ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS
DEPARTAMENTO DE POSGRADO

PROYECTO DE TITULACIÓN

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

“MAGÍSTER EN ESTADÍSTICA APLICADA”

TEMA:

Análisis Multivariante de las interacciones entre Factores Demográficos y Biomarcadores en pacientes con Linfoma no Hodgkin T

AUTOR:

EMILY ADRIANA ALCÍVAR TOALA

Guayaquil - Ecuador

2024

RESUMEN

El proyecto se enfoca en analizar pacientes diagnosticados con linfomas no Hodgkin T atendidos en SOLCA guayaquil, utilizando modelos de supervivencia como Kaplan-Meier y el modelo de regresión de Cox para evaluar el impacto de diversas covariables en la función de riesgo. Además de un análisis estadístico multivariante para determinar asociaciones entre las variables y sus categorías.

Los resultados indican que el 60% de los pacientes eran hombres y el 53% había fallecido, con una rápida disminución en la probabilidad de supervivencia en los primeros 10 meses y un tiempo mediano de supervivencia de aproximadamente 30 meses. Se identificaron factores significativos que afectan la supervivencia, como la edad, la velocidad de sedimentación globular, los niveles de lactato deshidrogenasa y la presencia de hepatomegalia. Los pacientes se clasificaron en tres grupos de riesgo (bajo, medio y alto), y se recomienda la identificación temprana de factores de riesgo para mejorar los resultados del tratamiento, así como la colaboración en investigaciones multicéntricas para optimizar el manejo de estos pacientes.

Palabras Clave: Cáncer, Linfomas T, Supervivencia, ACM.

ABSTRACT

The project focuses on analyzing patients diagnosed with non-Hodgkin T-cell lymphomas treated at SOLCA Guayaquil, using survival models such as Kaplan-Meier and the Cox regression model to assess the impact of various covariates on the risk function. Additionally, multivariate statistical analysis is conducted to determine associations between variables and their categories.

The results indicate that 60% of the patients were male and 53% had deceased, with a rapid decline in survival probability within the first 10 months and a median survival time of approximately 30 months. Significant factors affecting survival were identified, including age, erythrocyte sedimentation rate, lactate dehydrogenase levels, and hepatomegaly presence.

Patients were classified into three risk groups (low, medium, and high), and early identification of risk factors is recommended to improve treatment outcomes, along with collaboration in multicenter research to optimize patient management.

Palabras Clave: Cancer, Lymphoma T, Survival, ACM.

DEDICATORIA

A Dios ya que sin su guía y amor incondicional nada sería posible. A Virginia y Juan, cuyo apoyo constante han sido la base de mis logros. A Marlon por su aliento y compañía en los momentos de dificultad y alegría. A mis mentores y profesores, por guiarme y compartir su conocimiento, inspirándome a crecer y a superar mis propios límites. Este proyecto es un reflejo de todo lo que he aprendido y el resultado del esfuerzo conjunto de quienes han creído en mí.

AGRADECIMIENTO

Agradecer primeramente a Dios por permitir cada aprendizaje nuevo en mi vida, a mis padres por ser el soporte y motivación para continuar día a día, a Marlon por su ayuda incondicional, sus consejos han sido una de las razones por las que inicie esto.

A la Dra. Purificación por darme la oportunidad de formar parte de su equipo de trabajo. A mis compañeros de trabajo Gina y Mario que han sido parte del proceso con sus consejos.

A Fuad Huamán por su ayuda y acompañamiento en este proceso como tutor, gracias por toda la paciencia y tiempo que le ha dedicado a este proyecto. Por compartir sus conocimientos en la medicina con nosotros.

DECLARACIÓN EXPRESA

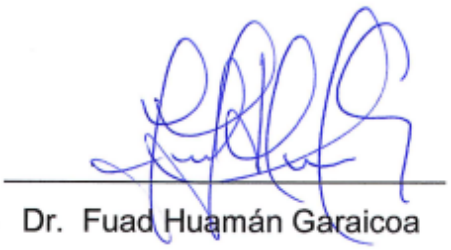
La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Titulación me corresponde exclusivamente y ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. El patrimonio intelectual del mismo corresponde exclusivamente a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

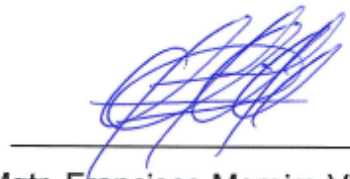
Emily Adriana Alcívar Toala

TRIBUNAL DE GRADUACIÓN

John Ramírez Figueroa, Ph.D
PRESIDENTE



Dr. Fuad Huamán Garaicoa
TUTOR



Mgtr. Francisco Moreira Villegas
DOCENTE EVALUADOR

ABREVIATURAS O SIGLAS

ACM	Análisis de correspondencia múltiple
AIC	Criterio de Akaike
B2M	Beta 2 Microglobulina
HB	Hemoglobina
LDH	Lactato deshidrogenasa
LH	Linfoma de Hodgkin
LNH	Linfoma no Hodgkin
LNH-T	Linfoma no Hodgkin de células T
OMS	Organización Mundial de Salud
SOLCA	Sociedad de lucha contra el cáncer
VSG	Velocidad de sedimentación globular

TABLA DE CONTENIDO

LISTADO DE FIGURAS

LISTADO DE TABLAS

CAPÍTULO 1

1. INTRODUCCIÓN

1.1. Antecedentes

La incidencia del cáncer ha aumentado a nivel mundial en los últimos años, lo que lo ha convertido en un importante problema de salud pública. Sin embargo, este incremento es atribuido a varios factores, como el estilo de vida, los cambios demográficos y los avances en la medicina que han mejorado la capacidad de diagnóstico. (Tanday, 2015)

Asimismo, la mejora en registros y recopilación de datos sobre el cáncer también han influido. Según la evaluación realizada con las cifras de incidencia y mortalidad del Observatorio Mundial del Cáncer (GLOBOCAN), se prevé que habrá un aumento de 28,4 millones de casos en la carga mundial del cáncer para el año 2040.(Sung et al., 2021)

Lo que pone en evidencia la necesidad de estrategias integrales de salud pública para gestionar y mitigar su impacto, especialmente en las poblaciones de alto riesgo y en las regiones con recursos limitados.

Tal es el caso de Ecuador en donde los estudios sobre incidencia, prevalencia y mortalidad de las Neoplasias son limitados. Sin embargo, los datos de mortalidad del 2019 indican que las neoplasias hematológicas, incluidos los linfomas de células T, afectan predominantemente a los adultos mayores, con una tasa bruta de mortalidad de 8,49 por cada 100 000 habitantes. (Garrido et al., 2021)

Destacando así la trascendencia del uso de las herramientas estadísticas, como una pieza importante en la investigación médica, ya que ayudan a fundamentar las decisiones clínicas. Tal como el análisis de supervivencia, que desempeña un papel crucial al examinar el impacto de los tratamientos o las covariables en el momento en que se produce un evento determinado, ya que ofrece información sobre cómo los diferentes factores ya sean clínicos o demográficos logran influir en la probabilidad de que ocurra el evento de interés.(Dugard et al., 2022)

De igual importancia tenemos al análisis de correspondencia múltiple (ACM) que es una metodología sólida y ampliamente utilizada en el ámbito de la investigación, que sirve para la identificar asociaciones entre diversas variables categóricas. Por ejemplo, en una investigación sobre la incidencia del cáncer en Lleida, se empleó el ACM para investigar las relaciones entre varios tipos de cáncer, diferentes grupos etarios, género y áreas poblacionales, lo que llevó a identificar asociaciones notables, como la influencia sustancial del cáncer colorrectal en los hombres de 80 años o más, así como el predominio del cáncer de pulmón entre las mujeres residentes en áreas urbanas dentro de la misma categoría de edad. (Florensa et al., 2021)

1.2. Descripción del problema

En términos generales, los linfomas no Hodgkin (LNH), que incluyen a los linfomas B, T y NK, según la (OMS, 2022) constituyen el décimo cáncer con mayor incidencia en hombres y mujeres (ver figura 1.1), siendo una de las neoplasias hematológicas malignas más comunes en todo el mundo, formando los linfomas B el subgrupo de mayor frecuencia y donde aproximadamente el 45% de ellos fallecieron. Por otro lado, los linfomas T son un grupo heterogéneo y menos frecuente.(Fitzmaurice et al., 2019)

Para América latina y el caribe el LNH ocupa el octavo lugar con 43128 casos reportados en 2022, en el caso de Ecuador tiene tasa bruta de Incidencia estandarizada por edad de 9.3 por cada 100 000 personas y una tasa de mortalidad de 4.6 ocupando el cuarto puesto de 32 países. (ver figura 2.1)

La supervivencia de los pacientes con linfoma T varía considerablemente en función de diversos factores, como el sexo, la edad, el estadio de la enfermedad, el tipo de tratamiento y las características biológicas del tumor.

A pesar de los avances en la comprensión de la biología subyacente de estas enfermedades, la identificación de factores predictivos precisos y biomarcadores asociados sigue siendo un área de investigación activa y crítica. (Rabasa, 2009)

El análisis del tiempo hasta un evento es un componente crucial en diversos campos de investigación, como la medicina, la epidemiología y las ciencias sociales. En este contexto, las curvas de supervivencia se consolidan como herramientas estadísticas de gran utilidad y la estimación puede realizarse por técnicas no paramétricas como las curvas de Kaplan-Meier. (Abraira, 1996)
Además, el análisis de la interacción entre variables y el desarrollo de modelos predictivos son aspectos fundamentales en diversas áreas del conocimiento. Las técnicas estadísticas multivariantes emergen como herramientas útiles para este fin, ya que posibilitan el estudio simultáneo de múltiples variables y la identificación de relaciones complejas entre ellas. (Shiker, 2012)

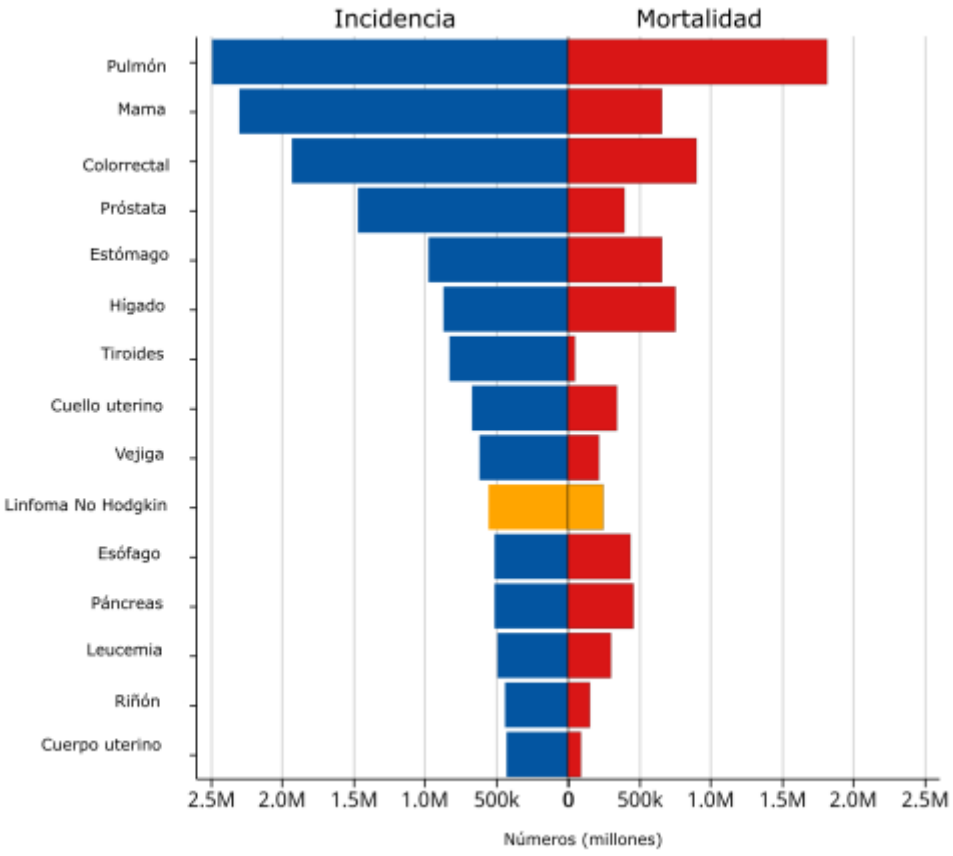


Figura 1.1 *Incidencia y Mortalidad del top 15 a nivel mundial de cáncer.*

Fuente: Cáncer TODAY | IARC- https://gco.iarc.who.int/today

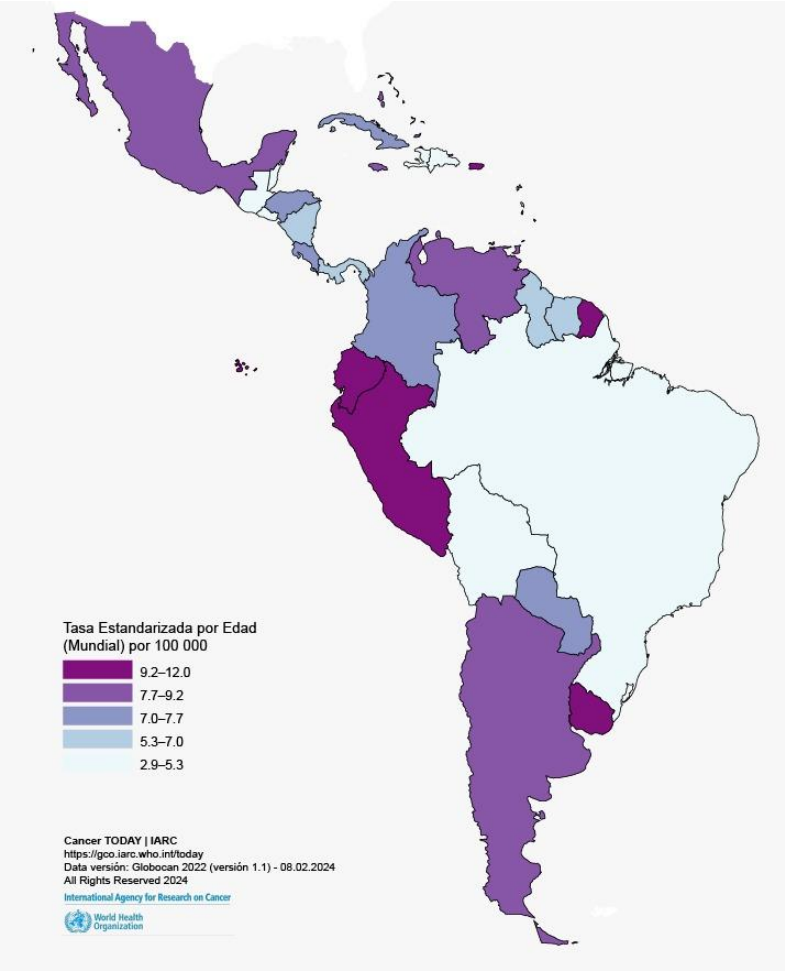


Figura 2.1 *Tasa estandarizada por edad de la incidencia de cáncer en América latina y el caribe.*

Fuente: Cáncer TODAY | IARC- https://gco.iarc.who.int/today

1.3. Objetivos

1.3.1. Objetivo General

Analizar la asociación entre diferentes variables clínicas, patológicas y demográficas de pacientes con Linfomas no Hodgkin T, utilizando técnicas Estadísticas Multivariantes.

1.3.2. Objetivos Específicos

- Determinar la sobrevida global de pacientes con linfomas T de acuerdo con sus características clínicas.
- Identificar y visualizar las relaciones entre los grupos de variables y sus categorías, mediante análisis de correspondencias múltiples.
- Aplicar técnicas de agrupación para establecer perfiles de riesgo en los pacientes con Linfoma no Hodgkin T, con el fin de guiar las decisiones clínicas.

1.4. Alcance

El estudio se enfocará en pacientes con diagnóstico de LNH-T confirmado, registrados en el Instituto Oncológico Nacional SOLCA Guayaquil en el periodo 2000-2018, que contiene datos sobre variables clínicas, demográficas, entre otras.

Se utilizarán técnicas estadísticas multivariantes para:

- Identificar asociaciones entre las variables.
- Desarrollar un modelo de la supervivencia de los pacientes.
- Evaluar la utilidad del modelo para la toma de decisiones clínicas.

CAPÍTULO 2

2. MARCO TEÓRICO

2.1. Linfomas

Las enfermedades hematológicas malignas comprenden una variedad de cánceres que afectan a la sangre, la médula ósea y los órganos linfoides, se clasifican como como leucemias, linfomas, neoplasias mieloproliferativas, discrasias de células

plasmáticas, tumores histiocíticos y neoplasias de células dendríticas. (Le Gall-Ianotto & Misery, 2016)

A lo largo del tiempo, se han propuesto diversos sistemas de clasificación para estas enfermedades, si bien la clasificación aceptada es la actualización de 2022 del sistema de clasificación de los linfomas realizada por la Organización Mundial de la Salud que demuestra los avances en este campo al clasificar los linfomas.

Estas neoplasias malignas suelen comenzar en los ganglios linfáticos, aunque tienen la capacidad de diseminarse a otros órganos linfáticos y sitios fuera de los ganglios (extraganglionares). La clasificación de los linfomas se divide ampliamente en linfoma de Hodgkin (LH) y linfomas no hodgkianos (LNH); estos últimos constan de más de 70 subtipos clasificados según el tipo de célula, las características físicas y el inmunofenotipo. Surgiendo de diferentes linajes celulares como células B, células T o células T N/K Figura 2.1. La incidencia del LNH ha aumentado considerablemente desde la década de 1960, especialmente entre las personas de edad avanzada. (Vinnicombe & Garg, 2023) Según la OMS (2022) constituyen el décimo cáncer con mayor incidencia.

En base a lo descrito por el Instituto Nacional de Cáncer (NCI) en su diccionario para términos asociados al cáncer, los LNH de células T son un cáncer que se origina en los linfocitos T, que son células del sistema inmunitario. Existen varios subtipos de LNH-T, como la micosis fungoide (MF), el linfoma anaplásico de células grandes (LACG), entre otros.(*NCI*, 2011)

En general, los linfomas cutáneos de células T, que incluyen la MF y el síndrome de Sezary (SS), se manifiestan principalmente en la piel y se distinguen por la infiltración de células T monoclonales malignas.(Dummer et al., 2021) Asimismo, suelen presentar un curso clínico más lento o inclusive indolente. Sin embargo otras localizaciones como ganglios linfáticos o tracto gastrointestinal tienden a ser más agresivos con una progresión más rápida y presencia de síntomas B. (Varghese & Alsubait, 2024)

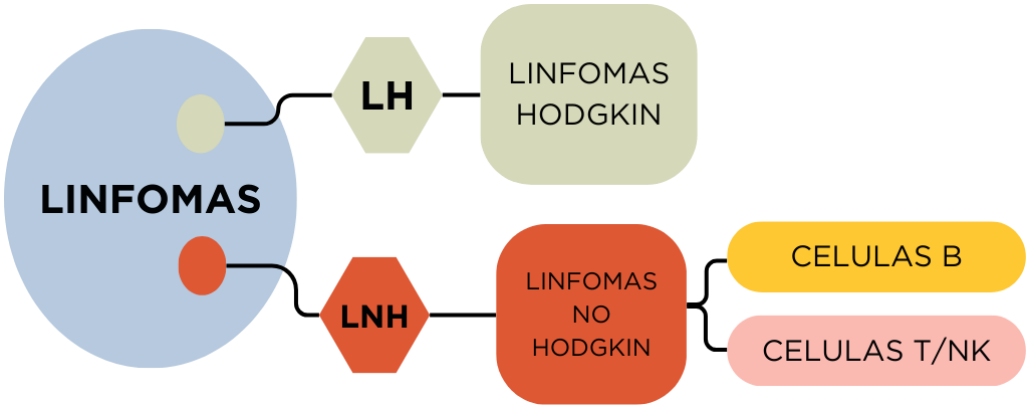


Figura 2.1 *Esquematización de los linfomas*

2.2. Modelos de Supervivencia

Estos modelos son particularmente útiles en la investigación médica, el que se emplea con frecuencia es el modelo de Kaplan-Meier, que se utiliza para estimar la función de supervivencia basándose en la información sobre la esperanza de vida y ayudar a evaluar las probabilidades de eventos en varios momentos. Otro modelo que suele usarse es el modelo de riesgos proporcionales de Cox, que investiga la correlación entre la duración de la supervivencia y una o más variables predictoras, lo que permite evaluar la forma en que los diversos factores influyen en la probabilidad de incidencia de los eventos. (Kovalchuk et al., 2023)

2.2.1. Kaplan Meier

La estimación de la supervivencia en el campo de la medicina puede estimarse mediante técnicas no paramétricas (curvas de Kaplan Meier). El modelo que detalla (Kaplan & Meier, 1958) es una herramienta útil para analizar el tiempo que tarda un evento en ocurrir, como la muerte o la recuperación de una enfermedad. Estas curvas nos muestran la probabilidad de que el evento ocurra en cada momento, junto con un indicador de la precisión de esa estimación (error estándar). Además, muestran información importante sobre los participantes en el estudio: Cuántos participantes siguen en el estudio, cuantos experimentaron el evento en cada momento y cuantos abandonaron el estudio antes de que finalizara.

Sea $S(t)$ la probabilidad de que un participante viva más allá de un tiempo t, en una muestra de la población de tamaño N, sean

$$t_{1\leq}t_{2\leq}t_{3\leq\cdots\leq}t_N \quad (1)$$

los tiempos que transcurren hasta la muerte de cada participante.

El estimador de Kaplan-Meier:

$$\hat{S}(t) = \prod_{t_i<t} \frac{n_i - d_i}{ni} \quad (2)$$

donde

d_i, es el número de eventos hasta el momento t_i

ni, cantidad de sujetos en riesgo antes de t_i

Existen elementos claves a considerar para realizar el modelo:

La fecha o momento de inicio del estudio, desde donde se empieza a incluir los pacientes. La fecha de entrada que será diferente para cada paciente, ya que es el momento en donde fue incluido, la cual es muy importante porque a partir de ahí se calcula el tiempo de supervivencia.

La fecha de la última observación, que puede ser hasta donde ocurre el evento de interés o hasta el último contacto con el paciente, de la misma manera es diferente para individuo. Los participantes que abandonan el estudio, se pierde el contacto o no presentan el evento de interés se consideran como “censurados”. (Martínez & Pérez, 2023)

La curva de Kaplan-Meier, caracterizada por un patrón escalonado que disminuye con cada aparición del evento de interés, sirve como representación visual de la probabilidad de supervivencia a medida que pasa el tiempo y es un componente fundamental en el campo del análisis de supervivencia. (Andrade, 2023)

Es difícil comparar la supervivencia entre dos grupos de forma directa debido a que es casi imposible encontrar dos grupos que sean completamente homogéneos en todos los factores que podrían afectar al resultado. Si se lograra tal homogeneidad, las muestras serían demasiado pequeñas para obtener resultados confiables. Los estudios de factores que condicionan la supervivencia pueden llevar a

interpretaciones erróneas si estos factores están interrelacionados entre sí (Fernández et al., 1996)

2.2.2. Regresión Cox

Los modelos de regresión son una herramienta útil para analizar la supervivencia, se emplean para examinar los datos relacionados con la duración hasta que ocurre un evento específico, particularmente cuando el evento de interés es binario y puede o no haber ocurrido al final del período de observación. Estos modelos estadísticos son muy valiosos en la investigación médica para evaluar la influencia de varios factores en las probabilidades de supervivencia de los pacientes, ya que permiten incorporar datos censurados, que incluyen información de personas que aún no han experimentado el evento al final del estudio o que lo han abandonado. (Andrade, 2023)

El modelo de regresión propuesto por Cox (Cox, 1972), también conocido como modelo proporcional, es un tipo de modelo semiparamétrico por no especificar la distribución de la función de riesgo . Sirve para evaluar el impacto de las covariables en la función de riesgo. Un elemento esencial del modelo de Cox es la función de riesgo, que se expresa como una combinación de una función de riesgo de referencia $h_0(t)$ y una función exponencial $e^{\beta \mathbb{X}}$ de las p covariables. Esta formulación permite calcular los coeficientes de riesgo sin necesidad de especificar el peligro de referencia. (Ramírez et al., 2017)

$$h(t;\mathbb{X}) = h_0(t)e^{\beta \mathbb{X}} = h_0(t)e^{(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_{p)}} \quad (3)$$

El modelo de riesgos proporcionales de Cox calcula la razón de riesgo (HR) para un parámetro de evaluación particular vinculado a ciertos factores de riesgo, incluidas variables continuas como la edad o variables categóricas como el sexo. Una premisa esencial del modelo de Cox es la proporcionalidad de los riesgos, lo que

indica que el impacto de varias variables en la supervivencia se mantiene constante a lo largo del tiempo y se acumula en una escala específica. (Abd ElHafeez et al., 2021)

2.3. Análisis de Correspondencia Múltiple

El análisis de correspondencia múltiple (ACM) es una técnica estadística de análisis exploratorio, que permite visualizar las relaciones entre múltiples variables en un espacio de baja dimensión, generalmente bidimensional, facilitando la interpretación visual de los datos. Esta técnica representa una expansión del análisis de componentes principales (ACP), pero está diseñado específicamente para variables categóricas, con el fin de reducir numerosos conjuntos de variables en unos de menos componentes que capturen la información esencial. (Mori et al., 2016)

Además, el ACM demuestra ser ventajoso cuando se trata de profundizar en las relaciones que existen entre las variables y las asociaciones entre sus respectivas categorías, dando así una perspectiva multidimensional a las similitudes que se puedan presentar entre diferentes individuos. (François & Julie, 2014)

2.4. Clúster Jerárquico

El análisis jerárquico de conglomerados (HCA) integrado con el análisis de correspondencia múltiple (ACM) constituye un marco metodológico sólido empleado para investigar e ilustrar las asociaciones entre variables categóricas, así como para clasificar as observaciones en distintos grupos en función de sus similitudes.

Esta combinación es particularmente útil en campos en los que los datos son categóricos y complejos, como las ciencias sociales, la epidemiología y la investigación de mercado. (Hjellbrekke, 2019)

El HCA, por otro lado, es una herramienta exploratoria de análisis de datos que agrupa observaciones similares en conglomerados, proporcionando una estructura jerárquica que se puede visualizar como un dendograma. (Mathai et al., 2022)

CAPÍTULO 3

3. METODOLOGÍA

El análisis de supervivencia es un conjunto de métodos estadísticos que se utiliza para estudiar el tiempo que tardan los individuos en experimentar un evento, como la muerte o la recuperación de una enfermedad. Esta información es crucial para comprender la historia natural de las enfermedades, evaluar la efectividad de los tratamientos y tomar decisiones informadas sobre la atención médica.

En este caso, se presenta una metodología paso a paso para realizar un análisis de supervivencia completo, desde la limpieza y transformación de variables, la construcción de un modelo multivariante de regresión de Cox, Análisis de Correspondencia múltiple y clúster. Para este estudio se cuenta con datos de pacientes que fueron diagnosticados con Linfomas no Hodgkin T, en una de las instituciones oncológicas del país SOLCA, desde el año 2000 al 2016 con un total de 60 pacientes y 26 características. Entre ellas demográficas, antecedentes, manifestaciones clínicas, exámenes de laboratorio, diagnóstico, tratamiento y estatus o condición final de paciente como vemos en la figura 3.1.

LINFOCITOS <dbl>	ERITROCITOS <dbl>	HTO <dbl>	HB <dbl>	TROMBOCITOS <dbl>	VSG <dbl>	LDH <dbl>	B2 MICROGLOBULINA <dbl>	EDAD <dbl>	SEXO <fctr>	SINTOMAS B <fctr>	ADENOPATÍAS <fctr>	NUM_TERRITORIOS <fctr>
28	4.60	31.0	13.8	262	35.0	156	2.70	38	F	NO	SI	UNO
18	3.50	32.0	12.0	340	27.0	734	8.50	81	M	SI	SI	3 O MAS
16	4.20	28.0	11.0	120	22.0	974	2.70	25	M	NO	SI	DOS
29	4.90	29.0	9.0	420	54.0	180	3.40	53	M	NO	NO	NO
18	5.10	22.0	8.0	343	88.0	542	7.60	70	F	SI	SI	UNO
29	4.43	36.0	12.0	262	54.5	184	1.80	41	M	NO	NO	NO
9	4.23	34.0	11.0	718	72.0	596	3.51	58	F	SI	SI	3 O MAS
33	3.77	32.0	10.8	200	58.0	248	4.20	72	M	SI	SI	UNO
20	4.10	37.0	11.0	224	55.0	232	6.10	28	F	SI	SI	UNO
31	4.90	42.0	13.0	190	24.0	169	2.14	48	F	NO	NO	NO
6	3.20	25.0	8.3	288	55.0	232	2.10	95	F	NO	SI	UNO
18	4.10	36.0	11.8	281	65.0	335	1.96	52	F	SI	SI	UNO
63	4.70	37.0	12.6	294	87.0	989	8.20	18	F	SI	SI	DOS
29	3.20	28.0	9.4	20	98.0	2058	8.50	13	M	SI	SI	3 O MAS
33	5.40	46.0	15.8	211	10.0	188	4.30	19	M	SI	SI	3 O MAS
29	4.43	36.0	12.0	262	54.5	234	2.80	59	F	NO	NO	NO
30	3.50	36.0	12.0	40	57.0	235	3.20	76	M	SI	SI	UNO
47	2.44	23.0	8.0	233	77.0	650	3.90	72	F	SI	SI	3 O MAS
9	4.60	38.6	12.3	200	85.0	441	8.50	82	F	SI	SI	3 O MAS
8	3.99	34.8	11.6	659	85.0	659	9.10	60	M	SI	SI	UNO

Figura 3.1 *Datos iniciales con las variables numéricas y nominales*

3.1. Limpieza y transformación de variables:

3.1.1. Revisión de datos:

Realizamos una exploración de los datos para identificar posibles errores, inconsistencias y valores perdidos. Se identificaron muchos datos faltantes por lo cual efectuamos una búsqueda en la historia clínica de los pacientes y se actualizaron los registros, con lo que se logró disminuir de aproximadamente un 30% a un 6% de data faltante.

De ese porcentaje de datos faltantes procedimos a imputar aquellos que eran numéricos por su mediana, y para las categóricas mediante el algoritmo MCA iterativo, tomando en consideración las variables que nos ayudarían a explicar el comportamiento de este tipo de lesiones con fundamentos clínicos.

3.1.2. Análisis exploratorio:

Se realizo un análisis descriptivo de las variables para comprender su distribución, rango y posibles relaciones entre ellas. Esto puede ayudar a identificar variables irrelevantes o redundantes que podrían afectar el modelo. Para las variables de laboratorio o numéricas se realiza una tabla resumen para ver su mínimo, máximo, media, etc. Por otro lado, para las cualitativas se realizaron gráficos de barras para identificar las frecuencias.

3.1.3. Transformación de variables:

Al explorar las variables numéricas se logra identificar que el rango es muy amplio ya que muchas de ellas son exámenes de laboratorio, lo que sugiere que existen muchos valores atípicos por ende muchos de ellos se verificaron para descartar el error en el ingreso de datos (ver Anexo1). Generando cierto ruido al intentar modelarlas, en base a esto se toma la alternativa de categorizar dichas variables.

Codificamos las variables continuas en variables categóricas como se observa en la figura 3.2, esto precisamente se lo realiza para las variables de laboratorio utilizando “los rangos normales” para cada uno de los indicadores. Creamos así 3 niveles.

“BAJO”, “NORMAL” y “ALTO”, para otras únicamente “ELEVADO” y “NORMAL” ya que clínicamente se conoce que mucho de estos indicadores tiende a elevarse ante la presencia de una neoplasia, y para aquellas que tienden a disminuir o bajar sus valores de los normales “BAJO” y “NORMAL”.

La variable edad se dividió en 4 grupos etarios, considerando que los pacientes reportados eran desde niños hasta adultos mayores. Además, algunas variables que ya eran nominales, se procedió a recodificarlas con el fin de resaltar la información que podría aportar a una mejor compresión.

CAT_LEU <fctr>	CAT_LIN <fctr>	CAT_ERI <fctr>	CAT_HTO <fctr>	CAT_HB <fctr>	CAT_TMB <fctr>	CAT_VSG <fctr>	CAT_LDH <fctr>	CAT_B2 <fctr>	CAT_EDAD <fctr>
NORMAL	NORMAL	NORMAL	BAJO	BAJO	NORMAL	ELEVADO	NO ELEVADO	ELEVADO	25–45
NORMAL	BAJO	BAJO	BAJO	BAJO	NORMAL	ELEVADO	ELEVADO	ELEVADO	Mayor A 65
BAJO	BAJO	NORMAL	BAJO	BAJO	BAJO	ELEVADO	ELEVADO	ELEVADO	25–45
ALTO	NORMAL	NORMAL	BAJO	BAJO	ALTO	ELEVADO	NO ELEVADO	ELEVADO	45–65
NORMAL	BAJO	NORMAL	BAJO	BAJO	NORMAL	ELEVADO	ELEVADO	ELEVADO	Mayor A 65
NORMAL	NORMAL	NORMAL	BAJO	BAJO	NORMAL	ELEVADO	NO ELEVADO	NO ELEVADO	25–45
NORMAL	BAJO	NORMAL	BAJO	BAJO	ALTO	ELEVADO	ELEVADO	ELEVADO	45–65
NORMAL	NORMAL	BAJO	BAJO	BAJO	NORMAL	ELEVADO	NO ELEVADO	ELEVADO	Mayor A 65
NORMAL	NORMAL	NORMAL	BAJO	BAJO	NORMAL	ELEVADO	NO ELEVADO	ELEVADO	25–45
NORMAL	NORMAL	NORMAL	NO BAJO	BAJO	NORMAL	ELEVADO	NO ELEVADO	ELEVADO	45–65

Figura 3.2 *Transformación de variables continuas en categóricas*

Finalmente se dividieron en 4 grupos de variables de interés para el análisis de la supervivencia, tomando como variable dependiente la condición del paciente.

Figura 3.3 *Estructura de los datos*

Tabla 1: Descripción de las variables

Variable	Descripción de la variable
Sexo	Características biológicas y fisiológicas que definen a hombres y mujeres
Edad	Años cumplidos del paciente al momento del diagnóstico
APF	Antecedentes patológicos familiares son el registro de las enfermedades y afecciones que se han dado en su familia
APP	Antecedentes patológicos personales
Enf. Auto	Afección por la que el sistema inmunitario del cuerpo ataca los tejidos sanos propios.
Adenopatías	Si el paciente presenta un ganglio linfático grande, hinchado o inflamado.
Sitio de biopsia	Parte del cuerpo de donde se extrae la muestra
Numero de territorios ganglionares	Partes del cuerpo donde el paciente presenta ganglios fuera del tamaño normal
Tamaño de ganglio	Tamaño del ganglio con mayor valor
Síntomas B	Si el paciente ha presentado un síntoma como, fiebre, vómito, fatiga
Hepatomegalia	Hígado agrandado
Esplenomegalia	Agrandamiento del bazo
Leucocitos	Tipo de glóbulo que se produce en la médula ósea y se encuentra en la sangre y el tejido linfático.
Linfocitos	Tipo de glóbulo blanco que es parte del sistema inmune.
Eritrocitos	Tipo de glóbulo sanguíneo que se usa para determinar la presencia de afecciones como la anemia, la deshidratación, la desnutrición y la leucemia

Hematocrito	Cantidad de sangre total compuesta de glóbulos rojos.
Hemoglobina	Proteína del interior de los glóbulos rojos que transporta oxígeno desde los pulmones a los tejidos y órganos del cuerpo.
Trombocitos	Glóbulos sanguíneos pequeños esenciales para la coagulación de la sangre.
Velocidad de sedimentación globular	Indica la presencia o remisión de una inflamación o infección.
Lactato deshidrogenasa	Indica que existe una lesión celular.
Beta 2 Microglobulina	Empleada como marcador tumoral.
TIPO DX	Si la afección está presente en los ganglios linfáticos.
DX Final	Diagnóstico final al que se llega después de obtener los resultados de pruebas, como análisis de sangre y biopsias.
Tratamiento	Medio que se empleó para tratar la afección.
Tipo de TX	El tratamiento que usted recibirá depende de su tipo de cáncer y de lo avanzado que esté.
Resp. TX	Como respondió el paciente con el tratamiento aplicado.
Estatus Final	Condición final del paciente que indica si el paciente falleció o sigue vivo hasta la última fecha de contacto.

3.2. Análisis univariante de su pervivencia con Kaplan-Meier:

3.2.1. Estimación de la Curva mediante Kaplan-Meier:

Se realizó la estimación de la curva mediante el modelo de Kaplan-Meier, para visualizar la probabilidad de supervivencia a lo largo del tiempo, el cual para fines de

este análisis se consideró en meses. Calculando desde que el paciente fue diagnosticado hasta su fecha de fallecimiento, fecha de última consulta o la fecha hasta donde se tuvo contacto y para las observaciones donde no se cumple el evento de interés las identificamos como censura.

3.2.2. Comparación de curvas entre grupos:

Se utilizaron pruebas estadísticas, como la prueba de log-Rank, para comparar las curvas de supervivencia entre grupos y determinar si existe una diferencia significativa en la supervivencia. Además de graficar y evidenciar como cambia la probabilidad y tiempo en base a los factores y las categorías o estratos de cada uno.

Hipótesis Nula (H_0): Todas las curvas son equivalentes.

Hipótesis Alternativa (H_1): No todas las curvas son equivalentes.

3.3. Análisis multivariante de supervivencia con el modelo de Cox:

3.3.1. Selección de variables:

Identificamos las variables explicativas más relevantes para incluir en el modelo de Cox. Inicialmente se utilizó la selección de variables paso a paso (Stepwise) para regresión de Cox. Aplicando la selección bidireccional, tomando como criterio de ajuste el de Akaie (AIC), además de considerar los criterios médicos en la cada una de las variables.

3.3.2. Modelo de Cox:

Se construyó un modelo de regresión de Cox para estimar el riesgo relativo de los pacientes fallecidos en función de las variables explicativas seleccionadas. Se interpretan los coeficientes de regresión para comprender cómo cada variable afecta el riesgo de supervivencia.

3.3.3. Evaluación del modelo:

Evaluamos el rendimiento del modelo de Cox utilizando medidas como la prueba de Schoenfeld para verificar la proporcionalidad de los riesgos.

Hipótesis nula (H_0): Las proporciones de riesgos son constantes a lo largo del tiempo.
Hipótesis alternativa (H_1): Las proporciones de riesgos no son constantes a lo largo del tiempo.

3.4. Análisis multivariante con Análisis de Correspondencia Múltiple

Para este análisis inicialmente se volvieron a tomar todas las variables para ver su comportamiento, sin embargo, los resultados no fueron tan favorables ya que la inercia era muy baja. Por ello nos basamos en las que incluimos en el modelo de Cox y otras que por criterio medico se sabía que podría presentar asociaciones.
Graficamos para ver el aporte de las variables en ambas dimensiones, después las categorías de las variables para ver sus relaciones o asociaciones y por último agregando el diagnóstico como una variable suplementaria e identificar qué características se asocian a cada uno de ellos.

3.5. Clúster jerárquico

En esta técnica nos basamos en los resultados del ACM, con el fin de corroborar que las agrupaciones que percibimos anteriormente ya que este método no necesita que se le indique un numero especifico de agrupaciones. Para caracterizar cada uno de los grupos.

3.6. Interpretación y presentación de resultados:

Interpretación de los resultados: Se interpretan los resultados del análisis en conjunto, considerando los hallazgos del análisis univariante y multivariante. Se identifican las variables que tienen un efecto significativo en la supervivencia y se describen sus relaciones con el riesgo de un evento. Además de las asociaciones presentes en cada variable que se seleccionó.

Presentación de resultados: Se presentan los resultados de manera clara y concisa, utilizando tablas y gráficos. Se comunica la información de manera que sea relevante para los investigadores, profesionales de la salud y otros interesados.

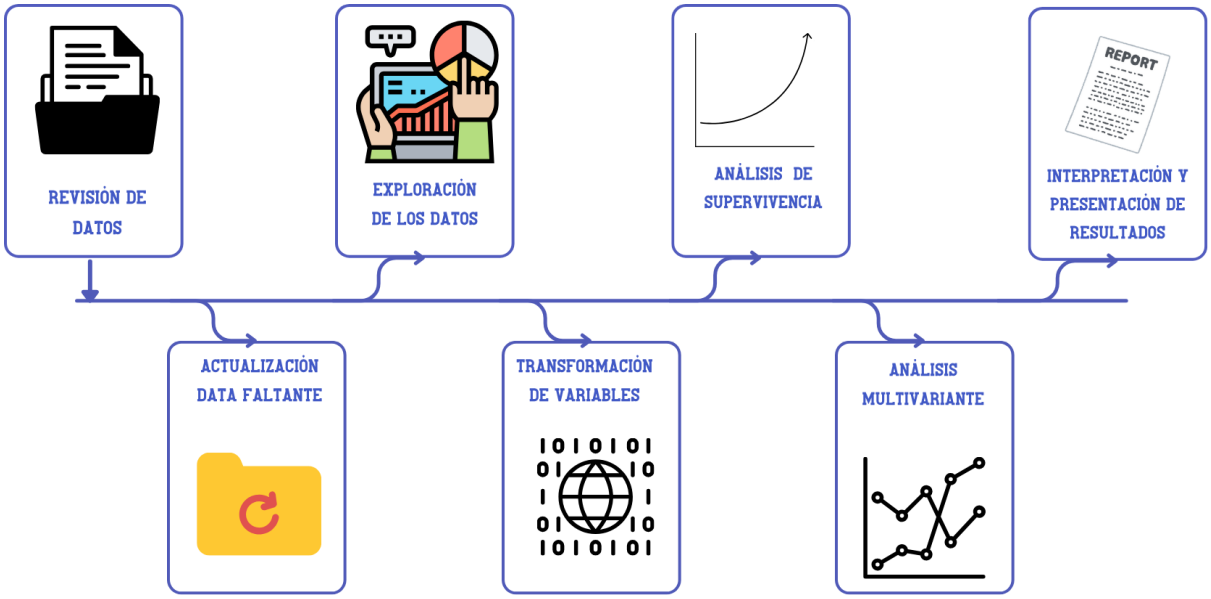


Figura 3.4 *Metodología del trabajo de investigación*

CAPÍTULO 4

4. RESULTADOS

4.1 Análisis Exploratorio

Para la exploración de los datos realizamos gráficos de frecuencias en porcentajes, mostrando de cada uno de los grupos de variables las relevantes. De las demográficas, la figura 4.1 nos muestra que de los pacientes diagnosticados el 60% de ellos son hombres. Además, aproximadamente el 36% de ellos se encuentra en un rango de edad de 45 a 65, seguido por los pacientes con un rango de 25 a 45 años con un 28.33%, representando más del 50% con ambas categorías de edad.

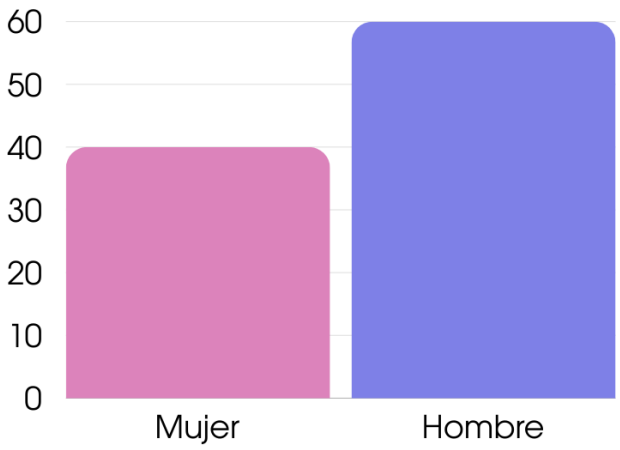


Figura 4.1 *Frecuencia en porcentajes de la variable sexo*

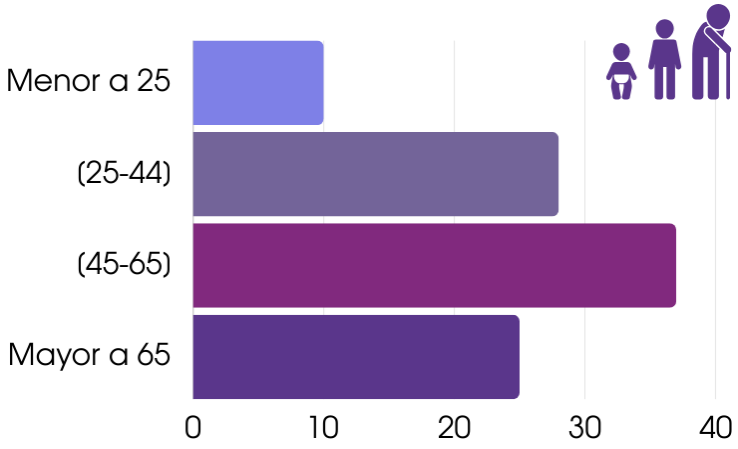


Figura 4.2 *Frecuencia en porcentajes de la variable Edad*

En cuanto al diagnóstico de los pacientes, tenemos según la figura 4.3 el 70% de los pacientes presentaron un diagnóstico de tipo ganglionar es decir que presentaba adenopatías y la biopsia fue realizado en un ganglio linfático.

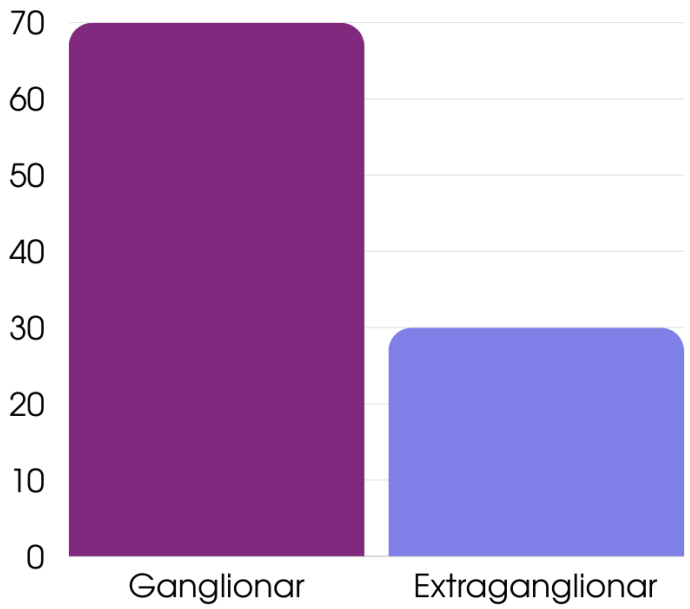


Figura 4.3 *Frecuencia en porcentajes de la variable tipo de diagnostico*

Ahora bien, analizando los diagnósticos vemos en la figura 4.4. que Linfoma T periférico y la Micosis Fungoide son los más comunes, ambos con frecuencias alrededor del 30%. Esto indica que aproximadamente una de cada tres personas diagnosticadas pertenece a una de estas dos categorías.

Sigue en frecuencia Linfoma extranodal NK Tipo Nasal, con un porcentaje menor del 15%, lo que lo convierte en el tercer diagnóstico más común, seguido del Linfoma anaplásico de células grandes ALK negativo. El resto de los diagnósticos son menos frecuentes, representando menos del 5% cada uno.

Por último, el estatus o condición final de los pacientes que se divide en aquellos que fallecieron y los vivos hasta la fecha de ultima consulta o contacto con el paciente, donde el 53% de ellos fueron reportados como fallecidos.

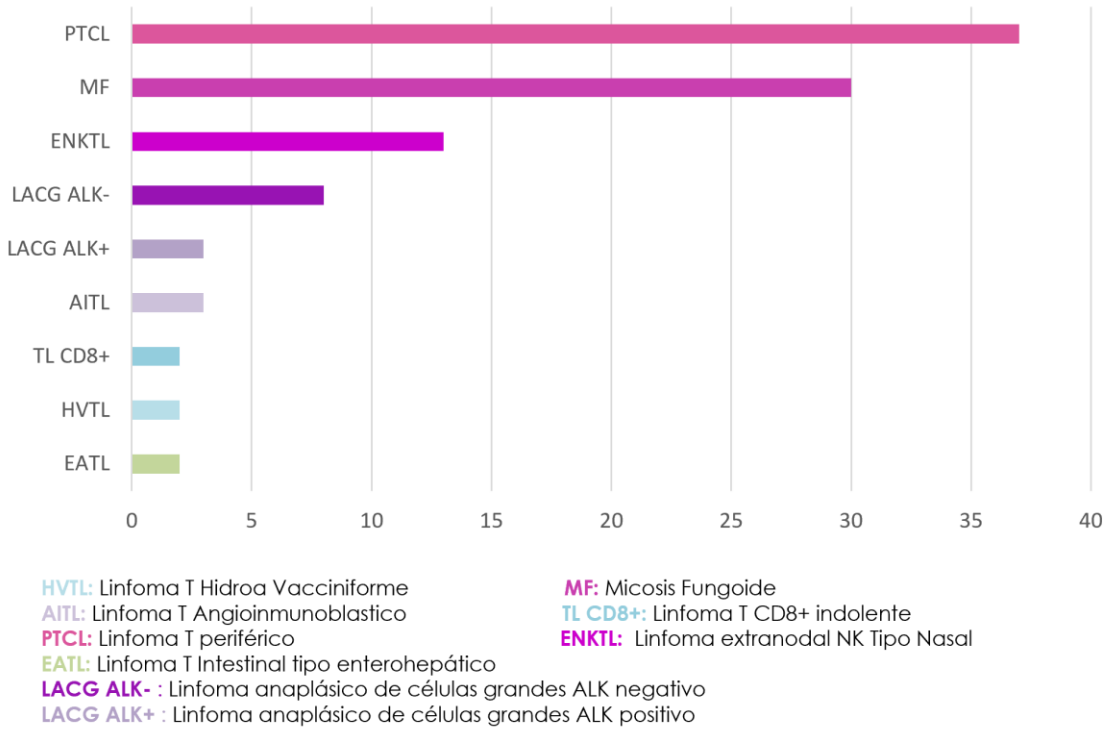


Figura 4.4 *Frecuencia en porcentajes de los diagnósticos.*

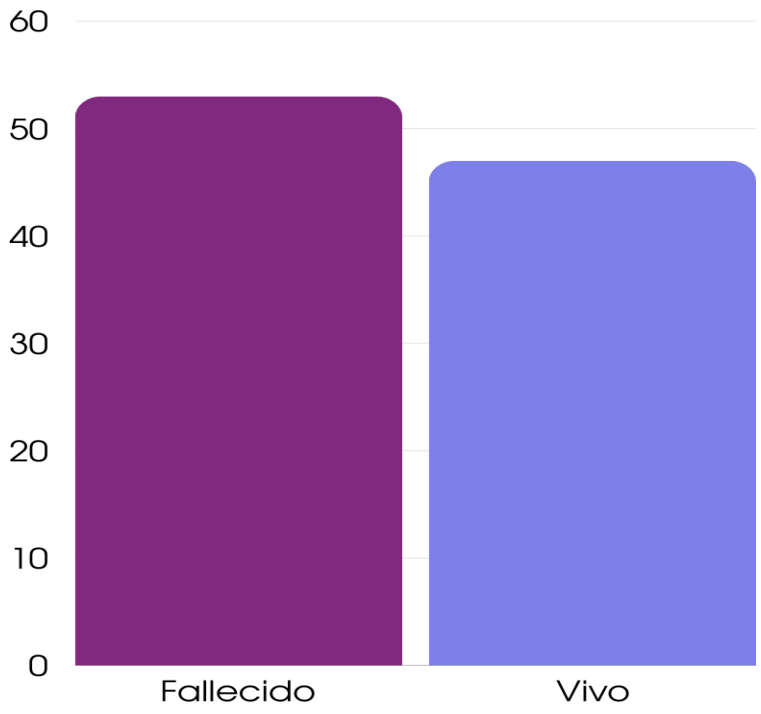


Figura 4.5 *Frecuencia en porcentajes de la condición final de los pacientes*

4.2 Modelo de Supervivencia Kaplan Meier

Para la compresión del análisis de la supervivencia, se realizó la curva de supervivencia mediante la estimación de Kaplan-Meier, que muestra la probabilidad de que un paciente sobreviva a lo largo del tiempo.

Se puede evidenciar que la curva desciende rápidamente en los primeros 10 meses, lo que nos sugiere que una proporción significativa de eventos ocurrió temprano en el seguimiento. Además, identificamos con una cruz los datos censurados, es decir aquellos pacientes que hasta ese momento no presentaron el evento de interés.

El tiempo mediano de supervivencia, en el gráfico está representado por las líneas punteadas, esto parece estar en aproximadamente 30 meses, lo que indica que el 50% de los individuos aún estaban vivos en ese tiempo.

Esto nos ayuda a tener una idea inicial de como la supervivencia de los pacientes diagnosticados con esta Neoplasia cambia considerablemente a través del tiempo. Sin embargo, no podemos dejar a un lado la importancia de otros factores o características que pueden influir en la disminución de la supervivencia de los individuos. Para esto se compararon las probabilidades de supervivencia entre diferentes grupos o características del paciente y así identificar aquellas con aporte significativo.

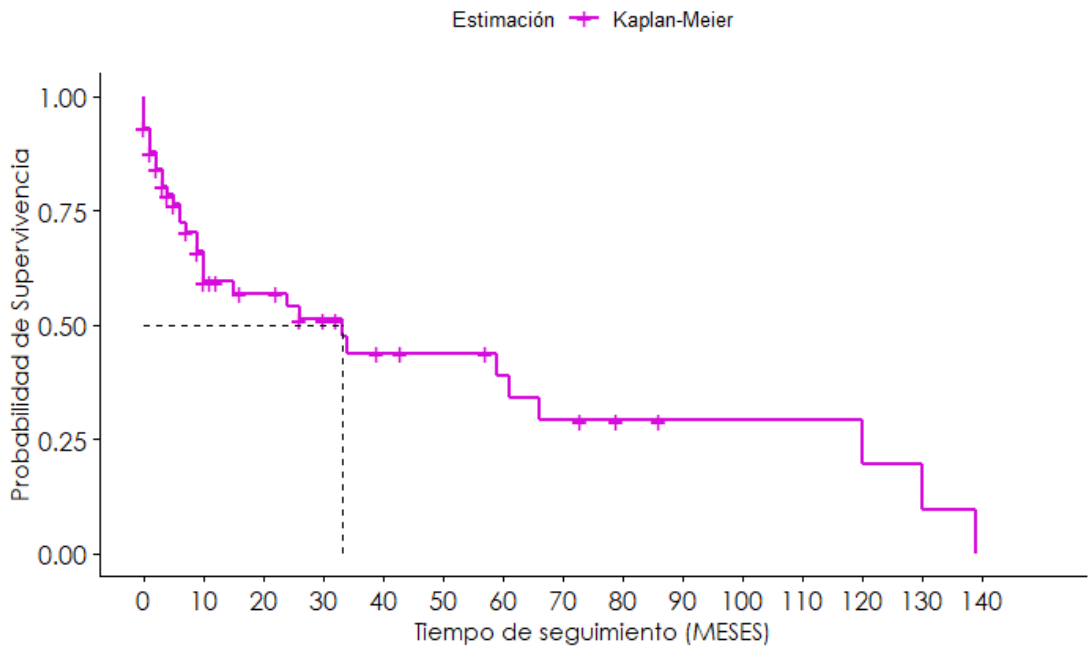


Figura 4.6 *Curva de la estimación de la supervivencia mediante Kaplan-Meier*

4.3 Comparación de curvas entre grupos o factores

Partiendo de la curva inicial, se realizó la integración de variable por variable al modelo con el fin de determinar aquellas con diferencias significativas entre sus grupos o categorías presentes. De las cuales 7 se destacaron, al mostrarse como la probabilidad de supervivencia disminuía drásticamente en algunos grupos.

El gráfico muestra varias curvas de Kaplan-Meier para diferentes grupos etarios, comparando la probabilidad de supervivencia a lo largo del tiempo. El p-valor de 0.00035 indica que hay evidencia estadística para rechazar la hipótesis nula, lo cual nos muestras que no todas las curvas son equivalentes.

Los grupos menores de 25 y mayores de 65 muestran una disminución rápida en la probabilidad de supervivencia en los primeros 10 meses, indicando que los individuos tienen una menor probabilidad de supervivencia comparada con los otros grupos.

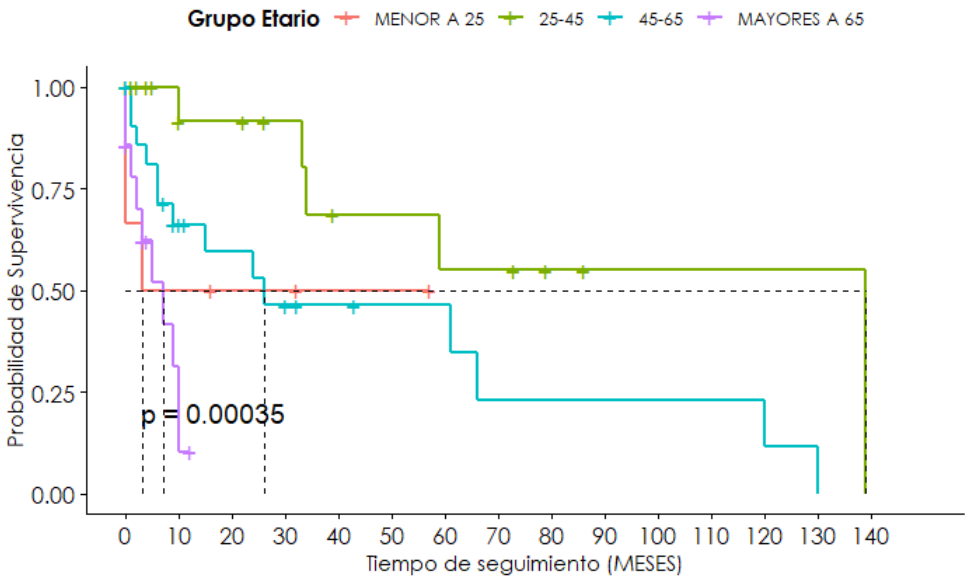


Figura 4.7 *Comparación de las curvas de supervivencia para los diferentes grupos etarios*

La velocidad de sedimentación globular que mide la actividad inflamatoria en el organismo se dividió en 2 grupos: aquellos que tenían un valor por encima de los 20 mm denominados como “elevado” y “normal” los menores iguales a ese valor.

La curva muestra una rápida disminución en la probabilidad de supervivencia en los primeros 20 meses, lo que sugiere que los individuos con VSG elevado tienen una menor probabilidad de supervivencia en comparación con el otro grupo. La mediana de supervivencia para este grupo está entre los 10 y 20 meses.

El p-valor de 0.0051 indica que se rechaza la hipótesis nula, por lo tanto, se asume que hay diferencias en las curvas de supervivencia para los grupos con VSG elevado y normal.

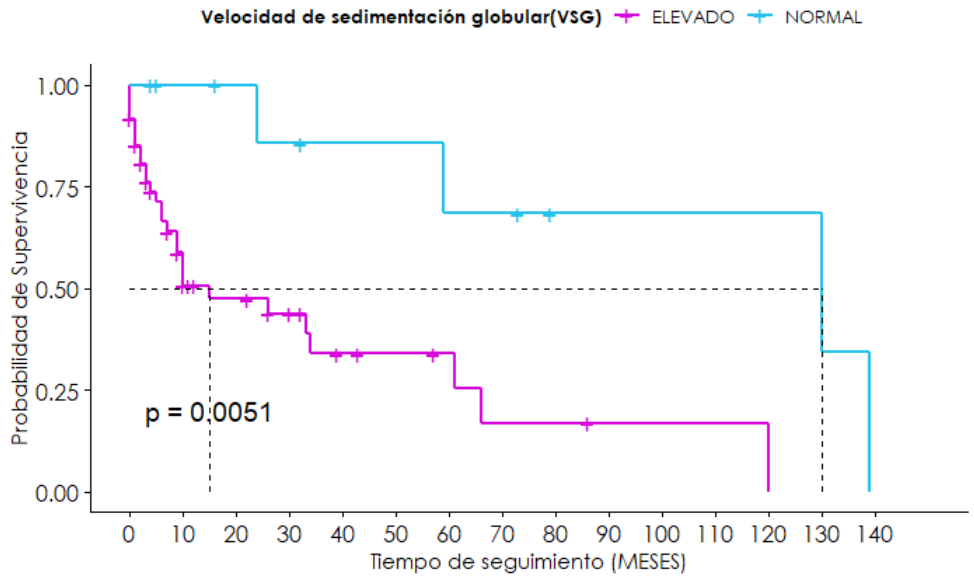


Figura ¡Error! No hay texto con el estilo especificado en el documento.**4.8** *Comparación de las curvas de supervivencia el factor VSG*

La lactato deshidrogenasa (LDH) se utiliza para detectar si existen daños o lesiones en el tejido, se considera aquellos valores superiores a 280 U/L como “ELEVADO”. El gráfico muestra que los niveles de LDH tienen un impacto significativo en la probabilidad de supervivencia.

Aproximadamente a los 10 meses, la probabilidad de supervivencia cae al 50% para el grupo con LDH elevado, lo que sugiere que los individuos tienen una menor probabilidad de supervivencia en comparación con el otro grupo.

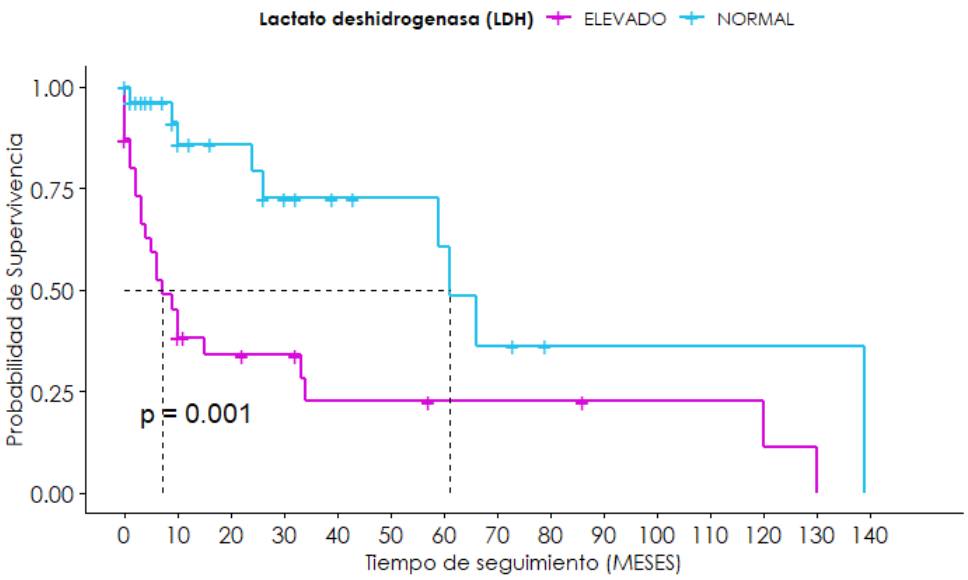


Figura 4.9 *Comparación de las curvas de supervivencia para LDH*

En base al diagnóstico presentado para cada individuo y el lugar donde se realizó la biopsia, se planteó un subtipo para identificar y separar aquellos con linfoma “extraganglionar”. Se denominan así a aquellos que se originan en un órgano diferente al ganglio.

La curva muestra una rápida disminución en la probabilidad de supervivencia en los primeros 10 meses, cae al 50% para el grupo ganglionar en comparación con aquellos con subtipo extraganglionar que desciende de manera gradual. Lo que nos indica que existen diferencias y los individuos con subtipo ganglionar tienen peores resultados.

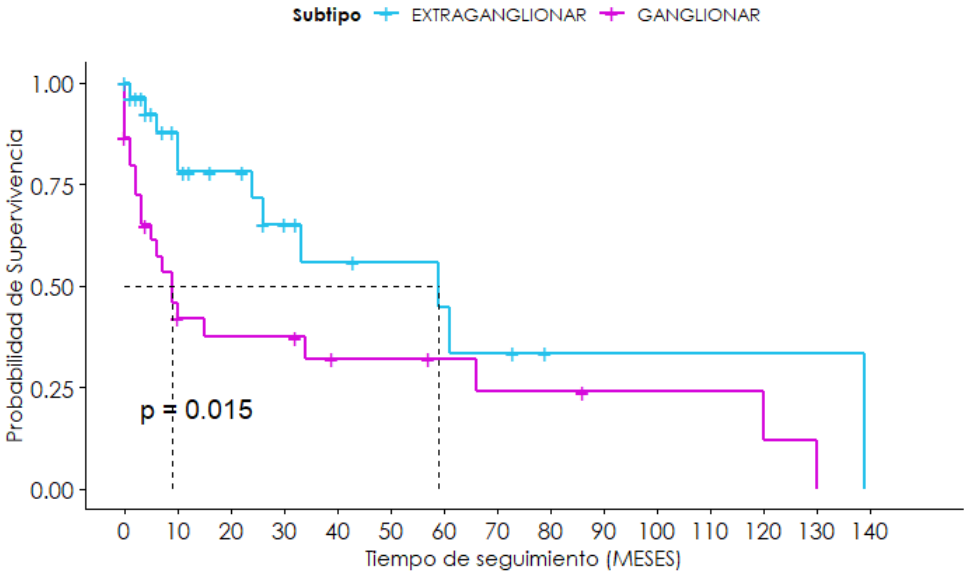


Figura 4.10 *Comparación de las curvas de supervivencia para las categorías de la variable subtipo*

Por último, tenemos 2 condiciones importantes que son la presencia de Hepatomegalia y Esplenomegalia, el gráfico muestra que ambas condiciones tienen un impacto en la probabilidad de supervivencia con los individuos que presentan esta condición, teniendo peores resultados de supervivencia a lo largo del tiempo.

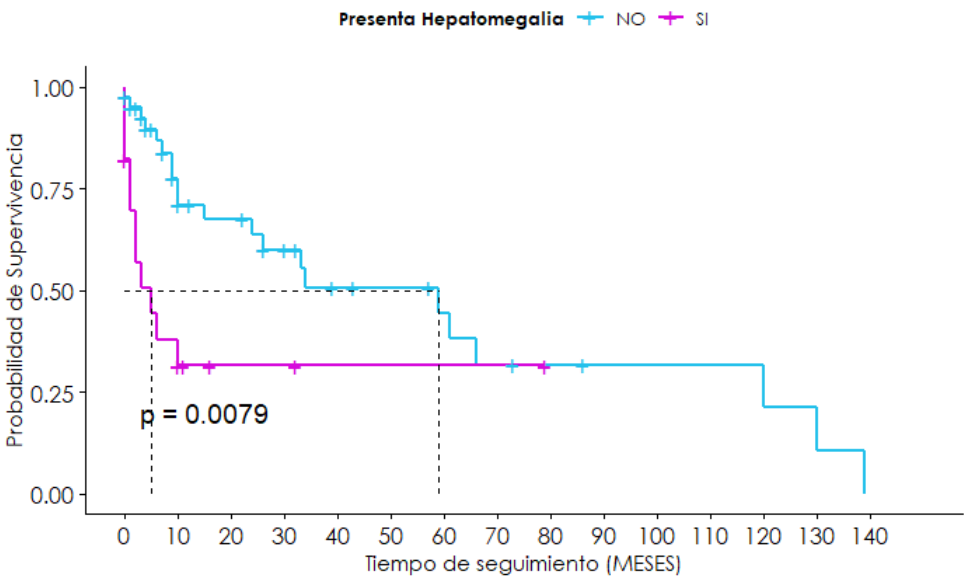


Figura 4.11 *Comparación de las curvas de supervivencia para la presencia de Hepatomegalia*

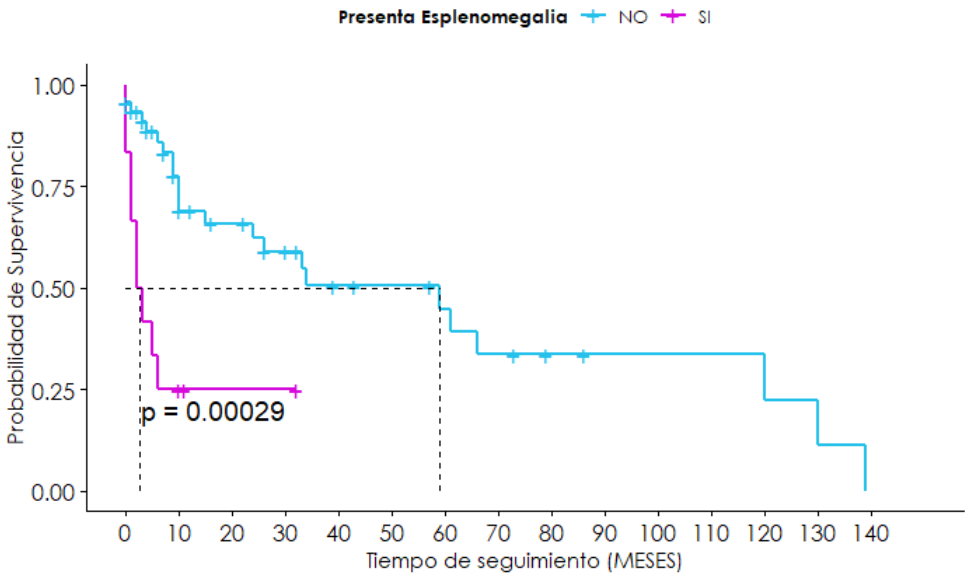


Figura 4.12 *Comparación de las curvas de supervivencia para la presencia de Esplenomegalia en los pacientes*

4.1. Análisis multivariante de supervivencia con el modelo de Cox

Con la finalidad de tener un modelo de fácil interpretación, eficiencia y determinar las covariables que son significativas o relevantes en el modelo se utilizó la selección de variables paso a paso (Stepwise) para regresión de Cox. Aplicando la selección bidireccional, tomando como criterio de ajuste el de Akaie (AIC), es decir que va agregando o eliminando variables en cada paso según mejore el criterio de ajuste.

Dando un modelo con buen ajuste conformado por 5 variables de las cuales 4 de ellas eran significativas, nos muestra la figura que los pacientes con hepatomegalia se asocian con un riesgo 4.887 veces mayor de ocurrencia del evento en comparación con aquellos que presentan este signo. De la misma manera tener niveles bajos de HB se asocia con un riesgo 5.881 veces mayor comparado con niveles normales de HB.

La VSG elevada se asocia con un riesgo 8.988 veces mayor, los niveles elevados de LDH se asocian con un riesgo 3.032 veces mayor, la B2-microglobulina se asocia a

un riesgo 5.754 veces, aunque en el modelo no muestre esta variable como significativa, evidenciamos que si existe mayor riesgo de muerte para los individuos que tienen niveles elevados. Por ultimo los individuos entre 25 y 45 años tienen un riesgo significativamente menor (HR = 0.095) de ocurrencia del evento en comparación con los menores de 25 años.

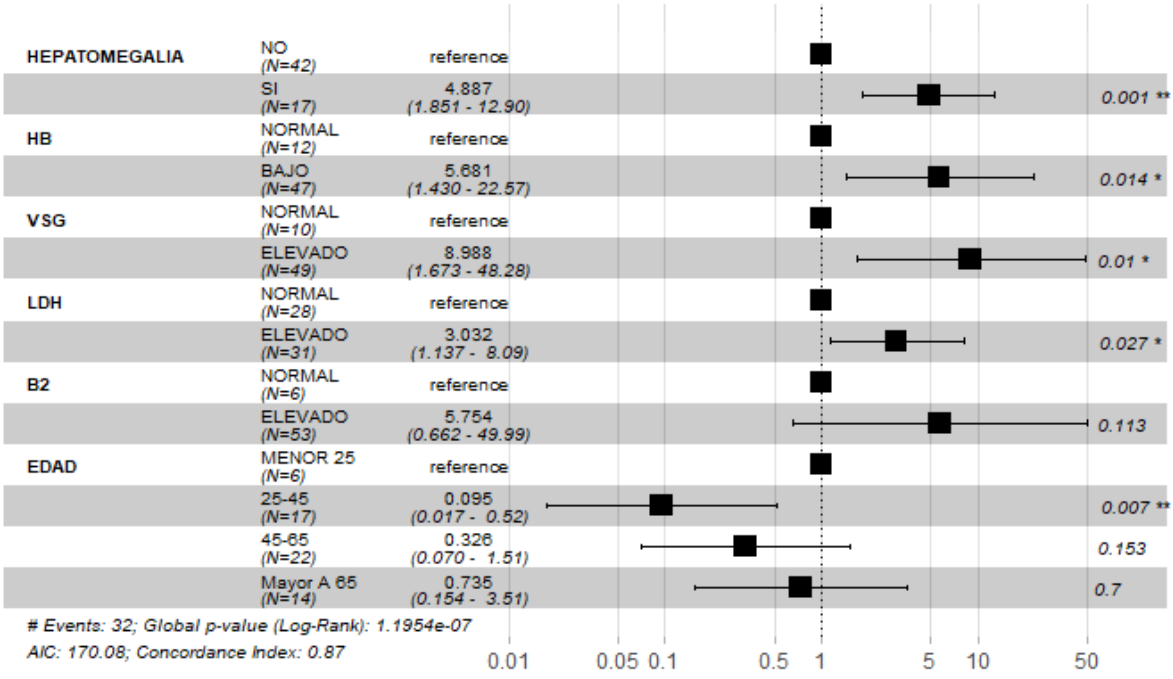


Figura 4.13 *Resumen del modelo de regresión Cox de la supervivencia con las variables de la selección paso a paso*

A pesar de que el modelo nos muestra resultados relevante y buen ajuste de las variables, viola el supuesto inicial de riesgo proporcional de una regresión Cox el cual lo verificamos utilizando medidas como la prueba de Schoenfeld.

Para la Hepatomegalia el valor p es 0.665, lo que indica que no hay evidencia significativa para rechazar el supuesto de riesgos proporcionales. Sin embargo, vemos que para la HB y la Edad presentan valores p menores a 0.05 lo que indica que hay evidencia significativa para rechazar el supuesto de riesgos proporcionales.

Tabla 2: *Prueba de Riesgos proporcionales para las covariables*

Variable	chisq	df	p
HEPATOMEGALIA	0.188	1	0.665
HB	4.786	1	0.029
VSG	0.214	1	0.644
LDH	0.651	1	0.42
B2	1.924	1	0.165
EDAD	9.122	3	0.028
GLOBAL	15.38	8	0.052

Partiendo de este modelo, se consideró la relevancia clínica de las variables seleccionadas y evaluando el supuesto de proporcionalidad se fueron probando diferentes combinaciones (ver Anexo 2 y 3). Eliminando aquellas variables que no eran significativas o no cumplen el supuesto de proporcionalidad para una regresión Cox.

Finalmente se obtuvo un modelo con un poco menos de covariables, ya que muchas de ellas no cumplían el supuesto. Quedando, así como el que mejor estima la supervivencia de los individuos que presentan esta condición, un modelo con 3 covariables.

Primero vemos que la presencia de Hepatomegalia, que se asocia con un riesgo 3 veces mayor a la muerte del paciente en la relación con los que no tienen esta condición, además de un valor p igual a 0.01 lo que nos muestra que es significativa al modelo con un 95% de confianza y ayuda a explicar la supervivencia de los pacientes. Para la Hemoglobina con niveles por debajo de lo normal presenta un riesgo 3.2 veces más grande y por último la LDH que también se asocia a un riesgo 3.5 veces mayor cuando se encuentra en niveles por encima del normal, para

ambas covariables se observa un valor p menor 0.05 determinando así su significancia al modelo.

El modelo en general tiene una buena capacidad predictiva con un índice de concordancia de 0.75. Además, que al analizar el supuesto de riesgos proporcionales se cumple para las 3 covariables, lo que hace que de manera general el modelo sea el óptimo para estimar la supervivencia de los pacientes en este análisis.

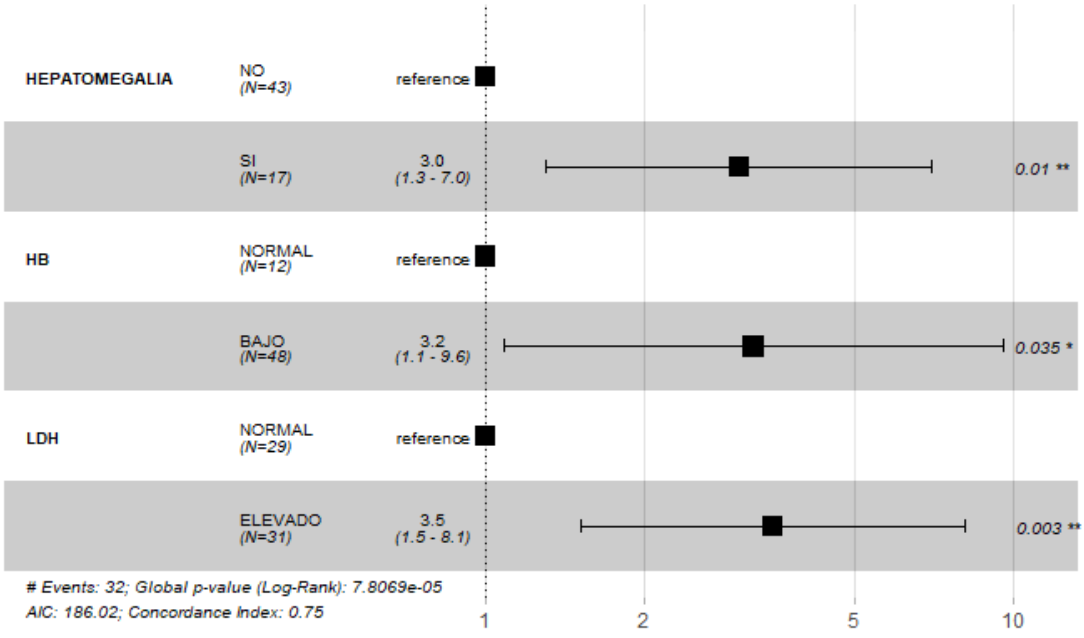


Figura 4.14 *Resumen del modelo final*

Tabla 3: *Prueba de Riesgo proporcional para las covariables del modelo final*

Variable	chisq	df	p
HEPATOMEGALIA	2.248	1	0.13
HB	0.988	1	0.32
LDH	1.499	1	0.22
GLOBAL	4.112	3	0.25

Finalmente, con la selección de covariables que nos ayuden a estimar la supervivencia de los pacientes con esta neoplasia, se graficó la curva del modelo Cox y las 3 covariables significativas. El punto donde la curva intercepta el valor de 0.5 en el eje nos indica el tiempo mediano de supervivencia (aproximadamente 18 meses en este caso). Este es el tiempo en el cual el 50% de los sujetos han experimentado el evento de interés.

En comparación con la curva inicial de la figura 4.6 vemos que, si existen cambios al ingresar las covariables al modelo, ya que se evidencia que la curva en la figura 4.15 cae o disminuye más rápido antes de los 20 meses. Lo que nos indica que los pacientes reportados con las condiciones mencionadas en la figura 4.14 fallecieron en menos de 2 años (15 meses). Esto nos ayuda a identificar que indicadores o biomarcadores tomar en cuenta al momento del diagnóstico en un paciente.

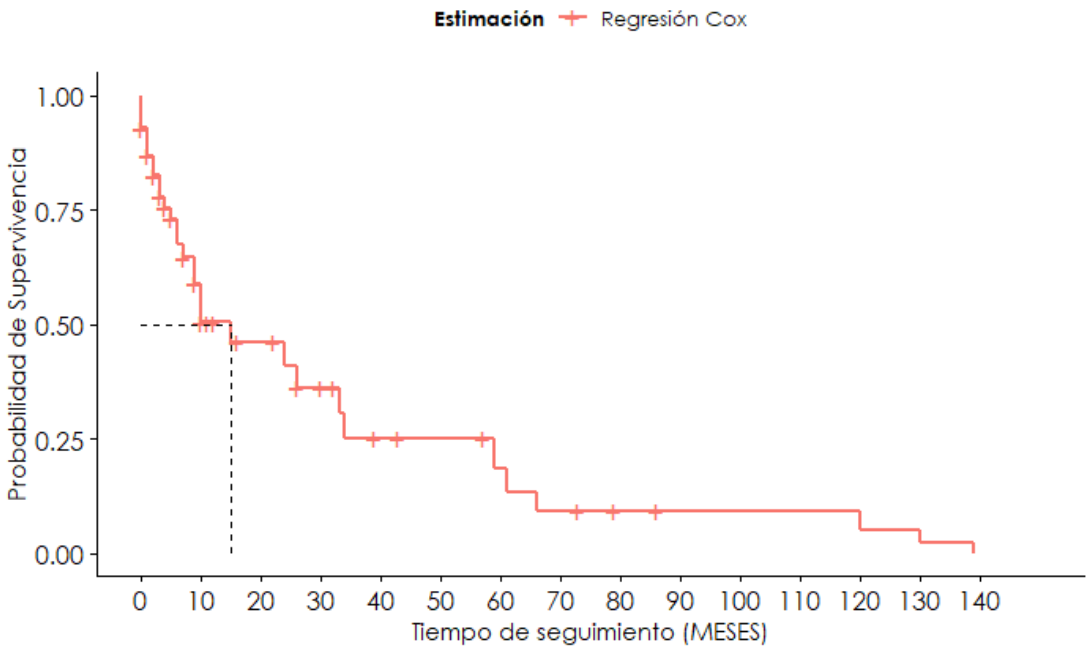


Figura 4.15 *Estimación de la curva con el modelo de Cox con las 3 covariables significativas*

4.2. Análisis de Correspondencia Múltiple y clúster jerárquico

Para el análisis de correspondencia múltiple dada sus características para el análisis simultaneo de numerosas variables categóricas y comprender las relaciones complejas, inicialmente se realizó un ACM donde se incluyeron todas las variables de los 4 segmentos incluida la del estatus final del paciente.
Como vemos en la figura 4.16, la primera dimensión absorbe un 13% de la inercia y vemos que la mayoría de las variables se representan en este ejes o dimensión. La absorción de la segunda dimensión tampoco es fuerte apenas es el 7.1% de la inercia.

Por ello en base al análisis previo que se realizó para la selección de variables en el modelo de regresión Cox, se probaron de la misma manera diferentes combinaciones de variables con el fin de buscar aquellas que nos ayuden a discriminar mejor el comportamiento de los individuos en base a ciertas características. Es decir, lograr que las dimensiones absorban la mayor cantidad de inercia posible.

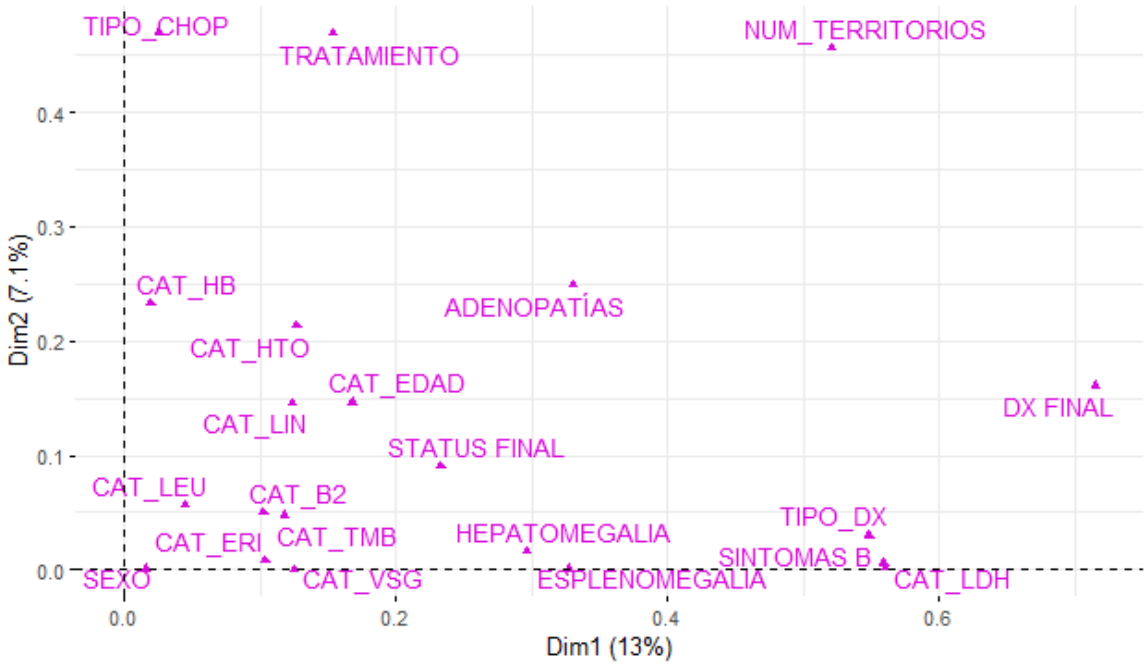


Figura 4.16 *Correlación entre variables y las dimensiones*

Dentro de las combinaciones de variables utilizadas se tomó aquella donde se representa variables de laboratorio como: HB, LDH, B2, del segmento de signos y síntomas: esplenomegalia, hepatomegalia y síntomas B, por último, de las demográficas se tomaron el sexo y los grupos etarios. Se tiene un 40% de absorción de la inercia con las 2 primeras dimensiones y la combinación de las 8 variables como vemos en la figura 4.19.

Además, en el gráfico de las contribuciones figura 4.17 y figura 4.18 se evidencia cuáles son las categorías que se representan mejor en las dimensiones, siendo así los signos y síntomas los que ayudan a representar mejor el primer eje.

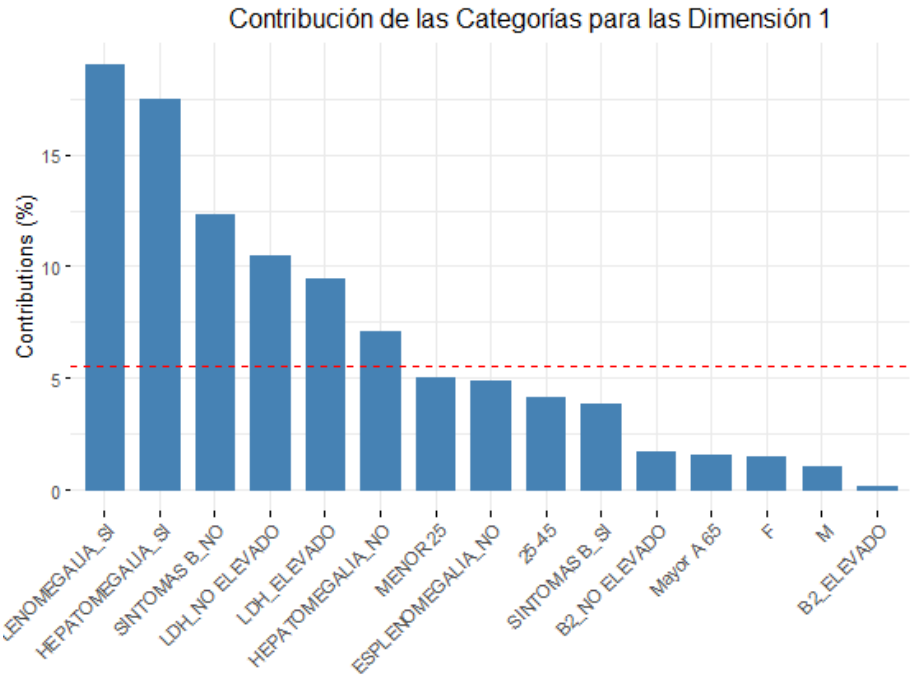


Figura 4.17 *Contribución de las categorías de las variables a las dimensiones 1*

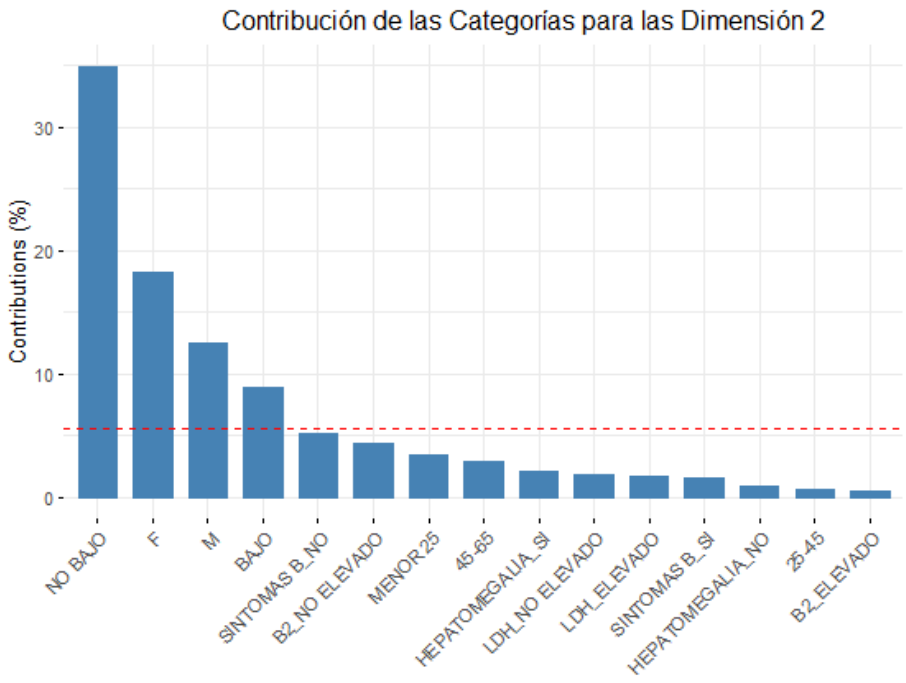


Figura 4.18 *Contribución de las categorías de las variables a las dimensiones 2*

En base a la figura 4.20 se muestran asociaciones en las categorías, si se agrupa a la derecha del gráfico, se evidencia una entre la hepatomegalia, esplenomegalia y el grupo etario menor a 25 años. De la misma forma si nos movemos un poquito a la izquierda encontramos otro grupo de asociaciones como la presencia de síntomas B, LDH elevado, B2M elevado y grupo etario mayor a 65 años.

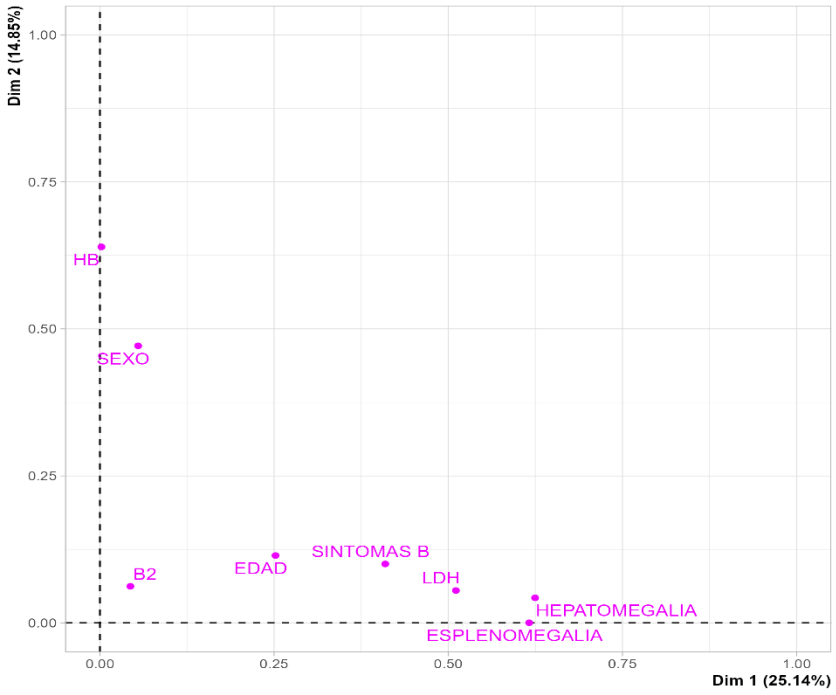


Figura 4.19 *Correlación entre variables y las dimensiones*

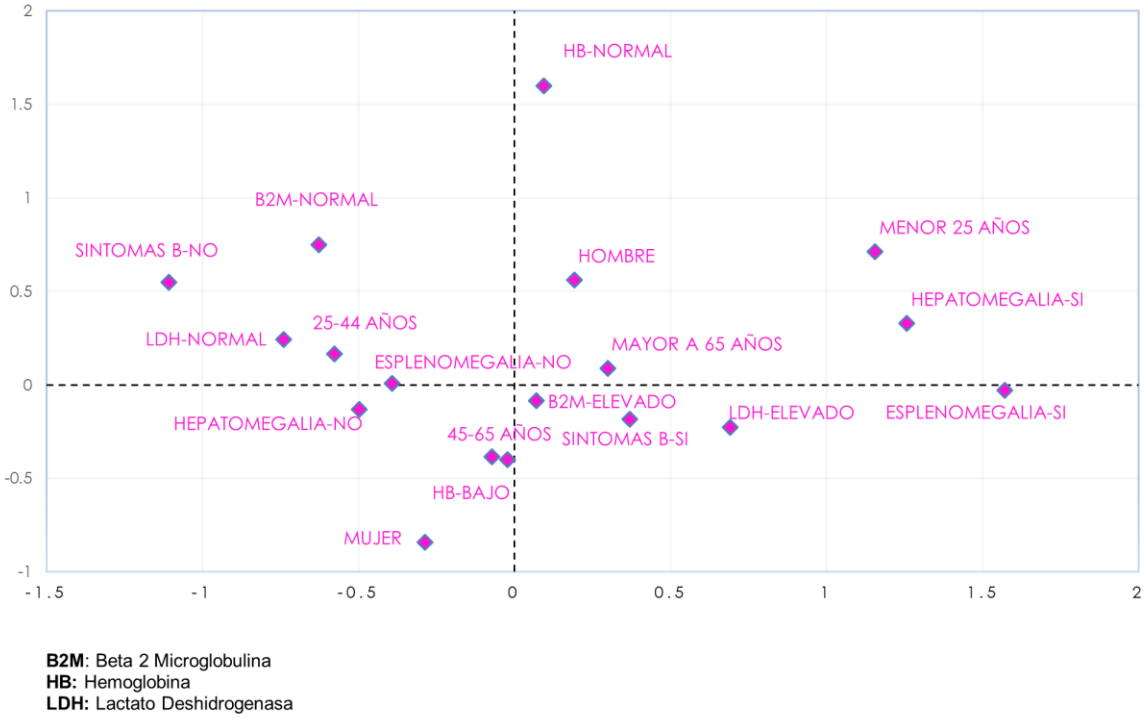
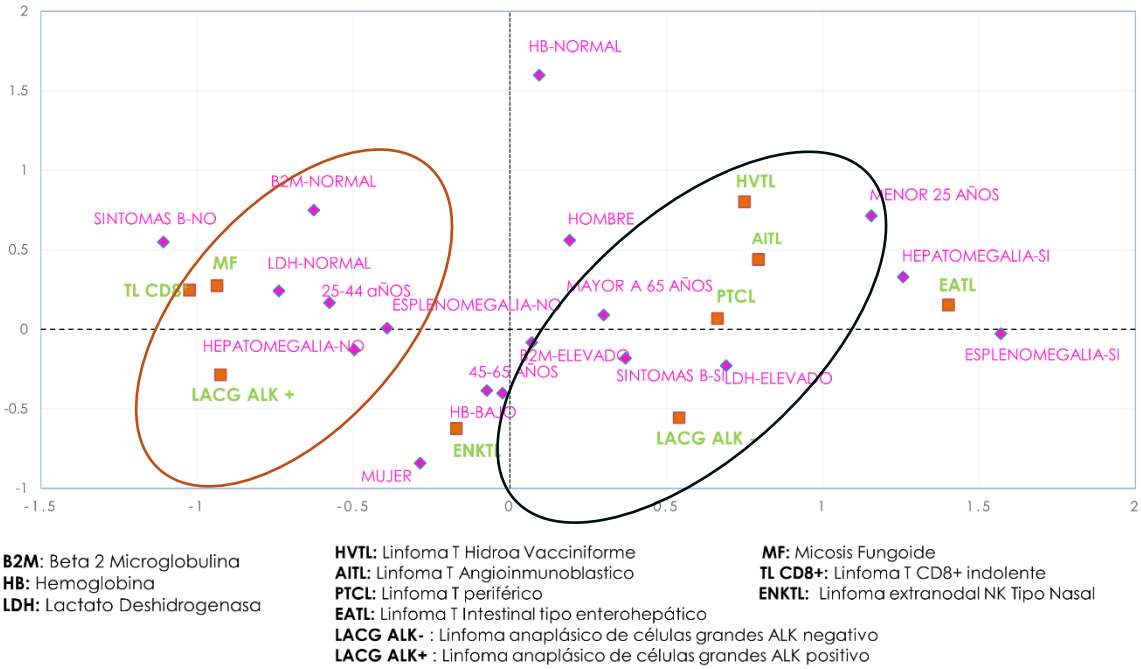


Figura 4.20 *Correlación categorías de las variables*

Para indagar un poco más como estas características se asocian con los diferentes subtipos de diagnósticos para los LNH-T, la añadimos como variables suplementarias. Lo cual nos da una apreciación más clara de cuales de estos subtipos de diagnóstico presentan características menos favorables.

En la figura 4.21 vemos como se divide en 2 grandes grupos con características particulares, la elipse azul nos muestra que son los individuos con exámenes de laboratorio con valor por encima de lo normal, donde los grupos etarios más vulnerables son los jóvenes, niños y mayores a 65 años, con signos y síntomas presentes. Todas estas características se asocian al subtipo de diagnóstico ganglionar o los que se conocen como los más agresivos como el LACG ALK positivo, el linfoma T intestinal.

Por otro lado, tenemos aquellos con características más indolentes que son del subtipo extraganglionar o cutáneo comprendido por el grupo etario de 24-45 y diagnostico como la Micosis Fungoide, LACG ALK negativo.



B2M: Beta 2 Microglobulina
HB: Hemoglobina
LDH: Lactato Deshidrogenasa

HVTL: Linfoma T Hidroa Vacciniforme
AITL: Linfoma T Angioinmunoblastico
PTCL: Linfoma T periférico
EATL: Linfoma T Intestinal tipo enterohepático
LACG ALK- : Linfoma anaplásico de células grandes ALK negativo
LACG ALK+ : Linfoma anaplásico de células grandes ALK positivo

MF: Micosis Fungoide
TL CD8+: Linfoma T CD8+ indolente
ENKTL: Linfoma extranodal NK Tipo Nasal

Figura 4.21 *Categorías de las variables con variable suplementaria (DX_FINAL)*

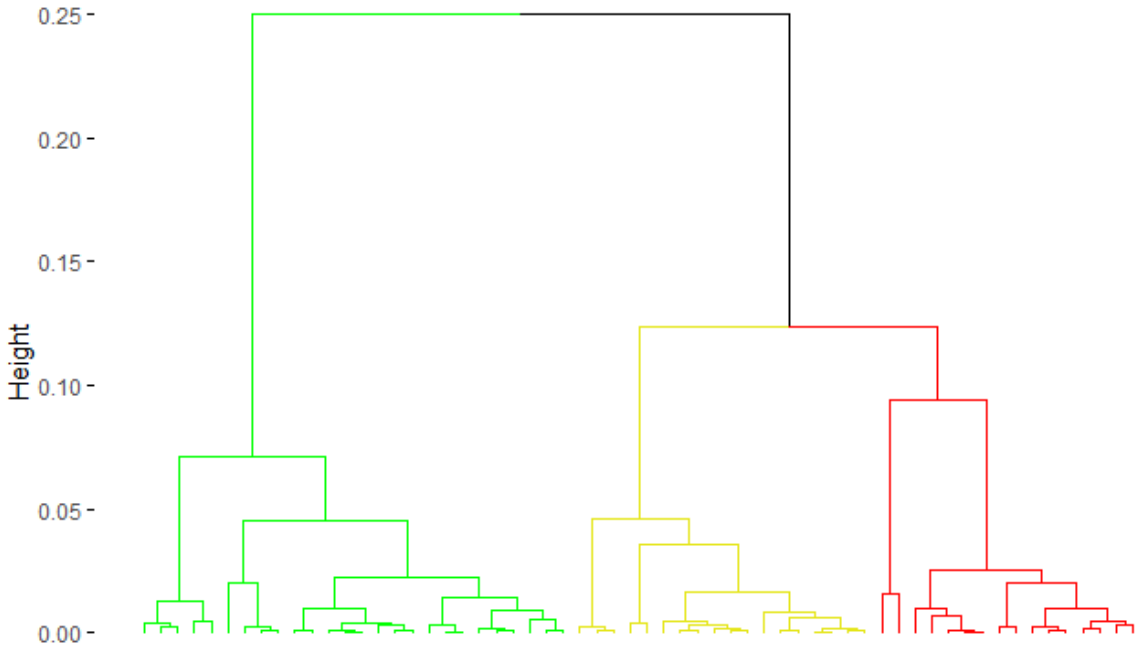


Figura 4.22 *Dendograma en base al análisis de correspondencia múltiple*

Las ramas del dendograma muestran cómo se agrupan las observaciones. Las ramas más cortas indican una mayor similitud entre las observaciones o individuos. El dendograma es útil para entender la estructura jerárquica de los datos y para decidir el número óptimo de clústeres. Basado en la figura 4.22, podrías decidir que tres será lo ideal.

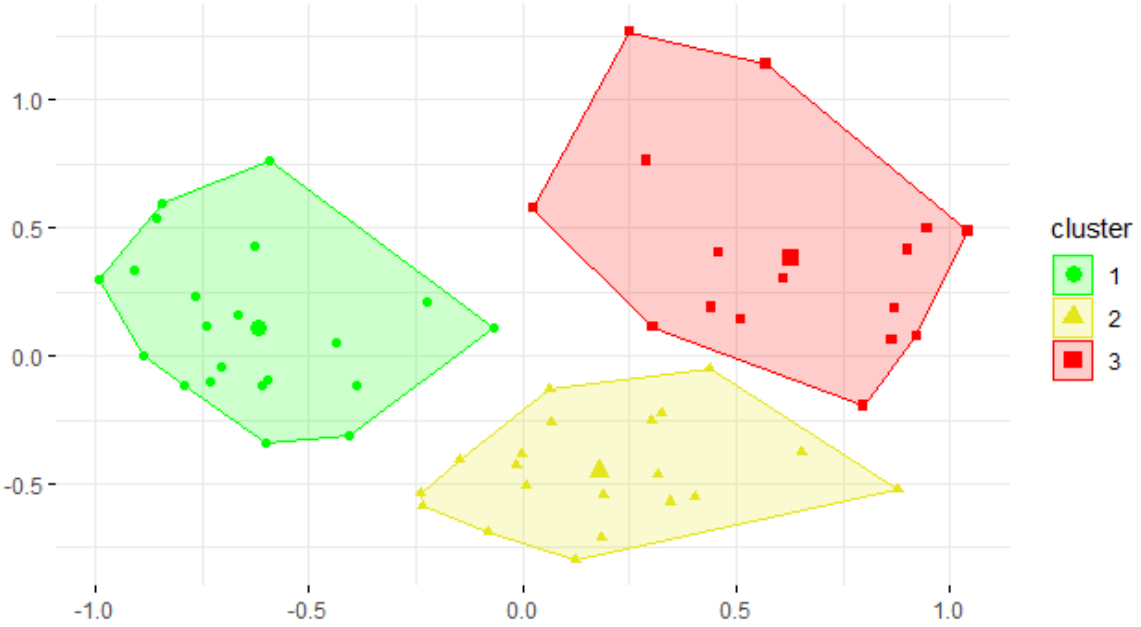


Figura 4.23 *Clúster de los individuos en base a los grupos de variables*

En las agrupaciones que vemos en la figura 4.23 podríamos clasificarla en 3 grupos de riesgo:

1. Grupo de riesgo bajo que presentan características como la LDH normal, ausencia de síntomas B, no presentan signos como la hepatomegalia y esplenomegalia y estos individuos tienen edades entre 25 y 45 años.

2. Grupo de riesgo medio donde los pacientes presentan HB baja, son mayormente mujeres y presentan síntomas B

3. Grupo de riesgo alto son los pacientes se podría decir con las características menos favorables, como la presencia de hepatomegalia, esplenomegalia, LDH elevado, Síntomas B, mayormente están representados por hombres en edades menores a 25 y mayores a 65.

CAPÍTULO 5

5. CONCLUSIONES Y RECOMENDACIONES

Se logró evidenciar que la supervivencia de los pacientes con linfomas T disminuye rápidamente en los primeros 10 meses, lo que sugiere una proporción significativa de eventos tempranos, además que las características clínicas como la hepatomegalia, niveles bajos de HB, y niveles elevados de LDH son factores determinantes en la sobrevida de los pacientes.

Los análisis revelaron que los individuos menores de 25 y mayores de 65 años tienen una menor probabilidad de supervivencia comparada con otros grupos etarios. Los signos, síntomas y biomarcadores clínicos tienen una fuerte asociación en el tipo de diagnóstico, lo que sugiere que una identificación temprana y precisa de estos factores podría mejorar los resultados de los tratamientos.

Se identificaron grupos clínicos distintivos, donde los pacientes diagnosticados con linfomas más agresivos como el LACG ALK negativo y linfoma T intestinal, presentan hepatomegalia y niveles elevados de LDH. Estos hallazgos subrayan la importancia de la estratificación del riesgo y el monitoreo intensivo en estos subgrupos, ya que estos marcadores clínicos están asociados con un pronóstico desfavorable. Además, la identificación temprana de estos linfomas agresivos permite una intervención más oportuna, lo que podría mejorar la respuesta al tratamiento y, en última instancia, la supervivencia de los pacientes.

La recolección sistemática de datos precisos y completos permiten realizar análisis robustos, facilitando la personalización del tratamiento y mejorando los resultados de los pacientes. Al mismo tiempo, un registro de datos minucioso permite una mejor monitorización del progreso del paciente y una detección temprana de posibles complicaciones, asegurando una intervención oportuna y eficaz.

Es importante fomentar investigaciones adicionales para explorar los mecanismos biológicos que vinculan las características clínicas y demográficas con la progresión

de los linfomas de T, así como realizar estudios longitudinales para evaluar el impacto de diferentes tratamientos en la sobrevida de pacientes y ajustar las estrategias terapéuticas en consecuencia.

Finalmente, la colaboración en redes de investigación multicéntricas y multidisciplinaria, que incluya oncólogos, hematólogos, y otros especialistas en salud, puede mejorar significativamente el manejo integral de los pacientes con linfomas T. Este enfoque colaborativo no solo facilita una atención más completa y personalizada, sino que también impulsa el descubrimiento de nuevos hallazgos y avances en el tratamiento.

6. Referencias

Abd ElHafeez, S., D'Arrigo, G., Leonardis, D., Fusaro, M., Tripepi, G., & Roumeliotis, S. (2021). Methods to Analyze Time-to-Event Data: The Cox Regression Analysis. *Oxidative Medicine and Cellular Longevity*, *2021*, 1-6. https://doi.org/10.1155/2021/1302811

Andrade, C. (2023). Survival Analysis, Kaplan-Meier Curves, and Cox Regression: Basic Concepts. *Indian Journal of Psychological Medicine*, *45*(4), 434-435. https://doi.org/10.1177/02537176231176986

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187-202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Definición de linfoma de células T - Diccionario de cáncer del NCI - NCI (nciglobal,ncienterprise). (2011, febrero 2). [nciAppModulePage]. https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/linfoma-de-celulas-t

Dugard, P., Todman, J., & Staines, H. (2022). Survival analysis. En P. Dugard, J. Todman, & H. Staines, *Approaching Multivariate Analysis* (2.ª ed., pp. 337-358). Routledge. https://doi.org/10.4324/9781003343097-14

Dummer, R., Vermeer, M. H., Scarisbrick, J. J., Kim, Y. H., Stonesifer, C., Tensen, C. P., Geskin, L. J., Quaglino, P., & Ramelyte, E. (2021). Cutaneous T cell lymphoma. *Nature Reviews Disease Primers*, *7*(1), 61. https://doi.org/10.1038/s41572-021-00296-9

Fernández, M., & Abraira, V. (s. f.). *Curvas de supervivencia y modelos de regresión: Errores y aciertos en la metodología de aplicación*.

Florensa, D., Godoy, P., Mateo, J., Solsona, F., Pedrol, T., Mesas, M., & Pinol, R. (2021). The Use of Multiple Correspondence Analysis to Explore Associations

Between Categories of Qualitative Variables and Cancer Incidence. *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS*, *25*(9).

François, Husson, & Julie, J. (2014). Multiple Correspondence Analysis. En *Visualization and Verbalization of Data* (1st Edition, p. 20). Chapman and Hall/CRC.

Garrido, D. I., Orquera, A., Rojas, J., & Granja, M. (2021). The Mortality burden of hematological malignancies in Ecuador. *Nepal Journal of Epidemiology*, *11*(2), 1040-1048. https://doi.org/10.3126/nje.v11i2.37057

Global Burden of Disease Cancer Collaboration, Fitzmaurice, C., Abate, D., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdel-Rahman, O., Abdelalim, A., Abdoli, A., Abdollahpour, I., Abdulle, A. S. M., Abebe, N. D., Abraha, H. N., Abu-Raddad, L. J., Abualhasan, A., Adedeji, I. A., Advani, S. M., Afarideh, M., Afshari, M., … Murray, C. J. L. (2019). Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncology*, *5*(12), 1749. https://doi.org/10.1001/jamaoncol.2019.2996

Hjellbrekke, J. (2019). *Multiple correspondence analysis for the social sciences*. Routledge, Taylor & Francis Group.

Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, *53*(282), 457-481. https://doi.org/10.1080/01621459.1958.10501452

Kovalchuk, O., Banakh, S., Masonkova, M., Moskaliuk, N., Rohatynska, N., & Pustovyi, O. (2023). Survival Analysis Models for Estimating and Predicting the Risks of Confession of Criminal Defendants. *2023 13th International Conference on Advanced Computer Information Technologies (ACIT)*, 46-51. https://doi.org/10.1109/ACIT58437.2023.10275450

Le Gall-Ianotto, C., & Misery, L. (2016). Pruritus in Hematological Diseases (Including Aquagenic Pruritus). En L. Misery & S. Ständer (Eds.), *Pruritus* (pp.

271-281). Springer International Publishing. https://doi.org/10.1007/978-3-319-33142-3_36

Martínez Pérez, J. A., & Pérez Martínez, P. S. (2023). Análisis de supervivencia. *Medicina de Familia. SEMERGEN*, *49*(5), 101986. https://doi.org/10.1016/j.semerg.2023.101986

Mathai, A., Provost, S., & Haubold, H. (2022). Chapter 15: Cluster Analysis and Correspondence Analysis. En A. M. Mathai, S. B. Provost, & H. J. Haubold, *Multivariate Statistical Analysis in the Real and Complex Domains* (pp. 845-886). Springer International Publishing. https://doi.org/10.1007/978-3-030-95864-0_15

Mori, Y., Kuroda, M., & Makino, N. (2016). Multiple Correspondence Analysis. En Y. Mori, M. Kuroda, & N. Makino, *Nonlinear Principal Component Analysis and Its Applications* (pp. 21-28). Springer Singapore. https://doi.org/10.1007/978-981-10-0159-8_3

Rabasa, M. P. (2009). Prognostic factors in lymphomas: Non-Hodgkin's lymphomas and Hodgkin's lymphoma. *Anales del Sistema Sanitario de Navarra*. https://doi.org/10.23938/ASSN.0441

Ramírez Montoya, J., Regino, E., & Guerrero, S. (2017). Comparación de Métodos de Estimación en Regresión de Cox. *Comunicaciones en Estadística*, *10*(1), 101. https://doi.org/10.15332/s2027-3355.2017.0001.05

Shiker, M. (2012). Multivariate Statistical Analysis. *British Journal of Science*, *6*, 55-66.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209-249. https://doi.org/10.3322/caac.21660

Tanday, S. (2015). Global cancer cases on the rise. *The Lancet Oncology*, *16*(7), e317. https://doi.org/10.1016/S1470-2045(15)00022-4

Varghese, M. T., & Alsubait, S. (2024). T-Cell Lymphoma. En *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK564354/

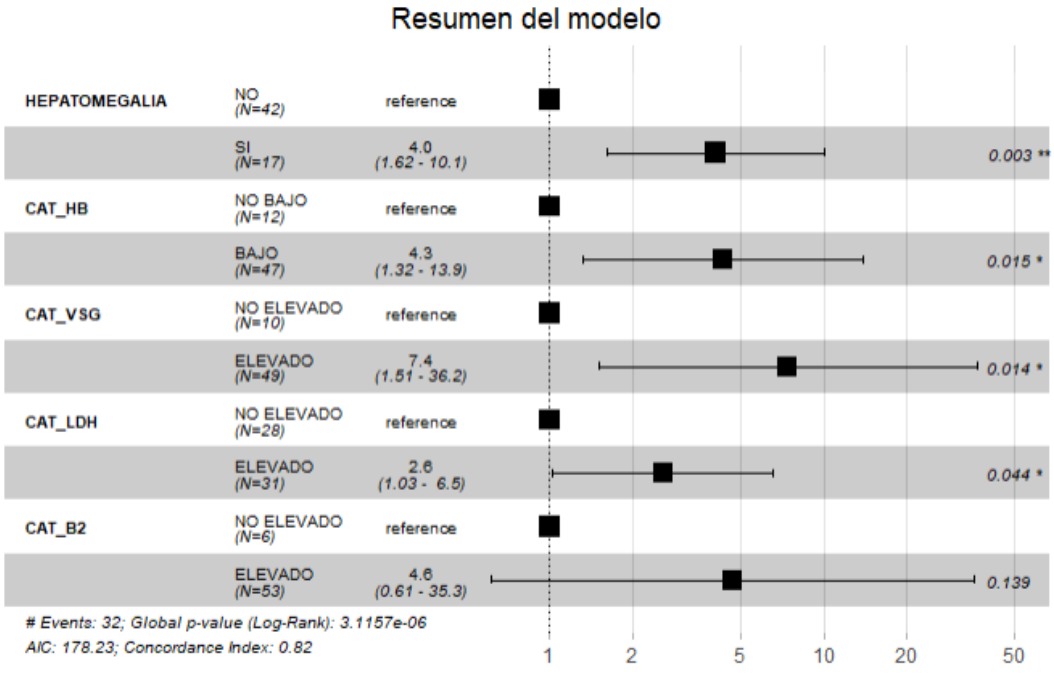
Vinnicombe, S. J., & Garg, N. (2023). Hematologic Malignancy: The Lymphomas. En *Oncologic Imaging: A Multidisciplinary Approach* (pp. 529-554). Elsevier. https://doi.org/10.1016/B978-0-323-69538-1.00030-6

7. Anexos

Anexo 1: Resumen de la estadística descriptiva de las variables de laboratorio.

```
   LEUCOCITOS       LINFOCITOS     ERITROCITOS         HTO              HB         TROMBOCITOS         VSG        
 Min.   : 1.510   Min.   : 6.00   Min.   : 2.440   Min.   :10.20   Min.   : 7.30   Min.   : 20.0   Min.   :  3.00  
 1st Qu.: 5.900   1st Qu.:18.00   1st Qu.: 3.505   1st Qu.:28.90   1st Qu.:10.72   1st Qu.:196.0   1st Qu.: 26.25  
 Median : 7.730   Median :29.00   Median : 4.430   Median :36.00   Median :12.00   Median :262.0   Median : 54.50  
 Mean   : 8.387   Mean   :28.45   Mean   : 4.598   Mean   :34.49   Mean   :12.11   Mean   :273.1   Mean   : 64.70  
 3rd Qu.: 9.740   3rd Qu.:38.50   3rd Qu.: 4.925   3rd Qu.:40.00   3rd Qu.:13.65   3rd Qu.:343.0   3rd Qu.: 85.00  
 Max.   :24.000   Max.   :63.00   Max.   :14.000   Max.   :46.00   Max.   :30.10   Max.   :718.0   Max.   :538.00  
 NA's   :9        NA's   :5       NA's   :5        NA's   :5       NA's   :8       NA's   :5       NA's   :4       
      LDH          B2 MICROGLOBULINA
 Min.   : 103.0   Min.   :1.500    
 1st Qu.: 202.5   1st Qu.:2.700    
 Median : 305.0   Median :4.800    
 Mean   : 534.1   Mean   :5.354    
 3rd Qu.: 545.0   3rd Qu.:8.300    
 Max.   :5010.0   Max.   :9.800    
 NA's   :1        NA's   :1        
```

Anexo 2: Resumen de modelo de Cox realizado que no cumplió los supuestos de riesgos proporcionales (Hepatomegalia, HB, VSG, LDH, B2M)



Anexo 3: Resumen de modelo de Cox realizado que no cumplió los supuestos de riesgos proporcionales (Hepatomegalia, HB, LDH, B2M)