

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE ESTUDIOS ESTADÍSTICOS

Máster en Minería de Datos e Inteligencia de Negocios



TRABAJO DE FIN DE MÁSTER

**ANÁLISIS DE SUPERVIVENCIA DE MATRIMONIOS DEL ECUADOR EN
BASE A LOS DATOS PUBLICADOS POR EL INEC DESDE 1997 HASTA 2015.**

Álvaro Marcelo Gutiérrez Burbano

Tutor: Javier Portela

Curso académico 2016-2017

Contenido

| | |
|---|----|
| Resumen | 4 |
| 1 Introducción, Metodología y Estado del Arte | 5 |
| 1.1 Introducción | 5 |
| 1.2 Metodología | 5 |
| 1.3 Estado del Arte | 7 |
| 2 Objetivos | 7 |
| 3 Fuente de datos..... | 8 |
| 4 Aspectos legales | 9 |
| 5 Análisis de supervivencia..... | 12 |
| 5.1 Objetivo del análisis de supervivencia | 12 |
| 5.2 Datos censurados..... | 12 |
| 5.3 Distribuciones de supervivencia..... | 13 |
| 5.3.1 Función de distribución acumulativa | 13 |
| 5.3.2 Función de densidad de probabilidad | 14 |
| 5.3.3 Función de riesgo | 14 |
| 5.4 Interpretación de la función de riesgo | 16 |
| 5.5 Método de supervivencia de Kaplan-Meier | 17 |
| 5.6 Modelo de regresión de Cox | 20 |
| 5.7 El modelo de riesgos proporcionales..... | 21 |
| 6 Resultados | 23 |
| 6.1 Extracción de variables de matrimonios y divorcios..... | 23 |
| 6.2 Exploración de variables de matrimonios y divorcios | 26 |
| 6.3 Emparejamiento de matrimonios y divorcios..... | 30 |
| 6.3.1 Variables de emparejamiento | 31 |
| 6.3.2 Procedimiento de emparejamiento | 32 |
| 6.3.3 Ejemplo de emparejamiento | 33 |
| 6.3.4 Diagrama de flujo del procedimiento de emparejamiento | 34 |
| 6.4 Curvas de supervivencia de Kaplan-Meier | 37 |
| 6.5 Curvas de supervivencia de Kaplan-Meier por grupos | 39 |
| 6.5.1 Por nivel de instrucción del hombre | 39 |
| 6.5.2 Por nivel de instrucción de la mujer | 42 |
| 6.5.3 Por edad del hombre..... | 45 |

| | | |
|-------|---|----|
| 6.5.4 | Por edad de la mujer..... | 47 |
| 6.5.5 | Por las 30 mejores combinaciones de edad y nivel de instrucción..... | 50 |
| 6.5.6 | Por matrimonios con menores de edad..... | 52 |
| 6.6 | Regresión de Cox | 55 |
| 6.7 | Regresión de Cox con interacciones | 57 |
| 6.8 | Exploración con árbol de decisión | 59 |
| 6.9 | Establecimiento de perfiles de supervivencia de matrimonios..... | 64 |
| 6.9.1 | Perfil alto de supervivencia de matrimonio..... | 67 |
| 6.9.2 | Perfil medio alto de supervivencia de matrimonio..... | 67 |
| 6.9.3 | Perfil medio bajo de supervivencia de matrimonio..... | 67 |
| 6.9.4 | Perfil bajo de supervivencia de matrimonio..... | 68 |
| 6.10 | Ajuste de la curva de supervivencia por mortalidad | 68 |
| 6.11 | Análisis de sensibilidad..... | 71 |
| 7 | Cronograma..... | 73 |
| 8 | Periodo y Lugar..... | 73 |
| 9 | Conclusiones | 74 |
| 10 | Trabajo futuro..... | 77 |
| 11 | Bibliografía | 78 |
| 12 | Índice de tablas..... | 79 |
| 13 | Índice de figuras | 80 |
| 14 | Anexos..... | 81 |
| 14.1 | Código para carga de datos de divorcios desde SPSS (fragmento de R) | 81 |
| 14.2 | Código para carga de datos de matrimonios desde SPSS (fragmento de R) | 81 |
| 14.3 | Código para emparejamiento de un divorcio con un matrimonio | 81 |

Resumen

La presente memoria tiene como objetivo analizar la supervivencia de los matrimonios de Ecuador. Existen dos eventos importantes para este análisis en la vida de una pareja: el matrimonio y el divorcio. El matrimonio es un evento que se lo realiza en el Registro Civil, institución que maneja la identidad de los ciudadanos, y luego se registra los datos importantes de este matrimonio en un formulario cuyo destino es el INEC. Otro suceso importante que se espera no se realice es el divorcio, que después de realizar su registro en el Registro Civil, se genera un formulario cuyo destino también es el INEC.

La primera parte de los resultados de esta memoria (puntos 6.1 y 6.2.) tiene que ver con la extracción y exploración de los datos de matrimonios y divorcios correspondientes a los años 1997 hasta 2015. El INEC procesa estas fuentes por cada año. Por lo que se definen diferentes formatos consignando variables que en determinadas ocasiones tienen diferentes nombres, tienen diferentes significados o tienen datos incompletos. En esta parte se extrae las variables comunes de matrimonios y divorcios de todos estos años y con los mismos significados.

La segunda parte de los resultados de esta memoria (punto 6.3) tiene que ver con el emparejamiento de divorcios con sus matrimonios. Si tenemos los datos de un divorcio entonces ¿cómo saber a qué matrimonio corresponde? La respuesta pudiera ser simple, en base a los nombres o cédulas del hombre y la mujer. Para proteger la identidad de las personas en estos formularios del INEC no se registran nombres, apellidos ni número de cédula, entonces la respuesta se complica. La solución no es sencilla, debido a esto se aplicó una técnica de emparejamiento tomando en cuenta la fecha del matrimonio que está registrada en el divorcio (aunque solo se pudo usar mes y año) como primer filtro, se toma en cuenta otras variables como edad y nivel de instrucción. Con la edad del hombre y de la mujer registrada en el divorcio se calculan sus posibles edades que tuvieron al momento de contraer matrimonio. Como segundo filtro se usó estas posibles edades y el nivel de instrucción. Como es difícil determinar el nivel de instrucción que tuvieron al momento del matrimonio se consideró que el nivel de educación sea el mismo del divorcio y se dio prioridad a la edad. Con estos antecedentes se construyó una variable distancia. El emparejamiento se lo realizó en base a estos dos filtros, y dando prioridad al de menor distancia. Cuando existen varios matrimonios para el mismo divorcio se usó una variable aleatoria para tomar un solo matrimonio.

La tercera parte de los resultados de esta memoria (puntos 6.4 en adelante) se encargan de realizar un análisis en base a las curvas de supervivencia de características atareas,

niveles de educación y otras características en el Ecuador. También se establece un perfil de supervivencia de matrimonio en el Ecuador.

1 Introducción, Metodología y Estado del Arte

1.1 Introducción

En la actualidad, un factor social preocupante es la tasa de divorcios que anualmente se está incrementando tanto en España como en América Latina.

El divorcio aparece en tiempos muy antiguos, justo con el nacimiento del matrimonio y en apariencia es la solución a los problemas conyugales de la pareja. Las causales de divorcio en el Ecuador tienen que ver con adulterio, violencia, falta de armonía y abandono.

Otro factor a considerar es la tasa de matrimonios que continua a la baja conforme pasan los años.

Según los últimos datos del Anuario de Estadísticas de Matrimonios y Divorcios publicado por el Instituto Nacional de Estadística y Censos (INEC) de Ecuador, en el 2015 se registraron 25,692 divorcios en contra posición a 11,725 registrados en el 2005, un incremento de 119,1%. Mientras, el número de matrimonios cayó 8,9% al pasar de 66,612 en el año 2005 a 60,636 en el año 2015.

El anuario indicado es una recopilación de los registros de matrimonios y divorcios que se generan en todas las oficinas de la Dirección General de Registro Civil, Identificación y Cedulación del Ecuador y la Corporación de Registro Civil de Guayaquil, captados en el momento que las parejas contraen matrimonio o inscriben legalmente la sentencia de divorcio.

1.2 Metodología

Para realizar un proceso de minería de datos se requiere una metodología. Vamos a usar CRISP-DM¹ que describe los enfoques comunes en minería de datos. Como herramienta de análisis se usará lenguaje R. No se usará la metodología SEMMA que es propietaria de SAS.

El siguiente diagrama muestra la relación entre las diferentes fases de la CRISP-DM.

¹ CRISP-DM, https://es.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

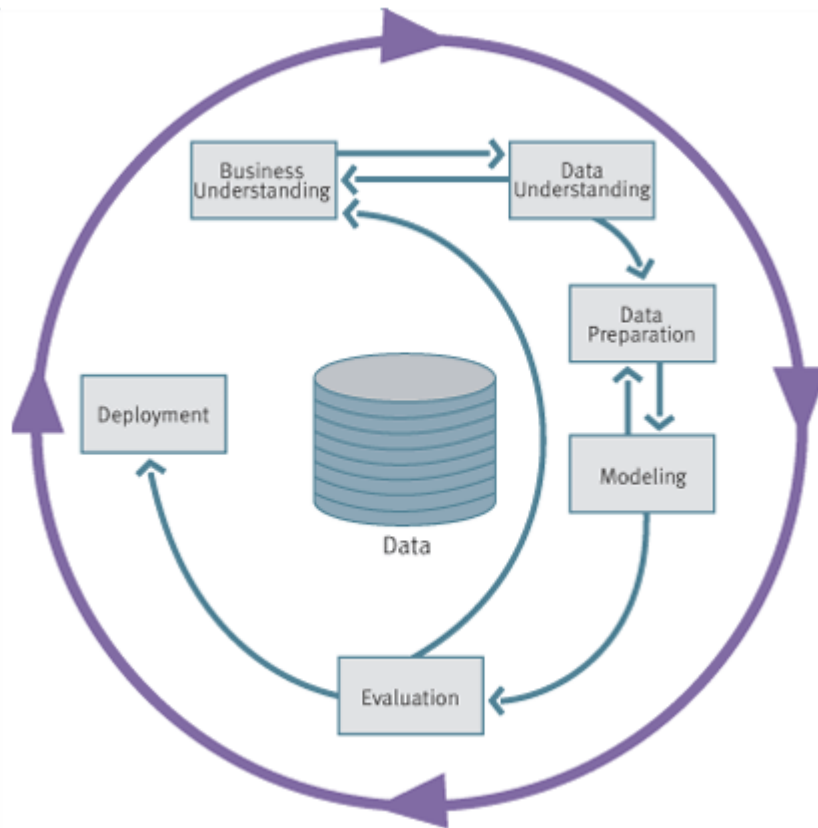


Ilustración 1-1 Diagrama de las fases CRISP-DM

Las flechas en el diagrama indican las dependencias más importantes. El círculo exterior simboliza la naturaleza cíclica del proceso de minería de datos en sí. Las lecciones aprendidas y resultados podrían provocar nuevas inquietudes de negocio (mas centradas) y los procesos posteriores se pueden modificar y ampliar.

Las seis fases principales que se desarrollaron con la metodología CRISP-DM son las siguientes:

En la fase de **comprensión del negocio** se establece los objetivos y requisitos, además se plantea una definición del problema de minería de datos. Esto se desarrolla en los puntos 2, 3 y 4.

En la fase de **comprensión de datos** se identifica las fuentes, se familiariza con los datos y se formula como trabajar con información oculta. Esto se desarrolla en el punto 6.1: Extracción de variables de matrimonios y divorcios.

En la fase de **preparación de datos** se selecciona, transforma los datos y se los prepara para el modelado. Esto se desarrolla en los puntos: 6.2 Exploración de variables de matrimonios y divorcios y 6.3 Emparejamiento de matrimonios y divorcios.

En la fase de **modelado** se seleccionan y aplican varias técnicas de modelado. Esto se desarrolla en los puntos: 6.4 Curvas de supervivencia de Kaplan-Meier, 6.8 Exploración con árbol de decisión y 6.9 Establecimiento de perfiles de supervivencia de matrimonios.

En la fase de **evaluación** se analiza el modelo obtenido. Esto se desarrolla en los puntos: 6.10 Ajuste de la curva de supervivencia por mortalidad y 6.11 Análisis de sensibilidad.

En la fase de **despliegue** se muestran los resultados. Esto se desarrolla en a lo largo de la elaboración de esta memoria.

1.3 Estado del Arte

En referencia a características sociales como lo son el matrimonio y el divorcio, no se ha encontrado referencias de análisis de supervivencia de matrimonios en el Ecuador en la web.

Un artículo interesante realiza un estudio de divorcio en USA y lleva como título: ¿Por qué han caído las tasas de divorcio?, el cual indica “Tres enfoques empíricos diferentes sugieren que el aumento de la edad de las mujeres en el matrimonio es la principal causa inmediata de la caída de las tasas de divorcio”². Las tasas de divorcio en Ecuador y en América Latina están subiendo por lo que en general no es aplicable lo analizado en USA.

2 Objetivos

El objetivo principal de esta memoria es analizar la supervivencia de los matrimonios de Ecuador según datos del INEC en el periodo desde 1997 hasta 2015 y compararlos con diferentes grupos de características atareas, niveles de educación y otras características de los matrimonios en el Ecuador. Como objetivos secundarios que se persiguen, se encuentran conocer el perfil de los matrimonios mediante un análisis descriptivo y analizar qué factores pueden influir en la supervivencia de los matrimonios y en qué medida lo hacen.

² Dana Rotz, Why Have Divorce Rates Fallen? The Role of Women’s Age at Marriage, 2015, <http://jhr.uwpress.org/content/early/2015/11/20/jhr.51.4.0214-6224R?related-urls=yes&legid=wpjhr;jhr.51.4.0214-6224Rv1>

3 Fuente de datos

El Instituto Nacional de Estadística y Censos (INEC) publica en forma anual el Anuario de Estadísticas Vitales: Matrimonios y Divorcios de la República del Ecuador. Al momento de elaborar esta memoria, el último anuario publicado es en mayo del 2016 y corresponde al periodo del 2015.

Este Anuario toma insumos de los informes estadísticos de matrimonios y divorcios que se generan en todas las oficinas de la Dirección General de Registro Civil, Identificación y Cedulación y la Corporación de Registro Civil Guayaquil, al momento que las parejas contraen matrimonio o cuando inscriben legalmente las sentencias de divorcio.

En general estos anuarios tienen información estadística de las principales características de estos hechos ocurridos y registrados en cada año, información proporcionada directamente por la Dirección General de Registro Civil, Identificación y Cedulación.

Los anuarios están disponibles en: http://www.ecuadorencifras.gob.ec/documentos/web-inec/Poblacion_y_Demografia/Matrimonios_Divorcios/Publicaciones/ y los datos están disponibles en: <http://www.ecuadorencifras.gob.ec/matrimonios-y-divorcios-bases-de-datos/>.

Según este anuario se tiene las siguientes definiciones:

Matrimonio

Es el acto, ceremonia o procedimiento por el cual se constituye la relación jurídica de marido y mujer. La legalidad de la unión de hecho puede establecerse por medios civiles, religiosos o de otra clase, reconocidos por las leyes de cada país.³

Divorcio

Es la disolución jurídica definitiva de un matrimonio, es decir la separación del marido y de la mujer que confiere a las partes el derecho a contraer nuevas nupcias según disposiciones civiles, religiosas o de otra clase de acuerdo con las leyes de cada país.

³ Anuario_Matrimonios_y_Divorcios2015.pdf

El procesamiento de la información según este anuario comprende 5 etapas:⁴

1. Crítica de la información

Consiste en la verificación de la consistencia y completitud de los datos contenidos en cada una de las secciones de los formularios estadísticos recibidos por las oficinas del Registro Civil, los mismos que son criticados por el personal del INEC, utilizando como herramienta el manual de crítica y codificación, con la finalidad de asegurar la calidad y confiabilidad de la información.

2. Codificación de la información

En esta fase se definen los códigos para cada categoría de respuesta. Para el caso de las variables de ubicación geográfica se utiliza el Clasificador Geográfico Estadístico 2015 (actualizado al 31 de diciembre de 2014).

3. Ingreso de la información

Se realiza el ingreso de la información de los formularios diligenciados a un sistema informático integrado. Una vez ingresada la información en el sistema, se procede a realizar un análisis de cobertura y validación de datos.

4. Validación de la información

La información ingresada se valida a través de sistemas informáticos, los mismos que contienen parámetros de consistencia. Al igual que en la fase de crítica se busca detectar posibles inconsistencias.

Cuando se determina que la información está incompleta o se han detectado inconsistencias, el personal de las Coordinaciones Zonales del INEC se encarga de solicitar la aclaración, a fin de completar o corregir los datos. Este proceso se lo hace vía telefónica o a través de visitas personales a las oficinas del Registro Civil.

5. Tabulación

Con la información validada y en base al plan de tabulación se construyen cuadros con información estadística a través del paquete estadístico SPSS (Statistical Package for Social Sciences). Los reportes son sometidos a análisis y revisión para su posterior publicación.

4 Aspectos legales

⁴ http://www.ecuadorencifras.gob.ec/documentos/web-inec/Poblacion_y_Demografia/Matrimonios_Divorcios/Publicaciones/

El código civil ecuatoriano indica la definición de matrimonio⁵:

“Art. 81.- Matrimonio es un contrato solemne por el cual un hombre y una mujer se unen con el fin de vivir juntos, procrear y auxiliarse mutuamente.”

Indica cuando termina un matrimonio:

“Art. 105.- El matrimonio termina:

- 1o.- Por la muerte de uno de los cónyuges;
- 2o.- Por sentencia ejecutoriada que declare la nulidad del matrimonio;
- 3o.- Por sentencia ejecutoriada que concede la posesión definitiva de los bienes del desaparecido; y,
- 4o.- Por divorcio.”

Indica que es el divorcio:

Art. 106.- El divorcio disuelve el vínculo matrimonial y deja a los cónyuges en aptitud para contraer nuevo matrimonio, salvo las limitaciones establecidas en este Código. De igual manera, no podrá contraer matrimonio, dentro del año siguiente a la fecha en que se ejecutorió la sentencia, quien fue actor en el juicio de divorcio, si el fallo se produjo en rebeldía del cónyuge demandado.

Reformas al Código Civil ecuatoriano

La Asamblea Nacional aprobó las reformas al Código Civil⁶, algunos aspectos del matrimonio cambian en el país. A partir de los 18 años los ciudadanos podrán casarse. Hasta ahora la edad mínima era 12 para las mujeres y 14 para los varones. Si un menor de 18 se casa ese compromiso será anulado. Durante los debates se revelaron que detrás de la unión de menores estaban matrimonios forzados, maltrato físico y psicológico y hasta casos de trata. Datos del Registro Civil revelan que entre el 2009 y el 2014, 6 740 niñas y adolescentes contrajeron matrimonio en el país. En el caso de los hombres, la cifra se reduce a 1 089 menores.

Las causas de divorcio cambiaron:

Art. 11.-Sustitúyase el artículo 110 por el siguiente:

“Art. 110.- Son causas de divorcio:

1. El adulterio de uno de los cónyuges.

⁵ EPN, Código Civil del Ecuador, Codigo-Civil1.pdf, <http://www.epn.edu.ec/wp-content/uploads/2015/06/Codigo-Civil1.pdf>

⁶ Ley reformativa al código civil, <http://www.oficial.ec/ley-reformativa-codigo-civil>

2. Los tratos crueles o violencia contra la mujer o miembros del núcleo familiar.
3. El estado habitual de falta de armonía de las dos voluntades en la vida matrimonial.
4. Las amenazas graves de un cónyuge contra la vida del otro.
5. La tentativa de uno de los cónyuges contra la vida del otro.
6. Los actos ejecutados por uno de los cónyuges con el fin de involucrar al otro o a los hijos en actividades ilícitas.
7. La condena ejecutoriada a pena privativa de la libertad mayor a diez años.
8. El que uno de los cónyuges sea ebrio consuetudinario o toxicómano.
9. El abandono injustificado de cualquiera de los cónyuges por más de seis meses ininterrumpidos.”

La Unión de Hecho, se convierte en estado civil, tal como los que ya existían: casado, soltero, divorciado y viudo. Para eso la unión debe ser "estable y monogámica (una sola pareja) entre dos personas mayores de edad, sin vínculo matrimonial".

Código orgánico general de procesos, COGEP

El código orgánico general de procesos COGEP⁷, regula la actividad procesal en todas las materias, excepto la constitucional, electoral y penal, con estricta observancia del debido proceso. Fue publicado en el Registro Oficial del Ecuador con Suplemento 506 el 22 de mayo de 2015 y entró en vigencia el 23 de mayo de 2016.

Las “Reformas al Código Civil ecuatoriano” entraron realmente a regir con el COGEP, el cual se innova principalmente por su principio de oralidad.

⁷ Código orgánico general de procesos, <http://www.ambiente.gob.ec/wp-content/uploads/downloads/2016/10/Codigo-Organico-General-de-Procesos.pdf>

5 Análisis de supervivencia

5.1 Objetivo del análisis de supervivencia

El objetivo del análisis de supervivencia es estudiar la distribución de los tiempos de supervivencia de un conjunto de individuos⁸. Es el único método estadístico que está más desarrollado en Medicina que en otros campos. Utiliza incluso observaciones incompletas: sujetos perdidos, cancelados o vivos al final del estudio (“censurado”).

Su supervivencia es al menos su tiempo de permanencia en el estudio. La observación de cada paciente se inicia al diagnóstico (tiempo = 0) y continua hasta la muerte o hasta que el tiempo de seguimiento se interrumpe. Cuando el tiempo de seguimiento termina antes de producirse la muerte o antes de completar el período de observación se habla de paciente “censurado”.

5.2 Datos censurados

No todos los datos de supervivencia tienen observaciones censuradas, la censura puede ocurrir en otras aplicaciones de análisis de supervivencia. Sin embargo, debido a que los datos de supervivencia censurados son tan comunes y debido a que la censura requiere un tratamiento especial, se requiere unificar los diferentes enfoques para el análisis de supervivencia.⁹

La censura viene en muchas formas y ocurre por muchas razones diferentes. La distinción más básica es entre la censura a la izquierda y la censura a la derecha. Una observación sobre una variable T es correcta censurada si todo lo que sabes sobre T es que es mayor que algún valor c . En el análisis de supervivencia, T es típicamente el momento de ocurrencia de algún evento, y los casos son censurados correctamente porque la observación termina antes de que ocurra el evento. Por lo tanto, si T es la edad de una persona al morir (en años), puede saber sólo que $T > 50$, en cuyo caso el tiempo de muerte de la persona es censurado correctamente a la edad de 50 años. Si solo sabe que los ingresos de una persona son mayores de \$ 75.000 por año, los ingresos se censuran correctamente a \$ 75.000.

Simétricamente, la censura a la izquierda ocurre cuando todo lo que sabes sobre una observación en una variable T es que es menor que un valor. Una vez más, puede aplicar esta noción a cualquier tipo de variable, no sólo un tiempo de evento. En el contexto de

⁸ Allison, Paul D. 2010. Survival Analysis Using SAS®: A Practical Guide, Second Edition. SAS Institute Inc.

⁹ Álvarez Cáceres, Rafael C. 2013. Análisis de la Supervivencia: Regresión del Cox. Ediciones Alfanova

los datos de supervivencia, la censura a la izquierda es más probable que ocurra cuando empiece a observar una muestra en un momento en que algunos de los individuos pueden haber experimentado ya el evento. Si usted está estudiando la menarquia (el inicio de la menstruación), por ejemplo, y usted comienza a seguir a las niñas a la edad de 12 años, puede encontrar que algunos de ellos ya han comenzado a menstruar. A menos que se pueda obtener información sobre la fecha de inicio de esas niñas, se dice que la edad de la menarquia se deja censurar a los 12 años. (En las ciencias sociales, la censura a la izquierda a menudo significa algo muy diferente. El tiempo de origen, no el tiempo del evento, se conoce sólo para ser menor que un cierto valor. Sin embargo, de acuerdo con las definiciones utilizadas aquí, estas observaciones están correctamente censuradas.)

Tanto en las ciencias naturales como en las sociales, la censura a la derecha es mucho más común que la censura a la izquierda, y la mayoría de los programas informáticos para el análisis de la supervivencia no permiten datos censurados a la izquierda.

5.3 Distribuciones de supervivencia

Todos los enfoques estándar para el análisis de supervivencia son probabilísticos o estocásticos. Es decir, se supone que los tiempos en que ocurren los acontecimientos son resultados de algún proceso aleatorio. Se deduce que T , el tiempo del evento para un individuo en particular, es una variable aleatoria que tiene una distribución de probabilidad. Hay muchos modelos diferentes para los datos de supervivencia, y lo que a menudo distingue un modelo de otro es la distribución de probabilidad para T . Antes de mirar estos diferentes modelos, es necesario entender tres maneras diferentes de describir las distribuciones de probabilidad.

5.3.1 Función de distribución acumulativa

Una forma que funciona para todas las variables aleatorias es la función de distribución acumulativa, o c.d.f (por sus siglas en inglés). El c.d.f. De una variable T , denotada por $F(t)$, es una función que nos dice la probabilidad de que la variable sea menor o igual que cualquier valor t que elijamos. Así,

$$F(t) = \Pr \{T \leq t\}.$$

Si conocemos el valor de F para cada valor de t , entonces sabemos todo lo que hay que saber sobre la distribución univariada de T . En el análisis de supervivencia, es más común trabajar con una función estrechamente relacionada llamada función superviviente, definida como

$$S(t) = \Pr \{T > t\} = 1 - F(t).$$

Si el evento de interés es una muerte, la función de sobreviviente da la probabilidad de sobrevivir más allá de t . Debido a que S es una probabilidad, sabemos que está delimitada por 0 y 1. Y porque T no puede ser negativa, sabemos que $S(0) = 1$. Finalmente, a medida que t se hace más grande, S nunca aumenta (y generalmente disminuye). Dentro de estas restricciones, S puede tener una gran variedad de formas.

A menudo, el objetivo es comparar funciones de supervivencia para diferentes subgrupos en una muestra. Si la función superviviente para un grupo es siempre superior a la función superviviente para otro grupo, entonces el primer grupo claramente vive más tiempo que el segundo grupo. Sin embargo, si las funciones de supervivencia se cruzan, la situación es más ambigua.

5.3.2 Función de densidad de probabilidad

Cuando las variables son continuas, otra forma común de describir sus distribuciones de probabilidad es la función de densidad de probabilidad, o p.d.f. (por sus siglas en inglés). Esta función se define como:

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}$$

Ecuación 1

Es decir, la p.d.f. es sólo la derivada o pendiente de la c.d.f. Aunque esta definición es considerablemente menos intuitiva que la del c.d.f., es la p.d.f. la que más directamente corresponde con nuestras nociones intuitivas de forma distributiva. Por ejemplo, la conocida curva en forma de campana que está asociada con la distribución normal se da por su p.d.f., no por su c.d.f.

5.3.3 Función de riesgo

Para los datos de supervivencia continua, la función de riesgo (Hazard Function), es en realidad más popular que el p.d.f como una forma de describir las distribuciones. La función de riesgo se define como:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

Ecuación 2

En lugar de $h(t)$, algunos autores denotan el riesgo por $\lambda(t)$ o $r(t)$. Debido a que la función de riesgo es tan central para el análisis de supervivencia, vale la pena tomar algún tiempo para explicar esta definición. El objetivo de la definición es cuantificar el riesgo

instantáneo de que un evento ocurra en el instante t . Debido a que el tiempo es continuo, la probabilidad de que un evento ocurra exactamente en el tiempo t es necesariamente 0. Pero podemos hablar de la probabilidad de que ocurra un evento en el pequeño intervalo entre t y $t + \Delta t$. También queremos condicionar esta probabilidad al individuo que sobrevive al tiempo t . ¿Por qué? Porque si los individuos ya han muerto (es decir, han experimentado el evento), claramente ya no están en riesgo del evento. Por lo tanto, queremos considerar sólo aquellos individuos que han llegado al comienzo del intervalo $[t, t + \Delta t)$. Estas consideraciones apuntan al numerador en la ecuación (2):

$$\Pr(t \leq T < t + \Delta t \mid T \geq t).$$

El numerador todavía no es exactamente lo que queremos, sin embargo. En primer lugar, la probabilidad es una función no decreciente de Δt : mientras más largo sea el intervalo, más probable es que ocurra un evento en ese intervalo. Para ajustarlo, dividimos por Δt , como en la ecuación (2). En segundo lugar, queremos que el riesgo de ocurrencia de eventos en exactamente el tiempo t , no en algún intervalo que comienza con t . Por lo tanto, reducimos el intervalo dejando que Δt se haga cada vez más pequeño, hasta que alcance un valor límite.

La definición de la función de riesgo en la ecuación (2) es similar a una definición alternativa de la función de densidad de probabilidad:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t)}{\Delta t}$$

Ecuación 3

La única diferencia es que la probabilidad en el numerador de la ecuación (3) es una probabilidad incondicional, mientras que la probabilidad en la ecuación (2) está condicionada a $T \geq t$. Por esta razón, la función de riesgo se describe a veces como una densidad condicional. Cuando los eventos son repetibles, la función de riesgo se refiere a menudo como la función de intensidad.

La función de supervivencia, la función de densidad de probabilidad y la función de riesgo son formas equivalentes de describir una distribución de probabilidad continua. Dado cualquiera de ellos, podemos reconstruir las otras dos. La relación entre la p.d.f. y la función superviviente se da directamente por la definición de la ecuación (1).

Otra fórmula simple expresa la función de riesgo en términos de la función p.d.f. y la función superviviente:

$$h(t) = \frac{f(t)}{S(t)}.$$

Ecuación 4

En conjunto, las ecuaciones (4) y (1) implican que

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

Ecuación 5

La integración de ambos lados de la ecuación (5) da una expresión para la función superviviente en términos de la función de riesgo:

$$S(t) = \exp\left[-\int_0^t h(u)du\right].$$

Ecuación 6

Junto con la ecuación (4), esta fórmula conduce a

$$f(t) = h(t) \exp\left[-\int_0^t h(u)du\right].$$

Ecuación 7

Estas fórmulas son extremadamente útiles en cualquier tratamiento matemático de modelos para el análisis de supervivencia porque a menudo es necesario pasar de una representación a otra.

5.4 Interpretación de la función de riesgo

Antes de seguir adelante, es necesario hacer tres aclaraciones:

- Aunque puede ser útil pensar en el riesgo como la probabilidad instantánea de un evento en el momento t , no es realmente una probabilidad porque el riesgo puede ser mayor que 1,0. Esto puede ocurrir debido a la división por Δt en la ecuación (1). Aunque el riesgo no tiene límite superior, no puede ser menor que 0.
- Porque el riesgo se define en términos de una probabilidad (que nunca se observa directamente), es en sí misma una cantidad no observada. Podemos estimar el riesgo con los datos, pero eso es sólo una estimación.
- Es muy útil pensar en el riesgo como una característica de los individuos, no de las poblaciones o de las muestras (a menos que todos en la población sean

exactamente iguales). Cada individuo puede tener una función de riesgo que es completamente diferente a la de cualquier otra persona.

5.5 Método de supervivencia de Kaplan-Meier

En biomedicina, el estimador de Kaplan-Meier (KM) es el método más utilizado para estimar las funciones de supervivencia. También conocido como el estimador del producto-límite, este método era conocido durante muchos años antes de 1958, cuando Kaplan y Meier demostraron que era, de hecho, el estimador no paramétrico de máxima verosimilitud. Esto dio al método una sólida justificación teórica.

Cuando no hay datos censurados, el estimador de KM es simple e intuitivo, la función de supervivencia $S(t)$ es la probabilidad de que un tiempo de evento sea mayor que t , donde t puede ser cualquier número no negativo. Cuando no hay censura, el estimador de KM: $\hat{S}(t)$ es sólo la proporción de observaciones en la muestra con eventos mayores que t . Por lo tanto, si el 75 por ciento de las observaciones tienen eventos mayores que 5, tenemos $\hat{S}(5) = .75$.

La situación es también muy simple en el caso de censura a la derecha (es decir, cuando todos los casos censurados son censurados al mismo tiempo c y todos los tiempos de eventos observados son menores que c). En este caso, para todo $t \leq c$, $\hat{S}(t)$ sigue siendo la proporción muestral de observaciones con eventos mayores que t . Para $t > c$, $\hat{S}(t)$ no está definido.

Las cosas se complican cuando algunos tiempos de censura son más pequeños que algunos tiempos de eventos. En ese caso, la proporción observada de casos con tiempos de evento mayores que t puede ser sesgada hacia abajo porque los casos censurados antes de t pueden, de hecho, haber "muerto" antes de t sin nuestro conocimiento. La solución es la siguiente. Supongamos que hay k tiempos de eventos distintos, $t_1 < t_2 < \dots < t_k$. En cada momento t_j , hay n_j individuos que se dice que están en riesgo de un evento. En riesgo significa que no han experimentado un evento ni han sido censurados antes del tiempo t_j . Si cualquier caso es censurado exactamente en t_j , también se considera que está en riesgo en t_j . Sea d_j el número de individuos que mueren en el tiempo t_j . El estimador de KM se define entonces como:

$$\hat{S}(t) = \prod_{j: t_j \leq t} \left(1 - \frac{d_j}{n_j} \right)$$

para $t_1 \leq t \leq t_k$.

En otras palabras, esta fórmula dice que para un tiempo dado t , toma todos los tiempos del evento que son menores que o iguales a t . Para cada uno de esos tiempos de evento, calcule la cantidad entre paréntesis, que puede interpretarse como una estimación de la probabilidad condicional de sobrevivir al tiempo $t_j + 1$, dado que uno ha sobrevivido hasta el tiempo t_j . Luego multiplique todas estas probabilidades condicionales juntas. Para t menor que t_1 (el tiempo de evento más pequeño), $\hat{S}(t)$ se define como 1,0.

Para t mayor que t_k , el tiempo de evento observado más grande, la definición de $\hat{S}(t)$ depende de la configuración de las observaciones censuradas. Cuando no hay tiempos censurados mayores que t_k , $\hat{S}(t)$ se pone a 0 para $t > t_k$. Cuando hay tiempos censurados mayores que t_k , $\hat{S}(t)$ es indefinido para t mayor que el mayor tiempo de censura.

Ejemplo con Kaplan-Meier:

Este ejemplo se basa en el estudio de 20 matrimonios¹⁰ y los resultados fueron los siguientes:

- 11 matrimonios se divorciaron a los 6, 8, 10, 11, 12, 13, 13, 22, 32, 34, 36 meses.
- 9 matrimonios permanecen al final del estudio contribuyendo 3, 7, 7, 11, 11,14, 16, 20, 20 meses de observación, sin haberse divorciado.

Con estos datos se construye la siguiente tabla:

| Columna 1 | Columna 2 | Columna 3 | Columna 4 | Columna 5 |
|----------------------------------|-----------------|--|-------------------------|-----------|
| Tiempo de supervivencia en meses | Numero de Orden | Orden de las observaciones no censuradas | $(n - r) / (n - r + 1)$ | |
| 3+ | 1 | | | |
| 6 | 2 | 2 | $18/19 = 0,95$ | 0.947368 |
| 7+ | 3 | | | |
| 7+ | 4 | | | |
| 8 | 5 | 5 | $15/16 = 0,94$ | 0.888158 |
| 10 | 6 | 6 | $14/15 = 0,93$ | 0.828947 |
| 11 | 7 | 7 | $13/14 = 0,93$ | 0.769737 |
| 11+ | 8 | | | |
| 11+ | 9 | | | |
| 12 | 10 | 10 | $10/11 = 0,91$ | 0.699761 |
| 13 | 11 | 11 | $9/10 = 0,90$ | 0.629785 |
| 13 | 12 | 12 | $8/9 = 0,89$ | 0.559809 |

¹⁰ Ejemplo basado en el Análisis de supervivencia. Autor: Pita Fernández, S. spita@canalejo.org. Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario-Universitario Juan Canalejo. A Coruña (España) Cad Aten Primaria 1995; 2: 130-135.

| | | | | |
|-----|----|----|--|----------|
| 14+ | 13 | | | |
| 16+ | 14 | | | |
| 20+ | 15 | | | |
| 20+ | 16 | | | |
| 22 | 17 | 17 | | 0.419856 |
| 32 | 18 | 18 | | 0.279904 |
| 34 | 19 | 19 | | 0.139952 |
| 36 | 20 | 20 | | 0.000000 |

Tabla 5-1 Ejemplo de construcción de una curva de supervivencia de matrimonios

Para calcular la proporción acumulativa que sobreviven hasta el tiempo t, o tasa de supervivencia acumulativa se realiza lo siguiente:

- Columna 1: Se hace una lista con todos los tiempos de supervivencia, censurada o no censurada, en orden de menor a mayor. Se coloca un signo positivo al lado de cada observación censurada. Para observaciones censuradas y no censuradas que tienen el mismo tiempo de supervivencia, se debe colocar la observación no censurada primero.
- Columna 2: Una vez ordenados de menor a mayor los datos, en esta columna se numeran las observaciones.
- Columna 3: Colocar el número de orden (rango) de las observaciones no censuradas (eventos, en este ejemplo divorcios).
- Columna 4: Calcular la proporción de pacientes que sobrevive a cada intervalo.

$$\frac{n - r}{n - r + 1}$$

donde n es el tamaño de la muestra y r el rango no censurado. Esta columna calcula la probabilidad de supervivencia para cada tiempo.

- Columna 5: Calcular el estimador de la proporción acumulativa que sobrevive. Se realiza multiplicando los valores de la columna anterior ($0,95 \cdot 0,94 = 0,89$).

De este modo, la probabilidad de seguir casado en un cierto período de tiempo (hasta el instante t) desde el principio del estudio, es el producto de la probabilidad acumulada de sobrevivir hasta el período del tiempo anterior a t, (t-1), multiplicado por la probabilidad de sobrevivir durante el intervalo (t-1; t).

La probabilidad de supervivencia puede representarse gráficamente como se muestra en la siguiente figura:

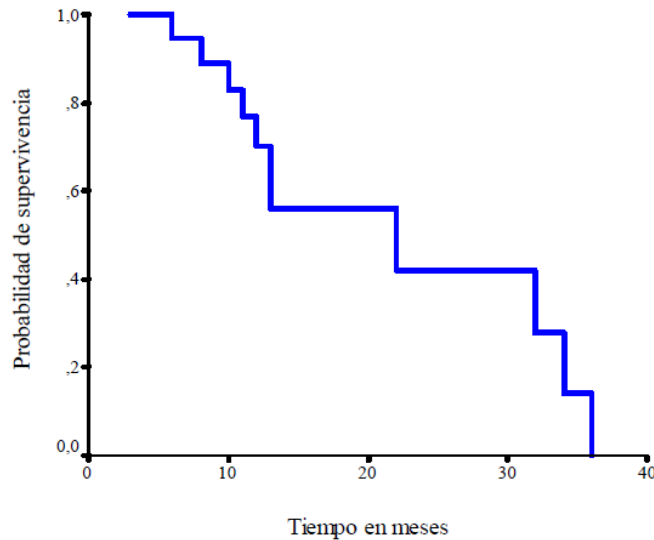


Ilustración 5-1 Ejemplo de Curva de Supervivencia de Matrimonios

5.6 Modelo de regresión de Cox

El método de regresión fue propuesto por primera vez en 1972 por el estadístico británico Sir David Cox en su famoso artículo "Regression Models and Life Tables" (Revista de la Sociedad Estadística Real, Serie B). Es difícil exagerar el impacto de este artículo. En el 2009 ISI Web of Science, este artículo fue citado más de 1.000 veces, por lo que es el artículo de revista más citados en toda la literatura de estadísticas. De hecho, Garfield (1990) informó que su recuento acumulativo de citas lo colocó entre los 100 mejores documentos en toda la ciencia. Estos números de citas indudablemente subestiman el uso real del método porque muchos autores no se molestan en citar el documento original.

¿Qué explica esta enorme popularidad? Quizás la razón más importante es que, a diferencia de los métodos paramétricos, el método de Cox no requiere que elija una distribución de probabilidad específica para representar los tiempos de supervivencia. Por eso se llama semiparamétrico. Como consecuencia, el método de Cox (a menudo denominado regresión de Cox) es considerablemente más robusto. Una segunda razón para la popularidad del papel es que la regresión de Cox hace que sea relativamente fácil incorporar covariables dependientes del tiempo (es decir, covariables que pueden cambiar de valor durante el período de observación).

Hay otras características atractivas de la regresión de Cox que son menos conocidas o apreciadas. La regresión de Cox permite un tipo de análisis estratificado que es muy

efectivo para controlar las variables molestas. Y la regresión de Cox hace que sea fácil de ajustar por períodos de tiempo en los que un individuo no está en riesgo de un evento. Finalmente, la regresión de Cox puede acomodar fácilmente la medición discreta y continua de los tiempos de evento.

A pesar de todas estas cualidades deseables, la regresión de Cox no debe considerarse como el método universal para el análisis de regresión de los datos de supervivencia. Hay momentos en que es preferible un método paramétrico.

5.7 El modelo de riesgos proporcionales

En su artículo de 1972, Cox hizo dos innovaciones significativas. En primer lugar, propuso un modelo que usualmente se conoce como el modelo de riesgos proporcionales. Sin embargo, ese nombre es engañoso, porque el modelo puede generalizarse fácilmente para permitir riesgos no proporcionales. En segundo lugar, propuso un nuevo método de estimación que más tarde se denominaría verosimilitud parcial o, más exactamente, máxima verosimilitud parcial. El término regresión de Cox se refiere a la combinación del modelo y el método de estimación. No tomó gran salto de imaginación formular el modelo de riesgos proporcionales, es una generalización relativamente directa de los modelos de Weibull y Gompertz que consideramos en el capítulo 2, "Conceptos básicos del análisis de supervivencia". Pero el método de verosimilitud parcial es algo completamente diferente. Los estadísticos tardaron años en comprender y apreciar plenamente este nuevo enfoque de la estimación.

Antes de discutir la probabilidad parcial, examinemos primero el modelo que se diseñó para estimar. Comenzaremos con el modelo básico que no incluye covariables dependientes del tiempo o riesgos no proporcionales. El modelo suele escribirse como:

$$h_i(t) = \lambda_0(t) \exp(\beta_1 x_{i1} + \dots + \beta_k x_{ik})$$

Ecuación 8

Esta ecuación dice que el riesgo para el individuo i en el tiempo t es el producto de dos factores:

- una función $\lambda_0(t)$ que no se especifica, excepto que no puede ser negativa
- una función lineal de un conjunto de k covariables fijos, que luego se eleva a un exponente.

La función $\lambda_0(t)$ puede considerarse como la función de riesgo para un individuo cuyas covariables tienen valores de 0. A menudo se denomina función de riesgo de línea de base.

Tomando el logaritmo de ambos lados, podemos reescribir el modelo como:

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

donde $\alpha(t) = \log \lambda_0(t)$. Si además especificamos $\alpha(t) = \alpha$, obtenemos el modelo exponencial. Si especificamos $\alpha(t) = \alpha t$, obtenemos el modelo de Gompertz. Finalmente, si especificamos $\alpha(t) = \alpha \log t$, tenemos el modelo de Weibull. Sin embargo, la gran atracción de la regresión de Cox es que tales elecciones son innecesarias. La función $\alpha(t)$ puede tomar cualquier forma cualquiera, incluso la de una función escalonada.

¿Por qué la ecuación (8) se llama modelo de riesgos proporcionales? Porque el riesgo para cualquier individuo es una proporción fija del riesgo para cualquier otro individuo. Para ver esto, tome la relación de los riesgos para dos individuos i y j , y aplique la ecuación (8):

$$\frac{h_i(t)}{h_j(t)} = \exp\{\beta_1(x_{i1} - x_{j1}) + \dots + \beta_k(x_{ik} - x_{jk})\}$$

Lo importante de esta ecuación es que $\lambda_0(t)$ se anula del numerador y del denominador. Como resultado, la relación de los riesgos es constante en el tiempo. Si representamos gráficamente los riesgos de registro para dos individuos, la propiedad de riesgos proporcionales implica que las funciones de riesgo deben ser estrictamente paralelas.

6 Resultados

6.1 Extracción de variables de matrimonios y divorcios

Los datos de divorcios y matrimonios se encuentran organizados en la fuente del INEC por año, ya que dependen del formulario en el que fueron recogidos. Cada año generalmente se establece un nuevo formulario. Los formularios más completos son de los años 2013, 2014 y 2015, los formularios de los años 2012 para atrás tienen datos más básicos.

La extracción de las variables se la realizó en lenguaje R desde las fuentes que se encuentran en formato SPSS.

Analizando los datos de matrimonios observamos lo siguiente:

- En los datos del 2015 existe la variable fecha de matrimonio, la cual tiene: año, mes y día del matrimonio. En cambio, en los datos del 2012 solo tenemos mes y día. Debido a esto se decidió trabajar la fecha del matrimonio solo con año y mes para todos los años.
- En los datos del 2015 tenemos las variables: fecha de nacimiento del hombre, fecha de nacimiento de la mujer, edad del hombre y edad de la mujer. En cambio, en los datos del 2012 solo tenemos las variables: edad del hombre, edad de la mujer. Debido a esto se decidió trabajar con edad del hombre y edad de la mujer.
- En los datos del 2015 existen las variables: provincia de inscripción, cantón de inscripción y parroquia de inscripción. En cambio, en los datos del 1999 solo tenemos provincia de inscripción. Debido a esto se decidió trabajar solo con provincia de inscripción.
- En los datos del 2015 existen las variables:
 - Auto identificación étnica del contrayente,
 - Sabe leer y escribir el contrayente,
 - Auto identificación étnica de la contrayente,
 - Sabe leer y escribir la contrayente,
 - Área de residencia habitual del contrayente,
 - Área de residencia habitual de la contrayente,
 - País del contrayente,

- País de la contrayente,
- Matrimonio con capitulación de bienes

En cambio, en los datos del 2001 para atrás no existen estas variables. Debido a esto se decidió no trabajar con estas variables.

Analizando los datos de divorcio observamos lo siguiente:

- En los datos del 2015 existen dos variables que son fecha de divorcio y fecha de matrimonio, las cuales tienen: año, mes y día. En cambio, en los datos del 2012 tenemos también las variables de fecha de divorcio y matrimonio, pero solo tenemos mes y día. Debido a esto se decidió trabajar la fecha del divorcio y matrimonio solo con año y mes para todos los años.
- En los datos del 2015 existen las variables: provincia de inscripción, cantón de inscripción y parroquia de inscripción. En cambio, en los datos del 1999 solo tenemos provincia de inscripción. Debido a esto se decidió trabajar solo con provincia de inscripción.
- En los datos del 2015 tenemos las variables: fecha de nacimiento del hombre, fecha de nacimiento de la mujer, edad del hombre y edad de la mujer. En cambio, en los datos del 2012 solo tenemos las variables: edad del hombre, edad de la mujer. Debido a esto se decidió trabajar con edad del hombre y edad de la mujer.
- En los datos del 2015 existe una variable importante que es: “Causas del divorcio”. Esta variable indica la causa del divorcio como por ejemplo por mutuo acuerdo, agresión y otras causas. En total son 10 causas. Lamentablemente esta variable no existe para los datos de los años 2012 para atrás.
- La variable DURACION_MATRIMONIO de los datos de divorcios se la descarta ya que no es consistente con la diferencia ANIO_DIVORCIO menos ANIO_MATRIMONIO. Al realizar el cálculo se ha encontrado diferencias de más de 3 años en varios casos. Se prefiere usar el resultado de la diferencia y se descarta la variable DURACION_MATRIMONIO original.
- Las variables: HIJOS_HOMBRE y HIJOS_MUJER se han descartado ya que no constan en los datos del matrimonio, solo constan en los datos del divorcio. Además, no se especifica si son hijos que provienen desde antes del matrimonio,

en esta pregunta del formulario del INEC generalmente los cónyuges no la contestan y la contestan sin la verdad.

- En los datos del 2015 existen las variables:
 - Nacionalidad del divorciado
 - Código del país del divorciado
 - Auto identificación étnica del divorciado
 - Sabe leer y escribir el divorciado
 - Área de residencia habitual del divorciado
 - Nacionalidad de la divorciada
 - Código del país de la divorciada
 - Auto identificación étnica de la divorciada
 - Sabe leer y escribir la divorciada
 - Área de residencia habitual de la divorciada
 - Matrimonio con capitulación de bienes

En cambio, en los datos del 2001 para atrás no existen estas variables. Debido a esto se decidió no trabajar con estas variables.

Las variables comunes para todos los años de observación que se ha llegado a determinar son las siguientes:

| MATRIMONIO | DIVORCIO |
|-----------------------|---------------------------|
| ANIO_INSCRIPCION | ANIO_INSCRIPCION |
| PROVINCIA_INSCRIPCION | PROVINCIA_INSCRIPCION |
| ANIO_MATRIMONIO | ANIO_MATRIMONIO |
| MES_MATRIMONIO | MES_MATRIMONIO |
| EDAD_HOMBRE | EDAD_HOMBRE |
| EDAD_MUJER | EDAD_MUJER |
| INSTRUCCION_HOMBRE | INSTRUCCION_HOMBRE |
| INSTRUCCION_MUJER | INSTRUCCION_MUJER |
| PARROQUIA_HOMBRE | PARROQUIA_HABITUAL_HOMBRE |
| PARROQUIA_MUJER | PARROQUIA_HABITUAL_MUJER |
| | ANIO_DIVORCIO |
| | MES_DIVORCIO |

| |
|---------------------|
| HIJOS_HOMBRE |
| HIJOS_MUJER |
| DURACION_MATRIMONIO |

Tabla 6-1 Variables de matrimonio y variables de divorcio

6.2 Exploración de variables de matrimonios y divorcios

Procedemos a realizar un análisis de las variables de matrimonios y divorcios.

Se ha encontrado lo siguiente en las variables de matrimonios:

- Al extraer los datos de matrimonios desde el año 1997 hasta el año 2015 tenemos inicialmente 1,303,073 matrimonios.
- Al analizar la variable de edad del hombre (EDAD_HOMBRE) encontramos que tiene 205 valores con 999, edad que es imposible de alcanzar. Además, existe un pico de la edad en 99 (con 1,627 valores). Esto se puede explicar debido a que en sistemas informáticos antiguos cuando se desconocía la edad se ponía 99 o 999. Con los números: 9 y 99 también había confusión ya que se los usaba para salir o regresar de un menú, ya que era la opción más alta. El gráfico siguiente muestra las frecuencias de las edades entre 80 y 101 años.

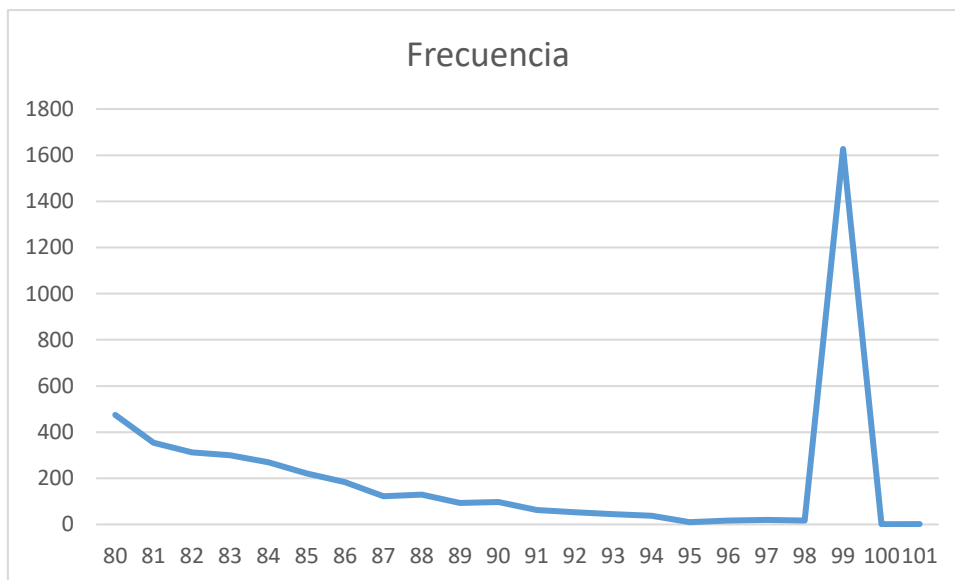


Ilustración 6-1 Frecuencia de la edad del hombre mayor que 80 en el matrimonio

Debido a esta razón se ha procedido a descartar los datos con edades mayores a 99 años. Se ha descartado 1,835 observaciones de matrimonios.

- Al analizar la variable de edad de la mujer (EDAD_MUJER) encontramos el mismo inconveniente de sistemas antiguos con la edad del hombre. Se ha descartado 1,138 observaciones de matrimonios.

- Al analizar la variable de nivel de instrucción del hombre (INSTRUCCION_HOMBRE) encontramos 32,606 observaciones con valor 9 que significa “Se ignora”. Esto se puede explicar debido a que en sistemas informáticos antiguos cuando se desconocía un valor se ponía 9 o 99 o 999. Con los números: 9 y 99 también había confusión ya que se los usaba para salir o regresar de un menú, ya que era la opción más alta.

Además, se encontraron 542 observaciones sin valor en esta variable. Para que no interfieran en la jerarquía ascendente del nivel de instrucción se decide descartar estas 33,148 observaciones.

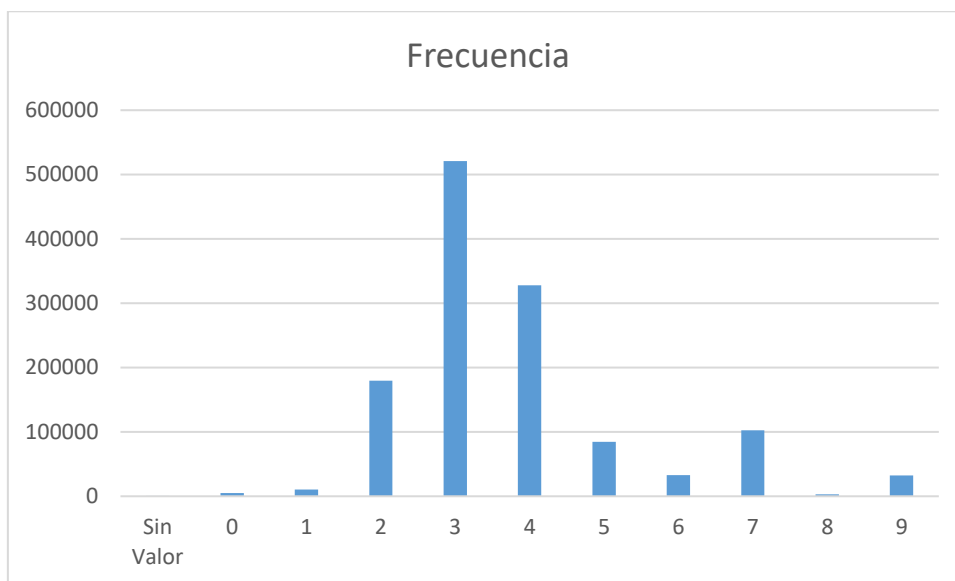


Ilustración 6-2 Frecuencia del nivel de instrucción del hombre en el matrimonio

- Al analizar la variable de nivel de instrucción de la mujer (INSTRUCCION_MUJER) encontramos el mismo inconveniente de sistemas antiguos con el nivel de instrucción del hombre. Para que no interfieran en la jerarquía ascendente del nivel de instrucción se decide descartar estas 15,835 observaciones.
- Finalmente nos quedan con 1,251,099 observaciones de matrimonios.

Se ha encontrado lo siguiente en las variables de divorcios:

- Al extraer los datos de los divorcios desde el año 1997 hasta el año 2015 tenemos inicialmente 287,903 divorcios.
- Al analizar la edad del hombre encontramos el mismo inconveniente de sistemas antiguos con la edad del hombre en el matrimonio. Se ha descartado 797 observaciones.

- Al analizar la edad de la mujer encontramos el mismo inconveniente de sistemas antiguos con la edad del hombre en el matrimonio. Se ha descartado 658 observaciones.
- Al analizar la variable nivel de instrucción del hombre encontramos el mismo inconveniente de sistemas antiguos con el nivel de instrucción del hombre en matrimonios. Para que no interfieran en la jerarquía ascendente del nivel de instrucción se decide descartar 25,749 observaciones.
- El mismo concepto anterior se aplica a 4,839 observaciones de la variable INSTRUCCION_MUJER.
- Se ha encontrado que la variable ANIO_DIVORCIO tiene 40 casos sin valor, por lo que se ha decidido descartarlos.
- Se ha encontrado que la variable MES_DIVORCIO tiene 21 casos sin valor, por lo que se ha decidido descartarlos.
- Se ha encontrado que la variable ANIO_MATRIMONIO tiene 1 caso sin valor, por lo que se ha decidido descartarlo.
- Se ha encontrado que la variable MES_MATRIMONIO tiene 20 casos sin valor, por lo que se ha decidido descartarlos.
- Se han encontrado 235 observaciones que tiene DURACION_MESES con un valor negativo. Esto lógicamente es una inconsistencia de los datos de ANIO_DIVORCIO, MES_DIVORCIO, ANIO_MATRIMONIO y MES_MATRIMONIO. Se decide descartar estos datos.
- Finalmente nos quedan con 255,543 observaciones de divorcios, se han eliminado 32,360 divorcios.

La edad se ha estratificado indicando características sociales que son: mayor o menor edad (18), tercera edad (65), edad económicamente activa (35).

| Rango de edades | | | | | | |
|------------------------|---------------------|--------------------|---------------------|------------------|---------------------|------------------|
| Valor | Categoría | Significado | Núm. Hombres | % Hombres | Núm. Mujeres | % Mujeres |
| 1 | Sin mayoría de edad | 12-17 | 22,323 | 2% | 122,424 | 10% |
| 2 | Adulto joven | 18-35 | 1,016,232 | 81% | 987,537 | 79% |
| 3 | Adulto maduro | 36-64 | 195,674 | 16% | 134,173 | 11% |
| 4 | Tercera edad | 65+ | 16,870 | 1% | 6,965 | 1% |
| | | | 1,251,099 | 100% | 1,251,099 | 100% |

Tabla 6-2 Estratificación de la edad por características sociales

El nivel de instrucción original es el siguiente:

| Nivel de instrucción original | |
|-------------------------------|------------------------------|
| Valor | Categoría |
| 0 | Ninguna |
| 1 | Centro de alfabetización |
| 2 | Primaria |
| 3 | Secundaria |
| 4 | Educación básica |
| 5 | Educación media/bachillerato |
| 6 | Ciclo posbachillerato |
| 7 | Superior |
| 8 | Posgrado |
| 9 | Se ignora |

Tabla 6-3 Nivel de Instrucción definida en el INEC

El nivel de instrucción se ha estratificado en lo siguiente:

| Rango de nivel de instrucción | | | | | | |
|-------------------------------|------------|-------------|--------------|-----------|--------------|-----------|
| Valor | Categoría | Significado | Núm. Hombres | % Hombres | Núm. Mujeres | % Mujeres |
| 1 | Ninguna | 0,1 | 15,605 | 1% | 18,604 | 1% |
| 2 | Primaria | 2,4 | 501,208 | 40% | 510,082 | 41% |
| 3 | Secundaria | 3,5,6 | 630,234 | 50% | 617,830 | 49% |
| 4 | Superior | 7,8 | 104,052 | 8% | 104,583 | 8% |
| | | | 1,251,099 | 100% | 1,251,099 | 100% |

Tabla 6-4 Estratificación del nivel de instrucción

Las variables que resultan al final son las siguientes:

| Núm. | VARIABLE | SIGNIFICADO |
|------|----------------|--|
| 1 | ID_MATRIMONIO | Identificación del matrimonio para control interno (secuencial). No proviene de la fuente de datos. |
| 2 | DIVORCIADO | 1 si es divorciado, 0 caso contrario. |
| 3 | DURACION_MESES | Numero de meses de duración del matrimonio. SI es divorciado, numero de meses desde el matrimonio hasta la fecha de divorcio. Si no es |

| | | |
|----|--------------------|--|
| | | divorciado es el número de meses desde el matrimonio hasta el 31 de diciembre del 2015 (Fin del análisis). |
| 4 | EDAD_RANGO_HOMBRE | Edad en rango del hombre |
| 5 | EDAD_RANGO_MUJER | Edad en rango de la mujer |
| 6 | NIVEL_RANGO_HOMBRE | Nivel de instrucción en rango del hombre |
| 7 | NIVEL_RANGO_MUJER | Nivel de instrucción en rango de la mujer |
| 8 | EDAD_HOMBRE | Edad del hombre |
| 9 | EDAD_MUJER | Edad de la mujer |
| 10 | INSTRUCCION_HOMBRE | Nivel de instrucción del hombre |
| 11 | INSTRUCCION_MUJER | Nivel de instrucción de la mujer |
| 12 | ANIO_MATRIMONIO | Año del matrimonio |

Tabla 6-5 Significado de las variables finales

6.3 Emparejamiento de matrimonios y divorcios

Un punto importante en este análisis es determinar si el matrimonio se ha terminado o continúa todavía.

La forma de establecer si un matrimonio continúa o no sería emparejando los datos del matrimonio con los datos del divorcio. La forma más práctica sería emparejar matrimonios y divorcios en base a la nombres, apellidos o números de cedula de identificación del hombre y de la mujer; sin embargo, estas variables por cuestiones de confidencialidad el INEC no las publica.

Al no tener estas variables en los datos de matrimonios y divorcios, necesitamos inventar una forma de emparejamiento, y se lo realizará en base a otras variables sustitutivas que pueden ayudar a emparejar un matrimonio con un divorcio. Al ser variables sustitutivas, no se garantiza un emparejamiento al ciento por ciento.

Las variables que tenemos al momento son:

| MATRIMONIO | DIVORCIO |
|-----------------------|-----------------------|
| ANIO_INSCRIPCION | ANIO_INSCRIPCION |
| PROVINCIA_INSCRIPCION | PROVINCIA_INSCRIPCION |
| ANIO_MATRIMONIO | ANIO_MATRIMONIO |
| MES_MATRIMONIO | MES_MATRIMONIO |
| EDAD_HOMBRE | EDAD_HOMBRE |

| | |
|--------------------|---------------------------|
| EDAD_MUJER | EDAD_MUJER |
| INSTRUCCION_HOMBRE | INSTRUCCION_HOMBRE |
| INSTRUCCION_MUJER | INSTRUCCION_MUJER |
| PARROQUIA_HOMBRE | PARROQUIA_HABITUAL_HOMBRE |
| PARROQUIA_MUJER | PARROQUIA_HABITUAL_MUJER |
| | ANIO_DIVORCIO |
| | MES_DIVORCIO |
| | HIJOS_HOMBRE |
| | HIJOS_MUJER |
| | DURACION_MATRIMONIO |

Tabla 6-6 Tabla de variables de matrimonio y de divorcio

6.3.1 Variables de emparejamiento

Las variables que vamos a usar para el emparejamiento son las siguientes:

- Fecha del matrimonio registrada en el divorcio.
- Fecha del divorcio.
- Duración del matrimonio, que en este caso sería la diferencia entre la fecha del divorcio y la fecha del matrimonio.
- Edad del hombre al momento del matrimonio.
- Edad del hombre al momento del divorcio. Con este dato calculamos la edad tentativa que tuvo el hombre al momento del matrimonio. Para este cálculo restamos la duración del matrimonio a la edad del hombre en el momento del divorcio.
- Edad de la mujer al momento del matrimonio.
- Edad de la mujer al momento del divorcio. Con este dato calculamos la edad tentativa que tuvo la mujer al momento del matrimonio. Para este cálculo restamos la duración del matrimonio a la edad de la mujer en el momento del divorcio.
- Nivel de instrucción del hombre y de la mujer registrados en el matrimonio.
- Nivel de instrucción del hombre y de la mujer registrados en el divorcio.

Con estos calculamos las siguientes variables

- La distancia de edad. En este caso se calcula la diferencia en valor absoluto de edades del hombre al momento del matrimonio y la edad tentativa del hombre al momento del matrimonio. A esto se suma la diferencia en valor absoluto de edades de la mujer al momento del matrimonio y la edad tentativa de la mujer al momento del matrimonio.

- La distancia del nivel de instrucción. En este caso se calcula la diferencia en valor absoluto del nivel de instrucción del hombre al momento del matrimonio y del nivel de instrucción del hombre al momento del divorcio. A esto se suma la diferencia en valor absoluto del nivel de instrucción de la mujer al momento del matrimonio y del nivel de instrucción de la mujer al momento del divorcio.

6.3.2 Procedimiento de emparejamiento

Para el emparejamiento seguimos los siguientes criterios:

- El **primer criterio** de emparejamiento es la fecha de matrimonio, que en este caso es año y mes de matrimonio. Por dar un ejemplo, hablando de enero del 1997 se debe emparejar 900 divorcios con 5,000 matrimonios.
- El **segundo criterio** de emparejamiento es el criterio “Distancia”, en el cual vamos a usar distancia de edad y distancia de nivel de instrucción, en ese orden. Se espera que la distancia de edad y del nivel de instrucción sea 0, en el caso de que no sea cero para emparejar el divorcio se tomaría el matrimonio con menor distancia de edad, y de persistir un no emparejamiento se tomaría el matrimonio con menor distancia de nivel de instrucción. En este caso se empareja la mayoría de casos, pero continúan poquísimos casos de un divorcio con dos o tres matrimonios por emparejar.
- El **tercer criterio** es usar una variable aleatoria para los poquísimos casos que no se pueden emparejar con el segundo criterio. Se enumeran los matrimonios en forma aleatoria usando una distribución uniforme (función `sample()` del lenguaje R), los ordenamos y tomamos el primer matrimonio. Con este hemos logrado emparejar el 100% de divorcios con sus matrimonios.

Inicialmente se pensó usar un procedimiento con el primer criterio descrito y el segundo criterio la distancia de hasta 1 año, en cuyo caso se cruzaba el 99% de los casos. Es decir, no se pudo emparejar todos los casos, por lo que no se usó ese procedimiento y se dio paso al procedimiento descrito de 3 criterios.

Como podemos observar en la tabla siguiente el 94% de emparejamientos se realiza con distancia 0, y el 99% de emparejamientos se realiza con distancia 0 y 1.

| Distancia Edad | Numero Emparejados | Porcentaje Emparejados |
|-------------------|-----------------------|---------------------------|
| 0 | 103,910 | 94% |

| | | |
|-----------|-------|----|
| 1 | 5,493 | 5% |
| 2 o mayor | 1,583 | 1% |

Tabla 6-7 Tabla de numero de emparejados por distancia

6.3.3 Ejemplo de emparejamiento

Para explicar de mejor manera, vamos a realizar el seguimiento de emparejamiento de un divorcio.

Supongamos que estamos analizando el primer mes de nuestro periodo de análisis, es decir estamos analizando enero del 2007. Primero filtramos los divorcios que tienen matrimonio en enero del 2007 y tomamos el primer divorcio con estas características. En este caso tenemos un divorcio compuesto por un hombre que tiene 25 años y de nivel de instrucción 2; con una mujer de 23 años y de nivel de instrucción 3, como lo indica la tabla que sigue.

Datos del divorcio a emparejar:

| | |
|--------------------------------|----|
| Edad Divorciado: | 25 |
| Edad Divorciada: | 23 |
| Instrucción Divorciado: | 2 |
| Instrucción Divorciada: | 3 |

Luego procedemos a filtrar los matrimonios correspondientes a enero del 2007. Supongamos que tenemos 10 matrimonios posibles, entonces procedemos a calcular la distancia de la edad y la distancia de la instrucción de cada uno de los 10 matrimonios con el divorcio seleccionado, además calculamos una variable aleatoria.

Matrimonios posibles:

| MATRIMONIO | | | | DISTANCIA CON DIVORCIO | | |
|-------------|------------|------------------------|------------------------|------------------------|-----------------------|-----------------|
| Edad Hombre | Edad Mujer | Instrucción Divorciado | Instrucción Divorciada | Distancia Edad | Distancia Instrucción | Valor Aleatorio |
| 29 | 34 | 2 | 3 | 15 | 0 | 4 |
| 30 | 29 | 2 | 2 | 11 | 1 | 8 |
| 26 | 24 | 2 | 3 | 2 | 0 | 5 |
| 35 | 28 | 2 | 1 | 15 | 2 | 2 |
| 24 | 22 | 1 | 2 | 2 | 2 | 10 |
| 31 | 21 | 2 | 3 | 8 | 0 | 1 |
| 18 | 19 | 3 | 3 | 11 | 1 | 7 |

| | | | | | | |
|----|----|---|---|---|---|---|
| 24 | 26 | 3 | 1 | 4 | 3 | 9 |
| 24 | 22 | 2 | 3 | 2 | 0 | 6 |
| 24 | 26 | 1 | 2 | 4 | 2 | 3 |

Matrimonios posibles ordenados por distancia y valor aleatorio:

| MATRIMONIO | | | | DISTANCIA CON DIVORCIO | | |
|----------------|---------------|---------------------------|---------------------------|------------------------|--------------------------|--------------------|
| Edad Hombre | Edad Mujer | Instrucción Divorciado | Instrucción Divorciada | Distancia Edad | Distancia Instrucción | Valor Aleatorio |
| 26 | 24 | 2 | 3 | 2 | 0 | 5 |
| 24 | 22 | 2 | 3 | 2 | 0 | 6 |
| 24 | 22 | 1 | 2 | 2 | 2 | 10 |
| 24 | 26 | 1 | 2 | 4 | 2 | 3 |
| 24 | 26 | 3 | 1 | 4 | 3 | 9 |
| 31 | 21 | 2 | 3 | 8 | 0 | 1 |
| 18 | 19 | 3 | 3 | 11 | 1 | 7 |
| 30 | 29 | 2 | 2 | 11 | 1 | 8 |
| 29 | 34 | 2 | 3 | 15 | 0 | 4 |
| 35 | 28 | 2 | 1 | 15 | 2 | 2 |

Tabla 6-8 Emparejamiento de un divorcio con un matrimonio

El matrimonio emparejado será el primero que resulte de ordenarlo en base a la distancia de edad, distancia de instrucción y orden aleatorio (color plomo en negrillas).

6.3.4 Diagrama de flujo del procedimiento de emparejamiento

A continuación, presentamos un diagrama de flujo con el procedimiento realizado para emparejar todos los divorcios desde el año 1997 hasta el año 2015.

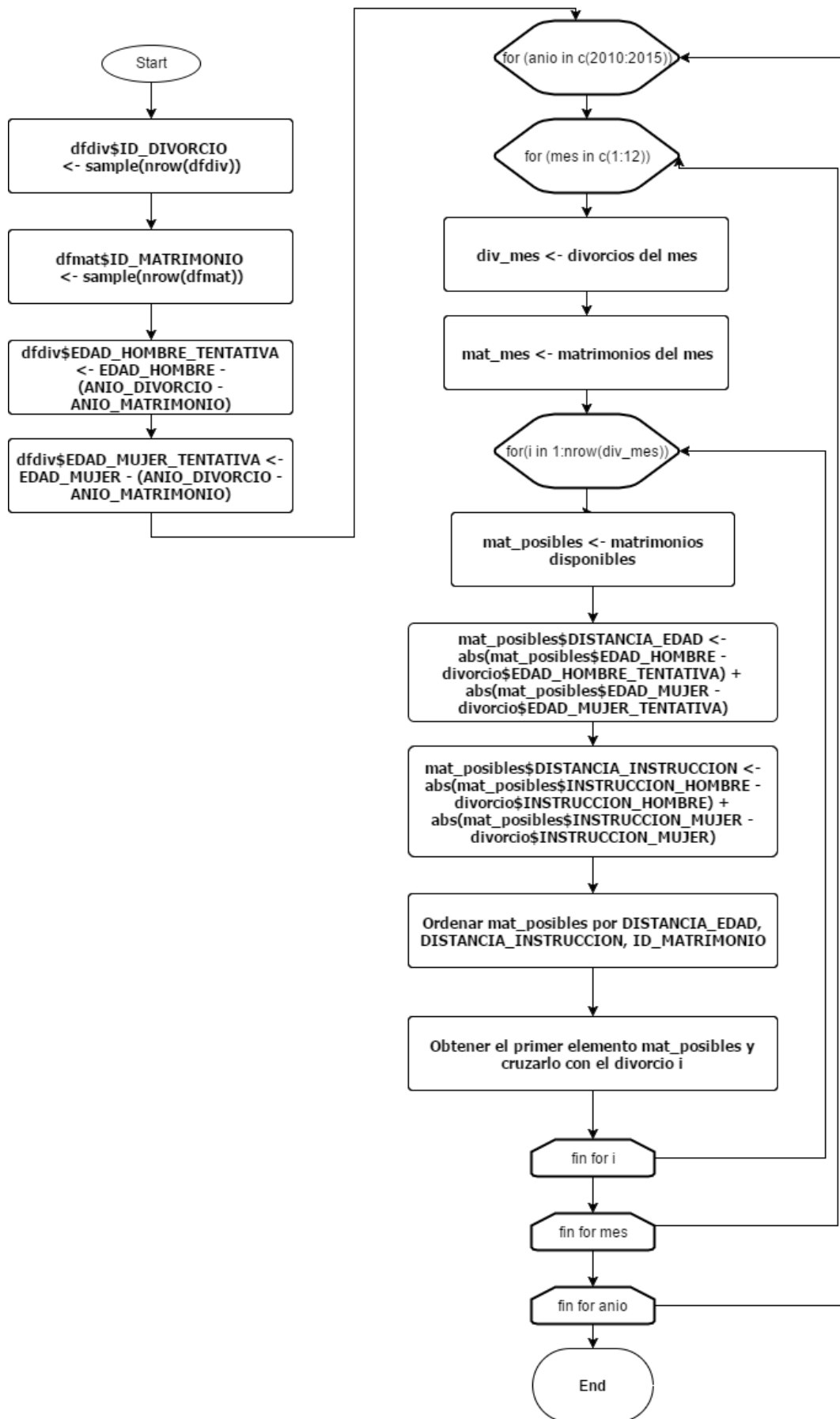


Ilustración 6-3 Diagrama de flujo del emparejamiento de divorcios con su matrimonio

Existen 255,543 observaciones de divorcios, de la cuales 110,986 corresponden a matrimonios realizados después de 1997, y 144,557 corresponden a matrimonios realizados antes de 1997.

Con este proceso se ha conseguido emparejar todos los 110,986 divorcios que corresponden a matrimonios realizados después de 1997, que representan todos los datos que poseemos actualmente.

En la siguiente grafico hemos confrontado los matrimonios y divorcios.

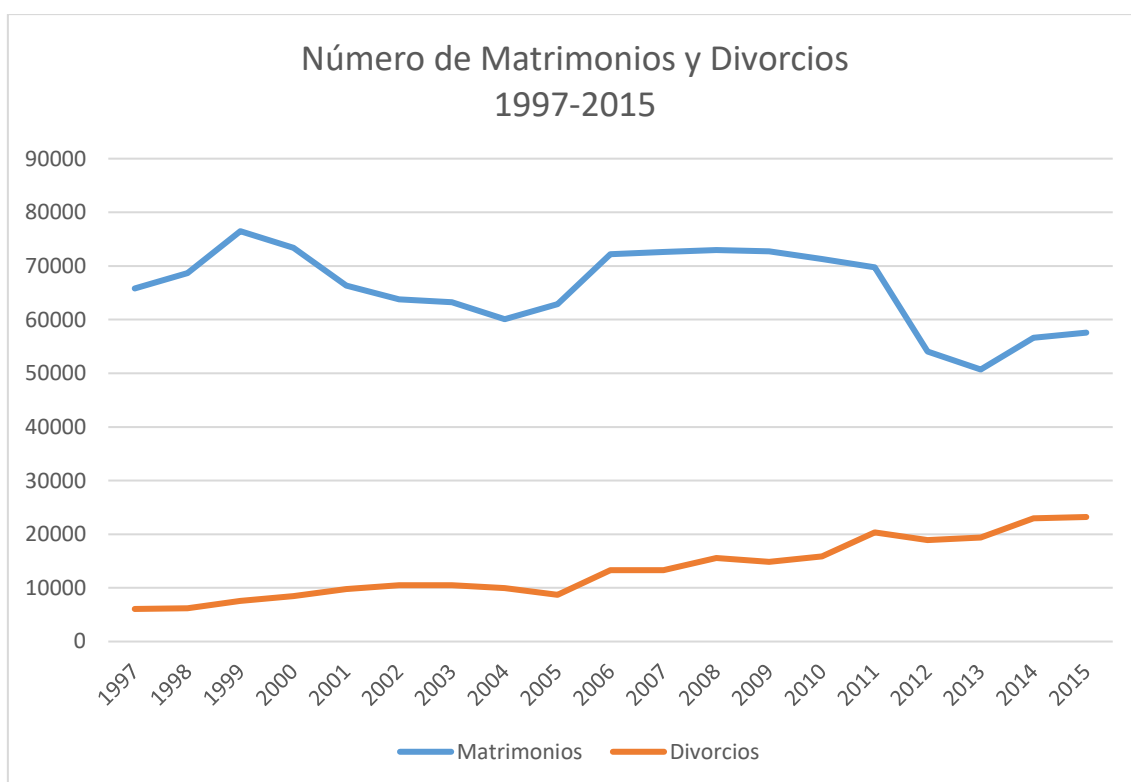


Ilustración 6-4 Número de Matrimonios y Divorcios en el Ecuador

Observamos que mientras el número de divorcios va creciendo con el pasar de los años, el número de matrimonios presenta una ligera tendencia a disminuir desde 1997 hasta el 2015. Aunque en este análisis deberíamos considerar la población que cada año va subiendo en el Ecuador.

6.4 Curvas de supervivencia de Kaplan-Meier

La línea de tiempo es de 228 meses, correspondiente a 19 años desde 1997 a 2015. A continuación, se muestra la gráfica con el código y resultado de SAS para la curva de supervivencia estimada por el método de Kaplan y Meier.

SAS

```
proc lifetest data=tfm.cruce notable outsurv=sur;  
time duracion_meses*estado(0);  
run;
```

```
The SAS System                               11:46 Friday, May 19, 2017  
  
The LIFETEST Procedure  
  
Summary of the Number of Censored and Uncensored Values  
  
Total Failed Censored Percent  
Censored  
1251099 110986 1140113 91.13
```

Ilustración 6-5 Código y Resultados SAS para estimar la función de supervivencia

El resultado de la ejecución de SAS indica el 91.13% de matrimonios censurados, es decir que en el periodo de análisis el 91.13% de parejas que se casaron continúan con su relación de matrimonio.

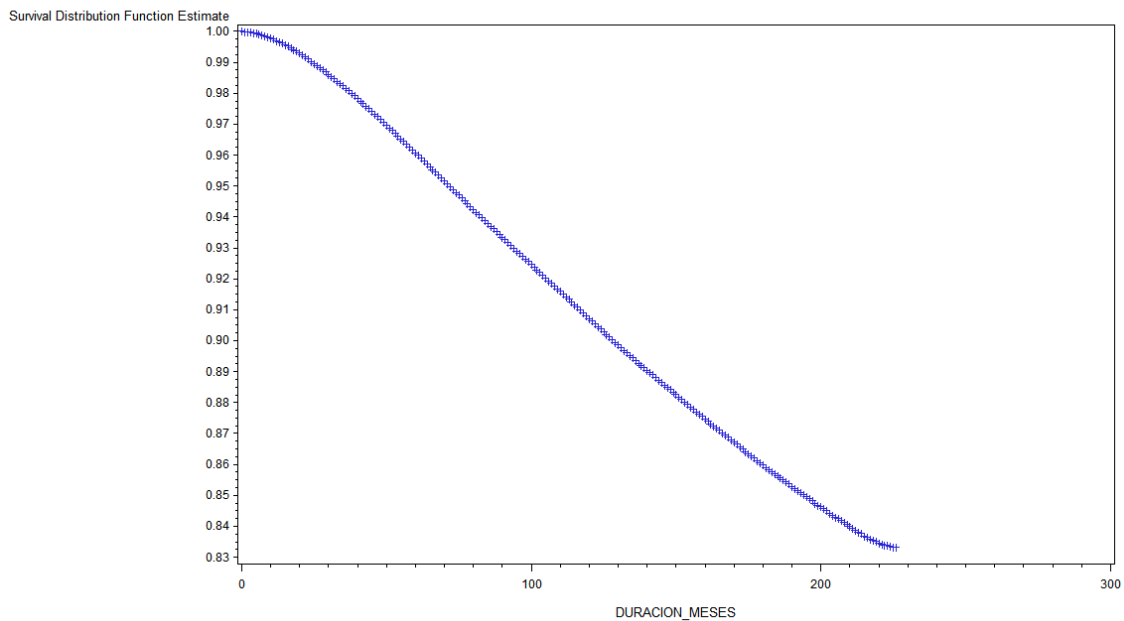


Ilustración 6-6 Curva de supervivencia para la totalidad de matrimonios realizada en SAS.

El grafico anterior que muestra la curva de supervivencia realizada en SAS indica que la supervivencia mínima se encuentra al final del periodo de observación con un valor de 83.31%. Es decir, considerando los matrimonios de 19 años (228 meses) podemos decir que el 83% sobrevive.

A continuación, se muestra la gráfica con el código y resultado de R para la curva de supervivencia estimada por el método de Kaplan y Meier.

R

```
fit <- survfit(Surv(DURACION_MESES, ESTADO==1) ~ 1, data = tabla_cruce)
```

| n | events | median | 0.95LCL | 0.95UCL |
|---------|--------|--------|---------|---------|
| 1251099 | 110986 | NA | NA | NA |

Ilustración 6-7 Código y Resultados R para estimar la función de supervivencia

La ejecución en R tiene los mismos resultados que en SAS. Además, podemos observar que la mediana (supervivencia = 0.50) no es alcanzada con esta curva de supervivencia, y es por eso que tenemos los valores NA en la gráfica anterior.

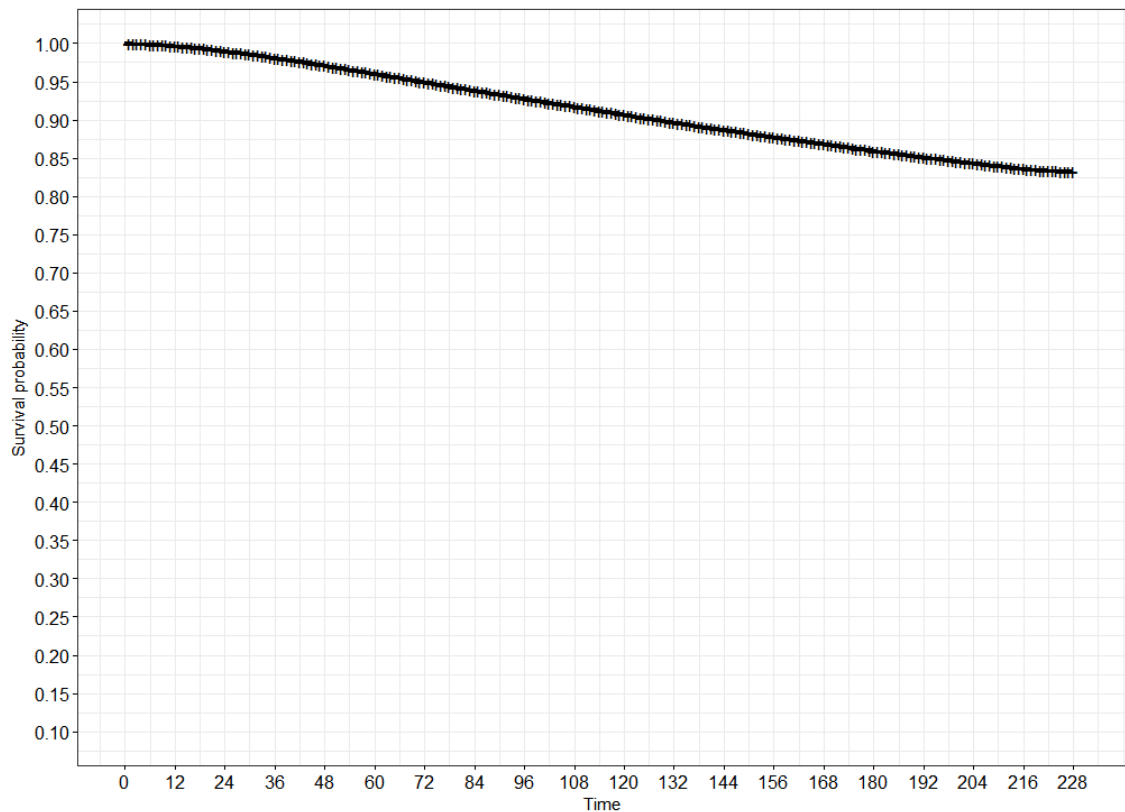


Ilustración 6-8 Curva de supervivencia para la totalidad de matrimonios realizada en R

El gráfico anterior que muestra la curva de supervivencia realizada en R es similar al obtenido en SAS, la supervivencia mínima se encuentra al final del periodo de observación con un valor de 83.31% para los matrimonios de 19 años.

Considerando los matrimonios de 5 años (60 meses) el 96.07% sobrevive. Considerando los matrimonios de 10 años (120 meses) el 90.71% sobrevive. Considerando los matrimonios de 15 años (180 meses) el 85.97% sobrevive.

En forma general se puede calcular la mediana de supervivencia trazando una línea horizontal en el grafico anterior en la supervivencia correspondiente a 0.50 (50%) y observamos en que momento cruza la curva de supervivencia y se comprueba a qué tiempo le corresponde; sin embargo, en este caso no se llega a obtener la mediana ya que los datos de supervivencia de los matrimonios pasan siempre arriba del 50%.

6.5 Curvas de supervivencia de Kaplan-Meier por grupos

Vamos a analizar la supervivencia de matrimonios por grupos de diferente naturaleza, usando la estimación de la función de supervivencia por el método de Kaplan y Meier.

6.5.1 Por nivel de instrucción del hombre

Para analizar la supervivencia por nivel de instrucción del hombre separamos en cuatro grupos correspondientes a los rangos del nivel de instrucción del hombre. El grafico de las curvas de supervivencia se muestra a continuación.

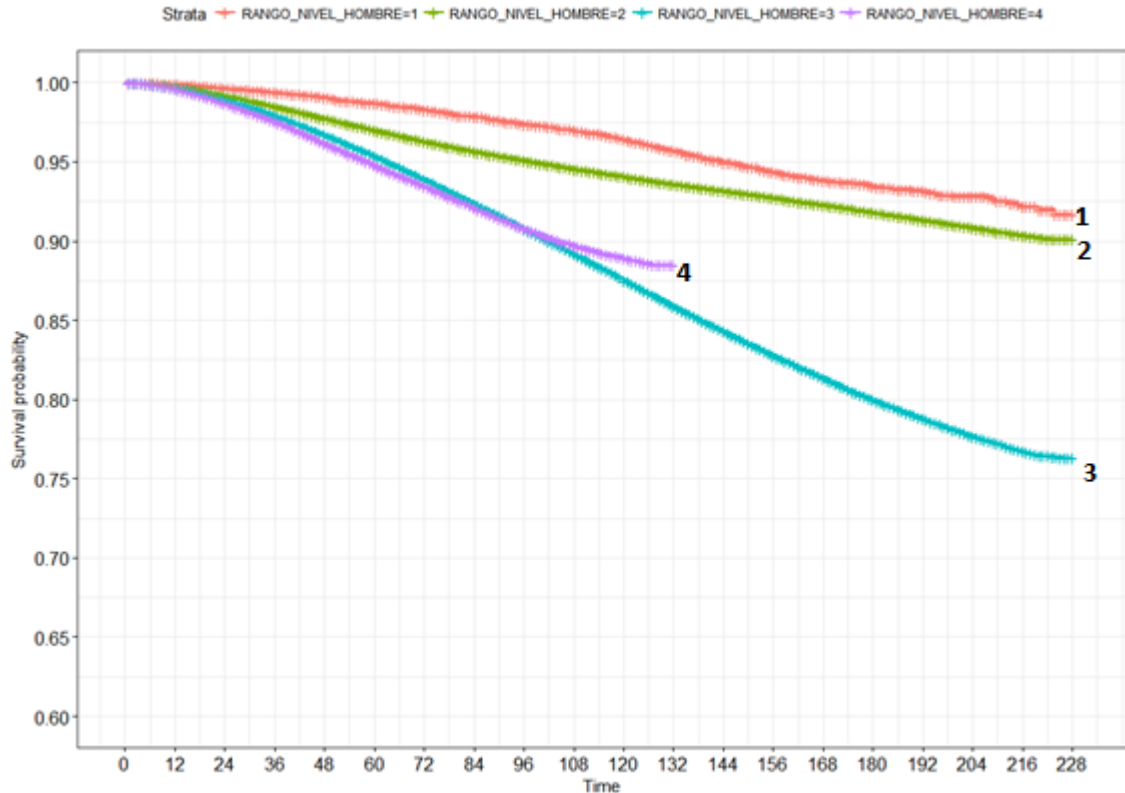


Ilustración 6-9 Curva de supervivencia de matrimonios por rangos de nivel de instrucción del hombre

La grafica puede parecer engañosa ya que muestra la escala de supervivencia desde 0.75 hasta 1.00, cuando la escala real debe ser de 0.00 a 1.00.

Observamos en el grafico anterior que la supervivencia de los matrimonios es mayor para los hombres de rango de nivel de instrucción: 1 y 2 (Ninguna y Primaria) en casi todo el intervalo de estudio, la curva de supervivencia del rango de nivel de instrucción: 3 (Secundaria) es la menor, la curva de supervivencia del rango de nivel de instrucción: 4 (Superior) desaparece a los 11 años (132 meses), posiblemente porque antes del 2004 había muy pocas personas siguiendo un master o carrera superior (2015 menos 11 igual a 2004).

También observamos que las curvas de supervivencia se cruzan entre sí, esto implica que no se cumple la condición de riesgos proporcionales. Las dos curvas más altas vamos a unir las en un solo grupo denominado: “Bajo” debido a que corresponde al nivel de instrucción: 1 y 2 (Ninguna y Primaria). Las curvas que corresponden a los niveles de instrucción: 3 y 4 (Secundaria y Superior) las uniremos en un solo grupo denominado: “Alto”. El grafico de las curvas de supervivencia con estas condiciones se muestra a continuación.

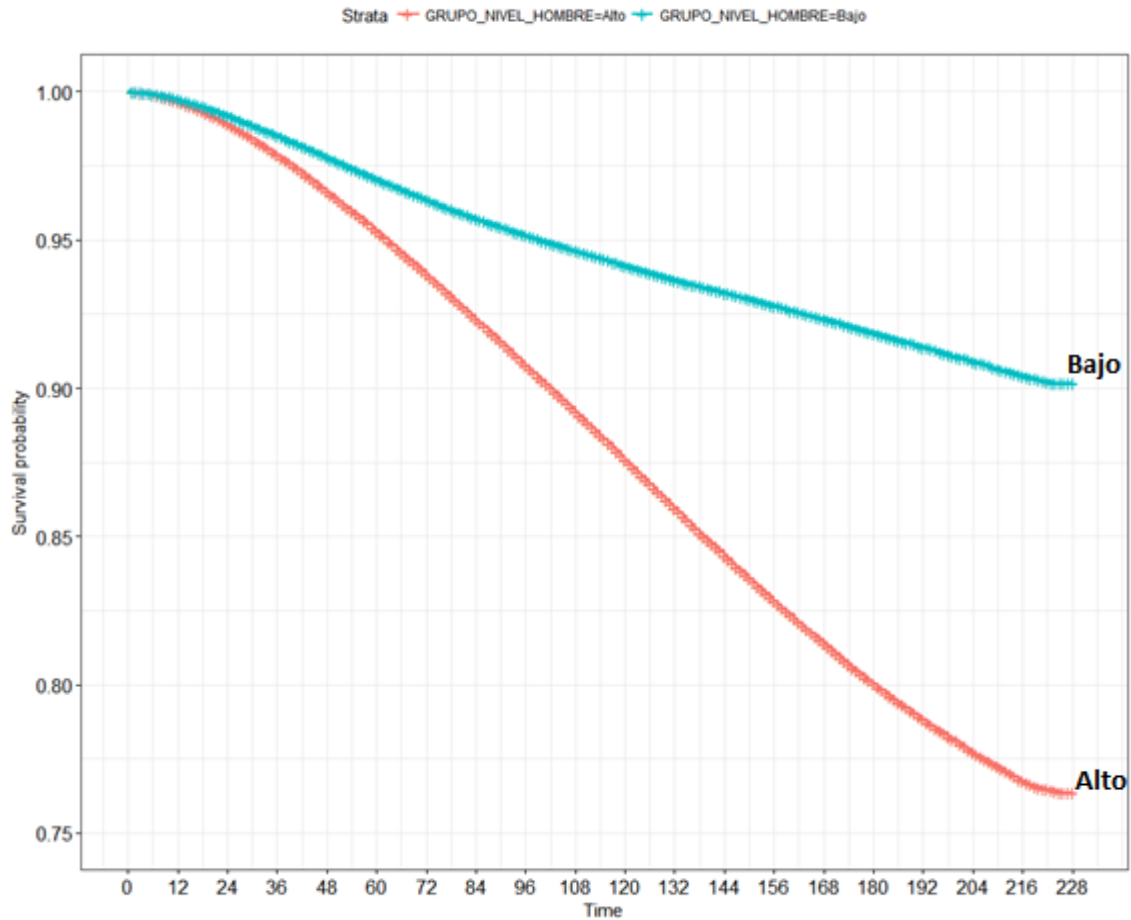


Ilustración 6-10 Curva de supervivencia de matrimonios por nivel de instrucción del hombre

Observamos en el gráfico anterior que la supervivencia de matrimonios es mayor para el nivel de instrucción del hombre: “Bajo” que corresponde al nivel de instrucción: 1 y 2 (Ninguna y Primaria). Adicionalmente observamos en el gráfico anterior que las curvas de supervivencia no se cruzan entre sí, es decir se cumple la condición de riesgos proporcionales. Para comprobarlo vamos a usar el modelo de regresión de Cox.

```
coxph(formula = surv(DURACION_MESES, ESTADO == 1) ~ GRUPO_NIVEL_HOMBRE,
      data = tabla_cruce)

n= 1251099, number of events= 110986

              coef exp(coef) se(coef)      z Pr(>|z|)
GRUPO_NIVEL_HOMBREBajo -0.856554  0.424623  0.006644 -128.9 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
GRUPO_NIVEL_HOMBREBajo  0.4246      2.355   0.4191   0.4302

Concordance= 0.586 (se = 0.001 )
Rsquare= 0.014 (max possible= 0.909 )
Likelihood ratio test= 18256 on 1 df,  p=0
Wald test               = 16618 on 1 df,  p=0
Score (logrank) test = 17635 on 1 df,  p=0
```

Ilustración 6-11 Resultados de Cox por nivel de instrucción del hombre

Observando el grafico anterior tenemos que:

- El modelo es significativo usando cualquiera de los tres criterios: test de razón de verosimilitud, test de Wald y test logrank.
- El nivel de instrucción del hombre: “Bajo” que corresponde al nivel de instrucción: 1 y 2 (Ninguna y Primaria) tiene un p-valor menor a 0.001 (***), que a su vez es menor que 0.05, lo que implica que esta característica es significativa a la hora de analizar la supervivencia de un matrimonio.
- El coeficiente es negativo (-0.856554) lo que indica que el nivel bajo de instrucción del hombre disminuye el riesgo de divorcio en 0.4246 veces. Un matrimonio con nivel bajo de instrucción del hombre tiene 0.4246 veces menos riesgo de divorcio que un matrimonio que no tenga el nivel bajo de instrucción del hombre. Si el riesgo de divorcio disminuye la supervivencia del matrimonio aumenta.

Analizando las curvas de supervivencia y el resultado de Cox por nivel de instrucción del hombre, podemos concluir que la supervivencia de un matrimonio es mayor para niveles de educación bajo del hombre.

6.5.2 Por nivel de instrucción de la mujer

Para analizar la supervivencia por nivel de instrucción de la mujer separamos en cuatro grupos correspondientes a los rangos del nivel de instrucción de la mujer. El grafico de las curvas de supervivencia se muestra a continuación.

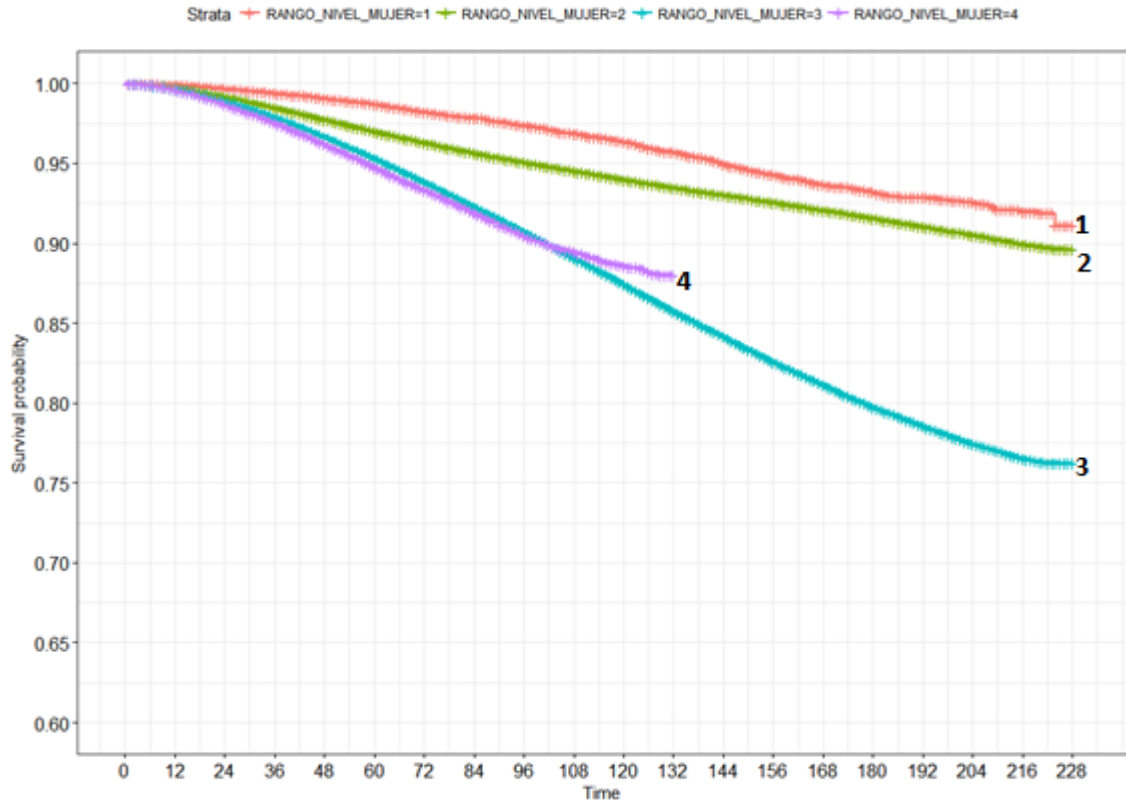


Ilustración 6-12 Curva de supervivencia de matrimonios por rangos de nivel de instrucción de la mujer

Observamos en el gráfico anterior que la supervivencia de los matrimonios es mayor para las mujeres de rango de nivel de instrucción: 1 y 2 (Ninguna y Primaria) en casi todo el intervalo de estudio, la curva de supervivencia del rango de nivel de instrucción: 3 (Secundaria) es la menor, la curva de supervivencia del rango de nivel de instrucción: 4 (Superior) desaparece a los 11 años (132 meses), posiblemente porque antes del 2004 había muy pocas personas siguiendo un master o carrera superior (2015 menos 11 igual a 2004).

También observamos que las curvas de supervivencia se cruzan entre sí, esto implica que no se cumple la condición de riesgos proporcionales. Las dos curvas más altas vamos a unir las en un solo grupo denominado: “Bajo” debido a que corresponde al nivel de instrucción: 1 y 2 (Ninguna y Primaria). Las curvas que corresponden a los niveles de instrucción: 3 y 4 (Secundaria y Superior) las uniremos en un solo grupo denominado: “Alto”. El gráfico de las curvas de supervivencia con estas condiciones se muestra a continuación.

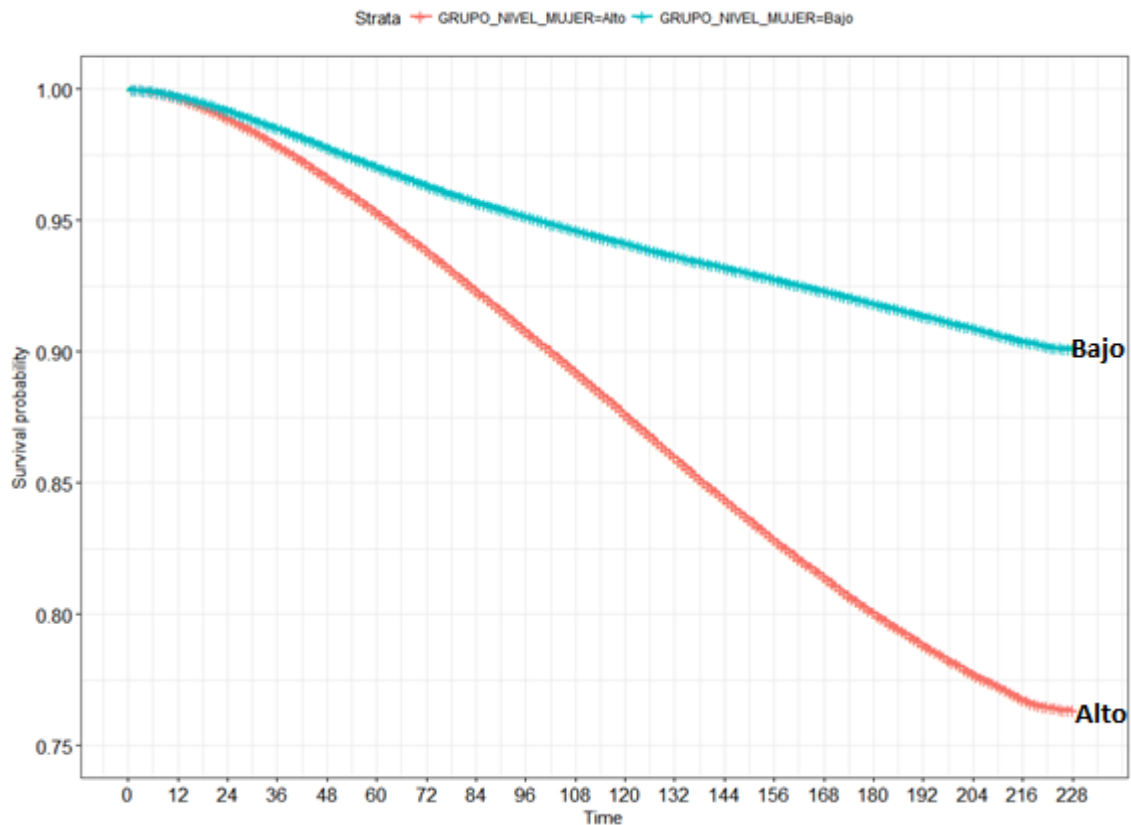


Ilustración 6-13 Curva de supervivencia de matrimonios por nivel de instrucción de la mujer

Observamos en el grafico anterior que la supervivencia de matrimonios es mayor para el nivel de instrucción de la mujer: “Bajo” que corresponde al nivel de instrucción: 1 y 2 (Ninguna y Primaria). Adicionalmente observamos que en el grafico anterior las curvas de supervivencia no se cruzan entre sí, es decir se cumple la condición de riesgos proporcionales. Para comprobarlo vamos a usar el modelo de regresión de Cox.

```
coxph(formula = Surv(DURACION_MESES, ESTADO == 1) ~ GRUPO_NIVEL_MUJER,
      data = tabla_cruce)

n= 1251099, number of events= 110986

              coef exp(coef) se(coef)      z Pr(>|z|)
GRUPO_NIVEL_MUJERBajo -0.852862  0.426193  0.006645 -128.3  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
GRUPO_NIVEL_MUJERBajo  0.4262      2.346  0.4207  0.4318

Concordance= 0.585 (se = 0.001 )
Rsquare= 0.014 (max possible= 0.909 )
Likelihood ratio test= 18086 on 1 df,  p=0
Wald test               = 16470 on 1 df,  p=0
Score (logrank) test = 17468 on 1 df,  p=0
```

Ilustración 6-14 Resultados de Cox por nivel de instrucción de la mujer

Observando el grafico anterior tenemos que:

- El modelo es significativo usando cualquiera de los tres criterios: test de razón de verosimilitud, test de Wald y test logrank.
- El nivel de instrucción de la mujer: “Bajo” que corresponde al nivel de instrucción: 1 y 2 (Ninguna y Primaria) tiene un p-valor menor a 0.001 (***), que a su vez es menor que 0.05, lo que implica que esta característica es significativa a la hora de analizar la supervivencia de un matrimonio.
- El coeficiente es negativo (-0.852862) lo que indica que el nivel bajo de instrucción de la mujer disminuye el riesgo de divorcio en 0.4262 veces. Un matrimonio con nivel bajo de instrucción de la mujer tiene 0.4262 veces menos riesgo de divorcio que un matrimonio que no tenga el nivel bajo de instrucción de la mujer. Si el riesgo de divorcio disminuye la supervivencia del matrimonio aumenta.

Analizando las curvas de supervivencia y el resultado de Cox por nivel de instrucción de la mujer, podemos concluir que la supervivencia de un matrimonio es mayor para niveles de educación bajo de la mujer.

6.5.3 Por edad del hombre

Para analizar la supervivencia por edad del hombre separamos en cuatro grupos correspondientes a los rangos de edad del hombre. El grafico de las curvas de supervivencia se muestra a continuación.

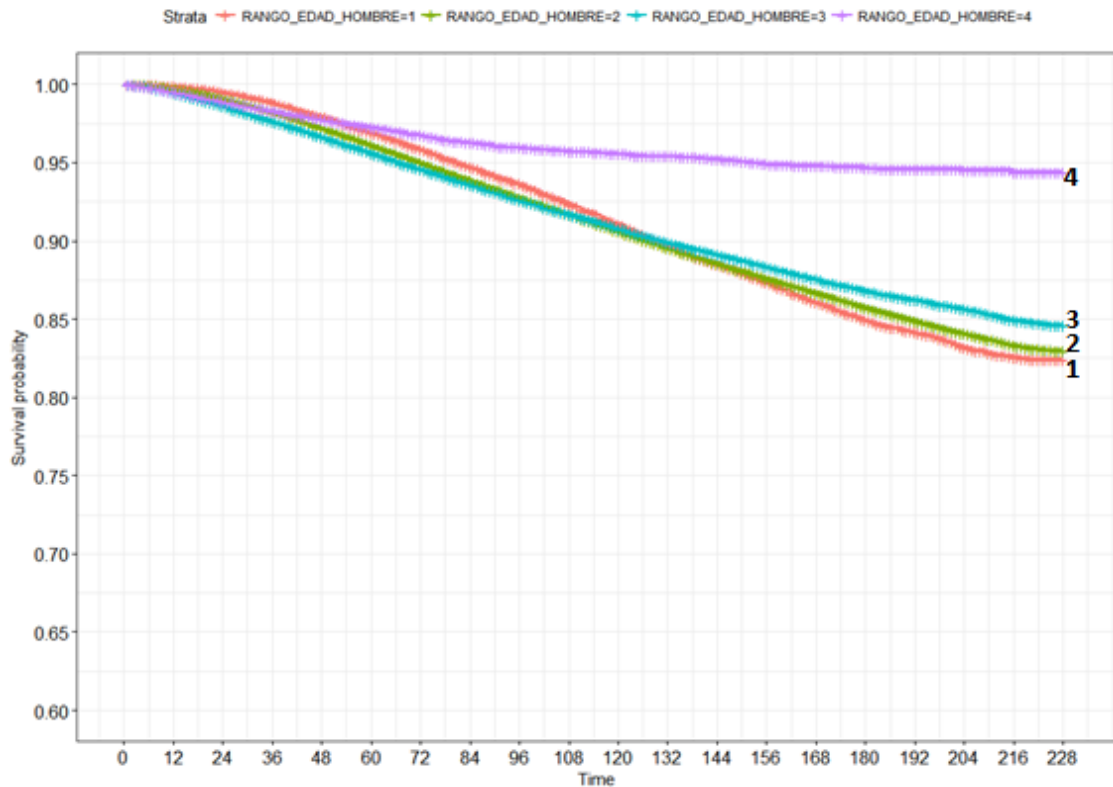


Ilustración 6-15 Curva de supervivencia de matrimonios por rangos de edad del hombre.

Observamos en el gráfico anterior que la supervivencia de los matrimonios es mayor para los hombres de rango de edad: 4 (mayores de 65 años), en casi todo el intervalo de estudio, los otros rangos de edad del hombre tienen curvas de supervivencia equivalentes.

También observamos que las curvas de supervivencia se cruzan entre sí, esto implica que no se cumple la condición de riesgos proporcionales. Las tres curvas más bajas vamos a unir las en un solo grupo denominado: “Bajo” debido a que corresponde al rango de edad: 1, 2 y 3 (12 a 17, 18 a 35 y 36 a 64) años construyendo un grupo de edades entre 12 y 64 años. La curva más alta que se diferencia notablemente de las otras 3 la vamos a dejar solo grupo denominado: “Alto” debido a que corresponde a los mayores de 65 años. El gráfico de las curvas de supervivencia con estas condiciones se muestra a continuación.

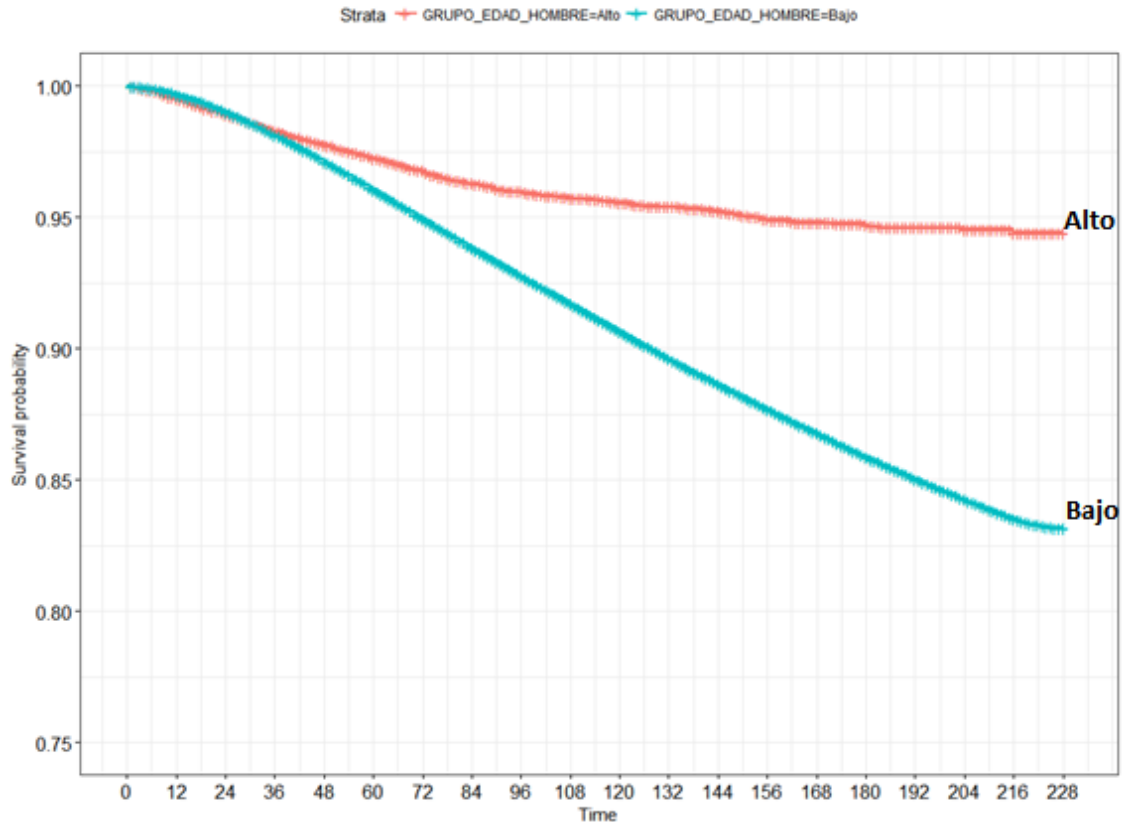


Ilustración 6-16 Curva de supervivencia de matrimonios por edad del hombre.

Observamos en el gráfico anterior que, la supervivencia de los matrimonios es mayor para el grupo de edad del hombre: “Alto” que corresponde a los mayores de 65 años. Adicionalmente observamos que en el gráfico anterior las curvas de supervivencia se continúan cruzando entre sí (desde los 6 a los 24 meses) a pesar de haber hecho dos grupos, es decir no se cumple la condición de riesgos proporcionales.

En estas condiciones no se puede obtener todavía ninguna conclusión de la edad del hombre con respecto a la supervivencia del matrimonio.

6.5.4 Por edad de la mujer

Para analizar la supervivencia por edad de la mujer separamos en cuatro grupos correspondientes a los rangos de edad de la mujer. El gráfico de las curvas de supervivencia se muestra a continuación.

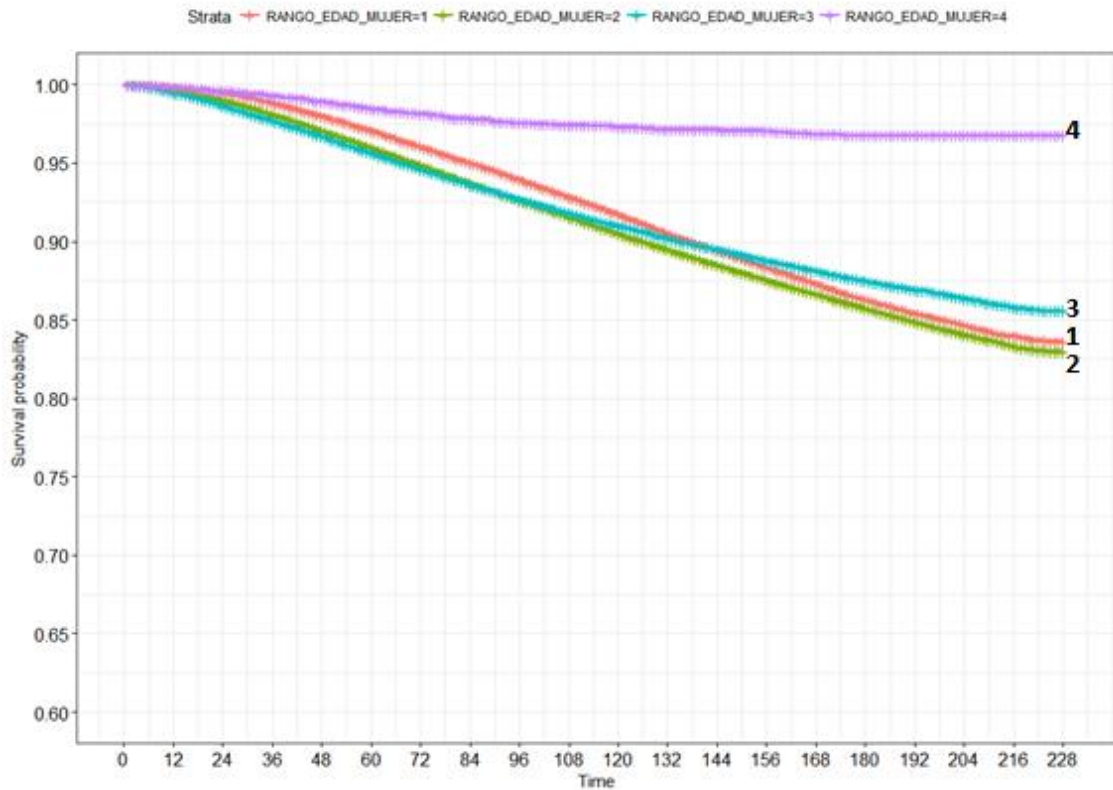


Ilustración 6-17 Curva de supervivencia de matrimonios por rangos de edad de la mujer

Observamos en el gráfico anterior que la supervivencia de los matrimonios es mayor para las mujeres de rango de edad: 4 (mayores de 65 años) en casi todo el intervalo de estudio, los otros rangos de edad de la mujer tienen curvas de supervivencia equivalentes.

También observamos que las curvas de supervivencia se cruzan entre sí, esto implica que no se cumple la condición de riesgos proporcionales. Las tres curvas más bajas vamos a unir las en un solo grupo denominado: “Bajo” debido a que corresponde al rango de edad: 1, 2 y 3 (12 a 17, 18 a 35 y 36 a 64) años construyendo un grupo de edades entre 12 y 64 años. La curva más alta que se diferencia notablemente de las otras 3 la vamos a dejar solo grupo denominado: “Alto” debido a que corresponde a las mayores de 65 años. El gráfico de las curvas de supervivencia con estas condiciones se muestra a continuación

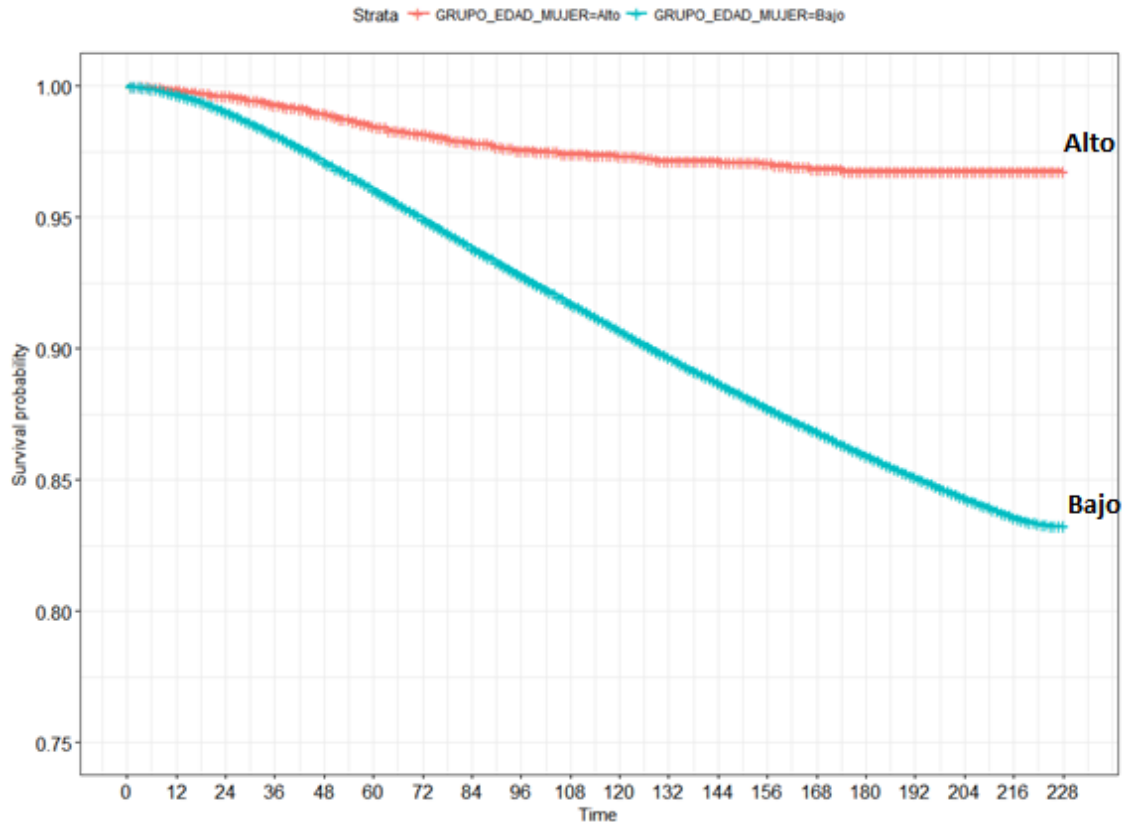


Ilustración 6-18 Curva de supervivencia de matrimonios por edad de la mujer

Observamos en el gráfico anterior que, la supervivencia de matrimonios es mayor por el grupo de edad de la mujer: “Alto” que corresponde a las mayores de 65 años. Adicionalmente observamos que en el gráfico anterior las curvas de supervivencia no se cruzan entre sí, es decir se cumple la condición de riesgos proporcionales. Para comprobarlo vamos a usar el modelo de regresión de Cox

```
coxph(formula = Surv(DURACION_MESES, ESTADO == 1) ~ GRUPO_EDAD_MUJER,
      data = tabla_cruce)

n= 1251099, number of events= 110986

              coef exp(coef) se(coef)      z Pr(>|z|)
GRUPO_EDAD_MUJERBajo 1.35904  3.89245  0.08457 16.07 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
GRUPO_EDAD_MUJERBajo  3.892    0.2569    3.298    4.594

Concordance= 0.502 (se = 0 )
Rsquare= 0 (max possible= 0.909 )
Likelihood ratio test= 427.9 on 1 df,  p=0
Wald test               = 258.2 on 1 df,  p=0
Score (logrank) test = 300.5 on 1 df,  p=0
```

Ilustración 6-19 Resultado de Cox por edad de la mujer

Observando el gráfico anterior tenemos que:

- El modelo es significativo usando cualquiera de los tres criterios: test de razón de verosimilitud, test de Wald y test logrank.
- La edad de la mujer: “Bajo” que corresponde a las edades entre 12 y 64 tiene un p-valor menor a 0.001 (***), que a su vez es menor que 0.05, lo que implica que esta característica es significativa a la hora de analizar la supervivencia de un matrimonio.
- El coeficiente es positivo (1.35904) lo que indica que cuando la mujer tiene una edad entre 12 y 64 años aumenta el riesgo de divorcio en 3.89245 veces. Un matrimonio con una mujer entre 12 y 64 años tiene 3.89245 veces más riesgo de divorcio que un matrimonio que tenga una mujer mayor de 65 años. Si el riesgo de divorcio aumenta la supervivencia del matrimonio disminuye.

Analizando las curvas de supervivencia y el resultado de Cox por edad de la mujer, podemos concluir que la supervivencia de un matrimonio es menor para mujeres entre 12 y 64 años; y la supervivencia de un matrimonio es mayor para mujeres mayores de 65 años.

6.5.5 Por las 30 mejores combinaciones de edad y nivel de instrucción

Para analizar la supervivencia vamos a separar en grupos correspondientes a las combinaciones entre edad del hombre, edad de la mujer, nivel de instrucción del hombre y nivel de instrucción de la mujer. De estos grupos vamos a tomar los 30 grupos mayoritarios con respecto a la cantidad de matrimonios.

Vamos a usar una notación abreviada para identificar a cada grupo. Por ejemplo: “EH2EM1NH3NM4” significa:

- Rango de edad del hombre igual a 2
- Rango de edad de la mujer igual a 1
- Rango de nivel de instrucción del hombre igual a 3
- Rango de nivel de instrucción de la mujer igual a 4

El gráfico de las curvas de supervivencia se muestra en la figura siguiente.

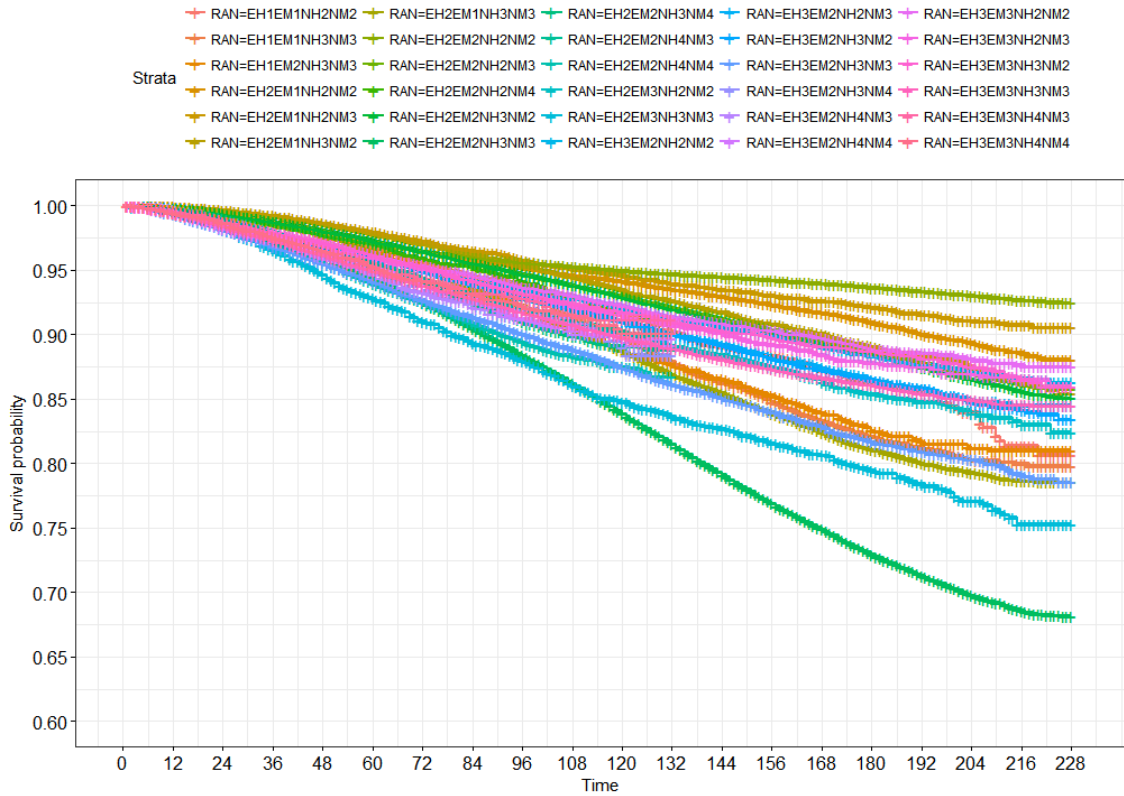


Ilustración 6-20 Función de supervivencia de los 30 grupos mayoritarios

Observamos en el grafico anterior que, la mayor curva de supervivencia comienza a diferenciarse a partir de los 10 años (120 meses). La menor curva se diferencia en forma considerable de las demás a partir de los 10 años también.

Al tener 30 curvas es difícil de apreciar cual es mayor o menor, debido a esto, se analiza los datos de las curvas y se muestra un gráfico solo con la mayor y menor curva.

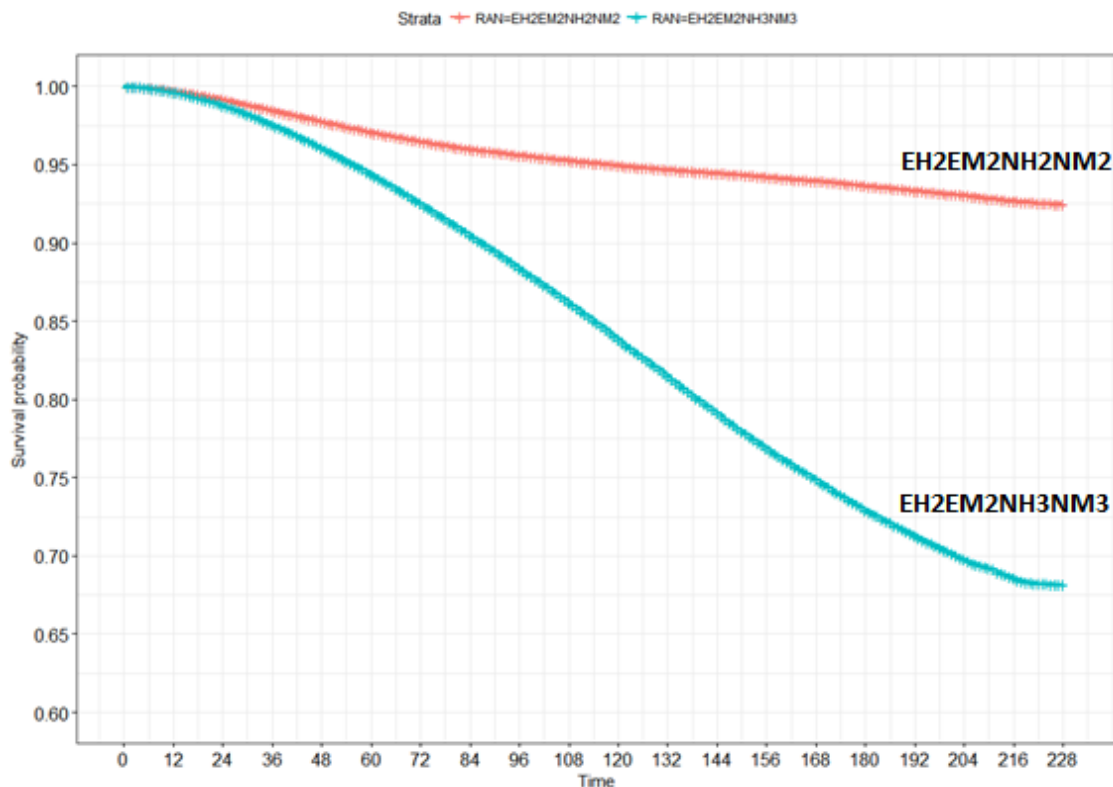


Ilustración 6-21 Función de supervivencia del mejor y peor grupo mayoritario

Observamos en el gráfico anterior que, la supervivencia de los matrimonios a partir de los 10 años es mayor para matrimonios con edad del hombre y de la mujer entre 18 y 35 años y nivel de educación primaria tanto del hombre y la mujer, la curva de supervivencia del grupo es menor para matrimonios con edad del hombre y de la mujer entre 18 y 35 años y nivel de educación secundaria tanto del hombre y la mujer.

6.5.6 Por matrimonios con menores de edad.

Se han detectado casos de matrimonios forzados, con maltrato físico y psicológico y hasta casos de trata de personas. Se ha asociado estos problemas con la edad del hombre y la mujer en el matrimonio. Datos del Registro Civil revelan que entre el 2009 y el 2014, 6 740 niñas y adolescentes contrajeron matrimonio en el país. En el caso de los hombres, la cifra se reduce a 1 089 menores. La edad mínima era 12 para las mujeres y 14 para los varones.

Debido a las razones anteriores, La Asamblea Nacional del Ecuador aprobó las reformas al Código Civil¹¹, indicando que la edad mínima para casarse es de 18 años tanto para hombres como para mujeres.

Vamos a realizar un estudio especial, separando los matrimonios en cuatro grupos:

- $H < M <$, indica matrimonios de hombres menores de 18 años con mujeres menores de 18 años.
- $H > M <$, indica matrimonios de hombres mayores de 18 años con mujeres menores de 18 años.
- $H < M >$, indica matrimonios de hombres menores de 18 años con mujeres mayores de 18 años.
- $H > M >$, indica matrimonios de hombres mayores de 18 años con mujeres mayores de 18 años.

Para analizar la supervivencia de los matrimonios en este caso separamos en los cuatro grupos indicados. El gráfico se muestra en la figura siguiente.

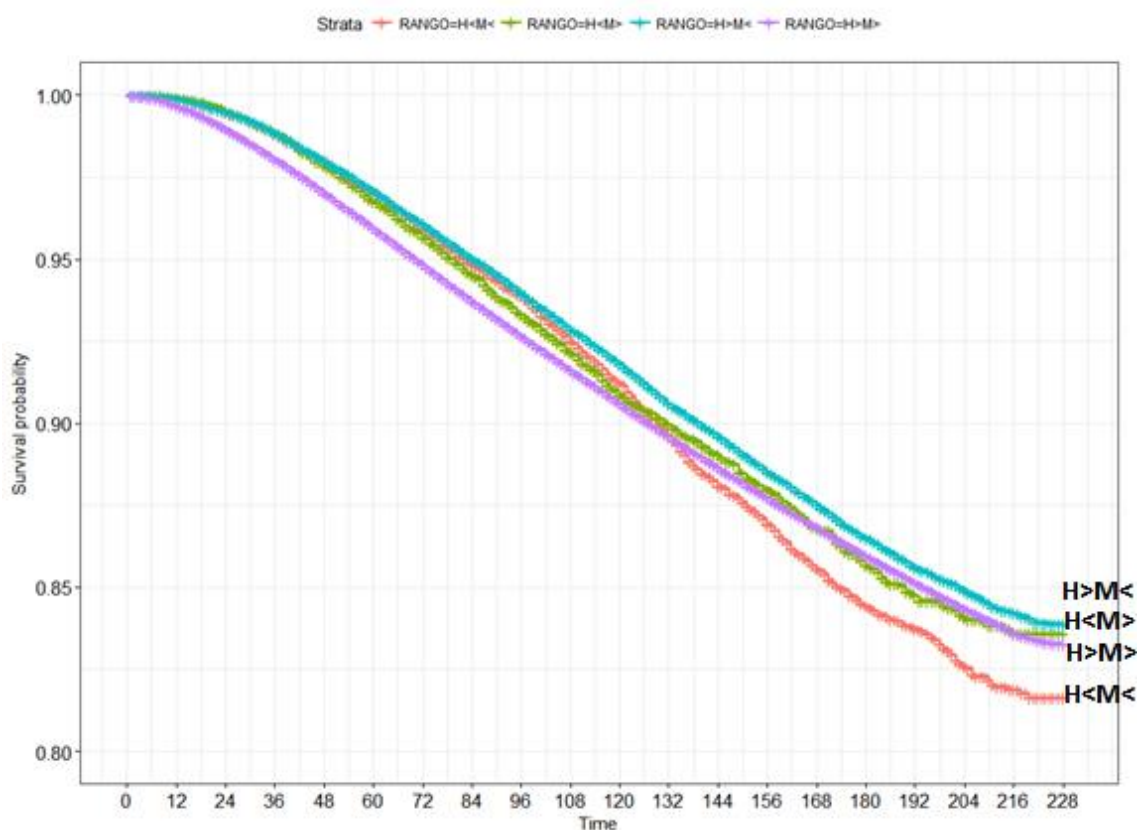


Ilustración 6-22 Curva de supervivencia de matrimonios por rangos para matrimonios con menores de edad

¹¹ Ley reformativa al código civil, <http://www.oficial.ec/ley-reformativa-codigo-civil>

Observamos en el gráfico anterior que la supervivencia de los matrimonios es mayor para el grupo H>M< de matrimonios de hombres mayores de edad casados con mujeres menores de edad. Adicionalmente observamos que después de los 12 años (144 meses), la curva del grupo H<M< es menor (matrimonios de hombres menores de 18 años con mujeres menores de 18 años), los otros grupos presentan una curva de supervivencia mayor; sin embargo, no tienen mucha diferencia visible.

También observamos que las curvas de supervivencia se cortan entre sí, esto implica que no se cumple la condición de riesgos proporcionales. De esta manera es difícil encontrar alguna relación de prevalencia de grupos, es decir en este caso no encontramos ninguna relación determinante a la hora de analizar la supervivencia de un matrimonio. Es difícil en esta situación establecer dos grupos.

En este caso podemos usar la regresión de Cox para hallar alguna relación.

```
coxph(formula = Surv(DURACION_MESES, ESTADO == 1) ~ RANGO, data = tabla_cruce2)
n= 1251099, number of events= 110986

              coef exp(coef) se(coef)      z Pr(>|z|)
RANGO<M> -0.06099  0.94083  0.04216 -1.447  0.148
RANGO<M> -0.12324  0.88405  0.02837 -4.343 1.4e-05 ***
RANGO<M> -0.02710  0.97327  0.02684 -1.009  0.313
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
RANGO<M>    0.9408      1.063    0.8662    1.0219
RANGO<M>    0.8841      1.131    0.8362    0.9346
RANGO<M>    0.9733      1.027    0.9234    1.0258

Concordance= 0.507 (se = 0.001 )
Rsquare= 0 (max possible= 0.909 )
Likelihood ratio test= 92.61 on 3 df,  p=0
Wald test               = 90.31 on 3 df,  p=0
Score (logrank) test = 90.38 on 3 df,  p=0
```

Ilustración 6-23 Regresión de Cox para matrimonios con menores de edad

Observando el gráfico anterior tenemos que:

- El modelo es significativo usando cualquiera de los tres criterios: test de razón de verosimilitud, test de Wald y test logrank.
- El grupo de matrimonios de hombres mayores de edad casados con mujeres menores (RANGO<M<) tiene un p-valor menor a 0.001, que a su vez es menor que 0.05, lo que implica que esta característica es significativa.
- El coeficiente es negativo (-0.12324) lo que indica que un matrimonio integrado con hombres mayores de edad casados con mujeres menores de edad disminuye

el riesgo de divorcio en 0.88405 veces. Un matrimonio integrado por hombres mayores de edad casados con mujeres menores de edad tiene 0.88405 veces menos riesgo de divorcio que un matrimonio que no se encuentre integrado por hombres mayores de edad casados con mujeres menores de edad. Si el riesgo de divorcio disminuye la supervivencia del matrimonio aumenta.

Analizando el resultado de Cox, podemos concluir que la supervivencia de un matrimonio es mayor en un matrimonio integrado de hombres mayores de edad casados con mujeres menores de edad.

Analizando las curvas de supervivencia de este estudio podemos concluir que después de 12 años de matrimonio entre 2 menores de edad, la curva de supervivencia es menor a los demás casos.

La formalización del divorcio ocurre cuando la pareja logra un acuerdo ante un notario o llega al final de un proceso en un juzgado. En ese momento se lo legaliza, se lo inscribe en el Registro Civil y se lo registra en un formulario del INEC. En muchas ocasiones esta formalización del divorcio no llega a ocurrir, debido a motivos como falta de acuerdo en la patria potestad de los hijos menores de edad o con capacidades especiales, falta de acuerdo en la separación de bienes y otros inconvenientes. El fracaso de un matrimonio no ocurre con el divorcio, generalmente ocurre varios meses atrás y debido a situaciones sociales y económicas no se inicia el proceso de divorcio sobre todo en parejas menores de edad. Estos aspectos están fuera del alcance de esta memoria que toma los datos proporcionados por el formulario del INEC.

Inclusive después de llegar de forma legal a un divorcio se tarda en registrarlo en el Registro Civil, debido a que muchas personas no consideran necesario hacerlo, sobre todo en las regiones rurales o en personas con nivel bajo de educación. Otras personas se demoran en registrar el divorcio por miedo a la etiqueta social “divorciado”, en especial en parejas menores de edad.

6.6 Regresión de Cox

Vamos a ajustar un modelo de Cox usando las variables más significativas:

- Edad del hombre
- Edad de la mujer
- Nivel de instrucción del hombre

- Nivel de instrucción de la mujer

El resultado de ejecutar el modelo de Cox es el siguiente.

```
coxph(formula = Surv(DURACION_MESES, ESTADO == 1) ~ EDAD_HOMBRE +
      EDAD_MUJER + RANGO_NIVEL_HOMBRE + RANGO_NIVEL_MUJER, data = tabla_cruce)

n= 1251099, number of events= 110986

      coef exp(coef) se(coef) z Pr(>|z|)
EDAD_HOMBRE -0.0048170 0.9951946 0.0005199 -9.266 < 2e-16 ***
EDAD_MUJER 0.0011824 1.0011831 0.0005878 2.012 0.0442 *
RANGO_NIVEL_HOMBRE2 0.3064872 1.3586441 0.0424608 7.218 5.27e-13 ***
RANGO_NIVEL_HOMBRE3 0.9013519 2.4629304 0.0421546 21.382 < 2e-16 ***
RANGO_NIVEL_HOMBRE4 0.7768986 2.1747171 0.0450988 17.227 < 2e-16 ***
RANGO_NIVEL_MUJER2 0.3833125 1.4671364 0.0377679 10.149 < 2e-16 ***
RANGO_NIVEL_MUJER3 0.9534124 2.5945481 0.0374572 25.453 < 2e-16 ***
RANGO_NIVEL_MUJER4 0.8902384 2.4357101 0.0408122 21.813 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
EDAD_HOMBRE 0.9952 1.0048 0.9942 0.9962
EDAD_MUJER 1.0012 0.9988 1.0000 1.0023
RANGO_NIVEL_HOMBRE2 1.3586 0.7360 1.2502 1.4766
RANGO_NIVEL_HOMBRE3 2.4629 0.4060 2.2676 2.6751
RANGO_NIVEL_HOMBRE4 2.1747 0.4598 1.9907 2.3757
RANGO_NIVEL_MUJER2 1.4671 0.6816 1.3625 1.5799
RANGO_NIVEL_MUJER3 2.5945 0.3854 2.4109 2.7922
RANGO_NIVEL_MUJER4 2.4357 0.4106 2.2485 2.6385

Concordance= 0.61 (se = 0.001 )
Rsquare= 0.02 (max possible= 0.909 )
Likelihood ratio test= 25323 on 8 df, p=0
Wald test = 22890 on 8 df, p=0
Score (logrank) test = 24438 on 8 df, p=0
```

Ilustración 6-24 Regresión de Cox de las variables más significativas

Con los resultados anteriores podemos observar lo siguiente:

- El modelo es significativo usando cualquiera de los tres criterios: test de razón de verosimilitud, test de Wald y test logrank.
- Analizando los resultados de Cox tenemos que casi todos los grupos de matrimonios tienen un p-valor menor a 0.001, lo que implica que estas características son significativas (***) según el gráfico. La edad de la mujer tiene un p-valor menor a 0.05, lo que implica que también es significativa (*).
- El gráfico anterior nos proporciona una información importante respecto a la estimación de los riesgos relativos a partir de los valores de “exp(coef)”:
 - La edad del hombre tiene 0.9951 veces el riesgo de divorcio de un hombre un año menor. Al aumentar la edad del hombre en 5 años, el riesgo de divorcio es 0.976202 ($\exp(5 \cdot -0.0048170)$) veces menos. Si el riesgo de divorcio disminuye la supervivencia del matrimonio aumenta.
 - La edad de una mujer tiene 1.0011 veces el riesgo de divorcio de una mujer un año menor. Al aumentar la edad de la mujer en 5 años, el riesgo de

divorcio es 1.00593 ($\exp(5 \cdot 0.0011824)$) veces más. Si el riesgo de divorcio aumenta la supervivencia del matrimonio disminuye.

- El coeficiente del nivel educación del hombre igual a primaria (2) es positivo (0.3064872) lo que indica que del nivel educación del hombre igual a primaria aumenta el riesgo de divorcio en 1.3586441 veces. Un matrimonio con nivel educación del hombre igual a primaria tiene 1.3586441 veces más riesgo de divorcio que un matrimonio que no tenga el nivel educación del hombre igual a primaria. Si el riesgo de divorcio aumenta la supervivencia del matrimonio disminuye.
- Bajo el mismo criterio anterior podemos deducir que:
 - Un matrimonio con nivel educación del hombre igual a secundaria tiene 2.46 veces más riesgo de divorcio que un matrimonio que no tenga el nivel educación del hombre igual a secundaria.
 - Un matrimonio con nivel educación del hombre igual a superior tiene 2.17 veces más riesgo de divorcio que un matrimonio que no tenga el nivel educación del hombre igual a superior.
 - Un matrimonio con nivel educación de la mujer igual a primaria tiene 1.46 veces más riesgo de divorcio que un matrimonio que no tenga el nivel educación de la mujer igual a primaria.

Las conclusiones serian entonces:

- Al aumentar la edad del hombre, el riesgo de divorcio disminuye, por lo que la supervivencia del matrimonio aumenta.
- Al aumentar la edad de la mujer, el riesgo de divorcio aumenta. por lo que la supervivencia del matrimonio disminuye.
- El nivel educación igual a primaria, secundaria y superior tanto del hombre como de la mujer aumenta el riesgo de divorcio por lo que la supervivencia del matrimonio disminuye.

6.7 Regresión de Cox con interacciones

Vamos a ajustar un modelo de Cox usando las variables más significativas y sus interacciones:

- Edad del hombre

- Edad de la mujer
- Nivel de instrucción del hombre
- Nivel de instrucción de la mujer

El resultado de ejecutar el modelo de Cox es el siguiente.

```
coxph(formula = Surv(DURACION_MESES, ESTADO == 1) ~ RANGO_EDAD_HOMBRE +
  RANGO_EDAD_MUJER + RANGO_NIVEL_HOMBRE + RANGO_NIVEL_MUJER +
  RANGO_EDAD_HOMBRE * RANGO_EDAD_MUJER + RANGO_EDAD_HOMBRE *
  RANGO_NIVEL_HOMBRE + RANGO_EDAD_HOMBRE * RANGO_NIVEL_MUJER +
  RANGO_EDAD_MUJER * RANGO_NIVEL_HOMBRE + RANGO_EDAD_MUJER *
  RANGO_NIVEL_MUJER + RANGO_NIVEL_HOMBRE * RANGO_NIVEL_MUJER,
  data = tabla_cruce)

n= 1251099, number of events= 110986

              coef exp(coef) se(coef)      z Pr(>|z|)
RANGO_EDAD_HOMBRE  1.712418  5.542345  0.055627  30.784 < 2e-16 ***
RANGO_EDAD_MUJER   0.761534  2.141559  0.055819  13.643 < 2e-16 ***
RANGO_NIVEL_HOMBRE  1.857136  6.405366  0.038833  47.823 < 2e-16 ***
RANGO_NIVEL_MUJER   1.659587  5.257137  0.038774  42.802 < 2e-16 ***
RANGO_EDAD_HOMBRE:RANGO_EDAD_MUJER -0.286523  0.750870  0.011543 -24.822 < 2e-16 ***
RANGO_EDAD_HOMBRE:RANGO_NIVEL_HOMBRE -0.229875  0.794633  0.015518 -14.814 < 2e-16 ***
RANGO_EDAD_HOMBRE:RANGO_NIVEL_MUJER -0.202254  0.816887  0.015467 -13.077 < 2e-16 ***
RANGO_EDAD_MUJER:RANGO_NIVEL_HOMBRE -0.044943  0.956052  0.015100  -2.976  0.00292 **
RANGO_EDAD_MUJER:RANGO_NIVEL_MUJER  0.024703  1.025011  0.015067  1.640  0.10110
RANGO_NIVEL_HOMBRE:RANGO_NIVEL_MUJER -0.312305  0.731759  0.007364 -42.407 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
RANGO_EDAD_HOMBRE      5.5423      0.1804  4.9699  6.1808
RANGO_EDAD_MUJER       2.1416      0.4669  1.9196  2.3891
RANGO_NIVEL_HOMBRE     6.4054      0.1561  5.9359  6.9119
RANGO_NIVEL_MUJER      5.2571      0.1902  4.8724  5.6722
RANGO_EDAD_HOMBRE:RANGO_EDAD_MUJER  0.7509      1.3318  0.7341  0.7681
RANGO_EDAD_HOMBRE:RANGO_NIVEL_HOMBRE  0.7946      1.2584  0.7708  0.8192
RANGO_EDAD_HOMBRE:RANGO_NIVEL_MUJER  0.8169      1.2242  0.7925  0.8420
RANGO_EDAD_MUJER:RANGO_NIVEL_HOMBRE  0.9561      1.0460  0.9282  0.9848
RANGO_EDAD_MUJER:RANGO_NIVEL_MUJER  1.0250      0.9756  0.9952  1.0557
RANGO_NIVEL_HOMBRE:RANGO_NIVEL_MUJER  0.7318      1.3666  0.7213  0.7424

Concordance= 0.618 (se = 0.001 )
Rsquare= 0.018 (max possible= 0.909 )
Likelihood ratio test= 22841 on 10 df,  p=0
Wald test               = 20673 on 10 df,  p=0
Score (logrank) test = 22100 on 10 df,  p=0
```

Ilustración 6-25 Regresión de Cox con interacciones de las variables más significativas

Con los resultados anteriores podemos observar lo siguiente:

- El modelo es significativo usando cualquiera de los tres criterios: test de razón de verosimilitud, test de Wald y test logrank.
- Analizando los resultados de Cox tenemos que:
 - Las interacciones más importantes tienen la variable rango de edad del hombre, tienen un p-valor menor a 0.001, lo que implica que estas características son significativas (***) según el gráfico.

- La interacción: RANGO_EDAD_MUJER:RANGO_NIVEL_HOMBRE tiene un p-valor menor a 0.01 lo que implica que estas características son también significativas (**) según el grafico.
- La interacción: RANGO_EDAD_MUJER:RANGO_NIVEL_MUJER tiene un p-valor mayor a 0.05 lo que implica que esta característica no es significativa según el grafico.
- Las variables: rango de edad tanto del hombre como de la mujer y las de nivel de instrucción tanto del hombre y de la mujer aumentan el riesgo de divorcio por lo que la supervivencia del matrimonio disminuye.
- Las interacciones con la variable rango de edad del hombre disminuye el riesgo de divorcio por lo que la supervivencia del matrimonio aumenta.

6.8 Exploración con árbol de decisión

Hasta el momento hemos creado grupos por edad basados en características sociales: mayor o menor edad (18), tercera edad (65), edad económicamente activa (35), niveles de instrucción, etc. Es necesario establecer puntos de quiebre de la edad, del nivel de instrucción desde el punto de vista estadístico. También tenemos varias conclusiones respecto al aumento o la disminución de la supervivencia del matrimonio, estas conclusiones son el resultado de análisis separados de variables, por lo que es necesario establecer un análisis conjunto.

La mejor manera de realizar un análisis conjunto es usando un árbol de decisión que tenga implícito el algoritmo de logrank en la construcción del modelo. Esto nos permite establecer segmentos de matrimonios que tengan mayor diferencia en cuanto a su curva de supervivencia.

El algoritmo de logrank en la construcción del modelo es necesario ya que permite crear nodos tomando en cuenta datos censurados (matrimonios continúan) y datos no censurados (divorcios).

Vamos a usar la función “rpart” para modelar un árbol de decisión ya que usa el algoritmo logrank cuando la entrada es una función de supervivencia “Surv”.

La prueba Log-Rank es el método más utilizado para comparar dos o más grupos de supervivencia. La hipótesis nula es que no hay diferencia en la supervivencia entre los dos grupos. El test Log-Rank es una prueba no paramétrica, que no hace suposiciones sobre las distribuciones de supervivencia. Esencialmente, el test Log-Rank compara el

número observado de eventos en cada grupo con lo que se esperaría si la hipótesis nula fuera verdadera (es decir, si las curvas de supervivencia eran idénticas). El estadístico de Log-Rank se distribuye aproximadamente como un test estadístico de chi-cuadrado.

Para un grupo 1 por ejemplo, el estadístico log-rank puede escribirse como:

$$\sum_{j=1}^r (d_{1j} - e_{1j})$$

donde la suma es sobre todos los tiempos de evento únicos, en un total de r veces.

d_{1j} es el número de muertes que ocurren en el grupo 1 en el tiempo j, y e_{1j} es el número esperado de eventos en el grupo 1 en el tiempo j.

El número esperado está dado por $n_{1j} d_j / n_j$, donde n_j es el número total de casos que están en riesgo justo antes del tiempo j, n_{1j} es el número en riesgo justo antes del tiempo j en el grupo 1 y d_j es el número total de muertes en el tiempo j en ambos grupos.

Las variables más significativas que usaremos para construir un árbol de decisión son:

- Edad del hombre
- Edad de la mujer
- Nivel de instrucción del hombre
- Nivel de instrucción de la mujer

A continuación, mostramos el código de lenguaje R para construir el árbol de decisión:

```
ModeloArbol<-rpart(Surv(DURACION_MESES, ESTADO==1) ~ EDAD_HOMBRE +
                  EDAD_MUJER + RANGO_NIVEL_HOMBRE + RANGO_NIVEL_MUJER,
                  data=tabla_cruce, control=rpart.control(cp=0.00105))
rpart.plot(ModeloArbol, type=1, extra=100, cex = .7)
summary(ModeloArbol)
print(ModeloArbol)
```

Ilustración 6-26 Código R para construcción de árbol de decisión de una función de supervivencia

El grafico del árbol de decisión es el siguiente:

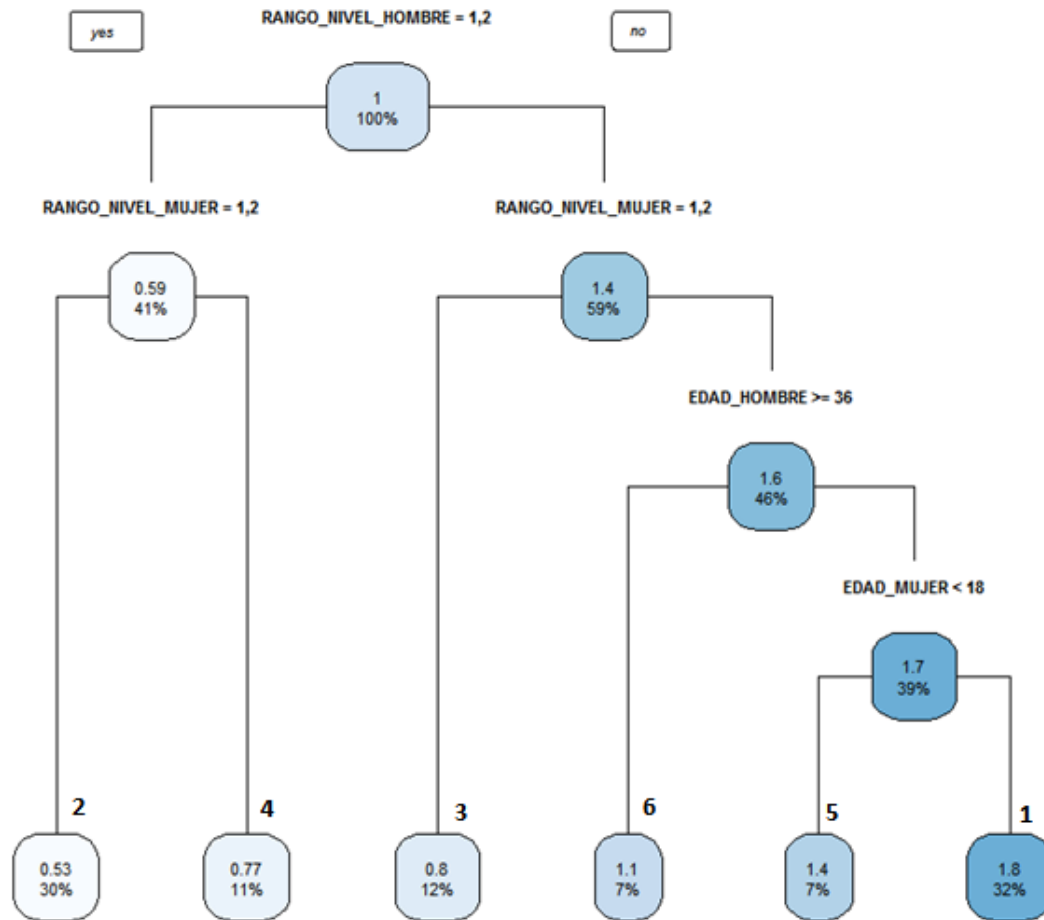


Ilustración 6-27 Árbol de decisión de la supervivencia de matrimonios

Analizando los resultados obtenemos el siguiente cuadro de los 6 segmentos realizados para la supervivencia de matrimonios por un árbol de decisión ordenados por mayor presencia.

| Perfil | Frecuencia | Porcentaje | Regla |
|--------|------------|------------|---|
| 1 | 406,383 | 32.48% | RANGO_NIVEL_HOMBRE=3,4 & RANGO_NIVEL_MUJER=3,4 & EDAD_HOMBRE<36 & EDAD_MUJER>=18 |
| 2 | 374,714 | 29.95% | RANGO_NIVEL_HOMBRE=1,2 & RANGO_NIVEL_MUJER=1,2 |
| 3 | 153,972 | 12.31% | RANGO_NIVEL_HOMBRE=3,4 & RANGO_NIVEL_MUJER=1,2 |
| 4 | 142,099 | 11.36% | RANGO_NIVEL_HOMBRE=1,2 & RANGO_NIVEL_MUJER=3,4 |
| 5 | 87,143 | 6.97% | RANGO_NIVEL_HOMBRE=3,4 & RANGO_NIVEL_MUJER=3,4 & EDAD_HOMBRE<36 & EDAD_MUJER<18 |
| 6 | 86,788 | 6.94% | RANGO_NIVEL_HOMBRE=3,4 & RANGO_NIVEL_MUJER=3,4 & EDAD_HOMBRE>=36 |
| | 1,251,099 | 100.00% | |

Tabla 6-9 Tabla de los segmentos del árbol de supervivencia de matrimonios

Las curvas de supervivencia resultado de la aplicación del árbol de decisión es la siguiente:

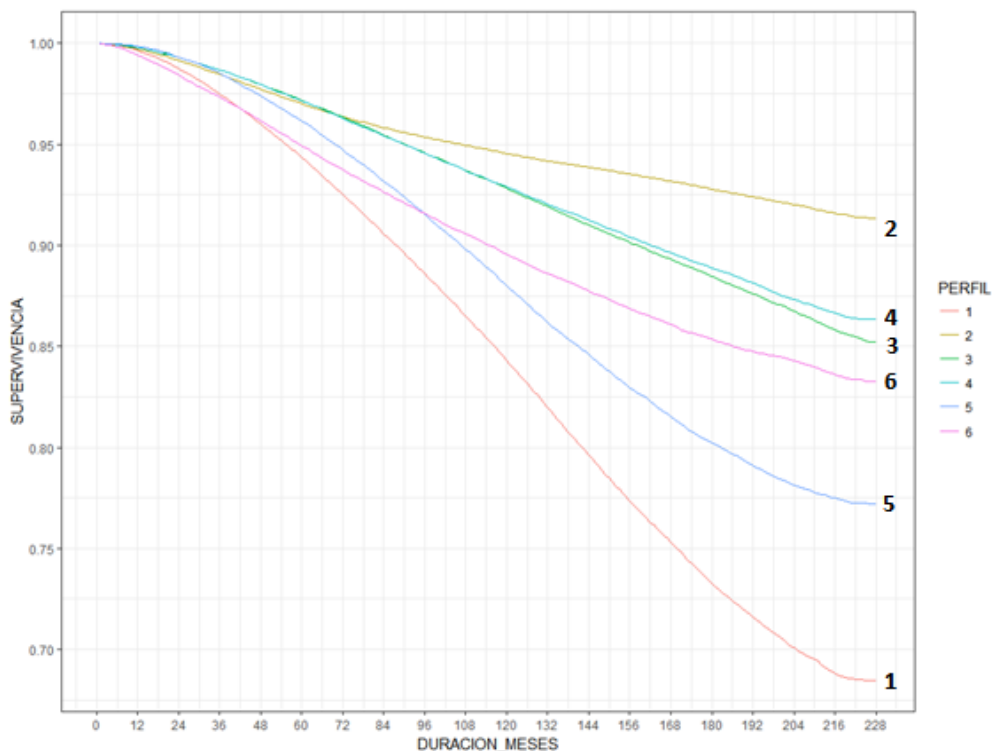


Ilustración 6-28 Curvas de supervivencia de los segmentos de matrimonio en base a los nodos del árbol de decisión

Se ha graficado las curvas de supervivencia en base a los perfiles.

Observamos en el gráfico anterior que la supervivencia de matrimonios es mayor para el perfil 2 que corresponde al segmento con las siguientes características: el hombre y la mujer tienen un nivel de educación: Ninguna, Centro de alfabetización, Primaria o Educación básica, sin importar la edad del hombre y la mujer.

La supervivencia de matrimonios es menor para el perfil 1 que corresponde al segmento con las siguientes características: el hombre y la mujer tienen un nivel de educación: Secundaria, Educación media/bachillerato, Ciclo posbachillerato, Superior o Posgrado; la edad del hombre es menor de 36 años y la mujer es mayor de edad.

Para comprobar los resultados anteriores vamos a usar el modelo de regresión de Cox.

```
coxph(formula = Surv(DURACION_MESES, ESTADO == 1) ~ PERFIL, data = tabla_cruce)
```

```
n= 1251099, number of events= 110986
```

| | coef | exp(coef) | se(coef) | z | Pr(> z) | |
|---------|-----------|-----------|----------|---------|----------|-----|
| PERFIL2 | -1.233538 | 0.291260 | 0.008253 | -149.46 | <2e-16 | *** |
| PERFIL3 | -0.824836 | 0.438307 | 0.010160 | -81.18 | <2e-16 | *** |
| PERFIL4 | -0.858104 | 0.423965 | 0.010847 | -79.11 | <2e-16 | *** |
| PERFIL5 | -0.294160 | 0.745158 | 0.010277 | -28.62 | <2e-16 | *** |
| PERFIL6 | -0.469441 | 0.625352 | 0.012936 | -36.29 | <2e-16 | *** |

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

| | exp(coef) | exp(-coef) | lower .95 | upper .95 |
|---------|-----------|------------|-----------|-----------|
| PERFIL2 | 0.2913 | 3.433 | 0.2866 | 0.2960 |
| PERFIL3 | 0.4383 | 2.282 | 0.4297 | 0.4471 |
| PERFIL4 | 0.4240 | 2.359 | 0.4150 | 0.4331 |
| PERFIL5 | 0.7452 | 1.342 | 0.7303 | 0.7603 |
| PERFIL6 | 0.6254 | 1.599 | 0.6097 | 0.6414 |

```
Concordance= 0.618 (se = 0.001 )  
Rsquare= 0.021 (max possible= 0.909 )  
Likelihood ratio test= 27180 on 5 df, p=0  
wald test = 26201 on 5 df, p=0  
Score (logrank) test = 28573 on 5 df, p=0
```

Ilustración 6-29 Resultados de Cox para perfiles de matrimonio

Observando el grafico anterior tenemos que:

- El modelo es significativo usando cualquiera de los tres criterios: test de razón de verosimilitud, test de Wald y test logrank.
- Los perfiles 2, 3, 4, 5 y 6 tienen un p-valor menor a 0.001 (***), que a su vez es menor que 0.05, lo que implica que esta característica es significativa a la hora de analizar la supervivencia de un matrimonio.
- El coeficiente del perfil 2 es negativo (-1.233538) lo que indica el perfil 2 de matrimonios disminuye el riesgo de divorcio en 0.2913 veces. Un matrimonio con perfil 2 tiene 0.2913 veces menos riesgo de divorcio que un matrimonio que no tenga el perfil 2. Si el riesgo de divorcio disminuye la supervivencia del matrimonio aumenta.
- El mismo criterio anterior se puede aplicar a los perfiles 3, 4, 5 y 6 ya que sus coeficientes son negativos.
- El perfil que disminuye el riesgo con mayor aceleración es el perfil 2 con un coeficiente de 0.2913. La supervivencia aumenta con mayor aceleración con el perfil 2.
- El perfil que disminuye el riesgo con menos aceleración es el perfil 5 con un coeficiente de 0.7452. La supervivencia aumenta con menor aceleración con el perfil 5.

Según el resultado de Cox, la supervivencia aumenta con mayor aceleración siguiendo el siguiente orden de perfiles: 2, 4, 3, 6 y 5. Esto concuerda con el gráfico de las curvas de supervivencia.

Analizando las curvas de supervivencia y el resultado de Cox por nivel de instrucción de la mujer, podemos concluir que la supervivencia de un matrimonio es mayor para el perfil 2.

La supervivencia de un matrimonio es menor para el perfil 1 según la curva de supervivencia. Según el resultado de Cox la supervivencia es menor para el perfil 5, esto se debe a que en el resultado de Cox no aparece el perfil 1, pero sí el perfil 5 que es el penúltimo en el análisis de las curvas de supervivencia.

La formalización del divorcio ocurre cuando la pareja logra un acuerdo ante un notario o llega al final de un proceso en un juzgado. En ese momento se lo legaliza, se lo inscribe en el Registro Civil y se lo registra en un formulario del INEC. En muchas ocasiones esta formalización del divorcio no llega a ocurrir, debido a motivos como falta de acuerdo en la patria potestad de los hijos menores de edad o con capacidades especiales, falta de acuerdo en la separación de bienes y otros inconvenientes. El fracaso de un matrimonio no ocurre con el divorcio, generalmente ocurre varios meses atrás y debido a situaciones sociales y económicas no se inicia el proceso de divorcio. Estos aspectos están fuera del alcance de esta memoria que toma los datos proporcionados por el formulario del INEC. Inclusive después de llegar de forma legal a un divorcio se tarda en registrarlo en el Registro Civil, debido a que muchas personas no consideran necesario hacerlo, sobre todo en las regiones rurales o en personas con nivel bajo de educación. Otras personas se demoran en registrar el divorcio por miedo a la etiqueta social “divorciado”. Inclusive existen casos que nunca llegan a registrar el divorcio.

6.9 Establecimiento de perfiles de supervivencia de matrimonios

Para el establecimiento de perfiles de matrimonio nos vamos a basar en las curvas de supervivencia de los segmentos de matrimonio en base a los nodos del árbol de decisión establecidos en el punto 6.8.

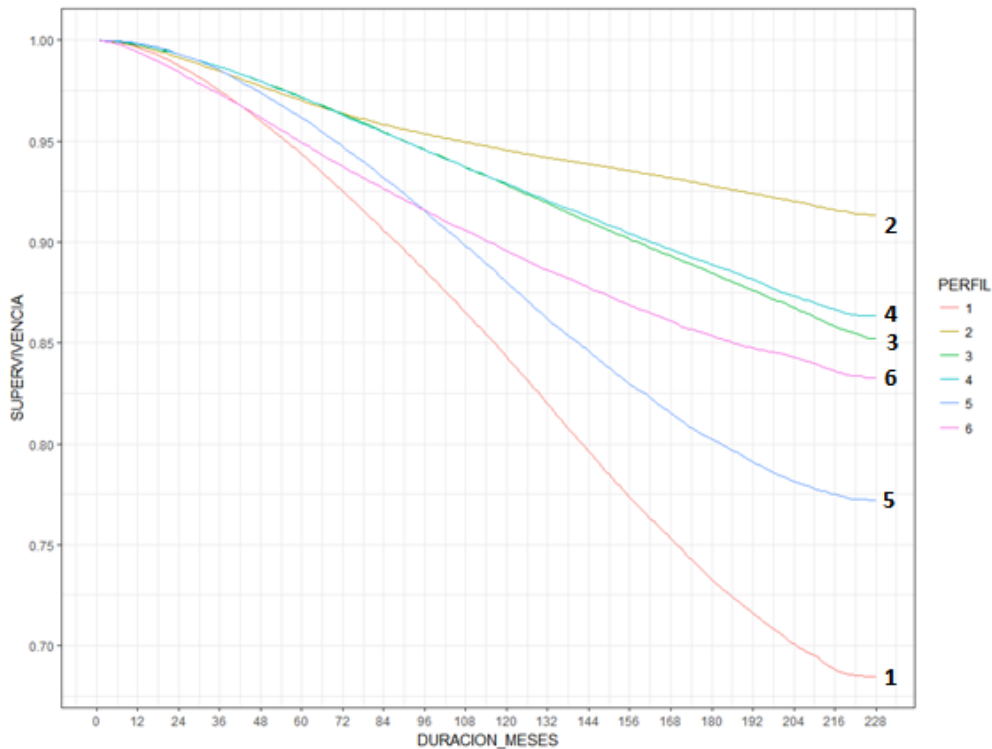


Ilustración 6-30 Curvas de Supervivencia por perfiles

Observamos lo siguiente en los perfiles iniciales 4 y 3:

- Las curvas de supervivencia del perfil inicial 4 y 3 son parecidas, casi idénticas.
- La regla del árbol para el perfil 4 indica que pertenecen a este grupo los matrimonios con nivel bajo de educación del hombre y nivel alto de educación de la mujer.
- La regla del árbol para el perfil 3 indica que pertenecen a este grupo los matrimonios con nivel alto de educación del hombre y nivel bajo de educación de la mujer.
- Debido a esto vamos a unir estos dos perfiles 4 y 3 en un solo grupo con una regla que diga que pertenecen a este grupo los matrimonios que tienen una diferencia entre el nivel de instrucción del hombre y la mujer es considerable.

Observamos lo siguiente en los perfiles iniciales 5 y 6:

- Son parecidos en cuanto a la tendencia de sus curvas de supervivencia.
- La conclusión del punto “6.5.1 Por nivel de instrucción del hombre” indica que la supervivencia de un matrimonio es mayor para niveles de educación bajo del hombre, lo que a su vez implica que la supervivencia de un matrimonio es menor para niveles de educación alta del hombre.

- La conclusión del punto “6.5.2 Por nivel de instrucción de la mujer” indica que la supervivencia de un matrimonio es mayor para niveles de educación bajo de la mujer, lo que a su vez implica que la supervivencia de un matrimonio es menor para niveles de educación alta de la mujer.
- En el punto “6.5.3 Por edad del hombre” se indica que no hay mayor conclusión en base a la edad del hombre.
- La regla del árbol para el perfil 5 indica que pertenecen a este grupo los matrimonios con nivel de instrucción alto tanto el hombre como la mujer, la edad del hombre menor a 36 años y la edad de la mujer menor a 18 años.
- La regla del árbol para el perfil 6 indica que pertenecen a este grupo los matrimonios con nivel de instrucción alto tanto el hombre como la mujer, la edad del hombre mayor a 36 años.
- Debido a esto vamos a unir estos dos perfiles 5 y 6 en un solo grupo con una regla que diga que pertenecen a este grupo los matrimonios que tienen nivel de instrucción alto tanto el hombre como la mujer, la edad de la mujer es menor a 18 años y no importa la edad del hombre

El análisis anterior da como resultado las curvas de supervivencia siguientes:

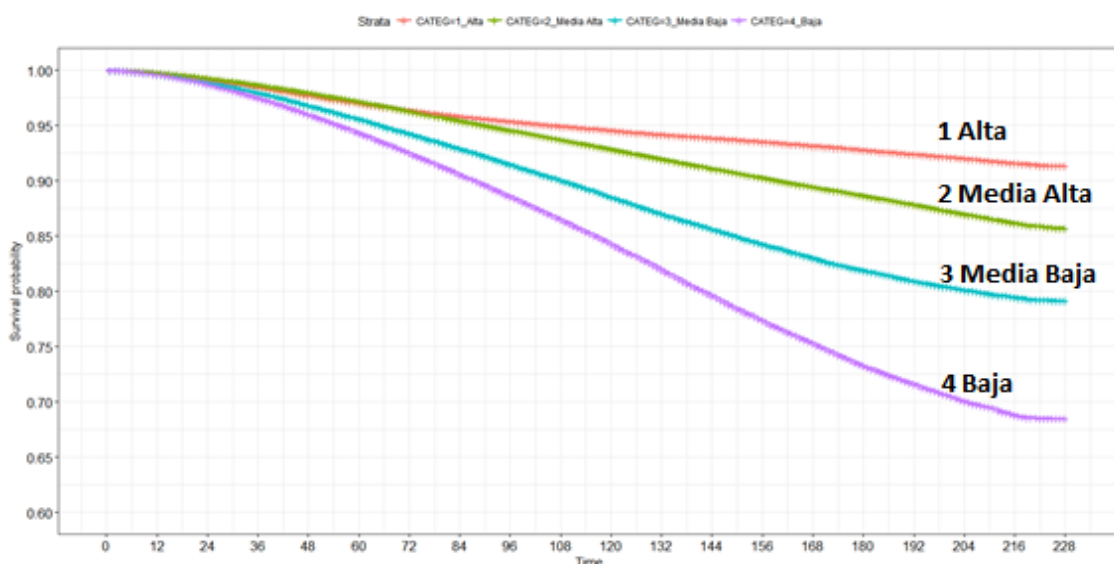


Ilustración 6-31 Curvas de supervivencia por categoría

A continuación, presentamos los perfiles ordenados en forma descendente en base a la supervivencia de matrimonio. Al aumentar la supervivencia disminuye el riesgo, es decir que los perfiles también están ordenados en base al riesgo en forma ascendente.

6.9.1 Perfil alto de supervivencia de matrimonio

En este perfil se encuentran los matrimonios que tienen una curva de supervivencia alta y un riesgo menor. Considerando un horizonte de 19 años en este perfil podemos decir que el 91% de matrimonios sobrevive.

Las características de estos matrimonio son: el hombre y la mujer tienen nivel de instrucción: Ninguna, Centro de alfabetización, Primaria o Educación básica, no importa la edad del hombre o de la mujer.

Este perfil corresponde al 29.95% de los matrimonios que alcanzaron una supervivencia de 19 años o más.

6.9.2 Perfil medio alto de supervivencia de matrimonio

En este perfil se encuentran los matrimonios que tienen una curva de supervivencia media alta y un riesgo medio bajo. Considerando un horizonte de 19 años en este perfil podemos decir que el 85% de matrimonios sobrevive.

Las características de estos matrimonios son: diferencia considerable entre el nivel de instrucción del hombre y la mujer. Es decir, cuando el hombre tiene nivel de instrucción alta la mujer tiene nivel de instrucción baja, de la misma manera cuando el hombre tiene nivel de instrucción baja la mujer tiene nivel de instrucción alta.

Este perfil corresponde al 23.67% de los matrimonios que alcanzaron una supervivencia de 19 años o más.

6.9.3 Perfil medio bajo de supervivencia de matrimonio

En este perfil se encuentran los matrimonios que tienen una curva de supervivencia media baja y un riesgo medio alto. Considerando un horizonte de 19 años en este perfil podemos decir que el 79% de matrimonios sobrevive.

Las características de estos matrimonios son: nivel de instrucción alto tanto el hombre como la mujer, la edad de la mujer es menor a 18 años y no importa la edad del hombre.

Este perfil corresponde al 13.91% de los matrimonios que alcanzaron una supervivencia de 19 años o más.

6.9.4 Perfil bajo de supervivencia de matrimonio

En este perfil se encuentran los matrimonios que tienen una curva de supervivencia baja y un riesgo mayor. Considerando un horizonte de 19 años en este perfil podemos decir que el 68% de matrimonios sobrevive.

Las características de estos matrimonios son: el hombre y la mujer tienen nivel de instrucción: Secundaria, Educación media/bachillerato, Ciclo posbachillerato, Superior o Posgrado, la edad del hombre menor a 36 años y la mujer es mayor de edad.

Este perfil corresponde al 32.48% de los matrimonios que alcanzaron una supervivencia de 19 años o más.

6.10 Ajuste de la curva de supervivencia por mortalidad

Hasta el momento tenemos datos de dos eventos que son el matrimonio y el divorcio y que se encuentran consignados en los formularios del INEC. Debemos considerar también que el matrimonio no se termina solamente con el divorcio sino también con el evento de la muerte de alguno de los integrantes.

Nos enfrentamos ahora con el problema de la mortalidad del hombre y de la mujer, por lo que necesitamos introducir esta situación en nuestro modelo de supervivencia y ajustarlo.

Vamos a calcular la tasa de mortalidad en base a los datos proporcionado por el INEC en el Anuario de Nacimientos y Defunciones del 2015¹², para lo cual usaremos las siguientes series correspondientes al 2015:

1. Número de defunciones por edades simples (horas) a nivel nacional. Periodo 1990 – 2015.
2. Número de defunciones por edades simples (días) a nivel nacional. Periodo 1990 – 2015.

¹² INEC 2016, Anuario Nacimientos y Defunciones 2015.xlsx,
http://www.ecuadorencifras.gob.ec/documentos/web-inec/Poblacion_y_Demografia/Nacimientos_Defunciones/Publicaciones/

3. Número de defunciones por edades simples (meses) a nivel nacional. Periodo 1990 – 2015.
4. Número de defunciones por edades simples (De 1 a 30 años) a nivel nacional. Periodo 1990 – 2015.
5. Número de defunciones por edades simples (De 31 a 60 años) a nivel nacional. Periodo 1990 – 2015.
6. Número de defunciones por edades simples (De 61 a 90 años) a nivel nacional. Periodo 1990 – 2015.
7. Número de defunciones por edades simples (De 91 a 116 años) a nivel nacional. Periodo 1990 – 2015.

Procedemos a calcular las tasas por edad según el número de defunciones por edad según datos del 2015. El total de defunciones es 64.790. La tasa de defunciones se calcula dividiendo el número de defunciones de la edad que corresponde dividido para el total. La tasa acumulada es la suma de las tasas correspondientes a los años anteriores más el año actual.

| DATOS DEL INEC | | | |
|-----------------------|------------------------------|-------------|-----------------------|
| EDAD AÑOS | NUMERO DE DEFUNCIONES | TASA | TASA ACUMULADA |
| ... | | | |
| 18 | 281 | 0.0043 | 0.0865 |
| 19 | 288 | 0.0044 | 0.0909 |
| 20 | 314 | 0.0048 | 0.0958 |
| 21 | 333 | 0.0051 | 0.1009 |
| 22 | 302 | 0.0047 | 0.1056 |
| 23 | 295 | 0.0046 | 0.1101 |
| 24 | 339 | 0.0052 | 0.1153 |
| 25 | 318 | 0.0049 | 0.1203 |
| 26 | 330 | 0.0051 | 0.1253 |
| 27 | 324 | 0.0050 | 0.1303 |
| 28 | 321 | 0.0050 | 0.1353 |
| 29 | 281 | 0.0043 | 0.1396 |
| 30 | 279 | 0.0043 | 0.1439 |
| ... | | | |
| Total | 64.790 | 1.0000 | |

Tabla 6-10 Tabla parcial de mortalidad por edad según datos del 2015

El ajuste de la supervivencia se calcula de la siguiente manera:

$$S'(t) = S(t) * P(t)$$

$S(t)$ es el valor de la supervivencia que ya hemos calculado.

$P(t)$ es la probabilidad de que el matrimonio sobreviva t meses.

$S'(t)$ es el valor de la supervivencia ajustado.

La probabilidad de mortalidad que hemos construido se encuentra calculada por años.

La probabilidad de que el matrimonio sobreviva t años es la probabilidad de que cada cónyuge no muera.

$$P(t) = P(\text{el no muera}) * P(\text{ella no muera})$$

La probabilidad de que alguien no muera es uno menos la probabilidad que muera.

$$P(t) = (1 - P(\text{el muera})) * (1 - P(\text{ella muera}))$$

La probabilidad de que alguien muera a los 4 años de matrimonio teniendo 40 de edad, sería.

$$P(\text{el muera en 4 años con edad inicial 40 años}) = A(44) - A(40)$$

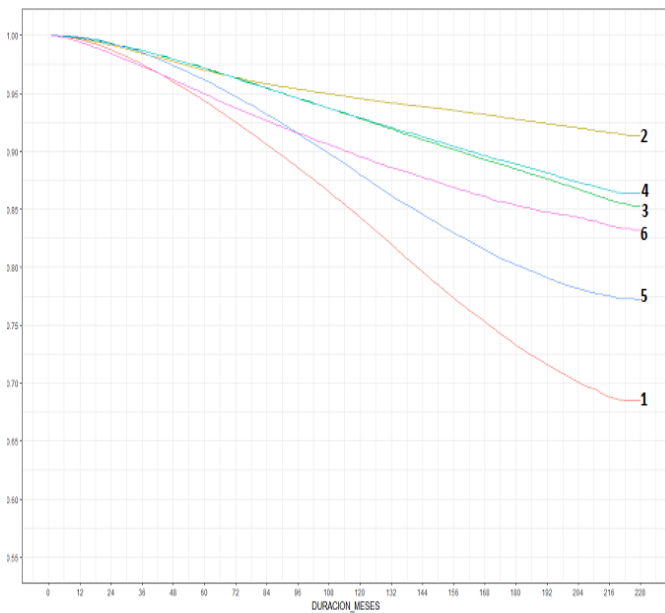
Donde $A(t)$ es la probabilidad acumulada de defunción que le hemos calculado en la tabla de mortalidad por edad en años. Aquí podemos observar, que la probabilidad acumulada de mortalidad es la misma para un matrimonio de duración de 12 a 23 meses y corresponde a la probabilidad de 1 año.

Una vez calculada $P(t)$ que es la probabilidad de que el matrimonio sobreviva t , donde t es la duración del matrimonio, procedemos a calcular la supervivencia ajustada realizando un promedio del producto de la supervivencia inicial por $P(t)$. El promedio lo realizamos debido a que existen varios matrimonios que tienen la misma duración.

$$S'(t) = \text{Promedio}(S(t) * P(t))$$

Aplicando el ajuste de la curva de supervivencia tenemos el siguiente gráfico:

SUPERVIVENCIA



SUPERVIVENCIA AJUSTADA

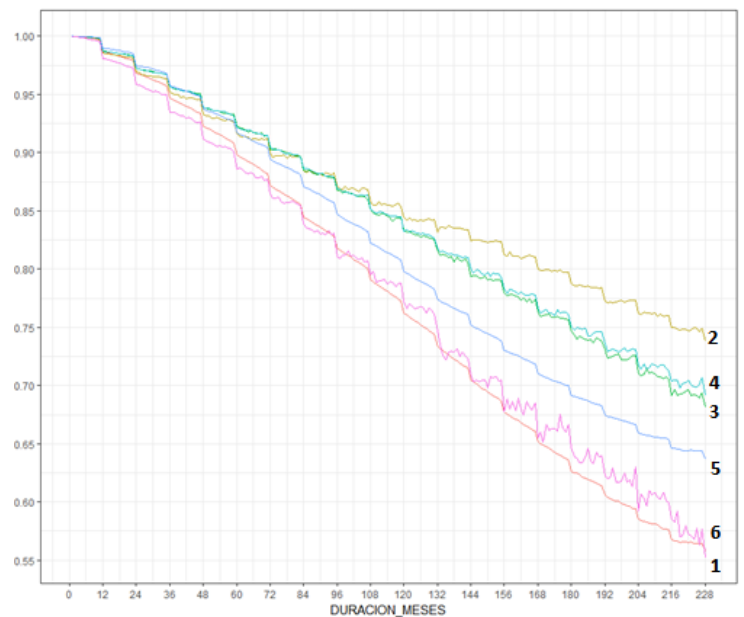


Ilustración 6-32 Comparación de supervivencia original y ajustada

Analizando el gráfico anterior podemos observar que la curva de la supervivencia ajustada tiene los valores siempre menores que la supervivencia original debido a que la probabilidad de que el matrimonio sobreviva a muertes y divorcio siempre es menor de que sobreviva a solo divorcio.

La curva de supervivencia ajustada presenta entrecortes con una longitud de un año, debido a que la mortalidad la tenemos por año. Existe una tasa de mortalidad por año y no una para cada mes del año.

En general las curvas con supervivencia ajustada tienen las mismas tendencias y tienen el mismo orden que las curvas de supervivencia original. La excepción es la curva de supervivencia con perfil 6 que tiene una supervivencia menor que el perfil 5. Esto es explicable debido a que el perfil 6 tiene a hombres mayores de 36 años que tienen una tasa de mortalidad mayor que los menores de 36 años que los tiene el perfil 5.

6.11 Análisis de sensibilidad

Como lo habíamos descrito en el apartado 6.3 respecto al “Emparejamiento de matrimonios y divorcios”, es importante determinar si el matrimonio se ha terminado o continúa todavía.

Para emparejar un divorcio con su matrimonio se usó la distancia y, en el caso de empate, asignación por ordenación aleatoria. Queremos saber ahora, que tan sensible es esta asignación aleatoria y si incide en la construcción del árbol de decisión. Por motivos de tiempo solo se ha hecho una prueba, pero parece claro que no afecta significativamente al resultado.

Hemos cambiando la semilla de 12345 a 55555 para generar un valor aleatorio diferente para el emparejamiento, y hemos obtenido el siguiente árbol.

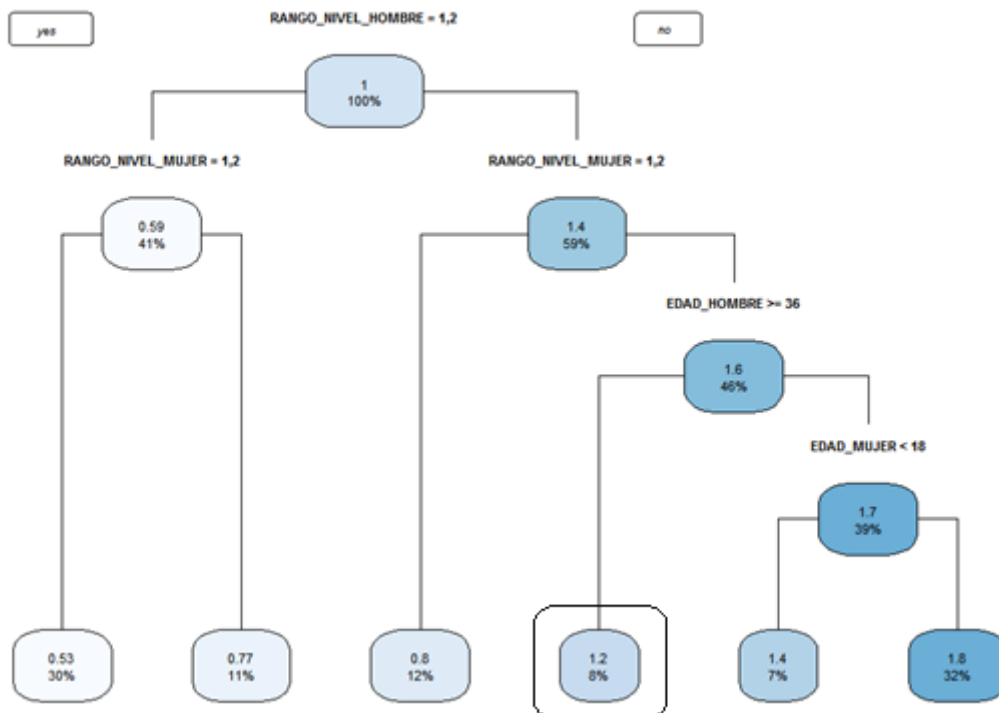


Ilustración 6-33 Árbol de decisión con nueva semilla

Comparando este árbol con el árbol original observamos que no existen diferencias en cuanto a las reglas. La única diferencia es un valor que cambia del 7% al 8% en la cuarta hoja. Este cambio es comprensible debido a efectos de redondeo de valores numéricos.

Podemos concluir entonces que el valor aleatorio no influye en el emparejamiento de los datos del divorcio con los datos del matrimonio.

7 Cronograma

La línea de tiempo está medida en semanas comenzando en la primera semana del año 2017 y acabando en la semana 20 que equivaldría a inicios de junio.

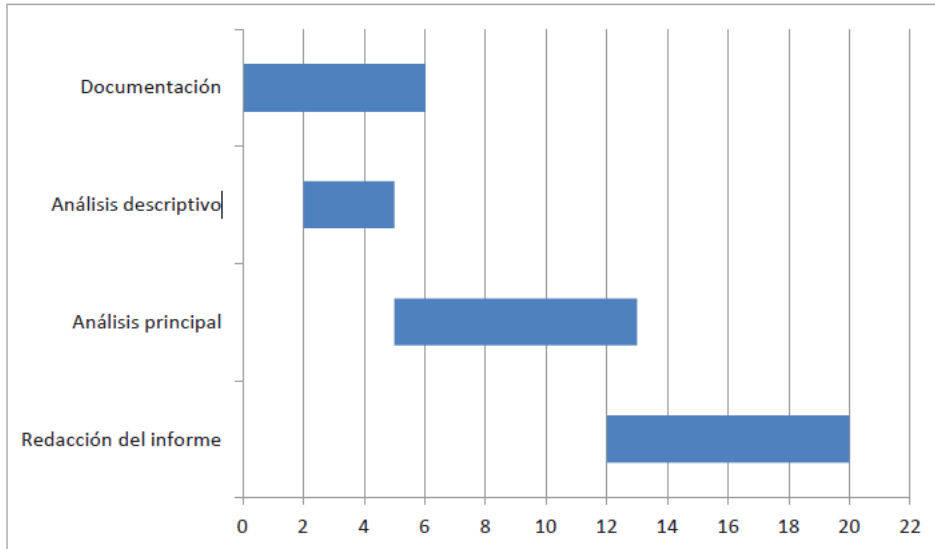


Ilustración 7-1 Cronograma de desarrollo del TFM

8 Periodo y Lugar

El trabajo será llevado a cabo en la ciudad de Madrid entre el 1 de febrero de 2017 y el 2 de junio de 2017.

9 Conclusiones

- Un árbol de decisión con la opción logrank construye segmentos de matrimonios que tienen mayor diferencia en cuanto a su curva de supervivencia.
- El árbol de decisión construye puntos de quiebre de la edad en 18 y 36 años, lo que se relaciona con factores sociales como: mayor o menor edad (18 años) y edad económicamente activa (35 años).
- Una manera de validar el resultado del árbol de decisión es incorporar datos de nuevos años de divorcios y matrimonios y comparar el resultado del árbol con el obtenido en esta memoria.
- Hemos usados datos imperfectos, como se ha indicado en la extracción, exploración y emparejamiento de matrimonios y divorcios, con estas características hemos construido un modelo aproximado de supervivencia de matrimonios. Debemos indicar que no se ha encontrado o no tenemos acceso a datos completos de cada matrimonio ni en España ni otros sitios. Además, debemos indicar que leyes de protección de datos personales nos impide analizar más características que desearíamos.
- Haciendo un análisis por nivel de instrucción del hombre, se llega a la conclusión de que la supervivencia de un matrimonio es mayor para niveles de educación bajo del hombre.
- Haciendo un análisis por nivel de instrucción de la mujer, se llega a la conclusión de que la supervivencia de un matrimonio es mayor para niveles de educación bajo de la mujer.
- Haciendo un análisis por edad de la mujer, se llega a la conclusión de que la supervivencia de un matrimonio es mayor para mujeres mayores de 65 años.
- Haciendo un análisis por matrimonios con menores de edad, se llega a la conclusión de que la supervivencia de un matrimonio es mayor para hombres mayores de edad casados con mujeres menores de edad.
- La formalización del divorcio ocurre cuando la pareja logra un acuerdo ante un notario o llega al final de un proceso en un juzgado. En ese momento se lo legaliza, se lo inscribe en el Registro Civil y se lo registra en un formulario del INEC. En muchas ocasiones esta formalización del divorcio no llega a ocurrir, debido a motivos como falta de acuerdo en la patria potestad de los hijos menores de edad o con capacidades especiales, falta de acuerdo en la separación de bienes y otros inconvenientes. El

fracaso de un matrimonio no ocurre con el divorcio, generalmente ocurre varios meses atrás y debido a situaciones sociales y económicas no se inicia el proceso de divorcio. Estos aspectos están fuera del alcance de esta memoria que toma los datos proporcionados por el formulario del INEC.

- Inclusive después de llegar de forma legal a un divorcio se tarda en registrarlo en el Registro Civil, debido a que muchas personas no consideran necesario hacerlo, sobre todo en las regiones rurales o en personas con nivel bajo de educación. Otras personas se demoran en registrar el divorcio por miedo a la etiqueta social “divorciado”. Inclusive existen casos que nunca llegan a registrar el divorcio. Personas con alto nivel de estudios consideran que es importante realizar el registro del divorcio en el Registro Civil.
- El Perfil alto de supervivencia de matrimonio, tiene las siguientes características: el hombre y la mujer tienen nivel de instrucción: Ninguna, Centro de alfabetización, Primaria o Educación básica, no importa la edad del hombre o de la mujer. Estos matrimonios tienen un riesgo menor.
- El Perfil medio alto de supervivencia de matrimonio tiene las siguientes características: diferencia considerable entre el nivel de instrucción del hombre y la mujer. Es decir, cuando el hombre tiene nivel de instrucción alta la mujer tiene nivel de instrucción baja, de la misma manera cuando el hombre tiene nivel de instrucción baja la mujer tiene nivel de instrucción alta. Estos matrimonios tienen un riesgo medio bajo.
- El Perfil medio bajo de supervivencia de matrimonio tiene las siguientes características: nivel de instrucción alto tanto el hombre como la mujer, la edad de la mujer es menor a 18 años y no importa la edad del hombre. Estos matrimonios tienen un riesgo medio alto.
- El Perfil bajo de supervivencia de matrimonio tiene las siguientes características: el hombre y la mujer tienen nivel de instrucción: Secundaria, Educación media/bachillerato, Ciclo posbachillerato, Superior o Posgrado, la edad del hombre menor a 36 años y la mujer es mayor de edad. Estos matrimonios tienen un riesgo alto.
- Se ha establecido un ajuste de las curvas de supervivencia debido a la mortalidad que da lugar a curvas de supervivencia similares. Para mejorar el ajuste de la supervivencia es recomendable tener tablas de mortalidad dependiendo del sexo y el

nivel de educación. Al momento tenemos un solo valor para una edad específica sin importar el sexo y el nivel de educación.

- En el análisis de sensibilidad, debemos destacar que a pesar de la asignación pseudo aleatoria por ordenación, el árbol y las curvas de supervivencia siguen siendo las mismas, aunque se varíe el sorteo.
- Para tener un análisis más completo de la supervivencia de matrimonios sería ideal tener datos de variables económicas del hombre y de la mujer, así como también su religiosidad que es un factor importante en Latinoamérica.

10 Trabajo futuro

- Realizar cambios en los scripts de código de R de manera que se encuentre automatizada la carga con un nuevo año de datos. Cada año el INEC publica un nuevo anuario de matrimonios y divorcios por lo que es necesaria una automatización en este caso. Al momento se debe cambiar e incorporar nuevo código el momento que se publican nuevos datos.
- Una importante mejora que se puede producir es la obtención de datos más exactos en el INEC. Desde el año 2012 para atrás no tenemos las fechas de divorcio y matrimonio con día, solamente tenemos con mes y año. Mejoraríamos la precisión al emparejar el divorcio con su matrimonio.
- Un trabajo importante que se debe realizar en la fuente del INEC es depurar los datos de matrimonios y divorcios con el objetivo de mejorar su calidad. Existen edades con valores 999 o 99 que como hemos visto no corresponden. El nivel de instrucción tiene también un considerable número de matrimonios con valor: “Se ignora”. Sería importante establecer estos valores para todos los matrimonios.

11 Bibliografía

- INEC 2016, Anuario_Matrimonios_y_Divorcios2015.pdf, http://www.ecuadorencifras.gob.ec/documentos/web-inec/Poblacion_y_Demografia/Matrimonios_Divorcios/Publicaciones/
- INEC 2016, Anuario Nacimientos y Defunciones 2015.xlsx, http://www.ecuadorencifras.gob.ec/documentos/web-inec/Poblacion_y_Demografia/Nacimientos_Defunciones/Publicaciones/
- INEC, <http://www.ecuadorencifras.gob.ec/el-numero-de-divorcios-en-ecuador-crecio-1191-en-diez-anos/>
- INEC, <http://www.ecuadorencifras.gob.ec/matrimonios-y-divorcios-bases-de-datos/>
- CRISP-DM, https://es.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining
- EPN, Código Civil del Ecuador, Codigo-Civil1.pdf, <http://www.epn.edu.ec/wp-content/uploads/2015/06/Codigo-Civil1.pdf>
- Ley reformatoria al código civil, <http://www.oficial.ec/ley-reformatoria-codigo-civil>
- Código orgánico general de procesos, <http://www.ambiente.gob.ec/wp-content/uploads/downloads/2016/10/Codigo-Organico-General-de-Procesos.pdf>
- Allison, Paul D. 2010. Survival Analysis Using SAS®: A Practical Guide, Second Edition. SAS Institute Inc.
- Álvarez Cáceres, Rafael C. 2013. Análisis de la Supervivencia: Regresión del Cox. Ediciones Alfanova.
- Dana Rotz, Why Have Divorce Rates Fallen? The Role of Women's Age at Marriage, 2015, <http://jhr.uwpress.org/content/early/2015/11/20/jhr.51.4.0214-6224R?related-urls=yes&legid=wpjhr;jhr.51.4.0214-6224Rv1>

12 Índice de tablas

| | |
|---|----|
| TABLA 5-1 EJEMPLO DE CONSTRUCCIÓN DE UNA CURVA DE SUPERVIVENCIA DE MATRIMONIOS..... | 19 |
| TABLA 6-1 VARIABLES DE MATRIMONIO Y VARIABLES DE DIVORCIO..... | 26 |
| TABLA 6-2 ESTRATIFICACIÓN DE LA EDAD POR CARACTERÍSTICAS SOCIALES | 29 |
| TABLA 6-3 NIVEL DE INSTRUCCIÓN DEFINIDA EN EL INEC..... | 29 |
| TABLA 6-4 ESTRATIFICACIÓN DEL NIVEL DE INSTRUCCIÓN..... | 29 |
| TABLA 6-5 SIGNIFICADO DE LAS VARIABLES FINALES..... | 30 |
| TABLA 6-6 TABLA DE VARIABLES DE MATRIMONIO Y DE DIVORCIO | 31 |
| TABLA 6-7 TABLA DE NUMERO DE EMPAREJADOS POR DISTANCIA..... | 33 |
| TABLA 6-8 EMPAREJAMIENTO DE UN DIVORCIO CON UN MATRIMONIO | 34 |
| TABLA 6-9 TABLA DE LOS SEGMENTOS DEL ÁRBOL DE SUPERVIVENCIA DE MATRIMONIOS | 62 |
| TABLA 6-10 TABLA PARCIAL DE MORTALIDAD POR EDAD SEGÚN DATOS DEL 2015 | 69 |

13 Índice de figuras

| | |
|--|----|
| ILUSTRACIÓN 1-1 DIAGRAMA DE LAS FASES CRISP-DM..... | 6 |
| ILUSTRACIÓN 5-1 EJEMPLO DE CURVA DE SUPERVIVENCIA DE MATRIMONIOS | 20 |
| ILUSTRACIÓN 6-1 FRECUENCIA DE LA EDAD DEL HOMBRE MAYOR QUE 80 EN EL MATRIMONIO | 26 |
| ILUSTRACIÓN 6-2 FRECUENCIA DEL NIVEL DE INSTRUCCIÓN DEL HOMBRE EN EL MATRIMONIO | 27 |
| ILUSTRACIÓN 6-3 DIAGRAMA DE FLUJO DEL EMPAREJAMIENTO DE DIVORCIOS CON SU MATRIMONIO | 35 |
| ILUSTRACIÓN 6-4 NÚMERO DE MATRIMONIOS Y DIVORCIOS EN EL ECUADOR..... | 36 |
| ILUSTRACIÓN 6-5 CÓDIGO Y RESULTADOS SAS PARA ESTIMAR LA FUNCIÓN DE SUPERVIVENCIA | 37 |
| ILUSTRACIÓN 6-6 CURVA DE SUPERVIVENCIA PARA LA TOTALIDAD DE MATRIMONIOS REALIZADA EN SAS. | 37 |
| ILUSTRACIÓN 6-7 CÓDIGO Y RESULTADOS R PARA ESTIMAR LA FUNCIÓN DE SUPERVIVENCIA | 38 |
| ILUSTRACIÓN 6-8 CURVA DE SUPERVIVENCIA PARA LA TOTALIDAD DE MATRIMONIOS REALIZADA EN R ... | 38 |
| ILUSTRACIÓN 6-9 CURVA DE SUPERVIVENCIA DE MATRIMONIOS POR RANGOS DE NIVEL DE INSTRUCCIÓN DEL HOMBRE | 40 |
| ILUSTRACIÓN 6-10 CURVA DE SUPERVIVENCIA DE MATRIMONIOS POR NIVEL DE INSTRUCCIÓN DEL HOMBRE | 41 |
| ILUSTRACIÓN 6-11 RESULTADOS DE COX POR NIVEL DE INSTRUCCIÓN DEL HOMBRE | 41 |
| ILUSTRACIÓN 6-12 CURVA DE SUPERVIVENCIA DE MATRIMONIOS POR RANGOS DE NIVEL DE INSTRUCCIÓN DE LA MUJER..... | 43 |
| ILUSTRACIÓN 6-13 CURVA DE SUPERVIVENCIA DE MATRIMONIOS POR NIVEL DE INSTRUCCIÓN DE LA MUJER | 44 |
| ILUSTRACIÓN 6-14 RESULTADOS DE COX POR NIVEL DE INSTRUCCIÓN DE LA MUJER..... | 44 |
| ILUSTRACIÓN 6-15 CURVA DE SUPERVIVENCIA DE MATRIMONIOS POR RANGOS DE EDAD DEL HOMBRE. | 46 |
| ILUSTRACIÓN 6-16 CURVA DE SUPERVIVENCIA DE MATRIMONIOS POR EDAD DEL HOMBRE. | 47 |
| ILUSTRACIÓN 6-17 CURVA DE SUPERVIVENCIA DE MATRIMONIOS POR RANGOS DE EDAD DE LA MUJER..... | 48 |
| ILUSTRACIÓN 6-18 CURVA DE SUPERVIVENCIA DE MATRIMONIOS POR EDAD DE LA MUJER | 49 |
| ILUSTRACIÓN 6-19 RESULTADO DE COX POR EDAD DE LA MUJER | 49 |
| ILUSTRACIÓN 6-20 FUNCIÓN DE SUPERVIVENCIA DE LOS 30 GRUPOS MAYORITARIOS | 51 |
| ILUSTRACIÓN 6-21 FUNCIÓN DE SUPERVIVENCIA DEL MEJOR Y PEOR GRUPO MAYORITARIO..... | 52 |
| ILUSTRACIÓN 6-22 CURVA DE SUPERVIVENCIA DE MATRIMONIOS POR RANGOS PARA MATRIMONIOS CON MENORES DE EDAD | 53 |
| ILUSTRACIÓN 6-23 REGRESIÓN DE COX PARA MATRIMONIOS CON MENORES DE EDAD | 54 |
| ILUSTRACIÓN 6-24 REGRESIÓN DE COX DE LAS VARIABLES MÁS SIGNIFICATIVAS | 56 |
| ILUSTRACIÓN 6-25 REGRESIÓN DE COX CON INTERACCIONES DE LAS VARIABLES MÁS SIGNIFICATIVAS..... | 58 |
| ILUSTRACIÓN 6-26 CÓDIGO R PARA CONSTRUCCIÓN DE ÁRBOL DE DECISIÓN DE UNA FUNCIÓN DE SUPERVIVENCIA..... | 60 |
| ILUSTRACIÓN 6-27 ÁRBOL DE DECISIÓN DE LA SUPERVIVENCIA DE MATRIMONIOS | 61 |
| ILUSTRACIÓN 6-28 CURVAS DE SUPERVIVENCIA DE LOS SEGMENTOS DE MATRIMONIO EN BASE A LOS NODOS DEL ÁRBOL DE DECISIÓN | 62 |
| ILUSTRACIÓN 6-29 RESULTADOS DE COX PARA PERFILES DE MATRIMONIO | 63 |
| ILUSTRACIÓN 6-30 CURVAS DE SUPERVIVENCIA POR PERFILES..... | 65 |
| ILUSTRACIÓN 6-31 CURVAS DE SUPERVIVENCIA POR CATEGORÍA..... | 66 |
| ILUSTRACIÓN 6-32 COMPARACIÓN DE SUPERVIVENCIA ORIGINAL Y AJUSTADA | 71 |
| ILUSTRACIÓN 6-33 ÁRBOL DE DECISIÓN CON NUEVA SEMILLA..... | 72 |
| ILUSTRACIÓN 7-1 CRONOGRAMA DE DESARROLLO DEL TFM | 73 |

14 Anexos

14.1 Código para carga de datos de divorcios desde SPSS (fragmento de R)

```
1 #Leer archivo SPSS
2 divorcios2015 <- import("Divorcios_2015.sav")
3
4 #Tomar las variables que necesitamos|
5 dfdiv2015 <- data.frame(
6   ANIO_INSCRIPCION = as.numeric(divorcios2015$anio_insc),
7   PROVINCIA_INSCRIPCION = as.numeric(divorcios2015$prov_insc),
8   DURACION_MATRIMONIO = as.numeric(divorcios2015$dur_mat),
9   EDAD_HOMBRE = as.numeric(divorcios2015$edad_hom),
10  EDAD_MUJER = as.numeric(divorcios2015$edad_muj),
11  HIJOS_HOMBRE = as.numeric(divorcios2015$hijos_hom),
12  HIJOS_MUJER = as.numeric(divorcios2015$hijos_muj),
13  INSTRUCCION_HOMBRE = as.numeric(divorcios2015$niv_insth),
14  INSTRUCCION_MUJER = as.numeric(divorcios2015$niv_instm),
15  PARROQUIA_HABITUAL_HOMBRE = as.character(divorcios2015$parr_habh),
16  PARROQUIA_HABITUAL_MUJER = as.character(divorcios2015$parr_habm),
17  ANIO_DIVORCIO = year(as.Date(divorcios2015$fecha_div)),
18  MES_DIVORCIO = month(as.Date(divorcios2015$fecha_div)),
19  ANIO_MATRIMONIO = year(as.Date(divorcios2015$fecha_mat)),
20  MES_MATRIMONIO = month(as.Date(divorcios2015$fecha_mat))
21 )
22
23 #Unir las observaciones. Union vertical
24 dfdiv <- rbind(dfdiv1997, dfdiv1998, dfdiv1999, dfdiv2000,
25               dfdiv2001, dfdiv2002, dfdiv2003, dfdiv2004,
26               dfdiv2005, dfdiv2006, dfdiv2007, dfdiv2008,
27               dfdiv2009, dfdiv2010, dfdiv2011, dfdiv2012,
28               dfdiv2013, dfdiv2014, dfdiv2015)
29
```

14.2 Código para carga de datos de matrimonios desde SPSS (fragmento de R)

```
1 #Leer variables desde SPSS por cada año
2 matrimonios2015 <- import("matrimonios_2015.sav")
3
4 #Tomar las variables que necesitamos
5 dfmat2015 <- data.frame(
6   ANIO_INSCRIPCION = as.numeric(matrimonios2015$anio_insc),
7   PROVINCIA_INSCRIPCION = as.numeric(matrimonios2015$prov_insc),
8   ANIO_MATRIMONIO = year(as.Date(matrimonios2015$fecha_insc)),
9   MES_MATRIMONIO = month(as.Date(matrimonios2015$fecha_insc)),
10  EDAD_HOMBRE = as.numeric(matrimonios2015$edad_hom),
11  EDAD_MUJER = as.numeric(matrimonios2015$edad_muj),
12  INSTRUCCION_HOMBRE = as.numeric(matrimonios2015$niv_insth),
13  INSTRUCCION_MUJER = as.numeric(matrimonios2015$niv_instm),
14  PARROQUIA_HOMBRE = as.character(matrimonios2015$parr_habh),
15  PARROQUIA_MUJER = as.character(matrimonios2015$parr_habm)
16 )
17
18 #Unir las observaciones. Union vertical
19 dfmat <- rbind(dfmat1997, dfmat1998, dfmat1999, dfmat2000,
20               dfmat2001, dfmat2002, dfmat2003, dfmat2004,
21               dfmat2005, dfmat2006, dfmat2007, dfmat2008,
22               dfmat2009, dfmat2010, dfmat2011, dfmat2012,
23               dfmat2013, dfmat2014, dfmat2015)
```

14.3 Código para emparejamiento de un divorcio con un matrimonio

```
#####
#Emparejamiento.R          2017-05-01          Alvaro Gutierrez
#Objetivo: Emparejar un divorcio con un matrimonio
#Entrada: Archivos de divorcios y de matrimonios
```

```

#Salida: Archivos de emparejamiento de divorcios y amatrmonios: matrimonios_cruce y divorcios_cruce
#####
library(sqldf)
library(tcltk)

dfdiv <- read.csv(file = "C:/Master/TFM/datos/divorcios.csv", header = TRUE, sep = ",", quote = "\"", colClasses = "character")
dfmat <- read.csv(file = "C:/Master/TFM/datos/matrimonios.csv", header = TRUE, sep = ",", quote = "\"", colClasses = "character")
dfdiv[, c(1:9,12:16)] <- sapply(dfdiv[, c(1:9,12:16)], as.integer)
dfmat[, c(1:8,11:11)] <- sapply(dfmat[, c(1:8,11:11)], as.integer)

#El cruce se registra en divorcios
dfmat$C_ID_DIVORCIO <- 0
dfdiv$EDAD_HOMBRE_TENTATIVA <- with(dfdiv, EDAD_HOMBRE - (ANIO_DIVORCIO - ANIO_MATRIMONIO))
dfdiv$EDAD_MUJER_TENTATIVA <- with(dfdiv, EDAD_MUJER - (ANIO_DIVORCIO - ANIO_MATRIMONIO))
dfdiv$C_ID_MATRIMONIO <- 0
dfdiv$C_CANDIDATOS <- 0
dfdiv$C_DIS_EDAD_HOMBRE <- 0
dfdiv$C_DIS_EDAD_MUJER <- 0
dfdiv$C_DIS_INST_HOMBRE <- 0
dfdiv$C_DIS_INST_MUJER <- 0

#C_ID_MATRIMONIO
#-1 "Este divorcio tiene una fecha de matrimonio menor a 1997. No existen matrimonios con esa característica."
#-2 "Este divorcio no se puede cruzar con matrimonios ya que no empata edad e instruccion."
dfdiv[dfdiv$ANIO_MATRIMONIO <= 1996, ]$C_ID_MATRIMONIO <- -1
#Divorcios con matrimonio menores de 1997, son censurados a la izquierda.
#No se pueden cruzar pero deben mantenerse.

print(paste(Sys.time(), " Inicio"))
duracion <- Sys.time()
for (anio in c(1997:2015))
{
  for (mes in c(1:12))
  {
    #Obtener divorcios del mes
    div_mes <- dfdiv[dfdiv$ANIO_MATRIMONIO == anio & dfdiv$MES_MATRIMONIO == mes, ]

    #Obtener matrimonios del mes
    mat_mes <- dfmat[dfmat$ANIO_MATRIMONIO == anio & dfmat$MES_MATRIMONIO == mes, ]

    #Emparejar divorcios con matrimonios
    for(i in 1:nrow(div_mes))
    {
      if (nrow(div_mes)>0) #Considerar que 1:0 es 1 0
      {
        #Tomar un divorcio
        divorcio <- div_mes[i,]
        #Tomar los matrimonios posibles que no han sido cruzados
        mat_posibles <- mat_mes[mat_mes$C_ID_DIVORCIO == 0, ]

        if (nrow(mat_posibles) >= 1)
        {
          #Calcular distancias
          mat_posibles$DISTANCIA_EDAD <- abs(mat_posibles$EDAD_HOMBRE - divorcio$EDAD_HOMBRE_TENTATIVA) +
abs(mat_posibles$EDAD_MUJER - divorcio$EDAD_MUJER_TENTATIVA)
          mat_posibles$DISTANCIA_INSTRUCCION <- abs(mat_posibles$INSTRUCCION_HOMBRE -
divorcio$INSTRUCCION_HOMBRE) + abs(mat_posibles$INSTRUCCION_MUJER - divorcio$INSTRUCCION_MUJER)
          #Ordenar en base a las distancias y variable aleatoria
          mat_posibles <- mat_posibles[order(mat_posibles$DISTANCIA_EDAD,
mat_posibles$DISTANCIA_INSTRUCCION, mat_posibles$ID_MATRIMONIO), ]
          #Tomar el primer matrimonio emparejado
          matrimonio <- mat_posibles[1,]
          #Poner marca en dfmat y mat_mes, para no volverlo a tomar
          dfmat[dfmat$ID_MATRIMONIO == matrimonio$ID_MATRIMONIO, ]$C_ID_DIVORCIO <- divorcio$ID_DIVORCIO
          mat_mes[mat_mes$ID_MATRIMONIO == matrimonio$ID_MATRIMONIO, ]$C_ID_DIVORCIO <-
divorcio$ID_DIVORCIO
          #Poner marca en dfdiv
          dfdiv[dfdiv$ID_DIVORCIO == divorcio$ID_DIVORCIO, ]$C_ID_MATRIMONIO <- matrimonio$ID_MATRIMONIO
          dfdiv[dfdiv$ID_DIVORCIO == divorcio$ID_DIVORCIO, ]$C_CANDIDATOS <- nrow(mat_posibles)
          dfdiv[dfdiv$ID_DIVORCIO == divorcio$ID_DIVORCIO, ]$C_DIS_EDAD_HOMBRE <-
abs(matrimonio$EDAD_HOMBRE - divorcio$EDAD_HOMBRE_TENTATIVA)
          dfdiv[dfdiv$ID_DIVORCIO == divorcio$ID_DIVORCIO, ]$C_DIS_EDAD_MUJER <- abs(matrimonio$EDAD_MUJER -
divorcio$EDAD_MUJER_TENTATIVA)
          dfdiv[dfdiv$ID_DIVORCIO == divorcio$ID_DIVORCIO, ]$C_DIS_INST_HOMBRE <-
abs(matrimonio$INSTRUCCION_HOMBRE - divorcio$INSTRUCCION_HOMBRE)

```

```

dfdiv[dfdiv$ID_DIVORCIO == divorcio$ID_DIVORCIO, ]$C_DIS_INST_MUJER <-
abs(matrimonio$INSTRUCCION_MUJER - divorcio$INSTRUCCION_MUJER)
}
else
{
dfdiv[dfdiv$ID_DIVORCIO == divorcio$ID_DIVORCIO, ]$CRUCE_ID_MATRIMONIO <- -2
}
} #if
} #for divorcio
print(paste(Sys.time(), "Cruzando para el Año: ", anio, " Mes: ", mes, " Num Mat: ", nrow(mat_mes), " Num Div: ",
nrow(div_mes), " Duracion: ", round(Sys.time() - duracion, 2)))
duracion <- Sys.time()
} #for mes
} #for anio
write.csv(dfdiv, file = "C:/Master/TFM/datos/divorcios_cruce.csv", row.names = FALSE)
write.csv(dfmat, file = "C:/Master/TFM/datos/matrimonios_cruce.csv", row.names = FALSE)

```