



ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Instituto de Ciencias Matemáticas

Ingeniería en Estadística Informática

“Validación y Calibración de Indicadores Discriminantes para Pacientes Hospitalarios con Problemas Cardiovasculares: Hospital Luis Vernaza de Guayaquil”

TESIS DE GRADO

Previa a la obtención del Título de:

INGENIERO EN ESTADÍSTICA INFORMÁTICA

Presentada por:

Abel Stalin Flores Laaz

GUAYAQUIL-ECUADOR

AÑO

2005

AGRADECIMIENTO

Al Mat. Fernando Guerrero por haberme brindado su amistad, por su gran apoyo en la realización de esta investigación y en otros temas de esta misma índole, que me han sido de mucha utilidad en el inicio de mi vida como profesional.

Al Mat. Fernando Sandoya que al igual que mi Director de Tesis me aportó con sus conocimientos, para que yo, pueda desenvolverme de manera eficaz y eficiente en temas investigativos.

A los múltiples funcionarios del Hospital Luis Vernaza, en especial a las personas que conforman el departamento de estadística. (Eduardo Moyano, Xavier Rosas y Livino Armijos).

A mis grandes amigos Michelle Echanique, Jorge Félix y Julio Buzetta por su gran ayuda en la recolección de datos.

A unos de los seres más espectaculares que Dios me pudo poner en mi camino "Guadalupe Bonilla", por ser mi soporte en todo momento.

A todos mis amigos que han vivido junto a mí, este fascinante mundo Universitario, les agradezco de corazón por estar siempre conmigo.

DEDICATORIA

Quisiera comenzar dedicando este trabajo a Dios, ya que si no fuera por él, no hubiera podido llevar a cabo este tema de investigación. También por darme el mejor de los regalos: Mi Familia.

A mi papá, Luis Abel Flores Arroba, por su gran confianza y ayuda en todo momento, por haberme enseñado e inculcado lo importante que es la responsabilidad y la honestidad. Por sus consejos los cuales me han ayudado de mucho en mi vida pre-profesional y especialmente al enseñarme afrontar cualquier obstáculo que se presente en mi vida.

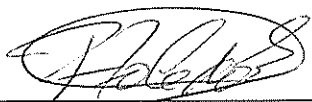
A mi mamá, Mariana Laaz, por su apoyo incondicional, sus consejos y su amor hacia mí durante toda la vida. Es el ángel más precioso que Dios me ha dado como madre.

A mis hermanos: Antonio , Pedro y Leonor, ya que nunca dudaron de mí y siempre estuvieron a mi lado.

A Samy, Denisse y Vale por ser las personitas más lindas del mundo, debido a que con la llegada de ellas mi familia y yo somos felices.

Y a toda mi familia por compartir este Triunfo conmigo.

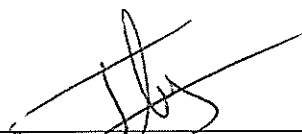
TRIBUNAL DE GRADUACIÓN



Ing. Roberto Toledo
PRESIDENTE DEL TRIBUNAL



Mat. César Fernando Guerrero
DIRECTOR DE TESIS



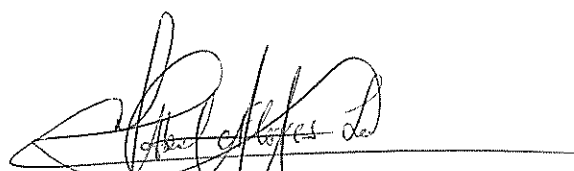
Ing. Soraya Solís
VOCAL



Ing. Enrique Bayot
VOCAL

DECLARACIÓN EXPRESA

“La responsabilidad del contenido de esta Tesis de Grado, me corresponde exclusivamente; y el patrimonio intelectual de la misma a la Escuela Superior Politécnica del Litoral”.



Abel Stalin Flores Laaz

RESUMEN

Para validar y calibrar los indicadores discriminantes para pacientes hospitalarios con problemas cardiovasculares se empleó el modelo FRAMINGHAM (USA-MASSACHUSSETS).

El objetivo de esta investigación es la adaptación del modelo Framingham a las características que presentan los pacientes del Hospital Luis Vernaza de Guayaquil; fundamentalmente, consiste en estimar la probabilidad de riesgo o la probabilidad que una persona reingrese al Hospital por problemas cardiovasculares. Por ello, este estudio es en base a la información obtenida en el establecimiento médico, consistente en las variables que utiliza el modelo antes mencionado.

Con los datos obtenidos se llevó a cabo la validación y calibración del modelo que se utilizó en la presente investigación, donde la validación consiste en evaluar el modelo para los coeficientes obtenidos en base a las características del país de origen; también se encontró los valores de supervivencia, debido a que en nuestro país no contamos con una tabla actuarial para la población específica que está siendo objeto de estudio (personas con problemas cardiovasculares), por lo tanto se ajustó los datos a la distribución de supervivencia de la Weibull, con la cual se obtiene los

valores de estos en base al tiempo. Se calibró el modelo debido a que los valores obtenidos en la evaluación aplicando el modelo Framingham, no se ajustan a los datos sometidos a estudio, porque posiblemente los hábitos de vida en nuestro país no son iguales que en Estados Unidos, siguiendo el procedimiento de calibración para encontrar los coeficientes utilizados en el modelo, se aplicaron las técnicas de regresión lineal múltiple y variables indicadoras.

El primer capítulo de la presente investigación trata el aspecto teórico de las enfermedades cardiovasculares, los tipos y factores de riesgo de las mismas.

El segundo capítulo se enfoca en el marco teórico, donde se presenta conceptos básicos de la bioestadística, del riesgo, supervivencia, regresión de Cox y la distribución de Weibull.

El tercer capítulo presenta una breve reseña sobre el origen de los modelos de riesgo cardiovascular y el modelo Framingham.

El cuarto capítulo describe las unidades de estudio, la forma de obtención de las mismas, y se hace un análisis de las variables utilizadas en el modelo para el cálculo de la probabilidad de re-ingreso al hospital por problemas cardiovasculares.

El quinto capítulo constituye un análisis de la validación y calibración del modelo Framingham.

Finalmente se presentan las diferentes conclusiones y recomendaciones.

ÍNDICE GENERAL

	Pág.
RESUMEN	VI
ÍNDICE GENERAL	IX
ABREVIATURAS	XII
SIMBOLOGÍA	XIII
ÍNDICE DE GRÁFICOS	XIV
ÍNDICE DE TABLAS	XV
INTRODUCCIÓN	XVI
CAPÍTULO 1	
1. Aspectos Básicos sobre los problemas Cardiovasculares	1
Enfermedades Cardiovasculares	1
Factores de Riesgo Cardiovascular	3
Prevención	8
CAPÍTULO 2	
2. Marco Teórico	9
2.1. Bioestadística	9
2.1.1. Variables	10
2.2. Riesgo y Supervivencia	12

2.2.1.	Análisis de Supervivencia	15
2.3.	Regresión Lineal Múltiple	18
2.4.	Regresión de Cox	24
2.5.	Distribución de Weibull	25

CAPÍTULO 3

3.	Modelos de Riesgo Cardiovascular	28
3.1.	Origen	28
3.2.	Modelos de Riesgo Cardiovascular	29
3.3.	Modelo de Framingham	31

CAPÍTULO 4

4.	Unidades de Estudio	36
4.1.	Recolección de la información y descripción de las variables..	36
4.2.	Codificación de las variables	39
4.3.	Análisis de cada variable	40

CAPÍTULO 5

5.	Validación y Calibración del Modelo Framingham	43
5.1.	Validación	43
5.2.	Calibración	54

5.3. Evaluación del Modelo Framingham en base a los coeficientes encontrados	60
--	----

CONCLUSIONES

RECOMENDACIONES

ANEXO

BIBLIOGRAFÍA

ABREVIATURAS

E.C.V.	Enfermedad cardiovascular.
Exp	Exponencial.
(f/q)	Fallecimiento o quiebra.
HDL – COLESTEROL	(Siglas en inglés de lipoproteína de alta densidad) – colesterol bueno.
H.L.V.	Hospital Luis Vernaza.
Mg/dl	Niveles en los que se mide el colesterol.
mmHg	Niveles en los que se mide la tensión arterial.
LDL – COLESTEROL	(Siglas en inglés de lipoproteína de baja densidad) colesterol malo.
PAD	Presión arterial diástolica.
PAS	Presión arterial sistólica.
S_H	Supervivencia base a 10 años para pacientes de sexo masculino.
S_M	Supervivencia base a 10 años para pacientes de sexo femenino.

SIMBOLOGÍA

%	Porcentaje.
\geq, \geq	Mayor o igual que.
\leq, \leq	Menor o igual que.
t	Tiempo de observación.
f(t)	Función densidad de probabilidad.
F(t)	Función de distribución de probabilidades.
S(t)	Función de supervivencia.
h(t)	Función de riesgo.
∞	Infinito.
R^2	Coefficiente de determinación del modelo, o potencia de explicación del modelo.
P(x)	Probabilidad de x.
#	Número.
e	Exponencial.
$\lim_{x \rightarrow +\infty}$	Límite cuando x tiende al infinito(+).
β_i	Coefficientes del modelo.
ε_i	Error aleatorio.
λ y γ	Parámetros de la función de supervivencia de la Weibull.
$N(0, \sigma^2)$	Distribución normal estándar con media 0 y varianza σ^2 .

ÍNDICE DE GRÁFICOS

	Pág.
Gráfico 1. Género de Pacientes.....	40
Gráfico 2. Ajuste para los Hombres (Weibull).....	48
Gráfico 3. Ajuste para los Mujeres (Weibull).....	49

ÍNDICE DE TABLAS

		Pág.
Tabla I	Niveles de Colesterol y HDL Colesterol	7
Tabla II	Niveles de Tensión Arterial	8
Tabla III	Coeficientes para el modelo de Framingham	33
Tabla IV	Valores Medios de las variables.....	41
Tabla V	Valores de la función de densidad y supervivencia para hombres.....	46
Tabla VI	Valores de la función de densidad y supervivencia para mujeres.....	47
Tabla VII	Riesgo medio a 10 años de acuerdo al estudio Framingham.....	53
Tabla VIII	Riesgo medio a 10 años en base a las características de los casos de estudio (ajuste Weibull).....	54
Tabla IX	Coeficiente para el modelo de Framingham en base a los datos sometidos a estudio.....	59

INTRODUCCIÓN

El manejo de la información sobre los niveles elevados de colesterol, HDL-Colesterol, tensión arterial, diabético, fumador; de personas con problemas cardiovasculares, son fundamentales para que los doctores ofrezcan a los pacientes mejores medidas correctivas, y de esta manera poder disminuir la probabilidad que se presente otro problema cardiovascular.

La presente investigación trata determinar la probabilidad de re-ingreso para personas con problemas cardiovasculares, para ello, usa el estudio creado en USA-MASSACHUSETTS (Modelo de Framingham). La hipótesis que se plantea es: “El modelo Framingham puede ser utilizado para calcular la probabilidad de re-ingreso”; ya que posiblemente influya mucho las diferentes características de vida (USA-ECUADOR), y de la cual se derivan las siguientes interrogantes: ¿Qué género es más propenso a re-ingresar al hospital por problemas cardiovasculares?, ¿Influye la edad en tener mayor probabilidad de re-ingresar al Hospital?, ¿Los casos de estudio tienen una alta o aceptable probabilidad de no re-ingresar al Hospital por problemas

cardiovasculares?, ¿Los parámetros de Framingham (USA-MASSACHUSETTS) sobre-estimarán la probabilidad de re-ingresos de los casos de estudio?.

Estudios de este tipo han sido realizados en otros países, por lo cual, la presente investigación es la primera en nuestra ciudad y país, por lo tanto sería conveniente tomarla como un estudio piloto, dejando sentada las bases para un estudio mayor, como podría ser a nivel de país.

CAPÍTULO 1

1.- ASPECTOS BÁSICOS SOBRE LOS PROBLEMAS CARDIOVASCULARES.

Este capítulo presentará algunos aspectos básicos de las enfermedades de tipo cardiovascular, también se darán a conocer los factores de riesgos cardiovasculares y los tipos de enfermedades. Además se detalla los niveles recomendables y no recomendables de la Tensión arterial, colesterol y HDL colesterol. Luego se mencionan los pasos a seguir para poder prevenir las enfermedades cardiovasculares.

1.1. Enfermedades cardiovasculares

La enfermedad cardiovascular se refiere a las enfermedades del corazón y enfermedades del sistema de vasos sanguíneos (arterias, capilares, venas) de todo el organismo, tales como el cerebro, las piernas y los pulmones. La palabra cardiovascular se descompone en **Cardio** que se refiere al corazón, y **Vascular** al sistema de vasos sanguíneos.

Bajo el término de enfermedades cardiovasculares se engloban un conjunto de procesos que afectan al corazón y al sistema vascular: enfermedades coronarias, valvulares, miocárdicas, congénitas, cerebrovasculares, vasculares periféricas y otras.

Por las características del corazón, sabemos que bombea sangre continuamente a través del sistema circulatorio que es la red de tubos elásticos, las cuales, permiten que la sangre fluya por todo el organismo. El sistema circulatorio comprende: corazón, pulmones y vasos sanguíneos (arterias, capilares y venas). Las arterias y capilares, transportan la sangre rica en oxígeno y nutrientes del corazón y los pulmones a todas las partes del cuerpo. Las venas regresan la sangre reducida en oxígeno y nutrientes al corazón y pulmones.

Los problemas del corazón y de los vasos sanguíneos no aparecen rápidamente, lo que sucede es que, con el tiempo las arterias que llevan la sangre al corazón y al cerebro pueden obstruirse, debido a la acumulación de células, grasas y colesterol (placa). La disminución en el flujo de sangre al corazón debido a obstrucciones en las arterias, ocasiona ataques cardíacos. El derrame cerebral se produce por la falta de flujo de sangre al cerebro, ocasionada por un coágulo de sangre o una hemorragia en el cerebro, debido a la rotura de los vasos sanguíneos.

Existen diversos tipos de enfermedades cardíacas y cardiovasculares, entre las que podemos mencionar: aterosclerosis, enfermedad cardíaca coronaria (o enfermedad de las arterias coronarias), angina de pecho, derrame cerebral, presión sanguínea elevada(o Hipertensión), insuficiencia cardíaca, infarto agudo del miocardio, arritmias cardíacas, entre otras.

1.2. Factores de riesgo cardiovascular.

Los factores de riesgos que intervienen en el desarrollo de las enfermedades cardiovasculares están determinados por las siguientes variables:

Sexo.-

El sexo es un indicador de diferenciación entre las personas con problemas cardiovasculares, ya que los hombres tienen mayor riesgo de enfermedad cardiovascular que las mujeres. Las personas de sexo femenino poseen una hormona que regula los ciclos menstruales, la cual, disminuye la concentración del Colesterol LDL, posible razón ya que las mujeres en edad menor que en la etapa de menopausia, son menos propensas a las E.C.V.

Edad.-

Los problemas cardiovasculares pueden aumentar en forma lineal con la edad, aunque este tipo de enfermedades no son causa directa del envejecimiento, sino, como consecuencia de un estilo de vida y de acumulación de múltiples factores de riesgo, ya que las arterias coronarias pueden tardar 20 y 30 años en estar completamente bloqueadas. Debido a esto, existen muchas personas de 90 años con corazones saludables y vigorosos.

Herencia o antecedentes familiares.-

Los miembros de familias con antecedentes de ataques cardíacos o enfermedades cardiovasculares, se consideran en una categoría de riesgo cardiovascular más alta que los otros.

Niveles elevados de colesterol total y LDL.-

Las personas que consumen grandes cantidades de colesterol y grasas saturadas, tienen niveles más altos de colesterol en la sangre y mayor probabilidad de padecer enfermedades cardiovasculares.

Niveles bajos de colesterol HDL.-

Los niveles bajos de HDL, son considerados como factor de riesgo, ya que no se encuentran dentro de los estándares establecidos. Por ejemplo, las mujeres en edad fértil tienen una menor incidencia en problemas cardiovasculares que los hombres, producto de la acción de estrógenos, que son hormonas femeninas que regulan los ciclos menstruales y elevan el HDL colesterol.

Tabaquismo.-

El consumo de tabaco aumenta el riesgo de sufrir enfermedades de tipo cardiovascular, por la nicotina y el monóxido de carbono que contiene. El cigarrillo produce un incremento en la adrenalina, ritmo cardíaco acelerado, elevación de presión sanguínea, falta de oxigenación de las células y daños en las paredes de las arterias.

Hipertensión.-

La presión arterial elevada incrementa la posibilidad de sufrir problemas cardiovasculares. Cuando la presión sanguínea elevada se mantiene de forma sostenida en el tiempo, produce múltiples efectos adversos en el sistema cardiovascular.

Diabetes.-

La diabetes aumenta notablemente el riesgo de ataques cardíacos y otras manifestaciones de enfermedades cardiovasculares, las personas con este tipo de enfermedad presentan complicaciones relacionadas con alta concentración de lípidos en la sangre, enfermedades coronarias, hipertensión y otros desórdenes circulatorios. También se pueden considerar como factores de riesgo al sedentarismo, obesidad y estrés.

Los niveles de colesterol total, HDL-colesterol y tensión arterial, son factores que inciden en gran proporción en el origen de las enfermedades cardiovasculares, por esto, es importante que las personas conozcan cuáles son los niveles en estado normal y crítico de los factores antes mencionados.

Niveles de Colesterol.-

Los lípidos o grasas son esenciales para una buena salud, ya que estos son componentes importantes en la estructura y funcionamiento de las células del organismo. Sin embargo cuando se acumula demasiado el colesterol LDL, se deposita en forma de placa en las paredes de los vasos sanguíneos, haciéndolos más gruesos, duros e

inflexibles. Los vasos sanguíneos pierden su facultad de transportar sangre de forma eficiente y como resultado se incrementa el riesgo del infarto, y otros tipos de enfermedades.

A continuación en la siguiente tabla, mostraremos los niveles: recomendables, de bajo y alto riesgo del colesterol y HDL Colesterol.

TABLA I
Niveles de Colesterol y HDL Colesterol.

Categorías	Colesterol Total	Colesterol HDL
Recomendable	<200	>40
Cercano a los límites elevados (Riesgo bajo)	200-239	>35
Elevado (Riesgo alto)	≥ 240	<35

Fuente: Hospital Luis Vernaza de Guayaquil, tesis 2005.

Elaboración: Abel Flores.

Niveles de Tensión arterial.-

La presión arterial es la fuerza con que la sangre empuja contra las paredes de los vasos sanguíneos. Esta se lee de la siguiente forma, por ejemplo: 112 / 78 mmHg, donde el primer número es la presión Sistólica,

se produce en las arterias cuando el corazón late. El segundo número es la presión Diastólica, que es cuando el corazón descansa entre latidos.

En la siguiente tabla se mostrará los niveles normales y altos de la tensión arterial.

TABLA II

Niveles de Tensión Arterial.

Categorías	Normal	Adulto
Presión Normal	< 120/80	< 130/85
Presión Alta	\geq 140/90	\geq 140/90

Fuente: Hospital Luis Vernaza de Guayaquil, tesis 2005.

Elaboración: Abel Flores.

1.3. Prevención.

Recomendaciones para el control de los Factores de Riesgo.

La prevención de riesgo, implica abstenerse en el consumo excesivo de alimentos y otros expresados a continuación: obesidad, comida chatarra, comida precocinada, comida congelada, grasas saturadas, azúcar y harinas refinadas, sal, bebidas alcohólicas, gaseosas, agua clorada, estimulantes, droga, tabaco, estrés. Como complemento se debe realizar más ejercicios físicos e incrementar el consumo de frutas y vegetales.

CAPÍTULO 2

2. MARCO TEÓRICO

El presente capítulo expresa la importancia de la bioestadística, los principales conceptos o teorías básicas que componen esta ciencia. También se presenta una amplia información acerca del riesgo y supervivencia, con los principales pasos que se deben tomar en cuenta al momento de llevar a cabo proyectos de esta índole, de la misma forma se habla acerca de la función de supervivencia utilizando la distribución Weibull y la regresión lineal múltiple.

2.1. Bioestadística

La bioestadística en forma general, es la aplicación de la estadística a la biología y de forma más frecuente a la medicina, debido que las cuestiones a investigar en biología y medicina son de naturaleza muy variada.

En la investigación médica es habitual encontrar situaciones, que un parámetro clínico permite observar las características del paciente durante determinado periodo de tiempo, recogiéndose información en momentos puntuales, los cuales pueden estar prefijados o no antes de comenzar el ensayo, por ejemplo los registros continuos de monitorización de la presión arterial, etc. Surge entonces la pregunta, qué método estadístico o matemático aplicar para analizar este tipo de datos, en planteamientos que pueden ser simplemente descriptivos, para resumir la información sobre el comportamiento de la muestra, o bien en estudios de contraste, por ejemplo para comparar la evolución de los pacientes que siguen dos tratamientos antihipertensivos diferentes. Todo esto resume lo fundamental de la Bioestadística.

La **Muestra**, es el grupo de pacientes u observaciones que se estudiará. Debe haberse elegido al azar y ser representativa de la población a la cual pertenece.

La **Población**, es el conjunto de elementos de la misma especie que se pretende estudiar en una investigación científica, donde podemos extraer muchas muestras.

2.1.1. Variables

Las **variables** en bioestadística, son las características o propiedades determinadas del individuo, sean estas medibles o no. Esta propiedad hace que las personas de un grupo difieran de otro grupo, tanto en la muestra o población de estudio.

-Las variables se clasifican en: cuantitativa y cualitativa.

a.- Variable Cuantitativa, es la que puede medirse, y se conoce como variable numérica o continua.

b.- Variable Cualitativa, son variables que representan cualidades de la muestra, también llamadas categóricas o discretas, por dividirse en categorías. Se clasifican en:

Variables Categóricas Dicotómicas, son las que tienen dos valores fijos y excluyentes entre sí, como la evolución, presencia o ausencia de una enfermedad o característica en la muestra.

Variables Categóricas Nominales, Son variables cualitativas las cuales no permiten establecer un orden, por ejemplo la raza, que puede ser blanca, negra, etc., o los tipos de sangre A, B, AB, O. También son excluyentes entre sí, esto es que cada paciente pertenece a una u otra categoría pero no a dos al mismo tiempo.

Variables Categóricas Ordinales, Estas permiten establecer un orden determinado. También son excluyentes entre ellas.

Además de lo mencionado en este capítulo, existe otra forma de clasificar a las variables como dependientes e independientes. Las variables dependientes son las de Interés, cuyos valores dependen de otras variables que pueden influir en ella. Por ejemplo: La supervivencia, respuesta al tratamiento, etc. Las variables independientes modifican de una u otra forma la variable dependiente, llamándose según el caso factor de riesgo o predictivo.

2.2. Riesgo y Supervivencia

El azar y los eventos –fenómenos- aleatorios son muy importantes en el diario vivir, debido a desconocer con exactitud que sucederá mañana; por lo tanto, estamos sujetos a incertidumbre, la cual está íntimamente ligada con el riesgo, que se lo puede estimar mediante estadística matemática.

El **riesgo**, es la probabilidad de que se produzca un evento de forma aleatoria. Para esta investigación el riesgo es la probabilidad de que el paciente vuelva a ingresar al hospital por problemas cardiovasculares. La **supervivencia** en la presente investigación, es la probabilidad de no re-ingresar al hospital por problemas cardiovasculares.

Al riesgo no se lo elimina, se lo controla, y una forma que las personas que padecen de este tipo de problemas pueden controlar o disminuir el peligro de sufrir estas dificultades, es cuidando su alimentación y hábitos de vida.

La variable tiempo de espera es aleatoria continua y no negativa, cuya función de probabilidad puede especificarse las siguientes maneras: Función densidad de probabilidad $f(t)$, la función de supervivencia $S(t)$ y la función de riesgo $h(t)$.

Cuando hablamos de supervivencia o riesgo, es fundamental definir los entes⁽¹⁾, los sucesos aleatorios y el evento inicial, tal como se muestra a continuación:

ENTE	SUCESO ALEATORIO	EVENTO INICIAL
Persona	Fallecimiento	Nacimiento
Empresa	Quiebra	Creación
Máquina	Pérdida, daño	Recién comprada (Nueva)

Se asocia cada ente en su evento inicial una variable aleatoria ξ que representa la edad del fallecimiento, quiebra o daño (f/q) del ente.

⁽¹⁾ La presente investigación se enfoca en el caso de personas.

Considerando a $x = 0, 1, 2, 3, \dots$ como los años de supervivencia de la persona u otros entes, definimos **F(x)** como la función de distribución de la variable aleatoria ξ que indica la Edad de (f/q), donde:

$$F(x) = P(\xi \leq x); \quad x = 0, 1, 2, 3, \dots$$

De igual forma se define a la función de supervivencia **S(x)** por:

$$S(x) = 1 - F(x) = P(\xi > x) \quad x = 0, 1, 2, 3, \dots$$

Dicha función nos proporciona la probabilidad de que el ente en un evento inicial alcance la edad x , por

lo tanto la función de supervivencia es decreciente a medida que el X tiende al infinito ($\lim_{x \rightarrow +\infty} S(x) = 0$).

Posteriormente, al hablar sobre la función de Supervivencia es fundamental mencionar a la **función de riesgo (h(t))**, con la cual se puede expresar la probabilidad para la variable tiempo de espera; por ejemplo, para la supervivencia a una intervención quirúrgica, la función de riesgo a 2 años, es la probabilidad de que la persona muera en este lapso de tiempo (después de la intervención). Esta función puede ser definida por: $h(t) = \frac{f(t)}{S(t)}$.

Por lo tanto se puede concluir: que si conocemos una de las funciones mencionadas anteriormente, fácilmente se podrán obtener las otras, debido a que están relacionadas entre ellas.

Existen diversos métodos para poder estimar la función de probabilidad y supervivencia, las cuales pueden ser estimadas por los siguientes procedimientos no paramétricos:

-Estimación de la Función de Probabilidad

$$\hat{f}(t) = \frac{\text{Número de Pacientes que fallecieron en el intervalo comenzando en } t}{(\# \text{ total de pacientes})(\text{Amplitud del Intervalo})}$$

- Estimación de la Función de Supervivencia

$$\hat{S}(t) = \frac{\# \text{ de pacientes con tiempos de supervivencia } \geq t}{\# \text{ total de pacientes}}$$

2.2.1. Análisis de Supervivencia.

Los datos proporcionados por estudios clínicos se expresan en múltiples ocasiones en términos de supervivencia, donde esta, es una medida de tiempo a una respuesta como fallo, muerte, recaída, que no queda limitada a los términos de vida o muerte, sino, a situaciones en la que se mide el tiempo que transcurre hasta que sucede un evento de interés como: tiempo de recurrencia, tiempo que dura la eficacia de una intervención, tiempo de aprendizaje determinado, etc.

La observación de cada paciente se inicia al diagnóstico (tiempo = 0) y continua hasta la muerte o hasta que el tiempo de seguimiento se interrumpe. Cuando el tiempo de observación termina antes de

producirse la muerte o antes de completar el período de estudio se habla de paciente “censurado”

El periodo de seguimiento puede terminar por las siguientes razones:

- a.-) El paciente decide no participar más en el estudio y lo abandona.
- b.-) El paciente se pierde y no tenemos información.
- c.-) El estudio termina antes de aparecer el evento.

Cuando los tiempos de supervivencia no se conocen con exactitud, los datos se consideran censurados, ya que no se conoce el tiempo hasta el suceso de interés (muerte, recaída), porque los individuos en el estudio pueden haberse perdido, retirado o el suceso puede no haber ocurrido durante el período de estudio. El seguimiento viene definido por una fecha de inicio y cierre, determinan el tiempo de seguimiento; estas fechas de inicio y cierre son diferentes para cada individuo, ya que los pacientes o personas incluidas en el estudio se incorporan en momentos diferentes.

En las observaciones censuradas el evento de interés no se ha producido, por lo siguiente: el estudio finalizó antes de la aparición del evento; el paciente decide abandonar y no participar en el estudio; perdemos al paciente por cambio en el lugar de residencia; y, muerte no relacionada con la investigación, etc.

El tiempo de supervivencia se define, como periodo transcurrido desde el acontecimiento o estado inicial hasta el estado final. El inicio debe ser definido, considerando la fecha que se produjo cada evento con exactitud: diagnóstico, intervención quirúrgica, inicio de la radioterapia o quimioterapia, etc. En igual forma el acontecimiento o suceso estudiado, debe estar perfectamente definido para determinar exactamente la fecha del mismo, este suceso casi siempre está asociado a la muerte del paciente, pero no necesariamente, sino, puede hacer referencia a la fecha de alta, etc.

Al estudiar la supervivencia, es importante el evento que se produce, esté relacionado con la enfermedad. Si consideramos una muerte no relacionada con la enfermedad introduciremos un sesgo de información, cuando tenemos un paciente fallecido por una

causa que no está vinculada al evento de interés, debe ser censurado y computar su tiempo como incompleto o perdido.

Es importante registrar el estado del sujeto y fecha de información de dicho estado, ya que el período de tiempo transcurrido entre la fecha de entrada y última observación, se conoce como tiempo de participación en el estudio.

Los factores que modifican la supervivencia de un paciente pueden ser variables fijas en el tiempo, como: el sexo, factores genéticos, entre otras. Las variables que se modifican en el tiempo son: intensidad de exposición a una dieta, consumo de cigarrillos diariamente, control médico, etc.

Los requisitos necesarios para disponer de datos adecuados para un análisis de supervivencia son: definir de forma apropiada el origen o inicio del seguimiento, definir la escala de tiempo y determinar apropiadamente el evento.

2.3. Teoría de Regresión Lineal Múltiple

Debemos conocer en ocasiones, es fundamental analizar datos y establecer relaciones que hagan posible predecir una o más variables en términos de otra. La regresión es el proceso que permite establecer la relación y/o predecir una o más variables en función de otras, todo esto se puede llevar a cabo mediante una ecuación matemática.

La ecuación de regresión lineal es de la forma:

$$y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i$$

así como, β_0 y β_1 son constantes llamados coeficientes y ε_i es el error aleatorio del modelo [$\varepsilon_i \approx N(0, \sigma^2)$]. De la ecuación matemática se pueden obtener algunos parámetros, como el R^2 es el coeficiente de determinación del modelo o la potencia de explicación del modelo.

Sin embargo, la presente investigación se basa en la regresión lineal múltiple. En ocasiones no se utiliza una variable explicativa, ya que parece razonable, que las predicciones mejoran si adicionalmente se consideran información relevante. Por ejemplo, debemos ser capaces en hacer mejores predicciones del desempeño de maestros recién contratados si no solo consideramos su educación, sino también, sus años de experiencias y su personalidad. En el caso de la presente investigación, para calcular el riesgo, no sólo se basa en la edad, sino

también, en niveles de colesterol total, HDL-colesterol, T.A.S, si es fumador y/o diabético.

Para efectuar un análisis de regresión múltiple, existen muchas fórmulas que nos sirven expresar la relación de “n” variables, sin embargo, la más usada es aquella, que parte de la ecuación de regresión lineal bivariada, y se generaliza para las variables con las que estamos trabajando.

Siendo este el caso, la ecuación inicial es la siguiente:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

En particular para cada observación se tendrá:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

Los $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ son los coeficientes de la regresión lineal múltiple, son constantes numéricas que deben determinarse a partir de datos observados, y de la misma manera en la regresión lineal simple, el ε representa al error del modelo.

Cuando hablamos de regresión bivariada, el método más usado para el análisis de datos es los **MINIMOS CUADRADOS**; sin embargo, al considerar que ahora tenemos “n” puntos de datos:

$\{(x_{1i}, x_{2i}, \dots, x_{ki}, y_i); i = 1, 2, \dots, n\}$, estimamos los valores de β basándonos en el criterio de mínimos cuadrados.

Las estimaciones de mínimos cuadrados de β son los valores $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$, tenemos que:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

o

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_k x_{ki} + e_i$$

donde, ε_i y e_i son los errores aleatorio y residual, respectivamente, asociados con la respuesta de y [$(y_i - \hat{y}_i) = \varepsilon_i = e_i$].

Para utilizar este criterio se debe minimizar la siguiente expresión y poder

encontrar los estimadores $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Se define a q como: $q = \sum_{i=1}^n \left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_k x_{ki}) \right]^2$, este

debe ser mínimo.

Luego de determinar la ecuación del mínimo de q , se procede a derivar parcialmente con respecto a los $\hat{\beta}_k$, posteriormente igualar todas esas derivadas a la constante 0.

$$\frac{\partial q}{\partial \hat{\beta}_0} = \sum_{i=1}^n (-2) \left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_K x_{ki}) \right] = 0$$

$$\frac{\partial q}{\partial \hat{\beta}_1} = \sum_{i=1}^n (-2) x_{1i} \left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_K x_{ki}) \right] = 0$$

$$\frac{\partial q}{\partial \hat{\beta}_2} = \sum_{i=1}^n (-2) x_{2i} \left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_K x_{ki}) \right] = 0$$

..

$$\frac{\partial q}{\partial \hat{\beta}_K} = \sum_{i=1}^n (-2) x_{ki} \left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_K x_{ki}) \right] = 0$$

Obtendremos las siguientes ecuaciones:

$$\sum y = \hat{\beta}_0 \cdot n + \hat{\beta}_1 \sum x_1 + \hat{\beta}_2 \sum x_2 + \dots + \hat{\beta}_k \sum x_k$$

$$\sum x_1 y = \hat{\beta}_0 \cdot \sum x_1 + \hat{\beta}_1 \sum x_1^2 + \hat{\beta}_2 \sum x_1 x_2 + \dots + \hat{\beta}_k \sum x_1 x_k$$

$$\sum x_2 y = \hat{\beta}_0 \cdot \sum x_2 + \hat{\beta}_1 \sum x_1 x_2 + \hat{\beta}_2 \sum x_2^2 + \dots + \hat{\beta}_k \sum x_2 x_k$$

.

$$\sum x_k y = \hat{\beta}_0 \cdot \sum x_k + \hat{\beta}_1 \sum x_k x_1 + \hat{\beta}_2 \sum x_k x_2 + \dots + \hat{\beta}_k \sum x_k^2$$

Con estas ecuaciones se pueden hallar los valores: $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$, por cualquier método apropiado para resolver sistemas de ecuaciones lineales. Nótese, que abreviamos la notación al escribir: $\sum_{i=1}^n x_{1i} = \sum x_1$,

$$\sum_{i=1}^n x_{1i} x_{2i} = \sum x_1 x_2$$

En este tipo de regresión, cuando existen más de dos variables es común usar la **notación matricial**, de tal forma, se puedan expresar resultados generales de manera compacta y así aprovechar las ventajas de la teoría matricial.

En **notación matricial** se puede escribir las ecuaciones:

$$y = X\beta + \varepsilon, \quad \text{donde:}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_3 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \cdot & & & & \\ \cdot & & & \dots & \\ \cdot & & & & \\ \cdot & & & \dots & \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\beta}_k \end{bmatrix}$$

Por lo tanto, para estimar los coeficientes $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$, por el método de mínimos cuadrados, estaría dado:

$$B = (X'X)^{-1} X'Y$$

$$X'X = \begin{pmatrix} n & \sum x_1 & \sum x_2 & \sum x_k \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 & \sum x_1 x_k \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 & \sum x_2 x_k \\ \cdot & \cdot & \cdot & \cdot \\ \sum x_k & \sum x_k x_1 & \sum x_k x_2 & \sum x_k^2 \end{pmatrix}$$

$$X'Y = \begin{bmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \\ \sum x_k y \end{bmatrix}$$

Con las técnicas referidas encontramos los coeficientes $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$, valores importantes en la investigación.

2.4. Regresión de Cox

En base a los modelos de regresión mencionados en el numeral anterior, es fundamental hablar de **regresión de COX**. Este es un método que permite estudiar datos de supervivencia, modelizando la función de azar de cada individuo como un producto de dos factores, dado que es

interesante modelar no sólo la relación entre la tasa de supervivencia y el tiempo, sino también la posible relación con diferentes variables registradas para cada sujeto. Por lo tanto se trata de calcular la tasa de mortalidad como una función del tiempo y de las variables pronóstico.

$$h(t) = h_0(t) \cdot e^{\{b_1x_1+b_2x_2+\dots+b_kx_k\}}$$

El primer factor $h_0(t)$ depende del tiempo, y el segundo factor $e^{\{b_1x_1+b_2x_2+\dots+b_kx_k\}}$ depende de las variables pronósticos que serán distintas para cada individuo, y no del tiempo.

Así, para un individuo para el que $e^{\{b_1x_1+b_2x_2+\dots+b_kx_k\}} = 2.54$, la función de supervivencia será $S(t) = S_0(t)^{2.54}$, donde $S_0(t)$ es la función de supervivencia base. A esta función base se la puede estimar por métodos como el Kaplan-Meier, la distribución Weibull, etc.

2.5. Distribución de Weibull

En la metodología estadística básica se señala la existencia de pruebas paramétricas y no paramétricas. Dentro del análisis de supervivencia, el estudio de datos puede ser realizado utilizando las pruebas antes mencionadas. Entre las **Paramétricas** las más frecuentes están:

distribución Exponencial, distribución **Weibull**, y Lognormal. Entre las **No paramétricas están**: Kaplan-Meier, Logrank y Regresión de Cox.

En la investigación es de vital importancia contar con la probabilidad de supervivencia a 10 años para los pacientes con problemas cardiovasculares, y debido a que en nuestro país, no existe este tipo de tablas para la cohorte antes mencionada, se hizo uso de la **distribución de supervivencia de la Weibull** para poder encontrar dichos valores, debido a que su distribución es igual a la de Gamma y exponencial, porque, se aplican a problemas de confiabilidad y prueba de duración de vida, tales como tiempo de falla o periodo de vida de un componente medido a partir de un tiempo determinado, hasta que se presente una falla.

La función de densidad de la distribución Weibull está dada por:

$$f(x) = \begin{cases} \lambda \gamma \cdot x^{\gamma-1} e^{-\lambda x^\gamma} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Esta función depende de dos parámetros λ y γ , evaluando la función para $\gamma=1$ la distribución Weibull se transforma en una distribución exponencial.

La función de la tasa de falla o también conocido como el modelo de función de **riesgo** denominado de Weibull, viene dado por la siguiente expresión:

$$h(t) = \lambda\gamma \cdot t^{\gamma-1}$$

Depende de dos parámetros λ y γ , ambos mayores que cero. Esta tasa de falla o riesgo disminuye con el tiempo si $\gamma < 1$, que puede servir por ejemplo para modelar la aparición de fallos. Para $\gamma > 1$ la función de riesgo aumenta con el tiempo que puede servir para modelar la supervivencia de un ente que envejece con el tiempo y, cuando $\gamma = 1$ la función permanece constante.

La supervivencia correspondiente a la función de riesgo de Weibull viene dada por la siguiente fórmula:

$$S(t) = \exp(-\lambda t^\gamma)$$

Nótese que calculando dos veces el logaritmo de la función de supervivencia, tendremos: $\ln[-\ln S(t)] = \ln \lambda + \gamma \ln t$, es decir las relaciones entre el logaritmo del logaritmo cambiando de signo de la supervivencia con el logaritmo de tiempo, son lineales.

Por lo tanto, si representamos en una gráfica $\ln(-\ln S(t))$ en función de $\ln(t)$, y el modelo de Weibull es adecuado, los puntos deberán estar próximo a una línea recta. En la estimación y ajuste en un modelo

Weibull de supervivencia para los datos de la presente investigación se obtienen los valores de λ y γ respectivamente, estos se utilizan para obtener los valores de supervivencia en el tiempo.

CAPÍTULO 3

3. MODELOS DE RIESGO CARDIOVASCULAR

A continuación detallamos el origen de los modelos de riesgo cardiovascular con sus respectivas características y ventajas, también se proporciona la información acerca del modelo para el cálculo de riesgo cardiovascular utilizado en esta investigación como lo es “EL MODELO DE FRAMINGHAM”.

3.1. Origen

La construcción de modelos de riesgo de aparición de un suceso es de gran importancia en medicina, tanto para intentar conocer las variables que influyen en que se presente ese suceso, como para analizar el mecanismo que lo produce y para predecir su aparición. En el primer

caso, el conocimiento de las variables que influyen nos permitirá establecer medidas preventivas o terapéuticas, y en el segundo mediante el modelo podemos efectuar cálculos relacionados con la aparición del suceso, por ejemplo para determinar las necesidades de recursos.

Precisamente la teoría matemática para el cálculo de modelos de riesgo tiene su origen probablemente en este último aspecto, y más concretamente en el campo de la ingeniería, donde la demanda creciente de equipos que funcionen cada vez mejor y a menor costo lleva aparejada la necesidad de disminuir las probabilidades de fallo de éstos, lo que posibilitó el estudio y desarrollo de modelos probabilísticos para analizar la naturaleza de esos fallos y minimizar así la probabilidad de que ocurran, lo que en ingeniería se conoce como **teoría de la fiabilidad** (*reliability*) y en medicina habitualmente como **análisis de supervivencia** (*survival analysis*).

Cualquier construcción matemática, por sencilla que ésta sea, constituye un modelo y como tal una simplificación de la realidad, útil pero simplificación al fin y al cabo. Por eso estos modelos constituyen sin ninguna duda una valiosísimas herramientas para el conocimiento, interpretación y en su caso modificación de los fenómenos, pero casi siempre se trata de modelos transitorios, sujetos a verificación y

perfeccionamiento, y como todo en el mundo de la ciencia solo pueden ser aceptados con una cierta dosis de escepticismo y con una mentalidad crítica.

3.2. Modelos de riesgo cardiovascular.

Puesto que las enfermedades cardiovasculares constituyen una de las principales causas de mortalidad y morbilidad en los países desarrollados, es lógico que sea de gran interés el desarrollo de modelos de predicción del riesgo de padecer este tipo enfermedades, tanto para intentar conocer los posibles mecanismos que afectan al aumento del riesgo, como para poder intervenir precozmente ante estas.

Resulta evidente que en muchos procesos dicha probabilidad depende del tiempo de exposición, aumentando a medida que éste transcurre, por lo tanto, el tiempo interviene en la ecuación como factor de riesgo o bien se utiliza un modelo específico en el que se tenga en cuenta esta característica, calculando ahora la probabilidad de que el suceso ocurra en un momento de tiempo determinado. Esto es precisamente lo que se hace en los modelos probabilísticos de supervivencia, siendo el método más conocido el denominado modelo de riesgos proporcionales o modelo de Cox. Sin embargo no es la única alternativa posible, existiendo otros posibles métodos de modelado denominados paramétricos, debido a que

suponen un tipo concreto de ecuación matemática para la función de riesgo, y que aunque en la industria son muy utilizados, sin embargo no es tan normal encontrarlos en la literatura médica, aunque como veremos más adelante, precisamente han sido utilizados recientemente en el modelo de riesgos propuesto por Framingham, entre éstos métodos paramétricos se encuentran los modelos basados en la función de **WEIBULL**.

Aunque existen gran número de trabajos relativos al estudio de los riesgos de enfermedad cardiovascular, el **estudio de Framingham** constituye un pilar básico, y en diferentes formas es ampliamente utilizado para la toma de decisiones terapéuticas en base a la estimación de riesgo proporcionada por el modelo al introducir las características de riesgo del paciente específico. Es tan popular que incluso existen calculadoras de bolsillo que implementan el algoritmo, y también diferentes páginas Web en las que se puede efectuar dicho cálculo.

Existen otros modelos además de los citados, pero quizá vale destacar otro trabajo anterior que incluye pacientes de Norte América y Europa, el cual basado en el proyecto INDANA, que también dispone de una calculadora on-line de riesgo.

3.3. Modelo de Framingham.

Para poder llevar a cabo el cálculo del riesgo mediante el modelo de Framingham que utiliza el valor del colesterol total, debemos tener presente las variables que intervienen en este modelo: **SEXO**, **EDAD**, el **COLESTEROL** sérico en mg/dl, fracción de colesterol ligado a lipoproteínas de alta densidad **HDL** en mg/dl, **PRESION SISTOLICA** en mmHg, **DIABETES** (No, Sí), **FUMADOR** (No, Sí), los cuales son factores de riesgo para problemas cardiovasculares.

Primero se debe calcular el valor de la siguiente expresión, en base a los factores de riesgo mencionados anteriormente:

Para los hombres: $L_H = b_{E1} \cdot EDAD + b_C + b_H + b_T + b_D + b_F$

Para las mujeres: $L_M = b_{E1} \cdot EDAD + b_{E2} \cdot EDAD^2 + b_C + b_H + b_T + b_D + b_F$

Los coeficientes correspondientes al factor de riesgo edad están dado por: b_{E1} es el coeficiente por género y b_{E2} es el coeficiente de la $(Edad)^2$, sólo para mujeres.

El coeficiente correspondiente al factor de riesgo colesterol, está dado por b_c obviamente por género, conformado por cinco coeficientes para los respectivos niveles de colesterol total agrupados por intervalos.

El coeficiente perteneciente al factor de riesgo HDL-Colesterol está representado por b_H , que a su vez está compuesto por cinco coeficientes para los respectivos niveles de HDL-Colesterol agrupados por intervalos.

De la misma forma, como los coeficientes mencionados anteriormente, b_r corresponde a los niveles de tensión arterial; y por último los coeficientes b_D y b_F corresponden a los factores de riesgo para pacientes diabéticos o fumadores respectivamente.

Por lo tanto, los coeficientes para sus respectivos niveles se muestran a continuación:

TABLA III

Coefficientes para el modelo de Framingham (Colesterol total)

Coeficiente	Hombres	Mujeres
$b_{E1} \times \text{Edad}$	0.04826	0.33766
$b_{E2} \times (\text{Edad})^2$	0	-0.00268
b_C Colesterol mg/dl		
< 160	-0.65945	-0.26138
160-199	0	0
200-239	0.17692	0.20771
240-279	0.50539	0.24385
≥ 280	0.65713	0.53513
b_H HDL-Col mg/dl		
< 35	0.49744	0.84312

35 - 44	0.24310	0.37796
45 - 49	0	0.19785
50 - 59	-0.05107	0
≥ 60	-0.48660	-0.42951
b_T Tensión arterial mmHg		
PAS < 120 PAD < 80	-0.00226	-0.53363
PAS <130 PAD < 85	0	0
PAS <140 PAD < 90	0.28320	-0.06773
PAS < 160 PAD < 100	0.52168	0.26288
PAS ≥160 PAD ≥100	0.61859	0.46573
b_D Diabetes		
NO	0	0
SI	0.42839	0.59626
b_F Fumador		
NO	0	0
SI	0.52337	0.29246

Fuente: Asociación de la Sociedad Española de Hipertensión
Elaboración: Luis M. Molinero

Una vez calculado el valor correspondiente de **L**, se le resta la cantidad **G** (función evaluada para los valores medios de las variables en el estudio) diferente para hombres o mujeres, donde:

$$\mathbf{G}_{\text{Hombres}} = 3.0975$$

$$\mathbf{G}_{\text{Mujeres}} = 9.92545$$

Exponenciamos ese valor calculado $\mathbf{B} = \exp(L - G)$ y determinamos el valor de la expresión $1 - S^{\mathbf{B}}$. **S** es la función de supervivencia base a 10 años, que es diferente para hombres y mujeres:

$$S_{\text{Hombres}} = 0.90015$$

$$S_{\text{Mujeres}} = 0.96246$$

-Ejemplo: Ponderaremos según este modelo el riesgo de una mujer de 57 años, con los siguientes datos: presenta 238 mg/dl de colesterol, 52 mg/dl de HDL, 150 / 92 mmHg de Tensión arterial, y no es diabética, ni fumadora.

Desarrollo:

$$L_M = 0.33766 \cdot 57 - 0.00268 \cdot 57^2 + 0.20771 + 0 + 0.26288 + 0 + 0$$

$$L_M = 11.00989$$

$$B = \exp(11.00989 - 9.92545)$$

$$B = \exp(1.08444)$$

$$B = 2.95778$$

El Riesgo o la probabilidad de un acontecimiento cardiovascular a los 10 años según este modelo está dado por:

$$R = 1 - 0.96246^{2.95778} = \mathbf{0.107}$$

$$R = 11\%$$

El riesgo de esta mujer de 57 años de presentar un problema cardiovascular es del 11%. Este modelo valora el posible riesgo de

enfermedad cardiovascular de un sujeto de acuerdo a una serie de características.

CAPÍTULO 4

4. UNIDADES DE ESTUDIO

Este capítulo detalla como se llevó a cabo la recolección de la información y el procesamiento de la misma, para determinar los datos que serán útiles en la presente investigación. Asimismo se hará la descripción de cada variable utilizada en el modelo Framingham con su respectivo análisis.

4.1. Recolección de la Información y descripción de las variables.

Para llevar a cabo la investigación, se acudió al Hospital Luis Vernaza de la ciudad de Guayaquil, de dicha institución se tomó la información de

pacientes que ingresaron en este establecimiento con problemas cardiovasculares, para así, efectuar la validación y calibración del modelo, el cual permita calcular el riesgo.

En el Hospital nos facilitaron una lista con los números correspondientes a fichas médicas de un periodo de 10 años (1994 – 2003). Se trató de obtener todos los datos y fichas de los pacientes que padecieron de dicha enfermedad, cabe recalcar que existieron muchas fichas extraviadas, repetidas y algunas historias clínicas incompletas, lo cual complicó al momento de extraer toda la información necesaria para este tipo de estudio.

Este modelo se basa en el cálculo de riesgo o en la probabilidad que una persona con problemas cardiovasculares reingrese al hospital por los mismos inconvenientes. Se obtuvo 155 fichas médicas correspondientes a este tipo de pacientes, en los cuales nos enfocamos, para de esta manera poder medir los días de reingresos por pacientes, este dato será fundamental para así encontrar la probabilidad de riesgo que estamos buscando.

Así como es importante los pacientes que re-ingresaron con problemas cardiovasculares, también fue necesario separar los datos por el género

de los pacientes, ya que en Framingham para poder calcular la probabilidad de riesgo se usan coeficientes distintos tanto para hombres y mujeres. La variable sexo nos indica el género de los pacientes del Hospital Luis Vernaza de Guayaquil.

Para iniciar esta recolección de datos, partimos de los indicadores que son utilizados en Framingham, por esto, las variables de interés dentro de la investigación son:

Variable # 1: Edad.

Es una variable cuantitativa que nos indica la edad del paciente que presentó problemas cardiovasculares, el coeficiente perteneciente a esta variable está dada por: b_{E1} para los hombres y b_{E1}, b_{E2} para las mujeres.

Variable # 2: Colesterol.

Esta variable cuantitativa mide el grado de colesterol total que tiene cada paciente, donde estos niveles influyen mucho en el origen de enfermedades cardiovasculares y está dado en mg/dl. El coeficiente para la variable por género se expresa por b_C .

Variable # 4: HDL Colesterol.

Variable cuantitativa que mide el grado de colesterol HDL que tiene el paciente. Los niveles de colesterol bueno, son importantes para poder prevenir enfermedades cardiovasculares y, están expresados en mg/dl.

El coeficiente para la variable por género está dado por: b_H

Variable #5: Tensión Arterial.

Es una variable cuantitativa, la cual mide el grado de tensión sistólica y diastólica de cada paciente. Los niveles de Tensión Arterial son primordiales en los pacientes que presentan este tipo de enfermedades y están expresados en mmHg. El coeficiente de esta variable por género está dado por: b_T .

Variable #6: Diabetes.

Variable cualitativa que indica si el paciente sufre o no de diabetes. Esta enfermedad es un factor de riesgo para problemas cardiovasculares y, el coeficiente de esta variable está dado por: b_D .

Variable #7: Fumador

Esta variable cualitativa nos indica si el paciente es o no fumador. Este hábito influye mucho en problemas cardiovasculares, debido a que los fumadores son más propensos a tener este tipo de problemas. El coeficiente de esta variable es: b_F .

4.2. Codificación de las variables.

Dos variables, de las cuales hemos obtenido la información son de tipo cualitativo, con el objeto de convertir estas variables en cuantitativas; para poder así utilizarlas al realizar la validación y calibración del modelo Framingham, hemos procedido a codificarlas de la siguiente manera:

Variable #6: Diabetes.	Codificación
Diabético	1
No Diabético	0

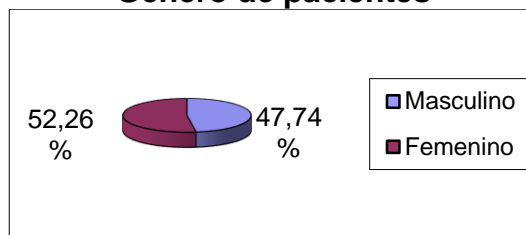
Variable #7: Fumador.	Codificación
Fumador	1
No Fumador	0

4.3. Análisis de Variables.

Como se menciona al inicio de este capítulo los casos de estudio en la presente investigación, son pacientes con problemas cardiovasculares del Hospital Luis Vernaza de Guayaquil.

De los pacientes que presentaron problemas cardiovasculares en el Hospital, tenemos que de un total de 155 pacientes: 74 eran hombres y 81 mujeres. Por lo tanto el 47.74% afecta a los hombres y el 52.26% a las mujeres.

GRÁFICO 1
Género de pacientes



Fuente: Validación y calibración del Modelo Framingham, tesis 2005.

Elaboración: Abel Flores.

En el presente análisis es de vital importancia los valores medios de cada variable, por este motivo en la siguiente tabla se proporcionará dichos valores, incluyendo otros parámetros correspondientes a factor o variable.

TABLA IV
Valores Medios

Variables	Hombres	Mujeres
Edad	63,1081	60,6173
Colesterol	190,0676	209,9877
HDL- Colesterol	40,5270	42,3086
T.A.S.		
Sistólica	125,6486	131,5432
Diastólica	75,6757	77,9506
Diabetes		
No	0,6622	0,5926
SI	0,3378	0,4074
Fumador		
No	0,3919	0,8148
Si	0,6081	0,1852

Fuente: Validación y calibración del Modelo Framingham, tesis 2005.
Elaboración: Abel Flores.

Observando los resultados en la Tabla IV podemos concluir:

- Los valores de colesterol presentados por los pacientes del Hospital Luis Vernaza están asociados a un nivel cercano a los límites elevados⁽³⁾, lo cual no es recomendable para este tipo de problemas. Se pudo observar que el valor promedio de colesterol presentado por los pacientes de sexo masculino es 190.0676 mg/dl, lo cual afirma que estos se aproximan a

(3) ver tabla I, en pág. 7

niveles de riesgo bajo y no son recomendables para personas con problemas cardiovasculares. En el caso femenino encontramos niveles de riesgo bajo, con un valor promedio de 209.9877 mg/dl, que apenas se

acerca a niveles elevados, lo cual es perjudicial para este tipo de pacientes

- Los valores promedios de HDL-Colesterol de los pacientes sometidos a estudio, no presentan mucha diferencia entre género.

- Los valores promedios de tensión arterial de personas de sexo masculino y femenino, se encuentran en niveles cercanos al riesgo ya que sobrepasan el nivel recomendable que es $< 120/80$ mmHg.

- En los casos de estudio, observamos que predominan las personas que no sufren de Diabetes tanto en hombres y mujeres. El 66.22% de los pacientes del sexo masculino y el 59.26% del sexo femenino no son diabéticos

- En los pacientes que son fumadores predominan los hombres con el 60.81%, lo cual los hace más propensos de sufrir problemas cardiovasculares. Esto no sucede con las mujeres, ya que existe una mayor proporción de no fumadoras.

CAPÍTULO 5

5. VALIDACIÓN Y CALIBRACIÓN DEL MODELO FRAMINGHAM.

Este capítulo explica cómo se lleva a cabo la validación, calibración y posterior evaluación, con los coeficientes obtenidos en la calibración del modelo Framingham; obviamente para los pacientes con problemas cardiovasculares del Hospital Luis Vernaza de Guayaquil.

5.1. Validación.

Para realizar la validación del modelo, es necesario contar con los datos de los pacientes con problemas cardiovasculares, los cuales deberán ordenarse (clasificarse) por género.

La validación consiste en evaluar los datos para las fórmulas utilizadas por el modelo de estudio, las cuales fueron mencionadas en el capítulo 3. Se evaluará con los coeficientes de la TABLA III, que son los coeficientes del modelo, los cuales fueron obtenidos en base a los datos pertenecientes a los pacientes de USA - Massachusetts, lugar donde se implementó el modelo por primera vez.

Lo primero para validar este modelo, es encontrar el **valor de G** tanto para hombres y mujeres. Esta es la **función evaluada para los valores medios de las variables de estudio** (diferente para hombres y mujeres); por lo tanto, se necesitará de los valores medios de cada variable que fueron presentados en la tabla IV.

Las funciones para el valor G (hombres y mujeres) está dada por las siguientes funciones:

$$G_H = b_{E1} \cdot \hat{Edad} + b_C + b_H + b_T + b_D + b_F \quad (1)$$

$$G_M = b_{E1} \cdot \hat{Edad} + b_{E2} \cdot \left(\hat{Edad} \right)^2 + b_C + b_H + b_T + b_D + b_F \quad (2)$$

Los coeficientes que se les asignarán a cada una de las variables, corresponden a los valores presentados en la Tabla III. Por ejemplo:

-Para el coeficiente de edad (b_{E1}) el valor para los hombres es 0.04826.

-Para el Coeficiente del Colesterol (b_C) de los cinco valores diferentes, elegimos el valor que en su intervalo contenga al valor promedio del colesterol para hombres (190.0676). Este valor se encuentra en el intervalo de 160 -199, cuyo coeficiente es 0.

-Para el coeficiente del HDL – Colesterol (b_H), al igual que en la variable de colesterol, observamos en que intervalo se encuentra el valor de

40.5270 correspondiente al valor promedio, este se encuentra en el rango de 35 – 44, que tiene un valor de 0.24310.

De esta forma, encontramos el valor de los coeficientes basados en la TABLA III. Los valores de G obtenidos serían:

- Para Hombres

$$G_H = (0.04826 \cdot 63.1081) + 0 + 0.24310 + 0 + (0.3378 \cdot 0.42839) + (0.6081 \cdot 0.52337)$$

$$G_H = 3.7517$$

- Para Mujeres

$$G_M = (0.33766 \cdot 60.6173) + (-0.00268 \cdot (60.6173)^2) + 0.20771 + 0.37796 + 0 + \\ + (0.4074 \cdot 0.59626) + (0.1852 \cdot 0.29246)$$

$$G_M = 11.5032$$

Después de obtener los valores de G, también es fundamental tener los valores de **supervivencia base a 10 años**, diferenciando su género. Para obtener ese valor de supervivencia, primero se debe ajustar los valores a una distribución Weibull y con eso se obtendrán los valores de λ y γ los cuales son los parámetros de esta función, y como la función de supervivencia de la Weibull está dada por: $S(t) = e^{-\lambda t^\gamma}$.

Sólo faltaría el valor de t , que es una variable cuantitativa medida por los días de re-ingreso por paciente, pues la estimación de la función de supervivencia será en base al tiempo.

Del seguimiento de las fichas de cada paciente, debemos contar los días de re-ingreso por cada uno, con lo cuál obtendremos la función de supervivencia. Estimamos la función de supervivencia, en base a las fórmulas expuesta en la sección 2 del capítulo 2; los resultados obtenidos se presentarán en las siguientes tablas:

TABLA V

Valores de la función de densidad y supervivencia para Hombres

Intervalo	Frecuencia	f(t)	S(t) Condicionado
(0-399]	50	0,00169	1
(399-798]	11	0,00037	0,32432
(798-1197]	5	0,00017	0,17568
(1197-1596]	3	0,00010	0,10811
(1596-1995]	1	0,00003	0,06757
(1995-2394]	2	0,00007	0,05405
(2394-2793]	1	0,00003	0,02703
(2793-3192]	1	0,00003	0,01351

Fuente: Validación y calibración del Modelo Framingham, tesis 2005.
Elaboración: Abel Flores.

TABLA VI

Valores de la función de densidad y supervivencia para Mujeres.

intervalo	Frecuencia	f(t)	S(t) Condicionado
-----------	------------	------	-------------------

(0-340]	45	0,00163399	1
(340-680]	18	0,00065359	0,44444
(680-1020]	6	0,00021786	0,22222
(1020-1360]	5	0,00018155	0,14815
(1360-1700]	2	7,2622E-05	0,08642
(1700-2040]	3	0,00010893	0,06173
(2040-2380]	1	3,6311E-05	0,02469
(2380-2720]	1	3,6311E-05	0,01235

Fuente: Validación y calibración del Modelo Framingham, tesis 2005.

Elaboración: Abel Flores.

Dado que $S(t)$ es condicionado porque los datos son en base a pacientes con re-ingreso, por este motivo existirán datos perdidos o censurados, puesto que existen personas que abandonaron el tratamiento, nunca más regresaron al Hospital, tal vez murieron o simplemente no sabemos que ha pasado con ellos. Por lo tanto, hacemos uso de técnicas matemáticas, para poder llevar a cabo el ajuste por medio de la función de Weibull.

Ante esto, se plantea la siguiente probabilidad:

$$P(S_{re-ingreso \leq T} > t) = P(S > t / S \leq T) = \frac{S(t) - c}{1 - c}$$

Donde c es una constante en la probabilidad ajustada condicionada, y reemplazando $S(t)$ por la función de supervivencia de Weibull en la probabilidad ajustada condicionada, tendremos que:

$$P(S_{re-ingreso \leq T} > t) = \frac{e^{-\lambda t^\gamma} - c}{1 - c}.$$

Basándose en esta probabilidad, se procede a hacer el ajuste a las tablas V y VI; utilizando también el modelo $\{ \exp(-\lambda t^\gamma) - C \} / \{ 1 - C \}$.

Los resultados obtenidos para los Hombres son los siguientes:

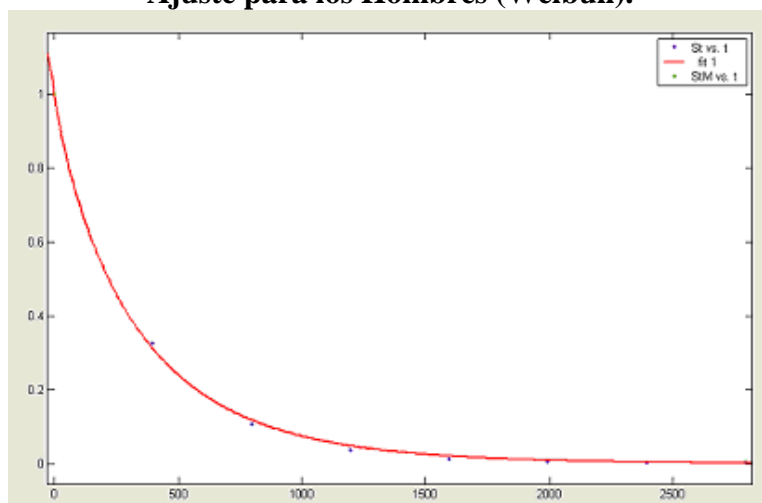
$$\begin{array}{ll} C = 0.00002644 & R^2 = 0.9989 \\ \lambda = 0.006508 & \gamma = 0.8672 \end{array}$$

Observando los resultados podemos decir que es un buen ajuste. Otra forma de confirmar esto, es observando la gráfica 2; además, porque el coeficiente R^2 se aproxima a 1.

Por lo tanto ya hemos encontrado los valores de λ y γ que son los necesarios para poder obtener los valores de Supervivencia.

GRÁFICO 2

Ajuste para los Hombres (Weibull).



Fuente: Validación y calibración del Modelo Framingham, tesis 2005.

Elaboración: Abel Flores.

Los resultados obtenidos para las mujeres son:

$$C = 0.005422$$

$$\lambda = 0.007537$$

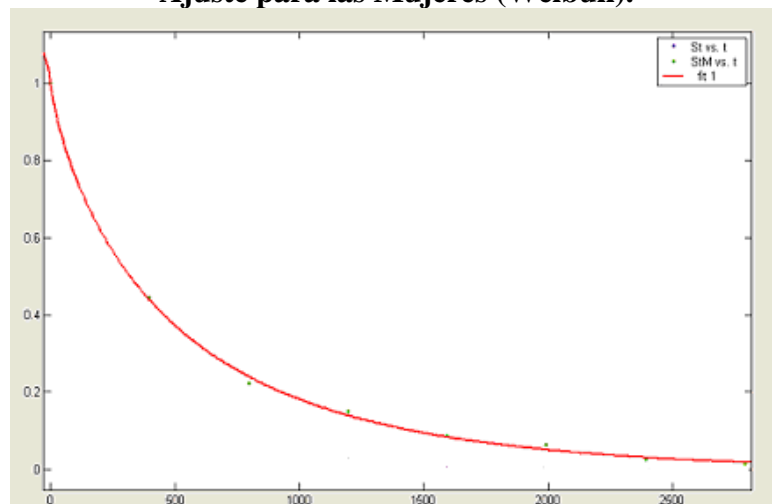
$$\gamma = 0.7829$$

$$R^2 = 0.9989$$

Similar a los hombres, podemos decir, que es un buen ajuste si observamos la gráfica 3 y además porque el coeficiente R^2 se aproxima a 1. Por lo tanto, hemos encontrado los parámetros de la función Weibull con la cual, podemos hallar los valores de supervivencia.

GRÁFICO 3

Ajuste para las Mujeres (Weibull).



Fuente: Validación y calibración del Modelo Framingham, tesis 2005.

Elaboración: Abel Flores.

En el siguiente cuadro, se mostraran las respectivas funciones de supervivencia por género:

S(t) Hombres	S(t) Mujeres
$S(t) = e^{-0.006508t^{0.8672}}$	$S(t) = e^{-0.007537t^{0.7829}}$

Obtenidos los valores de G y S(t), podemos validar el modelo aplicando la formula para el cálculo de riesgo, la cual detallamos a continuación:

$$\left\{ \begin{array}{l} R = 1 - S^{\exp(L-G)}, \text{ donde:} \\ S = \text{función de Supervivencia base a 10 años} \\ L = (L_H \quad \text{o} \quad L_M) \\ G = (\text{función evaluada para los valores medios de las variables en el estudio}) \end{array} \right.$$

Lo fundamental es calcular los valores de L_H o L_M , obviamente dependiendo del género.

Mediante ejemplos se explicará la evaluación de un paciente por género.

EJEMPLO 1

-Validemos según este modelo el riesgo de un hombre de 58 años de edad, que es diabético, pero no fuma, tiene 150/90 mmHg de presión arterial, 208 mg/dl de colesterol total, 38 mg/dl de HDL.

Encontramos L_H de la siguiente forma :

$$L_H = b_{E1} \cdot EDAD + b_C + b_H + b_T + b_D + b_F$$

Por lo tanto, de la misma forma que hicimos en el cálculo del valor G, tenemos que:

$$L_H = (0.04826 \cdot 58) + 0.17692 + 0.24310 + 0.52168 + 0.42839 + 0$$

$$L_H = 4.16917$$

A continuación, se hace uso del valor de G que fue calculado anteriormente: $G_H = 3.7517$.

Calculando $\exp(L_H - G_H)$, tenemos que:

$$\exp(4.16917 - 3.7517) = 1.518163317$$

Luego, usando la función de supervivencia base a 10 años antes encontrada, el valor de $S(t) = e^{-0.006508t^{0.8672}}$ será:

$$S_H = 0.95$$

Entonces la probabilidad que se produzca un incidente cardiovascular a 10 años será:

$$R = 1 - (0.95)^{1.518163317} = 0.0749168$$

$$\text{Riesgo} = 7.49 \%$$

EJEMPLO 2

- Encontramos el riesgo de una mujer de 58 años de edad, que es diabética, pero no fuma, tiene 190/110 mmHg de presión arterial, 166 mg/dl de colesterol total, 37 mg/dl de HDL.

Hallemos L_M :

$$L_M = b_{E1} \cdot EDAD + b_{E2} \cdot EDAD^2 + b_C + b_H + b_T + b_D + b_F$$

Consiguientemente nos queda que:

$$L_M = (0.33766 \cdot 58) + (-0.00268 \cdot (58)^2) + 0 + 0.37796 + 0.46573 + 0.59626 + 0$$

$$L_M = 12.00871$$

Haciendo uso del $G_M = 11.5032$

Calculando $\exp(L_M - G_M)$, tenemos que:

$$\exp(12.00871 - 11.5032) = 1.657830799$$

Utilizando la función de supervivencia base a 10 años que previamente

fue determinada: $S(t) = e^{-0.007537t^{0.7829}}$ y el valor de esta será:

$$S_M = \mathbf{0.96}$$

Entonces, la probabilidad que se produzca un suceso cardiovascular a 10 años será:

$$R = 1 - (0.96)^{1.657830795} = 0.06543673941$$

Riesgo = 6.54 %

El anexo A proporciona un conjunto de personas con diferentes características, a las cuales se les obtendrá la probabilidad de riesgo.

Para efectos comparativos de resultados, en la tabla VII se proporciona el riesgo medio a 10 años, de acuerdo al estudio Framingham.

TABLA VII

Riesgo Medio a 10 años de acuerdo al estudio Framingham

Edad	Mujeres	Hombres
15-25	< 0.09 %	< 0.51 %
26-36	1.29 %	2.35 %
37-47	2.11 %	4.87 %
48-58	7.05 %	8.68 %
59-69	11.51 %	24.01 %
70-80	9.02 %	25.73 %
81-91	10.23 %	23.25 %

Fuente: Validación y calibración del Modelo Framingham, tesis 2005.

Elaboración: Abel Flores.

La Tabla VII está basada en los coeficientes de USA-MASACHUSSETS, la cual debe ser comparada con la Tabla VIII que contiene los valores de

riesgo a 10 años según la tasa de re-ingreso (días) promedio ligados a los intervalos referentes a las edades de los casos de estudio.

TABLA VIII

Riesgo Medio a 10 años en base a las características de los casos de estudio.

Edad	Mujeres	Hombres
15-25	1.00 %	2.14 %
26-36	2.00 %	5.001 %
37-47	4.36 %	7.45 %
48-58	5.01 %	7.84 %
59-69	7.18 %	8.96 %
70-80	8.00 %	9.70 %
81-91	7.68 %	9.46 %

Fuente: Validación y calibración del Modelo Framingham, tesis 2005.
Elaboración: Abel Flores.

Observando las dos tablas antes expuestas, podemos darnos cuenta que los resultados en base a los coeficientes del modelo FRAMINGHAM-USA no se ajustan a los valores de riesgo real en base a los casos de estudio; por este motivo, tenemos que encontrar los coeficientes para las características de los pacientes con problemas cardiovasculares del Hospital. Este proceso de hallar los coeficientes, lo denominamos: calibración del modelo sometido a estudio que será explicado a continuación.

5.2. Calibración

Este proceso consistirá en encontrar los coeficientes del modelo Framingham en base a las características de los casos de estudio. Esto se debe a que los coeficientes utilizados en la primera validación son en base a la información obtenida en el país de origen del modelo (USA-MASACHUSSETS).

En base al modelo Framingham, se parte de: $S(t) = S^B$ (1), El objetivo, es hacer lineal la función (1), para de esta forma hacer uso de las técnicas de regresión lineal múltiple y encontrar los coeficientes del modelo. Aplicando logaritmos en ambos lados en la función (1) quedaría:

$$S(t) = S^B$$

$$\ln(S(t)) = B \cdot \ln(S)$$

$$(-\ln(S(t))) = B \cdot (-\ln(S))$$

$$\ln(-\ln(S(t))) = \ln[B \cdot (-\ln(S))]$$

$$\ln(-\ln(S(t))) = \ln(B) + \ln(-\ln(S))$$

$$\ln(-\ln(S(t))) = \ln(B) + b_0$$

$$\ln(-\ln(S(t))) = b_0 + \ln(B)$$

Dado que $B = \exp(L-G)$, reemplazando quedaría lo siguiente:

$$\ln(-\ln(S(t))) = b_0 + \ln(\exp(L-G))$$

$$\ln(-\ln(S(t))) = b_0 + (L-G), \text{ y por género sería: } \begin{pmatrix} b_0 + (L_H - G_H) \\ b_0 + (L_M - G_M) \end{pmatrix}$$

- Consecuentemente, el modelo de regresión a utilizar para encontrar los coeficientes utilizando el criterio de estimación de mínimos cuadrados, sería:

- Hombres:

$$\ln(-\ln(S(t))) = b_0 + b_{E1}(x_1 - \bar{x}_1) + b_C(x_2 - \bar{x}_2) + b_H(x_3 - \bar{x}_3) + b_T(x_4 - \bar{x}_4) + b_D(x_5 - \bar{x}_5) + b_F(x_6 - \bar{x}_6) \quad (2)$$

- Mujeres:

$$\ln(-\ln(S(t))) = b_0 + b_{E1}(x_1 - \bar{x}_1) + b_{E2}(x_2 - \bar{x}_2) + b_C(x_3 - \bar{x}_3) + b_H(x_4 - \bar{x}_4) + b_T(x_5 - \bar{x}_5) + b_D(x_6 - \bar{x}_6) + b_F(x_7 - \bar{x}_7) \quad (3)$$

-Ahora faltaría encontrar el valor de $\ln(-\ln(S(t)))$, para de esta forma obtener los coeficientes $b_{E1}, b_{E2}, b_C, b_H, b_T, b_D, b_F$.

Para encontrar el valor de $\ln(-\ln(S(t)))$ usamos la función de supervivencia de la Weibull ($S(t) = e^{-\lambda \cdot t^\gamma}$) y le aplicamos logaritmo en ambos lados, con lo cual tendremos:

$$\ln(S(t)) = \ln(e^{-\lambda \cdot t^\gamma})$$

$$\ln(S(t)) = -\lambda \cdot t^\gamma$$

$$-\ln(S(t)) = -\lambda \cdot t^\gamma$$

$$-\ln(S(t)) = \lambda \cdot t^\gamma$$

$$\ln(-\ln(S(t))) = \ln(\lambda \cdot t^\gamma)$$

$$\ln(-\ln(S(t))) = \ln \lambda + \ln t^\gamma$$

$$\ln(-\ln(S(t))) = \ln \lambda + \gamma \cdot \ln(t)$$

Entonces:

$$\ln(-\ln(S(t))) = \ln \lambda + \gamma \cdot \ln(t) \quad \begin{cases} \text{hom bres} & \ln(0.006508) + (0.8672) \cdot \ln(t) \\ \text{mujeres} & \ln(0.007537) + (0.7829) \cdot \ln(t) \end{cases}$$

Se ha encontrado los valores del lado izquierdo del modelo (2) y (3), para poder hallar los coeficientes.

-Pero aquí no termina la calibración, porque se debe hacer uso de las variables indicadoras “dummy”, para recodificar estas, debido a que si observamos las variables: Colesterol, HDL Colesterol y tensión arterial, estas están compuestas por 5 rangos en base a los niveles de las mismas; por ejemplo:

Rango (Niveles de Colesterol)	Rango (Niveles de HDL)	Rango (Niveles de T.A.S.)
<160	<35	PAS < 120 PAD <80

160-199	35-44	PAS < 130 PAD <85
200-239	45-49	PAS < 140 PAD <90
240-279	50-59	PAS < 160 PAD <100
≥ 280	≥ 60	PAS ≥ 160 PAD ≥ 100

Por consiguiente es necesario tener 5 coeficientes para cada variable antes mencionada, así obtener la tabla de estos en base a nuestros datos.

Esta recodificación se la realiza utilizando las variables indicadoras consistentes en:

Generar tantas variables como el total de la categoría menos uno, obviamente esto se puede efectuar cuando una variable presente más de dos categorías. Cada nueva variable tomará valor 1 para un determinado nivel y 0 el resto, de tal forma que los individuos en una misma categoría tomarán valor 1 en una misma variable y 0 en el resto.

La categoría no considerada o categoría referencia, estará representada por el valor 0 en todas las nuevas variables. En nuestro caso los niveles normales en colesterol, HDL y T.A.S., son considerados como referencia para los demás. Mediante este esquema de codificación, los coeficientes

de las nuevas variables reflejarán el efecto de las categorías representadas respecto al efecto de la categoría referencia. Por ejemplo las variables quedarían codificadas de la siguiente manera:

	x1	x2	x3	x4
bC1	1	0	0	0
bC2	0	0	0	0
bC3	0	1	0	0
bC4	0	0	1	0
bC5	0	0	0	1

En la tabla IX se muestran los coeficientes obtenidos, utilizando los modelos de regresión lineal múltiple (2) y (3), aplicando las variables indicadoras.

TABLA IX

Coeficientes para el modelo de Framingham en base a los datos sometidos a estudio

Coeficiente		Hombres	Mujeres
$b_{E1} \times \text{Edad}$	(bE1)	0.004144	0.0149
$b_{E2} \times (\text{Edad})^2$	(bE2)	0	-0.000102
b_C Colesterol mg/dl			
< 160	(bC1)	-0.273	-0.507
160-199	(bC2)	0	0
200-239	(bC3)	0.114	0.109
240-279	(bC4)	0.218	0.115
≥ 280	(bC5)	0.407	0.225
b_H HDL-Col mg/dl			
< 35	(bH1)	0.0367	0.600
35 – 44	(bH2)	0.01500	0.341
45 – 49	(bH3)	0	0.258
50 – 59	(bH4)	-0.0583	0
≥ 60	(bH5)	-0.291	-0.212
b_T Tensión arterial mmHg			

PAS < 120 PAD < 80 (bT1)	-0.717	-0.08040
PAS <130 PAD < 85 (bT2)	0	0
PAS <140 PAD < 90 (bT3)	0.05122	-0.0149
PAS < 160 PAD < 100 (bT4)	0.332	0.356
PAS ≥160 PAD ≥100 (bT5)	0.589	0.401
b_D Diabetes		
NO	0	0
SI	0.118	0.391
b_F Fumador		
NO	0	0
SI	0.09756	0.002838

Fuente: Validación y calibración del Modelo Framingham, tesis 2005.

Elaboración: Abel Flores.

5.3. Evaluación del Modelo Framingham con los coeficientes encontrados.

El objetivo de esta evaluación, es efectuar los cálculos del modelo Framingham con los coeficientes obtenidos en la sección 5.2 (Tabla IX). Por este motivo las funciones para el valor G (hombres y mujeres) estarán dadas por las siguientes funciones:

$$G_H = b_{E_1} \cdot \hat{Edad} + b_c + b_H + b_T + b_D + b_F \quad (1)$$

$$G_M = b_{E_1} \cdot \hat{Edad} + b_{E_2} \cdot \left(\hat{Edad} \right)^2 + b_c + b_H + b_T + b_D + b_F \quad (2)$$

- Para Hombres

$$G_H = (0.004144 \cdot 63.1081) + 0 + 0.01500 + 0 + (0.3378 \cdot 0.118) + (0.6081 \cdot 0.09756)$$

$$G_H = 0.3757066$$

- Para Mujeres

$$G_M = (0.0149 \cdot 60.6173) + (-0.000102 \cdot (60.6173)^2) + 0.109 + 0.341 + 0 + \\ + (0.4074 \cdot 0.391) + (0.1852 \cdot 0.002838) \\ G_M = 1.13822215$$

Se hace uso de los valores de supervivencia base a 10 años encontrados en la sección 5.1, que son:

$$S_H = 0.95 \quad \text{y} \quad S_M = 0.96$$

Conociendo, los valores de G y S(t), podemos validar el modelo aplicando la formula para el cálculo de riesgo, la cual detallamos a continuación:

$$R = 1 - S^{\exp(L-G)}, \text{ donde:}$$
$$\left\{ \begin{array}{l} S = \text{función de Supervivencia base a 10 años} \\ L = (L_H \quad \text{o} \quad L_M) \\ G = (\text{función evaluada para los valores medios de las variables en el estudio}) \end{array} \right.$$

Mediante ejemplos se explicará la evaluación de un paciente por género.

Ejemplo 3

- *Evaluar según este modelo, el riesgo de un hombre de 54 años de edad, que es diabético y fuma, tiene 150/70 mmHg de presión arterial, 190 mg/dl de colesterol total, 40 mg/dl de HDL.*

Encontramos L_H de la siguiente forma:

$$L_H = b_{E1} \cdot EDAD + b_C + b_H + b_T + b_D + b_F$$

Al igual que se hizo en el cálculo del valor G, tenemos que:

$$L_H = (0.004144 \cdot 54) + 0 + 0.01500 + 0.332 + 0.118 + 0.09756$$

$$L_H = 0.79$$

A continuación, se hace uso del valor G calculado anteriormente:

$$G_H = 0.3757066$$

Calculando $\exp(L_H - G_H)$, se tiene que:

$$\exp(0.79 - 0.3757066) = 1.5133$$

La probabilidad de un evento cardiovascular a los 10 años será:

$$R = 1 - (0.95)^{1.5133} = 0.075$$

$$\mathbf{Riesgo = 7.5\%}$$

Ejemplo 4

-Ahora de la misma forma, validemos el riesgo de una mujer de 74 años de edad, que es diabética, pero no fuma, tiene 160/100 mmHg de presión arterial, 220 mg/dl de colesterol total, 38 mg/dl de HDL.

Hallemos L_M :

$$L_M = b_{E1} \cdot EDAD + b_{E2} \cdot EDAD^2 + b_C + b_H + b_T + b_D + b_F$$

Consiguientemente nos queda que:

$$L_M = (0.0149 \cdot 74) + (-0.000102 \cdot (74)^2) + 0.109 + 0.341 + 0.401 + 0.391 + 0$$

$$L_M = 1.786048$$

Haciendo uso del $G_M = 1.13822215$

Calculando $\exp(L_M - G_M)$, tendremos que:

$$\exp(1.786048 - 1.13822215) = 1.91138068$$

Por lo tanto la probabilidad de un evento cardiovascular a los 10 años será:

$$R = 1 - (0.96)^{1.91138068} = 0.0750996597$$

Riesgo = 7.51%

CONCLUSIONES

Como conclusiones generales respecto a las características de los casos de estudio, tenemos que:

1. De los casos estudiados con problemas cardiovasculares (155 registros), la proporción entre hombres (47.74%) y mujeres (52.26%) son aproximadamente iguales.
2. El valor promedio de colesterol presentados por los hombres (190.0676 mg/dl) y mujeres (209.9877 mg/dl). Los hombres están cercanos a (200-239 mg/dl) nivel de bajo riesgo, y en el caso de las mujeres están en el nivel citado.
3. Los valores promedio de HDL-Colesterol presentados por los hombres (40.5270 mg/dl) y mujeres (42.3086 mg/dl), no presentan mucha diferencia entre género.
4. Los valores promedio de tensión arterial presentados por los hombres (125.6486/75.6757 mmHg) y mujeres (131.5432/77.9506 mmHg) se encuentran cercanos al riesgo ya que sobrepasan el nivel recomendable (< 120/80 mmHg).

5. En los casos de estudio con problemas cardiovasculares (74 hombres y 81 mujeres) predominan las personas que no son diabéticas, tanto hombres (66.22%) y mujeres (59.26%).
6. En el caso de pacientes fumadores predominan las personas del sexo masculino, el 60.81% tienen el hábito de fumar, lo cual los hace más propensos a sufrir problemas cardiovasculares.

Respecto a la validación y calibración del modelo Framingham se puede decir lo siguiente:

7. Los casos de estudio analizados son de pacientes con re-ingresos. por esto existen datos censurados. Para determinar la Supervivencia Real se utilizó un ajuste de Weibull condicionado, donde la probabilidad de supervivencia real no presenta mucha diferencia con la probabilidad de supervivencia condicionada ajustada.
8. Los valores del R^2 en el ajuste, para encontrar los parámetros de la función de supervivencia de la Weibull Condicionada son aceptables: hombres ($R^2=0.9989$) y mujeres ($R^2=0.9989$).

9. Se calibró el modelo Framingham debido a que el riesgo medio a 10 años en base a los coeficientes (USA-Massachusetts), sobrestima a los valores de riesgo real a 10 años en función a las características de los casos de estudio (mediante el ajuste Weibull).

10. Existe una gran variación entre los valores G según las características de (USA-Massachusetts, $G_H = 3.7517$ y $G_M = 11.5032$) y los casos de estudio ($G_H = 0.3757$, $G_M = 1.1382$), la cual, posiblemente se origina por los diferentes estilos de vida entre países (USA-ECUADOR).

Respecto a los resultados de la probabilidad de re-ingresar al Hospital se puede decir lo siguiente:

11. La probabilidad de no re-ingresar al Hospital por problemas cardiovasculares (base a 10 años) es aceptable referente a los casos de estudio: hombres (0.95) y mujeres (0.96).

12. A mayor edad la probabilidad de re-ingresar al hospital por problemas cardiovasculares aumenta, dependiendo de los factores de riesgo del paciente.

13. Según los resultados se afirma que el hombre tiene mayor probabilidad de re-ingresar al hospital por problemas cardiovasculares que la mujer.

14. Se comprobó, que la probabilidad de re-ingresar al hospital (problemas cardiovasculares) según las características de los casos analizados, es baja con respecto a los parámetros de USA-MASSACHUSETTS.

RECOMENDACIONES

1. Para que la recolección de datos o levantamiento de información, sea eficaz y eficiente, se recomienda al Hospital Luis Vernaza de Guayaquil, automatice sus servicios en el departamento de estadística para que posea un sistema de base de datos y permita acceder con facilidad a esta. Así obtendremos la información actualizada y requerida acerca de los pacientes con sus respectivas enfermedades que será de mucha utilidad para investigaciones futuras.
2. Es importante que los Hospitales, Clínicas y demás instituciones médicas, faciliten el ingreso de la información a la base de datos, deben tener diseñado un buen registro de información para que los médicos puedan llenar con rapidez todos los datos correspondientes al paciente, de forma clara y precisa. De esta manera la información será más útil para efectos de investigación.
3. Sería importante que las instituciones médicas incursionen más en investigaciones de bioestadística, éstas aportan de gran manera a la calidad de vida de las personas. Con los resultados que se obtengan de ellas, los médicos especialistas estarán en capacidad de dar mejores medidas correctivas al problema de los pacientes.

4. Las Instituciones Médicas deberían establecer convenios con el I.C.M., para poder llevar a cabo las investigaciones referidas en el punto 3. Esto sería favorable para ambas partes, debido a que las Instituciones Médicas tendrán a su disposición expertos en estadística, y los estudiantes del I.C.M. podrán poner en práctica sus conocimientos adquiridos a lo largo de su vida universitaria.

5. Se sugiere considerar esta investigación como base, para poder implementar el modelo Framingham en otros establecimientos médicos, ya que el estudio se realizó únicamente en el Hospital Luis Vernaza de Guayaquil.

6. Se considera necesario además, concientizar en las personas mejorar su estilo de vida, siendo la única forma de prevenir las enfermedades cardiovasculares y otras.

ANEXO

ANEXO A

Probabilidad de evento cardiovascular a 10 años de acuerdo al estudio de Framingham

Sexo	Edad	Diabetes	Fumador	T.A.S.	COLESTEROL		L_H	L_M	$B = e^{-L-g}$	S^B	$R = 1 - S^B$	Riesgo en %
					Total	HDL						
f	15	no	no	90/60	160	62	3,49876	0,000333976	0,99998637	1,36336E-05	0,0013634	
f	33	no	si	100/60	290	47	8,71607	0,061597746	0,99748862	0,002511384	0,2511384	
m	36	no	si	120/80	200	38	2,68075	0,34269353	0,98257571	0,01742429	1,742429	
f	44	no	no	190/100	200	38	10,71996	0,45692318	0,98152037	0,018479634	1,8479634	
m	45	si	si	90/60	234	47	3,29812	0,63536939	0,96793515	0,03206485	3,206485	
f	58	si	no	190/110	166	37	12,00871	1,6578308	0,93456326	0,065436739	6,5436739	
m	58	si	no	150/90	208	38	4,16917	1,51816332	0,9250832	0,0749168	7,49168	
m	69	si	si	120/80	240	34	5,28453	4,63140955	0,78854938	0,21145062	21,145062	
m	81	no	si	197/130	160	37	5,29412	4,67603842	0,78674632	0,21325368	21,325368	

Fuente: Validación y calibración del Modelo Framingham, tesis 2005.

Elaboración: Abel Flores.

BIBLIOGRAFÍA

- A. *Francisco Louzada Neto, Josmar Mazucheli y Jorge Alberto Achcar,* (2002). **Análise de Sobrevivencia e Confiabilidade.**-(Instituto de Matemática y Ciencias Afines, IMCA.)
- B. *John E., Irwin Millar y Maryless Millar,*(2000). **Estadística Matemática con Aplicaciones,** Pearson Educación – Sexta Edición.
- C. *Luis M. Molinero,* (2003). **Modelos de riesgo cardiovascular – Estudio de Framingham-** Asociación de la Sociedad Española de Hipertensión.
- D. *Luis M. Molinero,* (2004). **Función de Supervivencia de la función Weibull-** Asociación de la Sociedad Española de Hipertensión.
- E. *Luis M. Molinero,* (2003). **Variables Indicadoras Dummy-** Asociación de la Sociedad Española de Hipertensión.
- F. *L. Zapata,* “**Factores predictorios de sobre vida en pacientes con melona, mediante el modelo regresión de Cox**” (Tesis, Instituto de Ciencias Matemáticas, Escuela Superior Politécnica del litoral, 2004).

