

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL



**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS
DEPARTAMENTO DE POSTGRADO**

PROYECTO DE TITULACIÓN

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

**“MAGÍSTER EN ESTADÍSTICA CON MENCIÓN EN GESTIÓN
DE LA CALIDAD Y PRODUCTIVIDAD”**

TEMA:

**“DESARROLLO DE UN MODELO PARA ESTIMAR LA
PROBABILIDAD DE FALLECIMIENTO EN PACIENTES
DIAGNOSTICADOS CON ENFERMEDADES ISQUÉMICAS DEL
CORAZÓN Y EGRESADOS DE LOS SERVICIOS DE INTERNACIÓN
HOSPITALARIA A NIVEL NACIONAL”**

AUTOR:

GUSTAVO ROBALINO MUÑIZ

Guayaquil - Ecuador

2021

RESUMEN

Este proyecto se centra en el estudio de las enfermedades isquémicas del corazón en el Ecuador, esta enfermedad es la principal causa de muerte de los ecuatorianos, el estado ecuatoriano asigna importantes recursos económicos, médicos para ser tratada en los centros de salud que pertenecen al ministerio de salud pública.

El objetivo principal es predecir si la condición de egreso de los pacientes es vivo o muerto a causa de una enfermedad isquémica el corazón, para la investigación se utilizó un conjunto de variables: demográficas, sociales de los centros de salud y los pacientes, mediante la utilización de análisis bivariados y la implementación de un modelo de regresión logística multivariado.

La población de estudio son todos los pacientes que fueron atendidos, diagnosticados y dados el alta de alguna enfermedad isquémica del corazón, en todos los centros de salud pertenecientes a la red pública de salud en el año 2018, la cantidad de pacientes con esas características fueron 16094.

Como datos relevantes, se evidenció que los días de estadía que permanecen los pacientes en los centros de salud y la edad, son variables críticas, éstas aumentan la probabilidad de que la condición de egreso de los pacientes sea de fallecida.

Los resultados obtenidos de esta investigación, se espera sea el primer paso para la utilización de variables sociales, demográficas en el estudio de enfermedades como las isquemias de corazón.

Palabras claves:

Regresión logística, isquemias del corazón, variables sociales, Ecuador

ABSTRACT

This project focuses on the study of ischemic heart diseases in Ecuador, this disease is the main cause of death of Ecuadorians, the Ecuadorian state allocates important economic and medical resources to be treated in the health centers that belong to the ministry public health.

The main objective is to predict if the discharge condition of the patients is alive or dead due to an ischemic heart disease, for the research we will use a set of demographic and social variables of the health centers and patients, by using bivariate analysis and implementation of a multivariate logistic regression model.

The study population consists of all patients who were treated, diagnosed and discharged with an ischemic heart disease, in all health centers belonging to the public health network in 2018, the number of patients with these characteristics was 16094.

As relevant data, the data showed that the days of stay that patients remain in health centers and age are critical variables, these increase the probability that the discharge condition of patients is dead.

We hope the results obtained from this research will be the first step for the use of social and demographic variables in the study of diseases such as ischemia of the heart.

Keywords:

Logistic regression, ischemia of the heart, social variables, Ecuador

DEDICATORIA

A mi esposa, mi papá, mis hermanas, pero de manera especial a mi madre, se la debía.

AGRADECIMIENTO

Agradezco a Dios por todo, a mi familia por ser mi apoyo, al personal docente y administrativo de la Maestría por su esfuerzo, a los profesionales de la salud por su colaboración y a mi director del proyecto por sus conocimientos.

DECLARACIÓN EXPRESA

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

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




Gustavo David Robalino Muñiz

TRIBUNAL DE GRADUACIÓN



Mgr. Francisco Moreira Villegas
PRESIDENTE



Mgr. Mario Solórzano Carvajal
DIRECTOR



Mgr. Heydi Roa López
VOCAL 1



Ph.D. Sandra García Bustos
VOCAL 2

ABREVIATURAS O SIGLAS

INEC: Instituto Nacional Ecuatoriano de Censos.

EIC: Enfermedades Isquémicas del Corazón.

OMS: Organización Mundial de la Salud.

OPS: Organización Panamericana de la Salud.

ECU: Enfermedades Cardiovasculares.

RPIS: Red Publica Integral de Salud.

RC: Red Complementaria.

SpO2-I (2%): Índice de hipoxia

IMC: Índice de masa corporal.

PR-I: Índice de aceleración de la frecuencia del pulso.

CIE-10: Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud Décima Revisión.

CGE: Clasificador Geográfico Estadístico.

TABLA DE CONTENIDO

CAPÍTULO 1	1
1. INTRODUCCIÓN.....	1
1.1. Antecedentes.....	1
1.2. Descripción del problema	4
1.3. Objetivos.....	5
1.3.1 Objetivo General	5
1.3.2 Objetivos Específicos	5
1.4. Hipótesis.....	5
1.5. Alcance	6
CAPÍTULO 2	8
2. MARCO TEÓRICO.....	8
2.1. Marco Teórico	8
2.1.1. Prueba de razón de varianzas	8
2.1.2. Prueba t para muestras independientes.....	8
2.1.3. Tablas de Contingencia y prueba Ji – cuadrado.....	10
2.1.4. Regresión Logística	10
2.2. Estado del Arte	12
CAPÍTULO 3	15
3. METODOLOGÍA.....	15
3.1. Datos.....	16
3.2. Principales estadísticas descriptivas, de pacientes diagnosticados con isquemias del corazón del Ecuador en el 2018.	20
3.2.3. Variable: Sector.....	21
3.2.4. Variable: Sexo.....	21
3.2.5. Variable: Edad.....	22
3.2.6. Variable: Etnia.....	23
3.2.7. Variable: Provincia de residencia.....	23
3.2.8. Variable: Área de residencia.....	24
3.2.9. Variable: Días de estadía.	24
3.2.10. Variable: Región de Residencia	25
3.2.11. Variable: Condición de egreso.	26
CAPÍTULO 4	27
4. RESULTADOS.....	27
4.1. Introducción	27
4.2. Asociaciones Bivariantes	27

4.2.1. Condición de egreso vs Provincia de ubicación.....	27
4.2.2. Condición de egreso vs Área de ubicación.....	29
4.2.3. Condición de egreso vs Sector.....	29
4.2.4. Condición de egreso vs Sexo.....	30
4.2.5. Condición de egreso vs Edad.	30
4.2.6. Condición de egreso vs Etnia.	31
4.2.7. Condición de egreso vs Provincia de residencia.	32
4.2.8. Condición de egreso vs Área de residencia.....	33
4.2.9. Condición de egreso vs Días de estadía.	34
4.2.10. Condición de egreso vs región de residencia.	35
4.3. Regresión Logística	35
4.3.1. Selección de variables predictoras	36
4.3.2. Preparación de los datos.....	36
4.3.3. Modelos Logísticos	37
4.3.4. Resultados del modelo.....	39
4.3.5. Ajuste del modelo	42
4.3.6. Ejemplos prácticos del modelo	45
CAPÍTULO 5	47
5. CONCLUSIONES Y RECOMENDACIONES.....	47
5.1. Conclusiones.....	47
5.2. Recomendaciones.....	48
6. Referencias	50
7. Apéndices y anexos	52

LISTADO DE FIGURAS

<i>Figura 3.1 Número de egresos hospitalarios</i>	16
<i>Figura 3.2 Establecimientos de salud</i>	16
<i>Figura 3.3 Causas principales de la muerte general 2018</i>	17
<i>Figura 3.4 Establecimientos de salud</i>	22
<i>Figura 3.5 Histograma de la variable días de estadía</i>	25
<i>Figura 4.1 Gráfico ggpairs variables predictoras vs condición de egreso</i>	36
<i>Figura 4.2 Balanceo de clases en conjuntos de datos</i>	37
<i>Figura 4.3 Gráfico de violín del modelo</i>	43
<i>Figura 4.4 Gráfico del punto de corte</i>	44
<i>Figura 4.5 Gráfico curva AUC</i>	45

LISTADO DE TABLAS

<i>Tabla 2.1 Coeficientes modelo logística Universidad de Chile</i>	13
<i>Tabla 2.2 Comparación de modelos estadístico Universidad de la Costa</i>	14
<i>Tabla 2.3 Comparación de modelos estadístico Universidad de la Costa</i>	14
<i>Tabla 3. 1 Descripción de la variable dependiente</i>	17
<i>Tabla 3. 2 Descripción de las variables predictoras</i>	18
<i>Tabla 3. 3 Estadísticas descriptivas: variable edad</i>	23
<i>Tabla 3. 4 Estadísticas descriptivas: variable días de estadía</i>	25
<i>Tabla 4. 1 Tabla de Contingencia. Condición de egreso vs. Provincia de ubicación</i>	27
<i>Tabla 4. 2 Tabla de contingencia. Condición de egreso vs. Área de Ubicación</i>	29
<i>Tabla 4. 3 Tabla de contingencia. Condición de egreso vs. Sector</i>	29
<i>Tabla 4. 4 Tabla de contingencia. Condición de egreso vs. Sexo</i>	30
<i>Tabla 4. 5 Prueba de Levene. Condición de egreso vs. Edad</i>	31
<i>Tabla 4. 6 Prueba t de student. Variables Mortalidad vs. Edad</i>	31
<i>Tabla 4. 7 Tabla de contingencia. Condición de egreso vs Etnia</i>	31
<i>Tabla 4. 8 Tabla de contingencia. Condición de egreso vs. Provincia de residencia</i>	32
<i>Tabla 4. 9 Tabla de contingencia. Condición de egreso vs. Área de residencia</i>	34
<i>Tabla 4. 10 Prueba de Levene. Condición de egreso vs. Días de estadía</i>	34
<i>Tabla 4. 11 Prueba t de student. Condición de egreso vs. Días de estadía</i>	34
<i>Tabla 4. 12 Tabla de contingencia. Condición de egreso vs. Región de residencia</i>	35
<i>Tabla 4. 13 Modelos logísticos</i>	37
<i>Tabla 4. 14 Codificación de variables del modelo regresión logística</i>	38
<i>Tabla 4. 15 Coeficientes B estimados regresión logit</i>	39
<i>Tabla 4. 16 Oportunidades relativas</i>	40
<i>Tabla 4. 17 Medidas de bondad de ajuste del modelo</i>	43
<i>Tabla 4. 18 Tabla de clasificación</i>	44
<i>Tabla 4. 19 Ejemplos del modelo</i>	45

LISTADO DE CUADROS

Cuadro 3. 1 Estadísticas descriptivas: variable provincia de ubicación	20
Cuadro 3. 2 Estadísticas descriptivas: variable área de ubicación	21
Cuadro 3. 3 Estadísticas descriptivas: variable sector	21
Cuadro 3. 4 Estadísticas descriptivas: variable sexo	22
Cuadro 3. 5 Estadísticas descriptivas: variable etnia	23
Cuadro 3. 6 Estadísticas descriptivas: variable provincia de residencia	24
Cuadro 3. 7 Estadísticas descriptivas: variable área de residencia	24
Cuadro 3. 8 Estadísticas descriptivas: variable región de residencia	26
Cuadro 3. 9 Estadísticas descriptivas: variable mortalidad	26

CAPÍTULO 1

1. INTRODUCCIÓN

1.1. Antecedentes

En la actualidad las enfermedades isquémicas del corazón (EIC) lideran la primera causa de mortalidad de forma universal, cobran la vida de 17,7 millones de personas cada año, el 31% de todas las muertes mundiales (**OMS, 2017**), para el 2030, la Organización Mundial de la Salud deduce que se incrementará el número de defunciones situándolo en 23,6 millones de personas que morirán por alguna de las enfermedades cardiovasculares (**OPS, 2013**). Todo esto evidencia que las enfermedades isquémicas constituyen una seria complicación para salud mundial y mucho más para los países subdesarrollados continúa en ascenso hasta la presente (**Ramírez, 2010**), por lo que se han convertido en un reto el disminuir estas cifras alarmantes y el impacto en el sector hospitalario, buscando mecanismos que permitan tomar decisiones y a la vez la mitigación de la progresión de las enfermedades isquémicas.

Es necesario mencionar que se ha producido un desplazamiento de la carga de mortalidad de los países de ingresos altos a aquellos de ingresos medios y bajos tomando en consideración América Latina debido a:

- ✓ Una combinación de factores demográficos (como un rápido crecimiento poblacional y un proceso de envejecimiento);
- ✓ Procesos de urbanización;
- ✓ La globalización de comportamientos de riesgo cardiovascular que ha traído como consecuencia tendencias epidemiológicas divergentes. (INFOMED, 2020)

En América Latina y el Caribe las enfermedades isquémicas también son la primera causa de muerte, el 80 % de estas muertes ocurre en países de bajos y medianos ingresos, donde las personas desarrollan sus enfermedades a más corta edad, sobrellevan por más largo tiempo y fallecen más temprano que en los países de altos ingresos. Generando altos gastos destinados a la salud representando altos

costos en la economía de cada país y sobrecargando los servicios de salud, países como: México, Brasil, Argentina y Colombia ya presentan afectaciones económicas como resultado de dichas enfermedades cardiovasculares. (ESTEBAN, 2017).

En el transcurso de estas últimas décadas al incrementarse los gastos de salud se pudo evidenciar que el diseño tradicional del sistema de salud ha sido sobrepasado y más aún con el manejo inadecuado de las enfermedades isquémicas, es por eso sumamente importante un sistema de salud que tenga respuestas coherentes y sostenidas a lo actual, es necesario mencionar lo que plantea la OMS, teniendo como objetivo principal asegurar que los sistemas de salud se encuentren equipados y preparados para ofrecer cuidado continuo y de alta calidad para la sociedad, indicando claramente la importancia en la atención de las enfermedades cardiovasculares (ECV) (JoAnne Epping-Jordan, 2002).

Con respecto a Ecuador, las EIC han ocupado por varios años los primeros lugares dentro de las principales causas de defunción, es así que para el año 2016, las EIC fueron la primera causa de mortalidad en hombres y mujeres concentrando el 10,15% (n = 3.747) y 9,04% (n = 2.766) de las defunciones, respectivamente (INEC, INEC, 2018). En los últimos años, diversas estrategias se han planteado para reducir la alta mortalidad por las EIC y lograr el objetivo de disminuir en 25% el número de muertes prematuras asociadas a las enfermedades crónicas no transmisibles, propuesto por la OMS y Organización Panamericana de la Salud (OPS) (Salud, 2014) .

Para el 2018, según el Registro Estadístico de Camas y Egresos Hospitalarios del Ecuador, los centros de salud que forman parte de la Red Pública Integral de Salud (RPIS) y Red Complementaria (RC) en el país se registraron 1.164.660, de los cuales 1.145.302 egresaron vivos y 11.358 fallecieron, las personas que fueron diagnosticadas con enfermedades isquémicas del corazón y relacionadas fueron 16.144 de las cuales permanecen con vida 14.999 (hombres 9215- mujeres 5784) y fenecieron 1150 (hombres 632-mujeres 518) (INEC, Registro Estadístico de Camas y Egresos Hospitalarios del Ecuador, 2018).

Las enfermedades isquémicas del corazón, también son la primera causa de mortalidad según los datos del Registro Estadístico de Defunciones Generales emitido por el Instituto Nacional de Estadística y Censos de su estudio del 2018, en el cual se registra un total de muertes por dicha causa fue 7.862 personas, también se obtuvieron estimaciones con grupos específicos atribuyendo 4.479 a mortalidad masculina que representa el 11,47 % y 3.381 a mortalidad femenina que representa el 10,59 % del total de fallecimientos. (INEC, Registro Estadístico de Defunciones Generales, 2018)

Las enfermedades isquémicas, a más de ocupar el primer lugar como causa de mortalidad también representan una carga y factores de riesgo altos para los usuarios de la red hospitalaria. Gracias a las políticas de gobierno se destina un 2,5 % del presupuesto general de salud para enfermedades cardíacas, este estudio busca encontrar los grupos de riesgo en sufrirlas utilizando información personal, demográfica, receptada por los centros de salud, buscando mejorar la gestión de los recursos, tomando en cuenta que es la primera causa de muerte a nivel nacional. (Federation, 2016)

En Ecuador esta problemática radica en poder estimar la probabilidad de fallecimiento en pacientes con diagnósticos isquémicos, mostrando evidencia que reflejan el crecimiento significativo en cuanto a la tendencia de la salud pudiendo verse marcados periodos y grupos etarios incluidos en este estudio. Esta información alerta a la población ecuatoriana debido a que es necesario crear hábitos para disminuir el riesgo de mortalidad.

Esta investigación es importante y necesaria porque sirve para resaltar aquellas variables que aumentan o reducen las probabilidades de las enfermedades isquémicas que afectan a una gran parte de la población ecuatoriana.

Debido a esta problemática de las enfermedades isquémicas se planteó como objetivo general predecir la probabilidad de fallecimiento en pacientes diagnosticados con enfermedades isquémicas del corazón y egresos de los servicios de internación hospitalaria a nivel nacional, como indicador de fatalidad, para el establecimiento de acciones preventivas en la comunidad ecuatoriana.

1.2. Descripción del problema

La cardiopatía isquémica es la afección miocárdica producida por una desproporción entre el aporte del flujo coronario y los requerimientos miocárdicos, causado por cambios en la circulación coronaria. Comprende estadios agudos, temporales o crónicos por cambios funcionales o enfermedad orgánica, se excluye la isquemia por cambios hemodinámicas no coronarios, como sucede en la estenosis aórtica **(Jadraque, 1999)**. Se calcula que la incidencia de enfermedad coronaria en edades de entre 65 y 94 años se duplica en los varones y se triplica en las mujeres respecto a edades entre 35 y 64 años **(Barcelona, 2021)**.

Las causas de riesgo cardiovascular están ligadas con las características personales, condiciones de vida o dificultades de salud de cada persona que pueden deteriorar las arterias y causar arteriosclerosis. Mucho de estos componentes no se pueden modificar, ya sea la edad o el sexo, pero se debe prestar suma atención a la prevención y control del resto.

A finales del siglo XX los cambios en las formas vida fueron punto importante para la exposición masiva a las enfermedades cardiacas, llegó a sus más altos picos al finalizar este siglo, en los países desarrollados comenzó a descender mientras que en los subdesarrollados aún está en ascenso en el presente siglo. Todo esto abrió camino a las investigaciones permitiendo conocer los procesos complejos abarcando la genética de la sociedad, permitiendo diseñar tácticas para limitar los daños y así implementar estrategias preventivas para toda la población **(González Guzmán & Alcalá Ramírez, 2010)**.

Las EIC son enfermedades muy frecuentes las cuales afectan a una gran parte de la población de los países desarrollados y en mayor medida a los países en vía de desarrollo, como es el caso de Ecuador. Las EIC ocupan el primer lugar como causa de mortalidad general durante el periodo 2018 **(INEC, Registro Estadístico de Defunciones Generales, 2018)**

1.3. Objetivos

1.3.1 Objetivo General

Estimar la probabilidad de fallecimiento en pacientes diagnosticados con enfermedades isquémicas del corazón y egresos de los servicios de internación hospitalaria a nivel nacional, como indicador de fatalidad, para el establecimiento de acciones preventivas en la comunidad ecuatoriana.

1.3.2 Objetivos Específicos

- Determinar si las características de los centros de salud y del paciente inciden en el incremento o disminución de la probabilidad de fallecimiento en pacientes diagnosticados con enfermedades isquémicas del corazón.
- Establecer las variables que aumentan o reducen la probabilidad de fatalidad de los pacientes diagnosticados y egresados con enfermedades isquémicas del corazón en los centros de salud priorizando los grupos de riesgo.
- Desarrollar la función de mortalidad en cada categoría de las variables utilizadas en los centros de salud y características del paciente para estimar la probabilidad de fallecimiento.

1.4. Hipótesis

1. El sexo de los pacientes diagnosticados, tratados y dados el alta del sistema de salud nacional del Ecuador, inciden en la mortalidad a causa de una enfermedad isquémica del corazón.
2. La edad los pacientes diagnosticados, tratados y dados el alta del sistema de salud nacional del Ecuador, están relacionados en la mortalidad a causa de una enfermedad isquémica del corazón.

3. Los hospitales del sector público donde fueron diagnosticados, tratados y dados el alta, inciden en la mortalidad a causa de una enfermedad isquémica del corazón.
4. La cantidad de días que estuvieron hospitalizados los pacientes que fueron diagnosticados, tratados y dados el alta de un centro de salud del Ecuador, contribuyen en la mortalidad a causa de una enfermedad isquémica del corazón.

1.5. Alcance

Esta investigación busca obtener un modelo que permita calcular la probabilidad de fallecimiento en pacientes diagnosticados con enfermedades isquémicas del corazón y egresaron en los servicios de internación hospitalaria a nivel nacional desde el primero de enero del 2018 hasta el 31 de diciembre del mismo año. La población de estudio es a nivel nacional y los centros de salud forman parte de la Red Pública Integral de Salud (RPIS) y Red Complementaria (RC).

CAPÍTULO 2

2. MARCO TEÓRICO

2.1. Marco Teórico

Para la elaboración de este estudio se utilizaron varias técnicas estadísticas, conceptos teóricos y demás conceptos necesarios para el soporte de los resultados obtenidos, a continuación se presentan de forma breve:

2.1.1. Prueba de razón de varianzas

Es una prueba estadística inferencial utilizada para evaluar la igualdad de las varianzas para una variable calculada para dos o más grupos. La H_0 (hipótesis nula) representa la afirmación de que no existe asociación entre las dos variables estudiadas y la H_a (hipótesis alternativa) afirma que hay algún grado de relación o asociación entre las dos variables. (Rubio Hurtado & Berlanga Silvente, 2012).

El estadístico utilizado para esta prueba y para contrastar las hipótesis es F de Snedecor.

$$F = \frac{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}{\frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2} = \frac{S_1^2}{S_2^2}$$

2.1.2. Prueba t para muestras independientes

El procedimiento de la prueba utiliza las medias de los grupos de estudio a contrastar, y verifica si la media de una población es diferente estadísticamente a la media de otra población. Existen dos formas para calcular el estadístico de esta prueba, que dependen de si se asumen o no varianzas iguales (Zurita Herrera, 2008).

Cuando dos muestras independientes son extraídas de poblaciones que se asumen tienen varianzas idénticas, el estadístico t se calcula de la siguiente manera:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Con

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Identificando:

- \bar{x}_1 = Media de la primera muestra
- \bar{x}_2 = Media de la segunda muestra
- n_1 = Tamaño de la primera muestra
- n_2 = Tamaño de la segunda muestra
- s_1 = Desviación estándar de la primera muestra
- s_2 = Desviación estándar de la segunda muestra
- s_p = Desviación estándar agrupada

Al tener dos muestras independientes que son extraídas de poblaciones que se asumen tienen varianzas desiguales, el estadístico t se calcula de la siguiente manera:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Identificando:

- \bar{x}_1 = Media de la primera muestra
- \bar{x}_2 = Media de la segunda muestra
- n_1 = Tamaño de la primera muestra
- n_2 = Tamaño de la segunda muestra
- s_1 = Desviación estándar de la primera muestra
- s_2 = Desviación estándar de la segunda muestra

2.1.3. Tablas de Contingencia y prueba Ji – cuadrado

Las tablas de contingencia son matrices rectangulares que permiten establecer si existe alguna relación entre dos variables diferentes de una población, estas variables o características a su vez se encuentran divididas en categorías o criterios.

La tabla se compone de las frecuencias relativas observadas para cada variable con sus correspondientes categorías. La hipótesis nula (H_0) no existe ninguna asociación entre los dos atributos en la población, si H_0 es rechazada indicaría que existe asociación entre las variables analizadas. (Paul Newbold, 2008).

El estadístico de prueba Ji-cuadrado, compara las frecuencias observadas con las esperadas, para un nivel de significación α la hipótesis H_0 no se rechaza si cumple:

$$\sum_{i=1}^k \sum_{j=1}^m \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \leq \chi_{\alpha, (k-1)(m-1)}^2$$

Se rechaza en caso contrario.

2.1.4. Regresión Logística

La regresión logística es una técnica multivariada que tiene como objetivo analizar una determinada variable dicotómica, que toma el valor de “0” si el evento no ocurre y “1” en caso de que sí, utilizando conjunto de variables que se presumen tienen una incidencia sobre la variable dependiente.

Esta técnica también se utiliza para predecir la frecuencia de un determinado evento en función de un conjunto de causas que se presumen relevantes sobre la variable dependiente, y determinar cómo influyen cada una de las variables explicativas en el aumento o disminución de la probabilidad.

Las variables explicativas pueden ser de tipo cuantitativas o cualitativas, para el caso de las covariables categóricas es necesario que sean transformadas en variables indicadoras “dummy”, que son aquellas que toman el valor de 1 o 0 para

indicar la presencia o ausencia de un evento categórico que tal vez pueda provocar un efecto en el resultado (Everitt & Dunn, 2001).

El modelo de regresión logística se puede representar de la siguiente forma:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

Identificando:

$p_i =$ probabilidad de que ocurra el evento de interés.
 $1 - p_i =$ probabilidad de que no ocurra el evento de interés.

A través de los coeficientes de regresión β del modelo es posible obtener los *odds ratio*, los cuales reflejan el efecto del aumento o disminución de la variable independiente sobre la variable de estudio, es decir permiten cuantificar la magnitud entre la respuesta y el factor de interés.

La estimación de los coeficientes del modelo se realiza a través del cálculo de estimaciones de máxima verosimilitud, ejecutados a través de algoritmos iterativos como el de Newton Raphson, por lo que es necesario la aplicación de programas estadísticos para su ejecución. El contraste de los coeficientes se lo realiza a través de la prueba de Wald, cuya hipótesis a probar es que un coeficiente aislado es distinto de 0, lo cual indicaría que el modelo es adecuado para representar una determinada relación.

La regresión logística al igual que otros modelos de regresión, compara los valores observados de la variable de respuesta con los variables predichos obtenidos del modelo, en dos momentos: el primero en el modelo inicial, es decir considerando solo la constante, y el segundo al incluir una o más variables; esta comparación se realiza a través de la función de máxima verosimilitud, basada en la siguiente expresión (Hosmer, Lemeshow, & ed., 2000).

$$D = -2 \ln \left[\frac{(\text{probabilidad del modelo ajustado})}{(\text{probabilidad del modelo saturado})} \right]$$

El estadístico D es llamado desviación, y en el caso de la regresión logística realiza la función de la suma de cuadrados residual en la regresión lineal (Hosmer, Lemeshow, & ed., 2000).

Los modelos de regresión logística se ajustan utilizando el método de máxima verosimilitud, es decir, los parámetros estimados son aquellos valores que maximizan la probabilidad de que los resultados se hayan observado (McKenna & Smith, 2013). El R^2 de McFadden, es análogo al R^2 de la regresión lineal, y se define la siguiente manera:

$$R_{McFadden}^2 = 1 - \frac{\log(L_c)}{\log(L_n)}$$

Donde L_c es la función de verosimilitud del modelo completo, es decir con todas las variables explicativas, y L_n es la función de verosimilitud del modelo nulo (el modelo solo con el intercepto, sin covariables).

Otra forma de evaluar el modelo es a través de la prueba de Hosmer y Lemeshow, la cual evalúa la predicción real de la variable dependiente, dividiendo los casos en clases y comparando las frecuencias esperadas con las observadas en cada una de ellas (Hosmer, Lemeshow, & ed., 2000). El estadístico se define como:

$$c = \sum_{i=1}^g \frac{(O_i - n_i \bar{p}_i)^2}{n_i \bar{p}_i (1 - \bar{p}_i)}$$

Identificando:

- $g =$ es el número de grupos
- $n_i =$ es el número de grupos
- $O_i =$ es el número de observaciones en el i-ésimo grupo
- $\bar{p}_i =$ es el promedio de las p_i en el i-ésimo grupo

2.2. Estado del Arte

A nivel mundial existen investigaciones relacionadas a la mortalidad producida por enfermedades isquémicas del corazón, donde utilizan modelos de regresión logística para calcular la probabilidad de fatalidad por estas enfermedades.

Estudiantes de la Facultad de Ciencias Físicas y Matemáticas del Departamento de Ingeniería Industrial e Ingeniería en Computación de la Universidad de Chile, realizaron un estudio sobre: Selección, implementación y Validación de modelos de predicción de riesgos para enfermedades cardiovasculares en pacientes con problema de sueño. El modelo que se decide utilizar como referencia para la creación del módulo funcional y la evaluación económica y social es la regresión logística que incluye variables tradicionales y también variables derivadas de la oximetría, con el índice SPO2-I al 2 % y con la variable ProportionBelow90, el modelo predice si se presentará o no una ECV en los siguientes 10 años y este análisis es parte de su investigación. (Chile, 2017).

Los coeficientes del modelo son:

Tabla 2.1 Coeficientes modelo regresión logística Universidad de Chile

Intercepto	-2.189
ProportionBelow90	0.053
SpO2-I (2%)	-0.034
PR-I	0.009
Sexo	0.571
Log(edad)	1.039
Presión sistólica	-0.026
IMC	0.189
Tratamiento Presión	0.505
Fumador	0.799
Diabético	0.612

Fuente: Chile - 2017
Elaboración: Autor

Se aprecia con claridad que el peso de las variables derivadas de la oximetría es efectivamente menor al de otras variables. Las variables que más peso tienen en el modelo son la edad, seguida por si el paciente es o no fumador.

En Colombia, la Universidad de la Costa, en la ciudad de Barrquilla, realizó un estudio en el diagnóstico de enfermedades cardiovasculares, los datos fueron obtenidos del repositorio de la Universidad de Cleveland en Estados Unidos (2013). Utilizaron la herramienta para minería de datos Weka, para adelantar el entrenamiento y prueba de los clasificadores (árboles de decisión, máquinas de soporte vectorial, regresión logística, naivebayes, IBk y redes neuronales).

Podemos apreciar los siguientes gráficos, apreciamos los resultados de las diferentes técnicas estadísticas, se realizan estudios de pacientes con ECV menor al 50% y pacientes con ECV mayor al 50%.

Tabla 2.2 Comparación de modelos estadístico Universidad de la Costa (ECV menor 50%)

Diagnóstico de Enfermedad Cardíaca (menor 50%)				
	TP Rate	FP Rate	Precision	Recall
J48	0.805	0.281	0.772	0.805
SMO	0.86	0.216	0.825	0.86
SIMPLELOGISTIC	0.878	0.216	0.828	0.8798
NAIVEBAYES	0.872	0.209	0.831	0.872
IBk	0.78	0.266	0.776	0.78
MULTILAYER	0.902	0.899	0.542	0,902

*Fuente: Universidad de la Costa - 2015
Elaboración: Autor*

Tabla 2.3 Comparación de modelos estadístico Universidad de la Costa (ECV mayor 50%)

Diagnóstico de Enfermedad Cardíaca (mayor 50%)				
	TP Rate	FP Rate	Precision	Recall
J48	0.79	0.195	0.758	0.719
SMO	0.784	0.14	0.826	0.784
SIMPLELOGISTIC	0.784	0.112	0.845	0.784
NAIVEBAYES	0.791	0.128	0.84	0.791
IBk	0.734	0.22	0.739	0.734
MULTILAYER	0.101	0.098	0.467	0,101

*Fuente: Universidad de la Costa - 2015
Elaboración: Autor*

La técnica que logró mejores resultados en pacientes con ECV menor al 50% fue Regresión Logística con 83% de precisión y en pacientes con ECV mayor al 50% fueron Regresión Logística y NaiveBayes las cuales alcanzaron un 84% de precisión (Costa, 2015).

A nivel nacional la información es muy escasa, hay un estudio realizado por la Universidad Tecnológica Equinoccial, estudian la mortalidad por enfermedades isquémicas del corazón en el periodo del 2001 al 2016 mediante el uso de series temporales. (Clínica, 2018).

CAPÍTULO 3

3. METODOLOGÍA

La metodología utilizada para el desarrollo de esta investigación comprende inicialmente la compilación y limpieza de los datos disponibles de todos los pacientes diagnosticados y tratados con problemas isquémicos del corazón, según el CIE -10 se clasifica las enfermedades isquémicas entre las categorías I20-I25 **(OMS, 2018)** así como el análisis descriptivo de los resultados más importantes.

Cabe señalar en investigaciones realizadas en el pasado muestran problemas de sub-registro en datos sobre la mortalidad en el Ecuador en el año 2012, también existe evidencia en el registro de datos de defunciones en el período 1999-2010 una no correcta declaración de la edad de los pacientes, provocando que el registro de defunciones del sector rural sea subestimado porque muchos pacientes de este sector fallecen en centros urbanos. **(Vera, 2018)**.

En vista de los problemas de sub-registro se realizó la elaboración de tablas de contingencia para determinar si existe dependencia entre la mortalidad producida por enfermedades isquémicas del corazón y las variables relacionadas con los pacientes y los centros de salud donde fueron atendidos, diagnosticados y dados de alta, esto permitirá seleccionar del conjunto inicial de variables, aquellas que estén asociadas con la mortalidad de los mismos.

Posteriormente, a través de un modelo lineal generalizado de regresión logística se explica la variable dependiente binaria que indica si el paciente falleció a causa de enfermedad isquémica del corazón, y se determinan factores que reducen (paciente, hospital) y aumentan (paciente, hospital) la probabilidad de fallecer.

La mortalidad causada por enfermedades isquémicas del corazón, la estudiamos mediante el uso de variables sociales, demográficas.

La variable de respuesta a lo largo de este estudio es egreso hospitalario cuya definición es el retiro de paciente de los servicios de internación del hospital. Un egreso implica el término de la conclusión del período de hospitalización y la desocupación de una cama de hospital, ya sea por alta o fallecimiento. **(INEC, BOLETIN TECNICO No - 01 -2019 - ECEH, 2019)**

3.1. Datos

Para la realización de este estudio se consideraron datos históricos de los pacientes diagnosticados y con su respectivo egreso en el 2018 **(INEC, Registro Estadístico de Camas y Egresos Hospitalarios del Ecuador, 2018)** el indicador fue la cantidad de egresos hospitalarios en el año 2018 fueron 807.245.

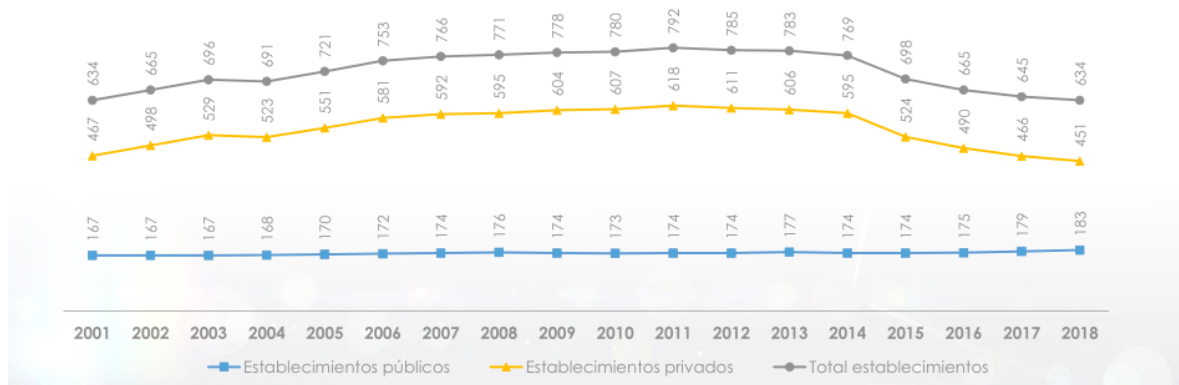
Figura 3.1 Número de egresos hospitalarios



Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: INEC

En el año 2018 existen 634 establecimientos de salud funcionando y reportaron información de sus actividades, de los cuales, 183 son públicos y 451 son privados.

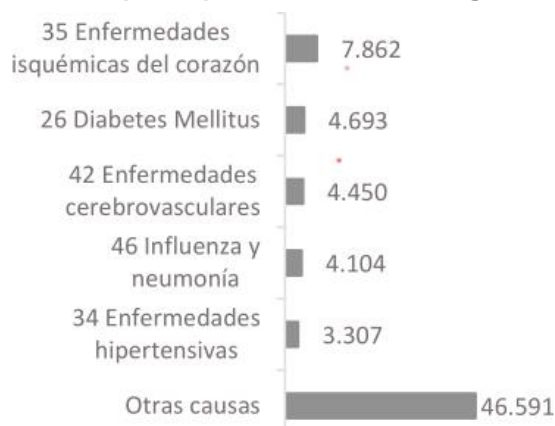
Figura 3. 2 Establecimientos de salud



Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: INEC

En el año 2018, las enfermedades isquémicas del corazón fueron la principal causa de muerte en hombres y mujeres con 7862 defunciones.

Figura 3. 3 Causas principales de la muerte general 2018



Fuente: Registro Estadístico de defunciones generales 2018 - INEC
Elaboración: INEC

De la fuente del INEN, utilizamos la variable: *causa221x* es una lista que se genera a partir de la agrupación de categorías y subcategorías establecidas en la CIE-10, utilizamos esta variable para encontrar a los pacientes diagnosticados y tratados con enfermedades isquémicas del corazón (**INEC, Boletín N°-01-2019-Registro Estadístico de Camas y Egresos Hospitalarios, 2018**). El total, de pacientes diagnosticados con isquemias del corazón y que fueron dados de alta (con egreso) vivos o muertos sin incluir neonatos y no podemos afirmar si en algunos casos el paciente fue tratado más de una vez, en el 2018 fueron 16.094.

La variable dependiente de este estudio es condición de egreso, descrita a continuación en la tabla 3.1.

Tabla 3. 1 Descripción de la variable dependiente

Código de la variable	Nombre de la variable	Definición de la variable	Formato de la variable	Categorías de la variable
con_egreso	Condición de egreso	Condición de egreso del paciente	numérica	1: muerto 0: vivo

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor

Las covariables a utilizar se describen en la tabla 3.2.

Tabla 3. 2 Descripción de las variables predictoras

Características centro de salud					
Código de la variable	Nombre de la variable	Definición de la variable	Formato de la variable	Categorías de la variable	Abreviatura
provi_ubi	Provincia de Ubicación	Provincia de Ubicación del establecimiento	cadena	1. AZUAY 2. BOLIVAR 3. CAÑAR 4. CARCHI 5. COTOPAXI 6.CHIMBORAZO 7.EL ORO 8.ESMERALDAS 9.GUAYAS 10.IMBABURA 11.LOJA 12.LOS RIOS 13.MANABI 14.MORONA SANTIAGO 15.NAPO 16.PASTAZA 17.PICHINCHA 18.TUNGURAHUA 19.ZAMORA CHINCHIPE 20.GALAPAGOS 21.SUCUMBIOS 22.ORELLANA 23.SANTO DOMINGO DE LOS TSCHILAS 24.SANTA ELENA	Azu Bol Cañ Car Coto Chi Oro Esm Gua Imb Loja Rios Man MorS Napo Pas Pich Tung Zam Galp Suc Orell StoD StaE
area_ubi	Área de ubicación	Área de ubicación del establecimiento	numérica	1: urbana 2: rural	urbana rural
sector	Sector	Sector al que pertenece el establecimiento	numérica	1: Privada con fines de lucro 2: Privada sin fines de lucro 3: Pública	PrivL PrivSL Pública
sexo	sexo	Sexo del paciente	numérica	1 : Masculino 0 : Femenino	M F
edad	edad	Edad del paciente	numérica	[1,104] años	
etnia	Etnia	Definición Étnica del paciente	numérica	1.indigena 2.afrodescendiente 3.negro 4.mulato 5.montubio 6.mestizo	Ind Afro Neg Mul Mon Mes

				7.blanco 8.otro 9.ignorado	Blan Otro Igno
prov_res	Provincia de residencia	Provincia residencia habitual del paciente	Cadena	1. AZUAY 2. BOLIVAR 3. CAÑAR 4. CARCHI 5. COTOPAXI 6.CHIMBORAZO 7.EL ORO 8.ESMERALDAS 9.GUAYAS 10.IMBABURA 11.LOJA 12.LOS RIOS 13.MANABI 14.MORONA SANTIAGO 15.NAPO 16.PASTAZA 17.PICHINCHA 18.TUNGURAHUA 19.ZAMORA CHINCHIPE 20.GALAPAGOS 21.SUCUMBIOS 22.ORELLANA 23.SANTO DOMINGO DE LOS TSCHILAS 24.SANTA ELENA	Azu Bol Cañ Car Coto Chi Oro Esm Gua Imb Loja Rios Man MorS Napo Pas Pich Tung Zam Galp Suc Orell StoD StaE
area_res	Área Residencia	Área de residencia del paciente	numérica	1: urbana 2: rural	
dia_estad	días de estadía	días de estadía en centro de salud	numérica	[1-104] años	
Reg_res	Región de residencia	Región de residencia del paciente	numérica	1: Costa 2: Insular 3 : Oriente 4: Sierra	

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor

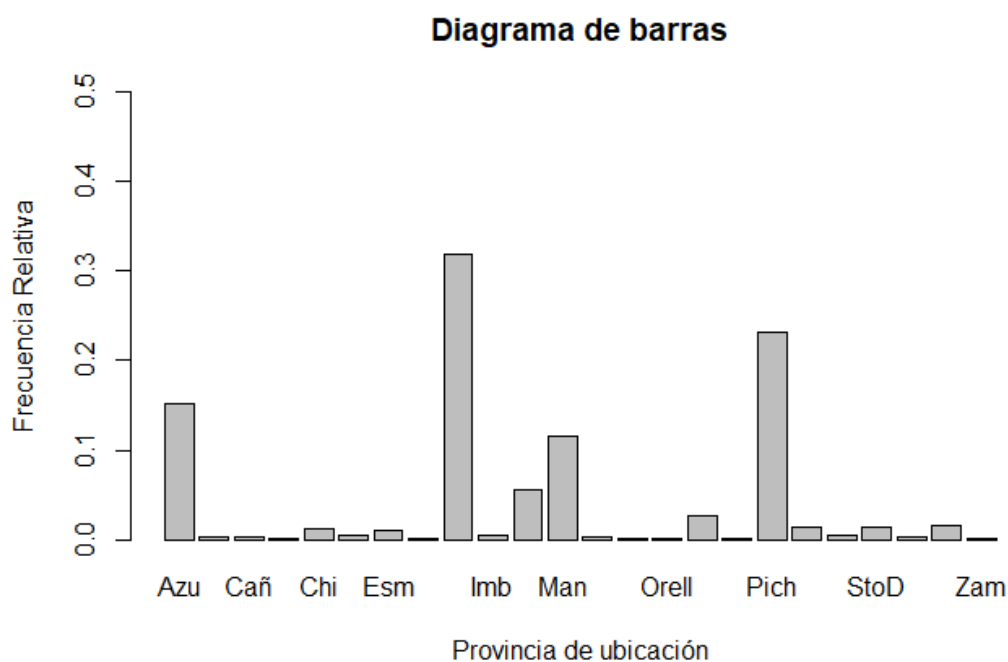
3.2. Principales estadísticas descriptivas, de pacientes diagnosticados con isquemias del corazón del Ecuador en el 2018.

3.2.1. Variable: Provincia de ubicación del centro de salud.

En cuanto a la provincia de ubicación de los centros de salud, se observa que Guayas con el 31,8%, Pichincha con el 23,2% y Azuay con el 15,3% son las provincias con más pacientes egresados con enfermedades isquémicas del corazón, como se muestra en el cuadro 3.1.

Cuadro 3. 1 Estadísticas descriptivas: variable provincia de ubicación

Provincias ubicación	Frecuencia Absoluta	Frecuencia Relativa
1 Rios	211	0,013
2 MorS	51	0,003
3 Pas	20	0,001
4 Coto	63	0,004
5 Suc	58	0,004
6 Orell	15	0,001
7 Boli	40	0,002
8 Tung	259	0,016
9 Napo	10	0,001
10 Esm	177	0,011
11 Man	1872	0,116
12 Loj	886	0,055
13 StoD	225	0,014
14 Oro	429	0,027
15 Gua	5125	0,318
16 Pich	3730	0,232
17 Galp	29	0,002
18 Imb	81	0,005
19 StaE	87	0,005
20 Chi	197	0,012
21 Cañ	40	0,002
22 Azuay	2455	0,153
23 Car	24	0,001
24 Zam	10	0,001
Total	16094	1,000



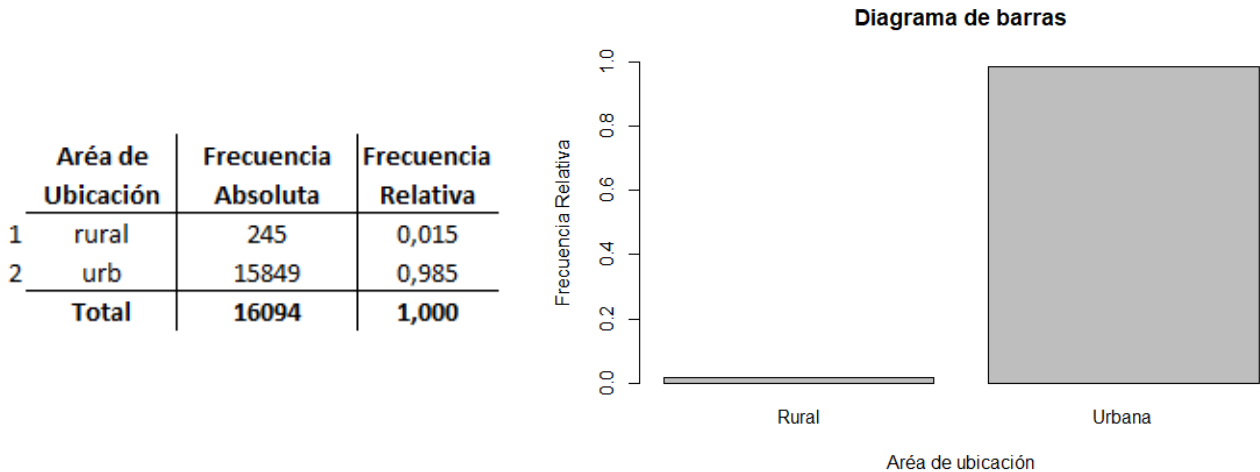
*Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor*

3.2.2. Variable: Área de ubicación del centro de salud.

La cantidad de centros de salud que atendieron a pacientes diagnosticados con enfermedades isquémicas del corazón el 98.5% pertenecen a zonas urbanas,

mientras los centros de salud ubicados en zonas rurales es del 1.5%, está información se la puede observar en el cuadro 3.2.

Cuadro 3. 2 Estadísticas descriptivas: variable área de ubicación

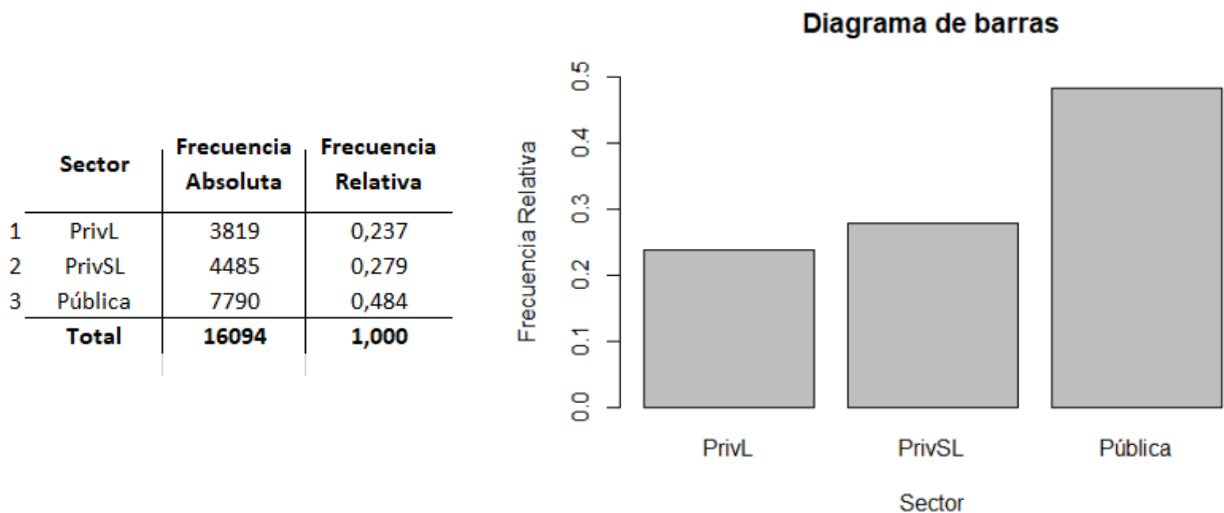


Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor

3.2.3. Variable: Sector.

En el año 2018 los centros de salud que atendieron a pacientes con enfermedades isquémicas del corazón el 48.4% fueron Públicos, Privados con fines de lucro el 23.7% y Privados sin fines de lucro el 27.9%, como lo muestra el cuadro 3.3.

Cuadro 3. 3 Estadísticas descriptivas: variable sector



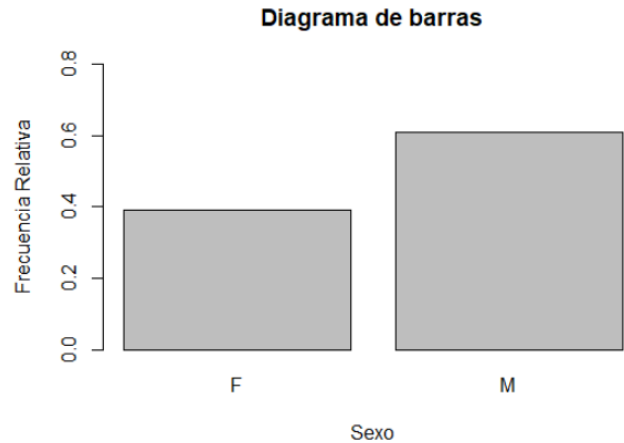
Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor

3.2.4. Variable: Sexo.

Como se observa en el cuadro 3.4, el 61% de los pacientes diagnosticados con enfermedades isquémicas del corazón son hombres y el 39% restante son mujeres.

Cuadro 3. 4 Estadísticas descriptivas: variable sexo

Sexo	Frecuencia Absoluta	Frecuencia Relativa
1 Femenino	6280	0,39
2 Masculino	9814	0,61
Total	16094	1,000

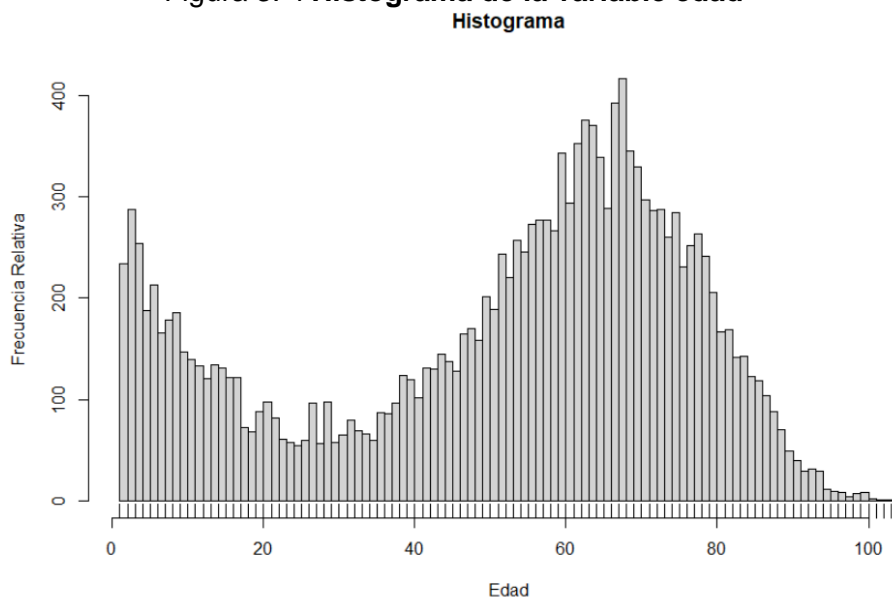


Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor

3.2.5. Variable: Edad.

En la figura 3.4, se puede apreciar el comportamiento de la variable edad, la mayor concentración de pacientes diagnosticados con enfermedades isquémicas es a partir de los 40 años.

Figura 3. 4 Histograma de la variable edad



Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor

El rango de las edades de los pacientes diagnosticados con enfermedades isquémicas del corazón va desde 1 año hasta los 104 años, teniendo los 52 años como la edad promedio para sufrir de esta enfermedad.

Tabla 3. 3 Estadísticas descriptivas: variable edad

	N	1er Q	Mediana	3er Q	Mínimo	Máximo	Media	Desv.
Edad	16094	35	59	70	1	104	52	25.08

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC

Elaboración: Autor

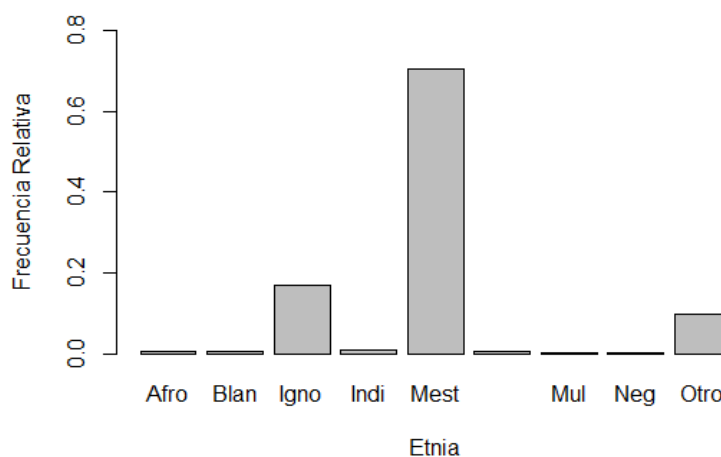
3.2.6. Variable: Etnia.

En el cuadro 3.5 nos informa que el grupo etnia con mayores pacientes diagnosticados con enfermedades isquémicas del corazón es el Mestizo con un 70,5%, cabe resaltar que los registros dicen que el 16,9% fueron ignorados y el 9,7% otros, según el INEC es un problema de registro en los centros de salud.

Cuadro 3. 5 Estadísticas descriptivas: variable etnia

	Etnia	Frecuencia Absoluta	Frecuencia Relativa
1	Afro	61	0,004
2	Blan	113	0,007
3	Igno	2719	0,169
4	Ind	162	0,010
5	Mest	11347	0,705
6	Mont	85	0,005
7	Mul	14	0,001
8	Neg	32	0,002
9	Otro	1561	0,097
	Total	16094	1,000

Diagrama de barras



Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC

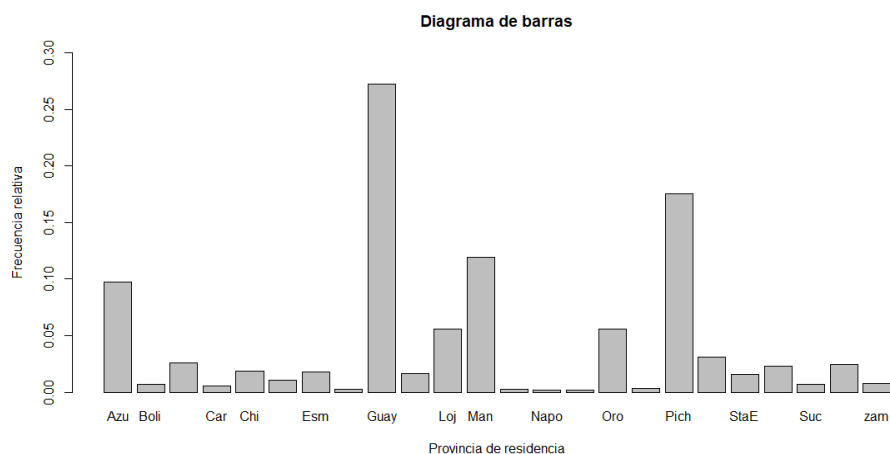
Elaboración: Autor

3.2.7. Variable: Provincia de residencia.

Los pacientes diagnosticados con enfermedades isquémicas del corazón provienen en gran porcentaje de las provincias del Guayas con 27,7%, Pichincha con 17.6%, Manabí con 11,9% y Azuay con 9,7%, como se muestra en el cuadro 3.6.

Cuadro 3. 6 Estadísticas descriptivas: variable provincia de residencia

Provincia residencia	Frecuencia Absoluta	Frecuencia Relativa
Azuay	1565	0,097
Boli	113	0,007
Cañ	341	0,021
Car	91	0,006
Chi	305	0,019
Coto	168	0,010
Esm	293	0,018
Galp	43	0,003
Gua	4383	0,272
Imb	266	0,017
Loj	897	0,056
Man	1921	0,119
MorS	128	0,008
Napo	33	0,002
Orell	33	0,002
Oro	895	0,056
Pas	53	0,003
Pich	2827	0,176
Rios	496	0,031
StaE	258	0,016
StoD	365	0,023
Suc	108	0,007
Tung	393	0,024
Zam	119	0,007
Total	16094	1,000



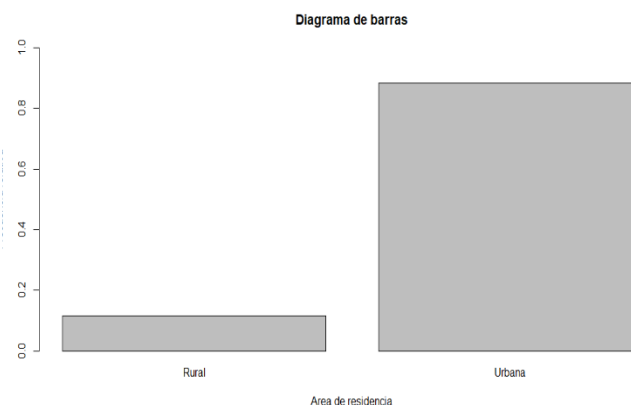
Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor

3.2.8. Variable: Área de residencia.

En el cuadro 3.7, se observa que el 88.4% de los pacientes diagnosticados con enfermedades isquémicas del corazón residen en zonas urbanas y el 11,6% restante en zonas rurales.

Cuadro 3. 7 Estadísticas descriptivas: variable área de residencia

Área de residencia	Frecuencia Absoluta	Frecuencia Relativa
rural	1869	0,116
urb	14225	0,884
Total	16094	1,000

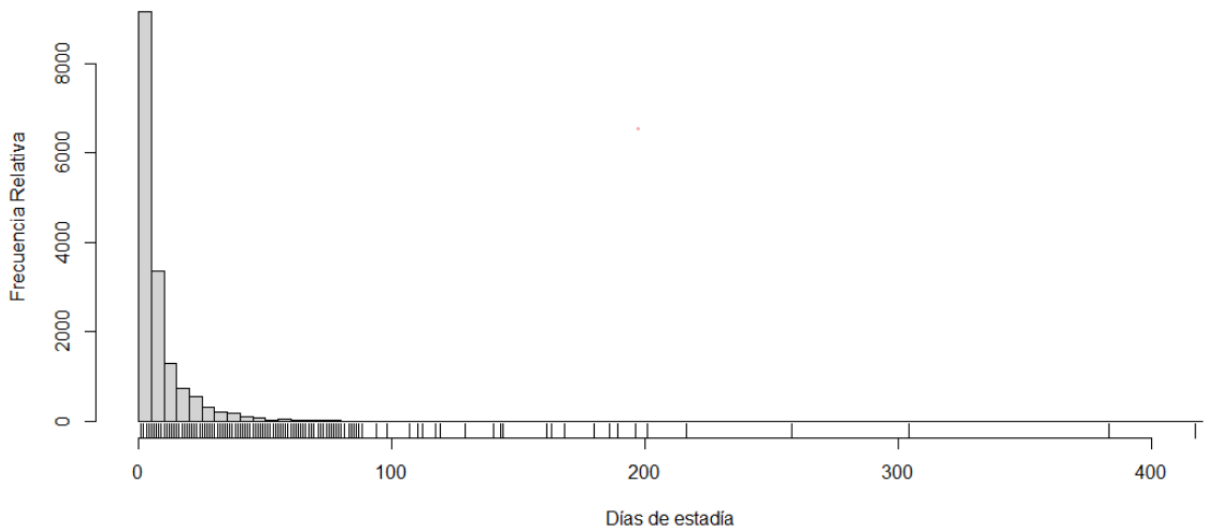


Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor

3.2.9. Variable: Días de estadía.

El comportamiento de la variable días de estadías lo apreciamos en la figura la figura 3.5, la mayor concentración de pacientes diagnosticados con enfermedades isquémicas permanece en los centros de salud pocos días.

Figura 3. 5 Histograma de la variable días de estadía
Histograma



Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor

El rango de los días de estadía en los centros de salud, los pacientes diagnosticados con enfermedades isquémicas del corazón desde 1 día hasta los 417, teniendo 8 días como el promedio de estadía, se puede observar en la tabla 3.4.

Tabla 3. 4 Estadísticas descriptivas: variable días de estadía

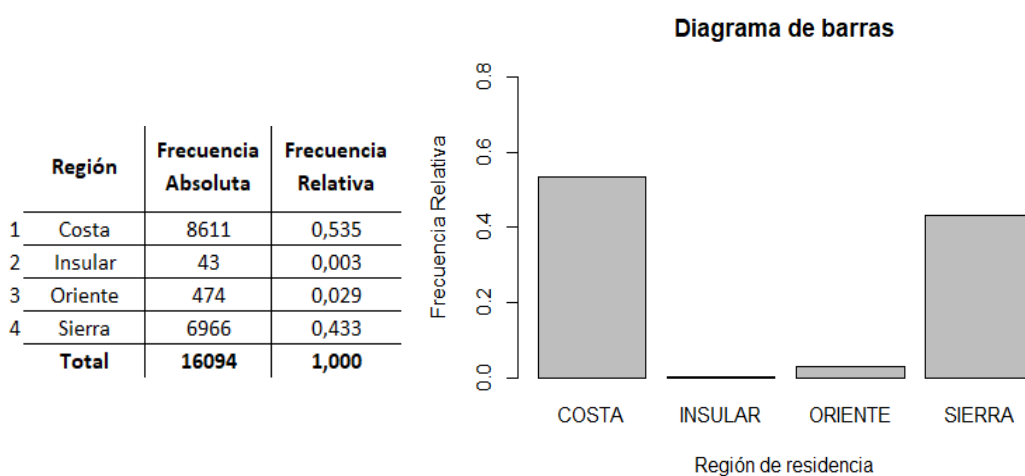
	N	1er Q	Mediana	3er Q	Mínimo	Máximo	Media	Desv.
Días de estadía	16094	2	5	8,296	1	417	8,269	25.08

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor

3.2.10. Variable: Región de Residencia

Como se puede observar en el cuadro 3.8, las regiones de la costa y la sierra concentran la mayor cantidad de pacientes con 53.3% y 43.4% respectivamente.

Cuadro 3. 8 Estadísticas descriptivas: variable región de residencia

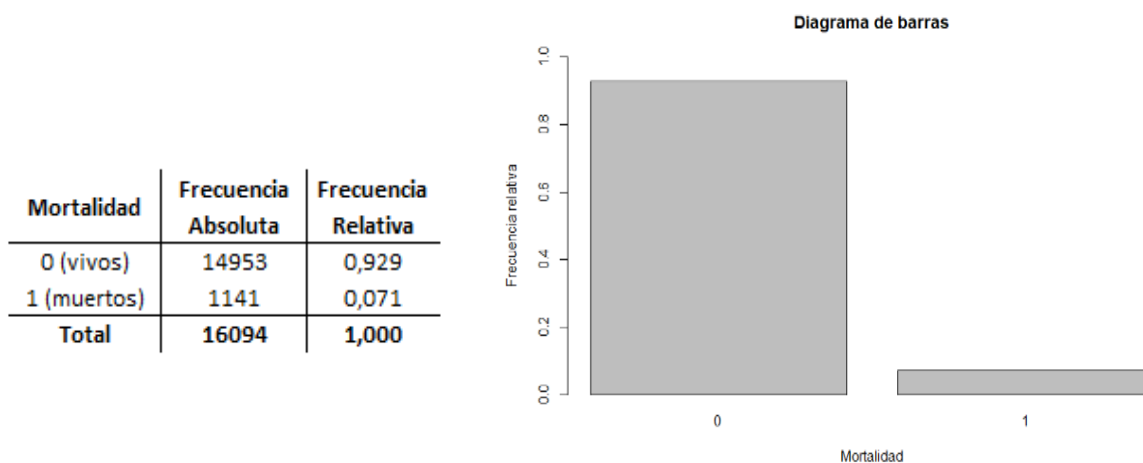


*Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor*

3.2.11. Variable: Condición de egreso.

En el cuadro 3.9, se observa que el 7.1% de los pacientes fallecieron a causa de una enfermedad isquémica del corazón, mientras el 92.9% permanecieron con vida.

Cuadro 3. 9 Estadísticas descriptivas: variable mortalidad



*Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor*

CAPÍTULO 4

4. RESULTADOS

4.1. Introducción

En este apartado se realizan diferentes pruebas estadísticas que permiten contrastar las hipótesis planteadas. Las variables independientes al ser analizadas a través de tablas de contingencia y pruebas de ji cuadrado, para las variables independientes cuantitativas utilizamos la prueba de t de student, al efectuar estas pruebas se obtendrá nuestro primer punto de vista de la relación de las variables explicativas y la variable que se pretende predecir.

A continuación se muestran los resultados de la implementación de un modelo de regresión logística multivariante, donde se puede predecir la condición de egreso de los pacientes que fueron diagnosticados con enfermedades isquémicas del corazón, tratados y dados de alta en los centros de salud de la red complementaria, en función de un conjunto de variables del paciente, del centro de salud.

4.2. Asociaciones Bivariantes

4.2.1. Condición de egreso vs Provincia de ubicación.

La tabla 4.1, muestra los resultados obtenidos de la condición de egreso de los pacientes según la provincia de ubicación del centro de salud, donde la prueba Chi-Cuadrado de Pearson arrojó un estadístico de 104.4851 con 23 grados de libertad.

Tabla 4. 1 Tabla de Contingencia. Condición de egreso vs. Provincia de ubicación

Provincia de Ubicación	Vivo	Muerto	Total
Azuay	2346 95.56%	109 4.4%	2455 15.3%
Boli	35 87.5%	5 12.5%	40 0.2%
Cañ	38 95%	2 5%	40 0.2%
Car	23 95.8%	1 4.2%	24 0.1%

Chi	186	11	197
	0,944	0,056	1.2%
Coto	54	9	63
	85.7%	14.3%	0.4%
Esm	160	17	177
	90.4%	9.6%	1.1%
Galp	27	2	29
	93.1%	6.9%	0.2%
Gua	4769	356	5125
	93.1%	6.9%	31.8%
Imb	76	5	81
	93.8%	6.2%	0.5%
Loj	819	67	886
	92.4%	7.6%	5.5%
Man	1715	157	1872
	91.6%	8.4%	11.6%
MorS	43	8	51
	84.3%	15.7%	0.3%
Napo	9	1	10
	90%	10%	0.1%
Orell	13	2	15
	86.7%	13.3%	0.1%
Oro	399	30	429
	93%	7%	2.7%
Pas	17	3	20
	85%	15%	0.1%
Pich	3471	259	3730
	93.1%	6.9%	23.2%
Rios	171	40	211
	81%	19%	1.3%
StaE	82	5	87
	94.3%	5.7%	0.5%
StoD	208	17	225
	92.4%	7.6%	1.4%
Suc	50	8	58
	86.2%	13.8%	0.4%
Tung	232	27	259
	89.6%	10.4%	1.6%
Zam	10	0	10
	1	0	0.1%
Total	14953	1141	16094
	92.91%	7.09%	100%

Chi² = 104.4851 d.f. = 23 p = 2.34e-12

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor

Para el análisis de la relación de estas variables se realiza el siguiente contraste de hipótesis:

Ho: Provincia de ubicación del centro de salud es independiente de la condición de egreso de los pacientes diagnosticados con enfermedades isquémicas del corazón.

Vs

H1: No es verdad Ho

Donde se obtuvo un nivel de significancia menor a 0.05, por lo cual se posee suficiente evidencia estadística para rechazar la hipótesis nula.

4.2.2. Condición de egreso vs Área de ubicación.

Los resultados en la tabla 4.2, muestran la condición de egreso de los pacientes según el área de ubicación del centro de salud, donde la prueba Chi-Cuadrado de Pearson arrojó un estadístico de 5.245629 con 1 grado de libertad.

Tabla 4. 2 Tabla de contingencia. Condición de egreso vs. Área de Ubicación

Área de Ubicación	Vivo	Muerto	Total
 rur 	218	27	245
	89.%	11%	1.5%
 urb 	14735	1114	15849
	93%	7%	98.5%
 Total 	14953	1141	16094
	92.91%	7.09%	100%
	Chi ² = 5.245629	d.f. = 1	p = 0.022

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor

Se realiza el siguiente contraste de hipótesis:

Ho: Área de ubicación del centro de salud es independiente de la condición de egreso de los pacientes diagnosticados con enfermedades isquémicas del corazón.

Vs

H1: No es verdad Ho

Se obtuvo un nivel de significancia menor a 0.05, por lo cual se posee suficiente evidencia estadística para rechazar la hipótesis nula.

4.2.3. Condición de egreso vs Sector.

La tabla 4.3, muestra los resultados obtenidos de la condición de egreso de los pacientes según el sector económico a los que pertenecen los centros de salud, donde la prueba Chi-Cuadrado de Pearson arrojó un estadístico de 8.141314 con 2 grados de libertad.

Tabla 4. 3 Tabla de contingencia. Condición de egreso vs. Sector

Sector	Vivo	Muerto	Total
 PrivL 	3551	268	3819
	93%	7%	23.7%

PrivSL	4205	280	4485
	93.8%	6.2%	27.9%
Pública	7197	593	7790
	92.4%	7.6%	48.4%
Total	14953	1141	16094
	92.91%	7.09%	100%
	Chi ² = 8.141314		d.f. = 2
			p = 0.0171

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor

El contraste de hipótesis planteamos Ho: sector de los centro de salud es independiente de la condición de egreso de los pacientes diagnosticados con enfermedades isquémicas del corazón.

Vs

H1: No es verdad Ho

Se obtuvo un nivel de significancia de 0,0171, por lo cual se posee suficiente evidencia estadística para rechazar la hipótesis nula.

4.2.4. Condición de egreso vs Sexo.

En la tabla 4.4, observamos los resultados obtenidos de la condición de egreso de los pacientes según el sexo de los pacientes, donde la prueba Chi-Cuadrado de Pearson presentó un estadístico de 19.02405 con 1 grado de libertad.

Tabla 4. 4 Tabla de contingencia. Condición de egreso vs. Sexo

Sexo	Vivo	Muerto	Total
F	5765	515	6280
	91.8%	8.2%	39%
M	9188	626	9814
	93.6%	6.4%	61%
Total	14953	1141	16094
	92.91%	7.09%	100%
	Chi ² = 19.02405		d.f. = 1
			p = 1.29e-05

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor

Ho: sexo de los pacientes es independiente de su condición de egreso.

Vs

H1: No es verdad Ho

Se obtuvo un nivel de significancia menor a 0.05, por lo cual se posee suficiente evidencia estadística para rechazar la hipótesis nula.

4.2.5. Condición de egreso vs Edad.

Al realizar la prueba de Levene para igualdad de varianzas (tabla 4.5) se obtuvo un valor p menor a 0.05, no podemos asumir que las varianzas son iguales; realizamos la prueba t de student para la igualdad de medias se obtiene un valor p menor a 0.05 lo que indica que sí existen diferencias significativas en las medias de las edades de los pacientes muertos vs vivos.

A continuación mostramos un análisis detallado, podemos observar la relación de la edad de los pacientes versus la condición de egreso, sean muertos o no muertos.

Tabla 4. 5 Prueba de Levene. Condición de egreso vs. Edad

	F	Sig.
Edad	15.244	0

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor

Tabla 4. 6 Prueba t de student. Variables Mortalidad vs. Edad

t	df	Sig. (bilateral)	Diferencia de Medias	Error tip. de la diferencia	95% intervalo de confianza para la diferencia	
					Inferior	Superior
Edad	-13,346	1338,6	2.2e-16		-11,26557	-8,37819

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor

4.2.6. Condición de egreso vs Etnia.

Los resultados obtenidos de la condición de egreso de los pacientes según su etnia, podemos observar en la tabla 4.7 la prueba Chi-Cuadrado de Pearson dio un estadístico de 34.98475 con 8 grados de libertad.

Tabla 4. 7 Tabla de contingencia. Condición de egreso vs Etnia

Etnia	Vivo	Muerto	Total
Afro	54	7	61
	88.5%	11.5%	0.4%
Blan	103	10	113
	91.2%	8.8%	0.7%
Igno	2590	129	2719
	95.3%	4.7%	16.9%
Indi	150	12	162
	92.6%	7.4%	1%
Mest	10482	865	11347
	92.4%	7.6%	7.05%
Mont	80	5	85

	94.1%	5.9%	0.5%
Mul	12	2	14
	85.7%	14.3%	0.1%
Neg	27	5	32
	84.4%	15.6%	0.2%
Otro	1455	106	1561
	93.2%	6.8%	9.7%
Total	14953	1141	16094
	92.91%	7.09%	100%
	Chi ² = 34.98475	d.f. = 8	p = 2.69e-05

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor

El contraste de hipótesis:

Ho: La etnia de los pacientes diagnosticados con enfermedades isquémicas del corazón es independiente de su condición de egreso.

Vs

H1: No es verdad Ho

Se obtuvo un nivel de significancia menor a 0.05, por lo cual se posee suficiente evidencia estadística para rechazar la hipótesis nula.

4.2.7. Condición de egreso vs Provincia de residencia.

La tabla 4.8, muestra los resultados obtenidos de la condición de egreso de los pacientes según la provincia de residencia de los pacientes, teniendo en la prueba Chi-Cuadrado de Pearson un estadístico de 75.02141 con 23 grados de libertad

Tabla 4. 8 Tabla de contingencia. Condición de egreso vs. Provincia de residencia

Provincia de residencia	Vivo	Muerto	Total
Azuay	1498	67	1565
	95.7%	4.3%	9,70%
Boli	101	12	113
	89.4%	10.6%	0.7%
Cañ	320	21	113
	93.8%	6.2%	0.21%
Car	87	4	91
	95.6%	4.4%	0.6%
Chi	290	15	305
	95.1	4.9%	1.9%
Coto	149	19	168
	88.7%	9.2%	1.8%
Esm	266	27	293
	90.8%	9.6%	1.1%
Galp	41	2	43
	95.3%	4.7%	0.3%
Gua	4072	311	4383
	92.9%	7.1%	27.2%

Imb	239	27	266
	89.8%	10.2%	1.7%
Loj	825	72	897
	92.4%	7.6%	5.5%
Man	1758	163	1921
	91.5%	8.5%	11.9%
MorS	116	12	128
	90.6%	9.4%	0.8%
Napo	30	3	33
	90.9%	9.1%	0.2%
Orell	31	2	33
	93.9%	6.1%	0.2%
Oro	849	46	895
	94.9%	5.1%	5.6%
Pas	47	6	53
	88.7%	11.3%	0.3%
Pich	2638	189	2827
	93.3%	6.7%	17.6%
Rios	438	58	496
	88.3%	11.7%	3.1%
StaE	239	19	258
	92.6%	7.4%	1.6%
StoD	348	17	365
	95.3%	4.7%	2.3%
Suc	98	10	108
	90.7%	9.3%	0.7%
Tung	359	34	393
	91.3%	8.7%	2.4%
Zam	114	5	119
	95.8	4.2%	0.7%
Total	14953	1141	16094
	92.91%	7.09%	100%
Chi ² = 75.02141			d.f. = 23
			p = 1.99e-07

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor

Para el análisis de la relación de estas variables realizamos el siguiente contraste de hipótesis:

Ho: Provincia de ubicación de los pacientes es independiente de la condición de egreso de los pacientes diagnosticados con enfermedades isquémicas del corazón.

Vs

H1: No es verdad Ho

Donde se obtuvo un nivel de significancia menor a 0.05, por lo cual se posee suficiente evidencia estadística para rechazar la hipótesis nula

4.2.8. Condición de egreso vs Área de residencia.

Los muestra los resultados en la tabla 4.9, nos muestra la relación de la condición de egreso de los pacientes según el área de su residencia, donde la prueba Chi-Cuadrado de Pearson arrojó un estadístico de 4.44611 con 1 grados de libertad.

Tabla 4. 9 Tabla de contingencia. Condición de egreso vs. Área de residencia

Área de residencia	Vivo	Muerto	Total
rural	1714	155	1869
	91.7%	8.3%	11.6%
urb	13239	986	14225
	93.1%	6.9%	88.4%
Total	14953	1141	16094
	92.91%	7.09%	100%
Chi ² = 4.44611			d.f. = 1
			p = 0.035

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor

Donde:

Ho: Área de residencia de los pacientes es independiente de la condición de egreso de los pacientes diagnosticados con enfermedades isquémicas del corazón.

Vs

H1: No es verdad Ho

Donde se obtuvo un nivel de significancia menor a 0.05, por lo cual se posee suficiente evidencia estadística para rechazar la hipótesis nula.

4.2.9. Condición de egreso vs Días de estadía.

Realizamos la prueba Levene (tabla 4.10) se obtuvo un valor p menor a 0.05, no podemos asumir que la varianza de las variables detalladas son iguales; al realizar la prueba t de student se obtiene un valor p menor a 0.05 lo que indica que sí existen diferencias significativas en los días de estadía en un centro de salud de los pacientes muertos vs. vivos.

Tabla 4. 10 Prueba de Levene. Condición de egreso vs. Días de estadía

	F	Sig.
Edad	9.7827	0

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor

Tabla 4. 11 Prueba t de student. Condición de egreso vs. Días de estadía

95% intervalo de

	t	df	Sig. (bilateral)	Diferencia de Medias	Error tip. de la diferencia	confianza para la diferencia Inferior Superior
Edad	-84.326	16107	2.2e-16			-8.416748 -8.034352

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor

4.2.10. Condición de egreso vs región de residencia.

Los resultados en la tabla 4.12 nos muestra la relación de la condición de egreso de los pacientes según el área de su residencia, donde la prueba Chi-Cuadrado de Pearson tiene un estadístico de 4.44611 con 1 grados de libertad.

Tabla 4. 12 Tabla de contingencia. Condición de egreso vs. Región de residencia

Región de residencia	Vivo	Muerto	Total
Costa	7970	641	8611
	53.3%	56.1%	53.5%
Insular	41	2	43
	0.27%	0.001%	0.26%
Oriente	436	38	474
	2.91%	3.3%	2.9%
Sierra	6506	460	6966
	43.5%	40.31%	43.28%
Total	14953	1141	16094
	92.91%	7.09%	100%
	Chi ² = 5.147398	d.f. = 3	p = 0.161

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor

Donde:

Ho: Región de residencia de los pacientes es independiente de la condición de egreso de los pacientes diagnosticados con enfermedades isquémicas del corazón.

Vs

H1: No es verdad Ho

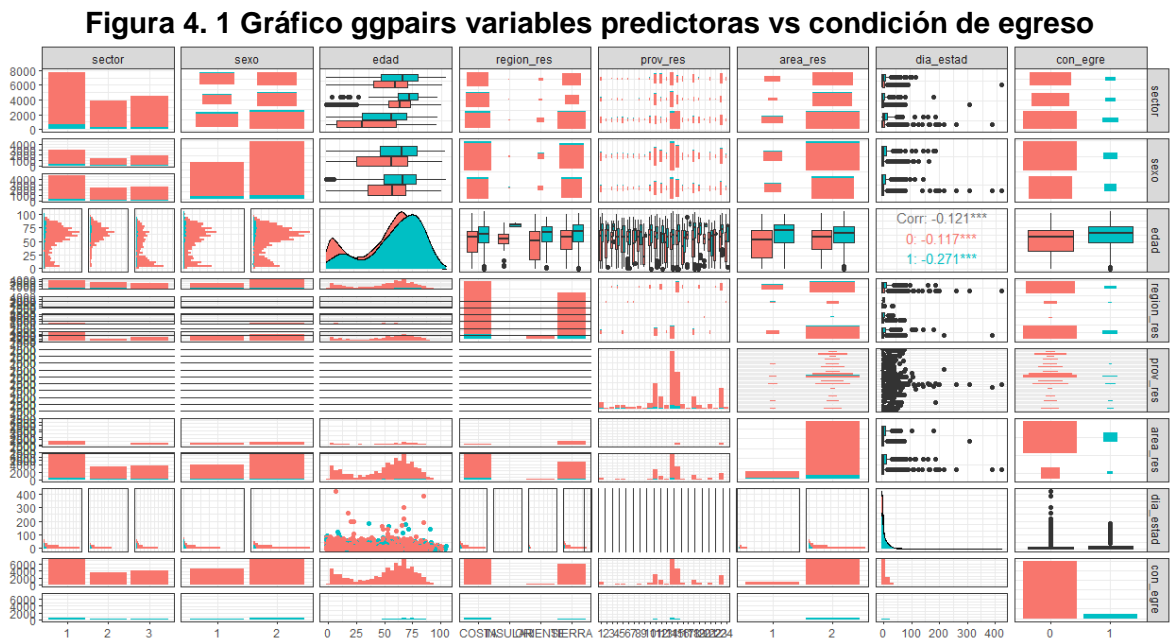
Donde se obtuvo un nivel de significancia mayor a 0.05, por lo cual se posee suficiente evidencia estadística para no rechazar la hipótesis nula.

4.3. Regresión Logística

Como se indicó en capítulos anteriores, la variable a explicar con esta técnica estadística será condición de egreso, la cual indica si un paciente diagnosticado, tratado de una enfermedad isquémica del corazón, es dado de alta vivo o muerto.

4.3.1. Selección de variables predictoras

Para la creación del modelo, se utilizarán variables que estén relacionadas con información del paciente y su entorno tales como: sexo, edad, área de residencia, provincia de residencia y región de residencia, también se incluirán variables relacionadas al centro de salud donde fueron atendidos los pacientes como sector y los días de estadía. Como se muestra en la figura 4.1.



Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor

Se puede observar la edad de los pacientes que fallecieron de enfermedades isquémicas del corazón en promedio es mayor a los que egresaron vivos.

La cantidad de mujeres y hombres que fallecieron por enfermedades isquémicas del corazón es muy parecida, la población de pacientes hombres es mayor a la de mujeres.

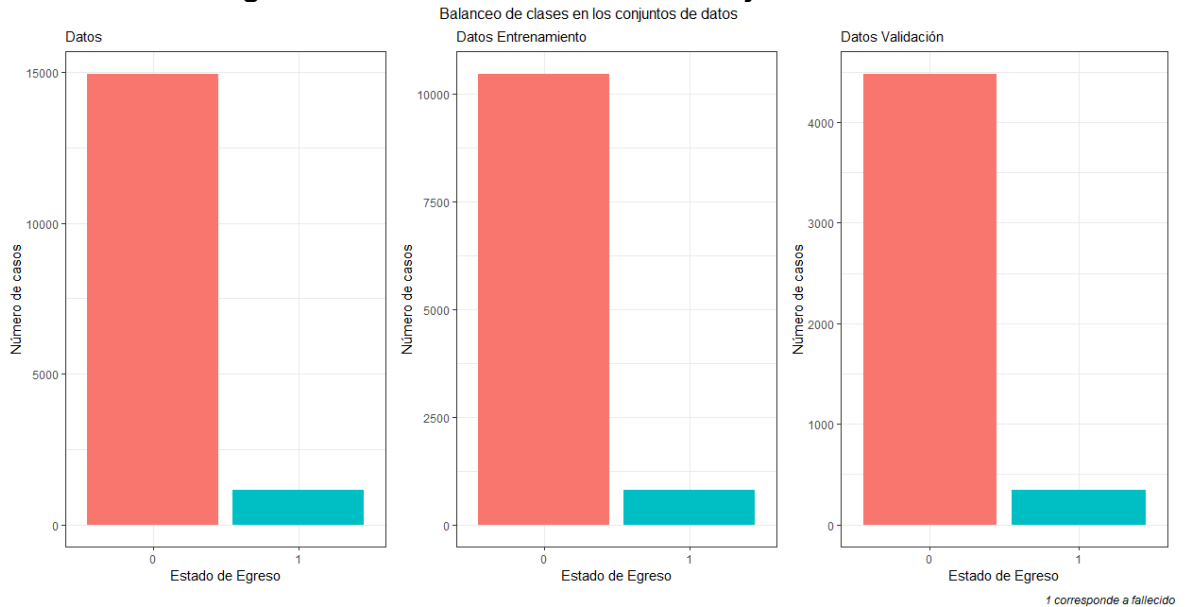
Los pacientes que permanecieron más días en los centros de salud, su condición de egreso fue de vivos.

4.3.2. Preparación de los datos

Se particionan en 2 grupos, el primer grupo que contiene el 70% se utilizará para entrenar el modelo y el 30% restante servirá para validar el modelo y encontrar el punto de corte óptimo.

En la figura 4.2, se observa el conjunto de datos globales, los datos utilizados para entrenar el modelo y los datos para la validación del mismo.

Figura 4. 2 Balanceo de clases en conjuntos de datos



Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor

4.3.3. Modelos Logísticos

El objetivo del modelo de regresión logística es encontrar el efecto de las variables del paciente y del centro de salud, sobre la probabilidad de fallecer de una enfermedad isquémica del corazón.

Cabe señalar que las variables provincia de residencia y región de residencia están correlacionadas, serán utilizadas en modelos diferentes.

Se desarrollaron 3 modelos logísticos y se eligió el mejor modelo que incluya a dichas variables predictoras, como se muestra en la tabla 4.13.

Tabla 4. 13 Modelos logísticos

Modelos	Variabes predictoras	AIC	Porcentaje de Deviance explicada
Modelo 1	sector, sexo, edad, provincia de residencia, área de residencia, días de estadía	5548.5	0.04857659
Modelo 2	sector, sexo, edad, área de residencia, días de estadía	5562.5	0.03816384
Modelo 3	sector, sexo, edad, región de residencia, días de estadía	5562.9	0.03913385

El mejor modelo de acuerdo al valor del AIC es el modelo 1, este incluye a las variables predictoras sector, sexo, edad, provincia de residencia, área de residencia y días de estadía.

En la tabla 4.14, se observa la codificación de las variables que utilizaremos para nuestro modelo final.

Tabla 4. 14 Codificación de variables del modelo regresión logística

Variables	Codificación
Variable dependiente	0: vivo 1: muerto
con_egre	
sector	1.Público 2.PrivL 3PrivSL
sexo	1 : Masculino 0 : Femenino
edad	[1,104] años
Variables predictoras	1. AZUAY 2. BOLIVAR 3. CAÑAR 4. CARCHI 5. COTOPAXI 6.CHIMBORAZO 7.EL ORO 8.ESMERALDAS 9.GUAYAS 10.IMBABURA 11.LOJA 12.LOS RIOS 13.MANABI 14.MORONA SANTIAGO 15.NAPO 16.PASTAZA 17.PICHINCHA 18.TUNGURAHUA 19.ZAMORA CHINCHIPE 20.GALAPAGOS 21.SUCUMBIOS 22.ORELLANA 23.SANTO DOMINGO DE LOS TSHILAS 24.SANTA ELENA
prov_res	
area_res	1: urbana 2: rural
dia_estad	[1-104] años

4.3.4. Resultados del modelo

El modelo de regresión logística estimado está compuesto por 7 variables independientes, los valores de B estimados, error estándar, estadístico de Wald y valor p los tenemos en la tabla 4.15.

Tabla 4. 15 Coeficientes B estimados regresión logit

Grupo de control	Covariable	B	Error Standar	Estadístico De Wald	Valor p
constante		-3,006	0.235345	-12.773	< 2e-16
Sector (público)	PrivL	-0,148	0.099723	-1.489	0.13650
	PrivSL	0,147	0,099	1496,000	0,135
Sexo (femenino)	masculino	-0,236	0,076	-3119,000	0,002
edad	edad	0,022	0,002	11,930	< 2e-16
Provincia de residencia (Los Ríos)	Pas	0,074	0,523	0,142	0,887
	Boli	0,002	0,402	0,005	0,996
	Imb	-0,138	0,302	-0,457	0,648
	Esm	-0,279	0,301	-0,927	0,354
	MorS	0,100	0,358	0,279	0,780
	Suc	-0,246	0,440	-0,560	0,576
	Coto	-0,451	0,405	-1,112	0,266
	Napo	0,113	0,770	0,147	0,883
	Tung	-0,477	0,287	-1,665	0,096
	Man	-0,251	0,202	-1,244	0,214
	Loj	-0,197	0,230	-0,854	0,393
	StaE	-0,353	0,350	-1,009	0,313
	Gua	-0,568	0,191	-2,966	0,003
	Pich	-0,602	0,196	-3,070	0,002
	Cañ	-0,376	0,316	-1,189	0,234
	Orell	-0,552	0,768	-0,718	0,473
	Oro	-0,858	0,256	-3,349	0,001
	Chi	-0,797	0,337	-2,360	0,018
	Galp	-1,026	1,037	-0,989	0,323
	StoD	-1,144	0,376	-3,040	0,002
	Car	-1,472	0,745	-1,975	0,048
	Azuay	-1,066	0,230	-4,629	0,000
	Zam	-0,803	0,548	-1,464	0,143
Sector de residencia (rural)	urb	-0,359	0,110	-3,264	0,001
Días de estadía	Dia_est	0,020	0,003	7,806	0,000

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC

Elaboración: Autor

Es importante mencionar que las categorías de las variables que se utilizó como grupo de control, todas ellas tienen el porcentaje más alto de pacientes cuya condición de egreso es muerto.

Al analizar los signos de coeficientes de la tabla 4.15, se encontraron variables con signos negativos, estas variables son sector, sexo, provincia de residencia y sector

de residencia, estas variables actúan como factores de protección a mortalidad causada por enfermedades isquémicas del corazón, mientras que la edad y los días de estadía son factores de riesgo para los pacientes.

La oportunidad relativa (Odds Ratio), es la comparación de la probabilidad de ocurrencia de un evento luego de la exposición a un determinado factor de riesgo, frente a la probabilidad de ocurrencia del mismo evento en una situación o estado de control (Andrade, 2015). En la tabla 4.16, se muestran las oportunidades relativas del modelo estimado, su intervalo de confianza, y su interpretación.

Tabla 4. 16 Oportunidades relativas

Grupo de control	Covariable	Exp(B)	Intervalo 95% para Exp(B)		Interpretación
			Inf.	Sup.	
Sector (público)	PrivL	0,862	0,708	1,047	Si bien el odd es de 0.862(menor a 1), al 95% de confianza no es posible concluir que la probabilidad se disminuye puesto que el límite superior del intervalo es mayor a 1.
	PrivSL	1,159	0,954	1,404	Siendo el odd de 1,159, al 95% de confianza no es posible concluir que la probabilidad aumente puesto que el límite inferior del intervalo es menor a 1.
Sexo (femenino)	masculino	0,790	0,681	0,917	Ser del sexo masculino disminuye 79% la posibilidad de fallecer de una enfermedad isquémica en relación a la referencia.
Edad	edad	1,022	1,018	1,026	La probabilidad de fallecer por una enfermedad isquémica del corazón aumenta por cada año cumplido.
Provincia de residencia (Los Ríos)	Pas	1,077	0,345	2,791	Siendo el odd de 1,077 al 95% de confianza no es posible concluir que la propensión aumentó puesto que el límite inferior del intervalo es menor a 1.
	Boli	1,002	0,431	2,122	Siendo el odd de 1,002 al 95% de confianza no es posible concluir que la propensión aumentó puesto que el límite inferior del intervalo es menor a 1.
	Imb	0,871	0,474	1,556	Si bien el odd es de 0.871(menor a 1), al 95% de confianza no es posible concluir que la propensión se disminuye puesto que el límite superior del intervalo es mayor a 1.
	Esm	0,757	0,412	1,349	Si bien el odd es de 0.757(menor a 1), al 95% de confianza no es posible concluir que la propensión se disminuye puesto que el límite superior del intervalo es mayor a 1.
	MorS	1,105	0,528	2,175	Siendo el odd de 1,105 al 95% de confianza no es posible concluir que la propensión aumente puesto que el límite inferior del intervalo es menor a 1.
	Suc	0,782	0,305	1,753	El odd es de 0.782(menor a 1), al 95% de confianza no es posible concluir que la propensión se disminuye puesto que el límite superior del intervalo es mayor a 1.

Coto	0,637	0,270	1,349	Siendo el odd de 0.637(menor a 1), al 95% de confianza no es posible concluir que la propensión se disminuye puesto que el límite superior del intervalo es mayor a 1.
Napo	1,120	0,173	4,136	Siendo el odd de 1,120 al 95% de confianza no es posible concluir que la propensión aumente puesto que el límite inferior del intervalo es menor a 1.
Tung	0,621	0,349	1,079	El odd es de 0.621(menor a 1), al 95% de confianza no es posible concluir que la propensión se disminuye puesto que el límite superior del intervalo es mayor a 1
Man	0,778	0,529	1,168	El valor de odd 0.778(menor a 1), al 95% de confianza no es posible concluir que la propensión se disminuye puesto que el límite superior del intervalo es mayor a 1
Loj	0,821	0,525	1,297	El odd es de 0.821(menor a 1), al 95% de confianza no es posible concluir que la propensión se disminuye puesto que el límite superior del intervalo es mayor a 1.
StaE	0,702	0,340	1,357	El odd es de 0.702(menor a 1), al 95% de confianza no es posible concluir que la propensión se disminuye puesto que el límite superior del intervalo es mayor a 1
Gua	0,567	0,394	0,835	Ser oriundo de la provincia del Guayas disminuye un 56,7% la propensión de fallecer de una enfermedad isquémica del corazón en relación a la provincia de los Ríos.
Pich	0,548	0,377	0,814	Ser oriundo de la provincia de Pichincha disminuye un 54,8% la propensión de fallecer de una enfermedad isquémica del corazón en relación a la provincia de los Ríos.
Cañ	0,687	0,361	1,255	El odd es de 0,687(menor a 1), al 95% de confianza no es posible concluir que la propensión se disminuye puesto que el límite superior del intervalo es mayor a 1
Orell	0,576	0,089	2,118	El odd es de 0.576(menor a 1), al 95% de confianza no es posible concluir que la propensión se disminuye puesto que el límite superior del intervalo es mayor a 1
Oro	0,424	0,255	0,699	Pertener a la provincia del Oro disminuye un 42,4% la propensión de fallecer de una enfermedad isquémica del corazón en relación a la provincia de los Ríos
Chi	0,451	0,225	0,852	Pertener a la provincia de Chimborazo disminuye un 45,1% la propensión de fallecer de una enfermedad isquémica del corazón en relación a la provincia de los Ríos
Galp	0,358	0,020	1,785	El odd es de 0.358(menor a 1), al 95% de confianza no es posible concluir que la probabilidad se disminuye puesto que el límite superior del intervalo es mayor a 1
StoD	0,319	0,144	0,641	Ser oriundo de la provincia de Santo Domingo disminuye un 31,9% la propensión de fallecer de una enfermedad isquémica del corazón en relación a la provincia de los Ríos.
Car	0,230	0,036	0,790	Ser de la provincia de Cañar disminuye un 23% la propensión de fallecer de una enfermedad isquémica del corazón en relación a la provincia de los Ríos.
Azuay	0,345	0,220	0,543	Ser oriundo de la provincia de Azuay disminuye un 34,5% la propensión de fallecer de una enfermedad isquémica del corazón en relación a la provincia de los Ríos.
Zam	0,448	0,130	1,178	El odd es de 0.448(menor a 1), al 95% de confianza no es posible concluir que la probabilidad se disminuye puesto que el límite superior del intervalo es mayor a 1.

Sector de residencia (rural)	urb	0,698	0,565	0,870	Los pacientes que residen en el sector urbano, tienen 69,8% menos propensión de fallecer de una enfermedad isquémica del corazón.
Días de estadía	Dia_est	1,020	1,015	1,256	Por cada día que permanecen los pacientes en los centros de salud, se incrementa en un 1,020 la propensión a fallecer.

*Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor*

Con los coeficientes B estimados de la tabla 4.16 se construye la ecuación del modelo, utilizamos las variables que son significativas, donde $y = 1$ es la probabilidad de que paciente ingresado a un centro de salud perteneciente a la red hospitalaria del Ecuador, sea diagnosticado de alguna enfermedad isquémica del corazón, tratado, dado de alta y fallezca por la enfermedad.

$$P(y = 1/x) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}}$$

Donde

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

$(y = 1 | x) = 1 / (1 + e^{(3,006 + 0,236 * \text{sexo masculino} - 0,022 * \text{edad} + 0,568 * \text{Provincia residencia Guayas} + 0,602 * \text{Provincia residencia Pichincha} + 0,858 * \text{Provincia residencia El Oro} + 0,787 * \text{Provincia residencia Chimborazo} + 1,144 * \text{Provincia residencia Santo Domingo} + 1,472 * \text{Provincia residencia Carchi} + 1,066 * \text{Provincia residencia Azuay} + 0,359 * \text{sector de residencia urbana} - 0,02 * \text{días de estadía})$

4.3.5. Ajuste del modelo

Una vez estimado el modelo, procedemos a evaluar su bondad de ajuste, para esto utilizaremos varios métodos; el primero es el pseudo R^2 , mide qué tan buenas predictoras son las variables independientes en relación a la dependiente. El pseudo R^2 de McFadden obtenido es de 0.0486, el ajuste del modelo no es bueno. Otra medida del ajuste es la significancia global del modelo. Esta prueba evalúa si el modelo con predictores se ajusta significativamente mejor que un modelo nulo. El estadístico de prueba es la diferencia entre la desviación residual del modelo con las variables predictoras y la desviación residual del modelo nulo, sigue una distribución Ji-cuadrado con grados de libertad iguales a las diferencias en grados de libertad entre el modelo actual y nulo. Para nuestro modelo el valor p obtenido

es 0, lo que indica que nuestro modelo se ajusta significativamente mejor que el modelo nulo.

Por otro lado, la prueba de Hosmer y Lemeshow compara por intervalos los valores observados con los esperados, sigue una distribución Ji-cuadrado y su hipótesis nula manifiesta que no existen diferencias entre los valores observados y los esperados; para nuestro modelo el valor p obtenido es de 2.2e-16, por lo que existe evidencia estadística suficiente para rechazar la hipótesis nula, el modelo no se ajusta bien, pues Ho: el modelo es adecuado. Cómo se muestra en la tabla 4. 17.

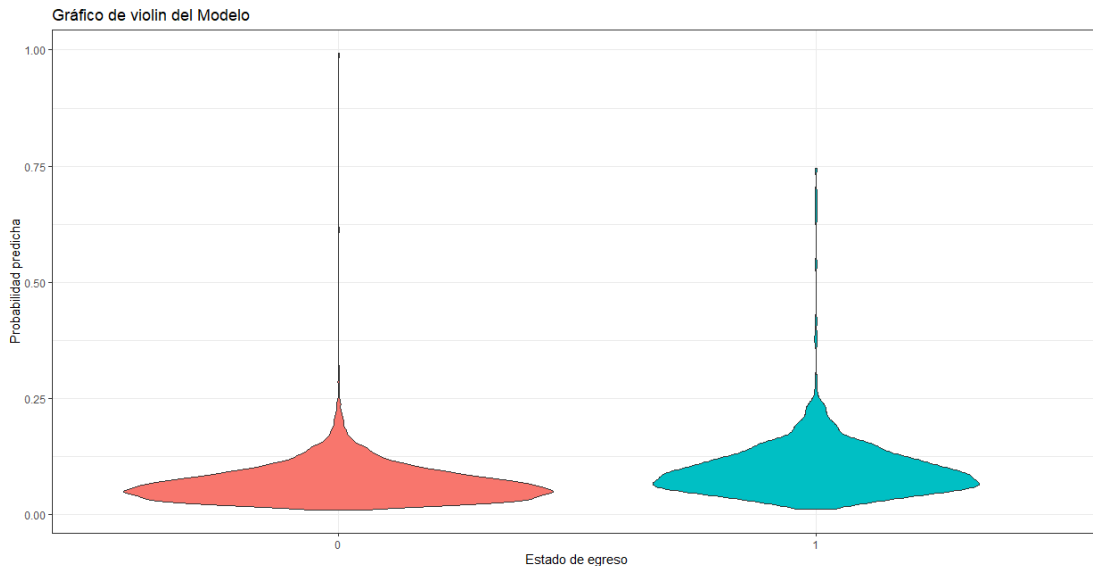
Tabla 4. 17 Medidas de bondad de ajuste del modelo

Pseudo R2 Mc Fadden	Significancia Global			Prueba de Hosmer - Lemeshow		
	X ²	gl.	Sig.	H-L	gl	Sig.
0.0486	280.2235	28	0	11267	8	2.2e-16

*Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor*

Además podemos observar la capacidad del modelo para clasificar los datos, a continuación en el gráfico 4.3, podemos apreciar el gráfico de violín del modelo.

Figura 4. 3 Gráfico de violín del modelo

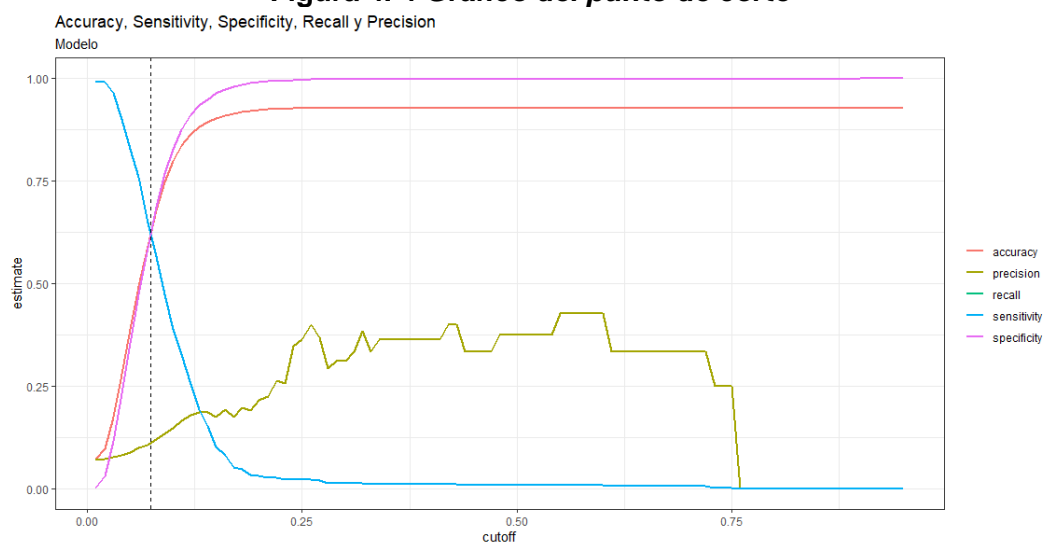


*Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor*

Además es posible evaluar la capacidad predictiva del modelo a través de la construcción de la tabla de confusión, en esta tabla se muestran los valores observados de la variable condición de egreso, distinguiendo entre vivos y muertos. Así como los valores estimados por el modelo para los cuales se establece un punto

de corte para las probabilidades, en el caso de nuestro modelo, el punto de corte no sigue los estándares asignándoles el valor de 0.5, dado que la probabilidad de éxito es baja, para poder calcularlo se realiza de manera gráfica, en la gráfica 4.4. se observa el gráfico de la función del punto de corte y obtenemos el valor de 0,074.

Figura 4. 4 Gráfico del punto de corte



Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2001- 2018 - INEC
Elaboración: Autor

El modelo estimado arroja un 63.7%% de aciertos de manera global , donde están bien clasificados el 95,61% de los pacientes que egresaron vivos y el 11,52% de los pacientes que egresaron muertos; estos valores se justifican, dado el porcentaje de personas que fallecen de una enfermedad isquémica es baja, por ese particular la capacidad del modelo de predecir la probabilidad de fallecer de una enfermedad isquémica del corazón si es considerable, teniendo en cuenta que utilizamos variables sociales, demográficas y solo la edad como dato del paciente.

Tabla 4. 18 Tabla de clasificación

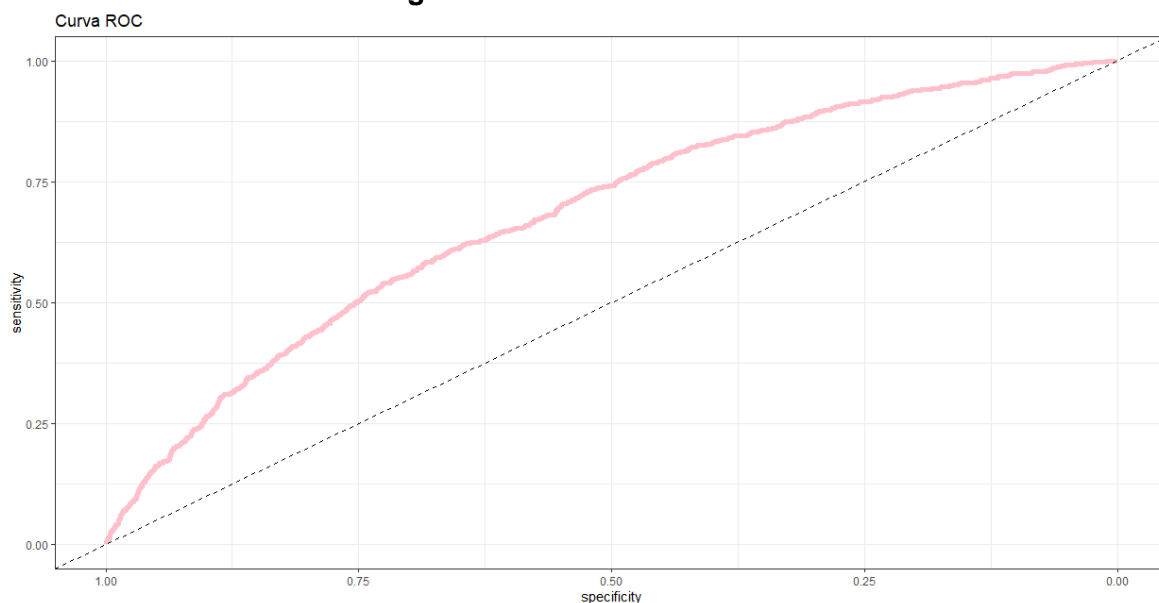
Observados	Pronosticado		Porcentaje correcto
	No	Si	
No	6692	307	95.61%
Si	3776	492	11.52%
Total	Porcentaje global		63,76%

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor

La curva ROC nos muestra gráficamente la sensibilidad y especificidad del Modelo, el área bajo de la curva se denomina AUC, y toma valores entre 0 y 1, mientras más cercano al valor de 1 nos indica que el modelo permite predecir de mejor forma

los dos posibles estados de la variable de respuesta. El AUC del modelo estimado es de 0.654, para el caso particular de nuestra variable de respuesta es aceptable.

Figura 4. 5 Gráfico curva AUC



Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor

4.3.6. Ejemplos prácticos del modelo

A continuación vamos a predecir la probabilidad de que un paciente llegue a un centro de salud, sea diagnosticado de alguna enfermedad isquémica del corazón, tratado y dado el alta y su condición de egreso sea muerto.

Para el ejemplo tenemos los datos de 3 pacientes, identificados como A,B y C, los pacientes A y B tendrán datos similares en su registro solo cambiará su sexo, el paciente C va a tener datos de registro de un paciente no típico, en la tabla 4.19 está detallado la ficha de los pacientes.

Tabla 4. 19 Ejemplos del modelo

	Sexo	Provincia de residencia	Área de residencia	Días de estadía	Probabilidad condición de egreso muerto
Paciente A	Masculino	Guayas	rural	4	2,46%
Paciente B	Femenino	Guayas	rural	4	3,14%
Paciente C	Femenino	Chimborazo	urbana	120	21,8%

Fuente: Registro Estadístico Camas y Egresos Hospitalarios 2018 - INEC
Elaboración: Autor

Podemos observar que los pacientes A y B, residentes de la misma provincia, comparten la misma área de residencia, estuvieron internados exactamente los mismos días pero son de diferentes sexo, su condición de egreso es diferente, la paciente A de sexo masculino tiene una probabilidad de que su condición de egreso sea muerta del 2.46% en cambio el paciente B de sexo femenino la probabilidad de que su condición de egreso sea muerto es de 3.14%.

El paciente C, pertenece a la provincia del Chimborazo, sexo femenino, de zona rural, permaneció 120 días en el centro de salud y cerebro su cumpleaños hasta que le dieron el alta, la probabilidad que el paciente C su condición de egreso sea muerto es del 21.8%.

CAPÍTULO 5

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

Las variables que resultaron estadísticamente significativas para el modelo nos permiten calcular su probabilidad de ocurrencia, intervalo de confianza en relación a su variable de control, mientras que las variables no significativas no sirven para el propósito del modelo.

El modelo de regresión permitió reconocer el efecto de las variables predictoras escogidas sobre la variable de respuestas, se concluye que las variables sexo, las provincias de residencia: Guayas, Pichincha, El Oro, Chimborazo, Santo Domingo, Carchi, Azuay y sector de residencia urbano, actúan como protectoras a mortalidad la causada por enfermedades isquémicas del corazón, mientras que la edad y los días de estadías son variables de alto riesgo, lo podemos apreciar en los ejemplos del modelo, como la edad y los días de estadías incrementan probabilidad de que la condición de egreso de un paciente sea muerto.

El sexo más propenso a que su condición de egreso sea muerto, es el sexo femenino, en las asociaciones bivariantes observamos que aunque la población de mujeres es menor que la de hombres, su porcentaje de muertas por enfermedades isquémicas del corazón es más elevado que el de los hombres y en modelo lo reafirmamos.

Los centros de salud Públicos no presentan la mayor probabilidad de que la condición de egreso de los pacientes sean muertos, los centro Privados sin fines de lucros tienen mayor probabilidad que los pacientes tengan una condición de egreso muertos y los menos probables son los centro Privados con fines de lucro.

Como se pudo observar en el modelo, en los ejemplos propuestos, la edad es un factor de alto riesgo, si una persona es diagnosticada con alguna enfermedad isquémica del corazón, la probabilidad de su condición de egreso aumenta **1,022** veces por cada año cumplido, lo podemos observar en la estadística descriptiva de la variable que edad, que mayor concentración de pacientes muertos es partir de los 50 años.

La variable más crítica y de mayor riesgo son los días de estadías de los pacientes diagnosticados, tratados de alguna enfermedad isquémica del corazón, cada día que permanece el paciente se incrementa la probabilidad en un 1,02% de que la condición de egreso del paciente sea muerto, en el ejemplo lo podemos observar claramente.

5.2. Recomendaciones

La investigación la realice con datos del INEC, me tomó mucho tiempo poder acceder a la base de datos, tener el asesoramiento adecuado, cabe señalar que las personas que me atendieron fueron muy amables, pero la parte técnica por más que solicite no tuve el soporte necesario. La investigación es sobre un tema de salud con variables sociales, demográficas, para entender la problemática y poder aterrizarla de la mejor manera en las herramientas estadísticas tuve que asesorarme en especialistas de la salud, recomiendo a los investigadores tener un especialista como asesor de sus proyectos.

Observamos problemas de sub-registro en la variable etnia existe un 26,6% de los pacientes que se capturaron mal los datos, es un error importante, no podemos ver si existe relación entre los pacientes diagnosticados con enfermedades isquémicas del corazón y su etnia. Mejorar la captura de los datos en los centros de salud, para disminuir el porcentaje de no registro.

Recomendación a la comunidad científica, universitaria del Ecuador, son necesarios cursos, seminarios sobre la preparación de los datos, la limpieza de los mismos, los protocolos que utilizan las instituciones sean privadas o públicas.

El proyecto utiliza variables sociales, demográficas, recomendando se realicen más investigaciones utilizando ese tipo de variables, tal vez los resultados iniciales no convencen a la comunidad científica, pero es el primer paso, en la rama donde realicemos una investigación, es más fácil acceder a este tipo de variables y al modelarlas podemos tener una primera idea del comportamiento de la variable de estudio.

6. Referencias

- Barcelona, C. (2021). *Clinic Barcelona - Hospital Universitario*. Obtenido de <https://www.clinicbarcelona.org/asistencia/enfermedades/cardiopatia-isquemica/definicion>
- Canavos, G. (1988). *Probabilidad y Estadística. Aplicaciones y Métodos*. México: McGraw-Hill.
- Chile, U. d. (2017). SELECCIÓN, IMPLEMENTACIÓN Y VALIDACIÓN DE MODELOS DE PREDICCIÓN. En M. J. ARANDA.
- Clínica, C. d. (2018). Mortalidad por enfermedades isquémicas del corazón en Ecuador, 2001-2016: estudio de tendencias. *Revista Medica de Chile*, 850 - 856.
- Costa, U. d. (2015). *ANÁLISIS DE TÉCNICAS DE MD EN DIAGNÓSTICO DE ENFERMEDADES CARDIOVASCULARES*. Barranquilla.
- ESTEBAN, L. A. (marzo de 2017). Las enfermedades crónicas y la ineludible transformación de los sistemas de salud en América Latina. *Revista Cubana Salud Pública*, 43(1), 68-74. Recuperado el 20 de febrero de 2021, de <http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-34662017000100007&lng=es&nrm=iso>
- Everitt, B., & Dunn, G. (2001). *Análisis de Datos Multivariados Aplicado*. Londres: Wiley.
- Federation, W. H. (3 de Junio de 2016). *World Heart Federation*. Obtenido de <https://www.world-heart-federation.org/wp-content/uploads/2017/05/spanish-press-release.pdf>
- González Guzmán , R., & Alcalá Ramírez, J. (Septiembre - Octubre de 2010). Enfermedad isquémica del corazón, epidemiología y prevención. *Revista de la Facultad de Medicina de la UNAM*, 53(5), 43. Obtenido de <https://www.medigraphic.com/pdfs/facmed/un-2010/un105h.pdf>
- Hosmer, D., Lemeshow, S., & ed., 2. (2000). *Applied Logistic Regression*. Jhon Wiley & Sons, Inc.
- INEC. (2018). Recuperado el enero de 12 de 2021, de INEC: https://www.ecuadorencifras.gob.ec/documentos/web-inec/Poblacion_y_Demografia/Nacimientos_Defunciones/2016/Presentacion_Nacimientos_y_Defunciones_2016.pdf
- INEC. (2018). *Boletín N°-01-2019-Registro Estadístico de Camas y Egresos Hospitalarios*. Quito. Obtenido de www.ecuadorencifras.gob.ec
- INEC. (2018). *Boletín técnico N°01-2019*. Quito. Obtenido de www.ecuadorencifras.gob.ec
- INEC. (2018). *Registro Estadístico de Camas y Egresos Hospitalarios del Ecuador*. INEC. Quito: INEC. Obtenido de <https://www.ecuadorencifras.gob.ec/camas-egresos/>
- INEC. (2018). *Registro Estadístico de Defunciones Generales*. Quito: INEC.
- INEC. (2019). *BOLETIN TECNICO No - 01 -2019 - ECEH*.
- INFOMED. (01 de 01 de 2020). Tendencia e impacto de la mortalidad por enfermedades cardiovasculares en México, 1990-2015. (C. A. Cervantes, F. L. México, & México, Edits.) *Revista Cubana de Salud Publica*, 45(4). Obtenido de <http://www.revsaludpublica.sld.cu/index.php/spu/article/view/1081>
- Jadraque, M. (1999). *Cardiopatía isquémica: Angina de pecho, infarto de miocardio*. La Habana: Capitel Científico - Técnica.
- JoAnne Epping-Jordan, S. P. (2002). *ORGANIZACIÓN PANAMERICANA DE LA SALUD*. Cuidado Innovador para las condiciones crónicas: Agenda para el cambio,

- Organización Mundial de la Salud. Obtenido de https://www.who.int/chp/knowledge/publications/iccc_spanish.pdf?ua=1
- McKenna, C., & Smith, T. (2013). *A Comparison of Logistic Regression Pseudo R Indices*.
- OMS. (2017). *ORGANIZACIÓN MUNDIAL DE LA SALUD*. Obtenido de https://www.who.int/cardiovascular_diseases/world-heart-day-2017/en/
- OMS. (2018). *CIE - 10*.
- OPS. (20 de Noviembre de 2013). *ORGANIZACIÓN PANAMERICANA DE LA SALUD*. Obtenido de https://www.paho.org/ecu/index.php?option=com_content&view=article&id=1072:noviembre-20-2013&Itemid=972
- Paul Newbold, W. L. (2008). *Estadística para Administración y Economía*. Madrid: PEARSON EDUCACIÓN, S.A.,.
- Ramírez, R. G. (2010). Enfermedad isquémica del corazón, epidemiología y prevención. *Revista de la Facultad de Medicina de la UNAM*.
- Rubio Hurtado, M. J., & Berlanga Silvente, V. (2012). Cómo aplicar las pruebas paramétricas bivariadas t de Student y ANOVA en SPSS. *REIRE. Revista d'Innovació i Recerca en Educació*, 5(2), 83-100. Obtenido de <http://www.ub.edu/ice/reire.htm>
- Salud, O. P. (2014). *PLAN DE ACCIÓN PARA LA PREVENCIÓN Y EL CONTROL DE LAS ENFERMEDADES NO TRANSMISIBLES EN LAS AMERICAS 2013-2019*. Obtenido de <https://www.paho.org/hq/dmdocuments/2015/plan-accion-prevencion-control-ent-americas.pdf>
- Vera, F. S. (2018). Tablas de mortalidad de Ecuador continental . *Papeles de población*.
- Zurita Herrera, G. (2008). Probabilidades y estadísticas. 515-518. Obtenido de <https://isbn.cloud/9789978310557/probabilidades-y-estadisticas/>

7. Apéndices y anexos

```
### LECTURA DE DATOS
library(readxl)
datos <- read_excel("basef.xlsx")

####PRINCIPALES ESTADISTICAS DESCRIPTIVAS
### variable: provincia de ubicación
str(datos)
attach(datos)
names(datos)
barplot(prop.table(table(datos$prov_ubi)),ylim=c(0,0.5),
        ylab = "Frecuencia Relativa", xlab="Provincia de ubicación",
        main="Diagrama de barras")
### variable: area de ubicación
barplot(prop.table(table(datos$area_ubi)),ylim=c(0,1),
        ylab = "Frecuencia Relativa", xlab="Área de ubicación",
        main="Diagrama de barras")
### variable: SECTOR
barplot(prop.table(table(datos$sector)),ylim=c(0,0.5),
        ylab = "Frecuencia Relativa", xlab="Sector",
        main="Diagrama de barras")
#### variable sexo
barplot(prop.table(table(datos$sexo)),ylim=c(0,1),
        ylab = "Frecuencia Relativa", xlab="Sexo",
        main="Diagrama de barras")
#### variable edad
hist(datos$edad,ylab = "Frecuencia Relativa", xlab="Edad", ylim=c(0,500),
     main="Histograma",breaks = 104)
rug(datos$edad)
summary(datos$edad)
library(psych)
describe(datos$edad)
"vars  n mean  sd median trimmed  mad min max range skew
X1  1 16094 51.65 25.08  59 53.35 22.24  1 104 103 -0.61
kurtosis se
X1  -0.77 0.2"
```



```

#### variable etnia
barplot(prop.table(table(datos$etnia)),ylim=c(0,1),
        ylab ="Frecuencia Relativa", xlab="Etnia",
        main="Diagrama de barras")
#### variable área de residencia
barplot(prop.table(table(datos$area_res)),ylim=c(0,1),
        ylab ="Frecuencia Relativa", xlab="Etnia",
        main="Área de residencia")
#### variable provincia de residencia
barplot(prop.table(table(datos$prov_res)),ylim=c(0,1),
        ylab ="Frecuencia Relativa", xlab="Etnia",
        main="Provincia de residencia")
#### variable región de residencia
barplot(prop.table(table(datos$reg_res)),ylim=c(0,1),
        ylab ="Frecuencia Relativa", xlab="Etnia",
        main="Región de residencia")
#### variable días de estadía
hist(datos$dia_estad,ylab ="Frecuencia Relativa", xlab="Días de estadía",
      main="Histograma",breaks = 104)
rug(datos$dia_estad)
summary(base$dia_estad)
" Min. 1st Qu. Median Mean 3rd Qu. Max.
1.000 2.000 5.000 8.296 9.000 417.000 "
describe(base$dia_estad)
" vars n mean sd median trimmed mad min max range skew
X1 1 16094 8.3 12.37 5 5.91 4.45 1 417 416 8.98
kurtosis se
X1 189.49 0.1 "

##### variable condición de egreso
barplot(prop.table(table(datos$con_egre)),ylim=c(0,1),
        ylab ="Frecuencia relativa", xlab="Condición de egreso",
        main="Diagrama de barras")

#### ASOCIACIONES BIVARIANTES
attach(datos)
names(datos)
library(descr)
datos$prov_ubi<-as.factor(datos$prov_ubi)
datos$area_ubi<-as.factor(datos$ area_ubi)
datos$prov_ubi<-as.factor(datos$prov_ubi)

```

```

datos$sector<-as.factor(datos$sector)
datos$etnia<-as.factor(datos$etnia)
datos$prov_res<-as.factor(datos$prov_res)
datos$area_res<-as.factor(datos$area_res)
datos$reg_res<-as.factor(datos$reg_res)

CrossTable(datos$prov_ubi, datos$con_egre, chisq = TRUE,
  prop.chisq = FALSE, prop.c = FALSE, prop.t = FALSE)
CrossTable(datos$area_ubi, datos$con_egre, chisq = TRUE,
  prop.chisq = FALSE, prop.c = FALSE, prop.t = FALSE)
CrossTable(datos$sector, datos$con_egre, chisq = TRUE,
  prop.chisq = FALSE, prop.c = FALSE, prop.t = FALSE)
CrossTable(datos$etnia, datos$con_egre, chisq = TRUE,
  prop.chisq = FALSE, prop.c = FALSE, prop.t = FALSE)
CrossTable(datos$sexo, datos$con_egre, chisq = TRUE,
  prop.chisq = FALSE, prop.c = FALSE, prop.t = FALSE)
CrossTable(datos$prov_res, datos$con_egre, chisq = TRUE,
  prop.chisq = FALSE, prop.c = FALSE, prop.t = FALSE)
CrossTable(datos$area_res, datos$con_egre, chisq = TRUE,
  prop.chisq = FALSE, prop.c = FALSE, prop.t = FALSE)
CrossTable(datos$reg_res, datos$con_egre, chisq = TRUE,
  prop.chisq = FALSE, prop.c = FALSE, prop.t = FALSE)

leveneTest(y = datos$con_egre ,group = datos$edad, center = "mean")

test <- t.test(datos$con_egre, datos$edad) # Prueba t de Student
print(test)
leveneTest(y = datos$con_egre ,group = datos$edad, center = "mean")
test <- t.test(datos$con_egre, datos$dia_estad) # Prueba t de Student
print(test)
library(kableExtra)
library(ggplot2)
library(tidyverse)
library(GGally)
library(caret)
library(gridExtra)
library(grid)
library(broom)
library(pROC)
library(caret)
library(dplyr)

##bf base para la regresión

```

```

bf<- read_excel("C:/Users/User/Desktop/base_reg_F.xlsx")
bf
#### GRAFICO PARA VARIABLES PREDICTORAS VS CONDICION DE
EGRESO
bf$sector=as.factor(bf$sector)
bf$sexo=as.factor(bf$sexo)
bf$prov_res=as.factor(bf$prov_res)
bf$area_res=as.factor(bf$area_res)
bf$region_res=as.factor(bf$region_res)
bf$con_egre=as.factor(bf$con_egre)
View(bf)
#kable(bf) %>% kable_styling(bootstrap_options = c("striped", "hover"))
summary(bf)
sector  sexo      edad      region_res  prov_res  area_res
1:7790  1:6280  Min. : 1.00  COSTA :8611  14      :4383  1: 1869
2:3819  2:9814  1st Qu.: 35.00  INSULAR: 43  15      :2827  2:14225
3:4485      Median : 59.00  ORIENTE: 474  11      :1921
Mean   : 51.65  SIERRA :6966  23      :1565
3rd Qu.: 70.00          12      : 897
Max.   :104.00          18      : 895
(Other):3606
dia_estad  con_egre
Min. : 1.000  0:14953
1st Qu.: 2.000  1: 1141
Median : 5.000
Mean   : 8.296
3rd Qu.: 9.000
Max.   :417.000

ggpairs(bf,cardinality_threshold = 24,mapping = aes(colour= con_egre)) +
theme(axis.text.x = element_text(angle = 90, hjust = 1)) + theme_bw()

##### PARTICIONAMOS LOS DATOS EN DOS GRUPOS UNO PARA EL
MODELO Y EL OTRO PARA VALIDAR Y ELEGIR EL PUNTO DE CORTE
trainIndex <- createDataPartition(bf$con_egre, p = .7,
list = FALSE,
times = 1)
tentrenamiento <- bf [ trainIndex,] %>% as_tibble()
tvalidacion <- bf[-trainIndex,] %>% as_tibble()

clases_train <- bf %>% group_by(con_egre) %>% summarise(numero_casos=n())
clases_entrenamiento <- tentrenamiento %>% group_by(con_egre) %>%
summarise(numero_casos=n())

```

```
clases_validacion <- tvalidacion %>% group_by(con_egre) %>%
summarise(numero_casos=n())
```

```
g1 <- ggplot(clases_train, aes(x=con_egre, y=numero_casos, fill = con_egre)) +
geom_col() + labs(subtitle = "Datos", x = "Estado de Egreso", y = "Número de
casos") + guides(fill = "none") + theme_bw()
g2 <- ggplot(clases_entrenamiento, aes(x=con_egre, y=numero_casos, fill =
con_egre)) + geom_col() + labs(subtitle = "Datos Entrenamiento", x = "Estado de
Egreso", y = "Número de casos") + guides(fill = "none") + theme_bw()
g3 <- ggplot(clases_validacion, aes(x=con_egre, y=numero_casos, fill =
con_egre)) + geom_col() + labs(subtitle = "Datos Validación", x = "Estado de
Egreso", y = "Número de casos") + guides(fill = "none") + theme_bw()
```

```
library(grid)
grid.arrange(
  g1,
  g2,
  g3,
  nrow = 1,
  top = "Balanceo de clases en los conjuntos de datos",
  bottom = textGrob(
    "1 corresponde a fallecido",
    gp = gpar(fontface = 3, fontsize = 9),
    hjust = 1,
    x = 1
  )
)
```

MODELOS LOGISTICOS

```
mf2<-glm(data = tentrenamiento, con_egre ~ sector + sexo + edad + prov_res +
area_res+ dia_estad, family = binomial)
```

```
summary(mf2)
```

Call:

```
glm(formula = con_egre ~ sector + sexo + edad + prov_res + area_res +
    dia_estad, family = binomial, data = tentrenamiento)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.9826	-0.4204	-0.3464	-0.2703	2.9161

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.006099	0.235345	-12.773	< 2e-16 ***
sector2	-0.148484	0.099723	-1.489	0.13650

```

sector3    0.147452  0.098567  1.496  0.13466
sexo2     -0.235694  0.075567 -3.119  0.00181 **
edad      0.021614  0.001812 11.930 < 2e-16 ***
prov_res2  0.074058  0.523272  0.142  0.88745
prov_res3  0.001922  0.402163  0.005  0.99619
prov_res4 -0.137765  0.301672 -0.457  0.64791
prov_res5 -0.278991  0.300963 -0.927  0.35393
prov_res6  0.099917  0.358345  0.279  0.78038
prov_res7 -0.246128  0.439720 -0.560  0.57566
prov_res8 -0.450724  0.405396 -1.112  0.26622
prov_res9  0.113135  0.769931  0.147  0.88318
prov_res10 -0.477110  0.286512 -1.665  0.09587 .
prov_res11 -0.250742  0.201628 -1.244  0.21365
prov_res12 -0.196681  0.230225 -0.854  0.39294
prov_res13 -0.353165  0.349894 -1.009  0.31281
prov_res14 -0.567550  0.191339 -2.966  0.00302 **
prov_res15 -0.601749  0.196006 -3.070  0.00214 **
prov_res16 -0.375815  0.316069 -1.189  0.23443
prov_res17 -0.551652  0.768489 -0.718  0.47286
prov_res18 -0.858040  0.256192 -3.349  0.00081 ***
prov_res19 -0.796590  0.337476 -2.360  0.01825 *
prov_res20 -1.025856  1.036993 -0.989  0.32254
prov_res21 -1.143611  0.376213 -3.040  0.00237 **
prov_res22 -1.471583  0.745076 -1.975  0.04826 *
prov_res23 -1.065562  0.230190 -4.629  3.67e-06 ***
prov_res24 -0.803072  0.548398 -1.464  0.14309
area_res2 -0.359062  0.110013 -3.264  0.00110 **
dia_estad  0.020168  0.002584  7.806  5.91e-15 ***

```

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 5768.7 on 11266 degrees of freedom
Residual deviance: 5488.5 on 11237 degrees of freedom
AIC: 5548.5

Number of Fisher Scoring iterations: 6

```

mf3<-glm(data = tentrenamiento, con_egre ~ sector + sexo + edad + area_res +
dia_estad, family = binomial)
> summary(mf3)

```

Call:

```

glm(formula = con_egre ~ sector + sexo + edad + area_res + dia_estad,

```

family = binomial, data = tentrenamiento)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.0274	-0.4219	-0.3599	-0.2804	2.7963

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.522027	0.160292	-21.973	< 2e-16 ***
sector2	-0.120496	0.094658	-1.273	0.20303
sector3	0.141123	0.092350	1.528	0.12648
sexo2	-0.229623	0.075137	-3.056	0.00224 **
edad	0.021033	0.001805	11.651	< 2e-16 ***
area_res2	-0.331922	0.105877	-3.135	0.00172 **
dia_estad	0.020303	0.002580	7.871	3.53e-15 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 5768.7 on 11266 degrees of freedom
Residual deviance: 5548.5 on 11260 degrees of freedom
AIC: 5562.5

Number of Fisher Scoring iterations: 6

```
mf4<-glm(data = tentrenamiento, con_egre ~ sector + sexo + edad + region_res +  
area_res + dia_estad, family = binomial)  
> summary(mf4)
```

Call:

```
glm(formula = con_egre ~ sector + sexo + edad + region_res +  
area_res + dia_estad, family = binomial, data = tentrenamiento)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.9997	-0.4216	-0.3586	-0.2798	2.8309

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.457651	0.171123	-20.206	< 2e-16 ***
sector2	-0.138811	0.096954	-1.432	0.15222
sector3	0.132746	0.092725	1.432	0.15226
sexo2	-0.235339	0.075205	-3.129	0.00175 **
edad	0.021203	0.001807	11.733	< 2e-16 ***
region_resINSULAR	-0.553849	1.023573	-0.541	0.58844
region_resORIENTE	0.288266	0.197723	1.458	0.14486

```

region_resSIERRA -0.120358  0.080347 -1.498  0.13414
area_res2      -0.353403  0.108225 -3.265  0.00109 **
dia_estad      0.020008  0.002579  7.759  8.59e-15 ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 5768.7 on 11266 degrees of freedom
Residual deviance: 5542.9 on 11257 degrees of freedom
AIC: 5562.9
Number of Fisher Scoring iterations: 6
##### PORCENTAJE DE DEVIANCE EXPLICADA

```

```

de2=(1-(mf2$deviance/mf2$null.deviance))
de2
0.04857659
de3=(1-(mf3$deviance/mf3$null.deviance))
de3
0.03816384
de4=(1-(mf4$deviance/mf4$null.deviance))
de4
0.0367927

```

```

#####INTERVALO DE CONFIANZA DE COEFICIENTES
Number of Fisher Scoring iterations: 6
confint(mf2)
Waiting for profiling to be done...
      2.5 %    97.5 %
(Intercept) -3.47843638 -2.55485589
sector2      -0.34555822  0.04554780
sector3      -0.04690457  0.33962852
sexo2        -0.38350839 -0.08719187
edad         0.01809738  0.02520180
prov_res2    -1.06415613  1.02656798
prov_res3    -0.84082054  0.75229672
prov_res4    -0.74674206  0.44193706
prov_res5    -0.88675008  0.29916967
prov_res6    -0.63897249  0.77680798
prov_res7    -1.18632958  0.56141716
prov_res8    -1.30764070  0.29919951
prov_res9    -1.75658154  1.41977626
prov_res10   -1.05215408  0.07617179
prov_res11   -0.63713065  0.15519124
prov_res12   -0.64489510  0.25998591
prov_res13   -1.07756912  0.30542117
prov_res14   -0.93193612 -0.18000112

```

```

prov_res15 -0.97629620 -0.20600495
prov_res16 -1.01988443 0.22687025
prov_res17 -2.41975791 0.75037131
prov_res18 -1.36556150 -0.35749442
prov_res19 -1.49289589 -0.15970466
prov_res20 -3.92523816 0.57915523
prov_res21 -1.93497820 -0.44532929
prov_res22 -3.31163230 -0.23628843
prov_res23 -1.51488585 -0.60986083
prov_res24 -2.04088655 0.16406551
area_res2 -0.57140081 -0.13981614
dia_estad 0.01518551 0.02531336

```

```

#####INTERVALO DE CONFIANZA DE ODDS
exp(cbind(OR = coef(mf2), confint(mf2,level=0.95)))

```

Waiting for profiling to be done...

```

      OR   2.5 %   97.5 %
(Intercept) 0.04948433 0.03085562 0.07770343
sector2      0.86201399 0.70782512 1.04660103
sector3      1.15887784 0.95417845 1.40442578
sexo2        0.79002210 0.68146636 0.91650123
edad         1.02184949 1.01826213 1.02552205
prov_res2    1.07686879 0.34501888 2.79146901
prov_res3    1.00192389 0.43135643 2.12186777
prov_res4    0.87130344 0.47390801 1.55571782
prov_res5    0.75654657 0.41199252 1.34873844
prov_res6    1.10507894 0.52783450 2.17452007
prov_res7    0.78182249 0.30533994 1.75315525
prov_res8    0.63716683 0.27045739 1.34877869
prov_res9    1.11978320 0.17263400 4.13619492
prov_res10   0.62057448 0.34918477 1.07914795
prov_res11   0.77822300 0.52880758 1.16788129
prov_res12   0.82145276 0.52471758 1.29691181
prov_res13   0.70246105 0.34042205 1.35719650
prov_res14   0.56691251 0.39379055 0.83526928
prov_res15   0.54785273 0.37670375 0.81382905
prov_res16   0.68672918 0.36063662 1.25466706
prov_res17   0.57599738 0.08894315 2.11778622
prov_res18   0.42399249 0.25523732 0.69942660
prov_res19   0.45086391 0.22472094 0.85239549
prov_res20   0.35848960 0.01973744 1.78453028
prov_res21   0.31866635 0.14442742 0.64061329
prov_res22   0.22956174 0.03645662 0.78955291

```



```
prov_res23 0.34453422 0.21983328 0.54342649
prov_res24 0.44795076 0.12991349 1.17829150
area_res2 0.69833085 0.56473380 0.86951809
dia_estad 1.02037243 1.01530140 1.02563646
```

```
##### AJUSTE DEL MODELO
```

```
### 1. PSEUDO R2
```

```
library(DescTools)
PseudoR2(mf2,c("Mac Fadden", "Nagel"))
PseudoR2(mf2)
PseudoR2(mf2,c("Mac Fadden", "Nagel"))
Nagelkerke
0.06130324
> PseudoR2(mf2)
  McFadden
0.04857659
```

```
# Bondad de ajuste del modelo Devianza y Chi2
```

```
> dev <- mf2$deviance
> nullDev <- mf2$null.deviance
> modelChi <- nullDev - dev
> modelChi
[1] 280.2235
> chigl <- mf2$df.null - mf2$df.residual
> chisq.prob <- 1 - pchisq(modelChi, chigl)
> chisq.prob
[1] 0
Chigl|
[1] 29
```

```
#### Hosmer y Lemeshow
```

```
library(ResourceSelection)
```

```
hoslem.test(tentrenamiento$con_egre,fitted(mf2))
Hosmer and Lemeshow goodness of fit (GOF) test
```

```
data: tentrenamiento$con_egre, fitted(mf2)
X-squared = 11267, df = 8, p-value < 2.2e-16
```

```
##### ELECCIÓN DE PUNTO DE CORTE CON DATOS DE VALIDACIÓN
```

```

p2=predict(mf2,type = "response",newdata = tvalidacion)
p2=as.data.frame(p2)

dataconfusion=c(tvalidacion,p2)
dataconfusion=as.data.frame(dataconfusion)
prediccion=if_else(dataconfusion$p2>0.5, 1, 0)
prediccion=as.factor(prediccion)
c1=confusionMatrix(tvalidacion$con_egre,prediccion,positive = "1")

c1%>%
  tidy()%>%
  select(term,estimate)%>%
  filter(term %in% c('accuracy', 'sensitivity', 'specificity', 'precision','recall'))

#####GRÁFICO DE VIOLIN
p1=as.data.frame(p1)
dataviolin=c(tentrenamiento,p1)
dataviolin=as.data.frame(dataviolin)
violin = ggplot(dataviolin, aes(x=dataviolin$con_egre, y=dataviolin$p1,
group=con_egre, fill=con_egre)) +
  geom_violin() +
  theme_bw() +
  guides(fill=FALSE) +
  labs(title='Gráfico de violín del Modelo ', x ='Estado de egreso', y='Probabilidad
predicha')
violin

##### FUNCIÓN PARA GRAFICAR PUNTO DE CORTE

prediction_metrics <- function(cutoff, predictions=dataconfusion){
  table <- predictions %>%
    mutate(predicted_class=if_else(dataconfusion$p2>cutoff, 1, 0) %>% as.factor(),
           con_egre= factor(dataconfusion$con_egre))

  confusionMatrix(table$predicted_class, table$con_egre, positive = "1") %>%
    tidy() %>%
    select(term, estimate) %>%
    filter(term %in% c('accuracy', 'sensitivity', 'specificity', 'precision','recall')) %>%
    mutate(cutoff=cutoff)
}

```

```

cutoffs = seq(0.01,0.95,0.01)

logit_pred = map_dfr(cutoffs, prediction_metrics)%>%
mutate(term=as.factor(term))

##### Probamos valores en xintercept hasta encontrar el punto de corte
ggplot(logit_pred, aes(cutoff, estimate, group=term, color=term)) +
geom_line(size=1) + theme_bw() +
  labs(title= 'Accuracy, Sensitivity, Specificity, Recall y Precision', subtitle=
'Modelo', color="") + geom_vline(xintercept=0.074, linetype="dashed", color =
"black")
## punto de corte es 0.074, probamos valores hasta encontrarlo

```

VALORES PREDICHOS MATRIZ DE CONFUSION

```

tentrenamiento
tentrenamiento$predict<-mf2$fitted.values
tentrenamiento$predict
View(tentrenamiento)

```

```

mo1<-tentrenamiento$con_egre
p1<-tentrenamiento$predict

```

```

a1<-data.frame(mo1,p1)
a1
View(a1)

```

```

confusionMatrix(mo1,p1,
  threshold = 0.074)

```

```

confusionMatrix(mo1,p1,
+   threshold = 0.074)
  0  1
0 6692 307
1 3776 492

```

```

obscorrectas<- 1 -
misClassError(tentrenamiento$con_egre,tentrenamiento$predict,
  threshold = 0.074)

```

```

obscorrectas
obscorrectas

```

```
[1] 0.6376
```

```
### porcentaje correcto egresos vivos
```

```
(6692)/(6692 + 307)
```

```
[1] 0.9561366
```

```
#### porcentaje correcto egresos muertos
```

```
492/(3376 + 492)
```

```
0.1271975
```

```
#### porcentaje global del modelo
```

```
(6692 + 492)/( 6692 + 307 + 3376 + 492)
```

```
0.6376142718
```

```
#####CURVA ROC
```

```
p1=predict(mf2,type = "response")
```

```
roc1=roc(response=tentrenamiento$con_egre,predictor=p1)
```

```
ggroc(roc1, color = "pink", size = 2) + geom_abline(slope = 1, intercept = 1,  
linetype ='dashed') + labs(title = "Curva ROC") + theme_bw()
```

```
roc1$auc
```

```
0.654
```