



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

**REGRESIÓN LOGÍSTICA PARA LA IDENTIFICACIÓN DE
FACTORES PRONÓSTICOS DE MORTALIDAD EN PACIENTES
CON CÁNCER DE MAMA EN SOLCA GUAYAQUIL, PERIODO
2015-2020**

Trabajo de Titulación

Tipo: Proyecto de investigación

Presentado para optar el grado académico de:

INGENIERO ESTADÍSTICO

AUTOR: JHONNATAN FERNANDO PINDUISACA ALLAICA

DIRECTOR: Ing. JOHANA ENITH AGUILAR REYES Msc.

Riobamba-Ecuador

2022

©2022, Jhonnatan Fernando Pinduisaca Allaica

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento, siempre y cuando se reconozca el Derecho de Autor.

Yo, JHONNATAN FERNANDO PINDUISACA ALLAICA, declaro que el presente trabajo de titulación es de mi autoría y los resultados del mismo son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autor asumo la responsabilidad legal y académica de los contenidos de este trabajo de titulación; El patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo.

Riobamba, 13 de enero de 2022



Jhonatan Fernando Pinduisaca Allaica

060550600-5

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

El Tribunal del Trabajo de Titulación certifica que: El Trabajo de Titulación; Tipo: Proyecto de Investigación, **REGRESIÓN LOGÍSTICA PARA LA IDENTIFICACIÓN DE FACTORES PRONÓSTICOS DE MORTALIDAD EN PACIENTES CON CÁNCER DE MAMA EN SOLCA GUAYAQUIL, PERIODO 2015-2020**, realizado por el señor: **JHONNATAN FERNANDO PINDUISACA ALLAICA**, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Titulación, El mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal Autoriza su presentación.

	FIRMA	FECHA
Dra. Jaqueline Balseca Msc. PRESIDENTE DEL TRIBUNAL		13-01-2022
Ing. Johanna Enith Aguilar Reyes Msc. DIRECTOR/A DEL TRABAJO DE TITULACION		13-01-2022
Ing. Nancy Chariguamán Maurisaca Msc. MIEMBRO DE TRIBUNAL		13-01-2022

DEDICATORIA

El éxito de una persona se mide por el sacrificio y dedicación con el que hace las cosas, por ello el presente trabajo de investigación, va dedicado en primer lugar a Dios por darme la sabiduría y ahínco necesario para culminar el presente trabajo de la mejor manera posible. A la par va dedicado a mi madre y hermanas, quienes han sido el pilar más fundamental, jugando siempre un papel importante en el desarrollo del presente trabajo. Va dedicado por el inmenso apoyo, cariño y aliento brindado; antes y durante el desarrollo del trabajo de titulación, por la confianza brindada y por la lucha constante a lo largo de este tiempo.

Jhonnatan

AGRADECIMIENTOS

Es preciso agradecer a Dios de primera mano, por darme la vida, sabiduría y constancia necesaria para concluir el presente trabajo de titulación y no desmayar en el camino.

Este trabajo es fruto de un esfuerzo en conjunto, por lo que agradezco principalmente a mi madre y hermanas por ser ese pilar fuerte en mi vida, que no me permitió desmayar, aprovechar para agradecer por el apoyo incondicional recibido de parte de ellas a lo largo de mi vida.

Es importante resaltar la ardua labor por parte de mi tutor de tesis Ing. Johana Enith Aguilar Reyes para el correcto desarrollo del trabajo, así mismo resaltar el trabajo del miembro de tesis Ing. Nancy Chariguamán Maurisaca por el apoyo directo en la resolución de este trabajo de titulación, con el fin de conseguir un trabajo de calidad, por lo que extiende mi agradecimiento.

Un agradecimiento especial para la Sociedad Ecuatoriana de Lucha contra el cáncer matriz Guayaquil y a cada uno de sus miembros, quienes me permitieron la ejecución del trabajo dentro de sus instalaciones, así como a los miembros del del Departamento de Bioestadística – Área Gestión de la Información y Productividad liderados de manera correcta por el Dr. Johnny Real Cotto y la Ing. Leyda Jaramillo, por encaminarme y brindarme todas las facilidades para la ejecución correcta del trabajo.

Quiero finalmente agradecer a todas y cada una de las personas, que de una u otra forma, siempre estuvieron ahí de una manera desinteresada, brindándome apoyo cuando necesitaba.

Jhonnatan

TABLA DE CONTENIDOS

ÍNDICE DE TABLAS.....	x
ÍNDICE DE FIGURAS.....	xii
ÍNDICE DE GRÁFICOS.....	xiii
ÍNDICE DE ANEXOS.....	xvi
ÍNDICE DE ABREVIATURAS.....	xvii
RESUMEN.....	xviii
SUMMARY.....	xix
INTRODUCCIÓN.....	1
OBJETIVOS.....	5

CAPÍTULO I

1. MARCO TEÓRICO REFERENCIAL.....	6
1.1. Cáncer y Tumor.....	6
1.2. Cáncer de mama.....	6
1.2.1. <i>Tipos de cáncer de mama.....</i>	<i>7</i>
1.2.2. <i>Partes de la mama.....</i>	<i>7</i>
1.2.3. <i>Síntomas del cáncer de mama.....</i>	<i>8</i>
1.2.4. <i>Tratamiento.....</i>	<i>9</i>
1.2.4.1. <i>Cirugía.....</i>	<i>9</i>
1.2.4.2. <i>Radioterapia.....</i>	<i>9</i>
1.2.4.3. <i>Terapias con medicamentos.....</i>	<i>9</i>
1.3. Factor pronóstico.....	10
1.3.1. <i>Relacionados con el paciente.....</i>	<i>10</i>
1.3.2. <i>Relacionados con la neoplasia.....</i>	<i>10</i>
1.3.3. <i>Relacionados con el tratamiento.....</i>	<i>10</i>
1.3.4. <i>Relacionados con el entorno.....</i>	<i>11</i>
1.3.5. <i>Principales factores pronósticos.....</i>	<i>11</i>
1.3.6. <i>Receptores hormonales y HER2.....</i>	<i>11</i>
1.3.7. <i>Localización de las metástasis.....</i>	<i>11</i>
1.3.8. <i>Carga tumoral.....</i>	<i>11</i>
1.3.9. <i>Edad.....</i>	<i>12</i>
1.4. Estadística.....	12
1.4.1. <i>Estadística Descriptiva.....</i>	<i>12</i>
1.4.1.1. <i>Variables.....</i>	<i>13</i>

1.4.1.2.	<i>Medidas de tendencia central</i>	13
1.4.1.3.	<i>Medidas de dispersión</i>	14
1.4.2.	Análisis de dependencia	15
1.4.2.1.	<i>Dependencia de variables cuantitativas</i>	15
1.4.2.2.	<i>Dependencia de variables cualitativas</i>	16
1.4.3.	Regresión logística	16
1.4.3.1.	<i>Regresión Logística Binaria</i>	17
1.4.3.2.	<i>Regresión Logística Múltiple</i>	17

CAPÍTULO II

2.	MARCO METODOLÓGICO	23
2.1.	Tipo de investigación	23
2.2.	Diseño de la investigación	23
2.3.	Identificación de variables	24
2.4.	Operacionalización de los objetivos	24
2.5.	Localización del Estudio	25
2.6.	Población de estudio	26
2.7.	Tamaño de la muestra	26
2.8.	Método de muestreo	26
2.9.	Técnicas de recolección de datos	27
2.10.	Modelo Estadístico	27

CAPÍTULO III

3.	RESULTADOS Y DISCUSIÓN DE LOS RESULTADOS	28
3.1.	Operacionalización de las variables	28
3.2.	Análisis Estadístico Descriptivo Exploratorio	30
3.2.1.	<i>Edad</i>	30
3.2.2.	<i>Raza</i>	31
3.2.3.	<i>Estatura</i>	32
3.2.4.	<i>Peso</i>	33
3.2.5.	<i>Cirugía</i>	34
3.2.6.	<i>Tiempo de evolución</i>	35
3.2.7.	<i>Tamaño tumoral</i>	36
3.2.8.	<i>Ubicación del tumor</i>	37

3.2.9.	<i>Tipo de tumor</i>	38
3.2.10.	<i>Estadio</i>	39
3.2.11.	<i>Estado T</i>	40
3.2.12.	<i>Estado N</i>	41
3.2.13.	<i>Estado M</i>	42
3.2.14.	<i>Estado</i>	42
3.3.	Análisis de Dependencia	43
3.3.1.	<i>Análisis de dependencia de variables cuantitativas</i>	43
3.3.2.	<i>Análisis de dependencia de variables cualitativas</i>	44
3.4.	Análisis Estadístico Logístico	45
3.4.1.	Modelo Logístico 1	46
3.4.1.1.	<i>Coefficientes de regresión del primer modelo</i>	47
3.4.1.2.	<i>Resumen de la regresión del primer modelo</i>	49
3.4.1.3.	<i>Variables significativas del primer modelo</i>	50
3.4.1.4.	<i>Evaluación del primer modelo</i>	51
3.4.1.5.	<i>Significancia del primer modelo</i>	51
3.4.1.6.	<i>Precisión del primer modelo</i>	51
3.4.1.7.	<i>Error de precisión del primer modelo</i>	53
3.4.1.8.	<i>Curva ROC del primer modelo</i>	53
3.4.1.9.	<i>Predicciones del primer modelo</i>	54
3.4.1.10.	<i>Predicción ejemplo con el primer modelo</i>	55
3.4.2.	Modelo Logístico 2	56
3.4.2.1.	<i>Coefficientes de regresión del segundo modelo</i>	57
3.4.2.2.	<i>Resumen de la regresión del segundo modelo</i>	58
3.4.2.3.	<i>Variables significativas del segundo modelo</i>	59
3.4.2.4.	<i>Evaluación del segundo modelo</i>	60
3.4.2.5.	<i>Significancia del segundo modelo</i>	60
3.4.2.6.	<i>Precisión del segundo modelo</i>	60
3.4.2.7.	<i>Error de precisión del segundo modelo</i>	62
3.4.2.8.	<i>Curva ROC del segundo modelo</i>	62
3.4.2.9.	<i>Predicciones del segundo modelo</i>	63
3.4.2.10.	<i>Predicción ejemplo con el segundo modelo</i>	64
3.4.3.	Modelo Logístico 3	65
3.4.3.1.	<i>Coefficientes de regresión del tercer modelo</i>	66
3.4.3.2.	<i>Resumen de la regresión del tercer modelo</i>	66
3.4.3.3.	<i>Variables significativas del tercer modelo</i>	67
3.4.3.4.	<i>Evaluación del tercer modelo</i>	68

3.4.3.5.	<i>Significancia del tercer modelo</i>	68
3.4.3.6.	<i>Precisión del tercer modelo</i>	68
3.4.3.7.	<i>Error de precisión del tercer modelo</i>	69
3.4.3.8.	<i>Curva ROC del tercer modelo</i>	70
3.4.3.9.	<i>Predicciones del tercer modelo</i>	71
3.4.3.10.	<i>Predicción ejemplo con el tercer modelo</i>	72
3.4.4.	<i>Comparación de modelos</i>	73
3.4.5.	<i>Factores Pronósticos Significativos</i>	74
3.4.5.1.	<i>Definidos con el modelo 1</i>	74
3.4.5.2.	<i>Definidos con el modelo 2</i>	75
3.4.5.3.	<i>Definidos con el modelo 3</i>	76
CONCLUSIONES		77
RECOMENDACIONES		78
BIBLIOGRAFÍA		
ANEXOS		

ÍNDICE DE TABLAS

Tabla 1-1:	Matriz de Confusión.	20
Tabla 2-1:	Criterio AUC.	22
Tabla 1-2:	Operacionalización de los objetivos.....	24
Tabla 1-3:	Operacionalización de las variables.	28
Tabla 2-3:	Estadísticas de la edad de pacientes diagnosticados con cáncer de mama.	31
Tabla 3-3:	Estadísticas de la raza de pacientes.	32
Tabla 4-3:	Estadísticas de la estatura de pacientes diagnosticados con cáncer de mama.	33
Tabla 5-3:	Estadísticas del peso de pacientes diagnosticados con cáncer de mama.	34
Tabla 6-3:	Estadísticas de la condición de cirugía del paciente.	35
Tabla 7-3:	Estadísticas del tiempo de evolución de la enfermedad.....	35
Tabla 8-3:	Estadísticas del tamaño de tumor.	36
Tabla 9-3:	Estadísticas de la ubicación del tumor.	37
Tabla 10-3:	Estadísticas de la condición de cirugía del paciente.	38
Tabla 11-3:	Estadísticas del estadio del paciente.....	39
Tabla 12-3:	Estadísticas del Estado T.....	40
Tabla 13-3:	Estadísticas del Estado N.	41
Tabla 14-3:	Estadísticas del Estado M.	42
Tabla 15-3:	Estadísticas del estado evolutivo del paciente.....	43
Tabla 16-3:	Matriz de correlación entre variables cuantitativas.	43
Tabla 17-3:	Coefficientes de relación de variables cualitativas.	44
Tabla 18-3:	Coefficientes de regresión Modelo 1.....	47
Tabla 19-3:	Resumen de regresión del modelo 1.....	49
Tabla 20-3:	Variables significativas del modelo 1.	50
Tabla 21-3:	Significancia del modelo 1.....	51
Tabla 22-3:	Matriz de confusión del modelo 1.....	52
Tabla 23-3:	Variación observaciones vs predicciones 1.....	54
Tabla 24-3:	Proporción de predichos modelos 1.	54
Tabla 25-3:	Coefficientes de regresión del modelo 2.	57
Tabla 26-3:	Resumen de regresión modelo 2.	59
Tabla 27-3:	Variables significativas del modelo 2.	59
Tabla 28-3:	Significancia del modelo 2.....	60
Tabla 29-3:	Matriz de confusión del modelo 2.....	60
Tabla 30-3:	Variación observaciones vs predicciones 2.....	63
Tabla 31-3:	Proporción de predichos modelos 2.	63
Tabla 32-3:	Coefficientes de regresión del modelo 3.	66

Tabla 33-3:	Resumen de regresión modelo 3	67
Tabla 34-3:	Variables significativas del modelo 3.	67
Tabla 35-3:	Significancia del modelo 3.....	68
Tabla 36-3:	Matriz de confusión del modelo 3.....	68
Tabla 37-3:	Variación observaciones vs predicciones 3.....	71
Tabla 38-3:	Proporción de predichos modelos 3.	71
Tabla 39-3:	Comparación de modelos.....	73
Tabla 40-3:	Factores pronósticos asumidos con el modelo 1.....	74
Tabla 41-3:	Factores pronósticos asumidos con el modelo 2.....	75
Tabla 42-3:	Factores pronósticos asumidos con el modelo 3.....	76

ÍNDICE DE FIGURAS

Figura 1-1:	Anatomía de la mama.....	7
Figura 2-1:	Partes de la mama.....	8
Figura 3-1:	Partes de la mama.....	8
Figura 1-2-:	Ubicación Geográfica del Hospital General SOLCA	26

ÍNDICE DE GRÁFICOS

Gráfico 1-3:	Distribución de la edad de pacientes diagnosticados con cáncer.	30
Gráfico 2-3:	Raza de pacientes diagnosticados con cáncer.	31
Gráfico 3-3:	Distribución de estatura de pacientes diagnosticados con cáncer.	32
Gráfico 4-3:	Distribución del peso de pacientes con cáncer.	33
Gráfico 5-3:	Condición de cirugía del paciente.	34
Gráfico 6-3:	Distribución del tiempo de evolución de la enfermedad.	35
Gráfico 7-3:	Distribución del tamaño del tumor.	36
Gráfico 8-3:	Ubicación del tumor en la paciente.	37
Gráfico 9-3:	Tipo de tumor del paciente.	38
Gráfico 10-3:	Estadio del paciente diagnosticado con cáncer.	39
Gráfico 11-3:	Tamaño y extensión del tumor principal.	40
Gráfico 12-3:	Invasión de ganglios linfáticos.	41
Gráfico 13-3:	Diseminación del cáncer principal.	42
Gráfico 14-3:	Estado evolutivo del paciente.	42
Gráfico 15-3:	Gráfico de correlación.	44
Gráfico 16-3:	Matriz de confusión del modelo 1.	52
Gráfico 17-3:	Curva ROC del modelo 1.	53
Gráfico 18-3:	Proporción de predicciones modelo 1.	55
Gráfico 19-3:	Matriz de confusión del modelo 2.	61
Gráfico 21-3:	Proporción de predicciones modelo 2.	64
Gráfico 22-3:	Matriz de confusión del modelo 3.	69
Gráfico 23-3:	Curva ROC del modelo 3.	70
Gráfico 24-3:	Proporción de predicciones modelo 3.	72

ÍNDICE DE ECUACIONES

Ecuación 1-1:	Media	13
Ecuación 2-1:	Mediana caso par.....	14
Ecuación 3-1:	Mediana caso impar	14
Ecuación 4-1:	Moda.....	14
Ecuación 5-1:	Rango.....	14
Ecuación 6-1:	Varianza	15
Ecuación 7-1:	Desviación Estándar.....	15
Ecuación 8-1:	Coefficiente de variación.....	15
Ecuación 9-1:	Coefficiente de Pearson.....	15
Ecuación 10-1:	Prueba de dependencia.....	16
Ecuación 11-1:	Ji-cuadrada	16
Ecuación 12-1:	Regresión logística.....	17
Ecuación 13-1:	Regresión logística múltiple.....	17
Ecuación 14-1:	Likelihood ratio.....	18
Ecuación 15-1:	Devianza	18
Ecuación 16-1:	Odd	18
Ecuación 17-1:	Odds ratio.....	19
Ecuación 18-1:	Odds ratio cambio en unidad.....	19
Ecuación 19-1:	Probabilidad	19
Ecuación 20-1:	Exactitud	21
Ecuación 21-1:	Precisión.....	21
Ecuación 22-1:	Error de precisión.....	21
Ecuación 1-3:	Prueba de dependencia.....	44
Ecuación 2-3:	Modelo 1	46
Ecuación 3-3:	Modelo 1 desglozado	46
Ecuación 4-3:	Modelo 1 simplificado	48
Ecuación 5-3:	Modelo 1 reemplazado.....	48
Ecuación 6-3:	Significancia de variables de modelo 1	49
Ecuación 7-3:	Significancia del modelo 1	51
Ecuación 8-3:	Precisión modelo 1	52
Ecuación 9-3:	Error modelo 1	53
Ecuación 10-3:	Modelo 1 a predecir	55
Ecuación 11-3:	Probabilidad modelo 1.....	55
Ecuación 12-3:	Modelo 2.....	56
Ecuación 13-3:	Modelo 2 desglozado	57

Ecuación 14-3:	Modelo 2 simplificado	58
Ecuación 15-3:	Modelo 2 reemplazado	58
Ecuación 16-3:	Significancia de variables de modelo 2	58
Ecuación 17-3:	Significancia del modelo 2	60
Ecuación 18-3:	Precisión modelo 2	61
Ecuación 19-3:	Error modelo 2	62
Ecuación 20-3:	Modelo 2 a predecir	64
Ecuación 21-3:	Probabilidaad modelo 2.....	65
Ecuación 22-3:	Modelo 3	65
Ecuación 23-3:	Modelo 3 reemplazado	66
Ecuación 24-3:	Significancia de variables de modelo 3	66
Ecuación 25-3:	Significancia del modelo 3	68
Ecuación 26-3:	Precisión modelo 3	69
Ecuación 27-3:	Error modelo 3	69
Ecuación 28-3:	Modelo 3 a predecir	72
Ecuación 29-3:	Precisión modelo 3.....	72

ÍNDICE DE ANEXOS

ANEXO A: DECLARACIÓN DE CONFIDENCIALIDAD

ANEXO B: DECLARACIÓN DE CONFIDENCIALIDAD PARTE 2

ANEXO C: BASE DE DATOS

ANEXO D: CÓDIGO EN R

ÍNDICE DE ABREVIATURAS

AUC	Area Under Curve
FN	Falso Negativo
FP	Falso Positivo
IV	Intravenoso
NC	Nivel de confianza
MLR	Multiple Logistic Regression
RLM	Regresión Logística Múltiple
ROC	Receiver Operating Characteristic
SOLCA	Sociedad de Lucha Contra el Cáncer
VN	Verdadero Negativo
VP	Verdadero Positivo

RESUMEN

El presente trabajo de investigación tuvo como objetivo principal, la identificación de aquellos factores pronósticos que influyen en la mortalidad de pacientes diagnosticados con cáncer de mama, con base en factores ya preestablecidos, analizando su significancia e influencia directa; así como relaciones que permitan generar modelos predictivos que anticipen el comportamiento de la enfermedad en un paciente. El desarrollo del trabajo se realizó en el Instituto Oncológico Nacional “Dr. Juan Tanca Marengo” (SOLCA Matriz Guayaquil), con base en los datos que reposan en el sistema propio de la institución. Se tomó en cuenta a todos los pacientes diagnosticados con cáncer de mama en el periodo 2015-2020, teniendo un total de 3062 pacientes, trabajando únicamente con 956 de ellos luego de una correcta depuración de datos. Se realizó una investigación mixta, combinando variables de carácter cualitativo y cuantitativo, estudiadas a través de modelos predictivos modelados a partir de regresión logística misma que se realizó en el software estadístico R. Además, es de carácter teórico, abarcando una investigación descriptiva y explicativa, de índole no experimental, basándose en una inferencia hipotética-deductiva. Con un 95% de confianza se plantearon tres modelos, que resultaron significativos; con una capacidad de precisión para predecir aceptable (mayor al 75%), definiendo los tres modelos como factores pronósticos influyentes al tamaño, peso, tiempo de evolución, tamaño tumoral, estadio, y los estados que conforman la estadificación (T, M, N). Además, con probabilidades aproximadamente aceptables se define al modelo uno, como el más idóneo para pronosticar el comportamiento de un paciente ante la enfermedad. El estudio de factores pronósticos es un tema importante a tratar a un futuro con un mayor número de pacientes, con el fin de conseguir modelos más precisos y con una mayor capacidad predictiva, que nos permitan delimitar factores pronósticos más influyentes y decisivos.

Palabras clave: <FACTORES PRONÓSTICOS>, <MODELO PREDICTIVO>, <PRECISIÓN>, <REGRESIÓN LOGÍSTICA MÚLTIPLE (RLM)>, <SIGNIFICANCIA>.

LEONARDO
FABIO MEDINA
NUSTE

Firmado digitalmente
por LEONARDO FABIO
MEDINA NUSTE
Fecha: 2021.10.07
12:23:41 -05'00'



1861-DBRA-UTP-2021

SUMMARY

The main objective of this research work was to identify those prognostic factors that influence the mortality of patients diagnosed with breast cancer, based on pre-established factors, analyzing their significance and direct influence; as well as relationships that allow generating predictive models that anticipate the behavior of the Disease in a patient. The development of the work was carried out at the National Oncological Institute “Dr. Juan Tanca Marengo” (SOLCA Guayaquil), based on the data that rest in the institution’s own system. All patients diagnosed with breast cancer in the period 2015-2020 were taken into account, with a total of 3062 patients, working only with 956 of them after correct data filtering. A mixed investigation was carried out, combining qualitative and quantitative variables, studied through predictive models modeled from the same logistic regression that was carried out in the statistical software R. In addition, it is of a theoretical type, encompassing a descriptive and explanatory investigation, of a non-experimental nature, based on a hypothetical-deductive inference. With 95% confidence, three models were proposed, which were significant; with an acceptable accurate ability to predict (greater than 75%), defining the three models as prognostic factors influencing size, weight, time of evolution, tumor size, stage, and the states that form the staging (T, M, N). In addition, with approximately acceptable probabilities, model one is defined as the most suitable for predicting the behavior of a patient before the disease. The study of prognostic factors is an important topic to deal with in the future with a bigger number of patients, in order to achieve more precise models with a greater predictive capacity, which allow us to define more influential and decisive prognostic factors.

Keywords: <FORECAST FACTORS>, <PREDICTIVE MODEL>, <PRECISION>, <MULTIPLE LOGISTIC REGRESSION (MLR)>, <SIGNIFICANCE>.

INTRODUCCIÓN

Durante años el cáncer ha sido una de las principales enfermedades causantes de muertes en el Ecuador, cada vez son más personas que padecen esta enfermedad tan morbosa y consumista, de esta manera no ha pasado de moda dentro de las tasas de mortalidad de la población ecuatoriana, sino al contrario ha tomado más fuerza y control, ayudada por el descuido, despreocupación y desconocimiento de la población. En los últimos años el cáncer de mama, próstata y piel, son aquellos que más han afectado a hombres y mujeres.

Teniendo así por ejemplo que, en el Ecuador, la incidencia de cáncer de mama fue de 28.058 nuevos casos, en promedio existen 165 casos de cáncer en todas sus variedades por cada 100.000 mujeres y 150 casos por cada 100.000 hombres (Ministerio de Salud, 2018, párr.1).

Es así como el cáncer de mama se coloca en el puesto número 1 como causa principal de muertes femeninas en el Ecuador, frente a esta situación la Sociedad de Lucha Contra el Cáncer del Ecuador, realiza esfuerzos continuos para la prevención, tratamiento y control de este padecimiento. Sin embargo, los esfuerzos son pocos, cuando la enfermedad está avanzada y se ha apoderado casi por completo del cuerpo del paciente, creando así una situación crucial en la cual muchas veces, a la recuperación positiva de un paciente se le asigna pocas probabilidades, obligando al mismo a la resignación y espera de una situación terminal.

Al ser el cáncer de mama uno de los que más afecta a la población ecuatoriana, en este caso a la población femenina, ha sido objeto de estudio dentro de la presente investigación. Teniendo así que a pesar de la tasa de decrecimiento de mujeres que padecen este tipo de cáncer en los últimos años, las cifras aún resultan alarmantes, por lo que la pronta prevención, y el adecuado tratamiento ayudan a mermar dicho padecimiento y de una manera positiva ayudan en el decrecimiento de las tasas de mortalidad por este tipo de cáncer. Es por ello por lo que el control y análisis de los factores pronósticos resulta importante para quienes no escatiman esfuerzos en la lucha de esta enfermedad maliciosa. Al ser los factores pronósticos aquellos detonantes e informantes claves para la pronosticación de la recuperación de la enfermedad de un paciente o por el otro lado de la recidivancia a la enfermedad, es correcto utilizarlos de forma positiva, lo que de manera directa o indirecta ayudará a la prevención secundaria y terciaria de la enfermedad, así como en la correcta aplicación de un tratamiento al paciente, que reduzca la probabilidad de no supervivencia de un individuo.

La evolución constante del cáncer de mama, ha concurrido en la evolución de estudios, en los cuales, con base científica, han apoyado en el desarrollo estratégico de tratamientos y medicamentos que ayuden a aminorar el índice de muertes a causa de este padecimiento, sin embargo, la evolución constante de la enfermedad, en este caso específicamente el cáncer de mama, hace que cualquier esfuerzo o estudio realizado, se vuelvan obsoletos.

La pronta prevención y correcto manejo de tratamientos es un apoyo fundamental en la lucha contra el cáncer, ya que permite tener una visión generalizada de la mayoría de persona que la padecen, al ser distinto el comportamiento de la enfermedad en cada persona, la convierte en aun más problemática. Los factores pronósticos han sido de mucha ayuda para hacerle frente a esta enfermedad, por lo que estudios de índole similar han sentado una base fundamental para que poco a poco se aborde más acerca de la relevancia de esta temática y así los estudios tomen mayor fuerza, aumentando la posibilidad de otorgar mejores tratamientos, indistintamente del caso, y de esta manera contribuir a la lucha de este mal.

La identificación de factores pronósticos de mortalidad de pacientes con cáncer es realmente importante, al momento de tratar a un paciente con cáncer, ya que son estos los que proveen de información importante. Los factores pronósticos, son un apoyo, y abren una gran fuente de información, para el apoyo en el tratamiento de un paciente, pues son estos aquellos que determinan probabilidades de evolución o de decaimiento de un paciente frente a la enfermedad, lo que favorece al momento de las decisiones respecto al tratamiento a utilizar con cierto paciente. Factores como: edad, peso, metástasis, localización central del tumor, patrón histológico, invasión vascular y linfática, tamaño tumoral, entre otros, considerados como importantes y manejados adecuadamente con técnicas estadísticas (regresión logística), permitirán la correcta identificación y validación de significancia de estos, así como relaciones que pudiesen existir, para de esta manera ayudar en la prevención que mitigue la tasa de mortalidad y que permita tener probabilidades más esperanzadoras en la recuperación de un paciente diagnosticado con este tipo de cáncer.

Como base fundamental se tiene un trabajo de (Flores, et al., 2008, pp.119-124) quienes estudiaron “Factores pronósticos relacionados con la supervivencia del cáncer de mama”, dicho estudio tuvo como objetivo evaluar los factores pronósticos del cáncer de mama en mujeres mexicanas, en el cual se incluyó a 432 mujeres con cáncer de mama admitidas de 1990 a 1999 en el Hospital General de México para evaluar la supervivencia mediante las técnicas de Kaplan Meier y los métodos de riesgos proporcionales de Cox. Resultados. La supervivencia global a cinco años fue de 58.9%. En donde se tuvo que a menor supervivencia corresponde a los estadios clínicos IIIB (47.5%), IIIA (44.2%) y IV (15%), la metástasis hematológica fue de 21.4% y el tumor positivo en bordes quirúrgicos de 12.5%. La invasión linfovascular (RR= 1.9; IC95% 1.3-2.8), el estadio clínico IV (RR=14.8; IC95% 5.5-39.7) y el tumor en bordes quirúrgicos (RR= 2.4; IC95% 1.2-4.8) fueron los principales factores pronósticos. Con el estudio realizado se dio consistencia a los criterios de diagnóstico y tratamiento de las mujeres con cáncer de mama atendidas en México y toman en consideración las condiciones de la tumoración, como extensión de la malformación, etapa clínica y estado de los ganglios linfáticos antes de decidir la conducta terapéutica inicial.

Un segundo estudio de (Fernandez, et al., 2009, pp.1-6) con título “Factores pronósticos en el cáncer de mama” se valoró la posible influencia pronostica de ciertos factores, como el oncogén c-erB-

2. El gen supresor p53 y la neovascularización en el cáncer de mama. Se tomó un total de 112 mujeres diagnosticadas de carcinoma ductal infiltrante de tipo clásico y tratadas mediante mastectomía radical modificada. Se hizo un seguimiento durante 60 meses y la detección de los factores pronósticos se hizo mediante inmunohistoquímica en el tejido tumoral en formol e incluido en parafina. Dentro de la serie de pacientes de este estudio se evidenció, que únicamente el estadio tumoral (TNM) y la neovascularización se comportaron como factores pronósticos independientes en el carcinoma ductal de mama. Además, apporto con la conclusión de que la detección de la densidad vascular en el tumor puede ser un factor pronóstico de importancia en las pacientes con cáncer de mama, sobre todo en aquellas sin metástasis ganglionares.

Un tercer estudio “Factores pronósticos de cáncer de mama”; en el que se analiza la importancia de que los factores pronósticos se deben diferenciar de los factores predictivos. (Ortega, et al., 2018, pp.130-136) refiere a que se tiene que un factor pronóstico es cualquier medición utilizable en el momento de la cirugía que correlaciona con el intervalo libre de enfermedad o supervivencia global en ausencia de tratamiento adyuvante sistémico y como resultado es capaz de correlacionar con la historia natural de la enfermedad. En contraste, un factor predictivo es cualquier medición asociada con respuesta a un tratamiento dado. Entre los factores pronósticos del cáncer de mama existen factores clínicos, histológicos, biológicos, genéticos y psicosociales. En esta revisión de los factores pronósticos psicosociales se demostró además que el estrés y la depresión son factores pronósticos negativos en las pacientes con cáncer de mama. Además, se debe recordar que la evaluación de un solo parámetro pronóstico ayuda, pero no es útil para la gestión clínica y terapéutica de la paciente.

La relevancia de este trabajo de investigación radica en que, a pesar de los años y el avance de herramientas tecnológicas, especialización de tratamientos, así como la innovación de nuevos medicamentos mejorados que ayudan mucho más en el proceso de control del padecimiento, aun cabe la necesidad de mejorar tratamientos como: quimioterapias, radioterapias, entre otros, en cuanto a eficacia; así como también en la identificación y mejor definición de estos factores pronósticos.

La Sociedad de Lucha Contra el Cáncer del Ecuador, Matriz Guayaquil y su Instituto Oncológico Nacional Dr. Juan Tanca Marengo constituye uno de los principales centros de prevención, diagnóstico, tratamiento y paliación de cáncer en la República del Ecuador, a éste, acuden pacientes de todo el territorio continental e insular. Enfocando todas sus energías y esfuerzos en la prevención, detección y tratamiento del cáncer. No permiten que distinción social o económica alguna evite que sus pacientes tengan los medios necesarios para derrotar al cáncer. Por lo que esta investigación trata de ser un apoyo en esa lucha, en la pronta detección, que permita contrarrestar esta enfermedad, esto a través de la identificación de los factores más incidentes en la mortalidad, a causa de esta enfermedad morbosa que no merma al contrario avanza. Por lo que el estudio de cada uno de estos factores pronósticos cobra sentido, al tratar de saber cuál es la

relación existente entre ellos, así como la importancia de cada uno de ellos, a través de técnicas estadísticas como la regresión logística se quiere dar solución a esta problemática, en búsqueda de una buena prevención secundaria y terciaria, así como de la aplicación de un tratamiento acorde y mejorado que permita tener probabilidades de recuperación más favorables frente a la enfermedad.

El estudio de factores pronósticos, trata principalmente de identificar factores pronósticos de mortalidad en pacientes con cáncer de mama en SOLCA Guayaquil, en el periodo 2015-2020, teniendo en cuenta estos factores delimitados y estudiados en la población ecuatoriana, basándose principalmente en el análisis de la significancia de estos factores acorde a la enfermedad, buscando además relaciones existentes entre factores pronósticos de mortalidad en pacientes con cáncer, así como predecir el comportamiento de la enfermedad en un paciente con cáncer de mama, todo ello a través de la aplicación de regresión logística y demás técnicas estadísticas acordes al tipo de estudio y tipo de datos relacionados al tema en estudio.

Dado que, para brindar mayor precisión y eficacia en la identificación de estos factores en pro de la prevención adecuada de esta enfermedad, es fundamental tener en cuenta características como: la fácil predicción de estos factores, así como su disponibilidad, sensibilidad y reproducción en cualquier situación clínica, el estudio de lo antes mencionado será de gran importancia y apoyo en búsqueda de mejores tratamientos o tratamientos más eficaces, así como de la prevención adecuada, que de manera directa ayuden en el control de las cifras de mortalidad alarmantes en la población ecuatoriana y dar probabilidades esperanzadoras a todas aquellas personas que estén afectadas por este padecimiento. En Ecuador son pocos los estudios acerca de la identificación de factores pronósticos de mortalidad en cáncer de mama; este estudio tendrá gran relevancia, como fuente de información, y pauta para posibles estudios futuros.

OBJETIVOS

General

Identificar factores pronósticos de mortalidad en pacientes con cáncer de mama en SOLCA Guayaquil, en el periodo 2015-2020.

Específicos

- Aplicar regresión logística para la identificación de factores pronósticos de mortalidad en pacientes con cáncer de mama en SOLCA guayaquil en el periodo 2015-2020.
- Analizar relaciones existentes entre factores pronósticos de mortalidad en pacientes con cáncer.
- Predecir el comportamiento de la enfermedad en un paciente con cáncer de mama.

CAPÍTULO I

1. MARCO TEÓRICO REFERENCIAL

1.1. Cáncer y Tumor

Ante el desconocimiento de temas similares en muchas ocasiones se tiende a inferir directamente que todo tumor o masa en el cuerpo, es cáncer sin embargo esto no es así, por lo que partiendo de lo más simple tumor, es una masa anormal de tejido corporal (Media Plus, 2020, párr.1). Estos tumores pueden ser benignos (aquellos que no resultan ser cancerosos, es decir no producen células cancerígenas), y los tumores malignos (aquellos que por la formación de células cancerígenas estos se vuelven cancerosos). Partiendo de esta base y según OMS (2007, párr.1):

El cáncer es un proceso de crecimiento y diseminación incontrolados de células. Puede aparecer prácticamente en cualquier lugar del cuerpo. Muchos tipos de cáncer se podrían prevenir evitando la exposición a factores de riesgo comunes como el humo de tabaco. Además, un porcentaje importante de cánceres pueden curarse mediante cirugía, radioterapia o quimioterapia, especialmente si se detectan en una fase temprana.

Los principales tipos de cáncer identificados son: Cáncer colorectal, Cáncer de estómago, Cáncer de hígado, Cáncer de mama, Cáncer de ovario, de las trompas de Falopio y peritoneal, Cáncer de próstata, Cáncer de pulmón de células no pequeñas, Cáncer de riñón, Cáncer oral y orofaríngeo, Linfoma no Hodgkin.

1.2. Cáncer de mama

Con base a ideas planteadas con anterioridad y desglosando el tema en lo directamente relacionado se indica que el cáncer de mama es una enfermedad en la cual las células de la mama se multiplican sin control. Existen distintos tipos de cáncer de mama. El tipo de cáncer de mama depende de qué células de la mama se vuelven cancerosas así se tiene que el cáncer de mama puede ser invasivo o no invasivo (ASCO, 2005, párr.1).

Refiriéndose al cáncer de mama invasivo como aquel cáncer que se disemina a los tejidos adyacentes y cáncer de mama no invasivo es aquel que no se extiende más allá de los conductos de la leche, ni los lobulillos de la mama. La mayoría de los tipos de cáncer de mama comienzan en los conductos o lóbulos, y se denominan carcinomas ductales o carcinomas lobulares (ASCO, 2014, párr.1).

1.2.1. Tipos de cáncer de mama

El cáncer de mama puede comenzar en distintas partes de la mama. La mayoría de los cánceres de mama comienzan en los conductos o en los lobulillos, así dependiendo de las diferentes áreas de formación se tiene carcinoma ductal, estos cánceres se originan en las células que recubren internamente los conductos de la leche y conforman la mayoría de los cánceres de mama, carcinoma ductal in situ, este cáncer se ubica únicamente en el conducto, carcinoma ductal invasivo o infiltrante, este cáncer se disemina fuera del conducto, carcinoma lobular invasivo, este cáncer se origina en los lóbulos. El cáncer de mama puede diseminarse fuera de la mama a través de los vasos sanguíneos y los vasos linfáticos. Cuando el cáncer de mama se disemina a otras partes del cuerpo, se dice que ha hecho metástasis (CDC, 2020, párr.5).

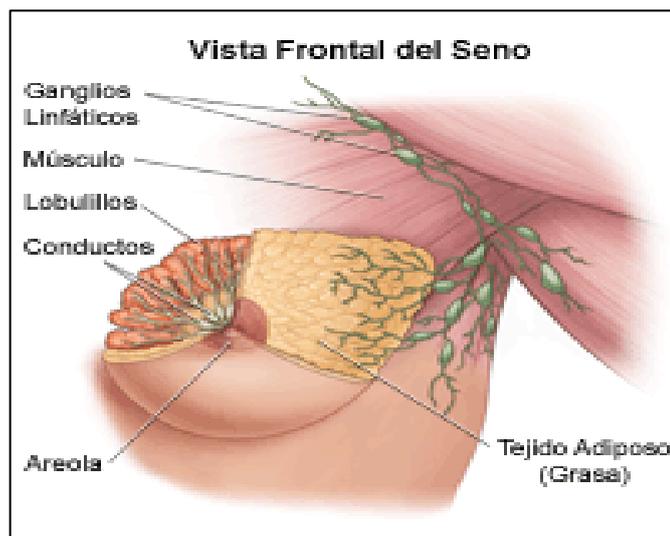


Figura 1-1: Anatomía de la mama.

Fuente: (SCH, 2008).

1.2.2. Partes de la mama

Desde el punto de vista anatómico, las mamas pueden presentar distintas formas y tamaños, dependiendo de la edad (niñez, pubertad, pleno desarrollo, menopausia, posmenopausia), talla, estado nutricional, así como de la morbilidad propia de la mama o comorbilidades; una mama es de mayor tamaño que la otra, pero esto no representa ninguna patología; desde el punto de vista descriptivo, a la mama se le divide en cinco elementos: cuatro cuadrantes y una cola; se trazan dos líneas imaginarias, una vertical y otra horizontal que se entrecruzan en el pezón de tal forma que se tiene un cuadrante superior externo, uno superior interno, otro inferior externo y un cuarto inferior interno, a los que se le agrega un área central circular, la areola (Ocampo & Archundia, 2010, párr.2).

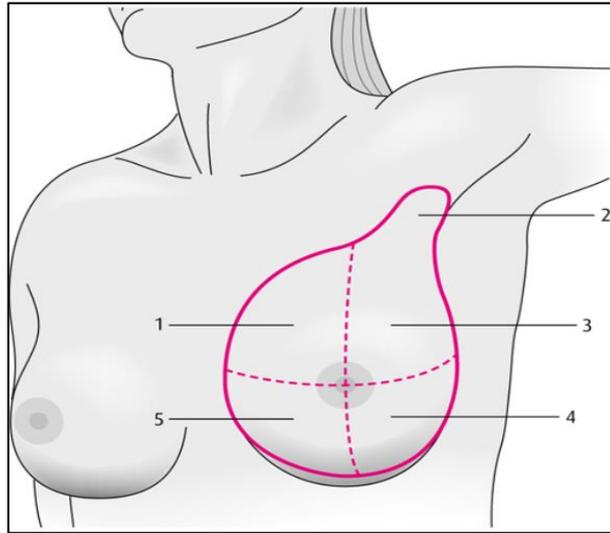


Figura 2-1: Partes de la mama.

Fuente: (Ocampo & Archundia, 2010).

1.2.3. Síntomas del cáncer de mama

Entre los síntomas más comunes dentro del cáncer de mama se puede presentar: Un bulto o engrosamiento en la mama que se siente diferente del tejido que la rodea, además cambio en cuanto a tamaño o forma de la mama, también cambios en la piel que se encuentra sobre la mama, como formación de hoyuelos, otro síntoma que se presenta, es la descamación, desprendimiento de la piel, formación de costras y pelado del área pigmentada de la piel que rodea el pezón (areola) o la piel de la mama, enrojecimiento o pequeños orificios en la piel que se encuentra sobre tu mama, como la piel de una naranja (Mayo Clinic, 2019, párr.4).



Figura 3-1: Partes de la mama.

Fuente: (Colon 15, 2015).

1.2.4. Tratamiento

Un tratamiento es un conjunto de medios que se utilizan para aliviar o curar una enfermedad, llegar a la esencia de aquello que se desconoce o transformar algo. El concepto es habitual en el ámbito de la medicina (Pérez Julián & Merino, 2013, párr.2). Se debe tener en cuenta que para tener un plan de tratamiento influye mucho la biología y el comportamiento del cáncer, dado que el tamaño del tumor puede variar, no hay descuidar el mismo ya que por más pequeños que suelen ser estos pueden llegar a desarrollarse y alcanzar mayor magnitud, de una manera acelerada, por lo contrario, también hay casos que resultan de mayor magnitud, pero su desarrollo es pequeño y de manera lenta. Las opciones y recomendaciones de tratamiento son muy personalizadas y dependen de varios factores (ASCO, 2014, párr.2). Los tratamientos más comunes y usados son:

1.2.4.1. Cirugía

“La cirugía es la extirpación del tumor y de parte del tejido circundante sano durante una operación. La cirugía también se utiliza para evaluar los ganglios linfáticos cercanos axilares, que se encuentran debajo del brazo” (ASCO, 2014, párr.16). El médico que se encarga de dicha acción se lo conoce como radioncólogo.

1.2.4.2. Radioterapia

“La radioterapia es el uso de rayos X u otras partículas con alta potencia para destruir las células cancerosas. El médico que se especializa en administrar radioterapia para tratar el cáncer se denomina radioncólogo” (ASCO, 2014, párr.41).

1.2.4.3. Terapias con medicamentos

La terapia sistémica es el uso de medicamentos para destruir las células cancerosas. Este tipo de medicamentos se administra a través del torrente sanguíneo para llegar a las células cancerosas en todo el cuerpo. El profesional que receta terapias sistémicas es un oncólogo clínico, un médico que se especializa en el tratamiento del cáncer con medicamentos. Los métodos frecuentes para administrar terapias sistémicas incluyen un tubo intravenoso (i.v.) que se coloca en una vena con una aguja, una inyección en un músculo o debajo de la piel, o una pastilla o cápsula que se traga (por vía oral) (ASCO, 2014, párr.67).

1.3. Factor pronóstico

El factor pronóstico es aquella situación, afección o característica del paciente que puede usarse para calcular la probabilidad de recuperación de una enfermedad o la probabilidad de que la enfermedad recidive (vuelva) (NIH, 2011, párr.1), que a diferencia del factor predictivo que se define como aquella situación, afección o característica del paciente que puede usarse para calcular la probabilidad de recuperación de una enfermedad o la probabilidad de que la enfermedad recidive (vuelva). Según (Sabadell, et al., 2017, pp.15-19) un factor pronóstico, para considerarse útil en el estudio del paciente oncológico debe cumplir una serie de requisitos; es así como debe aportar valor pronostico independiente, significativo y valido mediante estudios clínicos bien diseñados, su determinación debe ser factible, reproducible, ampliamente disponible y que supere los controles de calidad requeridos en cada caso particular, y que además su resultado sea fácilmente interpretable por el clínico. Los factores pronósticos de la enfermedad local de forma general y desde un punto de vista práctico se puede dividir.

1.3.1. Relacionados con el paciente

Son todos aquellos que se relacionan con las características de la persona que padece la enfermedad: edad, sexo, antecedentes, comorbilidades, estado general, etc. Son factores fundamentalmente clínicos (Sabadell, et al., 2017, pp.15-19).

1.3.2. Relacionados con la neoplasia

Son aquellos que depende de las características de la biología del tumor (Sabadell, et al., 2017, pp.15-19). En el cáncer de mama puede ser histología, los receptores hormonales, HER2, tamaño tumoral, grado histológico, grado nuclear, invasión linfovascular, Ki67, afectación ganglionar axilar, rotuna capsular e invasión de la grasa peri ganglionar, subtipo histológico, perfiles genéticos, etc.

1.3.3. Relacionados con el tratamiento

Son factores pronósticos muy importantes, ya que obviamente la evolución de muchos canceres dependerá del tratamiento realizado sea esta cirugía, radioterapia, quimioterapia, hormonoterapia o nuevas terapias anti-diana (Sabadell, et al., 2017, pp.15-19).

1.3.4. Relacionados con el entorno

Son aquellos que se refieren a las posibilidades de recibir distintos tratamientos por parte de las pacientes debido a su entorno social, nivel cultural económico y zona geográfica. Así, la supervivencia de las mujeres afectadas por esta enfermedad va a depender de forma directa de la zona geográfica en donde viven y las condiciones sanitarias de la misma (Sabadell, et al., 2017, pp.15-19).

1.3.5. Principales factores pronósticos

Los factores pronósticos son importantes por cuanto nos ayudan a tratar mejor a los pacientes en función de sus deseos y posibilidades de supervivencia y de respuesta a los distintos tratamientos, así tenemos algunos de los más importantes (Sabadell, et al., 2017, pp.15-19).

1.3.6. Receptores hormonales y HER2

Las principales decisiones de tratamiento en cáncer de mama dependen de estos biomarcadores. Según (Sabadell, et al., 2017, pp.15-19) el problema es que algunos tumores cambian su expresión de receptores y/o HER2 entre el tumor inicial y la enfermedad metastásica, por lo que se recomienda la realización de una biopsia de algunas de las metástasis, siempre que sea posible en el momento del diagnóstico de la enfermedad diseminada. En caso de cambios en la expresión de estos biomarcadores la decisión de tratamiento debería ajustarse de forma general, a la determinación más reciente.

1.3.7. Localización de las metástasis

Es uno de los factores pronósticos más importantes. La presencia de múltiples lesiones metastásicas viscerales implica una reducción de la supervivencia frente a las pacientes con metástasis óseas o ganglionares (Sabadell, et al., 2017, pp.15-19).

1.3.8. Carga tumoral

Se convierte un factor importante al ser obvio que a mayor carga de enfermedad peor pronóstico.

1.3.9. Edad

Se convierte en factor importante ya que en estudio los pacientes más mayores es posible que no puedan tolerar determinados tratamientos (Sabadell, et al., 2017, pp.15-19).

1.4. Estadística

“La estadística es la ciencia que se encarga de recopilar, organizar, procesar, analizar e interpretar datos con el fin de deducir las características de un grupo o población objetivo, pero esta sería solo una visión estrecha de lo que comprende esta rama del saber” (Fernandez, et al., 2020, párr.1). Esta ciencia engloba muchos conceptos que permiten conocer y entender esta rama de manera que pueda ser aplicable, lo más fundamental y desde donde parte todo es la unidad estadística aquellas personas u objetos que contienen cierta información que se desea estudiar y que pertenecen a la población en estudio (Fernandez, et al., 2020, párr.2). De esta manera cuando a estas unidades se las agrupa se tiene a la población que es el conjunto de individuos o elementos que cumplen ciertas propiedades y entre los cuales se desea estudiar un determinado fenómeno (Fernandez, et al., 2020, párr.29). Por la alta numerosidad de datos y la complejidad que muchas veces estos conllevan por diversos factores, conviene mejor trabajar sobre un subconjunto de la población denominado como muestra, en consecuente a lo antes mencionado se derivan términos como: estadístico, parámetro, estimador, que son respectivamente: una función definida sobre los valores numéricos de una muestra (Fernandez, et al., 2020, párr.30), una función definida sobre los valores numéricos de características medibles de una población (Fernandez, et al., 2020, párr.32), y una función de los valores de una muestra que se elabora para indagar el valor de un parámetro de la población de la que procede la muestra (Fernandez, et al., 2020, párr.33).

1.4.1. Estadística Descriptiva

Es un conjunto de técnicas numéricas y gráficas para describir y analizar un grupo de datos, sin extraer conclusiones (inferencias) sobre la población a la que pertenecen. En este tema se introducirán algunas técnicas descriptivas básicas, como la construcción de tablas de frecuencias, la elaboración de gráficas y las principales medidas descriptivas de centralización, dispersión y forma que permitirán realizar la descripción de datos (Fernandez, et al., 2020, párr.20).

1.4.1.1. Variables

Para un mejor estudio de la población o muestra en cuestión es primordial el uso de variables las mismas son característica observable que varía entre los diferentes individuos de una población. Las variables pueden dividirse en cualitativas y cuantitativas (Fernandez, et al., 2020, párr.35). Para las variables cualitativas se usan con datos que representan categorías que son mutuamente excluyentes, aunque se utilicen números para cada categoría no representan cantidades, para su medición usamos escalas (Fernandez, et al., 2020, párr.39). Por lo que las mismas pueden ser: nominales siendo aquellas en las que no hay relación entre las categorías. Una escala nominal sólo permite clasificar (no jerarquizar ni ordenar) (Fernandez, et al., 2020, párr.39), y ordinales donde sus posibles categorías se encuentran jerarquizadas y ordenadas (Fernandez, et al., 2020, párr.39). Por otro lado, tenemos las variables cuantitativas en las cuales se usan con datos que se expresan mediante cantidades numéricas que permiten hacer operaciones matemáticas (Fernandez, et al., 2020, párr.36), de manera similar que las variables cualitativas, las variables cuantitativas pueden ser: discretas sólo puede tomar valores enteros (Fernandez, et al., 2020, párr.37), y continuas si sus posibles valores están en un conjunto infinito. Las podemos categorizar en intervalos (Fernandez, et al., 2020, párr.38).

1.4.1.2. Medidas de tendencia central

Son medidas estadísticas que resumen en un único valor, a un conjunto de datos, representando la centralización de los mismos, así tenemos: a la media, mediana y moda.

Media

Según (Quevedo, 2011, párr.4) es una “medida de tendencia central una de las más conocida y utilizadas, también es conocida como promedio aritmético esta puede ser representada por la letra griega μ cuando se trata del promedio del universo o población y por \hat{x} (léase X barra) cuando se trata del promedio de la muestra”.

$$\bar{x} = \frac{\sum_i^N X_i}{N} \quad (1-1)$$

Donde:

X_i = Conjunto de observaciones

N = Total de observaciones

Mediana

Es otra medida de tendencia central. “La mediana es el valor de la variable que ocupa la posición central, cuando los datos se disponen en orden de magnitud. Es decir, el 50% de las observaciones tiene valores iguales o inferiores a la mediana y el otro 50% tiene valores iguales o superiores a la mediana” (Quevedo, 2011, párr.8).

En caso de que el número total de observaciones sea par:

$$\tilde{x} = X_{\frac{N}{2}} \quad (2-1)$$

Donde:

$$X_{\frac{N}{2}} = \text{Observación de la posición } \frac{N}{2}$$

N = Total de observaciones

En caso de que el número total de observaciones sea impar:

$$\tilde{x} = X_{\frac{N+1}{2}} \quad (3-1)$$

Donde:

$$X_{\frac{N+1}{2}} = \text{Observación de la posición } \frac{N+1}{2}$$

N = Total de observaciones

Moda

Se define “como el valor de la variable que más se repite. En un polígono de frecuencia la moda corresponde al valor de la variable que está bajo el punto más alto del gráfico. Una muestra puede tener más de una moda” (Quevedo, 2011, párr.10).

$$M_o = \text{maximo}(frecuencia(X_i)) \quad (4-1)$$

Donde:

X_i = Conjunto de observaciones

1.4.1.3. Medidas de dispersión

Otras medidas en estudio dentro de la estadística son las medidas de dispersión, según (Laguna, 2006, pp. 1-12) las “medidas de dispersión entregan información sobre la variación de la variable. Pretenden resumir en un solo valor la dispersión que tiene un conjunto de datos. Las medidas de dispersión más utilizadas son: Rango de variación, Varianza, Desviación estándar, Coeficiente de variación”.

Rango

El rango de variación se define como la diferencia entre el mayor valor de la variable y el menor valor de la variable (Laguna, 2006, pp. 1-12).

$$Rng = \text{maximo}(X_i) - \text{minimo}(X_i) \quad (5-1)$$

Donde:

X_i = Conjunto de observaciones

Varianza

“Se representa con el símbolo σ^2 (sigma cuadrada) para el universo o población y con el símbolo s^2 (s cuadrado), cuando se trata de la muestra. Esta se expresa en unidades de variable al cuadrado y la desviación estándar simplemente en unidades de variable” (Laguna, 2006, pp.1-12).

$$\sigma^2 = \frac{\sum_i^N (X_i - \bar{x})^2}{N} \quad (6-1)$$

Donde:

\bar{x} = Media de las observaciones

X_i = Conjunto de observaciones

N = Total de observaciones

Desviación estándar

“Es la raíz cuadrada de la varianza, se representa por σ (sigma) cuando pertenece al universo o población y por “s”, cuando pertenece a la muestra. σ^2 y σ son parámetros, constantes para una población particular; s^2 y s son estadígrafos, valores que cambian de muestra en muestra dentro de una misma población” (Laguna, 2006, pp.1-12).

$$\sigma = \sqrt{\frac{\sum_i^N (X_i - \bar{x})^2}{N}} \quad (7-1)$$

Donde:

\bar{x} = Media de las observaciones

X_i = Conjunto de observaciones

N = Total de observaciones

Coefficiente de variación

Es una medida de la dispersión relativa de los datos. Se define como la desviación estándar de la muestra expresada como porcentaje de la media muestral (Laguna, 2006, pp.1-12).

$$CV = \frac{\sigma_x}{|\bar{x}|} \quad (8-1)$$

Donde:

\bar{x} = Media de las observaciones

σ_x = Desviación estandar de las observaciones

1.4.2. Análisis de dependencia

1.4.2.1. Dependencia de variables cuantitativas

El estudio de dependencia de variables cuantitativas se puede estudiar, a través del coeficiente de Pearson, Spearman, T de Kendall; entre otros. El coeficiente de Pearson es una medida que cuantifica la relación entre dos variables, en base a criterios definidos (Laguna, 2006, pp. 22-26). Así se tiene:

$$\rho(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (9-1)$$

Donde:

\bar{x} = Media de las observaciones de x

\bar{y} = Media de las observaciones de y

x_i = Observaciones de x

y_i = Observaciones de y

En base al cálculo del coeficiente se tiene que serán dependientes siempre y cuando dicho coeficiente $\rho \cong \pm 1$; o por lo contrario si $\rho \cong 0$, están variables en estudio serán independientes; es decir no habrá relación entre sí.

1.4.2.2. Dependencia de variables cualitativas

Si se quiere analizar la relación que hay entre una y otra variable de carácter cualitativa, también hay pruebas que se focalizan a este tipo de estudio. Uno de la prueba más populares es la prueba de independencia de Chi-cuadrado (χ^2). Si la prueba no muestra relación entre las dos variables, esto indica que conocer el valor de una variable no proporciona información de la otra (Laguna, 2006, pp. 22-23).

Dicha prueba se basa en una prueba de hipótesis que contrasta:

H_0 = Las variables son independientes, no guardan relación. (10-1)

H_1 = Las variables son dependientes, guardan relación.

Estudiada a un nivel alfa (α) que se considere, se basa en el cálculo de la estadística de prueba, con la que en comparación del valor calculo o critico se comparan el rechazo o no rechazo de la hipótesis planteada, para ello usando los siguientes grados de libertad: $(n_1 - 1) * (n_2 - 1)$. Teniendo así:

$$\chi^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (11-1)$$

Donde:

O_{ij} = Frecuencias Observadas

E_{ij} = Frecuencias Esperadas

1.4.3. Regresión logística

Es una técnica multivariante que nos permite estimar la relación existente entre una variable dependiente no métrica, en particular dicotómica y un conjunto de variables independientes métricas o no métricas (Amat, 2016, párr.1). Para un mayor alcance en estudios en los que se necesite el uso de esta técnica se tiene por una parte a la regresión logística simple, desarrollada por David Cox en 1958, siendo un método de regresión que permite estimar la probabilidad de una variable

cualitativa binaria en función de una variable cuantitativa (Amat, 2016, párr.1), cuyo modelo se define como:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_j \quad (12-1)$$

Donde:

Y_i = Variable respuesta

X_i = Variable predictivas

β_i = Coeficientes de las variables predictivas

ε_j = Error

1.4.3.1. Regresión Logística Binaria

“Los modelos de regresión logística binaria se resumen a la estimación de un modelo, cuando se tiene como variable respuesta una variable de carácter dicotómica, en las que se tiene como probabilidades, el éxito o el fracaso” (Salcedo, 2011, pp.20-56).

1.4.3.2. Regresión Logística Múltiple

La regresión logística múltiple según (Amat, 2016, párr.37), es un método de regresión que permite estimar la probabilidad de una variable cualitativa en función de varias variables sean estas cualitativas o cuantitativas, con su modelo:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$$

Y_i = Variable respuesta

X_{ik} = Variable predictivas (variables dummy)

β_i = Coeficientes de las variables predictivas (variables dummy)

(13-1)

Ajuste del modelo

El modelo debe ser aquél más reducido que explique los datos (principio de parsimonia), y que además sea técnicamente congruente e interpretable. Hay que tener en cuenta que un mayor número de variables en el modelo implicará mayores errores estándar. Deben incluirse todas aquellas variables que se consideren técnicamente importantes para el modelo, no debería dejarse de incluir toda variable que en un análisis univariado previo demostrara una relación “suficiente” con la variable dependiente (Maurandi, et al., 2019, pp.100-120).

Bondad de Ajuste

“Se trata de definir criterio para establecer la significancia del modelo, así como de las variables que implican directamente en el modelo de regresión” (Maurandi, et al., 2019, pp.100-120).

Criterio de Máxima Verosimilitud

Recordemos que regresión logística lo que hacemos es predecir la probabilidad ($P(Y)$) de que un evento (Y) ocurra para una persona (i) dada, basado en las observaciones de si el evento ocurre o no para esa persona (denotamos esto como Y_i , el resultado real para la i -ésima persona). Así, para esa i -ésima persona el suceso Y toma los valores 0 (no ocurre) o 1 (ocurre), y el valor predicho, $P(Y)$, variará entre 0 (no hay probabilidad de que el evento ocurra) y 1 (el evento ocurre con seguridad) (Maurandi, et al., 2019, pp.100-120).

Por tanto, al igual que en regresión múltiple, podemos usar estos valores observados y predichos para evaluar el ajuste del modelo. La medida que usamos es log-likelihood (logaritmo de la razón de verosimilitud), se basa por tanto en la suma de las probabilidades asociadas con los resultados estimados y los valores reales (Maurandi, et al., 2019, pp.100-120).

El estadístico log-likelihood es análogo a la suma de cuadrados residual en la regresión múltiple en el sentido de que es un indicador cuánta información sin explicar queda en la variable respuesta tras haber ajustado el modelo. Grandes valores del log-likelihood indican un pobre ajuste del modelo, cuanto mayor sea este valor, más variabilidad sin explicar queda en el modelo (Maurandi, et al., 2019, pp.100-120).

$$\log - likelihood = \sum_{i=1}^N [Y_i \log P(Y_i) + (1 - Y_i) \log(1 - P(Y_i))] \quad (14-1)$$

Devianza

“Otro indicador importante para estudiar el ajuste del modelo logístico es la devianza que se define como el doble logaritmo del estadístico de verosimilitud” (Maurandi, et al., 2019, pp.100-120).

La devianza tiene una distribución χ^2 y compara los valores de la predicción con los valores observados en dos momentos: el modelo sin variables independientes, sólo con la constante y el modelo con las variables predictoras introducidas (Maurandi, et al., 2019, pp.100-120).

Simplemente tomamos la devianza del nuevo modelo y le restamos la devianza del modelo referencia. Esta diferencia se lo conoce como ratio-likelihood y tiene una distribución χ^2 con $k-1$ grados de libertad, el número de parámetros del nuevo modelo, menos el número de parámetros del modelo referencia que es siempre 1.

$$x^2 = 2LL(nuevo) - 2LL(referencial) \quad (15-1)$$

Donde:

$$gl: k_{nuevo} - 1$$

Odds, odds ratio y coeficientes

“El odds de un suceso es el cociente sus probabilidades de ocurrencia entre sus probabilidades de no ocurrencia, bajo unas determinadas condiciones” (Maurandi, et al., 2019, pp.100-120).

$$odds_c(evento) = \frac{P(evento)}{1 - P(evento)} \quad (16-1)$$

La medida más crucial para la interpretación del modelo logístico es el valor del odds ratio, que es la exponencial del valor B (e. d. $\exp(B)$) del modelo de regresión, y se define como el indicador del cambio en los odds resultante del cambio de una unidad en el predictor. Tenemos por tanto los odds antes y después de aumentar la variable predictora una unidad. Con todo esto es fácil ahora calcular el cambio proporcional en los odds simplemente dividiendo los odds después del cambio entre los odds antes del cambio, es lo que se conoce como odds-ratio (Maurandi, et al., 2019, pp.100-120).

$$\text{oddsRatio} = \frac{\text{odds tras cambio en una unidad } X_j}{\text{odds originales}} \quad (17-1)$$

En el caso de que tuviéramos un modelo logístico con más de una variable, (Maurandi, et al., 2019, pp.100-120) refiere a ello odds ratio del predictor X_j al cociente del odds de ocurrencia al aumentar X_j en una unidad con respecto a no aumentarla, cuando se mantienen constantes el resto de las predictoras X_1, \dots, X_k .

$$\text{oddsRatio}(X_j) = \frac{\text{odds tras cambio en una unidad } X_j}{\text{odds originales}} \quad (18-1)$$

La interpretación del odds-ratio es que valores mayores que 1 indican que si el predictor aumenta los odds de la variable dependiente crecen. Inversamente, un valor menor que 1 indica que tal como el predictor aumente el odds del resultado decrece (Maurandi, et al., 2019, pp.100-120).

Probabilidades

El cálculo de las probabilidades se resume, a la probabilidad que tomara cierto caso una vez obtenido el modelo suficientemente bueno para predecir.

$$P\left(Y = \frac{1}{X}\right) = \frac{\exp(b_0 + \sum_{i=1}^n b_i x_i)}{1 + \exp(b_0 + \sum_{i=1}^n b_i x_i)}$$

Donde:

$$P\left(Y = \frac{1}{X}\right): \text{Probabilidad de que } Y \text{ tome el valor } 1 \quad (19-1)$$

X : Conjunto de n covariables

b_0 : Constante del modelo o término independiente

b_i : Coeficientes de las covariables

Matriz de confusión

La matriz de confusión es una herramienta muy útil para valorar cómo de bueno es un modelo y su clasificación basado en aprendizaje automático. En particular, sirve para mostrar de forma explícita cuándo una clase es confundida con otra, lo cual nos, permite trabajar de forma separada con distintos tipos de error. Debido a que hay dos posibles valores reales y dos posibles valores de predicción o predichos. A partir de estas opciones podemos crear lo que se conoce como la matriz de confusión con 4 resultados posibles (Recuero, 2020, párr.1).

La matriz de confusión de un problema de clase n es una matriz $n \times n$ en la que las filas se nombran según las clases reales y las columnas, según las clases previstas por el modelo. Sirve para mostrar de forma explícita cuándo una clase es confundida con otra. Por eso, permite trabajar de forma separada con distintos tipos de error. De esta forma, la matriz de confusión para este modelo tiene etiquetadas sus filas con las clases reales, y sus columnas, con las predichas por el modelo (Recuero, 2018, párr.1).

Tabla 1-1: Matriz de Confusión.

Matriz de confusión	Estimado por modelo			
	Negativo (N)	Positivo (P)		
Real	Negativo	a: (VN)	b: (FP)	Precisión ("Recall") Porcentaje predicciones positivas detectadas
	Positivo	c: (FN)	d: (VP)	$\frac{d}{b + d}$
		Sensibilidad exhaustividad ("Recall") Porcentaje casos positivos detectados	Especificidad ("Specify") Porcentaje casos negativos detectados	Exactitud ("Accuracy") Porcentaje predicciones correctas (No sirve en datasets poco equilibrados)
		$\frac{d}{d + c}$	$\frac{a}{a + b}$	

Fuente: (Recuero, 2018).

Verdadero positivo

“El valor real es positivo y la prueba predice un positivo. La persona está enferma y la prueba así lo demuestra” (Recuero, 2018, párr.16).

Verdadero negativo

“El valor real es negativo y la prueba predice un negativo. La persona no está enferma y la prueba así lo demuestra” (Recuero, 2018, párr.16).

Falso negativo

“El valor real es positivo, y la prueba predice un negativo. La persona está enferma, pero la prueba dice de manera incorrecta que no lo está. También se denomina error tipo II en estadística” (Recuero, 2018, párr.17).

Falso positivo

“El valor real es negativo, y la prueba predice un positivo. La persona no está enferma, pero la prueba nos dice de manera incorrecta que si lo está. También se denomina error de tipo I en estadística” (Recuero, 2018, párr.17).

Exactitud

“Se refiere a lo cerca que está el resultado de una medición del valor verdadero. En términos estadísticos, la exactitud está relacionada con el sesgo de una estimación” (Recuero, 2018, párr.22).

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN}$$

Donde:

VP: verdadero positivo (20-1)

VF: verdadero falso

FP: falso positivo

FF: falso negativo

Precisión

“Se refiere a la dispersión del conjunto de valores obtenidos a partir de mediciones repetidas de una magnitud. Cuanto menor es la dispersión mayor la precisión” (Recuero, 2018, párr.23).

$$\text{Precisión} = \frac{VP}{VP + FP}$$

Donde: (21-1)

VP: verdadero positivo

FP: falso positivo

Error Precisión

Se refiere al error que se presenta en el modelo, al momento de clasificar los diferentes casos dados.

$$\text{Error} = \frac{FP + FN}{VP + VN + FP + FN}$$

Donde:

VP: verdadero positivo (22-1)

VF: verdadero falso

FP: falso positivo

FF: falso negativo

Curva ROC

Una mejor manera para evaluar la capacidad de un modelo de regresión logística binaria para clasificar con exactitud las observaciones es una curva de rendimiento diagnóstico (ROC). Una curva ROC se construye generando varias tablas de clasificación, para valores de corte que oscilan entre 0 y 1 y calculando la sensibilidad y la especificidad para cada valor. La sensibilidad se grafica en función de 1 - la especificidad para construir una curva ROC (Hosmer & Lemeshow, 2000, pp.156-164).

El área por debajo de la curva ROC (AUC) es una medida de discriminación; un modelo con un área por debajo de la curva ROC alta sugiere que el modelo puede predecir con exactitud el valor de la respuesta de una observación (Hosmer & Lemeshow, 2000, pp.156-164).

(Hosmer & Lemeshow, 2000, pp.156-164), facilitan reglas generales para interpretar valores de AUC. Así se tiene sus reglas se obtienen los lineamientos generales siguientes:

Tabla 2-1: Criterio AUC.

Valor	Descripción
$AUC=0.5$	No discriminación
$0.7 \leq AUC < 0.8$	Discriminación aceptable
$0.8 \leq AUC < 0.9$	Excelente discriminación
$AUC \geq 0.9$	Casi perfecta discriminación

Realizado por: Pinduisaca A, Jhonnatan F. 2022.

CAPÍTULO II

2. MARCO METODOLÓGICO

2.1. Tipo de investigación

En el presente estudio se utilizó un método de investigación mixta, es decir combinar una investigación cualitativa como una investigación cuantitativa, esto se da debido a que dentro de la investigación se estudió variables tanto de tipo numérico como por ejemplo la edad de los pacientes diagnosticados con cáncer de mama y de tipo categórico como por ejemplo la ubicación de las células cancerígenas, todo esto por la importancia e información que estas variables aportaron al estudio. Es importante aclarar que es un estudio que persiguió un objetivo teórico, más no aplicado; ya que todos los resultados y conclusiones que se obtuvieron del mismo, no pretende ser aplicado de manera directa e inmediato. El estudio abarco una investigación descriptiva y explicativa; ya que con el apoyo de técnicas basadas en este tipo de investigación se describió cada una de las variables en cuestión, para tener una ligera idea de lo pasa con todas y cada una de las variables utilizadas dentro del estudio y a la par explicar todas y cada una de las relaciones existentes, o alguna posible identificación de nuevos factores pronósticos que sirvan de apoyo fundamental para próximos estudios.

La identificación de factores pronósticos de mortalidad en pacientes con cáncer de mama es una investigación netamente no experimental en el que se tratara de realizar una inferencia hipotética-deductiva, tratando de partir de hipótesis como: ¿La edad es un factor pronostico importante para tomar en cuenta al momento de predecir la supervivencia o reincidencia a la enfermedad de un individuo?, ¿La ubicación del tumor determina una probabilidad alta de no supervivencia?, entre otras. Dado que la información que se utilizó para el estudio corresponde a fuentes secundarias, y que en ningún momento se pretende seguir al sujeto durante el estudio, se tiene una investigación de tipo transversal (Patten & Newhart, 2018).

2.2. Diseño de la investigación

La identificación de factores pronósticos de mortalidad en cáncer de mama en pacientes de Solca durante el periodo 2015-2020; es una investigación de tipo mixto; combinando investigación cuantitativa y cualitativa, por el carácter y tipo de variables que se manejaran y analizaran dentro del estudio; entre ellas: edad, sexo, tamaño tumoral, ubicación del tumor; entre otras. Definiendo un estudio de tipo no experimental, por el hecho de que no se involucrara cambios en variables o seguimiento de individuos durante el periodo que tarde el estudio (Berger et al, 2018).

Dentro de este estudio se trabajó principalmente con las variables: edad, estatura, peso, tamaño tumoral, ubicación del tumor, grado histológico, estadio clínico, invasión linfovascular, grado de afectación.

2.3. Identificación de variables

- Edad
- Raza
- Estatura
- Peso
- Cirugía
- Tiempo de evolución
- Tamaño tumoral
- Ubicación del tumor
- Tipo de cáncer
- Estadio clínico
- Estudio ganglionar
- Estudio metastásico
- Estadio de tamaño
- Estado paciente

2.4. Operacionalización de los objetivos

Tabla 1-2: Operacionalización de los objetivos.

Objetivo General	Objetivo Especifico	Concepto	Indicador	Instrumento
Identificar factores pronósticos de mortalidad en pacientes con cáncer de mama en SOLCA Guayaquil, en el periodo 2015-2020	Aplicar regresión logística para la identificación de factores pronósticos de mortalidad en pacientes con cáncer de mama en SOLCA guayaquil en el periodo 2015-2020 Analizar posibles relaciones existentes entre	Dentro de este presente estudio, y con base a los objetivos principalmente se quiere estudiar los factores pronósticos que determinen la mortalidad o capacidad de supervivencia de las personas ante el	Los objetivos se trabajarán a partir de frecuencias.	Los datos que se van a ocupar se reservan como base en las históricas clínicas patológicas de cada uno de los individuos puestos a estudiar en el presente trabajo, misma información reposa en una base generaliza con la información

	factores pronósticos de mortalidad en pacientes con cáncer Estudiar la significancia de las variables en observación y su influencia para las demás variables y el estudio en general Predecir el comportamiento de la enfermedad en un paciente con cáncer	cáncer de mama diagnosticado.		concatenada de dichos pacientes.
--	---	-------------------------------	--	----------------------------------

Realizado por: Pinduisaca A, Jhonnatan F. 2022.

2.5. Localización del Estudio

En Ecuador la lucha contra el cáncer, está a cargo únicamente de la Sociedad Ecuatoriana de Lucha contra el cáncer, con sus distintos hospitales ubicados en zonas estratégicas del país, por tal motivo el estudio de factores pronósticos de mortalidad en pacientes con cáncer de mama se desarrolló en las instalaciones en su matriz, el Instituto Oncológico Nacional “Dr. Juan Tanca Marengo”, ubicado en la provincia del Guayas, cantón Guayaquil, parte norte de la ciudad, en las calles Av. Pedro Menéndez Gilbert, entre Pasaje 11 y Callejón 11, con coordenadas de latitud y longitud (-2.17598,-79.88061) respectivamente.

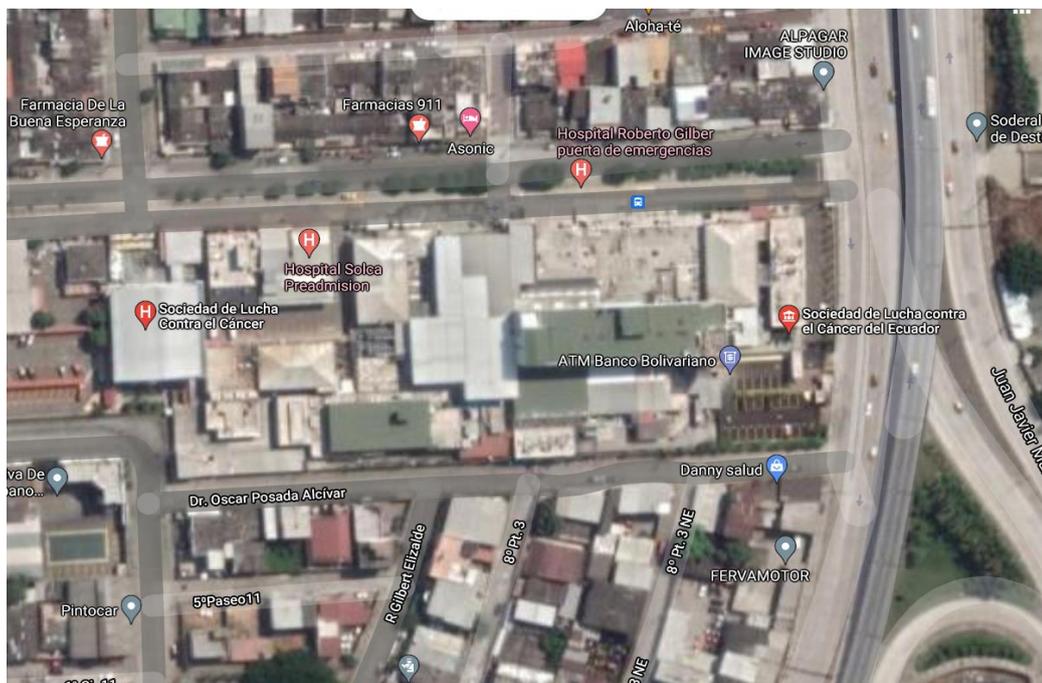


Figura 1-2-: Ubicación Geográfica del Hospital General SOLCA.

Fuente: (Google Maps 2021).

2.6. Población de estudio

El estudio tomo en cuenta a la población diagnosticada con cáncer de mama y tratada en el Hospital General de la Sociedad Ecuatoriana de Lucha Contra el cáncer (SOLCA) cede Guayaquil, además de aquella población que fue diagnosticado en otros hospitales o centros médicos adjuntos a SOLCA y que fue derivada al Hospital General de SOLCA cede Guayaquil, para su atención y tratamiento.

2.7. Tamaño de la muestra

Para el presente estudio se tomó en cuenta los datos de todos los pacientes diagnosticados con cáncer de mama y tratados en el Hospital General de la Sociedad Ecuatoriana de Lucha Contra el cáncer (SOLCA) cede Guayaquil, además de aquellos pacientes que fueron diagnosticados en otros hospitales o centros médicos adjuntos a SOLCA y que fueron derivados al Hospital General de SOLCA cede Guayaquil, para su atención y tratamiento, en el periodo 2015-2020.

2.8. Método de muestreo

En el estudio no se evidencio apoyo de ningún método de muestro, ya que fue utilizada toda población diagnosticada con cáncer de mama y tratada en el Hospital General de la Sociedad

Ecuatoriana de Lucha Contra el cáncer (SOLCA) sede Guayaquil, además de aquella población que fue diagnosticado en otros hospitales o centros médicos adjuntos a SOLCA y que fue derivada al Hospital General de SOLCA sede Guayaquil, para su atención y tratamiento, en el periodo 2015-2020.

2.9. Técnicas de recolección de datos

La información utilizada dentro del estudio es de tipo secundario, dicha información y datos de los pacientes diagnosticados con cáncer de mama, en el periodo 2015-2020, en Solca Guayaquil y demás centros adjuntos, que derivan a dicha sede; se recolecto primariamente a través de observación del sujeto, y fue constatada en historias clínicas, las mismas que se guardaron y reposan de manera digitalizada en las bases de datos del Hospital General de SOLCA Guayaquil, en el sistema institucional INTRANET, del cual se obtuvo la información de manera secundaria, tabulando cada uno de los datos dentro de Microsoft Excel, compactando todo en una sola base de datos, sujeta a depuración y análisis.

2.10. Modelo Estadístico

Para realizar un análisis y estudio correcto de los factores pronósticos de mortalidad en cáncer de mama; es indispensable primeramente realizar un análisis descriptivo de las variables en cuestión, utilizando indicadores descriptivos; tales como: medias de tendencia central, medias de dispersión; en las variables que sea pertinente dicho análisis, de este modo se tendrá una breve idea de cómo tratar a los datos de una mejor manera. También se utilizó pruebas de hipótesis estudiando la prueba de dependencia ji-cuadrado y estudio de coeficientes de correlación de Pearson, con el fin de estudiar la relación entre variables cualitativa y por otro lado la relación entre variables cuantitativas, respectivamente. Además se utilizó técnicas de regresión para analizar la importancia de los factores pronósticos de mortalidad por cáncer de mama estudiados con una posible estudio e implementación de variables dummy que ayuden a explicar de mejor manera ciertas variables de tipo cualitativo que van a ser usados dentro del estudio, que nos brinde un modelo adecuado con el cual sea factible y confiable realizar predicciones sobre la probabilidad de recuperación o reincidencia de cierto paciente hacia la enfermedad, que ira a la par del uso de técnicas de comprobación de supuestos, para corroborar dicha validez y confiabilidad del modelo planteado, lo que brindara mayor credibilidad y confiabilidad analítica e interpretativa.

CAPÍTULO III

3. RESULTADOS Y DISCUSIÓN DE LOS RESULTADOS

3.1. Operacionalización de las variables

Tras una rigurosa validación y depuración de la base de datos, se define una data catorce-variante con un total de 956 individuo estudiados.

Tabla 1-3: Operacionalización de las variables.

Variable	Etiqueta	Tipo	Medida	Categorías	Descripción
Edad	Edad	Cuantitativa discreta	Escala	No posee	Detallada la edad actual del paciente
Raza	Raza	Cualitativa nominal	Nominal	Soltera	Detallada la raza con la que se identifica el paciente
				Casada	
				Viuda	
				Divorciada	
				Unión Libre	
Estatura	Estatura	Cuantitativa continua	Escala	No posee	Detalla la estatura del paseante
Peso	Peso	Cuantitativa continua	Escala	No posee	Detalla el peso del paciente
Cirugía	Cirugía	Cualitativa dicotómica	Nominal	Si	Detalla si el paciente fue sometido a algún tipo de cirugía
				No	
Tiempo de evolución	Tiempo	Cuantitativa continua	Escala	No posee	Detalla el tiempo de evolución del paciente desde el inicio hasta la fecha de cierra del estudio
Tamaño tumoral	Tamaño	Cuantitativa continua	Escala	No posee	Detalla el tamaño tumoral alcanzado
Ubicación del cáncer	Ubicación	Cualitativa nominal	Nominal	Mama	Detalla la parte en la que se aloja el tumor
				Parte no especificada	
				Cola de Spence	
				Cuadrante superior interno de la mama	
				Cuadrante superior externo de la mama	
				Cuadrante inferior externo de la mama	

				Cuadrante inferior interno de la mama	
Tipo de cáncer	Tipo	Cualitativa nominal	Nominal	Carcinoma ductal in situ de la mama	Detalla que tipo de tumor es
				Tumor de comportamiento incierto de la mama	
Estadio clínico	Estadio	Cualitativa nominal	Nominal	0	Detalla el grado de estadio clínico del paciente
				1	
				2	
				3	
				4	
				x	
Estudio ganglionar	N	Cualitativa nominal	Nominal	x	Detalla el grado de propagación de los ganglios
				0	
				1	
				2	
				3	
				1a	
				1c	
				2a	
				2b	
				3a	
3b					
Estudio de Metástasis	M	Cualitativa nominal	Nominal	x	Detalla el grado de propagación del cáncer hacia el cuerpo
				0	
				1	
Estudio de tamaño	T	Cualitativa nominal	Nominal	x	Detalla el grado en cuanto a ubicación y tamaño del tumor principal
				0	
				1	
				2	
				3	
				4	
				1a	
				1b	
				1c	
				1d	
				2a	
				2b	
				2c	
				2d	
3a					
3b					

				3c	
				3d	
				4a	
				4b	
				4c	
				4d	
				is	
Estado paciente	Estado	Cualitativa dicotómica	Nominal	Venció	Detalla el estado del paciente en el tiempo de estudio
				No venció	

Realizado por: Pinduisaca A, Jhonnatan F. 2022.

3.2. Análisis Estadístico Descriptivo Exploratorio

Siempre una breve idea, de las variables inmiscuidas en el estudio, ayudara al análisis correcto de los datos, en búsqueda de una respuesta que conecte a todas estas variables, utilizando técnicas correctas, es por ello que un análisis descriptivo por variable es un paso importante antes de querer plantear un modelo, así se tiene:

3.2.1. Edad

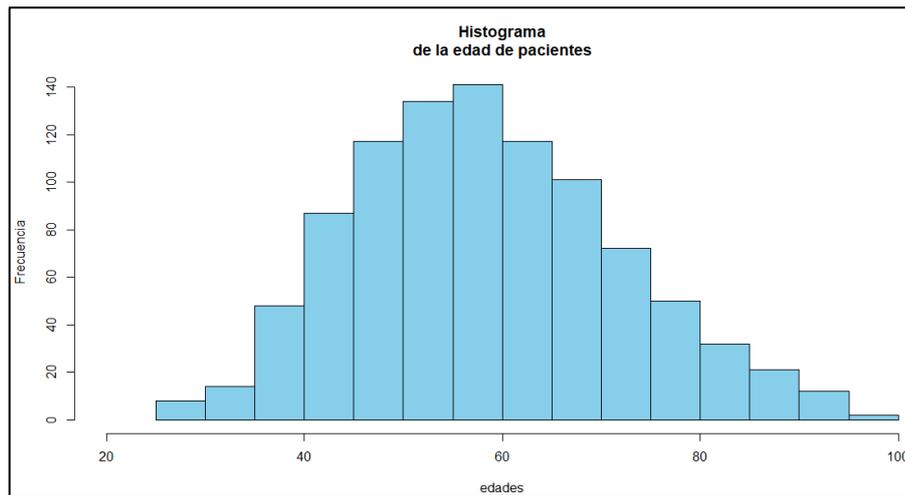


Gráfico 1-3: Distribución de la edad de pacientes diagnosticados con cáncer.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

De un total de 956 pacientes diagnosticados con cáncer de mama, las edades más representativas con un 18,83% del total de la población, son las que oscilan entre 53 y 60 años aproximadamente, aunque solo representan el 0,84% de la población total, es importante recalcar que hay pacientes con edades entre 92 a 99 años de edad aproximadamente, siendo este el grupo etario más avanzando dentro del estudio.

Tabla 2-3: Estadísticas de la edad de pacientes diagnosticados con cáncer de mama.

Limites	f	rf	rf(%)	cf	cf(%)
[26.73,33.298)	13	0.01	1.36	13	1.36
[33.298,39.866)	45	0.05	4.71	58	6.07
[39.866,46.435)	124	0.13	12.97	182	19.04
[46.435,53.003)	165	0.17	17.26	347	36.30
[53.003,59.571)	180	0.19	18.83	527	55.13
[59.571,66.139)	165	0.17	17.26	692	72.38
[66.139,72.707)	111	0.12	11.61	803	84.00
[72.707,79.275)	80	0.08	8.37	883	92.36
[79.275,85.844)	38	0.04	3.97	921	96.34
[85.844,92.412)	27	0.03	2.82	948	99.16
[92.412,98.98)	8	0.01	0.84	956	100.00
media					59
mediana					58
desviación					13,48

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.2.2. Raza

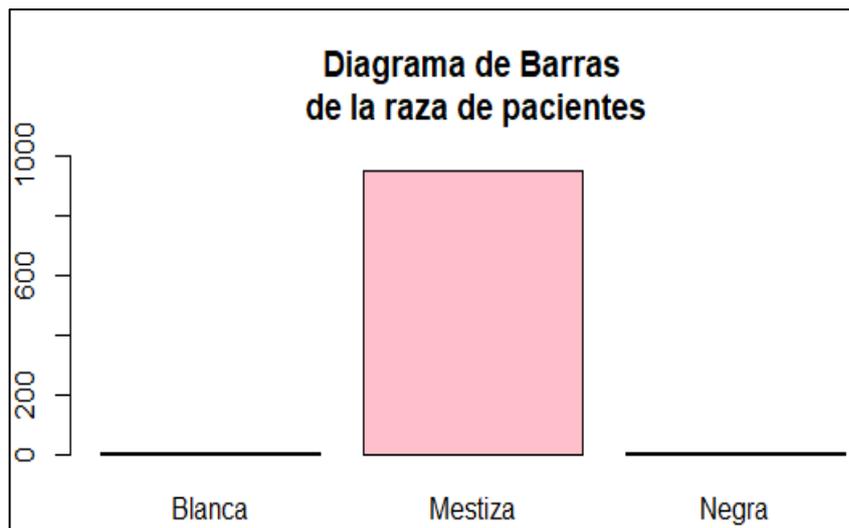


Gráfico 2-3: Raza de pacientes diagnosticados con cáncer.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

La población significativamente representativa es de raza mestiza siendo aproximadamente el 99.37% de la población total (950 de 956 pacientes), apenas 4 pacientes de raza blanca (aproximadamente 0.42% de la población) y 2 pacientes de raza negra (aproximadamente 0.21% de la población), fueron parte del estudio.

Tabla 3-3: Estadísticas de la raza de pacientes.

Medida	Categorías		
	blanca	mestiza	negra
n	4	950	2
proporción	0,42%	99,37%	0,21%
moda			mestiza

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.2.3. Estatura

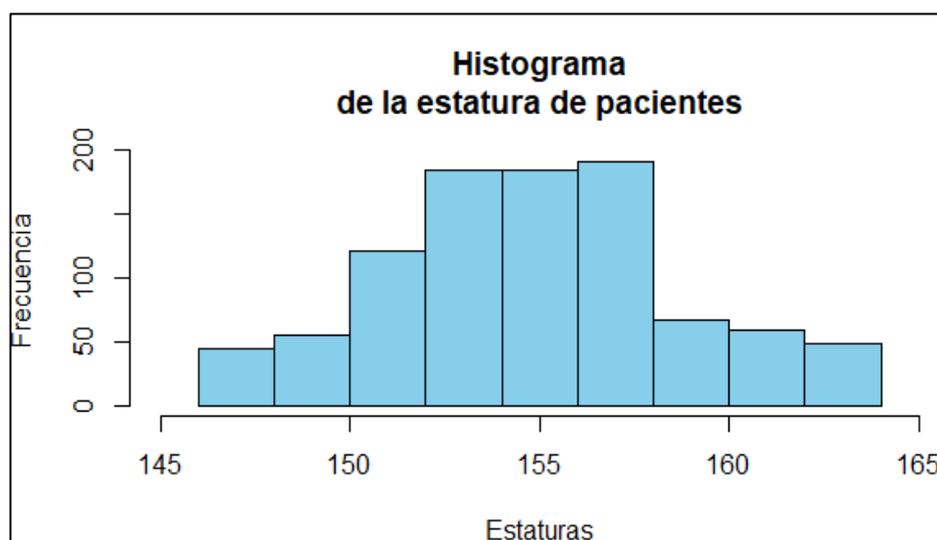


Gráfico 3-3: Distribución de estatura de pacientes diagnosticados con cáncer.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Dentro de la población se puede observar que la estatura de las pacientes diagnosticadas con cáncer de mama oscila aproximadamente en un rango de 145 cm a 165 cm; con una estatura promedio de 155.33 cm, en este caso se tiene que los datos no tienen mayor dispersión con respecto a la media, además se puede decir que la mayoría de pacientes representando aproximadamente el 58.68% del total de la población (561 de 956 pacientes) miden aproximadamente entre 153 a 158 cm, el porcentaje menos representativo corresponde al 1.88% de la población total, que son aquellas pacientes cuya estatura oscila aproximadamente entre 164 a 166 cm.

Tabla 4-3: Estadísticas de la estatura de pacientes diagnosticados con cáncer de mama.

Limites	f	rf	rf(%)	cf	cf(%)
[145.53,147.358)	19	0.02	1.99	19	1.99
[147.358,149.186)	68	0.07	7.11	87	9.10
[149.186,151.015)	48	0.05	5.02	135	14.12
[151.015,152.843)	86	0.09	9.00	221	23.12
[152.843,154.671)	185	0.19	19.35	406	42.47
[154.671,156.499)	185	0.19	19.35	591	61.82
[156.499,158.327)	191	0.20	19.98	782	81.80
[158.327,160.155)	67	0.07	7.01	849	88.81
[160.155,161.984)	27	0.03	2.82	876	91.63
[161.984,163.812)	62	0.06	6.49	938	98.12
[163.812,165.64)	18	0.02	1.88	956	100.00
media					155,33
mediana					155
desviación					3,89

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.2.4. *Peso*

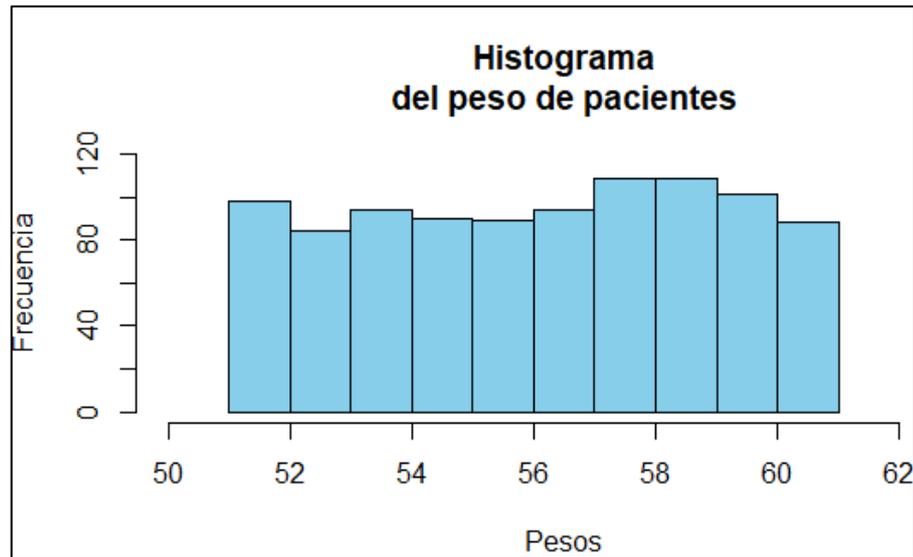


Gráfico 4-3: Distribución del peso de pacientes con cáncer.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Con un peso promedio de 56.16 Kg se puede observar que hay poca dispersión respecto a la media, en cuanto a los datos correspondientes al peso en Kg de las pacientes diagnosticadas con cáncer de mama, con esto se puede observar que, el 11.92% del total de la población pesan aproximadamente entre 56.5 kg a 57.6 kg, seguido del 11.19% con pesos de entre 58.6 kg a 59.6 kg aproximadamente, teniendo apenas un 4.50% de la población con pesos de entre 50.6 kg a 51.6

kg aproximadamente y un 3.56% de la población con peso que van desde los 60.6 kg a 61.6 kg aproximadamente.

Tabla 5-3: Estadísticas del peso de pacientes diagnosticados con cáncer de mama.

Limites	f	rf	rf(%)	cf	cf(%)
[50.589,51.591)	43	0.04	4.50	43	4.50
[51.591,52.593)	92	0.10	9.62	135	14.12
[52.593,53.595)	93	0.10	9.73	228	23.85
[53.595,54.597)	82	0.09	8.58	310	32.43
[54.597,55.599)	102	0.11	10.67	412	43.10
[55.599,56.6)	89	0.09	9.31	501	52.41
[56.6,57.602)	114	0.12	11.92	615	64.33
[57.602,58.604)	101	0.11	10.56	716	74.90
[58.604,59.606)	107	0.11	11.19	823	86.09
[59.606,60.608)	99	0.10	10.36	922	96.44
[60.608,61.61)	34	0.04	3.56	956	100.00
media					56,16
mediana					56,4
desviación					2,86

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.2.5. Cirugía

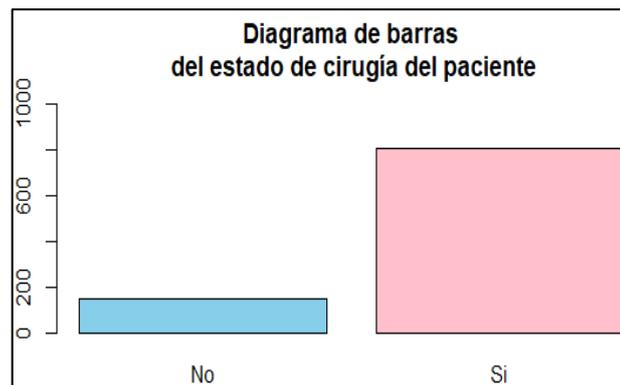


Gráfico 5-3: Condición de cirugía del paciente.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Se observa que un 84.62% de la población fue sometida algún tipo de cirugía (corresponde un total de 809 pacientes), mientras que el 15.38% de pacientes no fueron sometidos a ningún tipo de cirugía, es decir un total de 147 pacientes diagnosticados con cáncer no fueron sometidos a ningún tipo de cirugía, durante el periodo de enfermedad.

Tabla 6-3: Estadísticas de la condición de cirugía del paciente.

Medida	Categorías	
	SI	NO
n	809	147
proporción	84,62%	15,38%
moda		SI

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.2.6. Tiempo de evolución

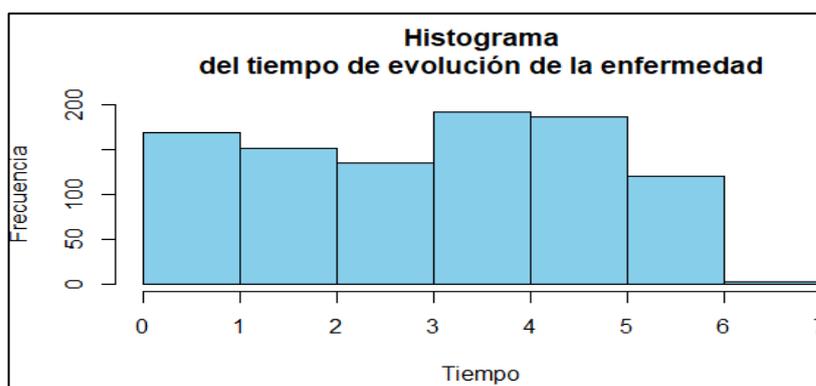


Gráfico 6-3: Distribución del tiempo de evolución de la enfermedad.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

En cuanto al tiempo de evolución la distribución es casi similar en todas las clases, teniendo así que la proporción de pacientes en cada una es aproximadamente semejante, a excepción de aquellos pacientes que han tenido una evolución mayor a 5 años que representan un porcentaje más bajo (6.49%).

Tabla 7-3: Estadísticas del tiempo de evolución de la enfermedad.

Limites	f	rf	rf(%)	cf	cf(%)
[0,0.8946)	135	0.14	14.12	135	14.12
[0.8946,1.789)	156	0.16	16.32	291	30.44
[1.789,2.684)	103	0.11	10.77	394	41.21
[2.684,3.578)	162	0.17	16.95	556	58.16
[3.578,4.473)	164	0.17	17.15	720	75.31
[4.473,5.367)	174	0.18	18.20	894	93.51
[5.367,6.262)	62	0.06	6.49	956	100.00
media					3
mediana					3,2
desviación					1,68

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.2.7. Tamaño tumoral

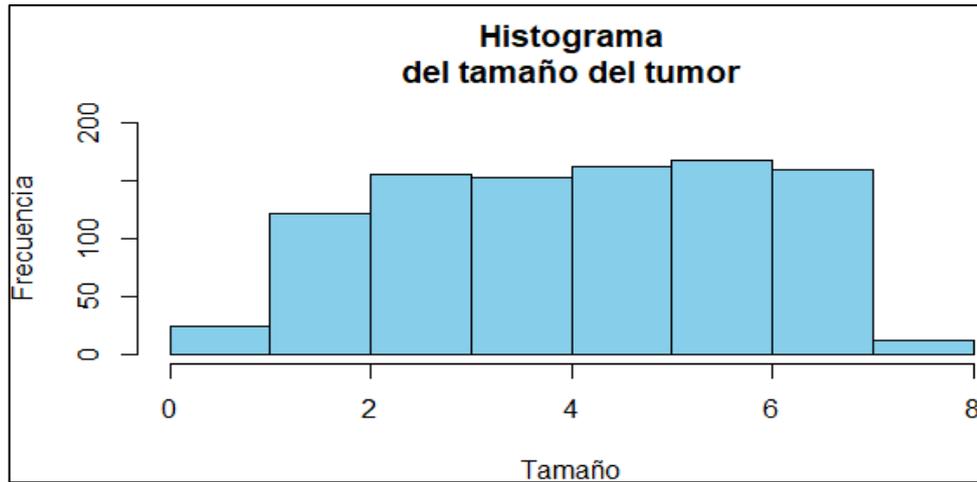


Gráfico 7-3: Distribución del tamaño del tumor.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Con un promedio de 4.15 cm, el tamaño tumoral más significativo dentro de la población oscila entre 5.8 cm a 6.8 cm aproximadamente, siendo el 20.19% de la población en general con un tumor con tamaño dentro del intervalo definido, además se puede observar que el 6.28% de la población tiene un tumor con tamaño que oscilan entre 0.5 cm a 1.5 cm y un 3.35% de la población tienen un tumor cuyo tamaño oscila entre 6.9 cm a 8.8 cm, siendo estos los porcentajes menos representativos del total de la población.

Tabla 8-3: Estadísticas del tamaño de tumor.

Limites	f	rf	rf(%)	cf	cf(%)
[0.495,1.55)	60	0.06	6.28	60	6.28
[1.55,2.604)	182	0.19	19.04	242	25.31
[2.604,3.659)	151	0.16	15.79	393	41.11
[3.659,4.714)	171	0.18	17.89	564	59.00
[4.714,5.769)	167	0.17	17.47	731	76.46
[5.769,6.823)	193	0.20	20.19	924	96.65
[6.823,7.878)	32	0.03	3.35	956	100.00
media					4,15
mediana					4,2
desviación					1,77

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.2.8. Ubicación del tumor

Gráficamente se puede observar que hay un claro grupo dominante (ubicación de tumor determinado en un paciente) con una diferencia significativa sobre el resto de los grupos (demás ubicaciones que alojan al tumor en un paciente). Resultado que se corrobora con los respectivos indicadores estadísticos encontrado en base a esta variable.

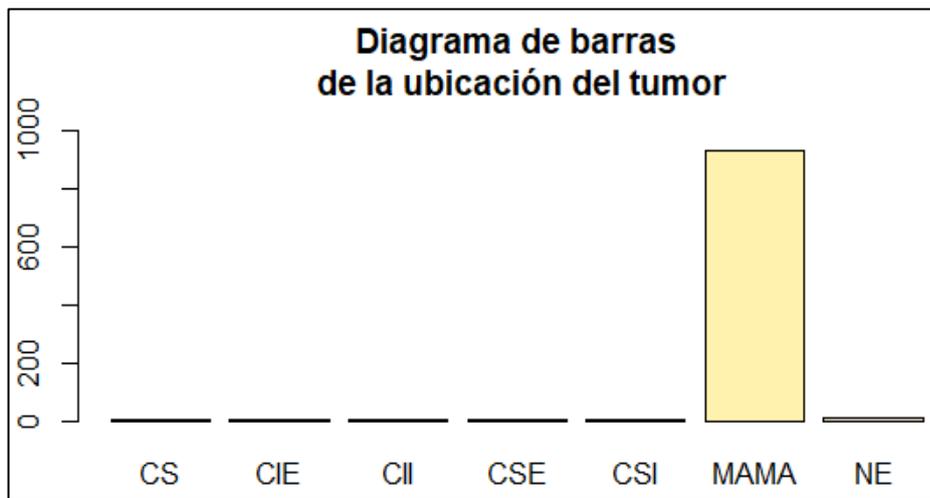


Gráfico 8-3: Ubicación del tumor en la paciente.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Se observa que en el 97.38% de los casos el tumor se aloja en toda la mama de la paciente, en el 0.31% de los casos de cáncer de mama el tumor se aloja en el cuadrante inferior de la mama, un 0.83% de los casos en el cuadrante superior de la mama, en un 1.36% de los casos no se ha podido especificar la zona en la que se ubica el tumor y apenas en un 0.10% de los casos el tumor se ubica en la denominada cola de Spence.

Tabla 9-3: Estadísticas de la ubicación del tumor.

Medida	Categorías						
	Mama	Parte no especificada (NE)	Cola de Spence (CS)	Cuadrante inferior interno (CII)	Cuadrante inferior externo (CIE)	Cuadrante superior interno (CSI)	Cuadrante superior externo (CSE)
n	931	13	1	2	1	3	5
proporción	97,38%	1,36%	0,10%	0,21%	0,10%	0,31%	0,52%
moda							Mama

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.2.9. Tipo de tumor

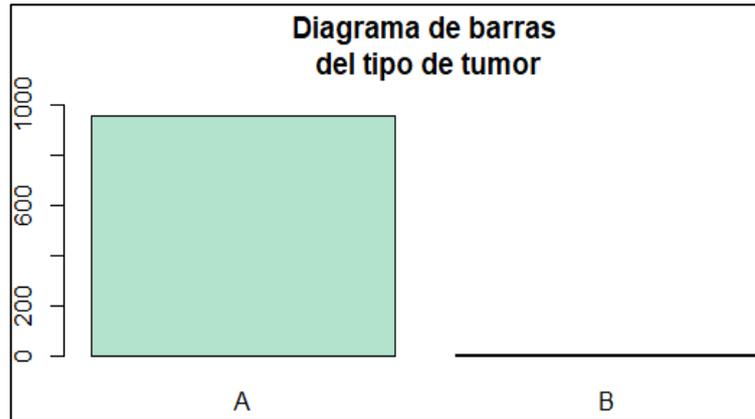


Gráfico 9-3: Tipo de tumor del paciente.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Se observa que el 99.90% de la población presentaron un tumor de tipo A (Carcinoma ductal in situ de la mama) y en el 0.10% de los casos restante no se pudo determinar con certeza el tipo de tumor por lo que presentan un tumor tipo B (Tumor de comportamiento incierto de la mama). Una única persona dentro de la población estudiada presentó un tumor de comportamiento incierto de mama, lo que induce que en los 955 casos restantes si se pudo verificar que tipo de cáncer padecía la paciente. Se puede decir además que en este caso es una variable totalmente desequilibrada en sus niveles.

Tabla 10-3: Estadísticas de la condición de cirugía del paciente.

Medida	Categorías	
	Carcinoma ductal in situ de la mama (A)	Tumor de comportamiento incierto de la mama (B)
n	955	1
proporción	99,90%	0,10%
moda		A

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.2.10. Estadio

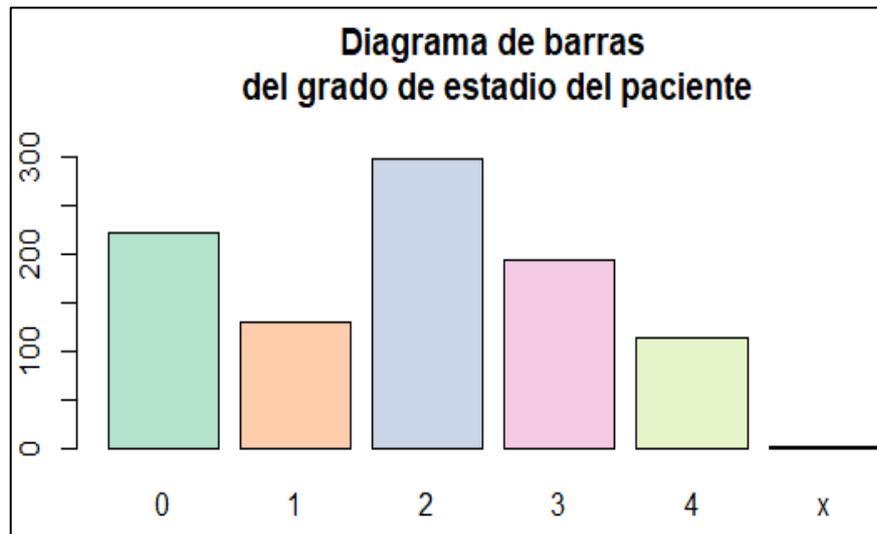


Gráfico 10-3: Estadio del paciente diagnosticado con cáncer.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Se observa que la mayoría de pacientes (298 de 957) han desarrollado un estadio de grado 2; 222 pacientes han desarrollado un estadio de grado 0; 193 pacientes presentaron un estadio grado 3; 133 pacientes presentaron un estadio grado 4; 129 pacientes presentaron estadio grado 1 y tan solo 1 paciente desarrollo estadio grado x, lo que permite deducir que hubo un caso en el cual al paciente no se pudo establecer el tipo de estadio que desarrollo. El porcentaje más representativo corresponde a los pacientes con estadio clínico grado 2 con un total del 31.17% de la población diagnosticada con cáncer de mama.

Tabla 11-3: Estadísticas del estadio del paciente.

Medida	Categorías					
	0	1	2	3	4	x
n	222	129	298	193	133	1
proporción	23,22%	13,49%	31,17%	20,19%	13,91%	0,10%
moda						2

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.2.11. Estado T

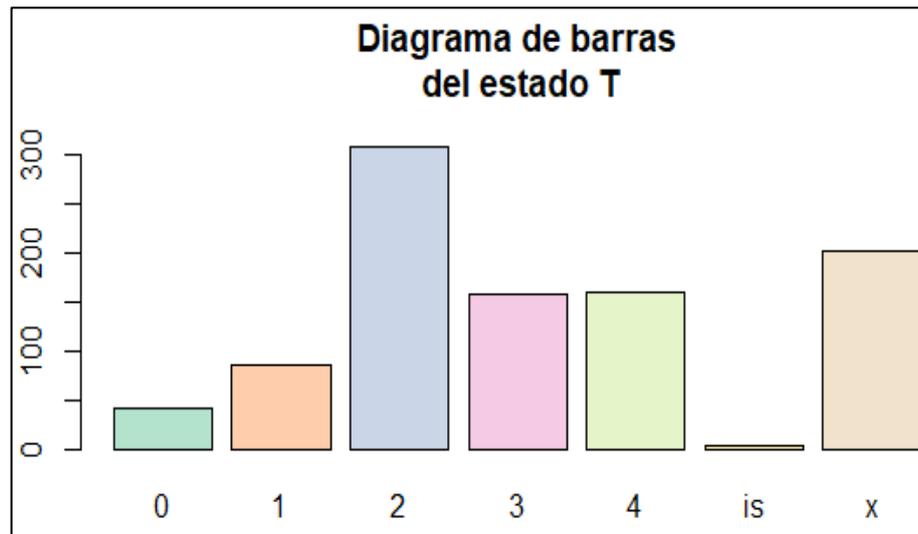


Gráfico 11-3: Tamaño y extensión del tumor principal.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Se tiene que un total de 308 pacientes tiene un estado T de grado 2, se observa que favorablemente 201 pacientes tienen un estado T de grado x; 160 pacientes presentan un estado T con grado 4; 158 pacientes presentan un estado T con grado 3, 85 pacientes tienen un estado T con grado 1 y 41 pacientes tiene un estado T con grado 0; finalmente se tiene que únicamente 3 pacientes tienen un estado T de grado is. Se tiene que el porcentaje más representativo de la población en lo que conlleva al estado T, es el 32.22% correspondientes a aquellas personas cuyo grado de estado T es igual a 2.

Tabla 12-3: Estadísticas del Estado T.

Medida	Categorías						
	0	1	2	3	4	is	x
n	41	85	308	158	160	3	201
proporción	4,29%	8,89%	32,22%	16,53%	16,74%	0,31%	21,03%
moda							2

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.2.12. Estado N

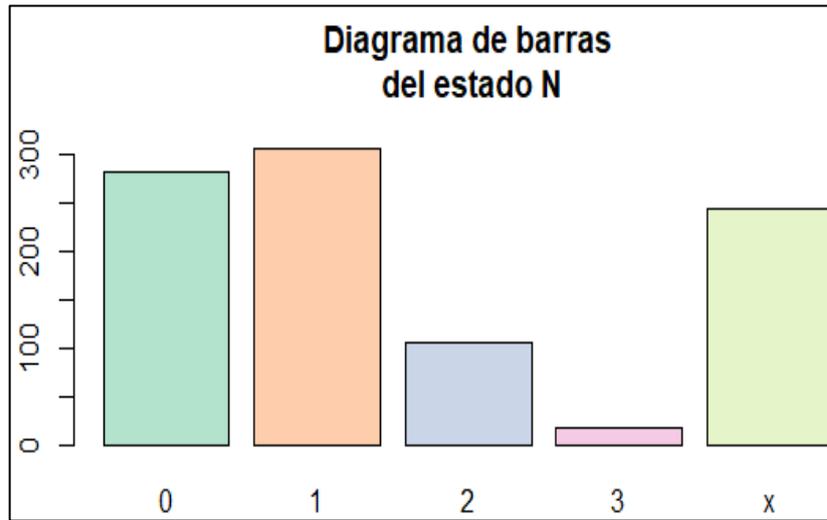


Gráfico 12-3: Invasión de ganglios linfáticos.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

En cuanto a proporción de pacientes respecto a la población general, se tiene que los estados que corresponde a la invasión de ganglios linfáticos, estarían de manera significativa de la siguiente forma; 1, 0, x, 2, 3; con proporciones de pacientes del 32.01%, 29.50%, 25.63%, 11.09% y 1.78% respectivamente.

Tabla 13-3: Estadísticas del Estado N.

Medida	Categorías				
	0	1	2	3	x
n	282	306	106	17	245
proporción	29,50%	32,01%	11,09%	1,78%	25,63%
moda					1

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.2.13. Estado M

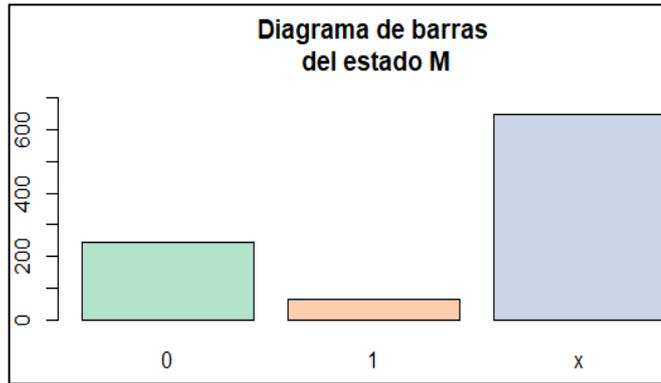


Gráfico 13-3: Diseminación del cáncer principal.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Se observa que los únicos estados presentados para el Estado M, son los de grado 0, 1, x; teniendo que un 67.78% de la población tuvo una diseminación de tumor principal de grado x, hay un 25.73% de pacientes que tuvieron grado 0 de diseminación y favorablemente un 6.49% del total de la población tuvo una diseminación de grado 1.

Tabla 14-3: Estadísticas del Estado M.

Medida	Categorías		
	0	1	x
n	246	62	648
proporción	25,73%	6,49%	67,78%
moda			x

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.2.14. Estado

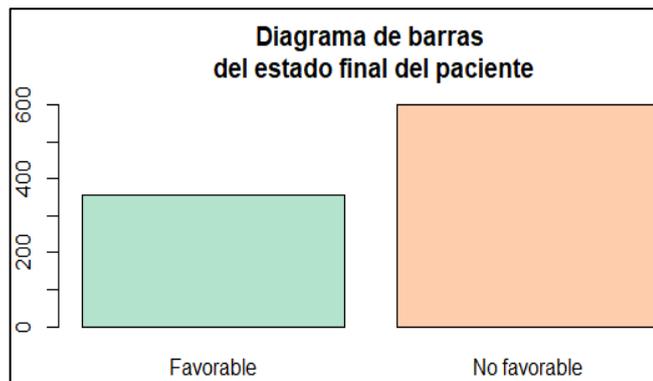


Gráfico 14-3: Estado evolutivo del paciente.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Se observa que un 62.87% de la población tiene un estado no favorable ante la enfermedad, por otro lado, tan solo el 37.13% de pacientes que han sido diagnosticados con cáncer de mama, han tenido un desarrollo evolutivo favorable ante la enfermedad.

Tabla 15-3: Estadísticas del estado evolutivo del paciente.

Medida	Categorías	
	0	1
n	355	601
proporción	37,13%	62,87%
moda		1

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.3. Análisis de Dependencia

3.3.1. Análisis de dependencia de variables cuantitativas

Para saber cuál de estas variables se relación entre si es necesario conocer si alguna depende de otra, en primer caso se tiene la relación para las variables de carácter cuantitativo, las mismas que se estudian a través de un análisis del coeficiente de correlación, apoyados en un gráfico de correlación que nos permitirá corroborar los resultados obtenidos con el test teórico así se tiene:

Tabla 16-3: Matriz de correlación entre variables cuantitativas.

	Edad	Tiempo evolución	Peso	Estatura	Tamaño
Edad	1	0,34	0,45	0,61	0,19
Tiempo evolución	0,34	1	0,72	0,81	0,88
Peso	0,45	0,72	1	0,82	0,73
Estatura	0,61	0,81	0,82	1	0,75
Tamaño	0,19	0,88	0,73	0,75	1

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Se pudo observar que la variable edad es independiente a las demás variables (Tiempo de evolución, Peso, Estatura, Tamaño tumoral). Así mismo se puede observar que las variables más dependientes son Tamaño tumoral con Tiempo de evolución, seguida de las relaciones entre: Estatura-Peso y Estatura-Tiempo de evolución, con coeficientes de correlación de 0.82 y 0.81 correspondientemente. Mismo que se corrobora con el gráfico en donde las relaciones más marcadas se evidencia un color de esfera más oscuro que los demás.

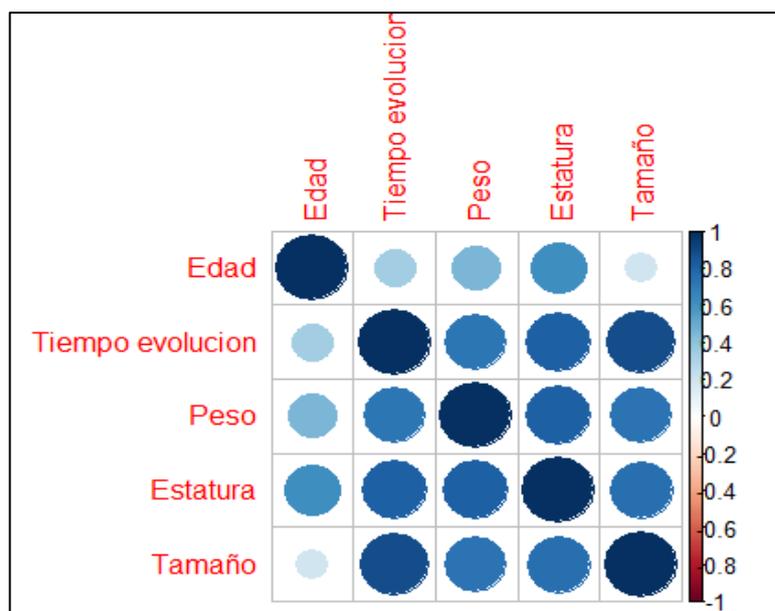


Gráfico 15-3: Gráfico de correlación.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.3.2. Análisis de dependencia de variables cualitativas

Sabiendo que hay variables cualitativas con sus respectivos niveles, también es necesario conocer si alguna de estas se relaciona entre sí, dicho estudio se realizó a través de la prueba de independencia ji-cuadrada, misma que se basa en un contraste de hipótesis en el que se analiza cierta hipótesis, con un nivel de significancia dado, partiendo de:

$$H_0: \text{Las variables son independientes entre sí.} \quad (1-3)$$

$$H_1: \text{Las variables no son independientes entre sí.}$$

Teniendo en cuenta lo siguiente y partiendo con base en un nivel de significancia $\alpha = 0.05$, se tiene lo siguiente:

Tabla 17-3: Coeficientes de relación de variables cualitativas.

	p value	Decisión	Conclusión
Estado civil / Raza	0,326	No Rechazo Ho	No hay dependencia
Estado civil / T	0,001	Rechazo Ho	Hay dependencia
Estado civil / N	0,001	Rechazo Ho	Hay dependencia
Estado civil / M	0,611	No Rechazo Ho	No hay dependencia
Estado civil / Estadio	0,038	Rechazo Ho	Hay dependencia
Estado civil / Tipo	0,856	No Rechazo Ho	No hay dependencia
Estado civil / Parte	0,577	No Rechazo Ho	No hay dependencia
Estado civil / Cirugía	0,036	Rechazo Ho	Hay dependencia
Raza / T	0,854	No Rechazo Ho	No hay dependencia
Raza / N	0,524	No Rechazo Ho	No hay dependencia
Raza / M	0,046	Rechazo Ho	Hay dependencia

Raza / Estadio	0,633	No Rechazo Ho	No hay dependencia
Raza / Tipo	0,997	No Rechazo Ho	No hay dependencia
Raza / Parte	0,999	No Rechazo Ho	No hay dependencia
Raza / Cirugía	0,131	No Rechazo Ho	No hay dependencia
T / N	0,001	Rechazo Ho	Hay dependencia
T / M	0,001	Rechazo Ho	Hay dependencia
T / Estadio	0,001	Rechazo Ho	Hay dependencia
T / Tipo	0,537	No Rechazo Ho	No hay dependencia
T / Parte	0,482	No Rechazo Ho	No hay dependencia
T / Cirugía	0,132	No Rechazo Ho	No hay dependencia
N / M	0,001	Rechazo Ho	Hay dependencia
N / Estadio	0,001	Rechazo Ho	Hay dependencia
N / Tipo	0,574	No Rechazo Ho	No hay dependencia
N / Parte	0,444	No Rechazo Ho	No hay dependencia
N / Cirugía	0,006	Rechazo Ho	Hay dependencia
M / Estadio	0,001	Rechazo Ho	Hay dependencia
M / Tipo	0,788	No Rechazo Ho	No hay dependencia
M / Parte	0,982	No Rechazo Ho	No hay dependencia
M / Cirugía	0,001	Rechazo Ho	Hay dependencia
Estadio / Tipo	0,556	No Rechazo Ho	No hay dependencia
Estadio / Parte	0,786	No Rechazo Ho	No hay dependencia
Estadio / Cirugía	0,017	Rechazo Ho	Hay dependencia
Tipo / Parte	0,999	No Rechazo Ho	No hay dependencia
Tipo / Cirugía	0,337	No Rechazo Ho	No hay dependencia
Parte / Cirugía	0,484	No Rechazo Ho	No hay dependencia

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Con un 95% de confianza, se puede concluir que en los casos correspondientes no se puede rechazar la hipótesis nula por lo que se dice que las variables no son dependientes; es así que algunas de las variables que no son dependientes son: Estado civil-Raza, Estadio-Tipo, Parte-Cirugía; entre otros. Por otro lado, se tiene que las variables que si son dependientes son, por ejemplo: Estadio-Cirugía, Estado civil-Estadio, Raza- Estudio M, Estudio T-Estudio N; entre otros.

3.4. Análisis Estadístico Logístico

Luego del correcto y minucioso análisis estadístico descriptivo exploratorio, y un análisis de dependencia entre variables, se procede a continuar con el planteamiento del respectivo modelo logístico que nos permita seguir con el estudio del tema en cuestión. Así se tiene:

3.4.1. Modelo Logístico 1

Se parte de primera mano con un primer modelo de regresión logística en el que se toma en cuenta todas las variables inmiscuidas en el estudio, con el fin de definir la significancia de todas o de cada una de dichas variables, o, por lo contrario, existe alguna o algunas que no aporta de información al modelo y que su estudio no es relevante por su baja significancia. Con lo que teniendo en cuenta a la variable cualitativa dicotómica “Estado” misma que tiene dos distintas categóricas “0=Evolución favorable frente a la enfermedad” y “1= Evolución no favorable frente a la enfermedad” como variable respuesta, frente a todas las demás como variables predictoras se tiene:

$$\begin{aligned} \widehat{Estado} = & \beta_0 + \beta_1 Edad + \beta_2 Raza + \beta_3 T + \beta_4 N + \beta_5 M + \beta_6 Estadio \\ & + \beta_7 Tipo + \beta_8 Parte + \beta_9 Cirugía + \beta_{10} Tiempo \\ & + \beta_{11} Tamaño + \beta_{12} Peso + \beta_{13} Estatura \end{aligned} \quad (2-3)$$

Con un modelo base planteado se puede desglosar sabiendo que el tipo de variables en estudio, no son únicamente de carácter cuantitativo, sino que también hay presentes variables de carácter cualitativo que tienen cada una categorías (niveles) distintos, con lo cual, pensando el hecho de que si se establece la significancia de una o algunas variables de este tipo, sería de suma importancia la especificidad para saber cuál o cuáles de dichos niveles de la variables cualitativa en cuestión tiene mayor significancia sobre el modelo en general. Así se plantearía el modelo desglosado:

$$\begin{aligned} \widehat{Estado} = & \beta_0 + \beta_1 Edad + \beta_2 Raza_{Mestiza} + \beta_3 Raza_{Blanca} \\ & + \beta_4 Raza_{Negra} + \beta_5 T_0 + \beta_6 T_1 + \beta_7 T_2 + \beta_8 T_3 + \beta_9 T_4 \\ & + \beta_{10} T_{is} + \beta_{11} T_x + \beta_{11} N_0 + \beta_{12} N_1 + \beta_{13} N_2 + \beta_{14} N_3 \\ & + \beta_{15} N_x + \beta_{16} M_0 + \beta_{17} M_1 + \beta_{18} M_x + \beta_{19} Estadio_0 \\ & + \beta_{20} Estadio_1 + \beta_{21} Estadio_2 + \beta_{22} Estadio_3 \\ & + \beta_{23} Estadio_4 + \beta_{24} Estadio_x + \beta_{25} Tipo_A + \beta_{25} Tipo_B \\ & + \beta_{26} Parte_{Mama} + \beta_{27} Parte_{NE} + \beta_{28} Parte_{CS} \\ & + \beta_{29} Parte_{CIE} + \beta_{30} Parte_{CII} + \beta_{31} Parte_{CSE} \\ & + \beta_{32} Parte_{CSI} + \beta_{33} Cirugía_{SI} + \beta_{34} Cirugía_{NO} \\ & + \beta_{35} Tiempo + \beta_{36} Tamaño + \beta_{37} Peso \\ & + \beta_{38} Estatura \end{aligned} \quad (3-3)$$

Una vez establecido el modelo logístico, y con la ayuda del software estadístico R, se procede a realizar la correspondiente regresión en este caso una regresión logística, poniendo a prueba un primer modelo con el que se tratara de explicar el comportamiento de las variables con respecto a los datos de la población en estudio en el que se inmiscuye a todas y cada una de las variables

especificadas con anterioridad; así como cada una de las categorías de las variables de carácter cualitativo, con lo después de la respectiva aplicación de obtuvo:

3.4.1.1. Coeficientes de regresión del primer modelo

Tabla 18-3: Coeficientes de regresión Modelo 1.

Variable	Coefficiente
(Intercept)	19,169
Edad	0,008
Raza Mestiza	-0,001
Raza Negra	-3,552
T1	-1,134
T2	-1,349
T3	-0,993
T4	-0,356
Tis	0,135
Tx	-0,266
N1	0,480
N2	0,585
N3	1,462
Nx	-0,142
M1	2,297
Mx	0,472
Estadio1	0,547
Estadio2	0,886
Estadio3	0,672
Estadio4	0,206
Estadios	-15,807
TipoTumor	12,897
Parte Cuadrante inferior externo de la mama	-30,786
Parte Cuadrante inferior interno de la mama	-16,228
Parte Cuadrante superior externo de la mama	-14,573
Parte Cuadrante superior interno de la mama	-16,071
Parte Mama	-14,780
Parte (Parte no especificada)	-14,844
Cirugía Si	-0,252
Tiempo	-0,515
Tamaño	-0,289
Peso	0,119
Estatura	-0,053

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Se observa cada uno de los coeficientes de regresión para cada uno de las variables estudiadas, tanto como para las de carácter cuantitativo, como para todas y cada una de las categorías de las variables de carácter cualitativo, en aquellos casos donde el coeficiente ha sido muy

insignificativo (es decir el valor del coeficiente de aproxima mucho o es igual a 0), el programa por defecto los excluye auto asumiendo poca significancia sobre el modelo planteado, mostrando así únicamente aquellos valores considerados en primera instancia como significativos, por lo tanto en base a los modelos planteados en específico el primer modelo desglosado (Ecuación 3-3), se puede asumir lo siguiente:

$$\begin{aligned}
 \widehat{Estado} = & 19.169 + 0.008Edad - 0.001Raza_{Mestiza} + 0Raza_{Blanca} \\
 & - 3.552Raza_{Negra} + 0T_0 - 1.134T_1 - 1.349T_2 \\
 & - 0.993T_3 - 0.356T_4 + 0.135T_{is} - 0.266T_x + 0N_0 \\
 & + 0.480N_1 + 0.585N_2 + 1.462N_3 - 0.142N_x + 0M_0 \\
 & + 2.297M_1 + 0.472M_x + 0Estadio_0 + 0.547Estadio_1 \\
 & + 0.886Estadio_2 + 0.672Estadio_3 + 0.206Estadio_4 \quad (4-3) \\
 & - 15.807Estadio_x + 0Tipo_A + 12.897Tipo_B \\
 & - 14.780Parte_{Mama} - 14.844Parte_{NE} + 0Parte_{CS} \\
 & - 30.786Parte_{CIE} - 16.228Parte_{CII} \\
 & - 14.573Parte_{CSE} - 16.071Parte_{CSI} \\
 & - 0.252Cirugía_{SI} + 0Cirugía_{NO} - 0.515Tiempo \\
 & - 0.289Tamaño + 0.119Peso - 0.053Estatura
 \end{aligned}$$

En general el primer modelo de regresión planteado es igual a lo presentado en la (Ecuación 2-3) respectivamente, pero como se aclaró, hay variables cuyos coeficientes de regresión (β_i) son poco significativos por lo que la omisión de los mismos ayudara a resumir el modelo a uno más claro y fácil de interpretar:

$$\begin{aligned}
 \widehat{Estado} = & 19.169 + 0.008Edad - 0.001Raza_{Mestiza} \\
 & - 3.552Raza_{Negra} - 1.134T_1 - 1.349T_2 - 0.993T_3 \\
 & - 0.356T_4 + 0.135T_{is} - 0.266T_x + 0.480N_1 + 0.585N_2 \\
 & + 1.462N_3 - 0.142N_x + 2.297M_1 + 0.472M_x \\
 & + 0.547Estadio_1 + 0.886Estadio_2 + 0.672Estadio_3 \\
 & + 0.206Estadio_4 - 15.807Estadio_x + 12.897Tipo_B \\
 & - 14.780Parte_{Mama} - 14.844Parte_{NE} \\
 & - 30.786Parte_{CIE} - 16.228Parte_{CII} \quad (5-3) \\
 & - 14.573Parte_{CSE} - 16.071Parte_{CSI} \\
 & - 0.252Cirugía_{SI} - 0.515Tiempo - 0.289Tamaño \\
 & + 0.119Peso - 0.053Estatura
 \end{aligned}$$

3.4.1.2. Resumen de la regresión del primer modelo

Preestablecido un primer modelo de regresión para dar validez e identificar aquellos factores que resulten importantes e influyentes en la mortalidad de los pacientes diagnosticados con cáncer de mama, se observa aquellas variables que son significativas estableciendo para ello un nivel de significancia con lo que ayude a rechazar o no rechazar la significancia de una variable dentro del modelo, en base a una prueba de comparación entre el valor de la probabilidad encontrada con respecto al nivel de significancia establecido, en este caso se trabaja con un nivel de confianza (NC) adecuado y que por defecto se trabaja en la mayoría de estudios estadístico del 95%, por lo tanto por lo contrario se asumirá un nivel de significancia del 5% (denotando al nivel de significancia con α , con lo que $\alpha = 0.05$.)

$$H_0: \text{La variable es significativa} \quad (6-3)$$

$$H_1: \text{La variable no es significativa}$$

Ayudados con el software y teniendo en cuenta que una variable será significativa siempre que $P_r(> |Z|) < \alpha$, por lo que presentada en la tabla 17-3 de resumen se denotará con (*) cuando la variable cumpla con la condición planteada, lo que conllevará a la significancia de la variable.

Tabla 19-3: Resumen de regresión del modelo 1.

Variable	Valor estimado	Std. Error	Valor Z	Pr(> z)	Significancia
(Intercept)	19,169	882,753	0,022	0,983	
Edad	0,008	0,006	1,367	0,172	
Raza Mestiza	-0,001	1,352	0,001	0,999	
Raza Negra	-3,552	2,248	-1,58	0,114	
T1	-1,134	0,055	-2,066	0,039	*
T2	-1,349	0,053	-2,524	0,012	*
T3	-0,993	0,059	-1,664	0,096	.
T4	-0,356	0,064	-0,557	0,578	
Tis	0,135	1,306	0,104	0,918	
Tx	-0,266	0,048	-0,558	0,577	
N1	0,48	0,021	2,17	0,029	*
N2	0,585	0,033	1,761	0,078	.
N3	1,462	1,138	1,286	0,199	
Nx	-0,142	0,031	-0,452	0,651	
M1	2,297	0,063	3,665	0,0002	***
Mx	0,472	0,019	2,399	0,016	*
Estadio1	0,547	0,039	1,383	0,167	
Estadio2	0,886	0,041	2,144	0,032	*
Estadio3	0,672	0,049	1,346	0,178	
Estadio4	0,206	0,059	0,349	0,727	
Estadiox	-15,807	882,744	-0,018	0,986	

Tipo Tumor de comportamiento incierto	12,897	882,743	0,015	0,988	
Parte Cuadrante inferior externo de mama	-30,786	1248,388	-0,025	0,98	
Parte Cuadrante inferior interno de mama	-16,228	882,745	-0,018	0,985	
Parte Cuadrante superior externo de mama	-14,573	882,744	-0,017	0,987	
Parte Cuadrante superior interno de mama	-16,071	882,745	-0,018	0,985	
Parte Mama	-14,78	882,743	-0,017	0,987	
Parte (Parte no especificada)	-14,844	882,744	-0,017	0,987	
Cirugía Si	-0,252	0,025	-1,007	0,314	
Tiempo	-0,515	0,005	-9,466	2x10-16	***
Tamaño	-0,289	0,006	-4,976	2x10-07	***
Peso	0,119	0,003	3,415	0,001	***
Estatura	-0,053	0,002	-2,625	0,009	**

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.4.1.3. Variables significativas del primer modelo

Así se tiene que al nivel plantado inicialmente ($NC = 95\%$, $\alpha = 0.05$) las variables de manera general consideradas como significativas son T (estudio del tamaño y ubicación del tumor principal), N (Estudio ganglionar), M (Estudio Metastásico), Estadio, Tiempo de evolución, Tamaño tumoral, Peso, Altura. Cabe recalcar que la significancia por categoría de las variables cuantitativas depende y varían, pero al ser el modelo inicial se tomara la variable en general asumiendo que la variable es significativa en su totalidad, para en base a ello partir y generar un próximo modelo mucho más específico que nos permita delimitar el comportamiento de forma más precisa acerca de todas y cada una de estas. Por lo que es permitente indicar que algunas de están su significancia están denotadas por “.” para aquellas que son significativas con un $NC=90\%$, “***” para aquellas que se puede considerar un nivel más estricto de $NC=99\%$. Todo ello será estudiado y finalmente comparado con el fin de elegir un modelo que permita generar predicciones más precisas.

Tabla 20-3: Variables significativas del modelo 1.

Variable	Nivel Significancia	Conclusión
T1	*	Significativa con $\alpha = 0.05$
T2	*	Significativa con $\alpha = 0.05$
T3	.	Significativa con $\alpha = 0.10$
N1	*	Significativa con $\alpha = 0.05$
N2	.	Significativa con $\alpha = 0.10$
M1	*, **	Significativa con $\alpha = 0.05$ o significativa con $\alpha = 0.01$
Mx	*	Significativa con $\alpha = 0.05$
Estadio2	*	Significativa con $\alpha = 0.05$
Tiempo	*, **	Significativa con $\alpha = 0.05$ o significativa con $\alpha = 0.01$
Tamaño	*, **	Significativa con $\alpha = 0.05$ o significativa con $\alpha = 0.01$
Peso	*, **	Significativa con $\alpha = 0.05$ o significativa con $\alpha = 0.01$

3.4.1.4. Evaluación del primer modelo

Es importante verificar si el modelo nos permitirá generar predicciones precisas, y que se apege a las observaciones originales, para ello un paso muy importante es evaluar el modelo de regresión, por tal motivo, teniendo un primer modelo de regresión se procederá a evaluarlo en base a: su significancia en base a la prueba de Likelihood ratio, la precisión que este tiene de discriminar las observaciones y predicciones en base a una matriz de precisión, analizando cual es el error al precisar y predecir, y finalmente analizando el área bajo de la curva ROC.

3.4.1.5. Significancia del primer modelo

Se toma en cuenta principalmente un nivel de Nivel de confianza (NC) y su respectivo nivel de significancia (α), con el cual se plantea una hipótesis, la cual es sometida respectivamente para aprobar la significancia del modelo.

$$H_0: \text{El modelo es significativo} \quad (7-3)$$

$$H_1: \text{El modelo no es significativo}$$

Con un 95% de confianza existe suficiente evidencia significativa para no rechazar la hipótesis nula por lo que se puede concluir, que el primer modelo de regresión planteado es significativo, por lo que se asume un modelo apto para hacer predicciones.

Tabla 21-3: Significancia del modelo 1.

Diferencia de residuos	Grados de libertad	Valor p	alfa	Decisión
238,4616	32	2.02×10^{-33}	0,05	Significativo

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.4.1.6. Precisión del primer modelo

Se busco que el modelo además de ser significativo tenga una buena capacidad de discriminación entre los distintos casos que se puede presentar al construir una matriz de confusión en donde nos interesa que tantos verdaderos positivos nos determina, y cuantos; así como los demás casos que se puedan estudiar, falsos positivos, falsos negativos, verdaderos negativos; reiterando que se centrara en la discriminación de los verdaderos positivos y cuan grande es esa entidad en comparación a la población en general.

Tabla 22-3: Matriz de confusión del modelo 1.

Comparación	Predicciones	
	0	1
Observaciones	0	1
	0	204
1	112	489

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

En base a la matriz de decisión y al gráfico presentado a continuación, se puede observar que la proporción de verdaderos positivos discriminados correctamente por el modelo es mayor a los demás casos, eso intuitivamente. Así que se procedió a calcular la precisión del modelo para detectar los casos positivos.

$$\begin{aligned}
 \text{Precisión} &= \frac{d}{b + d} * 100 \\
 &= \frac{489}{489 + 151} * 100 \\
 &= \frac{489}{640} * 100 \\
 &= 76.40625 \\
 &\cong 76.41\%
 \end{aligned}
 \tag{8-3}$$

El primer modelo planteado puede discriminar correctamente un 76.41% de predicciones positivas correctas, por lo que se puede considerar al modelo como apto para predecir, ya que las predicciones se aproximarán de buena manera a las observaciones.

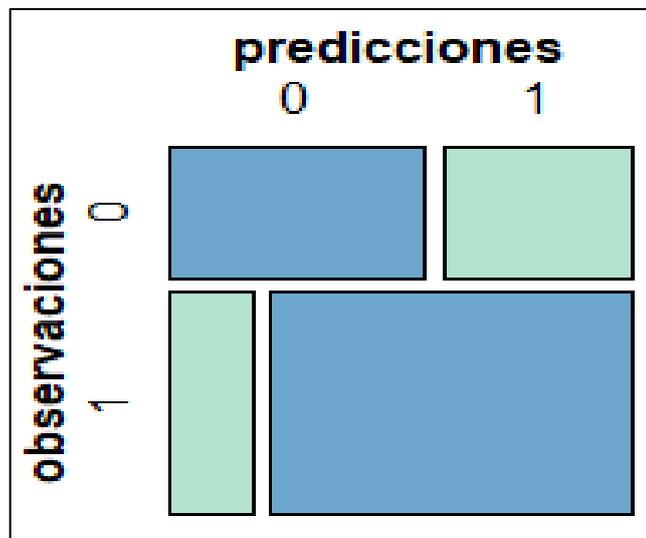


Gráfico 16-3: Matriz de confusión del modelo 1.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.4.1.7. Error de precisión del primer modelo

Es importante calcular por otro lado el error de precisión del modelo por lo que se basa en:

$$\begin{aligned} \mathbf{Error} &= \frac{b + c}{a + b + c + d} * 100 \\ &= \frac{151 + 112}{956} * 100 \\ &\cong 27.51\% \end{aligned} \tag{9-3}$$

Con un error de precisión del modelo del 27.51% se puede decir que el modelo es bueno para predecir, sin embargo, no es el mejor modelo para considerar.

3.4.1.8. Curva ROC del primer modelo

A pesar de lo ya comprobado anteriormente, se quiere refirmar esa capacidad del modelo, para lo cual se analizara el área bajo la curva ROC, en cuanto a los valores predeterminados dados ya por la prueba, con esto se estará seguros de que el modelo es buen discriminador lo que conlleva a que las predicciones son buenas, y se aproximan a las observaciones de buena manera, en este caso se sabe que el modelo de regresión tiene una capacidad de discriminación aceptable, con lo que se podrá calcular predicciones aceptables, con la opción de mejorar dichas predicciones.

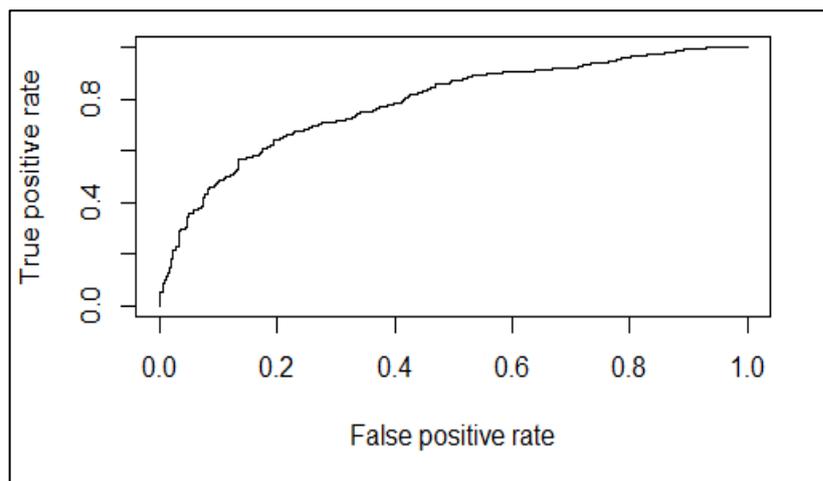


Gráfico 17-3: Curva ROC del modelo 1.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Con un área bajo la curva de 0.78 se admite que el modelo tiene una capacidad de discriminación aceptable, por ende, se corrobora lo ya dicho y se reafirma la capacidad buena del modelo para predecir.

3.4.1.9. Predicciones del primer modelo

Con el modelo ya planteado y validado su respectivo nivel de precisión, así como su significancia en cada una de las predicciones; se puede realizar las mismas, se observa que las dichas predicciones se aproximan de buena manera a las observaciones, no hay mayor variación en el total de observaciones con el total de predicciones que hizo el modelo. Sin embargo, por no ser del todo preciso el modelo habrá observaciones que tiendan a confundirse, para lo cual es muy importante estudiarlos correctamente.

Tabla 23-3: Variación observaciones vs predicciones 1.

Dato	Categorías	
	0	1
Observaciones	355	601
Predicciones	316	640
Variación		8,16%

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

De primera mano se tiene que el porcentaje que varía entre las observaciones y las predicciones es bueno únicamente el 8.16% de los resultados varia, lo que conllevaría a que el modelo está trabajando de buena manera teniendo así, pero el hecho de que haya un porcentaje de variación, es un llamado de atención para entrar en detalle del porque está variando.

Sabiendo que en síntesis el modelo era bueno y significativo, pero no era del todo perfecto, se va a tener errores en las predicciones, y por ende un porcentaje de variación en las mismas por lo que es importante analizar por secciones.

Tabla 24-3: Proporción de predichos modelos 1.

Dato	Categorías		
	0	NI	1
Predichos	252	114	590
Proporción	26,36%	11,92%	61,72%

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Se puede observar que dentro de las predicciones hay un 26.36% claro de que son individuos que tuvieron una evolución favorable ante la enfermedad, por el otro lado se tiene que un 61.72% son claros casos de pacientes que no tuvieron una evolución favorable ante la enfermedad, sin embargo, hay un 11.92% que no se puede identificar de manera correcta si son casos favorables o no favorables de evolución, dado que el modelo no es del todo preciso, a pesar de ser fuerte y significativo. Reafirmado en el gráfico de a continuación:

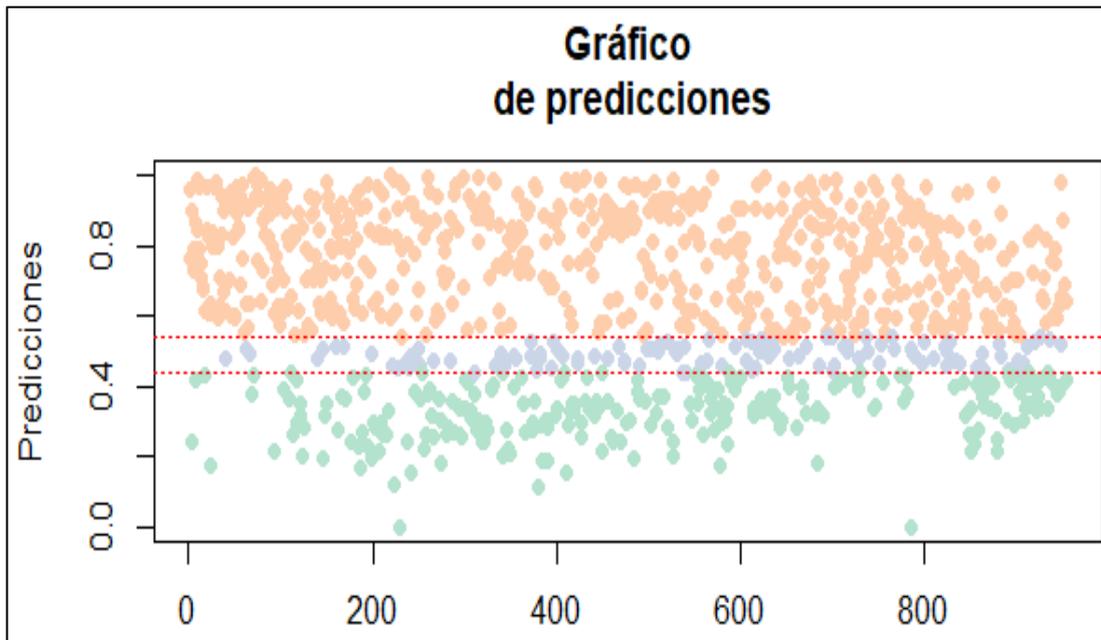


Gráfico 18-3: Proporción de predicciones modelo 1.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.4.1.10. Predicción ejemplo con el primer modelo

Ya teniendo un poco más claro cómo funciona el modelo se puede hallar probabilidades con el mismo, tratando de predecir la evolución que tendrá un sujeto que se puede presentar a distintas causas, así se plantea dos casos en el que se determinará la evolución del individuo frente a la enfermedad.

Se tiene una paciente diagnosticada con cáncer de mama cuya edad es de 48 años, al momento presenta un peso de 54.5 kg, una altura de 140 cm, es de raza blanca, con una estadificación general de N=3, M=1, T= 4, generando un estadio grado 2, reafirmada con el tipo de carcinoma in situ de mama, cuyo tiempo de evolución es de 2.5 años, desarrollando un tumor de tamaño 5.6.

$$\begin{aligned} \widehat{Estado} = & 19.169 + 0.008 * 48 - 0.356 * 1 + 1.462 * 1 + 0.886 * 1 \\ & - 0.515 * 2.5 - 0.289 * 5.6 + 0.119 * 54.5 - 0.053 \\ & * 140 \end{aligned} \quad (10-3)$$

Reemplazando dichos valores en el modelo y calculado el estado estimado resultado que este es igual a 5.02, se procede a calcular la probabilidad de que esta persona reaccione favorablemente a la enfermedad o por lo contrario no logre una evolución favorable, así se tiene:

$$P\left(Y = \frac{1}{X}\right) = \frac{\exp^{\beta_0 + \beta_1 X_1 + \dots + \beta_i x_i}}{1 + \exp^{\beta_0 + \beta_1 X_1 + \dots + \beta_i x_i}} \quad (11-3)$$

$$= \frac{\exp^{5.02}}{1 + \exp^{5.02}}$$

$$= 0.9934$$

$$\cong 1$$

Si se tiene una paciente diagnosticada con cáncer de mama cuya edad es de 48 años, al momento presenta un peso de 54.5 kg, una altura de 140 cm, es de raza blanca, con una estadificación general de N=3, M=1, T= 4, reafirmada con el tipo de carcinoma in situ de mama, cuyo tiempo de evolución es de 2.5 años, desarrollando un tumor de tamaño 5.6; la probabilidad de que esta sobreviva favorablemente es nula, dado que esta paciente evolucionara negativamente ante la enfermedad.

En base a los resultados hallados con un primer modelo, es importante un segundo análisis basándose en las variables consideradas únicamente como significativas, para ver si mejora en algo lo que ayudara a predicciones más aproximadas a las observaciones.

3.4.2. *Modelo Logístico 2*

Para el segundo modelo se tomara en cuenta todas y cada una de aquellas variables que resultaron significantes al realizar una regresión logística al modelo 1, en el caso de las variables cualitativas se tomó en cuenta a la variable en general como tal, sin importar cuál de sus niveles y cuan significativo resulto ser, lo que se pretende es verificar si de esta manera niveles que en el anterior modelo no dieron significativos en este si lo sean, o por otro lado reafirmar la no significancia de dichos niveles o variables. Con lo que teniendo nuevamente en cuenta a la variable cualitativa dicotómica “Estado” con categóricas “0=Evolución favorable frente a la enfermedad” y “1= Evolución no favorable frente a la enfermedad” como la variable que será estimada se tiene:

$$\widehat{Estado} = \beta_0 + \beta_1 T + \beta_2 N + \beta_3 M + \beta_4 Estadio + \beta_5 Tiempo \quad (12-3)$$

$$+ \beta_6 Tama\tilde{n}o + \beta_7 Peso + \beta_8 Estatura$$

Una vez más teniendo nuestro modelo base, se podría desglosar debido al carácter cualitativo de algunas variables todavía en estudio, lo que precisara, la búsqueda de la significancia de estas variables y de sus niveles, que acompañadas de las variables de carácter cuantitativo provean de una correcta información, convirtiéndose en un modelo preciso, que mejore al anterior y que permita aproximarse de mejor manera que se hizo con el modelo anterior, en cuanto a observaciones y ello valor predicho una vez planteado el modelo, con ello se tiene el siguiente modelo de regresión:

$$\begin{aligned}
\widehat{Estado} = & \beta_0 + \beta_1 T_0 + \beta_2 T_1 + \beta_3 T_2 + \beta_4 T_3 + \beta_5 T_4 + \beta_6 T_{is} + \beta_7 T_x \\
& + \beta_8 N_0 + \beta_9 N_1 + \beta_{10} N_2 + \beta_{11} N_3 + \beta_{12} N_x + \beta_{13} M_0 \\
& + \beta_{14} M_1 + \beta_{15} M_x + \beta_{16} Estadio_0 + \beta_{17} Estadio_1 \\
& + \beta_{18} Estadio_2 + \beta_{19} Estadio_3 + \beta_{20} Estadio_4 \\
& + \beta_{21} Estadio_x + \beta_{22} Tiempo + \beta_{23} Tamaño + \beta_{24} Peso \\
& + \beta_{25} Estatura
\end{aligned} \tag{13-3}$$

Una vez establecido el segundo modelo logístico, se procede a realizar la correspondiente regresión, poniendo a prueba el segundo modelo con el que se tratara de explicar el comportamiento de las variables frente a la supresión de aquellas que no resultaron significativas:

3.4.2.1. Coeficientes de regresión del segundo modelo

Tabla 25-3: Coeficientes de regresión del modelo 2.

Variable	Coefficiente
(Intercept)	3,822
T1	-1,113
T2	-1,388
T3	-1,059
T4	-0,4
Tis	0,129
Tx	-0,319
N1	0,429
N2	0,546
N3	1,492
Nx	-0,121
M1	2,264
Mx	0,453
Estadio1	0,533
Estadio2	0,882
Estadio3	0,696
Estadio4	0,268
Estadiox	-14,814
Tiempo	-0,509
Tamaño	-0,287
Peso	0,122
Estatura	-0,048

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Con los respectivos coeficientes de regresión, que corresponden a cada una de las variables significativas tanto aquella de carácter cuantitativo, como aquellas de carácter cualitativo con cada uno de sus niveles encontradas en el primer modelo, se tiene el segundo modelo desglosado.

$$\begin{aligned}
 \widehat{Estado} = & 3.822 + 0T_0 - 1.113T_1 - 1.388T_2 - 1.059T_3 - 0.400T_4 \\
 & + 0.129T_{is} - 0.319T_x + 0N_0 + 0.429N_1 + 0.546N_2 \\
 & + 1.492N_3 - 0.121N_x + 0M_0 + 2.264M_1 + 0.453M_x \\
 & + 0Estadio_0 + 0.533Estadio_1 + 0.882Estadio_2 \\
 & + 0.696Estadio_3 + 0.268Estadio_4 - 14.814Estadio_x \\
 & - 0.509Tiempo - 0.287Tamaño + 0.122Peso \\
 & - 0.048Estatura
 \end{aligned} \tag{14-3}$$

Dado que, en este modelo de igual forma, existen coeficientes considerados como poco significativos, se retiran las variables acompañantes, con el fin de reducir el modelo aún más lo que nos permita mejor manipulación del mismo.

$$\begin{aligned}
 \widehat{Estado} = & 3.822 - 1.113T_1 - 1.388T_2 - 1.059T_3 - 0.400T_4 + 0.129T_{is} \\
 & - 0.319T_x + 0.429N_1 + 0.546N_2 + 1.492N_3 - 0.121N_x \\
 & + 2.264M_1 + 0.453M_x + 0.533Estadio_1 \\
 & + 0.882Estadio_2 + 0.696Estadio_3 + 0.268Estadio_4 \\
 & - 14.814Estadio_x - 0.509Tiempo - 0.287Tamaño \\
 & + 0.122Peso - 0.048Estatura
 \end{aligned} \tag{15-3}$$

3.4.2.2. Resumen de la regresión del segundo modelo

Con el segundo modelo se procede de igual forma, que lo hecho con el primer modelo, recordando que se observa las variables que son significativas estableciendo para ello un nivel de significancia con lo que ayude a rechazar o no rechazar la significancia de una variable dentro del modelo, en base a una prueba de comparación entre el valor de la probabilidad encontrada con respecto al nivel de significancia establecido, se sigue trabajando con un nivel de confianza (NC) del 95%, por lo que nuevamente se asume un nivel de significancia del 5% (denotando al nivel de significancia con α , con lo que $\alpha = 0.05$.)

$$\mathbf{H_0: La variable es significativa} \tag{16-3}$$

$$\mathbf{H_1: La variable no es significativa}$$

A simple vista pareciera que la significancia de las variables no cambia, el número de variables significantes con el nivel de confianza propuestos siguen siendo los mismos, lo que no causaría mayor cambio al modelo, sin embargo, eso es de manera intuitiva, por lo que es preciso estudiar a fondo el modelo.

Tabla 26-3: Resumen de regresión modelo 2.

Variable	Estimación	Std. Error	Valor Z	Pr(> z)	Significancia
(Intercept)	3,822	3,485	1,097	0,273	
T1	-1,113	0,544	-2,046	0,04	*
T2	-1,388	0,529	-2,625	0,009	**
T3	-1,059	0,587	-1,805	0,071	.
T4	-0,4	0,631	-0,635	0,526	
Tis	0,129	1,309	0,098	0,923	
Tx	-0,319	0,472	-0,677	0,498	
N1	0,429	0,216	1,984	0,047	*
N2	0,546	0,329	1,657	0,097	.
N3	1,492	1,132	1,319	0,187	
Nx	-0,121	0,308	-0,394	0,693	
M1	2,264	0,623	3,632	0,0003	***
Mx	0,453	0,194	2,334	0,019	*
Estadio1	0,533	0,391	1,363	0,173	
Estadio2	0,882	0,408	2,16	0,031	*
Estadio3	0,696	0,491	1,417	0,156	
Estadio4	268	0,584	0,458	0,647	
Estadios	-14,814	535,411	-0,028	0,978	
Tiempo	-0,509	0,052	-9,767	2x10-16	***
Tamaño	-0,287	0,057	-5,039	4x10-07	***
Peso	0,122	0,034	3,541	0,0004	***
Estatura	-0,048	0,019	-2,446	0,014	*

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.4.2.3. Variables significativas del segundo modelo

Analizando ya cada una de las variables, se observa que la significancia cambia en algo, no mayoritariamente de una forma más palpable, que emocione a pensar que el modelo se convertirá en mucho más preciso y con una precisión mayor, sin embargo, el aumento de significancia para algunos niveles y variables es un buen predecesor para tomar en cuenta que podrá haber un mejoramiento en cuanto a precisión, y por ende a predicciones.

Tabla 27-3: Variables significativas del modelo 2.

Variable	Nivel Significancia	Conclusión
T1	*	Significativa con $\alpha = 0.05$
T2	*, **	Significativa con $\alpha = 0.05$ o significativa con $\alpha = 0.01$
T3	.	Significativa con $\alpha = 0.10$
N1	*	Significativa con $\alpha = 0.05$
N2	.	Significativa con $\alpha = 0.10$
M1	*, **	Significativa con $\alpha = 0.05$ o significativa con $\alpha = 0.01$
Mx	*	Significativa con $\alpha = 0.05$

Estadio2	*	Significativa con $\alpha = 0.05$
Tiempo	*, **	Significativa con $\alpha = 0.05$ o significativa con $\alpha = 0.01$
Tamaño	*, **	Significativa con $\alpha = 0.05$ o significativa con $\alpha = 0.01$
Peso	*, **	Significativa con $\alpha = 0.05$ o significativa con $\alpha = 0.01$
Estatura	*	Significativa con $\alpha = 0.05$

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.4.2.4. Evaluación del segundo modelo

Al igual a lo hecho con el primer modelo se procederá a evaluarlo en base a: su significancia en base a la prueba de Likelihood ratio, la precisión que este tiene de discriminar las observaciones y predicciones en base a una matriz de precisión, analizando cual es el error al precisar y predecir, y finalmente analizando el área bajo de la curva ROC.

3.4.2.5. Significancia del segundo modelo

Como ya se hizo se toma en cuenta principalmente un nivel de Nivel de confianza (NC) y su respectivo nivel de significancia (α), con el cual se plantea una hipótesis, la cual es sometida respectivamente para aprobar la significancia del modelo.

$$H_0: \text{El modelo es significativo} \quad (17-3)$$

$$H_1: \text{El modelo no es significativo}$$

Trabajando al 95% de confianza se observa suficiente evidencia estadística significativa para no rechazar la hipótesis nula y de esta manera concluir, que al igual que el primer modelo, este segundo modelo planteado tampoco es significativo.

Tabla 28-3: Significancia del modelo 2.

Diferencia de residuos	Grados de libertad	Valor p	alfa	Decisión
224.698	21	4.69×10^{-36}	0,05	Significativo

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.4.2.6. Precisión del segundo modelo

Aplicando el proceso realizado anteriormente realizado se resume a lo siguiente:

Tabla 29-3: Matriz de confusión del modelo 2.

Comparación	Predicciones	
Observaciones	0	1
0	198	157

Para este segundo modelo de regresión se observa que hay variación, aunque es mínima, puede resultar de mucha ayuda como mejoramiento al primer modelo ya planteado, o por su parte puede resultar que el segundo modelo no sea mejor que el anterior. Es importante observar que tanto mejora la precisión del modelo o que tanto disminuye, con respecto a la discriminación y a la capacidad de predicciones positivas correctas.

$$\begin{aligned}
 \text{Precisión} &= \frac{d}{b+d} * 100 \\
 &= \frac{490}{490 + 157} * 100 \\
 &= \frac{490}{647} * 100 \\
 &= 75.73415765 \\
 &\cong 75.73\%
 \end{aligned}
 \tag{18-3}$$

El segundo modelo planteado se observa que puede discriminar correctamente un 75.73% de predicciones positivas correctas, por lo que se puede considerar al modelo como apto para predecir, ya que las predicciones se aproximaran de buena manera a las observaciones.

Sin embargo, hay un leve decrementó con respecto al primer modelo en cuanto a la capacidad para precisar correctamente las predicciones positivas verdaderas (precisión del primer modelo = 76.41%). Esta leve variación pasa casi desapercibida en el gráfico de la matriz de precisión, sin embargo, matemáticamente se puede percibir dicha variación, con lo que hace pensar lo importante de la precisión de un modelo, y cuál es el error que este tiene para precisar, para en base a ello tomar decisiones, con la precisión calculada se tiene una idea que se corrobora con el respectivo cálculo del error.

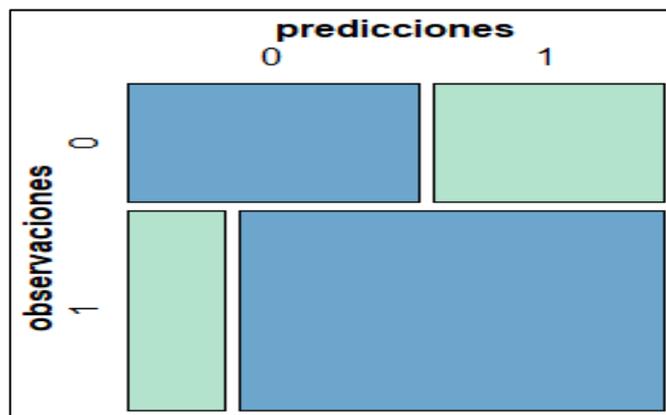


Gráfico 19-3: Matriz de confusión del modelo 2.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.4.2.7. Error de precisión del segundo modelo

Conociendo que la capacidad de discriminación y precisión del segundo modelo sufre una leve disminución, se intuye que el error al precisar aumentara.

$$\begin{aligned} \mathbf{Error} &= \frac{b + c}{a + b + c + d} * 100 \\ &= \frac{157 + 111}{956} * 100 \\ &= 28.03\% \end{aligned} \tag{19-3}$$

En efecto hay un leve aumento en cuanto al nivel de error al precisar, ahora se tiene un error de precisión del 28.03%, que es ligeramente mayor a 27.51% del primer modelo, a pesar de esto se puede considerar aun bueno al modelo, sin embargo, no tan idóneo para usar, e infiriendo que este segundo este no sería tan adecuado como el primer modelo.

3.4.2.8. Curva ROC del segundo modelo

Observando que el nivel de precisión bajo y el error aumento con una primera prueba, es factible comparar la siguiente prueba para ver cuál de las dos admite más variación, y de manera contigua, afirmar lo dicho o de otro lado contradecir, con respecto a la capacidad de precisión del segundo modelo.

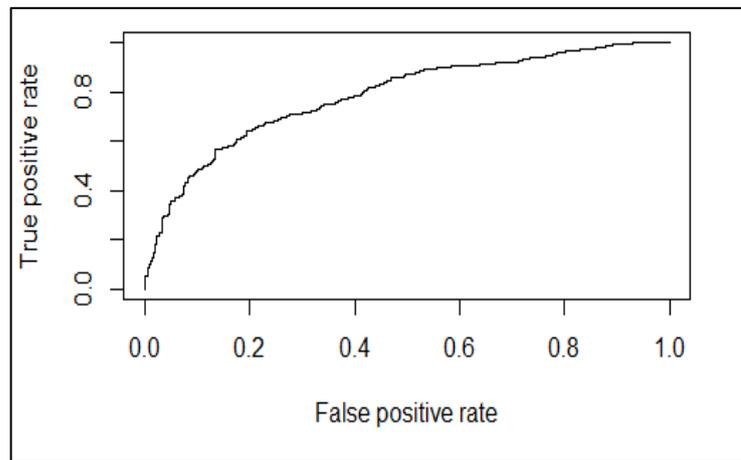


Gráfico 20-3: Curva ROC del modelo 2.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

En este caso el área bajo la curva es igual a 0.78, similar al área bajo la curva del primer modelo, por lo que se admitiría al segundo modelo de igual forma con una capacidad de discriminación aceptable, teniendo en cuenta que con el primer criterio hay una variación casi insignificante, lo que permitiría sugerir que el hecho de ocupar únicamente las significativas de manera general, y

dejando de lado aquellas que no eran significativas no influye mucho, lo que permite pensar que pasará si únicamente se trabajó con los niveles significativos.

3.4.2.9. Predicciones del segundo modelo

Al considerar al segundo modelo como apto para realizar predicciones, por su aceptable capacidad de precisión, se realiza de manera similar a lo ya hecho, comparando de esa manera tanto los valores de las observaciones como los valores de las predicciones realizadas, y se verifica la variación existente.

Tabla 30-3: Variación observaciones vs predicciones 2.

Dato	Categorías	
	0	1
Observaciones	355	601
Predicciones	309	647
Variación		9,62%

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

El porcentaje de variación de los valores predichos con respecto a los valores de observación, incrementa a 9.62%, algo que se veía venir, aunque es leve el incremento, es un indicador claro de que las predicciones tendrán muchos más errores, claramente revisado al saber que el error al precisar aumento.

A pesar de todo el segundo modelo igual sigue siendo un modelo aceptable para utilizarlo, y al igual que el modelo uno se verifico que pasa dentro de las predicciones.

Tabla 31-3: Proporción de predichos modelos 2.

Dato	Categorías		
	0	NI	1
Predichos	250	117	589
Proporción	26,15%	12,24%	61,61%

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Se observa que los valores para los casos específicos 0=evolución favorable del paciente y 1=evolución no favorable del paciente, sufren un leve decremento en su porcentaje total, teniendo así que ahora el 26.15% corresponden a casos positivos de predicción correcta con el valor 0, y que el 61.61% corresponden a los casos de predicciones correctas de los casos con valor 1. Esto permite observar que el porcentaje de los valores no nos dicen nada incrementa a 12.24%, lo que permite evidenciar que el número de predicciones inciertas incrementa.

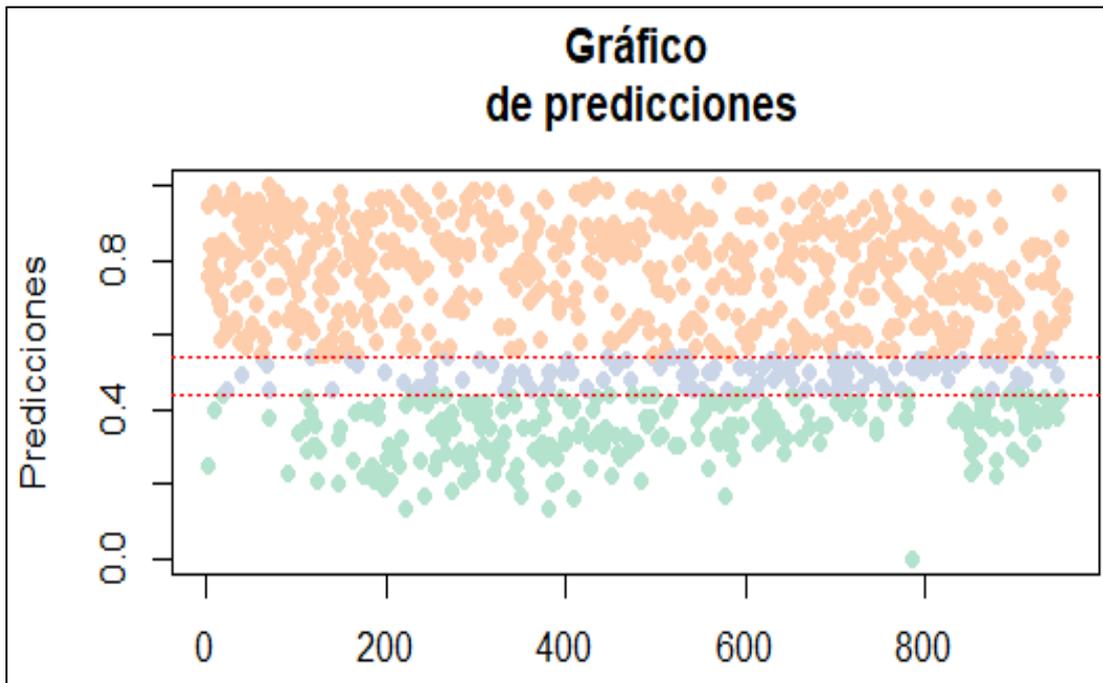


Gráfico 20-3: Proporción de predicciones modelo 2.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.4.2.10. Predicción ejemplo con el segundo modelo

La disminución de la precisión y por su contrario el aumento en el error de precisión, del segundo modelo, comparado con el primer modelo no son alarmantes, sin embargo, hay tener en cuenta que aumentan las posibilidades de generar error al calcular la probabilidad, se calculara la probabilidad en el segundo caso omitiendo aquellas variables que ya no se utilicen, para ver que tanto varían de terminar a si una decisión.

Se tiene una paciente diagnosticada con cáncer de mama cuya edad es de 48 años, al momento presenta un peso de 54.5 kg, una altura de 140 cm, es de raza blanca, con una estadificación general de N=3, M=1, T= 4, generando un estadio grado 2, reafirmada con el tipo de carcinoma in situ de mama, cuyo tiempo de evolución es de 2.5 años, desarrollando un tumor de tamaño 5.6.

$$\begin{aligned} \widehat{Estado} = & 3.822 + 1.493 * 1 + 2.264 * 1 - 0.400 * 1 + 0.882 * 1 \\ & - 0.509 * 2.5 - 0.287 * 5.6 + 0.122 * 54.5 - 0.048 \\ & * 140 \end{aligned} \quad (20-3)$$

Se obtuvo un estado estimado de 5.05, se procede a calcular la probabilidad de que esta persona reaccione favorablemente a la enfermedad o por lo contrario no logre una evolución favorable, así se tiene:

$$\begin{aligned}
P\left(Y = \frac{1}{X}\right) &= \frac{\exp^{\beta_0 + \beta_1 X_1 + \dots + \beta_i X_i}}{1 + \exp^{\beta_0 + \beta_1 X_1 + \dots + \beta_i X_i}} \\
&= \frac{\exp^{5.05}}{1 + \exp^{5.05}} \\
&= 0.9937 \\
&\cong 1
\end{aligned}
\tag{21-3}$$

Si se tiene una paciente diagnosticada con cáncer de mama cuya edad es de 48 años, al momento presenta un peso de 54.5 kg, una altura de 140 cm, es de raza blanca, con una estadificación general de N=3, M=1, T= 4, reafirmada con el tipo de carcinoma in situ de mama, cuyo tiempo de evolución es de 2.5 años, desarrollando un tumor de tamaño 5.6; se observa que la paciente tendrá una evolución no favorable ante la enfermedad.

Planteando un segundo modelo, esperando que resulte un nivel con mayor precisión al utilizar únicamente aquellas variables en general que resultaron significativas con el primer modelo, se pudo observar que este al contrario de mejorar, cayo levemente, a pesar de que siguió siendo un modelo apto, no pudo ser un mejor modelo. Observando que, tanto en el primer modelo como en el segundo modelo, hay variables significativas de carácter cuantitativo, y hay niveles de variables cualitativas significativos, es preciso estudiar apartando todo aquello que no sea significativo para ver si con ello, el modelo puede tener una mayor precisión y por ende una mejor capacidad predictiva.

3.4.3. Modelo Logístico 3

Estudiados tanto el modelo 1 como el modelo 2, se pretende estudiar otro modelo, uno que sea más específico, en el cual se utilice únicamente aquellas variables y niveles que se consideran significativas coincidentemente en los dos modelos anteriores planteado, con el fin de buscar una mejor precisión de un modelo, con mayor precisión y capacidad predictiva. Con lo que teniendo nuevamente en cuenta a la variable cualitativa dicotómica “Estado” con categóricas “0=Evolución favorable frente a la enfermedad” y “1= Evolución no favorable frente a la enfermedad” como la variable que será estimada se tiene:

$$\begin{aligned}
\widehat{Estado} &= \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_3 N_1 + \beta_4 M_1 + \beta_5 M_x + \beta_6 Estado_2 \\
&+ \beta_{22} Tiempo + \beta_{23} Tamaño + \beta_{24} Peso \\
&+ \beta_{25} Estatura
\end{aligned}
\tag{22-3}$$

Una vez establecido el tercer modelo logístico, se procede a realizar la correspondiente regresión, poniendo a prueba el tercer modelo con el que se tratara de verificar si el nuevo modelo puede superar en precisión a los dos antes planteados:

3.4.3.1. Coeficientes de regresión del tercer modelo

Tabla 32-3: Coeficientes de regresión del modelo 3.

Variable	Coefficiente
(Intercept)	3,361
T1	-0,489
T2	-0,626
N1	0,385
M1	2,189
Mx	0,255
Estadio2	0,295
Tiempo	-0,516
Tamaño	0,119
Peso	-0,048
Estatura	-0,277

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Realizada la correspondiente regresión, se determinó los coeficientes de regresión con lo que se puede plantear el modelo de mejor manera, ubicando cada uno de los coeficientes en las determinadas variables en estudio.

$$\begin{aligned} \widehat{Estado} = & 3.361 - 0.489T_1 - 0.626T_2 + 0.385N_1 + 2.189M_1 \\ & + 0.255M_x + 0.295Estadio_2 - 0.516Tiempo \\ & + 0.119Tamaño - 0.048Peso - 0.277Estatura \end{aligned} \quad (23-3)$$

3.4.3.2. Resumen de la regresión del tercer modelo

Teniendo en cuenta que las variables que son significativas estableciendo para ello un nivel de significancia con lo que ayude a rechazar o no rechazar la significancia de una variable dentro del modelo, en base a una prueba de comparación entre el valor de la probabilidad encontrada con respecto al nivel de significancia establecido, se sigue trabajando con un nivel de confianza (NC) del 95% para el tercer modelo planteado, asumiendo así un nivel de significancia del 5% (denotando al nivel de significancia con α , con lo que $\alpha = 0.05$.)

$$\begin{aligned} H_0: & \text{La variable es significativa} \\ H_1: & \text{La variable no es significativa} \end{aligned} \quad (24-3)$$

Se observa que en su mayoría las variables son significativas al nivel mínimo planteado, inclusive hay variables las cuales pueden ser significativas en un nivel más estricto de significancia.

Tabla 33-3: Resumen de regresión modelo 3.

Variable	Estimación	Std. Error	Valor Z	Pr(> z)	Significancia
(Intercept)	3,361	3,428	0,98	0,327	
T1	-0,489	0,266	-1,841	0,066	.
T2	-0,626	0,237	-2,647	0,008	**
N1	0,385	0,169	2,281	0,023	*
M1	2,189	0,555	3,943	8,03x10-05	***
Mx	0,255	0,168	1,517	0,129	
Estadio2	0,295	0,236	1,251	0,211	
Tiempo	-0,516	0,051	-10,122	2x10-16	***
Tamaño	0,119	0,034	3,539	0,0004	***
Peso	-0,048	0,019	-2,302	0,021	*
Estatura	-0,277	0,056	-4,951	7,37x10-07	***

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.4.3.3. Variables significativas del tercer modelo

Dado que se utilizaron las variables y niveles únicamente significativas denotadas en los dos modelos anteriormente planteadas, se supone el hecho de que todas resultaran significativas, sin embargo, hay variables que al ser modeladas en el conjunto (modelo de regresión 3), se vuelven no significativas o son significativas con un nivel más bajo de confianza, es de interés validar un modelo preciso, que sea sometido a niveles de error bajo, que haya suficiente precisión y confianza para establecer, los criterios y en base a ese modelo determinar predicciones que nos aproximen a las observaciones.

Tabla 34-3: Variables significativas del modelo 3.

Variable	Nivel Significancia	Conclusión
T1	.	Significativa con $\alpha = 0.10$
T2	*, **	Significativa con $\alpha = 0.05$ o significativa con $\alpha = 0.01$
N1	*	Significativa con $\alpha = 0.05$
M1	*, **	Significativa con $\alpha = 0.05$ o significativa con $\alpha = 0.01$
Mx		No significativa
Estadio2		No significativa
Tiempo	*, **	Significativa con $\alpha = 0.05$ o significativa con $\alpha = 0.01$
Tamaño	*, **	Significativa con $\alpha = 0.05$ o significativa con $\alpha = 0.01$
Peso	*	Significativa con $\alpha = 0.05$
Estatura	*, **	Significativa con $\alpha = 0.05$ o significativa con $\alpha = 0.01$

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.4.3.4. Evaluación del tercer modelo

Se procede de igual forma a los anteriores modelos en base a: su significancia en base a la prueba de Likelihood ratio, la precisión que este tiene de discriminar las observaciones y predicciones en base a una matriz de precisión, analizando cual es el error al precisar y predecir, y finalmente analizando el área bajo de la curva ROC.

3.4.3.5. Significancia del tercer modelo

Tomando en cuenta principalmente un nivel de Nivel de confianza (NC) y su respectivo nivel de significancia (α), con el cual se plantea una hipótesis, la cual es sometida respectivamente para aprobar la significancia del modelo.

$$H_0: \text{El modelo es significativo} \quad (25-3)$$

$$H_1: \text{El modelo no es significativo}$$

Al 95% de confianza se tiene suficiente evidencia estadística significativa para no rechazar la hipótesis nula y con ello concluir, que, de manera similar a los anteriores modelos, este modelo es significativo, por lo que se asume un modelo apto y significativo, con el que se puede trabajar.

Tabla 35-3: Significancia del modelo 3.

Diferencia de residuos	Grados de libertad	Valor p	alfa	Decisión
205.264	10	9.52×10^{-39}	0,05	Significativo

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.4.3.6. Precisión del tercer modelo

Aplicando el proceso realizado anteriormente realizado para los anteriores modelos se resume a lo siguiente:

Tabla 36-3: Matriz de confusión del modelo 3.

Comparación	Predicciones	
Observaciones	0	1
0	199	156
1	118	483

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

En cuanto al modelo, se observa que discrimina correctamente 483 verdaderos positivos, sin embargo, lo que se desea estudiar es la precisión con la que está discriminando ese número de verdaderos positivos, así como los demás casos.

$$\begin{aligned}
 \text{Precisión} &= \frac{d}{b + d} * 100 \\
 &= \frac{483}{483 + 156} * 100 \\
 &= \frac{483}{639} * 100 \\
 &= 75.58685446 \\
 &\cong 75.59\%
 \end{aligned}
 \tag{26-3}$$

El tercer modelo puede discriminar correctamente un 75.59% de predicciones positivas correctas, por lo que se puede considerar al modelo como apto para predecir, ya que las predicciones se aproximarán de buena manera a las observaciones, sin embargo, la capacidad de precisión de sus predicciones es inferior al modelo 1 (capacidad para precisar = 76.41%) y al modelo 2 (capacidad para precisar = 75.73%) respectivamente planteado anteriormente, entonces como paso en el anterior caso a menor porcentaje de precisión, mayor error al precisar, que será corroborado con el respectivo test.

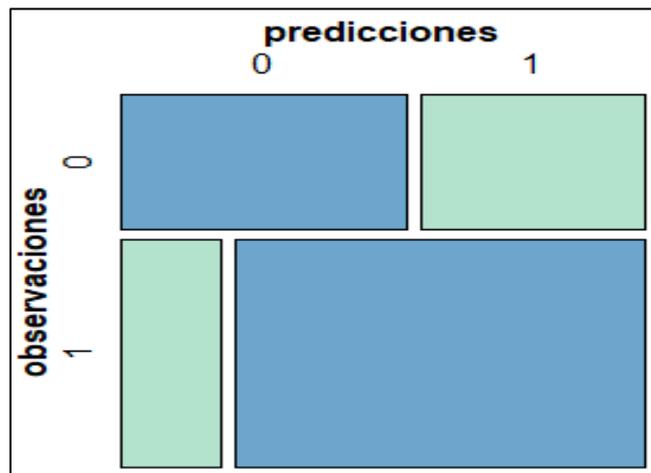


Gráfico 21-3: Matriz de confusión del modelo 3.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.4.3.7. Error de precisión del tercer modelo

Para el tercer modelo se tiene una capacidad de precisión por debajo de sus similares, modelo 1 y 2 respectivamente, por lo que se supone el error de precisar aumenta.

$$\text{Error} = \frac{b + c}{a + b + c + d} * 100
 \tag{27-3}$$

$$\begin{aligned}
&= \frac{156 + 118}{956} * 100 \\
&= \frac{274}{956} * 100 \\
&= 0.28661088 \\
&\cong 28.66\%
\end{aligned}$$

A pesar de que el porcentaje de error al precisar incrementa levemente, es preciso notar que no será conveniente utilizar este modelo a pesar de que es apto para poder usarlo, porque su capacidad de precisión super al 70%, sin embargo, no es lo más idónea, para conocer y aproximarse mejor a las observaciones.

3.4.3.8. Curva ROC del tercer modelo

Al bajar el nivel de precisión, es correcto suponer que el nivel de error subirá, sin embargo, no se puede discutir con una sola prueba, por lo que la base en un segundo fundamento, corroborará la idea planteada o por otro lado la desmentirá lo que hasta ahora se conocer del tercer modelo planteado.

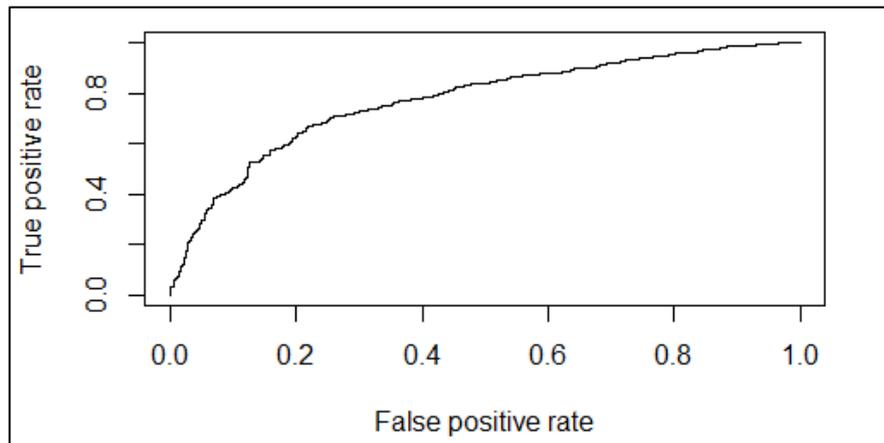


Gráfico 22-3: Curva ROC del modelo 3.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Con un valor de 0.77 respectivo al área existente bajo la curva, se concluye que el modelo es aceptable, que puede discriminar los datos de una manera aceptable, a la par que genera predicciones buenas.

3.4.3.9. Predicciones del tercer modelo

Se puede utilizar este tercer modelo para predecir, ya que es aceptable y apto, sin embargo, se corre el riesgo de que las predicciones no sean del todo correctas, y que en algo falle el tipo de discriminación que se realiza entre los verdaderos positivos, falsos positivos y demás.

Tabla 37-3: Variación observaciones vs predicciones 3.

Dato	Categorías	
	0	1
Observaciones	355	601
Predicciones	317	639
Variación		7.95%

Realizado por: Jhonatan F, Pinduisaca A. 2022.

La variación entre lo predicho y las observaciones, es inferior a lo visto en el modelo 2, sin embargo, no se puede confiar del todo, ya que de los 3 modelos este es el que menos capacidad y precisión tiene.

A pesar de todo el tercer modelo igual sigue siendo un modelo aceptable para utilizarlo, y al igual que el modelo uno se verificó que pasa dentro de las predicciones.

Tabla 38-3: Proporción de predichos modelos 3.

Dato	Categorías		
	0	NI	1
Predichos	248	125	583
Proporción	25,94%	13,08%	60,98%

Realizado por: Jhonatan F, Pinduisaca A. 2022.

Existe un incremento en la proporción de valores no específicos, por lo que se puede deducir que la variación entre lo predicho y lo observado en este modelo no es del todo correcta. Hay un 25.94% de datos predichos como casos favorables y un 60.98% como casos no favorables. Por otro lado 125 casos no pudieron ser identificados correctamente.

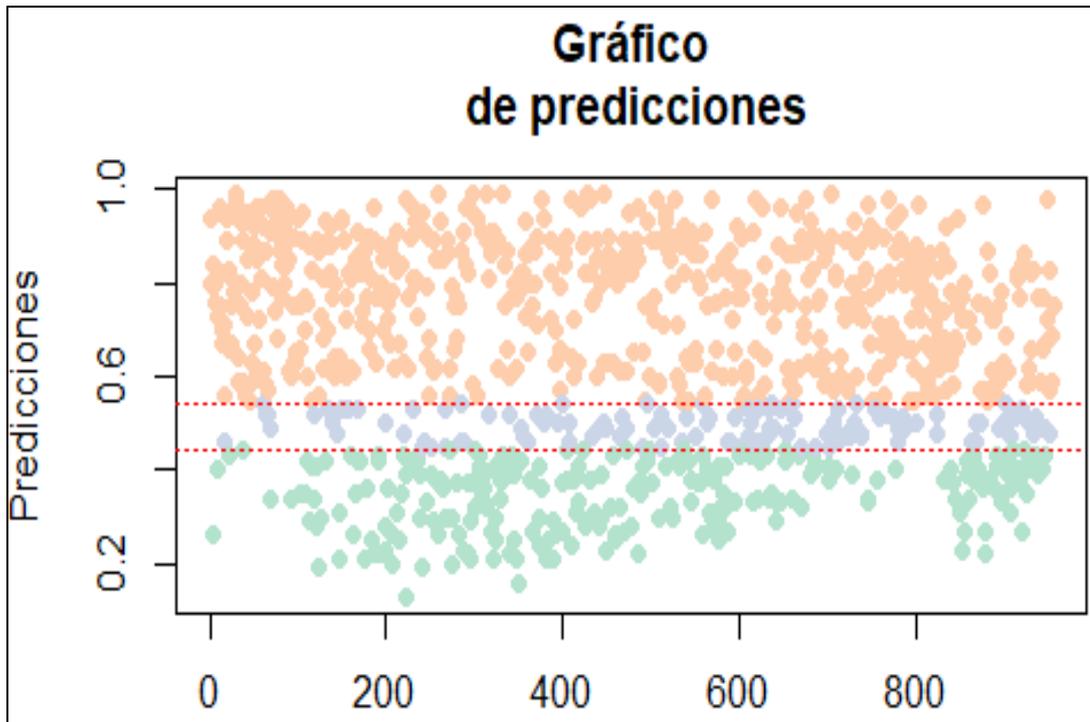


Gráfico 23-3: Proporción de predicciones modelo 3.

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

3.4.3.10. Predicción ejemplo con el tercer modelo

El tercer modelo resulto aceptable para predecir, a pesar de no ser el mejor, con él se prueba un ejemplo con el que se tratara de especificar una situación en la que nos sea factible aplicar este modelo.

Se tiene una paciente diagnosticada con cáncer de mama cuya edad es de 48 años, al momento presenta un peso de 54.5 kg, una altura de 140 cm, es de raza blanca, con una estadificación general de N=3, M=1, T= 4, generando un estadio grado 2, reafirmada con el tipo de carcinoma in situ de mama, cuyo tiempo de evolución es de 2.5 años, desarrollando un tumor de tamaño 5.6.

$$\begin{aligned}
 \widehat{\text{Estado}} &= 3.361 - 0.489 * 0 - 0.626 * 0 - 0.389 * 0 + 2.189 * 1 \\
 &\quad - 0.255 * 0 - 0.295 * 1 - 0.516 * 2.5 + 0.119 * 54.5 \\
 &\quad - 0.045 * 140 - 0.277 * 5.6
 \end{aligned} \tag{28-3}$$

Se obtuvo un estado estimado de 2.77, se procede a calcular la probabilidad de que esta persona reaccione favorablemente a la enfermedad o por lo contrario no logre una evolución favorable, así se tiene:

$$P\left(Y = \frac{1}{X}\right) = \frac{\exp^{\beta_0 + \beta_1 X_1 + \dots + \beta_i x_i}}{1 + \exp^{\beta_0 + \beta_1 X_1 + \dots + \beta_i x_i}} \tag{29-3}$$

$$= \frac{\exp^{2.77}}{1 + \exp^{2.77}}$$

$$= 0.9407$$

$$\cong 0.9$$

Si se tiene una paciente diagnosticada con cáncer de mama cuya edad es de 48 años, al momento presenta un peso de 54.5 kg, una altura de 140 cm, es de raza blanca, con una estadificación general de N=3, M=1, T= 4, reafirmada con el tipo de carcinoma in situ de mama, cuyo tiempo de evolución es de 2.5 años, desarrollando un tumor de tamaño 5.6; se observa que la paciente tendrá una evolución no favorable ante la enfermedad, según las probabilidades calculadas usando el modelo 3.

Se termina por observar que el modelo 3 más allá de resultar mejor que los anteriores modelos (1 y 2), fue el modelo menos preciso de los 3 por lo que su capacidad de precisión está limitada a pesar que el modelo sea aceptable y apto para poder predecir. Sus predicciones correrán el riesgo más que con los otros modelos de no aproximarse de la mejor forma a los datos de las observaciones.

3.4.4. Comparación de modelos

Una vez planteados los 3 modelos con los que principalmente se quería plantear un modelo que inmiscuya tantas variables como sea posible, y que de esa manera nos permita calcular probabilidad, y predecir el comportamiento de los pacientes diagnosticados con cáncer frente a la enfermedad, teniendo así.

Tabla 39-3: Comparación de modelos.

Parámetros	Modelos		
	Modelo 1	Modelo 2	Modelo 3
Significancia	Significativo	Significativo	Significativo
Precisión	76,40%	75,73%	75,59%
Error de precisión	27,51%	28,03%	28,66%
Variables inmiscuidas	Todas		T1
			T2
			N1
			M1
			Mx
			Estadio2
			Tiempo
Variables significativas	T N M Estadio Tiempo		Tiempo
			Tamaño
			Peso
			Altura
			T2
			N1
			M1

Tamaño	Tamaño	Peso
Peso	Peso	Altura
Altura	Altura	

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Se observa que los 3 modelos son significativos, y los 3 son aceptables para realizar predicciones por su aceptable porcentaje de precisión, cada uno tiene su modelo, sin embargo, concuerda al momento de utilizar las variables coinciden en unas cuantas, que resultan ser las más significativas dentro de los 3 modelos estudiados.

3.4.5. Factores Pronósticos Significativos

En base a las variables significativas encontradas en cada uno de los respectivos modelos planteados, se puede deducir a través de las mismas los factores pronósticos que se consideran significativos para el estudio y de manera directa influyen en la mortalidad en una paciente diagnosticada con cáncer.

3.4.5.1. Definidos con el modelo 1

Tabla 40-3: Factores pronósticos asumidos con el modelo 1.

NC	Nivel de significancia	Variable	Factor Pronostico
90%	0,1	T3	Estado T grado 3
		N2	Estado N grado 2
		T1	Estado T grado 1
		T2	Estado T grado 2
		N1	Estado N grado 1
		M1	Estado M grado 1
95%	0,05	Mx	Estado M grado x
		Estadio 2	Estadio grado 2
		Tiempo	Tiempo evolución
		Tamaño	Tamaño tumoral
		Peso	Peso paciente
		Estatura	Estatura paciente
		M1	Estado M grado 1
99%	0,01	Tiempo	Tiempo evolución
		Tamaño	Tamaño tumoral
		Peso	Peso paciente
		Estatura	Estatura paciente

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Trabajando con un 90% de confianza ($\alpha = 0.10$) hay suficiente evidencia estadística para definir como factores pronósticos significativos a Estado T grado 3 y Estado N grado 2, si se analiza al

95% de confianza ($\alpha = 0.05$) se tiene que se asume como factores pronósticos significativos a Estado T grado 1, Estado T grado 2, Estado N grado 1, Estado M grado 1, Estado M grado x, Estadio grado 2, Tiempo de evolución, Tamaño tumoral, Peso paciente, Estatura paciente y con un 99% de confianza ($\alpha = 0.01$) se pueden considerar como factores pronósticos significativos a Estado M grado 1, Tiempo de evolución, Tamaño tumoral, Peso paciente, Estatura paciente. Concluyendo en todos los casos que estos son significativos para el estudio.

3.4.5.2. Definidos con el modelo 2

Tabla 41-3: Factores pronósticos asumidos con el modelo 2.

NC	Nivel de significancia	Variable	Factor Pronóstico
90%	0,1	T3	Estado T grado 3
		N2	Estado N grado 2
		T1	Estado T grado 1
		T2	Estado T grado 2
		N1	Estado N grado 1
		M1	Estado M grado 1
95%	0,05	Mx	Estado M grado x
		Estadio 2	Estadio grado 2
		Tiempo	Tiempo evolución
		Tamaño	Tamaño tumoral
		Peso	Peso paciente
		Estatura	Estatura paciente
		T2	Estado T grado 2
99%	0,01	M1	Estado M grado 1
		Tiempo	Tiempo evolución
		Tamaño	Tamaño tumoral
		Peso	Peso paciente

Realizado por: Jhonnatan F. Pinduisaca A. 2022.

Con un 90% de confianza ($\alpha = 0.10$) hay suficiente evidencia estadística para definir como factores pronósticos significativos a Estado T grado 3 y Estado N grado 2, si se analiza al 95% de confianza ($\alpha = 0.05$) se asume como factores pronósticos significativos a Estado T grado 1, Estado T grado 2, Estado N grado 1, Estado M grado 1, Estado M grado x, Estadio grado 2, Tiempo de evolución, Tamaño tumoral, Peso paciente, Estatura paciente y con un 99% de confianza ($\alpha = 0.01$) se define como factores pronósticos significativos a Estado T grado 2, Estado M grado 1, Tiempo de evolución, Tamaño tumoral, Peso paciente.

3.4.5.3. Definidos con el modelo 3

Tabla 42-3: Factores pronósticos asumidos con el modelo 3.

NC	Nivel de significancia	Variable	Factor Pronóstico
90%	0,1	T1	Estado T grado 1
		T2	Estado T grado 2
		N1	Estado N grado 1
		M1	Estado M grado 1
		Tiempo	Tiempo evolución
		Tamaño	Tamaño tumoral
		Peso	Peso paciente
		Estatura	Estatura paciente
		T2	Estado T grado 2
		M1	Estado M grado 1
99%	0,01	Tiempo	Tiempo evolución
		Tamaño	Tamaño tumoral
		Estatura	Estatura paciente

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Con un 90% de confianza ($\alpha = 0.10$) hay suficiente evidencia estadística para definir como factores pronósticos significativos a Estado T grado 1, teniendo en cuenta un 95% de confianza ($\alpha = 0.05$) se define como factores pronósticos significativos a Estado T grado 2, Estado N grado 1, Estado M grado 1, Tiempo de evolución, Tamaño tumoral, Peso paciente, Estatura paciente y con un 99% de confianza ($\alpha = 0.01$) se considera como factores pronósticos significativos a Estado T grado 2, Estado M grado 1, Tiempo de evolución, Tamaño tumoral, Estatura paciente.

CONCLUSIONES

- Con el planteamiento de 3 modelos se tiene más oportunidad a tener variables más precisas, se observó dentro del estudio que de los factores pronósticos ya preestablecidos y tomados en cuenta para este estudio no todos resultados influyentes, por lo que no todos denotan mucha información al querer predecir sobre el comportamiento de la enfermedad sobre un paciente, algo que se vio evidente, a pesar de la diferencia de los 3 modelos, es que los 3 coincidían en exponer a ciertas variables como significativas, por lo que en conclusión, los factores más predominantes e influyentes son: peso, estatura, tamaño tumoral, tiempo de evolución, estadio, los estados completos de la estadificación, que conlleva el estudio del tamaño y ubicación del tumor principal, la invasión ganglionar, y el estudio de la metástasis generada por el tumor principal.
- Dentro de las relaciones entre variables, se evidencia claramente que las variables peso, estatura, tamaño tumoral, tiempo de evolución se relacionan más entre ellas, y estas generan mucha influencia sobre el estado de evolución del paciente, determinan y dan mucha información sobre si el paciente evoluciona favorablemente o por lo contrario recae en la enfermedad o por lo más trágico llega a fallecer, con las variables categóricas al no tener mucha significancia en todos sus niveles resulta que estas no generan una dependencia prioritaria, pero sin embargo influyen de manera indirecta sobre el estado final del paciente.
- Por lo que se vio en este caso que, a pesar de tener varias variables dentro del estudio, no todas resultaron significativas por lo que de manera relacional se induce a que los factores pronósticos representados por esas variables no son significativos, no influyen o no aportan de mucho al modelo. Con varias variables de carácter cualitativo, se tuvo que no todos los niveles de estas eran significativos por lo que influía a que las variables no lo sean; por lo que lo que se reitera en considerar significativas con un nivel de confianza de 95% a peso, estatura, tamaño tumoral, tiempo de evolución, estadio, los estados completos de la estadificación
- Se pudo constatar que los 3 modelos planteados fueron significativos, así como los 3 modelos tenían una capacidad de precisión aceptable, con lo que inducía a que estos modelos tenían una aceptable capacidad predictiva. Por ello de los 3 modelos, el mejor fue el primero en donde se tomó todas las variables para el estudio, este tiene mayor capacidad de precisión para predecir que los demás, y el error de precisar al predecir es menor al de sus contrincantes, a pesar de que algunas de las variables no resultaron significativas, es importante el uso de las mismas, porque puede ser que en otras condiciones lo sean.

RECOMENDACIONES

- El estudio de factores pronósticos es un tema muy amplio e interesante para tratar desde el ámbito estadístico, es un tema muy poco tratado por lo que un estudio posterior que sea una complementación al presente, nos permitirá tener una idea más clara sobre el tema desde la parte estadística, no obviando el hecho de poder ser un apoyo para la visión a partir del área médica, el estudio de dichos factores puede englobar mucho más, dado que la información recolectada no es suficiente para el correcto estudio del mismo, se pudiese esperar un tiempo para que haya más datos sentados completos, o a su vez contemplar más centro médicos donde poder recolectar, tratar y estudiar la información necesaria, para poder fundamentar más sobre el tema.
- La relación de dependencia entre las variables que inmiscuyen un estudio, permitirá en la mayoría de ellos casos generar y establecer modelos con mayor precisión y capacidad de predicción, siempre y cuando se estudie a todas las variables bajo las mismas condiciones, ya que esto es relevante dentro del estudio.
- La significancia de un modelo predictivo será de suma importancia antes de avanzar más en el estudio, no confiarse y parar cuando se consigue un modelo aceptable, tratar de llegar a un modelo que precise de información mucha más exacta y apegada a observaciones reales, una minuciosa validación de los modelos y la generación de otros con cual comparar, ayuda a que valores que deseamos predecir tengan mayor fiabilidad, con lo que reduzca el error y aumente la precisión.
- En las peores circunstancias es preferible tener un modelo que tenga una capacidad aceptable con algunas variables de cualquier tipo que permita experimentar con combinaciones que deriven a valores más pegados a los reales, que un modelo que tenga una mayor precisión pero que termine con pocas variables, lo que ocasiona que se vuelvan constantes las combinaciones omitiendo de esta manera la oportunidad de poder generar más casos apegados y aproximados a la realidad, que cohiba el hecho de poder experimentar.

BIBLIOGRAFÍA

- AMAT, J.**, 2016. Regresión logística simple y múltiple. [en línea]. [Consulta: 30 noviembre 2020]. Disponible en: https://www.cienciadedatos.net/documentos/27_regresion_logistica_simple_y_multiple.
- ASCO**, 2014. Cáncer de mama. *Cancer.Net* [en línea]. [Consulta: 30 noviembre 2020]. Disponible en: <https://www.cancer.net/es/tipos-de-cancer/c%C3%A1ncer-de-mama>.
- ASCO**, [sin fecha]. Tipos de Cáncer. *Cancer.Net* [en línea]. [Consulta: 30 noviembre 2020]. Disponible en: <https://www.cancer.net/es/tipos-de-c%C3%A1ncer>.
- Cap2.pdf** [en línea], [sin fecha]. S.l.: s.n. [Consulta: 24 agosto 2021]. Disponible en: https://sisbib.unmsm.edu.pe/bibvirtualdata/Tesis/Basic/Salcedo_pc/enPDF/Cap2.PDF.
- CDCESPANOL**, 2020. ¿Qué es el cáncer de mama? *Centers for Disease Control and Prevention* [en línea]. [Consulta: 30 noviembre 2020]. Disponible en: https://www.cdc.gov/spanish/cancer/breast/basic_info/what-is-breast-cancer.htm.
- Cifras de Ecuador – Cáncer de Mama – Ministerio de Salud Pública**. [en línea], 2018. [Consulta: 24 agosto 2021]. Disponible en: <https://www.salud.gob.ec/cifras-de-ecuador-cancer-de-mama/>.
- COLON 15**, [sin fecha]. Diagnóstico y seguimiento del cáncer de mama - Colón15 Sevilla - Centro Médico Avanzado. [en línea]. [Consulta: 25 agosto 2021]. Disponible en: <https://colon15.com/radiologia/diagnostico-seguimiento-cancer-mama/>.
- FERNANDEZ, M., GALINDO, M., ARAGONCILLO, P., MILLANA, M. y FARIÑA, J.**, 2009. *Factores pronósticos en el cáncer de mama*. 2009. S.l.: s.n.
- FERNANDEZ, S., CORDERO, J. y CORDOVA, A.**, 2020. Estadística: qué es, objetivos, tipos, partes, elementos • gestiopolis. *gestiopolis* [en línea]. [Consulta: 30 noviembre 2020]. Disponible en: <https://www.gestiopolis.com/que-es-estadistica-tipos-y-objetivos/>.

FLORES-LUNA, L., SALAZAR-MARTÍNEZ, E., DUARTE-TORRES, R.M., TORRES-MEJÍA, G., ALONSO-RUIZ, P. y LAZCANO-PONCE, E., 2008. Factores pronósticos relacionados con la supervivencia del cáncer de mama. *Salud Pública de México*, vol. 50, no. 2, pp. 119-125. ISSN 0036-3634. DOI 10.1590/S0036-36342008000200005.

GOOGLE MAPS, 2021. solca ubicacion - Buscar con Google. [en línea]. [Consulta: 24 agosto 2021]. Disponible en: [https://www.google.com/search?sxsrf=ALeKk00Np9qeVjmCLTN7pQ9RUYNaoPKGEg:1629820499089&q=solca+ubicacion&npsic=0&rflfq=1&rldoc=1&rllag=-2175032,-79880766,146&tbm=lcl&sxsrf=ALeKk00Np9qeVjmCLTN7pQ9RUYNaoPKGEg:1629820499089&sa=X&ved=2ahUKEwiq5JS_gsryAhViVTABHZIKCo8QtgN6BAgOEA M&biw=1536&bih=754#rfl=hd::si:9964731681063839988,1,Cg9zb2xjYSB1YmljYW Npb25luP6yxp2vgIAIWgsQABgAIgVzb2xjYZIBD3JIZ2lzdHJ5X29mZmljZQ;mv:\[\[-2.1709994000000004,-79.87728229999999\],\[-2.2235364,-79.925495\]\]](https://www.google.com/search?sxsrf=ALeKk00Np9qeVjmCLTN7pQ9RUYNaoPKGEg:1629820499089&q=solca+ubicacion&npsic=0&rflfq=1&rldoc=1&rllag=-2175032,-79880766,146&tbm=lcl&sxsrf=ALeKk00Np9qeVjmCLTN7pQ9RUYNaoPKGEg:1629820499089&sa=X&ved=2ahUKEwiq5JS_gsryAhViVTABHZIKCo8QtgN6BAgOEA M&biw=1536&bih=754#rfl=hd::si:9964731681063839988,1,Cg9zb2xjYSB1YmljYW Npb25luP6yxp2vgIAIWgsQABgAIgVzb2xjYZIBD3JIZ2lzdHJ5X29mZmljZQ;mv:[[-2.1709994000000004,-79.87728229999999],[-2.2235364,-79.925495]]).

HOMES, D. y LEMESHOW, S., 2020. Curva ROC (curva de rendimiento diagnóstico). [en línea]. [Consulta: 24 agosto 2021]. Disponible en: <https://support.minitab.com/es-mx/minitab/18/macro-library/macro-files/anova-and-regression-macros/roc-receiver-operating-characteristic-curve/>.

LAGUNA, C., [sin fecha]. INTRODUCCIÓN A LA ESTADÍSTICA. , pp. 12.

MAURANDI-LÓPEZ, A., DEL RÍO ALONSO, L., GONZÁLEZ-VIDAL, A., FERRE JAÉN, M.E. y HERNÁNDEZ VICENTE, Á., 2019. Fundamentos Estadísticos para Investigación. Introducción a R y Modelos. [en línea], [Consulta: 24 agosto 2021]. DOI 10.5281/ZENODO.2628915. Disponible en: <https://zenodo.org/record/2628915>.

MAYO CLINIC, [sin fecha]. Cáncer de mama - Síntomas y causas - Mayo Clinic. [en línea]. [Consulta: 25 agosto 2021]. Disponible en: <https://www.mayoclinic.org/es-es/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470>.

MEDIAPLUS, 2020. Tumor: MedlinePlus enciclopedia médica. [en línea]. [Consulta: 30 noviembre 2020]. Disponible en: <https://medlineplus.gov/spanish/ency/article/001310.htm>.

NIH, 2011. Definición de factor pronóstico - Diccionario de cáncer del NCI - Instituto Nacional del Cáncer. [en línea]. [Consulta: 30 noviembre 2020]. Disponible en: <https://www.cancer.gov/espanol/publicaciones/diccionario/def/factor-pronostico.nciglobal,ncienterprise>

OCAMPO, R., [sin fecha]. Tumor mamario | Cirugía 2 | AccessMedicina | McGraw Hill Medical. [en línea]. [Consulta: 25 agosto 2021]. Disponible en: <https://accessmedicina.mhmedical.com/content.aspx?bookid=1434§ionid=94946138>.

OMS, 2007. OMS | Cáncer. *WHO* [en línea]. [Consulta: 30 noviembre 2020]. Disponible en: <http://www.who.int/topics/cancer/es/>.

ORTEGA, J.M.G., WONG, M.M.M., LÓPEZ, Z. y VALDÉZ, M.D., 2011. Factores pronósticos del cáncer de mama. , pp. 9.

PEREZ, J. y MERINO, M., 2013. Definición de tratamiento — Definicion.de. *Definición.de* [en línea]. [Consulta: 30 noviembre 2020]. Disponible en: <https://definicion.de/tratamiento/>.

QUEVEDO, F., 2011. Medidas de tendencia central y dispersión. *Medwave* [en línea], vol. 11, no. 03. [Consulta: 30 noviembre 2020]. ISSN 0717-6384. DOI 10.5867/medwave.2011.03.4934. Disponible en: </link.cgi/Medwave/Series/MBE04/4934>.

RECUERO, P., 2020. Cómo interpretar la matriz de confusión: ejemplo práctico - Think Big Empresas. *Think Big* [en línea]. [Consulta: 24 agosto 2021]. Disponible en: <https://empresas.blogthinkbig.com/como-interpretar-la-matriz-de-confusion-ejemplo-practico/>.

SABADELL MERCADAL, M.D., IZQUIERDO SANZ, M., PRATS DE PUIGS, M., MODOLELL ROIG, A. y ADROVER CEBRIÁN, E., 2017. *Factores pronósticos y predictivos en cáncer de mama: una visión evolutiva de la morfología a la genética*. S.l.: s.n. ISBN 978-84-697-5070-4.

SCH, [sin fecha]. default - Stanford Children's Health. [en línea]. [Consulta: 25 agosto 2021]. Disponible en: <https://www.stanfordchildrens.org/es/topic/default?id=anatomadelossenos-85-P03255>.

VIDA”, P.R. de los S.E. en generación de contenidos tecnológicos para los canales digitales de T.T.A. of T.L. en C.F. y M. en T.E.A. por las “tecnologías para la y PEDAGOGÍA, L.Q.N.H.L.V.M.F.P.L., 2018. Machine Learning a tu alcance: La matriz de confusión - Think Big Empresas. *Think Big* [en línea]. [Consulta: 24 agosto 2021]. Disponible en: <https://empresas.blogthinkbig.com/ml-a-tu-alcance-matriz-confusion/>.

ANEXOS

ANEXO A: DECLARACIÓN DE CONFIDENCIALIDAD

DECLARACION DE CONFIDENCIALIDAD, NO USO Y DIVULGACIÓN DE LA INFORMACIÓN

En la ciudad de Guayaquil, a 03 de Junio del 2021, comparece Jhonnatan Pindisace Alloica (nombre y apellidos de quien solicita información), por sus propios y personales derechos y/o en nombre y representación de ESPOCH (nombre de la institución), con domicilio en la ciudad de Buena Vista a efectos de suscribir la presente Declaración de Confidencialidad, no Uso y Divulgación de la Información, contenida en los siguientes términos:

PRIMERA.- Antecedentes.- La Sociedad de Lucha Contra el Cáncer del Ecuador SOLCA, es una institución de derecho privado, con finalidad social y sin fines de lucro, a cuyo cargo tiene el Instituto Oncológico "Dr. Juan Tanca Marengo", cuyas instalaciones se encuentran ubicadas en la Avenida Pedro Menéndez Gilbert y Av. Atahualpa Chávez, de la ciudad de Guayaquil. Por su naturaleza la Sociedad de Lucha contra el Cáncer del Ecuador tiene información muy confidencial, que son de su exclusivo uso y propiedad.

Por su parte el/la declarante, Jhonnatan Pindisace Alloica manifiesta que requiere de esta información ya sea de forma oral, gráfica o escrita y, en estos dos últimos casos, esté contenida en cualquier tipo de documento, con ocasión de la realización de trabajo de investigación propia o en colaboración con otras entidades, tesis, tesinas, reportes científicos, presentación en congresos nacionales e internacionales u otros eventos técnicos médicos.

SEGUNDA.- Declaración.- Por lo expuesto, declaro expresamente:

1.- Que la información facilitada gratuitamente por la Sociedad de Lucha Contra el Cáncer del Ecuador únicamente la utilizaré para el fin mencionado en la cláusula anterior, comprometiéndome a mantener la más estricta confidencialidad respecto de dicha información, advirtiendo de dicho deber de confidencialidad y secreto a compañeros, asociados y, a cualquier persona que, por su relación para con el declarante, deba tener acceso a dicha información.

2. Que las personas mencionadas en el párrafo anterior no podrán reproducir, modificar, hacer pública o divulgar a terceros la información objeto de la presente Declaración sin previa autorización escrita y expresa de la Sociedad de Lucha Contra el Cáncer del Ecuador SOLCA.

3. De igual forma, me comprometo a adoptar respecto de la información objeto de esta Declaración las mismas medidas de seguridad que adoptaría normalmente respecto a la información confidencial de su propio interés y/o de la Empresa o Institución, evitando en la medida de lo posible su pérdida, robo o sustracción.

942	73	Casada	Mestiza	0	x	x	0	A	Parte no especificada	Si	1	0,5	55,3	153	6,6
943	42	Soltera	Mestiza	x	x	x	0	A	Mama	No	1	0,6	57,3	156	6,8
944	50	Soltera	Mestiza	x	x	x	0	A	Mama	Si	1	1,5	54,8	153	6,6
945	39	Unión Libre	Mestiza	2	1	x	1	A	Mama	Si	0	2,9	57,1	160	6,8
946	56	Unión Libre	Mestiza	x	x	x	0	A	Mama	Si	1	1,1	52,5	151	6,4
947	68	Soltera	Mestiza	x	x	x	0	A	Mama	Si	0	3,4	56,9	157	6,8
948	68	Viuda	Mestiza	2	1	x	2	A	Mama	Si	1	2,8	53,7	155	6,5
949	69	Divorciada	Mestiza	4	2	1	4	A	Mama	Si	1	1	57,7	152	6,8
950	52	Soltera	Mestiza	x	x	x	0	A	Parte no especificada	Si	1	0,8	53,6	152	6,5
951	68	Viuda	Mestiza	4	2	0	3	A	Mama	Si	1	0,5	52,9	149	6,4
952	47	Casada	Mestiza	x	x	x	0	A	Mama	Si	0	3,1	54,2	150	6,6
953	62	Soltera	Mestiza	4	1	x	3	A	Mama	Si	1	2,4	51,3	155	6,3
954	63	Unión Libre	Mestiza	x	x	x	0	A	Mama	Si	1	1,7	59,6	158	7
955	59	Unión Libre	Mestiza	2	1	x	3	A	Mama	Si	1	1,7	56,8	156	6,8
956	56	Soltera	Mestiza	3	1	0	1	A	Cuadrante inferior interno de la mama	No	0	0,1	57,2	162	6,8

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

ANEXO D: CÓDIGO EN R

```

Preparación de Datos

vcd
Paquetes y Librerías Utilizados: ROCR, paletteer, readxl, fdth, psych, corrplot
datos<-as.data.frame(read_excel("Base final.xlsx"))
dim(datos)
Carga de datos: names(datos), datos1<-datos[-c(1,3,5,7,11,15,16)]
dim(datos1)
names(datos1)

```

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

```

Análisis Descriptivo

Edad: e<-fdt(datos1$Edad,breaks = "Sturges")
hist(datos1$Edad, main = "Histograma\nde la edad de pacientes",
col = "skyblue",xlab = "edades",ylab = "Frecuencia",xlim = c(20,100))
mean(datos1$Edad)
median(datos1$Edad)
sd(datos1$Edad)

Raza: r<-table(datos1$Raza)
barplot(r,col = c("skyblue","pink","green"),ylim = c(0,1000),
main = "Diagrama de Barras\n de la raza de pacientes")

Estatura: es<-fdt(datos1$Estatura,breaks = "Sturges")

```

```

hist(datos1$Estatura, main = "Histograma\nde la estatura de pacientes",
     col = "skyblue",xlab = "Estaturas",ylab = "Frecuencia",xlim = c(145,165),
     ylim = c(0,195))
mean(datos1$Estatura)
median(datos1$Estatura)
sd(datos1$Estatura)
pe<-fdt(datos1$Peso,breaks = "Sturges")
hist(datos1$Peso, main = "Histograma\ndel peso de pacientes",
     col = "skyblue",xlab = "Pesos",ylab = "Frecuencia",xlim = c(50,63),
     ylim = c(0,120))
mean(datos1$Peso)
median(datos1$Peso)
sd(datos1$Peso)
Cirugia<-table(datos1$Cirugia)
Cirugía barplot(Cirugia,col = c("skyblue","pink"),ylim = c(0,1000),
     main = "Diagrama de barras\n del estado de cirugía del paciente")
tev<-fdt(datos1$`Tiempo evolucion`,breaks = "Sturges",k = 7)
hist(datos1$`Tiempo evolucion`, main = "Histograma\ndel tiempo de evolución de la
enfermedad",
     col = "skyblue",xlab = "Tiempo",ylab = "Frecuencia",xlim = c(0,7),ylim =
Tiempo de evolución c(0,200),nclass = 7)
mean(datos1$`Tiempo evolucion`)
median(datos1$`Tiempo evolucion`)
sd(datos1$`Tiempo evolucion`)
tt<-fdt(datos1$Tamaño,breaks = "Sturges",k = 7)
hist(datos1$Tamaño, main = "Histograma\ndel tamaño del tumor",
     col = "skyblue",xlab = "Tamaño",ylab = "Frecuencia",xlim = c(0,8),ylim = c(0,200),nclass
Tamaño tumoral = 7)
mean(datos1$Tamaño)
median(datos1$Tamaño)
sd(datos1$Tamaño)
ub<-table(datos1$Parte)
Ubicación del tumor barplot(ub,col = palette("pastel 2"),ylim = c(0,1000),
     main = "Diagrama de barras\n de la ubicación del tumor",names.arg =
c("CS","CIE","CII","CSE","CSI","MAMA","NE"))
tum<-table(datos1$Tipo)
Tipo de tumor barplot(tum,col = palette("pastel 2"),ylim = c(0,1000),
     main = "Diagrama de barras\n del tipo de tumor",names.arg = c("A","B"))
ei<-table(datos1$Estadio)
Estadio barplot(ei,col = palette("pastel 2"),ylim = c(0,320),
     main = "Diagrama de barras\n del grado de estadio del paciente")
gt<-table(datos1$T)
Estado T barplot(gt,col = palette("pastel 2"),ylim = c(0,320),
     main = "Diagrama de barras\n del estado T")
gn<-table(datos1$N)
Estado N barplot(gn,col = palette("pastel 2"),ylim = c(0,320),
     main = "Diagrama de barras\n del estado N")
gm<-table(datos1$M)
Estado M barplot(gm,col = palette("pastel 2"),ylim = c(0,700),
     main = "Diagrama de barras\n del estado M")
ef<-table(datos1$Estado)
Estado barplot(ef,col = palette("pastel 2"),ylim = c(0,600),
     main = "Diagrama de barras\n del estado final del paciente",
     names.arg = c("Favorable","No favorable"))

```

Realizado por: Jhonnatan F. Pinduisaca A. 2022.

Análisis de dependencia de variables	
Preparacion de datos	er<-as.data.frame(read_excel("Libro1.xlsx"))
Dependencia de variables cuantitativas	cuant<-er[,c(1,11,12,13,14)] cor(cuant) corrplot(cor(cuant))
Dependencia de variables cuantitativas	cual<-er[,c(1,11,12,13,14)] a<-table(cual[,c(1,2)]); depen(a) a<-table(cual[,c(1,3)]); depen(a)

```

a<-table(cual[,c(1,4)]); depen(a)
a<-table(cual[,c(1,5)]); depen(a)
a<-table(cual[,c(1,6)]); depen(a)
a<-table(cual[,c(1,7)]); depen(a)
a<-table(cual[,c(1,8)]); depen(a)
a<-table(cual[,c(1,9)]); depen(a)
a<-table(cual[,c(2,3)]); depen(a)
a<-table(cual[,c(2,4)]); depen(a)
a<-table(cual[,c(2,5)]); depen(a)
a<-table(cual[,c(2,6)]); depen(a)
a<-table(cual[,c(2,7)]); depen(a)
a<-table(cual[,c(2,8)]); depen(a)
a<-table(cual[,c(2,9)]); depen(a)
a<-table(cual[,c(3,4)]); depen(a)
a<-table(cual[,c(3,5)]); depen(a)
a<-table(cual[,c(3,6)]); depen(a)
a<-table(cual[,c(3,7)]); depen(a)
a<-table(cual[,c(3,8)]); depen(a)
a<-table(cual[,c(3,9)]); depen(a)
a<-table(cual[,c(4,5)]); depen(a)
a<-table(cual[,c(4,6)]); depen(a)
a<-table(cual[,c(4,7)]); depen(a)
a<-table(cual[,c(4,8)]); depen(a)
a<-table(cual[,c(4,9)]); depen(a)
a<-table(cual[,c(5,6)]); depen(a)
a<-table(cual[,c(5,7)]); depen(a)
a<-table(cual[,c(5,8)]); depen(a)
a<-table(cual[,c(5,9)]); depen(a)
a<-table(cual[,c(6,7)]); depen(a)
a<-table(cual[,c(6,8)]); depen(a)
a<-table(cual[,c(6,9)]); depen(a)
a<-table(cual[,c(7,8)]); depen(a)
a<-table(cual[,c(7,9)]); depen(a)
a<-table(cual[,c(8,9)]); depen(a)
depen función propia creada

```

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Regresión Modelo 1

Preparación de datos	<pre> Edade<-datos1\$Edad Razae<-as.factor(datos1\$Raza) TE<-as.factor(datos1\$T) NE<-as.factor(datos1\$N) Me<-as.factor(datos1\$M) Estadioe<-as.factor(datos1\$Estadio) Tipoe<-as.factor(datos1\$Tipo) Partee<-as.factor(datos1\$Parte) Cirugiae<-as.factor(datos1\$Cirugia) Estadoe<-as.factor(datos1\$Estado) Tiempoevolu<-datos1\$`Tiempo evolucion` Pesoe<-datos1\$Peso Estaturae<-datos1\$Estatura Tamañoe<-datos1\$Tamaño datos2<- data.frame(Edad=Edade,Raza=Razae,TE=TE,N=NE,M=Me,Estadio=Estadioe, Tipo=Tipoe,Parte=Partee,Cirugia=Cirugiae,Estado=Estadoe, Tiempo=Tiempoevolu,Tamaño=Tamañoe,Peso=Pesoe, Estatura=Estaturae) </pre>
Planteamiento del modelo 1	<pre> modlogit<-glm(datos2\$Estado~.,data = datos2,family = "binomial") summary(modlogit) sig.var<- summary(modlogit)\$coeff[-1,4] <0.05 </pre>

Variables significativas	<code>names(sig.var)[sig.var == TRUE]</code>
Significancia del modelo	<code>dif_residuos <- modlogit\$null.deviance - modlogit\$deviance</code> <code>df <- modlogit\$df.null - modlogit\$df.residual</code> <code>p_value <- pchisq(q = dif_residuos,df = df, lower.tail = FALSE)</code> <code>predicciones <- ifelse(test = modlogit\$fitted.values > 0.5, yes = 1, no = 0)</code> <code>matriz_confusion <- table(modlogit\$model\$`datos2\$Estado`, predicciones,</code> <code> dnn = c("observaciones", "predicciones"))</code>
Matriz de confusión	<code>mosaic(matriz_confusion, shade = T, colorize = T,</code> <code> gp = gpar(fill = matrix(c("skyblue3", palette("pastel 2"))[1], palette("pastel 2"))[1],</code> <code> "skyblue3"), 2, 2))</code>
Error de precisión	<code>error1 <- sum(matriz_confusion[1,2], matriz_confusion[2,1])/sum(matriz_confusion)</code> <code>n <- dim(datos2)[1]</code>
Curva ROC	<code>pred <- ROCR::prediction(pred1,datos2\$Estado)</code> <code>perf <- performance(pred, "tpr", "fpr")</code> <code>plot(perf)</code>
Precisión ROC	<code>AUCLog1=performance(pred, measure = "auc")@y.values[[1]]</code> <code>cat("AUC: ",AUCLog1,"n")</code> <code>log.odds <- predict(modlogit,</code> <code>data.frame(Edad=48,Raza="Blanca",N=as.factor(3),M=as.factor(1),</code> <code> Estadio=as.factor(2),Tipo="Carcinoma ductal in situ de la</code> <code>mama",</code> <code>Parte="Mama",Tiempo=2.50,Tamaño=5.6,Peso=54.5,Estatura=140,</code> <code> TE=as.factor(4),Cirugia="Si")</code>
Predicciones	<code>plot(df\$x, col = df\$grupo, pch = 16,xlab=" ",ylab="Predicciones",main="Gráfico\n de</code> <code>predicciones")</code> <code>abline(h=0.44,lty=3,col="red")</code> <code>abline(h=0.54,lty=3,col="red")</code>
Probabilidades	<code>erq <- exp(log.odds)/(1+exp(log.odds))</code>
Intervalo de confianza	<code>er <- data.frame(confint.default(modlogit))</code> <code>tu <- data.frame(confint(modlogit,level = 0.95))</code>

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Regresión Modelo 2

Preparación de datos	<code>names(sig.var)[sig.var == TRUE]</code> <code>datos2.2 <-</code> <code>data.frame(T=datos2\$TE,N=datos2\$N,M=datos2\$M,Estadio=datos2\$Estadio,Tiempo=datos</code> <code>2\$Tiempo,</code> <code>Tamaño=datos2\$Tamaño,Peso=datos2\$Peso,Estatura=datos2\$Estatura,Estado=datos2\$Estad</code> <code>o)</code>
Planteamiento modelo 2	<code>modlogit2 <- glm(datos2.2\$Estado~.,data = datos2.2,family = "binomial")</code> <code>summary(modlogit2)</code>
Variables significativas	<code>sig.var2 <- summary(modlogit2)\$coeff[,-1,4] < 0.05</code> <code>names(sig.var2)[sig.var2 == TRUE]</code>
Significancia del modelo	<code>dif_residuos2 <- modlogit2\$null.deviance - modlogit2\$deviance</code> <code>df.2 <- modlogit2\$df.null - modlogit2\$df.residual</code> <code>p_value2 <- pchisq(q = dif_residuos2,df = df.2, lower.tail = FALSE)</code> <code>predicciones2 <- ifelse(test = modlogit2\$fitted.values > 0.5, yes = 1, no = 0)</code> <code>matriz_confusion2 <- table(modlogit2\$model\$`datos2.2\$Estado`, predicciones2,</code> <code> dnn = c("observaciones", "predicciones"))</code>
Matriz de confusión	<code>mosaic(matriz_confusion2, shade = T, colorize = T,</code> <code> gp = gpar(fill = matrix(c("skyblue3", palette("pastel 2"))[1], palette("pastel 2"))[1],</code> <code> "skyblue3"), 2, 2))</code>
Error de precisión	<code>error1 <- sum(matriz_confusion[1,2],</code> <code>matriz_confusion[2,1])/sum(matriz_confusion)</code> <code>n <- dim(datos2)[1]</code>
Curva ROC	<code>pred22 <- ROCR::prediction(pred2,datos2.2\$Estado)</code> <code>perf2 <- performance(pred22, "tpr", "fpr")</code> <code>plot(perf2)</code>
Precisión ROC	<code>AUCLog2=performance(pred22, measure = "auc")@y.values[[1]]</code>

```

cat("AUC: ",AUCLog2,"n")
log.odds2 <- predict(modlogit2, data.frame(N=as.factor(3),M=as.factor(1),
Estadio=as.factor(2),Tiempo=2.50,Tamaño=5.6,Peso=54.5,Estatura=140,
T=as.factor(4)))
Predicciones plot(df2$x, col = df2$grupo, pch = 16,xlab=" ",ylab="Predicciones",main="Gráfico\n de
predicciones")
abline(h=0.44,lty=3,col="red")
abline(h=0.54,lty=3,col="red")
Probabilidades erq2<-exp(log.odds2)/(1+exp(log.odds2))
Intervalo de er2<-data.frame(confint.default(modlogit2))
confianza tu2<-data.frame(confint(modlogit2,level = 0.95))

```

Realizado por: Jhonnatan F, Pinduisaca A. 2022.

Regresión Modelo 3

```

Preparación de datososp3<-as.data.frame(read_excel("base auxiliar.xlsx"))
datos datos3<-datos3[-c(3,5)]
dim(datos3)
names(datos3)
Planteamiento modlogit3<-glm(datos3$Estado~.,data = datos3,family = "binomial")
modelo 3 summary(modlogit3)
Variables sig.var3<- summary(modlogit3)$coeff[-1,4] <0.05
significativas names(sig.var3)[sig.var3 == TRUE]
Significancia del dif_residuos3 <- modlogit3$null.deviance - modlogit3$deviance
modelo df.3 <- modlogit3$df.null - modlogit3$df.residual
p_value3 <- pchisq(q = dif_residuos3,df = df.3, lower.tail = FALSE)
predicciones3 <- ifelse(test = modlogit3$fitted.values > 0.5, yes = 1, no = 0)
matriz_confusion3 <- table(modlogit3$model$datos3$Estado`, predicciones3,
Matriz de dnn = c("observaciones", "predicciones"))
confusión mosaic(matriz_confusion3, shade = T, colorize = T,
gp = gpar(fill = matrix(c("skyblue3", palette("pastel 2")[1], palette("pastel 2")[1],
"skyblue3"), 2, 2)))
Error de error3<- sum(matriz_confusion3[1,2], matriz_confusion3[2,1])/sum(matriz_confusion3)
precisión n<- dim(datos3)[1]
Curva ROC pred33 <- ROCR::prediction(pred3,datos3$Estado)
perf3 <- performance(pred33, "tpr", "fpr")
plot(perf3)
Precisión ROC AUCLog3=performance(pred33, measure = "auc")@y.values[[1]]
cat("AUC: ",AUCLog3,"n")
log.odds3 <- predict(modlogit3, data.frame(N=as.factor(3),M=as.factor(1),
Estadio=as.factor(2),Tiempo=2.50,Tamaño=5.6,Peso=54.5,Estatura=140,
T=as.factor(4)))
Predicciones plot(df3$x, col = df3$grupo, pch = 16,xlab=" ",ylab="Predicciones",main="Gráfico\n de
predicciones")
abline(h=0.44,lty=3,col="red")
abline(h=0.54,lty=3,col="red")
Probabilidades erq3<-exp(log.odds3)/(1+exp(log.odds3))
Intervalo de er3<-data.frame(confint.default(modlogit3))
confianza tu3<-data.frame(confint(modlogit3,level = 0.95))

```

Realizado por: Jhonnatan F, Pinduisaca A. 2022.



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

**DIRECCIÓN DE BIBLIOTECAS Y RECURSOS DEL APRENDIZAJE
UNIDAD DE PROCESOS TÉCNICOS Y ANÁLISIS BIBLIOGRÁFICO Y DOCUMENTAL**

REVISIÓN DE NORMAS TÉCNICAS, RESUMEN Y BIBLIOGRAFÍA

Fecha de entrega: 04 / 01 / 2022

INFORMACIÓN DEL AUTOR/A (S)
Nombres – Apellidos: <i>Jhonnatan Fernando Pinduisaca Allaica</i>
INFORMACIÓN INSTITUCIONAL
Facultad: <i>Ciencias</i>
Carrera: <i>Estadística</i>
Título a optar: <i>Ingeniero Estadístico</i>
f. Analista de Biblioteca responsable: <i>Ing. Leonardo Medina Ñuste MSc.</i>

**LEONARDO
FABIO MEDINA
NUSTE**

Firmado digitalmente por LEONARDO
FABIO MEDINA NUSTE
Nombre de reconocimiento (DN): c=EC,
o=BANCO CENTRAL DEL ECUADOR,
ou=ENTIDAD DE CERTIFICACION DE
INFORMACION-ECIBCE, l=QUITO,
serialNumber=0000621485,
cn=LEONARDO FABIO MEDINA NUSTE
Fecha: 2022.01.04 10:20:03 -05'00'



1861-DBRA-UTP-2021