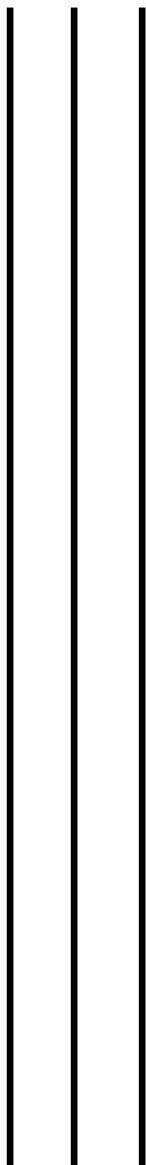




UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN



ANÁLISIS DE LA EVOLUCIÓN DE LA MORTALIDAD EN MÉXICO  
POR CAUSA DE MUERTE MEDIANTE EL ANÁLISIS  
MULTIVARIADO

TESINA  
QUE PARA OBTENER EL TÍTULO DE:  
ACTUARIO

PRESENTA:  
HUGO PIMENTEL COTA

ASESOR:  
ACT. MAHIL HERRERA MALDONADO



Acatlán, Estado de México

2011



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## DEDICATORIA

Para la persona que me dio la vida, y que luchó por que sus hijos fueran personas de bien, mi madre. Sin ti, jamás hubiera podido gozar de la felicidad plena por la que siempre luchaste para nosotros. Esta tesina es también resultado de tu gran trabajo de madre... Muy bien hecho.

Para la persona que me sigue dando vida, luz, cariño y amor... Motivos que solo tú podías darme para desear abrir los ojos... por todo tu apoyo, tu amor, tú mi compañera, mi esposa Pumita...

A la niña de mis ojos, Miztli, que nos hiciste los padres más felices y orgullosos del universo, y sé que cuando puedas leer estas palabras lo mismo estaremos pensando de ti, pero mucho más intenso. Te amo mi niña...

A mis hermanos, que tantos obstáculos sorteamos en el camino de la vida y que siempre hemos estado presentes, sobretodo, cuando más se ha necesitado, haciendo innegable un lazo irrompible. Gracias carnalita Moce, bro Hanz y compa Heri... para ustedes.

A mis amigos actuarios: el abuelo Vic, Luis, Edgardo, las ñoñas, Betoreje, Maribel y David, Evita, etc... profesores, alumnos, compañeros y gente que me dejó una pequeña semillita a lo largo de mi vida, para ellos...

## AGRADECIMIENTOS

Gracias Mahil, profesor, amigo, compañero y asesor.

Gracias a la UNAM, pues gracias a ella desperté y vi que es lo que quería...  
Me dio la vida que amo...

## CONTENIDO

	Página
Introducción.....	5
<b>I. Antecedentes y Conceptos .....</b>	<b>7</b>
A. Mortalidad.....	7
<i>i.</i> Definición .....	7
<i>ii.</i> Referencias históricas .....	13
<i>iii.</i> Leyes de mortalidad.....	16
B. Causa de Muerte.....	21
<i>i.</i> Necesidad de una clasificación por causa de muerte .....	21
<i>ii.</i> Definición del CEI-10.....	23
<b>II. Métodos del análisis multivariado de datos .....</b>	<b>30</b>
A. Componentes Principales .....	30
<i>i.</i> Determinación de las componentes .....	30
<i>ii.</i> Propiedades de las componentes .....	35
<i>iii.</i> Selección del número de componentes .....	36
B. Regresión Logística.....	37
<i>i.</i> Modelo de Regresión Logística .....	37
<i>ii.</i> Coeficientes de regresión.....	39
<i>iii.</i> Supuestos del modelo de regresión.....	40
C. Componentes Principales como variables endógenas de un Modelo de Regresión Lineal Múltiple.....	49
<b>III. Método de corrección de parámetros de la función de Gompertz-Makeham del Dr. Alejandro Mina Valdés .....</b>	<b>51</b>
<b>IV. Modelo de la Mortalidad Mexicana en función de las causas de muerte que la generan .....</b>	<b>56</b>
A. Modelo de la mortalidad mexicana usando las componentes principales como variables.....	56
B. Impacto de las causas de muerte sobre la mortalidad mexicana .....	58
<b>V. Método de proyección de la Mortalidad Mexicana a través del modelo lineal de parámetros diferenciales .....</b>	<b>61</b>
Conclusiones.....	65
Bibliografía .....	67

## Introducción

El análisis de la mortalidad humana ha ido evolucionando y encontrando nuevas vías de investigación para conocer más a fondo su comportamiento. Los investigadores relacionados con este tema han explorado nuevas formas de investigación refinando su análisis mediante el estudio por causa de muerte, uno de los canales de investigación de la Bio-demografía.

Uno de los objetivos principales en el estudio demográfico de la mortalidad ha sido conocer la forma en la que se extingue una población y la evolución de su comportamiento a través del tiempo. Cada causa de muerte impacta de manera diferente sobre la mortalidad poblacional y para entender su evolución a través del tiempo se hace necesario analizar de forma muy particular a cada una entendiendo el impacto que ejerce sobre la mortalidad poblacional.

Este trabajo pretende proponer una metodología para identificar los impactos de las causas de muerte en la mortalidad poblacional por edad así como estimar el impacto de cada una en los siguientes años.

La metodología que se propondrá utiliza una combinación entre la teoría actuarial de decrementos múltiples y el análisis de regresión planteando un modelo para su proyección en base al método publicado por el Dr. Alejandro Mina Valdés en su artículo *Funciones de Sobrevivencia empleadas en el análisis demográfico* (2001). En él, enuncia un método para la corrección de parámetros de la función de Gompertz-Makeham ampliada.

Sin embargo, el uso de este método en esta investigación será para modificar los parámetros utilizados y obtener los parámetros del modelo que estimará la mortalidad del año siguiente.

## I. Antecedentes y Conceptos

### A. Mortalidad

#### i. Definición

La mortalidad es un componente fundamental en la dinámica demográfica de la población y su comportamiento se ha estudiado bajo dos enfoques complementarios. Por un lado, la transición demográfica, que se refiere al tránsito entre la mortalidad y natalidad de la población así como la transición epidemiológica que analiza el cambio paulatino en el perfil de causas de muerte: las afecciones infecciosas y parasitarias, los eventos contingentes que causan perjuicios a la salud así como las enfermedades crónicas y/o degenerativas.

El fenómeno demográfico de la mortalidad está determinado esencialmente por la edad excepto en dos casos: la mortalidad infantil (concentrada no en el primer año de vida, sino, en los primeros días) y la mortalidad juvenil (sobretudo motivada por factores circunstanciales: accidentes, drogas, alcohol, etc., que poco tiene que ver con los factores biológicos). Es también el género una variable fundamental en el análisis demográfico, por ello, género y edad son las características de mayor significado actuarial y que mantiene una permanente interrelación con el comportamiento de la población, lo que da lugar a la elaboración de tablas de mortalidad para el caso de los hombres y de las mujeres.



El patrón moderno de mortalidad puede ser dividido en tres períodos: el primero de ellos contempla la mortalidad infantil (se distingue por una alta mortalidad en los primeros días de vida y un rápido decremento de las tasas de mortalidad para los años siguientes), la mortalidad juvenil (mencionado en párrafos anteriores motivada por factores circunstanciales: accidentes, drogas, alcohol, etc.) y la mortalidad geriátrica (la mortalidad incrementa de forma casi geométrica hasta antes de edad 80 disminuyendo su intensidad posterior a ésta).

El interés de los investigadores de las áreas de Demografía y de Medicina en encontrar factores que introduzcan mayores y mejores explicaciones del comportamiento de la mortalidad poblacional han ayudado a encauzar estudios dirigidos a analizar una nueva variable de estudio de enorme importancia en el entendimiento del comportamiento de la mortalidad humana: la causa de muerte. Esto permite exponer una nueva variable abriendo las posibilidades a que otras áreas del conocimiento como lo sería la medicina, la biología y algunas otras aportaran nuevas ideas en las líneas de investigación biodemográficas y entender mejor el impacto que cada una de ellas tiene en la mortalidad poblacional.

Al considerar la supervivencia humana en los estudios demográficos, el llamado modelo biométrico (expresión matemática que mide la vida), esta caracterizado por la edad de los individuos (período transcurrido desde el nacimiento hasta el momento actual). Así, en el análisis demográfico la Teoría

de la Supervivencia se fundamenta en la variable edad del individuo (tiempo biométrico del individuo) y se basa en la hipótesis de: homogeneidad, independencia y estacionalidad.

Teniéndose que la hipótesis de homogeneidad consiste en que todos los individuos objeto de estudio forman un grupo homogéneo respecto al fenómeno de la supervivencia, es decir, que el comportamiento estadístico de su edad de fallecimiento es idéntico; la hipótesis de independencia es que no existe interacción entre los individuos lo que implica que la probabilidad de supervivencia de un individuo a una determinada edad es independiente a la supervivencia de cualquier otro individuo del mismo grupo; y la hipótesis de estacionalidad en que las probabilidades de supervivencia de los individuos no dependen del tiempo físico (fecha de nacimiento) sino, únicamente, del tiempo biométrico (edad del individuo).

Destaca el hecho de que los modelos biométricos se expresan en función de dos sucesos complementarios a saber, el sobrevivir a una edad concreta ó no hacerlo, siendo la variable biométrica básica por excelencia la edad de fallecimiento de un recién nacido que se denota como  $x$ . Esta variante permite calcular distintas probabilidades, por ejemplo, que un recién nacido fallezca entre las edades  $x$  y  $x + t$ , que un recién nacido sobreviva a la edad  $x$  ó que una persona de edad  $x$  fallezca entre las edades  $x$  y  $x + t$ . Destacando en relación con las probabilidades la función de supervivencia y la tasa instantánea de mortalidad.

La función de supervivencia  $S(x)$  proporciona para cada edad  $x$  la probabilidad de que un recién nacido alcance con vida dicha edad. La tasa instantánea de mortalidad (o fuerza de mortalidad  $\mu(x)$  o  $\mu_x$ ) es uno de los elementos claves del estudio actuarial.

La relación entre las funciones  $S(x)$  y  $\mu(x)$  esta dada por la siguiente relación:

$$S(x) = \exp\left\{-\int_0^x \mu(s)ds\right\}, \text{ así como } \mu(x) = -\frac{S'(x)}{S(x)}$$

Las notaciones matemáticas utilizadas con mayor frecuencia en la demografía para el estudio de la mortalidad y en la ciencia actuarial son:

- ${}_n p_x$  : denota la probabilidad de que una persona de edad  $x$  fallezca después de la edad  $x + n$ .

Es posible expresar esta probabilidad en términos de la función de supervivencia  $S(x)$  y de la fuerza de mortalidad  $\mu(x)$ :

$${}_n p_x = \frac{S(x+n)}{S(x)} = \exp\left\{-\int_x^{x+n} \mu(s)ds\right\} = \exp\left\{-\int_0^n \mu(x+t)dt\right\}$$

- ${}_n q_x$  : denota la probabilidad de que una persona de edad  $x$  no fallezca después de la edad  $x + n$ . Expresamos esta probabilidad en términos de  ${}_n p_x$  y de la fuerza de mortalidad  $\mu(x)$ :

$${}_nq_x = 1 - {}_np_x = 1 - \exp\left\{-\int_0^n \mu(x+t)dt\right\} = \exp\left\{-\int_n^\infty \mu(x+t)dt\right\}$$

- $l_x$ : número de personas vivas de edad  $x$
- $d_x$ : número de personas fallecidas de edad  $x$  en el transcurso de 1 año
- $\omega = \{\min x | l_x = 0\}$

Ahora, en los modelos de Decrementos Múltiples se redefinen los valores actuariales que utilizaremos a lo largo de este trabajo:

- $d^{(j)}$ : personas fallecidas por la causa  $j \in \{1, 2, \dots, m\}$  pertenecientes a una cohorte o generación en estudio, de edades  $0, 1, 2, \dots, \omega - 1$ ,

$$d^{(j)} = \begin{bmatrix} d_0^{(j)} \\ d_1^{(j)} \\ d_2^{(j)} \\ \vdots \\ d_{\omega-1}^{(j)} \end{bmatrix}_{\omega \times 1}$$

- $d_x^{(j)}$ : personas fallecidas de edad  $x \in \{0, 1, 2, \dots, \omega - 1\}$  por la causa de muerte  $j$  en el transcurso del primer año.
- $l_x^{(\tau)}$ : personas de edad  $x \in \{0, 1, 2, \dots, \omega - 1\}$  que han logrado sobrevivir a todas las causas de muerte.

Relación importante:

$\sum_{j=1}^m d_x^{(j)} = d_x^{(\tau)}$ , donde  $\tau$  se entenderá como la “representación de todas las causas de muerte”, en consecuencia,  $d_x^{(\tau)}$  se explica como todas las personas fallecidas de edad  $x$  (por cualquier causa).

Por ende:

$q_x^{(j)}$ : es la probabilidad de que una persona de edad  $x$  fallezca por la causa  $j$  en el transcurso del primer año.

$$q_x^{(j)} = \frac{d_x^{(j)}}{l_x^{(\tau)}}$$

$q_x^{(\tau)}$ : probabilidad de que una persona de edad  $x$  fallezca por cualquier causa en el transcurso del primer año.

$$q_x^{(\tau)} = \sum_{j=1}^m q_x^{(j)}$$

$p_x^{(\tau)}$ : es la probabilidad de que una persona de edad  $x$  sobreviva a todas las causas de muerte al menos durante el primer año.

$$p_x^{(\tau)} = 1 - q_x^{(\tau)}$$

De los modelos de Decrementos Múltiples podemos obtener un modelo de Decremento Único mediante el siguiente planteamiento y usando la fuerza de cada decremento :

$\mu^{(j)}(x+t)$ : intensidad con la que causa baja de un grupo una persona de edad  $x+t$  por la causa  $j$ .

$$\mu^{(j)}(x+t) = \frac{1}{{}_t p_x^{(\tau)}} \cdot \frac{d {}_t q_x^{(j)}}{dt}$$

$$y \mu^{(\tau)}(x+t) = \sum_{j=1}^m \mu^{(j)}(x+t)$$

donde  $\mu^{(\tau)}(x+t)$  es la intensidad con la que causa baja de un grupo una persona de edad  $x+t$  por todas las causas.

$$\text{Por ende } {}_t p_x^{(\tau)} = \exp \left\{ - \int_0^t \mu^{(\tau)}(x+t) \right\} = \exp \left\{ - \int_0^t \sum_{j=1}^m \mu^{(j)}(x+t) \right\}$$

$$= \prod_{j=1}^m \exp \left\{ - \int_0^t \mu^{(j)}(x+t) \right\} = \prod_{j=1}^m {}_t p_x^{(j)}$$

Donde  ${}_t p_x^{(j)} = \exp \left\{ - \int_0^t \mu^{(j)}(x+t) \right\}$ , es la probabilidad de que una persona de edad  $x$  cause baja de un grupo en el transcurso de un tiempo  $t$ , sometido a una sola causa de baja.

Al igual que los valores actuariales definidos antes de redefinirlos para los modelos de decrementos múltiples, cada causa de muerte cuenta con su fuerza de decremento  $\mu_x^{(j)}$ , mediante el cual se pueden calcular las siguientes probabilidades y valores:

$${}_n q_x^{(\tau)} = 1 - {}_n p_x^{(\tau)} = 1 - \exp \left\{ - \int_0^n \mu^{(\tau)}(x+t) dt \right\}$$

$$\text{donde } \mu^{(\tau)}(x) = \sum_{j=1}^m \mu_x^{(j)}$$

después de esta última definición:

$$\begin{aligned}
 {}_n p_x^{(\tau)} &= \exp \left\{ - \int_0^n \sum_{j=1}^m \mu_{x+t}^{(j)} dt \right\} \\
 &= \exp \left\{ - \sum_{j=1}^m \int_0^n \mu_{x+t}^{(j)} dt \right\} = \prod_{j=1}^m \exp \left\{ - \int_0^n \mu_{x+t}^{(j)} dt \right\} = \prod_{j=1}^m {}_n p_x^{(j)}
 \end{aligned}$$

Los factores  ${}_n p_x^{(j)}$  corresponden a un modelo de decremento único, es decir, solo una causa de muerte es considerada.

Para estimar las  $q_x^{(j)}$  a partir de éstas últimas, se utiliza un método de aproximación que Bowers hace mención:

$$p_x^{(j)} = \left( p_x^{(\tau)} \right)^{q_x^{(j)} / q_x^{(\tau)}}$$

Despejando a  $q_x^{(j)}$  nos queda:

$$q_x^{(j)} = \frac{\ln(p_x^{(j)})}{\ln(p_x^{(\tau)})} \cdot q_x^{(\tau)}$$

## ii. Referencias históricas

En Roma aparece una de las mayores contribuciones al nacimiento y consolidación de los seguros: la organización de sociedades de enterramiento como forma rudimentaria de los actuales seguros de vida y de enfermedad. Además, es en Roma donde se ubican los antecedentes más importantes del seguro de vida en una norma por la que las viudas de los prestatarios de contratos de préstamos percibían una indemnización en forma de renta. En la Ley Falcidia (año 40 a.C.) aparece, por primera vez, el concepto de anualidad.

Los autores cuyas contribuciones se consideran esenciales en el desarrollo de las tablas de mortalidad son:

**Domitius Uloiano (230 d.C.).** Elaboró la conocida Tabla de Ulpiano, en la que aparecen reflejadas distintas edades asociadas a la esperanza de vida en años en cada una de ellas. Dicha tabla ha sido la más utilizada a lo largo de la historia para calcular las anualidades de rentas vitalicias.

**John Graunt (1662 d.C.).** En 1662, únicamente 5 años después de que Christsian Huygens publicara el primer texto escrito sobre la Teoría de Probabilidad (De Ratiociniis in Ludo AEDE), John Graunt publicó “Observations upon the Bills of Mortality”, trabajo posteriormente reconocido como el precursor de la Estadística Demográfica, en el que se incluye la primera tabla de mortalidad rudimentaria relativa a la población de Londres.

Los registros de mortalidad a los que tuvo acceso Graunt indicaban las causas de la muerte y el sexo de los difuntos pero no su edad. Por esto, registró la proporción de personas que morían de enfermedades infantiles (los cuales serían presumiblemente niños), añadiendo la mitad de las que morían de enfermedades como sarampión ó varicela que afectan tanto a niños como a adultos, concluyendo que 36 de cada 100 personas morían antes de los 6 años. El último dato de la tabla se le proporcionó la hipótesis de que nadie sobrevivía más de 76 años. Graunt no explica como obtuvo las filas intermedias pero Hacking (1995) considera la posible interpolación efectuada entre los 6 y los 76 años siguiendo una ley exponencial tomando como  $\mu = 0.047$  (tasa instantánea de mortalidad constante). Así, la función de supervivencia definida por este autor para las edades comprendidas entre 6 y 76 años es  $l(x) = 64 e^{-0.047(x - 6)}$

Debido a esta aportación, Graunt es conocido como el fundador de la Demografía.

**Edmund Halley (1693).** El famoso astrónomo, matemático y actuario inglés, Edmund Halley, quién calculo la órbita del cometa que lleva su nombre, fue el primero en construir en 1693 una tabla de mortalidad tal y como hoy en día las conocemos. Se basó en las estadísticas mortuorias (número de nacimientos y fallecimientos) de la ciudad alemana de Breslaw en un período de  $n$  años. Computó el número de personas de edad comprendida entre 0 y 1 año.

Lo mismo hizo para las personas comprendidas entre 1 y 2 años, y así sucesivamente. Posteriormente redujo los datos obtenidos en  $n$  años a un valor por cada período simplemente efectuando una media aritmética. El inconveniente que presenta el planteamiento de este autor es que supone la mortalidad constante, hipótesis falsa debido al efecto de los progresos médicos, higiénicos, etc. así como el factor de los movimientos migratorios (el de máxima importancia) que influyen y modifican seriamente la mortalidad.

Los cálculos de Halley fueron, después de los de De Wit, uno de los intentos más tempranos é importantes en el sentido de que han sido aplicados mucho más en la práctica. Este siglo XVII se considera enormemente fructífero para la Estadística Actuarial debido al desarrollo del cálculo de probabilidades y los avances en esta materia efectuados por Pascal, Fermat, Kepler, Galileo, Paccioli, Bernoulli, Bayes, Laplace, Markov y Kolmogorov, entre otros.



Durante los siglos siguientes (XVIII y XIX principalmente) cabe mencionar las aportaciones efectuadas a los estudios de mortalidad de Abraham De Moivre y los eminentes actuarios Gompertz y Makeham en el establecimiento y formulación de las leyes de mortalidad que llevan sus nombres.

De Moivre, en 1725, fue el primero en calcular una prima de seguros de vida, James Dobson, 50 años después, no solo calculaba primas para distintos tipos de seguros, sino, también reservas matemáticas. Estableció, por primera vez, un modelo global aplicable a la sistematización de una compañía de seguros de vida para garantizar su existencia y estabilidad.

### *iii.* Leyes de mortalidad

Las probabilidades de supervivencia  $S(x)$  permiten estimar funciones matemáticas que se resumen en modelos de comportamiento de las principales funciones biométricas que se expresan con base en la función de supervivencia y la tasa instantánea de mortalidad (fuerza de mortalidad). En la práctica actuarial se utilizan combinaciones de estas leyes aceptando diferentes modelos para distintos tramos de edades.

Las leyes de mortalidad son expresiones analíticas de la función de supervivencia que pretenden estimar el comportamiento de la mortalidad en función de la edad, siendo fundamental la elección de la función que mejor se adapte y represente adecuadamente la mortalidad, la que se hace según los

datos observados ó estableciendo ciertas hipótesis correspondientes a las características propias de la función de supervivencia.

Una constante a lo largo de la historia ha sido la búsqueda de una ley de mortalidad que sea válida para cualquier población humana, es decir, encontrar la “*Ley Universal de Mortalidad*” que probablemente no existe. Sin embargo, para determinadas poblaciones y ciertos tramos de edad, es posible encontrar el ajuste de alguna ley teórica.

Desde los tiempos de De Moivre (1725) se han propuesto expresiones matemáticas que representen la *fuera de mortalidad*, de las cuales la *Ley de Gompertz (1825)* podría ser la más usada y popular de ellas. Desde John Graunt (1620-1674) las tablas de mortalidad han sido construidas de forma empírica. Estas tablas representan la mortalidad durante la duración total de la vida humana y era natural preguntarse si las funciones matemáticas definidas a partir de las tablas de mortalidad pudieran ser descritas por leyes de mortalidad simples.

Siendo las *Leyes de mortalidad* intentos de agrupar las observaciones empíricas han sido íntimamente vinculadas con técnicas estadísticas para el análisis de los datos colectados en las tablas de mortalidad y pudieran ser descritas como formas paramétricas para datos cuantitativos resultantes de los modelos estadísticos.

Describiremos a continuación las principales *Leyes de mortalidad* y sus características:

La **Ley Exponencial**, supone que la tasa instantánea de mortalidad es constante:

$$\mu_x = \mu, \text{ donde } S(x) = e^{-\mu x} \quad \mu > 0, x \geq 0$$

Intuitivamente, es claro que la fuerza de mortalidad debe aumentar con la edad. Esto implica la imposibilidad de ajustar una ley exponencial a una población real, salvo en períodos de tiempo muy cortos. Sin embargo, la simplicidad de esta ley, la facilidad de sus cálculos y su importancia histórica hacen que sea comentada en primer lugar.

La **Ley de Moivre (1725)** incorpora el aumento de la fuerza de mortalidad con respecto a la edad mediante una fórmula matemática sencilla siendo la tasa instantánea de mortalidad:

$$\mu_x = \frac{1}{\omega - x} \text{ donde } S(x) = 1 - \frac{x}{\omega}, \text{ donde } \omega := \{\min x \mid l_x = 0\}$$

Con esta *Ley* la fuerza de mortalidad tiende hacia infinito cuando la edad lo hace hacia la edad límite  $\omega$ . La principal crítica que se le hace a esta ley es a su hipótesis de partida: número de fallecimientos constante (igual para todas las edades) é independiente de la edad.

La **primera Ley de Gompertz (1825)** asume que cada individuo presenta una resistencia a las enfermedades (y a fallecer por causas naturales) decreciente en función de la edad, por lo que la fuerza de mortalidad crece con la edad y su crecimiento relativo es constante. Por tanto, se deduce que dicha fuerza de mortalidad crece exponencialmente:

$$\mu_x = BC^x \text{ donde } S(x) = \exp\left\{-B \cdot \frac{(C^x-1)}{\ln(C)}\right\} \quad B > 0, C > 1$$

Posteriormente, Makeham enunció dos leyes de supervivencia. La **primera Ley de Makeham (1867)** considera la tasa instantánea de mortalidad añadiendo una constante arbitraria, que representa la mortalidad accidental (azar) independiente de la edad, a la fuerza de mortalidad de Gompertz. Por tanto, la muerte de un individuo es consecuencia de dos causas coexistentes: el azar y una resistencia (cada vez más débil) a la muerte conforme aumenta la edad, es decir, que además de considerar la mortalidad por causas naturales (igual que Gompertz) introduce la mortalidad accidental del individuo, independiente de la edad:

$$\mu_x = A + BC^x \text{ donde } S(x) = \exp\left\{-Ax - \left[B \cdot \frac{(C^x-1)}{\ln(C)}\right]\right\} \quad B > 0, C > 1, A > -B$$

Esta ley presenta buenos ajustes en edades intermedias (adultas), mientras que proporciona problemas en las edades extremas de la tabla, principalmente, en las edades más jóvenes puesto que en las edades infantiles la mortalidad es decreciente. Es considerada la ley más conocida y ampliamente utilizada para ajustar diversas tablas de supervivencia.

La primera ley de Makeham tiene problemas de ajuste para las edades más jóvenes, por lo que se formula la **segunda Ley de Makeham (1890)** más elástica y fundamentada que la anterior, añadiendo a la fuerza de mortalidad otro sumando proporcional a la edad:

$$\mu_x = A + Hx + BC^x \text{ donde } S(x) = \exp \left\{ -Ax - \left[ H \cdot \frac{x^2}{2} \right] - \left[ B \cdot \frac{(C^x - 1)}{\ln(C)} \right] \right\}$$

$$B > 0, C > 1, A > -B, H > 0$$

La **Ley de Weibull (1951)** que establece la siguiente expresión para la tasa instantánea:

$$\mu_x = Ax^B \text{ donde } S(x) = \exp \left\{ -A \cdot \frac{x^{B+1}}{B+1} \right\} \quad A > 0, B > 0$$

La **Ley Geométrica Doble (autor desconocido)** que considera a la fuerza de mortalidad:

$$\mu_x = A + BC^x + MN^x \text{ donde } S(x) = \exp \left\{ -Ax - \left[ B \cdot \frac{(C^x - 1)}{\ln(C)} \right] - \left[ M \cdot \frac{(N^x - 1)}{\ln(N)} \right] \right\}$$

$$B > 0, M > 0, A > -(B + M), C > 1, N > 1$$

Durante el siglo XIX podemos encontrar otras Leyes de mortalidad que destacan de forma importante por proponer una expresión matemática para representar a  $S(x)$  en vez de  $\mu_x$  como son las de **Babbage (1823)**, **Littrow (1832)** y **Moser (1839)**. Leyes de mortalidad más recientes las podemos

encontrar en el siglo XX como son las de **Perks (1932)**, **Beard (1931)**, **Barnett (1974)** y **Wilkie (1976)**, además de otras valiosas contribuciones a la Teoría de la Supervivencia.

## B. Causa de Muerte

### *i.* Necesidad de una clasificación por causa de muerte

El estudio de la mortalidad por causas constituye un elemento fundamental para la planeación y evaluación de los servicios y programas de salud. Identifica los principales problemas de salud de una población, mediante la magnitud y los efectos sobre las condiciones de vida de ésta.

El estudio de la mortalidad visto como el resultado de una suma de fallecimientos por cada causa de muerte que afectan a las poblaciones del mundo ha sido el interés no solo de las ciencias médicas o demográficas, sino, también de las económicas, políticas, financieras, etc. Basta citar un ejemplo como lo es la clasificación por causas de muerte propuesta por el Banco Mundial que las separa en tres grupos de acuerdo al tipo de intervenciones que se requieren: 1) Transmisibles, Maternas y Perinatales; 2) No Transmisibles; y 3) Lesiones y Accidentes. En el primer grupo se concentran las enfermedades susceptibles de ser reducidas con acciones de bajo costo y alta efectividad como vacunas, medidas sanitarias y diversas medidas comunitarias, generalmente proporcionadas por el primer nivel de atención (Centros de Salud y Brigadas Móviles). Dentro del segundo grupo se incluyen las enfermedades

crónico-degenerativas, mismas que requieren tratamientos más costosos y prolongados que corresponden al segundo y tercer nivel de atención (organismos de salud con profesionales y servicios más especializados), y que implican modificaciones en los estilos de vida de la población. Y las lesiones o accidentes que son causas potencialmente prevenibles mediante programas específicos promovidos y/o apoyados por el sector salud.

Una clasificación de enfermedades puede definirse como un sistema de categorías a las cuales se asignan entidades morbosas<sup>1</sup> de acuerdo con criterios establecidos. O como lo definió William Farr en 1856: *“Una clasificación es un método de generalización. Por lo tanto, diferentes clasificaciones pueden utilizarse con ventaja, según el usuario. Así, el médico, el patólogo o el jurista, cada uno con su propio punto de vista, puede legítimamente clasificar las enfermedades y las causas de defunción de una manera que, a juicio suyo, facilite sus investigaciones y permita obtener resultados generales”*.

Una clasificación estadística de enfermedades debe estar conformada por un número limitado de categorías mutuamente excluyentes, capaces de abarcar todo el rango de condiciones morbosas. Las categorías deben ser estructuradas de tal forma que permitan facilitar el estudio estadístico del fenómeno de la enfermedad. Una enfermedad específica que es de particular importancia en salud pública, o que se presenta frecuentemente, debe tener su propia categoría. Otras categorías pueden ser asignadas a grupos de

---

<sup>1</sup> Síntomas específicos de una enfermedad diferenciados de otros síntomas, ya sean, peculiares del paciente o de otra enfermedad.

afecciones distintas pero relacionadas. Cada enfermedad o entidad mórbida debe tener un lugar bien definido en las listas de categorías. Consecuentemente, a lo largo de la clasificación, debe haber categorías residuales para otras afecciones y para una miscelánea de afecciones que no pueden ser ubicadas en las categorías más específicas. Las categorías residuales deben ser tan pocas como sean posibles.

Una clasificación estadística puede permitir diferentes niveles de detalle si tiene una estructura jerárquica con subdivisiones. Una clasificación de esta naturaleza debe permitir la presentación de datos específicos hasta la agrupación de ellos de manera que permita que la información obtenida sea útil y comprensible.

#### *ii.* Definición del CEI-10

El propósito de la CIE (Clasificación Internacional de Enfermedades) es permitir el registro sistemático, el análisis, la interpretación y la comparación de los datos de mortalidad y morbilidad recolectados en diferentes países o áreas, y en diferentes épocas. La CIE se utiliza para convertir los términos diagnósticos y de otros problemas de salud, de palabras a códigos alfanuméricos que permiten su fácil almacenamiento y posterior recuperación para el análisis de la información.

En la práctica, la CIE se ha convertido en una clasificación diagnóstica estándar internacional para todos los propósitos epidemiológicos generales y muchos otros de administración de salud. Esto incluye el análisis de la



situación general de salud de grupos de población y el seguimiento de la incidencia y prevalencia de enfermedades y otros problemas de salud en relación con otras variables, tales como las características y circunstancias de los individuos afectados. La CIE no se propone ni es adecuada para indizar entidades clínicas individuales. Existen también ciertas limitaciones en cuanto el uso de la CIE en estudios de aspectos financieros, tales como asignación de recursos o facturación por servicios de salud.

La CIE puede utilizarse para clasificar enfermedades y otros problemas de salud consignados en muchos tipos de registros vitales y de salud. Originalmente su uso se limitó a clasificar las causas de mortalidad tal como se mencionan en los registros de defunción. Más tarde, su campo se extendió para incluir diagnósticos de morbilidad. Es importante destacar que, aunque la CIE se diseñó primariamente para clasificar enfermedades y traumatismos con un diagnóstico formal, no se puede categorizar de esta manera cada problema o razón para entrar en contacto con los servicios de salud. En consecuencia, la CIE ofrece alternativas adicionales para una amplia variedad de signos, síntomas, hallazgos anormales, quejas y circunstancias de tipo social que pueden ocupar el lugar del diagnóstico en los registros de salud. Por lo tanto, la Clasificación puede utilizarse para clasificar información registrada bajo denominaciones tales como diagnósticos, razones para la admisión, afecciones tratadas y motivo de la consulta, las que aparecen en una amplia variedad de registros, a partir de los cuales se derivan muchas estadísticas y otras informaciones sobre la situación de salud. Se ha sugerido también que la CIE debiera incluir clasificaciones que permitan la codificación de la información

adicional relacionada con el estado de salud o con la atención de la salud y por tal razón se ha propuesto la idea de una “familia” de clasificaciones de enfermedades y de otros problemas relacionados con la salud, publicada en volúmenes separados de la CIE.

La “parte central”, es decir, el núcleo de la Clasificación Internacional de Enfermedades, está constituido por los códigos de tres caracteres, los cuales son el mínimo obligatorio que exige la OMS para formar la base de datos sobre mortalidad, y para hacer comparaciones internacionales. Las categorías de cuatro caracteres, aunque no son obligatorias para la información internacional, se recomiendan para varios propósitos y forman una parte integral de la CIE.

Hay dos tipos principales de clasificación. El primer grupo cubre datos relacionados con los diagnósticos y el estado de salud. También incluye clasificaciones complementarias que permiten la ubicación de diagnósticos utilizando ejes diferentes de clasificación, tales como el de la morfología de los tumores. El segundo grupo de clasificaciones cubre aspectos relacionados con problemas de salud, que no son generalmente diagnósticos formales o afecciones existentes, así como las clasificaciones relacionadas con la atención de la salud. Este grupo incluye clasificaciones de discapacidades, de procedimientos médicos y quirúrgicos, y de razones para el contacto con los proveedores de servicios de salud.

La Clasificación Internacional de Enfermedades que hoy día tenemos (CIE-10) se debe al Instituto Internacional de Estadística, sucesor del Congreso

Internacional de Estadística, en su reunión en Viena en 1891 en la cual se definió formar una clasificación por causa de muerte (que más tarde quedaría definida como la CIE: Clasificación Internacional de Enfermedades) denominada *Clasificación de Causas de Defunción de Bertillon*, misma que fue usada por primera vez en Norte América por Jesús E. Monjarás, en las estadísticas de San Luis Potosí, México en 1898. En ese mismo año, en una reunión celebrada en Canadá por parte de la Asociación Americana de Salud Pública recomendó que los registros civiles de Canadá, México y Estados Unidos adoptaran la clasificación de Bertillon con una supervisión de esta última cada 10 años.

Más adelante, se consideró que en ausencia de una clasificación uniforme de enfermedades que pudiera utilizarse satisfactoriamente para estadísticas de morbilidad, muchos países consideraron necesario preparar sus propias listas realizando agrupaciones de las causas de defunciones definidas. A través de los siguientes años, los países que contribuyeron de mayor forma en la definición de una lista más completa fueron EUA, Canadá e Inglaterra a través de las asociaciones Médicas de cada uno de estos países. Para 1975, durante la Conferencia Internacional para la Novena revisión de la CIE, en esta ocasión convocada por la OMS, se centro en la modificación de la clasificación de enfermedades e introduciendo posibilidades especiales de codificación ya que varios países habían adoptado la CIE para su propio uso. Se manejaba que era necesario una clasificación más completa para cada especialidad ya que se consideraba que varias partes no estaban organizadas adecuadamente: se

requería más detalle en el tipo de afección y realizar la agrupación en base a los sistemas orgánicos afectados.

Para la décima revisión de la CIE, uno de los desafíos de la Estrategia Mundial de Salud para Todos en el Año 2000 es proveer información de apoyo para la atención primaria de salud. En países que tienen información incompleta o de baja calidad, se necesita una variedad de enfoques para adaptar, suplementar o reemplazar el uso convencional de la CIE.

Desde el final de la década de 1970, varios países han experimentado con la recolección de información por personal no médico (lego). La notificación por personal no médico ha sido subsecuentemente extendida a un concepto amplio denominado "*métodos no convencionales*". En diferentes países estos métodos, que cubren una variedad de enfoques, han evolucionado como medios de obtener información acerca del estado de salud de la población donde los métodos convencionales (censos, encuestas, estadísticas vitales y estadísticas de morbilidad) han resultado inadecuados.

Uno de tales enfoques, denominado "*información basada en la comunidad*", implica la participación de la comunidad en la definición, recolección y utilización de la información relacionada con la salud. El grado de participación de la comunidad va desde el trabajo de recolección de los datos solamente hasta el diseño, el análisis y la utilización de la información. La experiencia obtenida en algunos países ha demostrado que este enfoque es más que un marco teórico.

Se dio cuenta a la Conferencia Internacional para la Décima Revisión de la Clasificación Internacional de Enfermedades (CIE) de la experiencia de varios países en la generación y aplicación de información de salud basada en la comunidad que comprendía problemas y necesidades de salud, factores de riesgo relacionados y recursos. Esto apoyaba el concepto de crear métodos no convencionales en la comunidad como una manera de colmar las lagunas de información en los países y de fortalecer sus sistemas de información. Se subrayó que, lo mismo para los países en desarrollo que para los desarrollados, dichos métodos o sistemas deben crearse localmente y que, a causa de factores como los patrones de morbilidad y las variaciones de idioma y culturales, no debe intentarse su transferencia a otras zonas o países.

Dados los resultados alentadores de este enfoque en muchos países, la Conferencia acordó que la OMS debería continuar orientando el desarrollo de esquemas locales y apoyando los progresos de esta metodología. En este sentido, la OMS promueve el desarrollo de adaptaciones que extienden tanto la utilidad de la CIE como la comparación de las estadísticas de salud. El papel de la OMS en el desarrollo de nuevas clasificaciones, adaptaciones y glosarios, es el de proveer un liderazgo cooperativo y el de actuar como una entidad que da orientación técnica y apoyo donde sea necesario. Los interesados en preparar adaptaciones de la CIE-10 deben consultar con la OMS tan pronto se tengan claros los objetivos de la adaptación. De esta manera se evitarán duplicaciones innecesarias, mediante un enfoque coordinado para el desarrollo de los varios componentes de la familia de clasificaciones.

La CIE-10 (10<sup>ma</sup> revisión de la CIE) comprende tres volúmenes: el Volumen 1 contiene las clasificaciones principales, el Volumen 2 provee orientación a los usuarios de la CIE y el Volumen 3 es el Índice alfabético de la clasificación.

## II. Métodos del análisis multivariado de datos

### A. Componentes Principales

#### i. Determinación de las componentes

Es frecuente el caso en que se tiene una colección de registros, cada uno de cuyos integrantes puede ser descrito por un vector  $D$ , de dimensión  $1 \times m$ , es decir, de  $m$  variables o características de cada registro. En tales casos, es también frecuente que entre las diferentes características o variables componentes del vector  $D$  exista cierta correlación, que, en el caso más extremo, haría que alguna de estas variables fuera combinación lineal exacta de otra u otras. En tales casos, surge de modo natural la pregunta de si no sería más útil tomar un subconjunto de las variables originales, o quizá un número reducido de variables compuestas, transformadas de las originales que describiera el colectivo sin gran pérdida de información. Un ejemplo de un vector  $D$  sería el compuesto por los fallecidos por cada causa de muerte:

Fallecidos por Cáncer	Fallecidos por Neumonía	Fallecidos por Accidente	Fallecidos por un EVC <sup>2</sup>
$d^{(1)}$	$d^{(2)}$	$d^{(3)}$	$d^{(4)}$

En el caso de tomar el subconjunto propuesto de estas variables la pérdida de información se quisiera fuera la menor para la descripción del fenómeno en estudio, que en este caso se trata de la mortalidad. Si obtuviéramos nuevas variables, mismas que fueran una transformación de las primeras, aparte de interesarnos no perder información significativa, nos interesaría saber cuáles de las nuevas variables escoger.

---

<sup>2</sup> EVC: evento vascular cerebral

Supongamos que la siguiente matriz  $R$  es la matriz de correlaciones de las variables arriba definidas:

$$R = \begin{bmatrix} 1 & 0.25 & 0.41 & 0.09 \\ 0.25 & 1 & 0.93 & 0.87 \\ 0.41 & 0.93 & 1 & 0.43 \\ 0.09 & 0.87 & 0.43 & 1 \end{bmatrix}$$

Podemos observar que las variables<sup>3</sup>  $d^{(2)}$  y  $d^{(3)}$  de dimensión  $\omega \times 1$  poseen una alta correlación lineal infiriendo que pudieran explicarse una con la otra de forma lineal, así como las variables  $d^{(2)}$  y  $d^{(4)}$ . Concluiríamos que sería posible simplificar la información de tres de estas variables por la de una sola y obtener una gran cantidad de información utilizando solo dos variables en total en vez de cuatro. El problema que esto conlleva es que omitir variables, en ocasiones, no es viable pues por ejemplo en esta investigación se quiere determinar el efecto que cada causa de muerte tiene sobre la mortalidad poblacional, así como su evolución a través del tiempo.

Para obtener las componentes principales de una tabla de fallecimientos por causa de muerte usaremos el vector estandarizado de  $d^{(j)}$  definido como:

$${}_z d^{(j)} = \frac{d^{(j)} - \bar{d}^{(j)}}{S_{d^{(j)}}}$$

$$\text{donde } S_{d^{(j)}} = \sqrt{\frac{\sum_{x=0}^{\omega-1} (d_x^{(j)} - \bar{d}^{(j)})^2}{\omega - 1}}$$

$$\bar{d}^{(j)} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{\omega \times 1} \cdot \frac{\bar{d}^{(j)}}{\omega} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{\omega \times 1} \cdot \frac{\sum_{x=0}^{\omega-1} d_x^{(j)}}{\omega}$$

<sup>3</sup> Ver *Tabla A* para visualizar un ejemplo de esta información.



Consideraremos a los componentes principales como las siguientes variables conformadas por:

$$\begin{aligned}
 C^1 &= [ {}_z d^{(1)} \quad {}_z d^{(2)} \quad \dots \quad {}_z d^{(m)} ]_{\omega \times m} \cdot \alpha_1 = \alpha_{11} {}_z d^{(1)} + \alpha_{21} {}_z d^{(2)} + \dots + \alpha_{m1} {}_z d^{(m)} \\
 C^2 &= [ {}_z d^{(1)} \quad {}_z d^{(2)} \quad \dots \quad {}_z d^{(m)} ]_{\omega \times m} \cdot \alpha_2 = \alpha_{12} {}_z d^{(1)} + \alpha_{22} {}_z d^{(2)} + \dots + \alpha_{m2} {}_z d^{(m)} \\
 &\quad \vdots \\
 C^{m'} &= [ {}_z d^{(1)} \quad {}_z d^{(2)} \quad \dots \quad {}_z d^{(m)} ]_{\omega \times m} \cdot \alpha_{m'} = \alpha_{1m'} {}_z d^{(1)} + \alpha_{2m'} {}_z d^{(2)} + \dots + \alpha_{mm'} {}_z d^{(m)} \\
 &\quad \text{donde } D = [ {}_z d^{(1)} \quad {}_z d^{(2)} \quad \dots \quad {}_z d^{(m)} ]_{\omega \times m}, \text{ es decir: } C^j = D \cdot \alpha_j \\
 &\quad \text{y por tanto: } [C^1, C^2, \dots, C^{m'}] = D \cdot [\alpha_1, \alpha_2, \dots, \alpha_{m'}] = D \cdot \alpha = C, \text{ donde } m \geq m'
 \end{aligned}$$

Una vez planteadas la matrices  $\alpha$  de dimensión  $m \times m'$  y  $C$  de dimensión  $\omega \times m'$  de los componentes principales, el problema radica en la elección de los vectores de coeficientes  $\alpha_j$  de dimensión  $m \times 1$  que permitan obtener a los  $C^j$  como combinaciones lineales de las variables originales  ${}_z d^{(j)}$ .

El primer componente  $C^1$  se define como la combinación lineal de las variables originales que tiene **varianza máxima**. Como las variables  ${}_z d^{(j)}$  tienen media 0, entonces su varianza se definirá como:

$$\frac{1}{m} \cdot C^{1T} C^1 = \frac{1}{m} \cdot \alpha_1^T D^T D \alpha_1 = \alpha_1^T S \alpha_1$$

donde  $S$  es la matriz de varianzas y covarianzas de  $D$ . Es claro que podemos maximizar la varianza sin límite aumentando el módulo<sup>4</sup> del vector  $\alpha_1$ . Para maximizar a  $\alpha_1^T S \alpha_1$  debemos imponer una restricción al módulo del vector  $\alpha_1$  tal que  $\alpha_1^T \alpha_1 = 1$ . Introduciremos esta restricción mediante el multiplicador de Lagrange:

$$M = \alpha_1^T S \alpha_1 - \lambda (\alpha_1^T \alpha_1 - 1)$$

---

<sup>4</sup> Módulo de un vector  $X_{1 \times n}$  se define como  $\|X\|^2 = \sqrt{X \cdot X^T}$

y maximizaremos esta expresión derivando con respecto a los componentes de  $\alpha_1$  e igualando a 0. Entonces:

$$\frac{\partial M}{\partial \alpha_1} = 2S\alpha_1 - 2\lambda\alpha_1 = 0$$

cuya solución es:

$$S\alpha_1 = \lambda\alpha_1$$

que implica que  $\alpha_1$  es un vector propio de la matriz  $S$  y  $\lambda$  su correspondiente valor propio. Para determinar que valor propio es la solución de la relación anterior, multiplicamos por la izquierda por  $\alpha_1^T$ :

$$\alpha_1^T S \alpha_1 = \lambda \alpha_1^T \alpha_1 = \lambda$$

y concluimos que  $\lambda$  es la varianza de  $C^1$ . Como esta es la cantidad que queremos maximizar,  $\lambda$  será el mayor valor propio de la matriz  $S$ . Su vector asociado  $\alpha_1$ , define los coeficientes de cada variable en el primer componente principal.

Para el cálculo del 2<sup>do</sup> componente, ahora lo realizaremos modificando la función objetivo donde la suma de las varianzas de  $C^1 = D \cdot \alpha_1$  y  $C^2 = D \cdot \alpha_2$  sea máxima. La función objetivo la definimos así:

$$\phi = \alpha_1^T S \alpha_1 + \alpha_2^T S \alpha_2 - \lambda_1(\alpha_1^T \alpha_1 - 1) - \lambda_2(\alpha_2^T \alpha_2 - 1)$$

que incorpora las restricciones de que los vectores  $\alpha_1$  y  $\alpha_2$  deben de tener módulos unitarios, es decir,  $\alpha_1^T \alpha_1 = 1$  y  $\alpha_2^T \alpha_2 = 1$ .

Derivando con respecto a  $\alpha_1$  y  $\alpha_2$  e igualando a 0 obtenemos:

$$\frac{\partial \phi}{\partial \alpha_1} = 2S\alpha_1 - 2\lambda_1\alpha_1 = 0$$

$$\frac{\partial \phi}{\partial \alpha_2} = 2S\alpha_2 - 2\lambda_2\alpha_2 = 0$$

La solución de este sistema es:

$$S\alpha_1 = \lambda_1\alpha_1$$

$$S\alpha_2 = \lambda_2\alpha_2$$

que indica que  $\alpha_1$  y  $\alpha_2$  deben ser vectores propios de  $S$ . Tomando los vectores propios de norma 1 y sustituyendo el máximo en nuestra función objetivo obtenemos  $\phi = \lambda_1 + \lambda_2$ . Es claro que  $\lambda_1$  y  $\lambda_2$  deben ser los dos autovalores (o eigenvalores) mayores de la matriz  $S$ ;  $\alpha_1$  y  $\alpha_2$  sus correspondientes autovectores. La covarianza entre  $C^1$  y  $C^2$ , dada por  $\alpha_1^T S \alpha_2$  es 0 ya que  $\alpha_1^T \alpha_2 = 0$ , y las variables  $C^1$  y  $C^2$  estarán incorreladas.

La generalización de este método resulta en que la matriz  $D$  de rango  $m$ , y por tanto la matriz  $S$  también, a través de esta última se obtendrán sus valores propios o raíces características  $\lambda_1, \dots, \lambda_p$ :

$$|S - \lambda I| = 0$$

y sus vectores propios asociados  $\alpha_1, \dots, \alpha_p$ :

$$(S - \lambda_i I)\alpha_i = 0$$

Los términos  $\lambda_i$  son reales, al ser la matriz  $S$  simétrica, y positivos, ya que  $S$  es definida positiva. Por ser  $S$  simétrica si  $\lambda_j$  y  $\lambda_h$  son dos raíces distintas sus vectores asociados son ortogonales. Si  $S$  fuese semidefinida positiva de rango  $r < m$ , lo que ocurriría si  $m-r$  variables fuesen combinación lineal de las demás, habría solamente  $r$  raíces características positivas y el resto serían ceros.

Llamando  $C$  a la matriz cuyos componentes en los  $\omega$  observaciones, estas nuevas variables están relacionadas con las originales mediante:

$$C_{\omega \times m} = DA$$

Como  $A_{m \times m} = [\alpha_1, \dots, \alpha_m]$ ; por tanto  $A^T A = I$ . Calcular los componentes principales equivale a aplicar una transformación ortogonal  $A$  a las variables en  $D$  para obtener unas nuevas variables en  $C$  incorreladas entre sí.

## ii. Propiedades de las componentes

Las propiedades de los componentes obtenidos son:

1. Conservan la variabilidad inicial: la suma de las varianzas de los componentes es igual a la suma de las varianzas de las variables originales, y la varianza generalizada de los componentes es igual a la original.
2. La proporción de la variabilidad explicada por un componente es el cociente entre su varianza, el valor propio asociado al vector propio que lo define, y la suma de los valores propios de la matriz.
3. Las covarianzas entre cada componente principal y las variables  $z d^{(j)}$  vienen dadas por el producto de las coordenadas del vector propio que define el componente por su valor propio:

$$Cov(C^j; z d^{(1)}, \dots, z d^{(m)}) = \lambda_j \alpha_j = (\lambda_j \alpha_{j1}, \dots, \lambda_j \alpha_{jm})$$

donde  $\alpha_j$  es el vector de coeficientes del la componente  $C^j$ .

4. La correlación entre un componente principal y una variable  $z d^{(j)}$  es proporcional al coeficiente de esa variable en la definición del

componente, y el coeficiente de proporcionalidad es el cociente entre la desviación típica del componente y la desviación típica de la variable.

5. Las  $r$  componentes principales ( $r < m$ ) proporcionan la predicción lineal óptima con  $r$  variables del conjunto de variables  ${}_z d^{(j)}$ .

Si estandarizamos los componentes principales, dividiendo cada uno por su desviación típica, se obtiene la estandarización multivariante de los datos originales.

### *iii.* Selección del número de componentes

Se sugieren distintas reglas para seleccionar el número de componentes a utilizar:

- a) Realizar un gráfico de  $\lambda_i$  frente a  $i$ . Comenzar seleccionando componentes hasta que los restantes tengan aproximadamente el mismo valor de  $\lambda_i$ . La idea es buscar un “codo” en el gráfico, i.e., un punto a partir del cual los valores propios son aproximadamente iguales. El criterio es quedarse con un número de componentes que excluya los asociados a valores pequeños y aproximadamente del mismo tamaño.
- b) Seleccionar componentes hasta cubrir una proporción determinada de varianza, como el 80 o 90%. Esta regla es arbitraria y debe aplicarse con cierto cuidado. Por ejemplo, es posible que un único componente recoja el 90% de la variabilidad y, sin embargo, pueden existir otros componentes que sean muy adecuados para explicar variabilidad adicional.

- c) Desechar aquellos componentes asociados a valores propios inferiores a una cota, que suele fijarse como la varianza media,  $\sum \lambda_i/p$ . En particular, cuando se trabaja con la matriz de correlación, el valor medio de las componentes es 1, y esta regla lleva a seleccionar los valores propios mayores que la unidad. De nuevo, esta regla es arbitraria: una variable que sea independiente del resto suele llevarse un componente principal y puede tener un valor propio mayor que la unidad. Sin embargo, si está incorrelada con el resto puede ser una variable poco relevante para el análisis y no aportar mucho a la comprensión del fenómeno global.

## B. Regresión Logística

### i. Modelo de Regresión Logística

Los modelos de regresión se han convertido en una herramienta integral de cualquier análisis de datos en el que se busque encontrar la relación entre una variable de respuesta o dependiente y una o más variables explicatorias o independientes.

El objetivo principal de los modelos de regresión es construir un modelo que ajuste a los datos observados conservando el principio de *parsimonia*, que sea un modelo razonablemente descriptivo de la relación existente entre la variable dependiente o de respuesta y la o las variables predictoras o explicatorias. Los modelos de regresión logística siguen los mismos principios que los modelos

de regresión lineal, con la diferencia de que la variable de respuesta que se quiere explicar será discreta tal como sigue:

$P(Y = 1|X_i) = q_i = \pi(X_i)$ , donde  $\pi(X_i)$  denota la probabilidad de que una persona de edad  $x$  fallezca entre las edades  $x$  y  $x + 1$ , bajo las características del vector fila  $X_i$ .

$$Y = \begin{cases} 1, & \text{si } K(x) = 0 \\ 0, & \text{si } K(x) > 0 \end{cases}$$

$K(x)$ : variable aleatoria discreta que describe el número de años enteros sobrevividos por una persona de edad  $x$ , donde  $K(x) = 0, 1, 2, \dots$

El modelo de regresión logística está dado por la ecuación:

$$g(X_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$$

en cuyo caso el modelo de regresión logística es:

$$\pi(X_i) = \frac{1}{1 + e^{-g(X_i)}}$$

Si  $Y$  es codificada como 0 o 1 entonces la expresión para  $\pi(X_i)$  denota la probabilidad condicional que  $Y = 1$  dado  $X_i$ . Esto se denotará como  $P(Y = 1|X_i)$ . Se sigue que entonces  $1 - \pi(X_i)$  sería  $P(Y = 0|X_i)$ .

De esta manera, para los pares  $(X_i, Y_i)$ , donde  $Y_i = 1$  entonces es  $\pi(X_i)$ , y para los pares en los que  $Y_i = 0$ , es  $1 - \pi(X_i)$ .

Supongamos que tenemos una muestra de  $n$  observaciones independientes  $(X_i, Y_i)$ ,  $i = 1, 2, \dots, n$ . Para obtener el modelo de regresión logística es necesario

estimar los coeficientes  $\beta^T = [\beta_0, \beta_1, \dots, \beta_p]$ . El método de estimación para este vector será el de máxima verosimilitud.

Sea  $X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1,p-1} \\ 1 & x_{21} & x_{22} & \dots & x_{2,p-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{n,p-1} \end{bmatrix}$  la matriz de datos asociada a las

observaciones de cada  $\pi(X_i)$ , donde  $X_i = [1, x_{i1}, x_{i2}, \dots, x_{i,p-1}]$ , entonces para

estimar el vector  $\beta^T = [\beta_0, \beta_1, \dots, \beta_{p-1}]$  del modelo de regresión logística

$\pi(X_i) = \frac{1}{1+e^{-g(X_i)}}$ , es necesario realizar una transformación a  $\pi(X_i)$  de la

siguiente manera:

$$G_i = -\ln\left(\frac{1 - \pi(X_i)}{\pi(X_i)}\right) = -\ln\left(\frac{P(Y = 0|X_i)}{P(Y = 1|X_i)}\right) = g(X_i)$$

$g(X_i)$  puede ser expresado como  $X_i\beta = \beta_0 + \beta_1x_{i1} + \beta_2x_{i2} + \dots + \beta_{p-1}x_{i,p-1}$ , por

lo que ahora, el objetivo es calcular el vector  $\beta$  para el cual se minimicen los errores.

## ii. Coeficientes de regresión

Nuestro nuevo modelo es  $G_i = X_i\beta + \epsilon_i$ , donde  $\epsilon_i$  es el error entre el dato observado y el dato estimado.

De manera más general, nuestro modelo será  $G = X\beta + \epsilon$ , donde:

- $G$  es el vector de  $(n \times 1)$  de la transformación de los datos observados
- $X$  es la matriz de  $(n \times p)$ , donde cada  $X_i \perp X_j \forall i \neq j$
- $\beta$  es el vector de parámetros de  $(p \times 1)$
- $\epsilon$  el vector de los errores de  $(n \times 1)$



donde  $E(\epsilon) = 0$ ,  $V(\epsilon) = I\sigma^2$ , por tanto los elementos de  $\epsilon$  están no correlacionados, i.e.,  $Corr(\epsilon_i, \epsilon_j) = 0 \forall i \neq j$ , y  $\epsilon \sim Nor(0, I\sigma^2)$ . Como  $E(\epsilon) = 0$ , una forma alterna de escribir el modelo es  $E(G) = X\beta$ .

La suma de los cuadrados de los errores es entonces:

$$\begin{aligned}\epsilon^T \epsilon &= (G - X\beta)^T (G - X\beta) \\ &= G^T G - \beta^T X^T G - G^T X\beta + \beta^T X^T X\beta \\ &= G^T G - 2\beta^T X^T G + \beta^T X^T X\beta\end{aligned}$$

Usando el método de máxima verosimilitud y diferenciando sobre el vector  $\beta$ , la solución que minimiza  $\epsilon^T \epsilon$  es:

$$\hat{\beta} = (X^T X)^{-1} X^T G$$

Esta solución tiene las siguientes propiedades:

1. Es el estimado de  $\beta$  que minimiza la suma del cuadrado de los errores.
2. Los elementos de  $\beta$  son funciones lineales de las observaciones  $G_1, G_2, \dots, G_n$  y son estimadores insesgados de mínima varianza de los  $\beta_i$ .
3. Dado que  $Corr(\epsilon_i, \epsilon_j) = 0 \forall i \neq j$ , entonces  $\hat{\beta}$  es el estimador máximo verosímil de  $\beta$ .

### iii. Supuestos del modelo de regresión

El modelo general de regresión lineal posee algunos supuestos, sin los cuales no sería posible considerar el modelo propuesto como el indicado. Los supuestos que se deben de cumplir son:

1. Linealidad, es decir, el modelo especifica una relación lineal entre  $G$  y  $x_1, x_2, \dots, x_p$ , donde  $x_j$  es un vector columna de  $n \times 1$ :

$$G_i = X_i\beta + \epsilon_i$$

2. Rango completo de la matriz de datos  $X$ . No existe una relación lineal exacta entre ninguna de las variables independientes.
3. Exogeneidad de las variables independientes, es decir, que el valor esperado de cualquier error  $i$  observado en la muestra no es función de las variables independientes observadas, ni siquiera la observación  $i$ . Esto significa que las variables independientes no llevan consigo información valiosa para pronosticar a  $\epsilon_i$ :

$$E[\epsilon_i | x_{ij}, j = 1, \dots, p] = 0$$

4. Homocedasticidad y no existencia de autocorrelación, i.e., cada  $\epsilon_i$  tiene la misma varianza finita  $\sigma^2$  y no está correlacionada con cualquier otro  $\epsilon_j$ .
5. Normalidad de los errores, i.e.,  $\epsilon \sim N(\bar{0}, \sigma^2 I)$ .

Este último punto implica también implica el cumplimiento del punto 4 de los supuestos.

Sin el cumplimiento de los supuestos antes mencionados no es posible considerar a nuestro modelo como el correcto para pronosticar o proyectar por lo que será necesario detectar, y en su caso, corregir mediante diversas técnicas estadísticas nuestros datos observados, tanto nuestra matriz  $X$  como nuestra matriz  $G$ .

Para el primer supuesto, es obvio que si nuestro modelo no tiene un suficiente ajuste a los datos observados sería un error considerarlo para representarlos. Algunos estadísticos nos ayudarán a evaluar el ajuste de nuestro modelo y el cumplimiento de los supuestos antes mencionados.

El primero de ellos el estadístico  $R^2$  denominado como coeficiente de determinación y definido por:

$$R^2 = \frac{\sum_{i=1}^n (\hat{G}_i - \bar{G})^2}{\sum_{i=1}^n (G_i - \bar{G})^2}$$

Con este estadístico conocemos el cuadrado del coeficiente de correlación múltiple entre  $\hat{G}$  y  $G$ , donde  $0 \leq R^2 \leq 1$ ; mientras más tienda a 0, menos ajusta el modelo a los datos observados, caso contrario, mientras más tienda a 1, mejor ajusta el modelo a los datos.

El problema del estadístico  $R^2$  es que puede ser engañoso ya que pudiéramos incluir en el modelo cuantas variables explicativas obtuviéramos y poder lograr un modelo con un estadístico  $R^2 = 1$ . Por esta situación tenemos un estadístico más confiable para medir el ajuste del modelo, este es la  $R^2$  *ajustada*.

Supongamos que  $p$  es el total de parámetros en un modelo de regresión, entonces el estadístico  $R^2$  *ajustada* es definido por:

$$R_a^2 = 1 - (1 - R^2) \left( \frac{n-1}{n-p} \right)$$

El ajuste hecho para este estadístico es sobre los grados de libertad. La idea central es poder hacer un comparativo entre diferentes modelos usando

colecciones de datos diferentes para estimar un conjunto de datos observados, de ahí conservar el principio de parsimonia, un modelo sencillo pero efectivo.

Para el supuesto de que la matriz de datos  $X$  sea de rango completo, deriva que de no serlo,  $(X^T X)^{-1}$  no existiría dado que algunas filas de  $X$  pudieran ser combinación lineal de otras. De aquí surge en el modelo el problema de la multicolinealidad, considerando el modelo  $G = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_{p-1}$ , donde  $X_j = [x_{ij}]_{i=1}^n$  matriz de  $n \times 1$ . La varianza del j-ésimo coeficiente de regresión estimado es:

$$Var[\hat{\beta}_j] = \sigma^2 \left( \frac{1}{1 - R_j^2} \right) \left( \frac{1}{S_{X_j X_j}} \right)$$

donde  $S_{X_j X_j} = \sum_{i=1}^n (x_{ij} - \bar{x}_{.j})^2$  y  $\bar{x}_{.j} = \sum_{i=1}^n x_{ij}$ . El coeficiente de determinación  $R_j^2$  representa el ajuste de un modelo de regresión lineal de  $X_j$  contra todas las demás variables independientes o predictoras. El factor  $\left( \frac{1}{1 - R_j^2} \right)$  es llamado el j-ésimo Factor de inflación de varianza o  $VIF_j$ . Si  $R_j^2$  es cercano a 1 entonces la varianza de  $\hat{\beta}_j$  aumentará grandemente. El  $VIF_j$  representa el incremento en la varianza debido a la presencia de multicolinealidad. Una variable predictora con un  $VIF_j$  mayor a 10 (esto es equivalente a aceptar un  $R^2 = 0.90$  es un indicador de una buena relación lineal) puede causar multicolinealidad. Los  $VIF$  son los elementos de la diagonal de la matriz de correlaciones de las variables predictoras.

Para diagnosticar la multicolinealidad se pueden seguir los siguientes pasos:

1. Cotejar si hay coeficientes de regresión con valores bien grandes o de signo opuesto a lo que se esperaba que ocurriera.
2. Cotejar si las variables predictoras que se esperaban sean importantes tienen valores de  $t$  pequeños para las hipótesis de sus coeficientes.
3. Cotejar si la eliminación de una fila o columna de la matriz  $X$  produce grandes cambios en el modelo ajustado.
4. Cotejar las correlaciones entre todas las parejas de variables predictoras para detectar las que son bastante altas.
5. Examinar el  $VIF$ . Si el  $VIF$  es grande, mayor que 10, entonces puede haber multicolinealidad.
6. Usar el número condición de la matriz correlación  $X^{*T}X^*$ , la cual es de la forma:

$$\begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}$$

donde  $r_{ij}$  representa la correlación entre las variables  $X_i$  y  $X_j$ . La matriz  $X^*$  es obtenida restando a cada columna de  $X$  la media de cada variable columna y dividiendo luego entre la raíz de la suma de cuadrados corregida por la media de la misma columna.

El número condición de la matriz  $X^*$  está definida por:

$$K = \left( \frac{\text{Mayor eigenvalue}}{\text{Menor eigenvalue}} \right)^{\frac{1}{2}}$$

Los “*eigenvalues*” son de la matriz de correlación. Notar que  $K \geq 1$ . Se sugiere que si  $K > 30$  entonces hay multicolinealidad.

7. Otra alternativa para detectar multicolinealidad es proponer la hipótesis estadística  $H_0: \beta_j = \beta_i + \beta_k + \dots + \beta_r$ , y si  $H_0$  no se rechaza entonces es posible que la variable fila  $X_j$  sea combinación lineal de  $X_i, X_k, \dots, X_r$ .

Para remediar el problema de la multicolinealidad podemos recurrir a las siguientes opciones:

- a. Transformación de los datos de la variable columna  $X_j$  mediante una función  $f(x_{ij})$  tal que evitemos la relación lineal de  $X_j$  con otras variables.
- b. Regresión Ridge.
- c. Componentes principales, el cual utilizaremos en este trabajo no solo para evitar el problema de la multicolinealidad, sino, para reducir la cantidad de variables en el análisis y para evitar la no inclusión de alguna de las variables originales.
- d. Mínimos cuadrados parciales.

En el supuesto donde se pide que el valor esperado de cualquier error  $i$  observado en la muestra no es función de las variables independientes observadas, i.e.,  $E[\epsilon_i | x_{ij}, j = 1, \dots, p] = 0$ , en caso que no fuera así el problema, entre otros más que surgen al estimar los coeficientes de regresión sería que si  $E[\epsilon_i | x_{ij}, j = 1, \dots, p] \neq 0$  entonces como  $y_i = \hat{y}_i + \epsilon_i$  implica que:

$$\begin{aligned} E[y_i | x_{ij}, j = 1, \dots, p] &= E[\hat{y}_i + \epsilon_i | x_{ij}, j = 1, \dots, p] \\ &= E[\hat{y}_i | x_{ij}, j = 1, \dots, p] + E[\epsilon_i | x_{ij}, j = 1, \dots, p] = X_i \beta + X_i E \end{aligned}$$

donde  $E$  es el vector de coeficientes del modelo de regresión:

$$\epsilon_i = E_0 + E_1 x_{i1} + E_2 x_{i2} + \dots + E_{p-1} x_{i,p-1} + r_i$$

Esto implica que  $E[y_i|x_{ij}, j = 1, \dots, p] = X_i(\beta + E)$ . Lo que quiere decir es que la matriz  $\beta$  inicialmente estimada no pondera de manera correcta a cada una de las variables predictoras, que en caso de que  $E[r_i|x_{ij}, j = 1, \dots, p] \neq 0$ , el proceso sería iterativo para  $r_i$ .

Por tanto, los coeficientes de la matriz  $\beta$  no representan correctamente el peso de cada una de las variables predictoras  $x_j$ .

Otra forma de encontrar la no dependencia de los errores con las variables predictivas y al mismo tiempo detectar heterocedasticidad, i.e., varianza no constante de los errores del modelo, es formulando la siguiente hipótesis estadística  $H_0: Var(\epsilon|x_j, j = 1, \dots, p) = \sigma^2$ . Dado que  $E(\epsilon|x_j, j = 1, \dots, p) = 0$ , lo mismo es probar la hipótesis nula  $H_0: E(\epsilon^T \epsilon|x_j, j = 1, \dots, p) = E(\epsilon^T \epsilon) = \sigma^2$ , lo cual quiere decir que nuestra hipótesis nula supone la no existencia de heterocedasticidad.

Queremos usar nuestros datos para rechazar la hipótesis nula, de tal manera que si existe algún modelo que pueda representar a la varianza de los errores del modelo  $\sigma^2$  mediante el modelo condicional sobre  $x_1, x_2, \dots, x_p$ :

$$\sigma^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_p x_p$$

La existencia de homocedasticidad implica que  $\alpha_1 = \alpha_2 = \dots = \alpha_p = 0$  y por tanto:

$$E(\epsilon^T \epsilon|x_j, j = 1, \dots, p) = \alpha_0$$

Para estimar los coeficientes de este modelo tomaremos los errores cuadráticos del modelo para cada observación  $X_i$  tal que:

$$\hat{\sigma}_i^2 = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \dots + \alpha_p x_{ip} + \varepsilon_i$$

y probar si los coeficientes son iguales a 0:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_p = 0 \text{ vs } H_1: \alpha_j \neq 0$$

Podemos probar nuestra hipótesis nula usando una prueba  $F$ :

$$F \equiv \frac{(SSR_r - SSR_u)/p}{SSR_u/(n - p - 1)}$$

$$\text{donde } SSR_u = \sum_{i=1}^n (\hat{\sigma}_i^2 - \alpha_0 - \alpha_1 x_{i1} - \alpha_2 x_{i2} - \dots - \alpha_p x_{ip})^2$$

$$SSR_r = \sum_{i=1}^n (\hat{\sigma}_i^2 - \alpha_0)^2$$

Si el estadístico  $F$  es mayor a un cierto  $c$  determinado para un valor estadístico de significancia, rechazamos nuestra hipótesis nula de no heterocedasticidad.

Otras pruebas para detectar heterocedasticidad son las de *Breusch – Pagan*, la de *White* o la de *Goldfeld – Quandt*.

Si hay indicación de que la varianza poblacional  $\sigma^2$  no es constante entonces hay dos remedios posibles:

1. Usar mínimos cuadrados ponderados o generalizados donde los pesos que se usan son hallados en base a los datos tomados.
2. Transformar la variable de respuesta  $Y$  para estabilizar la varianza. Algunas transformaciones útiles se enuncian en la siguiente tabla:

Transformación	Situación
$\sqrt{y}$	$Var(\varepsilon_i) \propto E(y_i)$



$\sqrt{y} + \sqrt{y+1}$	$Var(\epsilon_i) \propto E(y_i)$
$\ln(y)$	$Var(\epsilon_i) \propto [E(y_i)]^2$
$\ln(y+1)$	$Var(\epsilon_i) \propto [E(y_i)]^2$
$1/y$	$Var(\epsilon_i) \propto [E(y_i)]^4$
$1/(y+1)$	$Var(\epsilon_i) \propto [E(y_i)]^4$
$\sin^{-1}(\sqrt{y})$	$Var(\epsilon_i) \propto E(y_i)[1 - E(y_i)]$

Otro de los supuestos que se debe de cumplir es la no existencia de la autocorrelación, i.e., que  $Corr(\epsilon_t, \epsilon_{t-s}) = 0$ .

Si los errores de nuestro modelo estuvieran autocorrelacionados los coeficientes de regresión de la matriz  $\beta$  no serían estimadores eficientes, i.e.:

$$(X^T X)^{-1} X^T \Omega X (X^T X)^{-1} = Var(\hat{\beta}') > Var(\beta') = \sigma^2 (X^T X)^{-1}$$

donde  $\Omega = \{Cov(\epsilon_i, \epsilon_j)\}$ , es la matriz de covarianzas de los errores del modelo y  $\hat{\beta}'$  es el estimador de los coeficientes de regresión del modelo con autocorrelación, y  $\hat{\beta}$  del modelo sin autocorrelación.

La detección puede ser gráficamente mediante la observación de algún patrón de comportamiento contrastando los  $\epsilon_i$ 's contra las observaciones  $y_i$ 's, también graficando  $\epsilon_t$  vs  $\epsilon_{t-s}$ , i.e.,  $s$  períodos de rezago.

Estadísticamente, podemos detectar la autocorrelación mediante los estadísticos de prueba *Durbin-Watson*, *H-Durbin* o *Breusch-Godfrey* que lo que hacen es detectar si existiera alguna correlación significativa de los errores con períodos de rezago.

Para remediar el problema de la autocorrelación en un modelo enunciamos 2 métodos para ello:

1. Método de Cochrane-Orcutt.
2. Mínimos cuadrados generalizados.

Por último sería necesario comprobar la normalidad de los  $\epsilon_i$  mediante pruebas de normalidad o pruebas de bondad de ajuste con la distribución  $Nor(\bar{0}, \sigma^2 I)$ . Este supuesto lo usaremos para generar intervalos de confianza, hipótesis estadísticas sobre los parámetros del modelo como otros usos más.

### c. Componentes Principales como variables endógenas de un Modelo de Regresión Lineal Múltiple

El objetivo más citado del uso de los componentes principales es hacer una reducción dimensional. En este estudio también utilizaremos esta técnica de análisis multivariado para no dejar fuera del modelo a ninguna de las causas de muerte y para evitar la multicolinealidad que pudiera presentar el modelo.

Consideremos el modelo de regresión lineal múltiple:

$$g(X) = \beta_0 + \beta_1 d^{(1)} + \beta_2 d^{(2)} + \dots + \beta_m d^{(m)} + \epsilon$$

donde la matriz  $X$  es igual a un vector de 1's de  $\omega \times 1$  pegado con la matriz de los fallecidos por edad y por causa de muerte de  $\omega \times m$ .

Estandarizando las variables predictoras tenemos:

$$g(X^*) = \beta_0 + \beta_1 z d^{(1)} + \beta_2 z d^{(2)} + \dots + \beta_m z d^{(m)} + \epsilon$$

$$\text{donde } z d^{(j)} = \frac{d^{(j)} - \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{\omega \times 1} \cdot \frac{\sum_{x=0}^{\omega-1} d_x^{(j)}}{\omega}}{S_{d^{(j)}}}$$

$$\text{y } X^* = [\bar{1}, D]$$

La idea de usar los componentes principales como variables predictoras es crear las nuevas variables que sean ortogonales entre si, tal cual se describe en páginas anteriores.

La j-ésima componente principal  $C_j$  se puede escribir como:

$$C_j = \alpha_{j1} z d^{(1)} + \alpha_{j2} z d^{(2)} + \dots + \alpha_{jm} z d^{(m)}$$

Una vez que se obtienen los componentes hay dos opciones para utilizarlas en un modelo de regresión lineal múltiple:

1. La opción clásica: hacer la regresión solamente con las componentes principales que han sido seleccionadas, o
2. La opción moderna: hacer la regresión con todas las componentes principales y luego aplicar técnicas de selección de variables para elegir el mejor modelo.

Al hacer uso de los componentes principales como variables predictivas los  $VIF$ 's son iguales a 1, lo cual indica la no existencia de multicolinealidad. El  $R^2$  de esta modelo y del modelo con las variables originales también es el mismo ya que el coeficiente de determinación es invariante a transformaciones lineales.

### III. Método de corrección de parámetros de la función de Gompertz-Makeham del Dr. Alejandro Mina Valdés

Dado que la función de Gompertz-Makeham ampliada se denota como

$y(x) = Ka^x w^{x^2} b^{d^x}$  (donde  $y(x)$  es el estimado de los vivos a edad  $x$ ) para

toda  $x = 0, 1, 2, \dots, 4m - 1$ , se obtiene el logaritmo natural de la función  $y(x)$ .

Es posible realizar variaciones en ellos con el objetivo de calcular una mejor aproximación a los valores observados.

Dado que  $y(x) = Ka^x w^{x^2} b^{d^x}$  para toda  $x = 0, 1, 2, \dots, 4m - 1$ , se obtiene el logaritmo natural de la función  $y(x)$ :

$$\begin{aligned} \ln y(x) &= \ln(Ka^x w^{x^2} b^{d^x}) \\ &= \ln K + x \ln a + x^2 \ln w + d^x \ln b \end{aligned} \quad (15)$$

Hay que calcular la derivada de la expresión (15):

$$\frac{\partial}{\partial y(x)} \ln y(x) = \frac{\partial}{\partial u} (\ln K + x \ln a + x^2 \ln w + d^x \ln b) \quad (16)$$

Al calcular la derivada se considera que:

$$\frac{\partial}{\partial y(x)} \ln y(x) = \frac{1}{y(x)} dy(x) \quad (17)$$

mientras que la derivada del miembro derecho se puede expresar como:

$$\frac{\partial}{\partial y(x)} \ln y(x) = \frac{\partial}{\partial K} \ln K + \frac{\partial}{\partial a} x \ln a + \frac{\partial}{\partial w} x^2 \ln w + \frac{\partial}{\partial b} d^x \ln b + \frac{\partial}{\partial d} d^x \ln b$$

$$= \frac{1}{K} dK + \frac{x}{a} da + \frac{d^x}{b} db + \ln b \frac{\partial}{\partial d} d^x + \frac{x^2}{w} dw \quad (18)$$

El factor  $\frac{\partial}{\partial d} d^x$  del penúltimo término de la expresión (18) se puede presentar de acuerdo al razonamiento siguiente: de acuerdo a las propiedades de los logaritmos se puede expresar:

$$\ln d^x = x \ln d$$

Al obtener la derivada de la expresión anterior se observa que:

$$\frac{\partial}{\partial d} \ln d^x = \frac{\partial}{\partial d} x \ln d$$

$$\frac{1}{d^x} d(d^x) = \frac{x}{d} d(d)$$

Por lo tanto, la derivada de  $d^x$  con respecto a  $d$  es:

$$d(d^x) = x d^x \frac{\partial d}{d} \quad (19)$$

Dado lo anterior, la expresión (18) se puede escribir como:

$$\frac{1}{y(x)} dy(x) = \frac{1}{k} dk + \frac{x}{a} da + \frac{d^x}{b} db + x d^x \ln b \frac{\partial d}{d} + \frac{x^2}{w} dw \quad (20)$$

en consecuencia, la derivada de  $y(x)$  es:

$$dy(x) = \frac{y(x)}{k} dk + \frac{xy(x)}{a} da + \frac{d^x y(x)}{b} db + x d^x y(x) \ln b \frac{\partial d}{d} + \frac{x^2 y(x)}{w} dw \quad (21)$$

Para calcular los valores de los parámetros a partir de la expresión (21) se procede a linealizar dicha expresión, para ello se denota como:

$$x_1 = dy_x \quad x_2 = y_x \quad x_3 = x(x_2) \quad x_4 = x_2 d^x \quad x_5 = x_3 d^x \quad x_6 = x^2 y_x$$

$$c_2 = \frac{\partial k}{k} \quad c_3 = \frac{\partial a}{a} \quad c_4 = \frac{\partial b}{b} \quad c_5 = \ln b \frac{\partial d}{d} \quad c_6 = \frac{\partial w}{w}$$

Una vez hecho lo anterior, se sustituye en (21) estas variables, por lo que puede expresarse en forma de regresión múltiple lineal como se presenta a continuación:

$$x_1 = c_2 x_2 + c_3 x_3 + c_4 x_4 + c_5 x_5 + c_6 x_6$$

Se denotan las diferencias entre los valores observados y los valores teóricos como  $dy_x$ , de tal forma que sea posible calcular los coeficientes de regresión:

$c_2, c_3, c_4, c_5, c_6$  a través de las siguientes ecuaciones normales:

$$\sum_{i=1}^{4m} x_{1,i} x_{2,i} = c_2 \sum_{i=1}^{4m} x_{2,i} x_{2,i} + c_3 \sum_{i=1}^{4m} x_{2,i} x_{3,i} + c_4 \sum_{i=1}^{4m} x_{2,i} x_{4,i} + c_5 \sum_{i=1}^{4m} x_{2,i} x_{5,i} + c_6 \sum_{i=1}^{4m} x_{2,i} x_{6,i}$$

$$\sum_{i=1}^{4m} x_{1,i} x_{3,i} = c_2 \sum_{i=1}^{4m} x_{2,i} x_{3,i} + c_3 \sum_{i=1}^{4m} x_{3,i} x_{3,i} + c_4 \sum_{i=1}^{4m} x_{3,i} x_{4,i} + c_5 \sum_{i=1}^{4m} x_{3,i} x_{5,i} + c_6 \sum_{i=1}^{4m} x_{3,i} x_{6,i}$$

$$\sum_{i=1}^{4m} x_{1,i} x_{4,i} = c_2 \sum_{i=1}^{4m} x_{2,i} x_{4,i} + c_3 \sum_{i=1}^{4m} x_{3,i} x_{4,i} + c_4 \sum_{i=1}^{4m} x_{4,i} x_{4,i} + c_5 \sum_{i=1}^{4m} x_{4,i} x_{5,i} + c_6 \sum_{i=1}^{4m} x_{4,i} x_{6,i}$$

$$\sum_{i=1}^{4m} x_{1,i} x_{5,i} = c_2 \sum_{i=1}^{4m} x_{2,i} x_{5,i} + c_3 \sum_{i=1}^{4m} x_{3,i} x_{5,i} + c_4 \sum_{i=1}^{4m} x_{4,i} x_{5,i} + c_5 \sum_{i=1}^{4m} x_{5,i} x_{5,i} + c_6 \sum_{i=1}^{4m} x_{5,i} x_{6,i}$$

$$\sum_{i=1}^{4m} x_{1,i} x_{6,i} = c_2 \sum_{i=1}^{4m} x_{2,i} x_{6,i} + c_3 \sum_{i=1}^{4m} x_{3,i} x_{6,i} + c_4 \sum_{i=1}^{4m} x_{4,i} x_{6,i} + c_5 \sum_{i=1}^{4m} x_{5,i} x_{6,i} + c_6 \sum_{i=1}^{4m} x_{6,i} x_{6,i}$$

Para resolver este sistema de ecuaciones lineales simultáneas de cinco incógnitas puede emplearse el método de matrices (Mendenhall, Scheaffer. 1986); donde para encontrar la solución al sistema es necesario conocer la matriz integrada por los coeficientes de la regresión.

Las ecuaciones normales antes presentadas, pueden expresarse en forma matricial de la siguiente manera:

$$\begin{array}{l}
 \sum_{i=1}^{4m} x_{2,i} x_{2,i} \quad \sum_{i=1}^{4m} x_{2,i} x_{3,i} \quad \sum_{i=1}^{4m} x_{2,i} x_{4,i} \quad \sum_{i=1}^{4m} x_{2,i} x_{5,i} \quad \sum_{i=1}^{4m} x_{2,i} x_{6,i} \quad c_2 \quad \sum_{i=1}^{4m} x_{1,i} x_{2,i} \\
 \sum_{i=1}^{4m} x_{2,i} x_{3,i} \quad \sum_{i=1}^{4m} x_{3,i} x_{3,i} \quad \sum_{i=1}^{4m} x_{3,i} x_{4,i} \quad \sum_{i=1}^{4m} x_{3,i} x_{5,i} \quad \sum_{i=1}^{4m} x_{3,i} x_{6,i} \quad c_3 \quad \sum_{i=1}^{4m} x_{1,i} x_{3,i} \\
 A = \quad \sum_{i=1}^{4m} x_{2,i} x_{4,i} \quad \sum_{i=1}^{4m} x_{3,i} x_{4,i} \quad \sum_{i=1}^{4m} x_{4,i} x_{4,i} \quad \sum_{i=1}^{4m} x_{4,i} x_{5,i} \quad \sum_{i=1}^{4m} x_{4,i} x_{6,i} \quad V = \quad c_4 \quad G = \quad \sum_{i=1}^{4m} x_{1,i} x_{4,i} \\
 \sum_{i=1}^{4m} x_{2,i} x_{5,i} \quad \sum_{i=1}^{4m} x_{3,i} x_{5,i} \quad \sum_{i=1}^{4m} x_{4,i} x_{5,i} \quad \sum_{i=1}^{4m} x_{5,i} x_{5,i} \quad \sum_{i=1}^{4m} x_{5,i} x_{6,i} \quad c_5 \quad \sum_{i=1}^{4m} x_{1,i} x_{5,i} \\
 \sum_{i=1}^{4m} x_{2,i} x_{6,i} \quad \sum_{i=1}^{4m} x_{3,i} x_{6,i} \quad \sum_{i=1}^{4m} x_{4,i} x_{6,i} \quad \sum_{i=1}^{4m} x_{5,i} x_{6,i} \quad \sum_{i=1}^{4m} x_{6,i} x_{6,i} \quad c_6 \quad \sum_{i=1}^{4m} x_{1,i} x_{6,i}
 \end{array}$$

Para encontrar los coeficientes de la matriz  $V$ , se realiza la siguiente operación:

$$V = A^{-1} G$$

De esta manera se calculan los valores de las  $c_j$  y por consiguiente las primeras correcciones a los parámetros  $k, a, b, d$  y  $w$  de la función Gompertz-Makeham ampliada. Estas correcciones permiten obtener nuevas aproximaciones para los parámetros. Por lo tanto, los nuevos valores para estos son:

$$k_1 = k(1 + c_2), a_1 = a(1 + c_3), b_1 = b(1 + c_4), d_1 = d \left( 1 + \frac{c_5}{\ln b} \right), w_1 = w(1 + c_6)$$

A partir de estos valores se obtienen nuevos valores teóricos y por lo tanto nuevas diferencias  $dy_x$ . Lo anterior lleva un proceso iterativo que permitirá ir obteniendo aproximaciones cada vez más satisfactorias. Es decir, el proceso deberá repetirse hasta que la magnitud de las correcciones alcancen un valor reducido tal que no logren cambiar sensiblemente los valores teóricos obtenidos usando los valores de los parámetros hasta esa iteración.

En general, se observa que si  $k_i, a_i, b_i, d_i$  y  $w_i$  son valores de la iteración  $i$ , los valores de esos parámetros a la iteración  $i + 1$  serán:

$$k_{i+1} = k_i(1 + c_{2,i+1}), a_{i+1} = a_i(1 + c_{3,i+1}), b_{i+1} = b_i(1 + c_{4,i+1}), d_{i+1} = d_i \left( 1 + \frac{c_{5,i+1}}{\ln b} \right), w_{i+1} = w_i(1 + c_{6,i+1})$$



#### IV. Modelo de la Mortalidad Mexicana en función de las causas de muerte que la generan

##### A. Modelo de la mortalidad mexicana usando las componentes principales como variables

Los coeficientes de regresión utilizando componentes principales no son fáciles de interpretar porque en este caso las variables predictoras son combinaciones lineales de las predictoras originales. Lo que hay que hacer es tratar de expresar la regresión en términos de las variables originales para luego hacer la interpretación.

El modelo de regresión suponiendo que usáramos  $r$  de los  $m'$  componentes obtenidos se denotaría de la siguiente manera:

$$\begin{aligned} g(X^C) = G &= X^C \beta + \epsilon \\ &= \beta_0 + \beta_1 C^1 + \beta_2 C^2 + \dots + \beta_r C^r + \epsilon \\ \text{donde } X^C &= [\bar{1}, C] = [\bar{1}, D \cdot \alpha] \\ \text{entonces } g(X^C) &= \beta_0 + \beta_1 D \alpha_1 + \beta_2 D \alpha_2 + \dots + \beta_r D \alpha_r + \epsilon \end{aligned}$$

La estimación del vector de coeficientes  $\beta$  es la siguiente:

$$\hat{\beta} = (X^{CT} X^C)^{-1} X^{CT} G$$

Para entender los coeficientes de regresión obtenidos, expresaremos el modelo en términos de las variables originales. Para ello, comenzaremos a desarrollar las componentes principales incluidas en nuestro modelo:

$$\begin{aligned}
g(X^C) &= \beta_0 + \beta_1 \sum_{j=1}^m \alpha_{j1} z d^{(j)} + \beta_2 \sum_{j=1}^m \alpha_{j2} z d^{(j)} + \dots + \beta_r \sum_{j=1}^m \alpha_{jr} z d^{(j)} + \epsilon \\
&= \beta_0 + \sum_{k=1}^r \sum_{j=1}^m \beta_k \alpha_{jk} z d^{(j)} = \beta_0 + \sum_{j=1}^m z d^{(j)} \sum_{k=1}^r \beta_k \alpha_{jk}
\end{aligned}$$

Denotaremos a la suma-producto de coeficientes de regresión y de los coeficientes de la  $j$ -ésima fila de la matriz  $\alpha$  de los componentes como  $\sum_{k=1}^r \beta_k \alpha_{jk} = \delta_j$ , por lo que nuestro modelo queda como:

$$g(X^C) = \beta_0 + \sum_{j=1}^m \delta_j z d^{(j)} = \beta_0 + \delta_1 z d^{(1)} + \delta_2 z d^{(2)} + \dots + \delta_m z d^{(m)}$$

Por último, ya que  $z d^{(j)} = \frac{d^{(j)} - \bar{d}^{(j)}}{S_{d^{(j)}}}$ , entonces redefinimos nuestro modelo:

$$\begin{aligned}
g(X^C) &= \beta_0 + \sum_{j=1}^m \delta_j \frac{d^{(j)} - \bar{d}^{(j)}}{S_{d^{(j)}}} = \beta_0 + \sum_{j=1}^m \frac{\delta_j}{S_{d^{(j)}}} d^{(j)} - \sum_{j=1}^m \frac{\delta_j}{S_{d^{(j)}}} \bar{d}^{(j)} \\
&= \beta_0^* + \sum_{j=1}^m \beta_j^* d^{(j)}
\end{aligned}$$

$$\text{donde } \beta_0^* = \beta_0 - \sum_{j=1}^m \beta_j^* \bar{d}^{(j)} \text{ y } \beta_j^* = \frac{\delta_j}{S_{d^{(j)}}}$$

Si tomamos una observación  $X_i^C$  de la matriz original  $X^C$ , entonces obtendríamos el valor esperado para la observación  $i$ :

$$g(X_i^C) = \beta_0^* + \beta_1^* d_i^{(1)} + \beta_2^* d_i^{(2)} + \dots + \beta_m^* d_i^{(m)}$$

Al obtener el modelo de regresión lineal múltiple de la variable dependiente  $g(X_i^C)$  en función de las causas de muerte solo resta expresar al modelo de regresión logística que buscamos ajustar a las  $q_i$ , mediante las estimaciones  $\hat{q}_i$ :

$$\pi(X_i^C) = \frac{1}{1 + e^{-g(X_i^C)}} = \frac{1}{1 + e^{-(\beta_0^* + \beta_1^* d_i^{(1)} + \beta_2^* d_i^{(2)} + \dots + \beta_m^* d_i^{(m)})}} = \hat{q}_i$$

## B. Impacto de las causas de muerte sobre la mortalidad mexicana

Con este modelo es posible entender el impacto en la mortalidad poblacional total de cada causa de muerte. Cada una de ellas se encuentra ponderada en el modelo obtenido y tiene una importancia significativa en la obtención de las probabilidades de fallecimiento por causa de muerte.

Para calcular las probabilidades estimadas  $\hat{q}_i^{(j)}$  para la causa de muerte  $j$  desarrollaremos el modelo obtenido de  $\hat{q}_i$  basándonos en la teoría descrita de Decrementos Múltiples:

$\hat{q}_i = \hat{q}_i^{(\tau)}$ , i.e., la probabilidad de fallecer es igual a la probabilidad de que una persona de edad  $x$  fallezca por cualquier causa de muerte.

Como  $\hat{q}_i^{(\tau)} = 1 - \hat{p}_i^{(\tau)}$ , entonces:

$$\hat{q}_i^{(\tau)} = \frac{1}{1 + e^{-g(X_i^C)}}$$

$$\text{entonces } \hat{p}_i^{(\tau)} = 1 - \frac{1}{1 + e^{-g(X_i^C)}} = \frac{e^{-g(X_i^C)}}{1 + e^{-g(X_i^C)}} = \prod_{j=1}^m p_i^{(j)}$$

Desarrollaremos convenientemente a  $g(X_i^C)$ :

$$g(X_i^C) = \beta_0^* + \beta_1^* d_i^{(1)} + \beta_2^* d_i^{(2)} + \dots + \beta_m^* d_i^{(m)}$$

$$= \beta_0 + \beta_1^* (d_i^{(1)} - \bar{d}^{(1)}) + \beta_2^* (d_i^{(2)} - \bar{d}^{(2)}) + \dots + \beta_m^* (d_i^{(m)} - \bar{d}^{(m)})$$

Entonces:

$$\begin{aligned}\hat{p}_i^{(\tau)} &= \frac{e^{-\{\beta_0 + \beta_1^*(a_i^{(1)} - \bar{a}^{(1)}) + \beta_2^*(a_i^{(2)} - \bar{a}^{(2)}) + \dots + \beta_m^*(a_i^{(m)} - \bar{a}^{(m)})\}}}{1 + e^{-g(x_i^c)}} \\ &= \frac{e^{-\beta_0}}{1 + e^{-g(x_i^c)}} \cdot e^{-\{\beta_1^*(a_i^{(1)} - \bar{a}^{(1)}) + \beta_2^*(a_i^{(2)} - \bar{a}^{(2)}) + \dots + \beta_m^*(a_i^{(m)} - \bar{a}^{(m)})\}} \\ &= \prod_{j=1}^m \left( \frac{e^{-\beta_0}}{1 + e^{-g(x_i^c)}} \right)^{\beta_j^* / \sum_{j=1}^m \beta_j^*} \cdot e^{-\beta_j^*(a_i^{(j)} - \bar{a}^{(j)})} = \prod_{j=1}^m p_i'^{(j)}\end{aligned}$$

$$\text{donde } p_x'^{(j)} = \left( \frac{e^{-\beta_0}}{1 + e^{-g(x_i^c)}} \right)^{\beta_j^* / \sum_{j=1}^m \beta_j^*} \cdot e^{-\beta_j^*(a_i^{(j)} - \bar{a}^{(j)})}$$

Ahora, una vez que definimos  $p_x'^{(j)}$ , nos queda definir a  $\hat{q}_i^{(j)}$  tal que  $\hat{q}_i^{(\tau)} =$

$\sum_{j=1}^m \hat{q}_i^{(j)}$ . Una aproximación que menciona *Bowers* nos servirá para calcular las

$\hat{q}_i^{(j)}$  usando las  $p_x'^{(j)}$ :

$$p_i'^{(j)} = \left( \hat{p}_i^{(\tau)} \right)^{\hat{q}_i^{(j)} / \hat{q}_i^{(\tau)}}$$

Resta despejar a  $\hat{q}_i^{(j)}$  para obtener la relación que necesitamos:

$$\hat{q}_i^{(j)} = \frac{\ln(p_i'^{(j)})}{\ln(\hat{p}_i^{(\tau)})} \cdot \hat{q}_i^{(\tau)}$$

$$\begin{aligned}\text{donde } \sum_{j=1}^m \hat{q}_i^{(j)} &= \sum_{j=1}^m \frac{\ln(p_i'^{(j)})}{\ln(\hat{p}_i^{(\tau)})} \cdot \hat{q}_i^{(\tau)} = \frac{\hat{q}_i^{(\tau)}}{\ln(\hat{p}_i^{(\tau)})} \sum_{j=1}^m \ln(p_i'^{(j)}) \\ &= \frac{\hat{q}_i^{(\tau)}}{\ln(\hat{p}_i^{(\tau)})} \cdot \ln \left( \prod_{j=1}^m p_i'^{(j)} \right) = \frac{\hat{q}_i^{(\tau)}}{\ln(\hat{p}_i^{(\tau)})} \cdot \ln(\hat{p}_i^{(\tau)}) = \hat{q}_i^{(\tau)}\end{aligned}$$

Por tanto,  $\hat{q}_i^{(j)}$  queda expresado en términos del modelo obtenido y de los fallecidos por cada causa de muerte como:

$$\hat{q}_i^{(j)} = \frac{\ln \left[ \left( \frac{e^{-\beta_0}}{1 + e^{-g(x_i^c)}} \right)^{\beta_j^* / \sum_{j=1}^m \beta_j^*} \cdot e^{-\beta_j^* (a_i^{(j)} - \bar{a}^{(j)})} \right]}{\ln \left( \frac{e^{-g(x_i^c)}}{1 + e^{-g(x_i^c)}} \right)} \cdot \frac{1}{1 + e^{-g(x_i^c)}}$$

## V. Método de proyección de la Mortalidad Mexicana a través del modelo lineal de parámetros diferenciales

El método de corrección de parámetros que el Dr. Alejandro Mina propone en su artículo *Funciones de Supervivencia empleadas en el análisis demográfico* (2001) lo utilizaremos con el fin de hacer correcciones en los coeficientes  $\beta_j^*$ . Entiéndase estas correcciones como el ajuste que requieren los coeficientes  $\beta_j^*$  para poder estimar la mortalidad por causa de muerte para los siguientes años.

Dado que:

$$\hat{q}_i^{(j)} = \frac{\ln \left[ \left( \frac{e^{-\beta_0}}{1 + e^{-g(x_i^c)}} \right)^{\beta_j^* / \sum_{j=1}^m \beta_j^*} \cdot e^{-\beta_j^* (d_i^{(j)} - \bar{d}^{(j)})} \right]}{\ln \left( \frac{e^{-g(x_i^c)}}{1 + e^{-g(x_i^c)}} \right)} \cdot \frac{1}{1 + e^{-g(x_i^c)}}$$

y resolviendo los logaritmos de esta última expresión la podemos simplificar por:

$$\hat{q}_i^{(j)} = \left\{ \left( \frac{\beta_j^*}{\sum_{j=1}^m \beta_j^*} \right) [-\beta_0 + \ln(\hat{q}_i^{(\tau)})] - \beta_j^* (d_i^{(j)} - \bar{d}^{(j)}) \right\} \cdot \frac{\hat{q}_i^{(\tau)}}{\ln(1 - \hat{q}_i^{(\tau)})}$$

Obtengamos ahora la parcial de  $\partial \hat{q}_i^{(j)} / \partial \beta_j^*$ :

Recordando que  $D(fh) = Df \cdot h + f \cdot Dh$ ,

$$\text{si } f(\beta_j^*) = \left( \frac{\beta_j^*}{\sum_{j=1}^m \beta_j^*} \right) [-\beta_0 + \ln(\hat{q}_i^{(\tau)})] - \beta_j^* (d_i^{(j)} - \bar{d}^{(j)}) \text{ y}$$

$$h(\beta_j^*) = \frac{\hat{q}_i^{(\tau)}}{\ln(1 - \hat{q}_i^{(\tau)})}$$

entonces vamos a obtener tanto de  $f(\beta_j^*)$  como de  $h(\beta_j^*)$  las parciales con respecto a los coeficientes  $\beta_j^*$  con  $j = 1, 2, \dots, m$ .

Obtenemos de  $f$  lo siguiente:

$$\frac{\partial f(\beta_j^*)}{\partial \beta_j^*} = \left( \frac{1}{\sum_{j=1}^m \beta_j^*} - \frac{\beta_j^*}{(\sum_{j=1}^m \beta_j^*)^2} \right) \left[ \left( \frac{1}{\hat{q}_i^{(\tau)}} \right)^2 \cdot \frac{\partial g(X_i^C)}{\partial \beta_j^*} \cdot e^{-g(X_i^C)} \right]$$

donde  $e^{-g(X_i^C)} = \frac{\hat{p}_i^{(\tau)}}{\hat{q}_i^{(\tau)}}$ , y  $\frac{\partial g(X_i^C)}{\partial \beta_j^*} = \frac{\partial g(X_i^C)}{\partial a_i^{(j)}} \cdot \frac{\partial a_i^{(j)}}{\partial \beta_j^*} = \beta_j^* \cdot \frac{1}{d\beta_j^*}$  por lo cual:

$$\frac{\partial f(\beta_j^*)}{\partial \beta_j^*} = \left( \frac{1}{\sum_{j=1}^m \beta_j^*} - \frac{\beta_j^*}{(\sum_{j=1}^m \beta_j^*)^2} \right) \left[ \frac{\beta_j^*}{d\beta_j^*} \cdot \frac{\hat{p}_i^{(\tau)}}{\hat{q}_i^{(\tau)^3} \right]$$

Ahora, obtenemos de  $h$  lo siguiente:

$$\frac{\partial h(\beta_j^*)}{\partial \beta_j^*} = \frac{\partial \hat{q}_i^{(j)}}{\partial \beta_j^*} \cdot \left[ \frac{\frac{\hat{q}_i^{(\tau)}}{\hat{p}_i^{(\tau)}} + 1}{\ln(\hat{p}_i^{(j)})} \right] = \frac{\partial \hat{q}_i^{(j)}}{\partial \beta_j^*} \cdot \left[ \frac{e^{g(X_i^C)} + 1}{\ln(\hat{p}_i^{(j)})} \right] = \frac{\partial \hat{q}_i^{(j)}}{\partial \beta_j^*} \cdot \left( \frac{1}{\ln(\hat{p}_i^{(j)}) \hat{p}_i^{(\tau)}} \right)$$

$$\text{donde } \hat{q}_i^{(\tau)} / \hat{p}_i^{(\tau)} = e^{g(X_i^C)} \text{ y } (e^{g(X_i^C)} + 1) = \frac{1}{\hat{p}_i^{(\tau)}}$$

Por lo tanto:

$$\begin{aligned} \frac{\partial \hat{q}_i^{(j)}}{\partial \beta_j^*} = & \left[ \left( \frac{\beta_j^*}{\sum_{j=1}^m \beta_j^*} \right) [-\beta_0 + \ln(\hat{q}_i^{(\tau)})] - \beta_j^* (d_i^{(j)} - \bar{d}^{(j)}) \right] \cdot \left( \frac{1}{\ln(\hat{p}_i^{(j)}) \hat{p}_i^{(\tau)}} \right) \frac{\partial \hat{q}_i^{(j)}}{\partial \beta_j^*} \\ & + \frac{\hat{q}_i^{(\tau)}}{\ln(1 - \hat{q}_i^{(\tau)})} \cdot \left( \frac{1}{\sum_{j=1}^m \beta_j^*} - \frac{\beta_j^*}{(\sum_{j=1}^m \beta_j^*)^2} \right) \left[ \frac{\beta_j^*}{d\beta_j^*} \cdot \frac{\hat{p}_i^{(\tau)}}{\hat{q}_i^{(\tau)^3} \right] \end{aligned}$$

Entonces:

$$\frac{\partial \hat{q}_i^{(j)}}{\partial \beta_j^*} = \frac{\frac{\hat{q}_i^{(\tau)}}{\ln(1 - \hat{q}_i^{(\tau)})} \cdot \left( \frac{1}{\sum_{j=1}^m \beta_j^*} - \frac{\beta_j^*}{(\sum_{j=1}^m \beta_j^*)^2} \right) \left[ \frac{\beta_j^*}{d\beta_j^*} \cdot \frac{\hat{p}_i^{(\tau)}}{\hat{q}_i^{(\tau)^3} \right]}{1 - \left[ \left( \frac{\beta_j^*}{\sum_{j=1}^m \beta_j^*} \right) [-\beta_0 + \ln(\hat{q}_i^{(\tau)})] - \beta_j^*(d_i^{(j)} - \bar{d}^{(j)}) \right] \cdot \left( \frac{1}{\ln(\hat{p}_i^{(j)}) \hat{p}_i^{(\tau)}} \right)}$$

$$= \frac{\left( \frac{1}{\sum_{j=1}^m \beta_j^*} - \frac{\beta_j^*}{(\sum_{j=1}^m \beta_j^*)^2} \right) \left( \frac{\hat{p}_i^{(\tau)}}{\hat{q}_i^{(\tau)^3} \right)^2 \frac{\beta_j^*}{d\beta_j^*}}{\ln(\hat{p}_i^{(j)}) \hat{p}_i^{(\tau)} - \left[ \left( \frac{\beta_j^*}{\sum_{j=1}^m \beta_j^*} \right) [-\beta_0 + \ln(\hat{q}_i^{(\tau)})] - \beta_j^*(d_i^{(j)} - \bar{d}^{(j)}) \right]}$$

Sea  $a_{ij} = \frac{\left( \frac{1}{\sum_{j=1}^m \beta_j^*} - \frac{\beta_j^*}{(\sum_{j=1}^m \beta_j^*)^2} \right) \left( \frac{\hat{p}_i^{(\tau)}}{\hat{q}_i^{(\tau)^3} \right)^2 \frac{\beta_j^*}{d\beta_j^*}}{\ln(\hat{p}_i^{(j)}) \hat{p}_i^{(\tau)} - \left[ \left( \frac{\beta_j^*}{\sum_{j=1}^m \beta_j^*} \right) [-\beta_0 + \ln(\hat{q}_i^{(\tau)})] - \beta_j^*(d_i^{(j)} - \bar{d}^{(j)}) \right]}$ , entonces:

$$\frac{\partial \hat{q}_i^{(j)}}{\partial \beta_j^*} = a_{ij} \cdot \frac{\beta_j^*}{d\beta_j^*} = a_{ij} \cdot \frac{1}{d\beta_j^* / \beta_j^*}$$

A  $d\beta_j^* / \beta_j^*$  lo entenderemos como la tasa de cambio que tendrá el coeficiente  $\beta_j^*$  para formar parte del modelo que estimará la mortalidad por causa de muerte del siguiente año. En párrafos siguientes describiremos el proceso de iteraciones que seguiremos para esto.

Ya que obtuvimos el resultado anterior, lo usaremos para calcular a  $\frac{\partial \hat{q}_i^{(\tau)}}{\partial u}$  donde

$$u = (\beta_1^*, \beta_2^*, \dots, \beta_m^*):$$

$$\frac{\partial \hat{q}_i^{(\tau)}}{\partial u} = \sum_{j=1}^m \frac{\partial \hat{q}_i^{(\tau)}}{\partial \beta_j^*} = \sum_{j=1}^m \frac{\partial \hat{q}_i^{(j)}}{\partial \beta_j^*} = \sum_{j=1}^m a_{ij} \cdot \frac{\beta_j^*}{d\beta_j^*}$$



A  $\frac{\partial \hat{q}_i^{(\tau)}}{\partial u}$  lo tomaremos como las diferencias entre las  $\hat{q}_i^{(\tau)}$  del año en el que se calculo y las del año anterior, i.e., si definimos a  $\hat{q}_{i,n}^{(\tau)}$  como la probabilidad que tiene una persona de edad  $i$  de fallecer antes de cumplir más de  $i + 1$  por cualquier causa en el año  $n$ , entonces:

$$\frac{\partial \hat{q}_i^{(\tau)}}{\partial u} = \hat{q}_{i,n}^{(\tau)} - \hat{q}_{i,n-1}^{(\tau)}$$

Ahora, formaremos un modelo lineal con los parámetros diferenciales  $\beta_j^*/d\beta_j^*$  como coeficientes de un modelo de regresión lineal de la siguiente manera:

$$a_i = \frac{\partial \hat{q}_i^{(\tau)}}{\partial u} \text{ y } \theta_{j,1} = \frac{\beta_j^*}{d\beta_j^*}, \text{ donde el subíndice } (j,1) \text{ indica el } j\text{-ésimo parámetro}$$

diferencial de la 1<sup>ra</sup> iteración.

Entonces:

$$a_i = a_{i1}\theta_{1,1} + a_{i2}\theta_{2,1} + \dots + a_{im}\theta_{m,1} + \epsilon_i$$

Definimos el modelo de forma matricial:

$$\text{Sea } a = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{\omega-1} \end{bmatrix}, A = \begin{bmatrix} a_{01} & a_{02} & \dots & a_{0m} \\ a_{11} & a_{12} & \dots & a_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{\omega-1,1} & a_{\omega-1,2} & \dots & a_{\omega-1,m} \end{bmatrix}, \theta_1 = \begin{bmatrix} \theta_{1,1} \\ \theta_{2,1} \\ \vdots \\ \theta_{m,1} \end{bmatrix} \text{ y } \epsilon = \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_{\omega-1} \end{bmatrix}$$

entonces  $a = A\theta + \epsilon$ , donde  $\theta = (A^T A)^{-1} A^T a$

Una vez obtenidos los coeficientes  $\theta_{j,1}$ , recordemos que  $\theta_{j,1} = \beta_j^*/d\beta_j^*$ . Por tanto,  $1/\theta_{j,1} = d\beta_j^*/\beta_j^*$ . Definiremos a  $\beta_{j,n}^*$  como los coeficientes de nuestro modelo correspondientes al año  $n$ . Por consiguiente:

$$\beta_{j,2}^* = \beta_{j,1}^* \cdot (1 + 1/\theta_{j,1}), \beta_{j,3}^* = \beta_{j,2}^* \cdot (1 + 1/\theta_{j,2}), \dots, \beta_{j,n+1}^* = \beta_{j,n}^* \cdot (1 + 1/\theta_{j,n})$$

## Conclusiones

Este trabajo contribuye de manera significativa a las líneas de investigación que existen en la Demografía con respecto al conocimiento del comportamiento futuro de cada causa de muerte. La metodología propuesta ayudará a conocer el impacto que las causas de muerte tengan sobre la población bajo estudio en siguientes años teniendo la ventaja que en nuestro modelo podemos estimar el comportamiento futuro de todas las causas de muerte, además de requerir poca información histórica.

Además, la propuesta es muy elástica, es decir, la metodología no es estricta en la modificación al uso exclusivo del modelo de parámetros diferenciales en el modelo de análisis de regresión. Ni tampoco en cuanto al uso del modelo de regresión logística. Contribuye a una visión más del uso de la teoría actuarial en la que existen algunas investigaciones enunciando ésta como punto de partida para el análisis demográfico del fenómeno de la mortalidad.

En cuanto a probar esta metodología, se requerirá de más tiempo para probarla en un ejercicio práctico así como los ajustes que requiera el modelo. Técnicamente, el modelo respeta las propiedades de la teoría matemática y estadística utilizada pero que, sin embargo, dejamos pendiente el estudio de los resultados que obtengamos para compararlos con el comportamiento de la mortalidad observada hasta ahora y así poder detectar inconsistencias en la metodología.

En lo personal, dejo abierta la posibilidad de profundizar más en el tema de proyecciones de mortalidad, tema que adquiere cada vez mayor importancia por la planeación estratégica de los sistemas de salud, áreas financieras y demás sectores que se pueden ver impactados por los cambios en la mortalidad por causa de muerte.

## Bibliografía

- A. Fernández, Edgar. *Análisis de Regresión*. Puerto Rico (2003)
- Booth, Heather y Tickle, Leonie. Artículo: *Mortality modelling and forecasting: A review of methods*. Australia (2008)
- C. Canavos, George. *Applied probability and statistical methods*. Estados Unidos (1984)
- C. Elizalde, Rosario y F. Ham, Patricia. Artículo: *La mortalidad por causas. Las tendencias demandan modificaciones al sistema de salud*. México (1993)
- Christensen, Kaare. Artículo: *Human biodemography: Some challenges and possibilities for aging research*. Alemania (2008)
- de Jong, Piet y Marshall, Claymore. Artículo: *Mortality projection based on the Wang transform*. Australia (2007)
- F., Tussel. *Análisis Multivariante*. España (2004)
- H. Greene, William. *Econometric Analysis*. Estados Unidos (2002)
- INEGI. Artículo: *Clasificación de Instituciones de Salud*. México (2005)
- Larousse *Diccionario Pocket Español-Inglés*. México (1995)
- M. Apostol, Tom. *Calculus Vol. II*. Estados Unidos (1969)
- Mina V., Alejandro. Artículo: *Funciones de sobrevivencia empleadas en el análisis demográfico*. México (2001)
- Mood, A. McFarlane. *Introduction to the theory of statistics*. Estados Unidos (1974)
- Newton L., Bowers. *Actuarial Mathematics*. Estados Unidos (1986)
- O. Forfar, David. Artículo: *Mortality Laws*. Estados Unidos (2002)
- Peña, Daniel. *Análisis de datos multivariantes*. España (2002)
- R. Draper, Norman y Smith, Harry. *Applied regression analysis Third edition*. Estados Unidos (1998)
- Secretaría de Salud. Artículo: *El Sistema de Salud Mexicano*. México (2009)
- Secretaría de Salud. *Lista mexicana para la selección de las principales causas*. México (2005)
- W. Hosner, David y Lemeshow, Stanley. *Applied logistic regression Second edition*. Estados Unidos (2000)
- W. Jordan, Chester Jr. *Life Contingencies*. Estados Unidos (1967)
- World Health Organization. *International Statistical Classification of Diseases and Related Health Problems Vol. 1, 2 & 3*. Suiza (2004)
- [www.ineqi.gob.mx](http://www.ineqi.gob.mx)