\left(t,\mathbf{X}\right)=h_{0g}\left(t\right)\exp\left(\sum_{j=1}^{p}\beta_{jg} X_j\right), \tag{37}$$

con $g=1,\dots,k^*$ estratos definidos por Z^*. Se trata de averiguar si $\beta_{jg}=\beta_j$ para $j=1,\dots,p$ y $g=1,\dots,k^*$.

Para definir el modelo de Cox estratificado "con interacción" primero definiremos las variables que nos permitirán incluir los términos de interacción. Construiremos k^*-1 variables binarias, $Z_1^*,\dots,Z_{k-1}^*$, a partir de las k^* categorías de Z^* y con ellas construiremos los términos de interacción incluyendo las interacciones de primer orden entre estas variables y los p predictores del modelo, $\mathbf{X}=\left(X_1,\dots,X_p\right)$, que sí cumplen la hipótesis de proporcionalidad. De este modo, el modelo de Cox estratificado "con interacción" se define como:

$$h_g\left(t,\mathbf{X}\right)=h_{0g}\left(t\right)\exp\left(\sum_{j=1}^{p}\beta_j X_j+\sum_{k=1}^{k^*-1}\sum_{j=1}^{p}\beta_{jk} Z_k^* X_j\right), \tag{38}$$

con $g=1,\dots,k^*$ estratos definidos por Z^*. Puesto que es un modelo de Cox calculado con la formulación del modelo estratificado, observamos que el subíndice g únicamente afecta a la función de riesgo basal y no a los parámetros, que en este modelo son: $\boldsymbol{\beta}=\left(\beta_1,\dots,\beta_p,\beta_{11},\dots,\beta_{p1},\beta_{12},\dots,\beta_{p2},\cdots,\beta_{1(k^*-1)},\dots,\beta_{p(k^*-1)}\right)$ en total $p+p\times\left(k^*-1\right)$.

Para hacer el test de si los términos de interacción son significativos o no, podemos utilizar el test de razón de verosimilitud dónde la hipótesis nula es:

$$H_0:\beta_{11}=\cdots=\beta_{p1}=\beta_{12}=\cdots=\beta_{p2}=\cdots=\beta_{1(k^*-1)}=\cdots=\beta_{p(k^*-1)}=0.$$

Si L_p indica la verosimilitud parcial evaluada en el modelo de Cox "sin interacción" (34) con p parámetros y $L_{p+p\times(k^*-1)}$ la verosimilitud parcial evaluada en el modelo de Cox "con interacción" (38) con $p+p\times(k^*-1)$ parámetros, entonces $LR=-2\log L_p-\left(-2\log L_{p+p\times(k^*-1)}\right)$ sigue una distribución chi-cuadrado con $p\times(k^*-1)$ grados de libertad.

4. Extensión del modelo de Cox de riesgos proporcionales para variables dependientes del tiempo

Como ya hemos visto, la hipótesis de riesgos proporcionales la podemos validar mediante tres metodologías: 1) utilización de un método gráfico (log-log ó "observadas" vs "esperadas"); 2) contrastación con un test estadístico; 3) inclusión de variables explicativas dependientes del tiempo mediante un modelo de Cox extendido.

En el caso 3) en el que se incluyen variables dependientes del tiempo en un modelo de Cox extendido, se extiende el modelo en el sentido de que se incorporan los términos del producto (i.e., la interacción) de la variable predictora independiente del tiempo por alguna función del tiempo. Recordemos que pusimos a modo de ejemplo el del sexo en la fórmula (28).

En caso de que después de evaluar la hipótesis de riesgos proporcionales, obtengamos que no se cumple dicha hipótesis para uno ó mas predictores, tenemos dos caminos: 1) aplicar el modelo de Cox estratificado, estratificando por los predictores que no cumplen la hipótesis de proporcionalidad y manteniendo los que sí la cumplen; 2) utilización de variables dependientes del tiempo. En esta sección nos ocuparemos de describir esta segunda opción.

Las variables dependientes del tiempo pueden ser: **intrínsecamente dependientes del tiempo** o bien **definidas** para la evaluación de la hipótesis de riesgos proporcionales a partir de predictores independientes del tiempo. Por ejemplo, la variable raza, $RACE$, es independiente del tiempo, pero $RACE \times t$ no lo es. Si $RACE$ toma los valores 0 y 1, entonces:

- cuando $RACE = 1 \Rightarrow RACE \times t = t$
- cuando $RACE = 0 \Rightarrow RACE \times t = 0$

Supongamos que E denota una variable de exposición que toma los valores $(0,1)$, podemos definir, por ejemplo, la variable $E \times (\log t - 3)$, donde el tiempo entra en el modelo a través de una función $g(t) = (\log t - 3)$ en lugar de entrar únicamente a través de $g(t) = t$ como ocurría en el ejemplo anterior.

Un ejemplo de función que se utiliza en la contrastación de la hipótesis de riesgos proporcionales es la denominada de "cambio de punto":

$$g(t) = \begin{cases} 1 & si\ t \geq t_0 \\ 0 & si\ t < t_0. \end{cases} \tag{39}$$

Otro tipo de variables dependientes del tiempo son aquellas que intrínseca o internamente dependen del tiempo, ya que sus valores varían por las propias características del sujeto. Por ejemplo, el nivel de exposición en el tiempo t, $E(t)$, estado de empleo en el tiempo t, $EMP(t)$, estado de fumador en el tiempo t, $SMK(t)$, o nivel de obesidad en el tiempo t, $OBS(t)$.

Hay otro tipo de variables dependientes del tiempo ya que sus valores varían a lo largo de este, pero no por causas intrínsecas al individuo sino debido a características externas. A estas variables se las denomina "**auxiliares**". Por ejemplo, el índice del nivel de polución en el tiempo t, y también el estado de empleo en el tiempo t, $EMP(t)$, siempre que el estado de empleado o desempleado dependa de las condiciones económicas más que de las características individuales del sujeto.

Otro ejemplo, que puede considerarse en parte dependiente intrínsecamente y en parte dependiente auxiliarmente, es la variable estado del trasplante de corazón (HT) en el tiempo t para una persona identificada idónea o elegible para un trasplante por tener una enfermedad de corazón seria. El valor de esta variable en el tiempo t será 1, $HT = 1$, si la persona ya ha recibido un trasplante en algún tiempo, t_0, anterior al tiempo t analizado, y el valor será es 0, $HT = 0$, si en t la persona aún no ha recibido un trasplante. Podemos escribir esta variable como:

$$HT(t) = \begin{cases} 1 & t \geq t_0 \\ 0 & t < t_0. \end{cases} \tag{40}$$

Esta variable puede ser considerada esencialmente una variable intrínseca porque los rasgos individuales de un receptor de trasplante elegible son los determinantes importantes en la decisión de realizar una cirugía de trasplante. Sin embargo, la disponibilidad de un corazón adecuado para el receptor puede ser considerada una característica "auxiliar" externa al receptor.

En resumen, podemos tener variables dependientes del tiempo: definidas, intrínsecas y auxiliares.

Formulación del modelo de Cox extendido para variables dependientes del tiempo

Supongamos que disponemos de p_1 predictores independientes del tiempo, $X_1,\ldots,X_{p_1}$, y p_2 predictores dependientes del tiempo, $X_1(t),\ldots,X_{p_2}(t)$, con lo que si denominamos al conjunto de predictores $\mathbf{X}(t)=\left(X_1,\ldots,X_{p_1},X_1(t),\ldots,X_{p_2}(t)\right)$, el modelo de Cox extendido se define como:

$$h\left(t,\mathbf{X}(t)\right)=h_0(t)\exp\left(\sum_{j=1}^{p_1}\beta_j X_j+\sum_{j=1}^{p_2}\delta_j X_j(t)\right). \tag{41}$$

En este modelo la estimación de los parámetros también se realiza por máxima verosimilitud parcial. La diferencia con el modelo de Cox clásico es que los modelos que intervienen en los conjuntos a riesgo de los diferentes tiempos son más sofisticados que los de un modelo con únicamente predictores independientes del tiempo. Deberemos tener en cuenta que los datos estarán en formato "counting process" o "(start, stop)".

Los métodos estadísticos para realizar inferencia son básicamente los mismos que para el modelo de Cox de riesgos porporcionales. Podemos utilizar el test de Wald y tests basados en razones de verosimilitud, así como calcular intervalos de confianza, ya que la estimación está basada en la maximización de la función de verosimilitud parcial.

Modelo de Cox extendido y retardado

Observamos en (41) que la función de riesgo en el tiempo t depende del valor que toman los p_2 predictores $X_j(t)$ justamente en el instante t. Sin embargo, es posible modificar la definición de las variables dependientes del tiempo para tener en cuenta un efecto "de tiempo de retraso" L_j para $j=1,\ldots,p_2$. En tal caso, el modelo se formula como:

$$h\left(t,\mathbf{X}(t)\right)=h_0(t)\exp\left(\sum_{j=1}^{p_1}\beta_j X_j+\sum_{j=1}^{p_2}\delta_j X_j\left(t-L_j\right)\right). \tag{42}$$

Formulación de los ratios de riesgo, HR

Dados dos conjuntos de predictores $\mathbf{X}^*(t)=\left(X_1^*,\ldots,X_{p_1}^*,X_1^*(t),\ldots,X_{p_2}^*(t)\right)$ y $\mathbf{X}(t)=\left(X_1,\ldots,X_{p_1},X_1(t),\ldots,X_{p_2}(t)\right)$, definimos el $HR(t)$ como:

$$HR(t)=\frac{h\left(t,\mathbf{X}^*(t)\right)}{h\left(t,\mathbf{X}(t)\right)}=\frac{h_0(t)\exp\left(\sum_{j=1}^{p_1}\beta_j X_j^*+\sum_{j=1}^{p_2}\delta_j X_j^*(t)\right)}{h_0(t)\exp\left(\sum_{j=1}^{p_1}\beta_j X_j+\sum_{j=1}^{p_2}\delta_j X_j(t)\right)},$$

y simplificando,

$$HR(t)=\exp\left(\sum_{j=1}^{p_1}\beta_j\left(X_j^*-X_j\right)+\sum_{j=1}^{p_2}\delta_j\left(X_j^*(t)-X_j(t)\right)\right). \tag{43}$$

Un ejemplo suponiendo una única variable E de exposición que tome los valores $(0,1)$, y utilizando $g(t)=t$ tenemos el modelo de Cox extendido:

$$h\left(t,\mathbf{X}(t)\right)=h_0(t)\exp\left(\beta E+\delta(t\times E)\right).$$

El HR lo calculamos como:

$$HR(t)=\frac{h(t,E=1)}{h(t,E=0)}=\exp(\beta+\delta t).$$

Observamos como el HR no cumple la hipótesis de proporcionalidad de riesgos.

En general, el modelo de Cox extendido no cumple la hipótesis de proporcionalidad de los riesgos, a no ser que los coeficientes δ_j para $j=1,\ldots,p_2$ sean estadísticamente cero. Por otro lado, cabe notar que cada δ_j es independiente del tiempo y representa el efecto global de la variable a lo largo del tiempo, es decir, teniendo en cuenta todas las observaciones de la variable a lo largo del tiempo.

Veamos otro ejemplo, supongamos la variable $E(t)$ que mide semanalmente el estado de exposición químico del sujeto, y toma valores 0 si no está expuesto esa semana y 1 si esa semana el sujeto sí lo está. El modelo de Cox extendido que incluye únicamente dicha variable es:

$$h\left(t,E(t)\right)=h_0(t)\exp\left(\delta\times E(t)\right).$$

El HR no significará exactamente lo mismo que una variable de estado que valoraría expuestos vs no expuestos, pero igualmente calculamos el HR para las dos categorías:

$$HR(t) = \frac{h(t, E=1)}{h(t, E=0)} = \exp(\delta).$$

Observamos que nos ha quedado un valor constante, $\exp(\delta)$, sin embargo para este modelo no se cumple la hipótesis de riesgos proporcionales, ya que $HR(t)$ es dependiente del tiempo debido a que $E(t)$ también lo es. En este caso el HR se interpreta como la razón de riesgos para un momento del tiempo t en que $E(t) = 1$.

Evaluación de la hipótesis de riesgos proporcionales para variables independientes del tiempo a partir del modelo de Cox extendido

En general incluiremos un término para cada una de las variables independientes del tiempo y de las cuales queremos evaluar la hipótesis de riesgos proporcionales tal y como ya habíamos indicado en (28), de este modo:

$$h(t, \mathbf{X}) = h_0(t) \exp\left(\sum_{j=1}^{p} \beta_j X_j + \sum_{j=1}^{p} \delta_j \left(X_j \times g_j(t) \right) \right),$$

donde la función $g_j(t)$ será la función del tiempo t para el predictor *j*-ésimo para $j = 1, \ldots, p$.

Algunos ejemplos de funciones $g_j(t)$ son:

<u>Ejemplo 1</u>. Si $g_j(t) = 0$, sería el caso de un modelo de Cox sin variables dependientes:

$$h(t, \mathbf{X}) = h_0(t) \exp\left(\sum_{j=1}^{p} \beta_j X_j \right).$$

<u>Ejemplo 2</u>. Si $g_j(t) = t$, tendríamos:

$$h(t, \mathbf{X}) = h_0(t) \exp\left(\sum_{j=1}^{p} \beta_j X_j + \sum_{j=1}^{p} \delta_j \left(X_j \times t \right) \right).$$

<u>Ejemplo 3</u>. Si $g_j(t) = \ln t$, tendríamos:

$$h(t,\mathbf{X}) = h_0(t)\exp\left(\sum_{j=1}^{p}\beta_j X_j + \sum_{j=1}^{p}\delta_j\left(X_j \times \ln t\right)\right).$$

Con esta función del tiempo la expresión queda:

$$h(t,\mathbf{X}) = h_0(t)\exp\left(\sum_{j=1}^{p}\beta_j X_j\right)t^{\sum_{j=1}^{p}\delta_j X_j}.$$

Ejemplo 4. Si $g_j(t) = \begin{cases} 1 & si\ t \geq t_0 \\ 0 & si\ t < t_0 \end{cases}$, la función de cambio de punto, (39), tendríamos la formulación general $h(t,\mathbf{X}) = h_0(t)\exp\left(\sum_{j=1}^{p}\beta_j X_j + \sum_{j=1}^{p}\delta_j\left(X_j \times g_j(t)\right)\right)$. En breve vemos con detalle este ejemplo, en el que los HR son constantes a tramos del tiempo.

Contraste: Para realizar el contraste de si se cumple o no la hipótesis de riesgos proporcionales realizaremos el contraste de razón de verosimilitud para los parámetros correspondientes a las variables dependientes. La hipótesis nula será: $H_0 : \delta_1 = \cdots = \delta_p = 0$, y la contrastaremos con $LR = -2\log L_p - \left(-2\log L_{2p}\right)$ que sigue una distribución chi-cuadrado con p grados de libertad, donde L_p indica la verosimilitud parcial evaluada en el modelo de Cox únicamente incluyendo las p variables independientes del tiempo y L_{2p} la verosimilitud parcial evaluada en el modelo completo. De este modo, si aceptamos la hipótesis nula, el modelo cumplirá la hipótesis de riesgos proporcionales. En caso de que los coeficientes sean significativos, será preferible el modelo de Cox ampliado y los HR no cumplirán la hipótesis de proporcionalidad, ya que serán dependientes del tiempo.

Función de cambio de punto

La función de cambio de punto (39), $g(t) = \begin{cases} 1 & si\ t \geq t_0 \\ 0 & si\ t < t_0 \end{cases}$, tiene unas características peculiares que a continuación explicamos. La idea es que antes y después de t_0 se obtienen ratios de riesgo constantes en el tiempo, es decir que sí cumplen la hipótesis de riesgos proporcionales.

Por ejemplo, supongamos que tenemos una variable de estado E que toma los valores 0 y 1, y que incorporamos la función de cambio de punto en un modelo con únicamente dicha variable. El modelo de Cox extendido sería:

$$h\big(t,E(t)\big)=h_0(t)\exp\Big(\beta E+\delta\big(E\times g(t)\big)\Big).$$

Cuando la función de cambio de punto se evalúa en $t\geq t_0$, tenemos el modelo:

$$h\big(t,E(t)\big)=h_0(t)\exp\left(\beta E+\delta E\right)\text{ y }HR=\exp\left(\beta+\delta\right),$$

y cuando se evalúa en $t<t_0$, tenemos:

$$h\big(t,E(t)\big)=h_0(t)\exp\left(\beta E\right)\text{ y }HR=\exp\left(\beta\right).$$

Observamos que el HR es constante en el tiempo con una discontinuidad de salto en t_0.

Podemos formular el ejemplo anterior de otro modo y obtener matemáticamente los mismos resultados. Se trata de formular el modelo de Cox extendido con dos funciones de cambio de punto y sin incluir el término independiente del tiempo en el modelo, de este modo:

Supongamos que tenemos dos funciones de cambio de punto:

$$g_1(t)=\begin{cases}1 & si\ t\geq t_0\\ 0 & si\ t<t_0\end{cases}\text{ y }g_2(t)=\begin{cases}1 & si\ t<t_0\\ 0 & si\ t\geq t_0\end{cases},$$

y que formulamos el modelo extendido siguiente:

$$h\big(t,E(t)\big)=h_0(t)\exp\Big(\delta_1\big(E\times g_1(t)\big)+\delta_2\big(E\times g_2(t)\big)\Big),$$

cuando la función de cambio de punto se evalúe en $t\geq t_0$, tendremos el modelo:

$$h\big(t,E(t)\big)=h_0(t)\exp\left(\delta_1 E\right)\text{ y }HR=\exp\left(\delta_1\right),$$

y cuando se evalúe en $t<t_0$, tendremos:

$$h\big(t,E(t)\big)=h_0(t)\exp\left(\delta_2 E\right)\text{ y }HR=\exp\left(\delta_2\right).$$

La equivalencia entre los dos modelos será que los HR serán iguales, es decir que se cumple que:

$$HR=\exp\left(\beta+\delta\right)\text{ será igual a }HR=\exp\left(\delta_1\right)\text{ y por lo tanto }\hat{\beta}+\hat{\delta}=\hat{\delta}_1;$$

$HR = \exp(\beta)$ será igual a $HR = \exp(\delta_2)$ y por lo tanto $\hat{\beta} = \hat{\delta}_2$.

Una forma de buscar el cambio de punto que mejor se adapte a los datos puede ser seleccionar aquél que maximice el logaritmo de la verosimilitud parcial para diferentes valores de t_0. También nos puede ayudar la información visual desprendida de los análisis gráficos de las curvas de supervivencia y de riesgo utilizadas en la contrastación de la hipótesis de riesgos proporcionales.

Este ejemplo se puede generalizar al caso de tener HR constantes en más de dos intervalos. Supongamos que queremos tener cuatro intervalos diferentes, con puntos de corte t_1, t_2, t_3. Y que definimos las siguientes funciones de cambio de punto:

$$g_1(t) = \begin{cases} 1 & si\ t \geq t_1 \\ 0 & si\ t < t_1 \end{cases},\quad g_2(t) = \begin{cases} 1 & si\ t \geq t_2 \\ 0 & si\ t < t_2 \end{cases}\ \text{y}\ g_3(t) = \begin{cases} 1 & si\ t \geq t_3 \\ 0 & si\ t < t_3 \end{cases},$$

el modelo de Cox extendido lo formularíamos como:

$$h(t, E(t)) = h_0(t)\exp\left(\beta E + \delta_1\left(E \times g_1(t)\right) + \delta_2\left(E \times g_2(t)\right) + \delta_3\left(E \times g_3(t)\right)\right).$$

O bien, alternativamente como hemos visto antes, definimos las siguientes funciones de cambio de punto escalonadas, las cuales nos definirán los intervalos,

$$g_1(t) = \begin{cases} 1 & si\ 0 \leq t < t_1 \\ 0 & \text{en otro caso} \end{cases},\quad g_2(t) = \begin{cases} 1 & si\ t_1 \leq t < t_2 \\ 0 & \text{en otro caso} \end{cases},$$

$$g_3(t) = \begin{cases} 1 & si\ t_2 \leq t < t_3 \\ 0 & \text{en otro caso} \end{cases}\ \text{y}\ g_4(t) = \begin{cases} 1 & si\ t \geq t_3 \\ 0 & \text{en otro caso} \end{cases},$$

entonces el modelo de Cox extendido equivalente al anterior sería:

$$h(t, E(t)) = h_0(t)\exp\left(\delta_1\left(E \times g_1(t)\right) + \delta_2\left(E \times g_2(t)\right) + \delta_3\left(E \times g_3(t)\right) + \delta_4\left(E \times g_4(t)\right)\right),$$

y los HR constantes a tramos serían:

$$HR = \begin{cases} 0 \leq t < t_1 & \exp(\delta_1) \\ t_1 \leq t < t_2 & \exp(\delta_2) \\ t_2 \leq t < t_3 & \exp(\delta_3) \\ t \geq t_3 & \exp(\delta_4) \end{cases}.$$

Bibliografía

Breslow, N. (1974). Covariance Analysis of Survival Data under the Proportional Hazards Model. *International Statistical Review*, 43, 43–54.

Cox, D. R. (1972). Regression Models and Life Tables. *Journal of the Royal Statistical Society, Series B*, 34, 187–220.

Cox, D. R. and Oakes, D. (1984). *Analysis of survival data*. Chapman and Hall.

Efron, B. (1977). The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association*, 72, 557–565.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data*. Wiley-Interscience.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis techniques for censored and truncated data*. Springer.

Kleinbaum, D. G. and Klein, M. (2012). *Survival Analysis: A Self-Learning Text, Third Edition (Statistics for Biology and Health)*. Springer.

Lee, E. T. and Wang, J. W. (2003). *Satistical Methods for suvirval data analysis*. Wiley Series.

Therneau, T. M. and Grambsch, P. M. (2000). *Modelling Survival Data: Extending the Cox model*. Springer-Verlag.

Therneau, T. M. (2024). *A Package for Survival Analysis in R*. R package version 3.7-0, https://CRAN.R-project.org/package=survival.