



UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

FACULTAD DE CIENCIAS

EFFECTO DE LA CONTAMINACION AMBIENTAL SOBRE LAS
MUERTES PROVOCADAS POR ISQUEMIA EN PERSONAS
MAYORES DE 65 AÑOS EN LA CIUDAD DE MEXICO EN
1994.

UNA APLICACION DE ANALISIS DE REGRESION POISSON.

T E S I S
QUE PARA OBTENER EL TITULO DE
A C T U A R I O
P R E S E N T A
MARCO ANTONIO VEGA GUERRERO

DIRECTOR DE TESIS: M. EN C. JOSE ANTONIO FLORES DIAZ



MEXICO, D.F.

2004



Universidad Nacional
Autónoma de México



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.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

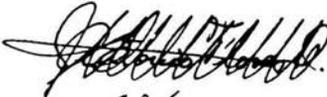
ACT. MAURICIO AGUILAR GONZÁLEZ
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:
 Efecto de la contaminación ambiental sobre las muertes provocadas por
 isquemia en personas mayores de 65 años en la Ciudad de México en 1994.
 Una aplicación de análisis de regresión Poisson
 realizado por Vega Guerrero Marco Antonio

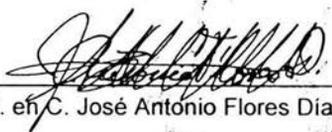
con número de cuenta 8513607-8 , quien cubrió los créditos de la carrera de: Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

A t e n t a m e n t e

- | | | |
|----------------------------------|---|--|
| Director de Tesis
Propietario | M. en C. José Antonio Flores Díaz |  |
| Propietario | Dr. Luis Antonio Rincón Solís |  |
| Propietario | M en A. P. María del Pilar Alonso Reyes |  |
| Suplente | M. en C. Aarón Emigdio Marentes Baéz |  |
| Suplente | Act. Jaime Vázquez Alamilla |  |

Consejo Departamental de Matemáticas


 M. en C. José Antonio Flores Díaz

SECRETARÍA DE EDUCACIÓN PÚBLICA
 INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS
 DIRECCIÓN GENERAL DE ESTADÍSTICA Y CENSOS
 DIRECCIÓN DE ESTADÍSTICA DE LA EDUCACIÓN
 DIRECCIÓN DE ESTADÍSTICA DE LA SALUD
 DIRECCIÓN DE ESTADÍSTICA DE LA ECONOMÍA
 DIRECCIÓN DE ESTADÍSTICA DE LA CULTURA Y DE LOS RECURSOS CULTURALES
 DIRECCIÓN DE ESTADÍSTICA DE LA INDUSTRIA Y DE LOS SERVICIOS
 DIRECCIÓN DE ESTADÍSTICA DE LA AGRICULTURA, GANADERÍA Y PESQUERÍA
 DIRECCIÓN DE ESTADÍSTICA DE LA ENERGÍA Y DE LOS RECURSOS NATURALES
 DIRECCIÓN DE ESTADÍSTICA DE LA DEMOGRAFÍA Y DE LOS RECURSOS HUMANOS
 DIRECCIÓN DE ESTADÍSTICA DE LA ECONOMÍA Y DE LOS RECURSOS NATURALES
 DIRECCIÓN DE ESTADÍSTICA DE LA ECONOMÍA Y DE LOS RECURSOS NATURALES

Efecto de la contaminación ambiental sobre las muertes
provocadas por isquemia en personas mayores de 65 años
en la Ciudad de México en 1994.

Una aplicación de análisis de regresión Poisson.

Marco Antonio Vega Guerrero

Abril 2004

Agradecimientos

A mi familia:

A mi padre **José Ascensión**, por educarme, darme su afecto, paciencia, fortaleza y ser mi ejemplo como persona, amigo, cabeza de familia y por su intenso amor a mi madre.

A mi madre **Catalina**, que me dedico su tiempo, amor, confianza y que junto con papá lograron mantener una familia unida.

A mis hermanos: **Miguel Angel**, el mayor, agradezco el sentimiento de protección y confianza que siempre tuve. **Juan Carlos**, a quien admiro por su determinación y valentía al enfrentar la vida. **Francisco Javier**, quien siempre ha sido un dechado de inteligencia y calma sin titubeos. **Gabriela**, que ha sido y será la causa de todos nosotros. **Raúl**, una persona con una infinidad de valores y cualidades. **Luis Christian**, el menor de todos y a quien he tenido la oportunidad de ver crecer.

A mis maestros:

José Antonio Flores Díaz y **María del Pilar Alonso Reyes**, amigos incondicionales a los cuales respeto y admiro, por sus atinados consejos en los momentos más difíciles de decisión y por apoyarme en este trabajo.

Al **Dr. Luis Antonio Rincón Solís** y al **Act. Jaime Vázquez Alamilla**, por su sencillez, confianza y ayudarme a concretar este proyecto.

Javier Fernández García, mi amigo, maestro y la persona que causo una infinita influencia de construir un criterio propio para entender las cosas.

A mis amigos

Aide Gutiérrez Zermeño, **Raquel Buenrostro Sánchez**, **Víctor Javier Arrieta Rosas**, **Víctor Manuel Barbosa Martínez** que el destino me permitió conocerlos y aprender de ellos y a **Diana Díaz Ledezma** en cualquier lugar donde se encuentre, la recordaré por siempre.

Angélica Loza Orozco, lo mejor que me ha ocurrido en la vida.

Al señor **Héctor Loza Avilés**, una de las personas más sabias que conozco, que admiro y respeto y de quien he aprendido a ver la vida con un gran sentido común.

Adriana Loza Orozco que estimo y quiero mucho.

Ah! por supuesto a mis mejores amigos que fueron, son y serán para toda la vida: **Fernando Del Río Ortíz, Rosendo Porfirio Osorio Pérez (†), Héctor Reyes Vela, Germán Hernández García, Aide Camargo Victoriano, Aarón E. Marentes Báez.**

A todas las personas que no cito y que han compartido algún momento de mi vida.

A **Dios**, que aunque pocas veces le he solicitado ayuda, me ha permitido llegar hasta aquí. Gracias.

Marco Antonio Vega Guerrero
México, D.F.
Abril 2004

Índice General

Presentación	1
Introducción	3
1 Generalidades	5
1.1 Contaminación	5
1.1.1 Marco teórico	5
1.1.2 Marco epidemiológico	7
1.1.3 Antecedentes	8
1.2 Área de estudio	10
1.3 Monitoreo de los contaminantes atmosféricos	11
1.4 Índices de la calidad del aire	12
1.5 Planteamiento del problema	14
2 Modelos Lineales Generalizados	15
2.1 Introducción	15
2.2 Aplicaciones	16
2.3 La familia de distribuciones exponencial	17
2.4 Componentes de los modelos lineales generalizados	18
2.5 Inferencia y análisis de residuales	23
2.5.1 Uso de los residuales	26
2.6 Regresión Poisson	29
2.6.1 Estimadores de máxima verosimilitud para la regresión Poisson	30
2.6.2 Devianza para el modelo Poisson	30
2.6.3 Sobredispersión	31
2.7 Medidas Pseudo R^2	32
2.8 Interpretación de los coeficientes	33
2.9 Anexo: Identificación de parámetros de la familia exponencial	36
2.9.1 Distribución normal	36
2.9.2 Distribución Poisson	37
2.9.3 Distribución binomial	37

ÍNDICE GENERAL

3	Análisis Estadístico	39
3.1	Análisis exploratorio de datos	39
3.2	Construcción de modelos	48
3.2.1	Modelos con efectos principales	49
3.2.2	Modelos con interacciones	53
3.3	Resultados	54
3.3.1	Modelo 6: variables indicadoras como efectos principales y con transformaciones de retraso	56
3.3.2	Modelo 7: variables indicadoras como efectos principales, con transformaciones de retraso y la temperatura como medida de elasticidad	57
3.3.3	Modelo 9: variables indicadoras como efectos principales, con transformaciones de retraso, interacciones y la temperatura como medida de elasticidad	60
	Conclusiones	79
	Modelos	79
	Limitaciones y sugerencias	81
A	Mínimos cuadrados iterativos ponderados	83
	Índice de Tablas	85
	Índice de Figuras	87
	Bibliografía	89

Presentación

En años recientes, muchos estudios reportan la relación entre la exposición a las variaciones diarias de contaminantes del aire, afecciones respiratorias y causas de muerte; principalmente en grupos de edad considerados vulnerables, como la población infantil, las personas de la tercera edad y personas con padecimientos respiratorios y/o cardiovasculares. En la mayoría de dichos estudios se encuentran relaciones relativamente altas en la respuesta y efectos positivos en casi todos los contaminantes junto con los factores meteorológicos; aunque en la revisión de estos artículos no se encontró alguno que ajustara la muerte por isquemia, proporcionaron algunas ideas de cómo abordar este tema.

Esta tesis se compone de tres capítulos, el primero provee una descripción de las fuentes y consecuencias que provocan los contaminantes, las normas de calidad del aire y la situación geográfica de la Ciudad de México como un caso especial de acumulación de contaminantes; además se define un estado isquémico en órganos relacionados al problema. En el segundo capítulo se introducen los modelos lineales generalizados y el caso del análisis de regresión Poisson, como caso particular de esta familia de modelos; es importante señalar que la sección 2.8 es fundamental ya que proporciona una guía para el entendimiento de los coeficientes de regresión, útil para la comprensión de los resultados. En el capítulo 3 se presentan los resultados del análisis de los datos, mostrando tres enfoques en el ajuste, que dependieron en esencia de la transformación de las variables independientes y la interpretación de las mismas.

Introducción

La Ciudad de México -junto con el área metropolitana- es la urbe más grande e importante del país, sede corporativa y eje financiero, concentra gran parte del mercado interno y la industria, como consecuencia convergen un gran número de vías de comunicación. Su crecimiento demográfico va en aumento y como muchas de las grandes ciudades tiene que pagar un precio muy alto por el desarrollo que se refleja en problemas de sobrepoblación, seguridad pública, vivienda, generación de empleo, servicios urbanos básicos, vialidad, transporte y contaminación; esta última es un problema de interés, no sólo por los factores que la originan sino por sus repercusiones en la salud.

La contaminación se genera a partir de las características demográficas, las actividades productivas y el tránsito vehicular y junto con los factores geográficos, climatológicos y meteorológicos es determinante para trastornos en la salud humana, principalmente los cardiovasculares y respiratorios.

En la Ciudad de México, la calidad del aire se evalúa con base en las mediciones realizadas sobre cada uno de los principales contaminantes como lo son: bióxido de azufre, monóxido de carbono, óxidos de nitrógeno, ozono, plomo, partículas suspendidas totales y algunos hidrocarburos, estableciendo el denominado Índice Metropolitano de la Calidad del Aire (IMECA) [Ciencias(1991a)], que determina las normas mínimas y rangos convenientes de condensación de contaminantes, proporcionando una idea del grado de peligrosidad que ocasionan a la salud [Riveros(1990)].

Las normas de calidad del aire fijan valores máximos permisibles de concentración de contaminantes, con el propósito de proteger la salud de la población en general, principalmente a los grupos de mayor susceptibilidad, para esto se incluye un margen adecuado de seguridad. Como en la mayoría de los países, en México no se cuenta con la experiencia de estudios epidemiológicos, toxicológicos y de exposición a contaminantes en animales y seres humanos, por lo que se establecieron las normas tomando en cuenta los criterios y estándares adoptados en otros países del mundo [DDF(1996)].

Actualmente la Secretaría de Salud realiza estudios epidemiológicos que valoran la relación dosis/respuesta entre los diferentes contaminantes y la salud de la población del Valle de México para la futura actualización de los criterios que establecen las normas que regulan la calidad del aire. Los síntomas que presenta la

Introducción

salud de la población guardan una clara correlación positiva con el aumento en el nivel del índice IMECA. Los síntomas comúnmente observados son: disnea (dificultad para respirar), cefalea, conjuntivitis, irritación de las mucosas respiratorias y tos productiva [DDF(1996)].

En la actualidad los grupos de la población de mayor vulnerabilidad ante este problema son los niños, ancianos y enfermos. Para las personas en edad avanzada, los problemas de salud son más frecuentes; por lo general, se presentan estados isquémicos en los órganos relacionados con los padecimientos de tipo cardiovascular, pulmonar y/o respiratorio, agudizándose el problema por las variaciones de los niveles diarios de contaminación, incrementando la posibilidad de muerte en este grupo de la población.

El objetivo principal de este trabajo es presentar una técnica que establece la relación de las mediciones diarias de un conjunto de contaminantes con la causa de muerte generada por isquemia. Considerando las características propias de la variable respuesta, cuya naturaleza es de conteo, se sugiere utilizar un modelo lineal generalizado cuya distribución es Poisson.

Se presentarán los resultados sobre esta técnica en la asociación entre el número diario de muertes causadas por algunos padecimientos respiratorios y cardiovasculares ocasionados por estados isquémicos en los órganos relacionados y sus vínculos con los niveles diarios de la contaminación del aire en población de 65 años o más de la Ciudad de México durante el año de 1994. De acuerdo con esta información, desde el punto de vista estadístico, los niveles de contaminación significan riesgos importantes para la salud dentro de este segmento de la población y además propone una vía para evaluar estas relaciones no sólo para este grupo de edad, sino para otros que sean susceptibles al problema de la contaminación del aire.

Con la realización de esta investigación, se propone un método que ante los problemas de salud de la población, provocados por la contaminación del aire proporcione una alternativa que produzca información valiosa y oportuna, que apoye a las instituciones encargadas de regular la calidad del aire de esta ciudad para el diseño de mejores estrategias.

Capítulo 1

Generalidades

1.1 Contaminación

1.1.1 Marco teórico

La contaminación del aire es uno de los problemas ambientales más importantes, las principales causas que la originan son: las actividades industriales, comerciales, domésticas y agropecuarias; la combustión que se emplea para obtener calor, generar energía eléctrica o movimiento, da lugar al proceso de emisión de los contaminantes más significativos, además, la fundición y la producción de sustancias químicas pueden provocar el deterioro de la calidad del aire si se realizan sin controles efectivos [Reyes(1990)].

El aire puro es una mezcla gaseosa compuesta por 78% de nitrógeno, 21% de oxígeno y 1% de diferentes compuestos tales como el argón, el bióxido de carbono y el ozono. Se entiende por **contaminación atmosférica** cualquier cambio en el equilibrio de estos componentes que altere las propiedades físicas y químicas del aire.

La contaminación es una combinación de componentes químicos, que se generan en su conjunto por una gran variedad de fuentes naturales y artificiales. Los contaminantes del aire se clasifican en: **primarios**, los que son emitidos directamente por la fuente y que permanecen en la atmósfera, para fines de evaluación de la calidad del aire se consideran los óxidos de azufre y nitrógeno, monóxido de carbono, hidrocarburos y partículas suspendidas en general; y en **secundarios**, los que han estado sujetos a cambios químicos o son el producto de la reacción de dos o más contaminantes primarios en la atmósfera, destacan los oxidantes fotoquímicos y algunos radicales de corta existencia como el ozono [Suárez(1991)]. A su vez se pueden clasificar en gases y partículas, además en orgánicos e inorgánicos [Suárez(1991)].

Para abordar adecuadamente el problema, es fundamental conocer las principales características de los contaminantes de mayor importancia, es decir, cómo se producen, cómo se miden en la atmósfera y los efectos que provocan en la salud. A

continuación se presenta un breve resumen de los contaminantes del aire:

El **ozono** (O_3) se forma a partir de la reacción fotoquímica de la radiación solar con los compuestos orgánicos volátiles y óxidos de nitrógeno emitidos principalmente por los automóviles. Es un gas fuertemente oxidante, y como tal, puede reaccionar de forma inmediata con cualquier molécula biológica; debido a su alta reactividad, sólo entrará en contacto con tejidos superficiales del organismo, puede penetrar en las vías respiratorias e irritar severamente las mucosas y los tejidos pulmonares. Según sea la concentración y la duración a su exposición, los efectos en la salud serán más o menos graves; principalmente se observa irritación de los ojos, resequedad en general, cefaleas y dolores pectorales al respirar profundamente, entre otros. A concentraciones elevadas puede provocar inflamación pulmonar, efectos isquémicos en órganos blandos distantes del pulmón, como por ejemplo el hígado, hiperreactividad bronquial y disminuciones importantes de la función pulmonar [DDF(1996)].

El **monóxido de carbono** (CO) es uno de los contaminantes principales de las grandes ciudades y lo produce básicamente la combustión incompleta de la gasolina y el diesel, los vehículos aportan el 98% y el resto se debe a la emisión de procesos industriales. La mayor concentración se presenta en la cercanía de los ejes de circulación con altos flujos vehiculares [DDF(1996)]. A lo largo del día, sus concentraciones en la zona metropolitana del Valle de México varían en forma directamente proporcional a las fluctuaciones de tránsito; en la semana sus niveles presentan una tendencia ligeramente ascendente de lunes a viernes, pero decrecen significativamente los fines de semana, básicamente el domingo. Es un gas incoloro e inodoro que se combina con la hemoglobina para formar la carboxihemoglobina [Reyes(1990)] que afecta principalmente al sistema nervioso central provocando cambios funcionales cardíacos y pulmonares, dolor de cabeza, fatiga, somnolencia, fallos respiratorios y en casos graves la muerte.

Las **partículas suspendidas totales** (PST) son pequeños fragmentos de polvo, agua, carbón, cenizas, hollín, polen, silicatos y óxidos metálicos e incluyen todas las partículas que se pueden mantener en la atmósfera o en una corriente de gases de combustión. El tiempo que pueden estar en suspensión varía de unos cuantos segundos hasta meses y está en función de su tamaño, donde las más pequeñas permanecen más tiempo suspendidas hasta que la lluvia las precipita o el viento las dispersa. Sus fuentes de emisión son muy variadas y pueden ser naturales (erupciones volcánicas, incendios, tolvaneras y otros procesos de erosión de áreas descubiertas) o generadas por el hombre (procesos de combustión de carbón, de refinamiento y quema de petróleo y sus derivados, fundición de metales y humo de tabaco) [Fuentes(1991)]. Pueden tener un tamaño que va de 2.5 a 10 micras (μm) de diámetro. Las partículas menores de 10 μg constituyen la llamada fracción respirable, que por su tamaño pueden llegar hasta la parte alveolar del aparato respiratorio, se estima que más de la mitad de las partículas comprendidas entre 0.01 y 0.1 μm penetran los pulmones. La exposición a las partículas suspendidas

puede causar reducción en las funciones pulmonares, que contribuye a aumentar la frecuencia de las enfermedades respiratorias. En concentraciones muy elevadas, ciertas partículas (como el asbesto) pueden provocar cáncer pulmonar y muerte prematura. Las partículas en conjunción con el bióxido de azufre provocan respiración agitada, disminución del volumen de las vías respiratorias, dificultad para respirar e irritación en las vías respiratorias, de leve a severa [DDF(1996)].

El **bióxido de azufre** (SO_2) se produce de la combustión de cualquier sustancia que contenga azufre, se estima que la aportación de este gas por parte de la industria es de aproximadamente el 97%, el resto corresponde a la emisión de vehículos automotores que utilizan diesel y otras fuentes como la combustión de carbón, el refinamiento de petróleo y sus derivados, la fundición de metales con alto contenido de azufre y las erupciones volcánicas. Es soluble en agua y al hidratarse da lugar a la formación de ácido sumamente agresivo, cuando se mezcla con la humedad de las mucosas conjuntivas y respiratorias constituye un riesgo al producir irritación e inflamación aguda o crónica, y con la acción conjunta de partículas suspendidas, da lugar a un riesgo superior. Cuando se inhala o se disuelve con gotas de agua y penetran por vía respiratoria se puede alojar en los pulmones y se convierte en un agente irritante del tracto respiratorio inferior. Sus principales repercusiones son: la reducción de las funciones pulmonares, complicación de enfermedades respiratorias (asma, bronquitis crónica, enfisema pulmonar) y el incremento de muertes en enfermos crónicos del corazón y vías respiratorias.

El **bióxido de nitrógeno** (NO_2) se produce por la oxidación del nitrógeno atmosférico durante los procesos de combustión, se estima que los vehículos aportan el 62% a la atmósfera y la industria el 38%, además, puede ser producido en menor escala por tormentas eléctricas [DDF(1996)]. Es potencialmente dañino para la salud, se estima que es 4 veces más tóxico que el monóxido de nitrógeno. Su acumulación en el cuerpo humano constituye un riesgo para las vías respiratorias, ya que se ha comprobado que puede alterar la capacidad de respuesta de las células en el proceso inflamatorio en células polimorfonucleares, macrófagos alveolares y los linfocitos, observándose con mayor frecuencia en casos de bronquitis crónica; causa además problemas de percepción olfativa, molestias y dolores respiratorios agudos, edema pulmonar (acumulación de fluido) y en casos graves la muerte. De los daños producidos a los materiales, el más común es el decoloramiento de pinturas; en las plantas provoca la caída prematura de las hojas e inhibición del crecimiento [DDF(1996)].

1.1.2 Marco epidemiológico

Para evaluar los riesgos de la exposición a la contaminación y para fijar las normas de calidad del aire, lo ideal sería conocer con precisión la relación dosis/respuesta para cada uno de los contaminantes atmosféricos y sus combinaciones, para los diferentes grupos de la población expuesta. Sin embargo, en la actualidad no se dispone de

la información necesaria para la mayoría de estos grupos, por lo que para fijar las normas se utilizan algunos coeficientes de seguridad según el grado de protección que se le quiere dar a la población en general.

Resultados de numerosos estudios indican que la exposición a diferentes contaminantes puede ocasionar reducción en las funciones pulmonares, es decir, cambios agudos en la función, estructura y metabolismo pulmonar y efectos isquémicos en otros órganos (corazón, cerebro, hígado, entre otros) contribuyendo a aumentar la frecuencia de las enfermedades respiratorias, así como el incremento de la mortalidad en enfermos crónicos del corazón y vías respiratorias [DDF(1996)].

En los últimos años se ha publicado un número importante de artículos informando sobre los efectos en la salud de los contaminantes del aire a niveles muy cercanos a las normas actuales, en especial, para el ozono y partículas suspendidas, considerados como buenos indicadores para medir la calidad del aire. Desde una perspectiva de análisis del riesgo, el hecho de que en la zona metropolitana del Valle de México se rebasan prácticamente todos los días del año la norma de ozono y que estos excedentes a la norma ocurren en toda la zona urbana, permite considerar que el 100% de la población de la ciudad se ve expuesta con frecuencia y por periodos de una o más horas, a concentraciones de ozono superiores a la norma actual [DDF(1996)].

Un **estado isquémico** o simplemente **isquemia** se presenta, por lo general, en los padecimientos de tipo pulmonar, respiratorio y cardiovascular, y se manifiesta como una detención o disminución de circulación de la sangre a través de las arterias a cierta zona del órgano afectado, o bien, se presenta como un estado de sufrimiento celular, por ausencia de oxigenación o de materias nutritivas en la parte afectada del órgano, los casos más comunes de isquemia son la apoplejía (cerebro) y la angina de pecho (corazón).

1.1.3 Antecedentes

La Ciudad de México se ha ido extendiendo paulatinamente desde su formación, primero en detrimento de los antiguos lagos, posteriormente en el de excelentes tierras y bosques. El crecimiento demográfico y territorial de la ciudad data desde la época de los Aztecas, teniendo numerosos altibajos, pero a partir del siglo XX, en especial en los setentas, es cuando se hace más notorio.

Varios estudios señalan las causas que motivan el crecimiento demográfico de la Ciudad de México, entre las que pueden mencionarse: el hecho de que los programas de control de crecimiento demográfico se establecieron en el país varias décadas después de cuando eran necesarias; la centralización de las actividades políticas, económicas, educativas, de investigación, artísticas, culturales y el gran desarrollo de la actividad industrial que se viene dando desde los últimos cuarenta años, además, hay que sumar la creciente migración de las áreas rurales hacia la ciudad en busca de

empleo y mejores condiciones de vida, tal crecimiento demográfico va extendiéndose hacia las áreas industriales del Estado de México [Calderón(1987)].

La consecuencia asociada al crecimiento de la mancha urbana en la Ciudad de México, por sus dimensiones y la magnitud de su población, es que las necesidades de sus habitantes no sean cubiertas de manera suficiente en cuanto a servicios; por ejemplo en el uso del agua se observan dos grandes problemas, por un lado no se cuenta con una cuenca con la calidad y cantidad adecuada para satisfacer la gran demanda y el indispensable desalojo de los grandes volúmenes que se desechan y se precipitan; finalmente la contaminación ambiental provocada por los múltiples procesos productivos de la ciudad (transportación, industria, etc.) [Ezcurra(1990)].

La contaminación del aire y la basura son problemáticas de gran trascendencia en la Ciudad de México. El problema de la calidad del aire se refiere a los contaminantes que se emiten a la atmósfera, que son producidos por las más de 30,000 fábricas en el área metropolitana de la ciudad, que comprenden principalmente refinerías, industrias químico-farmacéuticas, alimentarias, cementeras, incineradores industriales, termoeléctricas, donde más de 250 manejan residuos peligrosos y cerca de 4,000 desarrollan procesos de combustión o transformación con emisiones significativas a la atmósfera. Otras fuentes naturales y artificiales de contaminación de diferente magnitud son: las áreas erosionadas, el fecalismo al aire libre, tortillerías, baños públicos, talleres, fábricas de cerveza, cigarros, tabiqueras, etc. Estas fuentes de contaminación son responsables de la emisión de la mayor parte de las partículas sólidas, el bióxido de azufre y los óxidos de nitrógeno que contaminan la ciudad [Legorreta(1991)].

Los automóviles son los principales responsables de los contaminantes atmosféricos en la Ciudad de México, producen la mayor parte de monóxido de carbono y de los residuos de hidrocarburos, debido a que su funcionamiento es mediante gasolina y otros combustibles fósiles. Se calcula que representan alrededor del 85% de los contaminantes del aire, mientras que el resto se debe a la industria, desechos orgánicos y tolvaneras [Bravo(1987)].

La altura de la Ciudad de México está directamente relacionada con la eficiencia de la combustión de los vehículos, a 2,240 metros sobre el nivel del mar un m^3 de aire contiene 212g de oxígeno (O_2), mientras que al nivel del mar el mismo volumen contiene 275 g; esto determina la eficiencia de combustión, ya que, al nivel del mar un auto bien afinado desarrolla la combustión al 92% mientras que en la Ciudad de México ésta es del 69%, esto explica que 3 millones de autos que circulan en la ciudad liberen una cantidad de contaminantes similar a la que emitirían 11.63 millones de vehículos al nivel del mar.

Se estima que el parque vehicular de la Ciudad de México es mayor a 3 millones de autos de los cuales cerca del 60% son modelos anteriores a 1982, de estos el 86% corresponde a particulares, 4% a taxis y colectivos y el 10% a autobuses y camiones. El crecimiento de vehículos es mucho más rápido que el crecimiento poblacional y la

expansión urbana; entre 1950 y 1980 el número de vehículos en el D.F. aumentó 24 veces. En 1940 existía un vehículo por cada 36 habitantes, para 1989 la proporción era de 1:6 [Bravo(1987)].

Las dimensiones de la ciudad determinan que grupos considerablemente grandes de personas se tengan que desplazar a través de distancias importantes para llegar a sus centros de trabajo. Se calcula que se realizan aproximadamente 21 millones de viajes-persona al día en la ciudad, de estos, 4 corresponden a autos particulares, 2 a taxis y colectivos, 6 al metro, 6 a autobuses urbanos y 3 a autobuses suburbanos. Se contempla que la ciudad tiene 7,000 unidades de autobuses urbanos, cada uno efectúa 860 viajes-persona al día, mientras que los autos particulares hacen solamente 2 viajes-persona al día, esto implica que el autobús consume 160 litros de diesel, realizando en un día de trabajo lo que realizarían 570 autos particulares que consumen alrededor de 3,700 litros de gasolina [Peralta(1989)].

1.2 Área de estudio

La Ciudad de México se sitúa en una latitud tropical (19° latitud norte y $98^\circ 15'$ longitud este) sus características climatológicas y topográficas están asociadas al hecho de situarse en el interior de una cuenca muy elevada (2,240 metros sobre el nivel del mar (msnm) al sur y 2,390 msnm en el norte), rodeada de importantes cadenas montañosas, esta situación determina que con frecuencia aumente la concentración de contaminantes [Jáuregui(1988)].

El Valle de México está integrado por el Distrito Federal y parte de los estados de México, Hidalgo, Tlaxcala, y Puebla [Reyes(1990)]. La cuenca que lo forma, es rodeada por montañas que sobrepasan en su mayoría los 3,000 metros al sur y los 2,800 m al norte, sobresaliendo el Iztaccihuatl (5,826 m) y el Popocatepetl (5,483 m); y esta delimitada por sierras: al sur Chichinautzin y el Ajusco, al sureste las Cruces, Monte Alto y Monte Bajo; al noroeste Tepotzotlán y Tezontlalpan, al norte la de Pachuca; al noreste por Chichucuaatla, al este Tepozán y Calpulalpan, comunicándose esta última con las sierras de Río Frío y la Nevada [Reyes(1990)]. En el interior de la cuenca se encuentran otras formaciones orográficas, como las sierras, al este de la Ciudad de México la de Santa Catarina, al norte y noroeste las de Guadalupe, Patlachique y las Pitayas; así como algunas elevaciones aisladas, como los cerros del Peñón de los Baños, Peñón de Marqués, Xihuingo, el Gordo y el de la Estrella entre los más importantes.

El clima del Valle de México muestra ciertas particularidades por su situación, altitud y topografía. Los climas que se pueden encontrar en la región son de cuatro tipos: semiseco, templado semihúmedo, frío y muy frío con varios subtipos que dependen de la altitud, además prevalecen las lluvias de tipo orográfico.

En la ciudad de México, el régimen climático puede dividirse en dos estaciones: la seca que va de noviembre a abril y la de lluvias, de mayo a octubre. Por su

1.3 Monitoreo de los contaminantes atmosféricos

gran altitud sobre el nivel del mar intervienen algunas características de las zonas templadas y frías, como la disminución de la temperatura, los vientos altos y secos del oeste en invierno y el calentamiento intenso del aire en verano que origina movimientos convectivos. La precipitación pluvial promedio es de 7.8 mm al día. La humedad relativa es aproximadamente de 45% en marzo y 75% en diciembre.

El hecho de que la ciudad se encuentre rodeada de montañas propicia una circulación de vientos provenientes de noroeste a noreste, con una alta incidencia de calmas e inversiones térmicas. El número de días despejados al año está entre 100 y 200.

Una observación interesante es que cuando la cuenca del Valle de México se empezó a poblar, el área correspondiente a la Ciudad de México se encontraba en su mayor parte ocupada por un sistema lacustre que cubría aproximadamente 1,500 km² que era formada por cinco lagos minúsculos conectados de norte a sur: Tzompamco, Xaltocan, Texcoco, Xochimilco y Chalco, los cuales se fueron desecando de manera directamente proporcional al crecimiento de la ciudad. Climatológicamente esto es importante, pues los límites del noreste de la ciudad tienen características semiáridas, pues el lecho arcilloso del ex lago de Texcoco es fácilmente erosionado por los vientos, especialmente durante los tres primeros meses del año, que es cuando se producen las llamadas tolvaneras; la superficie de estas áreas prácticamente carece de vegetación, lo que resulta excelente para absorber la energía solar, elevándose notablemente la temperatura en días soleados [Jáuregui(1988)].

1.3 Monitoreo de los contaminantes atmosféricos

En el ámbito nacional, la contaminación atmosférica se limita a las zonas de alta densidad demográfica o industrial. Las emisiones anuales de contaminantes en el país son superiores a los 16 millones de toneladas, de las cuales el 65% es de origen vehicular, en la Ciudad de México se genera 23.5% de dichas emisiones, en Guadalajara el 3.5%, y en Monterrey el 3%, el 70% restante lo generan los otros centros industriales del país.

La Ciudad de México, cuenta desde los setenta con una red manual de 16 estaciones de muestreo del aire con las que se miden las partículas totales, bióxido de azufre y algunos metales como el plomo. A partir de los ochenta (1986) se pone en funcionamiento el sistema denominado Red Automática de Monitoreo Atmosférico (RAMA), que es un conjunto de estaciones automáticas en las que se miden parámetros meteorológicos que pueden afectar la precipitación y dispersión de los contaminantes, tales como la dirección y velocidad del viento, humedad relativa, altura de la capa de inversión; además tres unidades de sonar que realizan sondeos atmosféricos con el fin de registrar los gradientes de temperatura en la zona metropolitana.

La RAMA cuenta con 25 estaciones a lo largo y ancho de la ciudad, donde las

Generalidades

estaciones del Cerro de la Estrella, Merced, Pedregal, Xalostoc y Tlalnepantla son las más completas, ya que monitorean cinco contaminantes y cuatro parámetros meteorológicos e informa sobre los valores atmosféricos de los siguientes contaminantes: bióxido de azufre, monóxido de carbono, óxidos de nitrógeno, ozono, partículas suspendidas totales, plomo, ácido sulfhídrico y últimamente algunos hidrocarburos y otros contaminantes como cadmio, cobre, cobalto y níquel.

La RAMA registra concentraciones muy pequeñas de contaminantes por separado, de acuerdo a un tiempo de muestreo que esté relacionado con los criterios de la calidad del aire. Desde 1982, se reportan las concentraciones finales que se tienen para la Ciudad de México. Los valores máximos recomendables de concentración atmosférica de los contaminantes (ver Tabla 1.1) para que no se afecte la salud de ningún grupo de la población se determinan por estudios epidemiológicos y experimentales, sobre los efectos de la contaminación aérea en la salud a corto y largo plazo [Reyes(1990)].

Para determinar la concentración de una sustancia química en volumen se utilizan principalmente las partes por millón (PPM), partes por billón (PPB) o su correspondiente equivalencia en microgramos sobre metro cúbico ($\mu\text{g}/\text{m}^3$). Las PPM y PPB significan que se divide el volumen en un millón o un billón de partes iguales. Cada fracción de parte de este volumen, correspondiente a la sustancia de interés, se considera una parte por millón o billón de la sustancia. Estas medidas se utilizan para determinar concentraciones muy pequeñas de gases en la atmósfera.

Tabla 1.1: Criterios de la calidad del aire.

Contaminante	Concentración		Tiempo de muestreo
	$\mu\text{g}/\text{m}^3$	PPM	
Ozono	216	0.11	1 h (promedio máximo)
Monóxido de Carbono		0.11	8 h (promedio)
Partículas Suspendidas Totales	150	-	24 h (promedio)
Bióxido de Azufre	340	0.13	24 h (promedio móvil)
Bióxido de Nitrógeno	395	0.21	1 h (promedio máximo)

1.4 Índices de la calidad del aire

La calidad del aire se evalúa según las mediciones realizadas sobre cada uno de los contaminantes y se expresa en función del denominado Índice Metropolitano de Calidad del Aire (IMECA) que se define como un valor representativo de los niveles de contaminación atmosférica y sus efectos en la salud, dentro de una región determinada.

La función principal del IMECA es mantener informada a la población sobre la calidad del aire en la Ciudad de México, así como observar el comportamiento de los distintos contaminantes y comparar la calidad del aire entre zonas que utilicen

1.4 Índices de la calidad del aire

índices similares. Tal indicador se elabora usando datos de las 18:00 horas del día anterior hasta las 18:00 horas del día en curso, por lo que se promedia la noche anterior y el día de la fecha del informe.

El informe se basa en una metodología de cálculo a partir de dos puntos de quiebre, que son valores estadísticamente conocidos: los umbrales críticos se definen como IMECA 100 y 500, que diagnostican las posibles alteraciones significativas en la fisiología de los humanos. Las rectas que unen los puntos de quiebre sirven para convertir valores y concentración de contaminantes del aire en valores de una escala arbitraria que va de 0 a 500 puntos IMECA (ver tabla 1.2), que proporcionan una idea del grado de peligrosidad asociado a los niveles de contaminación del aire. Los índices obtenidos de estas rectas (conocidas como funciones linealmente segmentadas) miden la calidad del aire [Ezcurra(1990)] con respecto a:

1. Partículas en suspensión.
2. Bióxido de azufre.
3. Ozono.
4. Monóxido de carbono.
5. Óxidos de nitrógeno.
6. La acción sinérgica del bióxido de azufre y partículas suspendidas totales.

Tabla 1.2: Índice Metropolitano de la Calidad del Aire (IMECA)

IMECA	Calidad del aire	Efectos
0 - 100	Satisfactoria	Situación favorable para la realización de todo tipo de actividades
101 - 200	No satisfactoria	Aumento de molestias menores en personas sensibles
201 - 300	Mala	Aumento de molestias e intolerancia relativa al ejercicio en personas con afecciones respiratorias
301 - 500	Muy mala	Aparición de diversos síntomas e intolerancia al ejercicio de la población en general.

Además de generar informes diarios de la calidad del aire, se genera un pronóstico de situaciones críticas que permite prever con anticipación la presencia de fenómenos meteorológicos y actuar de acuerdo a un Programa de Atención a Contingencias Ambientales, para prevenir y atender los episodios de contaminación atmosférica. Este programa contempla la suspensión de actividades industriales y de la circulación vehicular en caso de eventos extraordinarios, para reducir la emisión de contaminantes.

1.5 Planteamiento del problema

Este trabajo estima la relación a corto plazo del número de muertes ocasionadas por trastornos isquémicos de la población de la Ciudad de México mayor a 65 años, durante el año de 1994.

Por la naturaleza de la variable respuesta, conteos de muertes en la población, la variable es entera no negativa y los valores que representa son pequeños con relación a la población de estudio, el análisis que se propone es el desarrollo de un modelo de regresión Poisson, además el objetivo de la investigación es presentar una asociación entre la variabilidad observada de las muertes diarias como consecuencia de la contaminación ambiental, empleando la técnica antes mencionada.

Capítulo 2

Modelos Lineales Generalizados

2.1 Introducción

Un modelo es una abstracción de la realidad, este puede ser clasificado como determinístico o probabilístico. El primero se caracteriza porque no tiene ningún elemento aleatorio y su solución es exacta, en el segundo ocurre todo lo contrario, tiene elementos aleatorios los cuales sólo permiten resolverlo óptimamente de forma estocástica.

Una clase de los modelos probabilísticos son los que se establecen por una relación lineal, donde existe un conjunto de variables de las cuales una es la respuesta y el resto influyen o explican aproximadamente este resultado de manera lineal; así, la respuesta carece de una relación exacta debido a que contiene al menos un elemento aleatorio que no se puede controlar y se presume tiene algún patrón distribucional conocido. El ejemplo más conocido es el que involucra errores que se asocian con la distribución normal y es el más sencillo de ajustar; no obstante, por la naturaleza de los datos, estas desviaciones pueden tener un comportamiento diferente.

Los modelos lineales generalizados fueron introducidos por Wedderburn y Nelder [Nelder(1972)], son una clasificación que incorporan una rica familia de distribuciones con errores normales y no normales y se ajustan a respuestas univariadas y a una clase de distribuciones llamada familia exponencial, que incluye la normal, binomial, Poisson, geométrica, binomial negativa, exponencial, gamma, normal inversa entre otras. Estos se caracterizan por tres elementos:

1. Una distribución de la variable respuesta (llamada la estructura del error).
2. Un predictor lineal que involucra a las covariables.
3. Una función liga que relaciona al predictor lineal con la media de la variable respuesta.

Usualmente también se asume que las observaciones son independientes.

2.2 Aplicaciones

Las aplicaciones que tienen estos modelos concierne en esencia a la manera de cómo una respuesta depende de una combinación de las variables explicativas y muy en particular de su naturaleza; como por ejemplo, los conteos que involucran defectos o eventos raros, tal como pacientes con un padecimiento en particular o la ocurrencia de algún fenómeno natural; respuestas binarias cuya respuesta es acierto o error. Comúnmente se emplean en las ciencias de la ingeniería; hay respuestas no normales siendo continuas, en general ejemplificadas en fallas de componentes mecánicos, tiempo de falla de componentes electrónicos o sistemas. Estos tipos de respuestas en todos los casos son no negativas y exhiben un elevado sesgo a la derecha.

Desde su desarrollo se han implementado modelos en áreas de investigación con una gran variedad de nombres como: análisis histórico de eventos (sociología); análisis de relatividad y de tiempo de falla (ingeniería); análisis de duración y transición (economía) y en otras disciplinas con diferentes tipo de eventos, tanto en áreas de las ciencias sociales o naturales, incluyendo ocurrencia de terremotos, accidentes en automóviles, caídas de mercados de acciones, contaminación, revoluciones, jubilaciones, casamientos, divorcios, nacimientos, promociones y arrestos entre otros.

Lo que se ha realizado en la actuaria y las finanzas concierne a eventos que pueden provocar pérdidas financieras como:

1. La destrucción de propiedades por fuego o catástrofes naturales (tormentas, granizadas, inundaciones, aludes, terremotos o erupciones volcánicas) es usualmente considerada un evento aleatorio en el cual la pérdida puede ser medida en términos monetarios.
2. Una indemnización por daño impuesta por un juez como el resultado de un hecho negligente en es ocasiones considerado como un evento aleatorio con una pérdida monetaria como resultado.
3. Una enfermedad prolongada que ocurra inesperadamente, resulta una pérdida en términos del ingreso reducido y el costo extra en cuidados de salud.
4. La muerte de un adulto joven puede ocurrir mientras su familia tiene compromisos a largo plazo (pago de hipotecas, financiamientos, etc) que no pueden ser cubiertos.
5. La supervivencia en edad avanzada puede agotar los recursos de las personas, considerando el costo de vida e incluyendo cuidados de salud.

En particular, para las clases que se incluyen en los modelos lineales generalizados (regresión lineal, análisis de varianza, log-lineales para tablas de contingencia, logit para datos binarios en la forma de proporciones entre otros) hay un buen número de aplicaciones prácticas en el trabajo actuarial como son:

2.3 La familia de distribuciones exponencial

1. En supervivencia: modelación y graduación con respecto a la edad en presencia de duplicaciones; con respecto a la edad y el tiempo; con respecto a la edad y la duración de la póliza; empleando cantidades basadas en los datos; o modelación dual.
2. Modelos de estados múltiples o jerárquicos en seguros de salud.
3. Ciertos ajustes de distribuciones de pérdida por severas reclamaciones en seguros de no vida.
4. Clasificación de riesgo como la modelación de: la mortalidad para personas con deterioro de vida que pueda asegurarse; de exceso de mortalidad entre fumadores; tasas de seguro de vida en un intervalo de tiempo.
5. Fijación de primas en seguros de no vida basados en la frecuencia y severidad de reclamos.
6. Reclamos de reservas en seguros de no vida.

Cada una de estas aplicaciones envuelve diferentes tipos de datos y modelos, antes de entrar al análisis de los datos de este estudio se dará una descripción de la naturaleza y estructura de los modelos lineales generalizados.

2.3 La familia de distribuciones exponencial

Una importante base que unifica el concepto de los modelos lineales generalizados es la familia de distribuciones exponencial cuya forma general es:

$$f(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \quad (2.1)$$

para una respuesta observada y , donde $a(\cdot)$, $b(\cdot)$ y $c(\cdot)$ son funciones específicas y θ y ϕ son los parámetros de localización y dispersión respectivamente, el segundo también es conocido como ruido. La función $a(\phi)$ es igual a $\phi\omega$ donde ω es una constante conocida y para algunas distribuciones el parámetro de dispersión $\phi = 1$, al final de este capítulo se mostrará la identificación de los componentes de la forma general para las distribuciones normal, Poisson y binomial. Para el caso de sobredispersión de la Poisson, en el análisis se necesita seleccionar un parámetro adicional que se detallará en la sección 2.6.3.

A partir de las funciones específicas que componen la forma general de la familia exponencial y empleando los fundamentos básicos de inferencia estadística se puede identificar la media y la varianza de cada miembro de esta familia.

Partiendo de los siguientes resultados:

$$E \left[\frac{\partial \ln L(y; \theta)}{\partial \theta} \right] = 0 \quad (2.2)$$

$$E \left[\frac{\partial^2 \ln L(y; \theta)}{\partial \theta^2} \right] + E^2 \left[\frac{\partial \ln L(y; \theta)}{\partial \theta} \right] = 0 \quad (2.3)$$

donde $L(y; \theta)$ es la función de verosimilitud de las observaciones dado el parámetro θ , y considerando la ecuación (2.1) en (2.2) se obtiene que:

$$b'(\theta) = \mu$$

Considerando nuevamente (2.1) en (2.3) se tiene que

$$E \left[\frac{\partial}{\partial \theta} \left(\frac{y - b'(\theta)}{a(\phi)} \right) \right] = -\frac{1}{a^2(\phi)} E \left[(y - b'(\theta))^2 \right]$$

como $E[y] = b'(\theta) = \mu$ entonces

$$b''(\theta) = \frac{Var[y]}{a(\phi)}$$

por lo tanto $Var[y] = a(\phi)b''(\theta)$, visto de otra forma, se tiene que la varianza se puede expresar como:

$$Var[y] = \frac{d^2 b(\theta)}{d\theta^2} a(\phi) = \frac{d\mu}{d\theta} a(\phi)$$

con excepción de $a(\phi)$, se ve la dependencia de la varianza de la respuesta con su media. Si se denota como Var_{μ}

$$Var_{\mu} = \frac{Var[y]}{a(\phi)} = \frac{d\mu}{d\theta}$$

y así también

$$\frac{d\theta}{d\mu} = \frac{1}{Var_{\mu}}$$

2.4 Componentes de los modelos lineales generalizados

En esta sección se describen los modelos lineales generalizados como una extensión de los modelos lineales clásicos, esto es consecuencia de la forma de la familia exponencial definida por la distribución de probabilidad dada en la ecuación (2.1), la estructura es como sigue:

1. Se cuentan con y_1, y_2, \dots, y_n observaciones de la respuesta independientes con $\mu_1, \mu_2, \dots, \mu_n$ respectivamente.

2. Cada observación y_i tiene una misma distribución que es miembro de la familia exponencial, aunque pueden tener diferentes parámetros.
3. La porción sistemática del modelo involucra regresores o variables independientes que influyen sobre la respuesta denotados como x_1, x_2, \dots, x_k .
4. El modelo se constituye alrededor de un predictor lineal

$$\eta = \mathbf{x}'\boldsymbol{\beta} = \beta_0 + \sum_{i=1}^k \beta_i x_i$$

donde los β_i son parámetros que usualmente son desconocidos y serán estimados con los datos, es decir, con las respuestas y los regresores, además dicho así, el predictor lineal sugiere una terminología de los modelos lineales generalizados.

5. El modelo se busca a través del uso de una función liga

$$\eta_i = g(\mu_i) \quad i = 1, 2, \dots, n$$

el término liga es derivado del hecho de que la función es la "liga" entre la media y el predictor lineal. Nótese que la esperanza de la respuesta es:

$$E[y_i] = g^{-1}(\eta_i) = g^{-1}(x_i'\boldsymbol{\beta})$$

De hecho en el modelo de regresión múltiple

$$\mu_i = \eta_i = x_i'\boldsymbol{\beta} \quad i = 1, 2, \dots, n$$

lo que sugiere el caso especial donde $g(\mu_i) = \mu_i$ y la función liga que se emplea es la identidad.

6. La función liga es monótona derivable. [Wedderburn(1972)]
7. La varianza σ_i^2 ($i = 1, 2, \dots, n$) es una función de la media. [Wedderburn(1974)]

Los puntos anteriores simplemente mencionan la caracterización de los modelos lineales generalizados, en el artículo de Nelder y Wedderburn [Nelder(1972)] o en el título de McCullagh y Nelder [McCullagh(1983)] se profundiza en detalle la estructura de ellos.

Como se menciona, la función liga relaciona el predictor lineal con la media de la variable respuesta, su elección depende en particular de su distribución que debe pertenecer a alguna familia exponencial específica, para cada una existe una función liga llamada canónica o natural, que relacionan el parámetro natural directamente con el predictor lineal:

$$\theta = \theta(\mu) = \eta = \mathbf{x}'\boldsymbol{\beta}$$

Tabla 2.1: Ligas canónicas para los modelos lineales generalizados

Distribución	Función liga	Nombre
normal	$\eta_i = \mu_i$	liga identidad
binomial	$\eta_i = \ln\left(\frac{p_i}{1-p_i}\right)$	liga logística
Poisson	$\eta_i = \ln \mu_i$	log-liga
gamma	$\eta_i = \frac{1}{\mu_i}$	liga recíproca
normal inversa	$\eta_i = \frac{1}{\mu_i^2}$	liga recíproca

En la tabla 2.1 se muestran la función liga canónica de las distribuciones más conocidas en los modelos lineales generalizados.

Hay otras funciones liga que se pueden usar en un modelo lineal generalizado:

1. La liga probit:

$$\eta_i = \Phi^{-1}(E[y_i])$$

donde Φ representa la función de distribución acumulativa de una normal estándar.

2. La liga log-log complementaria:

$$\eta_i = \ln(\ln[1 - \mu_i])$$

3. La liga de la familia potencia:

$$\eta_i = \begin{cases} \mu_i^\lambda & \lambda \neq 0 \\ \ln \mu_i & \lambda = 0 \end{cases}$$

La idea básica de los modelos lineales generalizados contempla dos componentes: la distribución de la respuesta y la función liga.

Se puede ver la selección de la función liga como una transformación de la respuesta, es importante saber que modifica la media poblacional y no a los datos. El no usar una liga adecuada puede generar problemas en cuanto al ajuste del modelo lineal generalizado.

La ventaja de la función liga sobre otro tipo de transformaciones, es que esta directamente relacionada con las características distribucionales de las respuestas, por lo tanto las elecciones inapropiadas se reflejan en la significancia global del modelo.

El método de máxima verosimilitud es la base teórica para la estimación de los parámetros en estos modelos; no obstante, la obtención de los resultados por esta vía se basa en un algoritmo denominado “mínimos cuadrados ponderados iterativos” (iterative reweighted least squares).

2.4 Componentes de los modelos lineales generalizados

Considerando el método de máxima verosimilitud aplicado a los modelos lineales generalizados y suponiendo que se emplea la liga canónica a la función de log-verosimilitud:

$$l = \ln L(y; \beta) = \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)$$

para la liga canónica se tiene que $\mu_i = \eta_i = x_i' \beta$, por lo tanto:

$$\begin{aligned} \frac{\partial L}{\partial \beta} &= \frac{\partial L}{\partial \theta_i} \frac{\partial \theta_i}{\partial \beta} \\ &= \sum_{i=1}^n \frac{1}{a(\phi)} \left[y_i - \frac{d}{d\theta_i} b(\theta_i) \right] \\ &= \sum_{i=1}^n \frac{1}{a(\phi)} [y_i - \mu_i] x_i \end{aligned}$$

entonces se pueden encontrar los estimadores de máxima verosimilitud de los parámetros resolviendo el siguiente sistema de ecuaciones:

$$\sum_{i=1}^n \frac{1}{a(\phi)} (y_i - \mu_i) x_i = 0$$

en la mayoría de los casos se asume que $a(\phi)$ es constante por lo que el sistema se reduce a:

$$\sum_{i=1}^n (y_i - \mu_i) x_i = 0$$

Hay que recordar que el sistema anterior parte de que μ_i es una combinación lineal de las x_1, \dots, x_k , por lo tanto el sistema tiene $p = k + 1$ ecuaciones, una por cada parámetro desconocido β_j ($j = 1, \dots, p$), la forma matricial es:

$$\mathbf{X}'(y - \mu) = \mathbf{0} \tag{2.4}$$

donde $\mu' = (\mu_1, \mu_2, \dots, \mu_n)$. La solución del sistema son los estimadores de máxima verosimilitud, al sistema (2.4) se le conoce como la función de "puntaje" o "score" y como consecuencia de la no linealidad que se presenta, se recurre al proceso iterativo de los mínimos cuadrados ponderados, tal procedimiento producirá $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ estimadores de $\beta_0, \beta_1, \dots, \beta_k$.

Se tiene que $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$ es el valor final que el algoritmo produce como estimación de β y si los supuestos del modelo incluyendo la elección de la función liga es correcta, entonces se puede mostrar normalidad asintótica en $E[\hat{\beta}] = \beta$, donde $\hat{\beta}$ es la solución resultante del proceso iterativo de las ecuaciones de puntaje.

Ahora la matriz de información definida por $I[\hat{\beta}]$ de los estimadores, proporciona la varianza de la ecuación de puntaje como sigue:

$$I[\hat{\beta}] = \text{Var} \left[\frac{1}{a(\phi)} [\mathbf{X}'(y - \mu)] \right] = \frac{\mathbf{X}'\mathbf{V}\mathbf{X}}{[a(\phi)]^2}$$

donde $\mathbf{V} = \text{diag}(\sigma_i^2)$ y σ_i^2 es una función de μ_i que a su vez depende de la distribución en cuestión. La matriz de varianzas y covarianzas asintótica de $\hat{\beta}$ este dada por:

$$\text{Var}(\hat{\beta}) = I^{-1}[\hat{\beta}] = [\mathbf{X}'\mathbf{V}\mathbf{X}]^{-1} \times [a(\phi)]^2 \quad (2.5)$$

Los errores estándar de los coeficientes provienen de las raíces cuadradas de los elementos de la diagonal de la matriz de la expresión (2.5) con elementos de \mathbf{V} reemplazados por estimadores apropiados.

De la reciente discusión, se considera nuevamente las distribuciones que se manejan al final del capítulo. Para el caso normal las $\sigma_i^2 = \sigma^2$ y $a(\phi) = \sigma^2$, entonces:

$$\text{Var}(\hat{\beta}) = [\mathbf{X}'\mathbf{X}]^{-1} \sigma^2$$

para los modelos Poisson y binomial, $a(\phi) = 1$ por lo que

$$\text{Var}(\hat{\beta}) = [\mathbf{X}'\mathbf{V}\mathbf{X}]^{-1}$$

para el caso Poisson:

$$\sigma_i^2 = e^{\mathbf{x}_i'\boldsymbol{\beta}}$$

y para el caso binomial:

$$\sigma_i^2 = n_i P_i (1 - P_i) = \frac{n_i e^{-\mathbf{x}_i'\boldsymbol{\beta}}}{(1 + e^{-\mathbf{x}_i'\boldsymbol{\beta}})^2}$$

Como se mencionará en el apartado 2.6.3, las varianzas en los elementos de la diagonal de la matriz (2.5) pueden ser infladas, a este efecto se le conoce como sobredispersión.

Toda la discusión anterior parte de considerar que la liga canónica es la transformación más natural dada una distribución, sin embargo, esto no significa que una función liga que no sea canónica no se considere; en la regresión lineal múltiple estándar la elección de esta es equivalente al ejercicio de seleccionar el modelo en este análisis. Una forma es intentar determinar la liga y así, el modelo podrá describir de mejor manera los datos. Si los rangos de los regresores son suficientemente limitados, el modelo lineal es suficiente, entonces la liga identidad deberá ser considerada.

Ahora supónganse que no se emplea la liga canónica, es decir, que $\eta_i \neq \theta_i$ entonces la derivada de la log-verosimilitud es

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \frac{\partial L}{\partial \theta_i} \frac{\partial \theta_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}}$$

teniendo que calcular por separado usando la expresión (2.1):

$$\frac{\partial L}{\partial \theta_i} = \sum_{i=1}^n \frac{1}{a(\phi)} \left[y_i - \frac{\partial}{\partial \theta_i} b(\theta_i) \right] = \frac{1}{a(\phi)} \sum_{i=1}^n [y_i - \mu_i] \quad \text{y} \quad \frac{\partial \eta_i}{\partial \boldsymbol{\beta}} = \mathbf{x}_i$$

se tiene por lo tanto la ecuación de puntaje:

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^n \frac{[y_i - \mu_i]}{a(\phi)} \frac{\partial \theta_i}{\partial \eta_i} x_i \quad (2.6)$$

como en el caso donde se usa la liga canónica, se asume nuevamente que $a(\phi)$ es una constante, entonces se puede expresar (2.6) de forma matricial como:

$$\mathbf{X}' \Delta (y - \mu) = 0 \quad (2.7)$$

donde $\Delta = \text{diag} \left(\frac{\partial \theta_i}{\partial \eta_i} \right)$

La matriz de información que resulta de la máxima verosimilitud de $\hat{\beta}$ y que ahora estima a β es

$$I[\hat{\beta}] = \frac{\mathbf{X}' \Delta \mathbf{V} \Delta \mathbf{X}}{[a(\phi)]^2}$$

La matriz de varianzas y covarianzas asintótica de $\hat{\beta}$ es

$$\text{Var}(\hat{\beta}) = [\mathbf{X}' \Delta \mathbf{V} \Delta \mathbf{X}]^{-1} \times [a(\phi)]^2$$

mientras que la ecuación de puntaje para la liga canónica es más sencilla que para el caso general como en la expresión de (2.7), se observa que μ y $\partial \theta_i / \partial \eta_i$ son casi siempre no lineales y la complejidad obtenida no tiene ningún sentido.

Considérese ahora el siguiente ejemplo, para la distribución Poisson se tiene que la liga canónica es:

$$\sum_{i=1}^n [y_i - e^{x_i' \beta}] x_i = 0 \quad (2.8)$$

se observa que no se puede resolver de manera puntual porque la ecuación (2.8) no es lineal en β , si se supone que se decide por una liga identidad que justificará de este modo linealidad aún cuando ésta en el modelo no sea canónica, se tiene que $\mu_i = x_i' \beta$ y $\partial \theta_i / \partial \eta_i = 1/x_i' \beta$ con la ecuación (2.7) se obtendrá la siguiente ecuación de puntaje:

$$\sum_{i=1}^n \frac{(y_i - x_i' \beta) x_i}{x_i' \beta} = 0$$

que sería resuelta y aunque el modelo es lineal es inevitable la estimación de los parámetros por un proceso iterativo, por lo que no se obtiene ninguna ganancia al cambiar la liga canónica.

2.5 Inferencia y análisis de residuales

Para pruebas de hipótesis en coeficientes individuales, la estadística de Wald hace el uso de normalidad asintótica del estimador de máxima verosimilitud. Así para probar:

$$H_0 : \beta_j = 0 \quad \text{vs.} \quad H_0 : \beta_j \neq 0$$

la prueba estadística a considerar es:

$$\left(\frac{\hat{\beta}_j}{es(\hat{\beta}_j)} \right)^2$$

donde $es(\hat{\beta}_j)$ es el error estándar y tiene una distribución χ_1^2 para muestras grandes.

Los intervalos de confianza sobre la media de la respuesta y predicción de la estimación de los coeficientes para una sola observación, con una aproximación al $100(1 - \alpha)\%$, Wald propone para el parámetro β_j

$$\hat{\beta}_j + Z_{\alpha/2} es(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j - Z_{\alpha/2} es(\hat{\beta}_j)$$

con $Z_{\alpha/2}$ el cuantil de la probabilidad acumulada de una distribución normal estándar.

El uso de la estadística de Wald permite realizar pruebas de hipótesis para el modelo completo o sobre los parámetros individuales en un modelo lineal generalizado, tales pruebas emplean la normalidad asintótica en los estimadores máximo verosimiles y se utiliza muy a menudo para la inclusión o no de los parámetros del modelo comparando las log verosimilitudes de dos modelos, el completo contra el reducido, es decir:

$$-2[\ln L(\text{modelo reducido}) - \ln L(\text{modelo completo})] = -2 \ln \frac{L(\text{modelo reducido})}{L(\text{modelo completo})}$$

de hecho, la expresión anterior se distribuye asintóticamente como una χ_{Δ}^2 , donde Δ es la diferencia del número de parámetros del modelo reducido y completo.

Por ejemplo, supónganse que se considera el predictor lineal:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

y se esta interesado en probar que:

$$H_0 : \beta_1 = \beta_2 = 0$$

la estadística de prueba esta dada por la razón de verosimilitud:

$$-2[\ln L(\beta_0, \beta_3) - \ln L(\beta_0, \beta_1, \beta_2, \beta_3)]$$

donde $L(\beta_0, \beta_3)$ es la verosimilitud para el modelo en el cual la hipótesis nula es que β_1 y β_2 son cero. Como resultado la hipótesis se rechaza si la log verosimilitud se incrementa significativamente por el remplazo de β_1 y β_2 en el modelo en la presencia de β_0 y β_3 . La distribución que se emplea en el límite superior de la prueba es una χ_2^2 en este caso. El uso de la razón de verosimilitud deriva a una prueba empleada en una expresión de nombre devianza y permite determinar en

teoría si el valor ajustado para un modelo es significativamente mejor que el modelo saturado dado por:

$$E[y_i] = \mu_i \quad i = 1, \dots, n$$

donde el modelo saturado requiere la estimación de n parámetros independientes.

Así, el término saturado tiene a y_i como estimador de μ_i . Como resultado de lo anterior hay cero grados de libertad en los residuales después de la estimación, exhibiendo además una verosimilitud no mayor comparando al correspondiente en el modelo de regresión ajustado y de manera análoga en el caso de errores normales, la devianza para un modelo lineal generalizado se define como:

$$D(\beta) = -2 \ln \left[\frac{L(\beta)}{L(\mu)} \right] \quad (2.9)$$

donde $L(\beta)$ y $L(\mu)$ corresponden a las verosimilitudes de los modelos ajustado y saturado, este último considera n parámetros y se emplea habitualmente como una estadística de bondad de ajuste, $D(\beta)$ tiene una distribución asintótica χ^2 con $n - p$ grados de libertad.

Un valor insignificante de $D(\beta)$ en el límite superior de la prueba implica que el modelo no es significativamente peor que el saturado, así un valor relativamente chico de devianza es favorable para el ajuste del modelo. Se puede utilizar una regla de dedo: considerando que la cantidad del ajuste es razonable si $\frac{D(\beta)}{(n-p)}$ no es apreciablemente más grande que 1. Sin embargo, hay ejemplos en los cuales el resultado asintótico no es satisfactorio para muestras pequeñas [Myers(2001)]. Se hace notar que en la definición (2.9), una razón de verosimilitud, por ejemplo $-2 \ln[L(\text{reducido}) - L(\text{completo})]$, se puede escribir en términos de la devianza como $D(\text{reducido}) - D(\text{completo})$, desde que $L(\mu)$ se cancela en la diferencia. Como el ejemplo que se desarrolló en esta sección donde se quería probar $H_0 : \beta_1 = \beta_2 = 0$ en un modelo de cuatro parámetros, la razón de verosimilitud con el concepto de devianza puede escribirse como:

$$2[\ln L(\beta_0, \beta_1, \beta_2, \beta_3) - \ln L(\beta_0, \beta_3)] = D(\beta_0, \beta_3) - D(\beta_0, \beta_1, \beta_2, \beta_3)$$

donde se simplifica y encuentra un mejor ajuste con el empleo de un método tipo "stepwise".

A continuación se proporcionan algunas expresiones de devianza de las distribuciones más comunes:

$$\text{normal} \quad \sum_{i=1}^n (y - \mu)^2$$

$$\text{Poisson} \quad 2 \sum_{i=1}^n \left[y \ln \frac{y}{\hat{\mu}} - (y - \hat{\mu}) \right]$$

$$\text{binomial} \quad 2 \sum_{i=1}^n \left\{ y \ln \frac{y}{\hat{\mu}} + (m - y) \ln \left(\frac{m - y}{m - \hat{\mu}} \right) \right\}$$

$$\text{gamma} \quad 2 \sum_{i=1}^n \left[-\ln \frac{y}{\hat{\mu}} + \frac{(y - \hat{\mu})}{\hat{\mu}} \right]$$

Si el predictor lineal contiene una constante, la devianza para los casos Poisson y gamma se reduce a

$$D(\beta) = \sum_{i=1}^n y \ln \left(\frac{y_i}{\hat{\mu}_i} \right)$$

Otra importante medida de ajuste de un modelo, es la estadística χ^2 de Pearson generalizada que toma la forma:

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\text{Var}(\hat{\mu}_i)} \quad (2.10)$$

donde la $\text{Var}(\hat{\mu}_i)$ es la varianza estimada para la distribución en cuestión.

Para modelos con errores normales la χ^2 es de nuevo la suma de cuadrados de los residuales, mientras que para las distribuciones Poisson y binomial es la estadística χ^2 de Pearson original. Como la devianza, la χ^2 de Pearson generalizada es asintóticamente distribuida como una χ^2 con $n - p$ grados de libertad.

La devianza tiene una ventaja como medida de ajuste: la presentación de aditividad para conjuntos de datos anidados, si los estimadores de máxima verosimilitud fueron empleados; mientras que la estadística χ^2 de Pearson generalizada en algunas ocasiones no, sin embargo puede utilizarse porque su interpretación es más directa que el uso de la devianza.

2.5.1 Uso de los residuales

Los residuales son la desviación de los valores observados contra los ajustados de la variable independiente, con frecuencia se emplean para detectar la violación de los supuestos, como la no homogeneidad de la varianza y la especificación del modelo; además de las observaciones aisladas o con ajuste pobre; también se utilizan para encontrar valores de influencia o con un gran impacto en el modelo.

El análisis de residuos, es particularmente visual e indica en potencia la naturaleza del error y sirve para sugerir las formas de corrección; además, proporciona

un sentido de la magnitud del efecto del error en el modelo. En contraste, las pruebas estadísticas formales los cataloga por cajas negras, produciendo solamente un número que puede ser comparado con el valor crítico, por otra parte, existen pruebas con el mismo nivel de significancia sin hacer caso al tamaño de la muestra y en algún modelo real invariablemente se rechaza una muestra suficientemente grande aunque el ajuste hecho a los datos sea correcto.

Hay tres tipos de residuales apropiados para los modelos lineales generalizados:

1. El residual natural o puro

$$r_i = (y_i - \hat{\mu}_i)$$

es importante observar que estas desviaciones son técnicamente inapropiadas y esto se debe a que la $Var(y_i)$ no es constante por la distribución teórica de los datos; sería deseable que fuera simétricamente distribuido alrededor del cero con varianza constante, para casos más generales $(y - \mu)$ es asimétrico y heterocedástico.

La corrección para la heterocedasticidad es la de Pearson,

$$r_p = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{w}_i}}$$

donde \hat{w}_i es un estimador de la varianza w_i de y_i . La suma de los cuadrados de estos residuales es la estadística de Pearson definida en la ecuación (2.10).

2. Residuales de devianza. En este tipo de residuo de observaciones y cuya función de densidad pertenece a la familia exponencial, se emplea la devianza residual y su valor se escribe de la siguiente manera:

$$\sum_{i=1}^n d_i = D(\beta)$$

los componentes individuales se definen como residuales de devianza y están dados por

$$\begin{aligned} d_{i,r} &= [sgn(y_i - \hat{\mu}_i)] \sqrt{d_i} \\ &= [sgn(y_i - \hat{\mu}_i)] \sqrt{2[l(y_i) - l(\hat{\mu}_i)]} \end{aligned}$$

recuérdese que $l(\hat{\mu}_i)$ y $l(y)$ corresponden a las log-verosimilitudes de y evaluadas en $\hat{\mu}_i$ y y respectivamente, nótese que la suma de las devianzas residuales es igual a la definida en la ecuación (2.9), es decir:

$$D(\beta) = \sum_{i=1}^n d_{i,r}^2$$

3. El residual de Ascombe se define como una transformación de y que se aproxima más a la normalidad estándar. En ocasiones cuando el residual de Pearson es marcadamente sesgado, el residual de Ascombe es una aproximación normal mucho mejor que otra transformación. Ascombe definió un residual usando una función $A(y)$ en lugar de y , donde $A(\cdot)$ se elige para hacer a la distribución de $A(Y)$ "tan normal como sea posible". Wedderburn mostró que para las funciones de verosimilitud en los modelos lineales generalizados la función $A(\cdot)$ esta dada por:

$$A(y) = \int \frac{d\mu}{\sqrt[3]{\mu}}$$

Seguramente la manera más fructífera de identificar los residuales es graficarlos contra otras variables, inclusive contra las predicciones de la respuesta; como por ejemplo, para ver si el ajuste es pobre para valores grandes o pequeños de la variable dependiente; o contra los regresores omitidos para ver si hay alguna relación e incluirlos a los residuales; o contra las variables independientes consideradas en el modelo para ver si estas deben estar por medio de una forma funcional diferente a la que esta especificada.

McCullagh y Nelder recomiendan graficar la devianza residual contra los valores ajustados, transformando a la escala de la varianza constante y contra los regresores. Estas gráficas son exactamente análogas a las de los residuales comunes utilizadas en la regresión múltiple y tienen exactamente la misma interpretación. También sugieren graficar los valores absolutos de las devianzas residuales contra los valores ajustados en la escala de la varianza constante. Una función mal elegida de la varianza producirá una tendencia en la gráfica. El empleo de la escala de la varianza constante para ambos conjuntos de gráficas, en general, es por razones estéticas. Para la mayor parte, esta transformación separa los datos hacia fuera a través del eje horizontal.

La tabla 2.2 da una transformación apropiada en la escala de la constante de información del error de las distribuciones más comunes. McCullagh y Nelder también recomiendan una gráfica de probabilidad normal de las devianzas residuales, otra vez la interpretación es exactamente análoga a la correspondiente en la regresión múltiple estándar.

Tabla 2.2: Transformaciones a información constante

Distribución	Transformación
Normal	$\hat{\mu}$
Binomial	$2 \times \text{sen}^{-1} \sqrt{\hat{\mu}}$
Poisson	$2 \times \sqrt{\hat{\mu}}$
Exponencial	$2 \times \log \hat{\mu}$
Gamma	$2 \times \log \hat{\mu}$

Algunas observaciones importantes de los modelos lineales generalizados son:

1. En la mayoría de los experimentos, cuando se analizan los datos y se utiliza una transformación, por lo general en una primera instancia se utilizan mínimos cuadrados ordinarios, para el ajuste del modelo en la escala transformada.
2. Cuando se reconoce varianza no constante se emplean mínimos cuadrados ponderados como la base de estimación de los parámetros.
3. Se sugiere que deberían llevar un análisis estándar que dependa de las transformaciones cuando un problema queda al final con varianza constante.
4. La devianza puede usarse para probar el ajuste del modelo saturado, la diferencia de esta medida entre el modelo completo y uno reducido se emplea para probar la hipótesis acerca de los subconjuntos de parámetros.
5. La inferencia de Wald se puede aplicar para probar las hipótesis y construcción de intervalos de confianza acerca de los parámetros individuales en el modelo.

2.6 Regresión Poisson

Para datos cuya naturaleza es de conteo, se emplean modelos con distribuciones asociadas Poisson o binomial negativa, la primera es la más simple para el ajuste de estos tipos de respuesta y parte de considerar una variable independiente y_i distribuida Poisson dado un vector de regresores \mathbf{x}_i , con densidad:

$$f[y_i|\mathbf{x}_i] = \frac{\mu_i^{y_i}}{y_i!} e^{-\mu_i}, \quad \text{para } y_i = 0, 1, 2, \dots$$

donde la media se expresa como:

$$\mu_i = e^{\mathbf{x}_i' \boldsymbol{\beta}}$$

$\boldsymbol{\beta}$ es un vector de parámetros de $k \times 1$. El proceso de conteo provee una motivación para elegir la distribución Poisson, porque la exponencial de $\mathbf{x}_i' \boldsymbol{\beta}$ garantiza que el parámetro μ_i es no negativo. Este modelo a su vez implica que la media condicional esta dada por

$$E[y_i|\mathbf{x}_i] = e^{\mathbf{x}_i' \boldsymbol{\beta}} \quad (2.11)$$

y tiende con frecuencia al cambio de la media condicional debido a los regresores, esto implica una forma particular de homoscedasticidad, bajo la igualdad de la varianza con la media condicional:

$$\text{Var}[y_i|\mathbf{x}_i] = e^{\mathbf{x}_i' \boldsymbol{\beta}}$$

2.6.1 Estimadores de máxima verosimilitud para la regresión Poisson

Ahora considere la función de log-verosimilitud de y_1, y_2, \dots, y_n , independiente de las observaciones y con el estimador estándar en el modelo:

$$L[y; \beta] = \ln \prod_{i=1}^n \left[\frac{\mu_i^{y_i}}{y_i!} e^{-\mu_i} \right] = \sum_{i=1}^n y_i \mathbf{x}_i' \beta - e^{\mathbf{x}_i' \beta} - \log y_i! \quad (2.12)$$

diferenciando la ecuación anterior con respecto a β se obtiene el estimador de máxima verosimilitud de $\hat{\beta}$ como la solución de la ecuación:

$$\sum_{i=1}^n [y_i - e^{\mathbf{x}_i' \beta}] x_i = 0 \quad (2.13)$$

se tienen k ecuaciones que no son lineales con k β 's desconocidas, que no cuentan con una solución cerrada para $\hat{\beta}$, por lo que se recurre al uso de métodos iterativos, como el de Newton-Rapson o de puntaje o "score" para calcular $\hat{\beta}$ [Wedderburn(1974)].

Otra consecuencia de que no exista una solución cerrada para $\hat{\beta}$ es que los resultados distribucionales exactos son muy difíciles de obtener, su inferencia se basa en resultados asintóticos. Hay varias maneras de proceder:

1. Se puede ver a $\hat{\beta}$ como el estimador que maximiza la ecuación (2.12) y se aplica la teoría de máxima verosimilitud.
2. $\hat{\beta}$ es definida en (2.13), entonces las ecuaciones tienen una interpretación similar a la de mínimos cuadrados ordinarios, donde el residual no ponderado ($y_i - \mu_i$) es ortogonal a los regresores, esto es posible para los mínimos cuadrados ordinarios y la inferencia puede ser desarrollada bajo supuestos sobre la media y posiblemente la varianza, que es una aproximación a un modelo lineal generalizado.
3. La ecuación (2.11) implica que $E[(y_i - e^{\mathbf{x}_i' \beta}) x_i] = 0$, entonces se puede definir un estimador que es la solución a la condición del correspondiente momento sobre la muestra, que es la solución de (2.13), esto es una aproximación a los modelos basados en momentos.

2.6.2 Devianza para el modelo Poisson

En el caso Poisson, el estimador máximo verosímil de μ_i para el modelo saturado es y_i , la i -ésima observación porque se asume que las n observaciones Poisson son independientes y así no está afectada por los regresores, la log-verosimilitud del modelo saturado es:

$$L[\mu; y] = - \sum_{i=1}^n y_i + \sum_{i=1}^n y_i \ln y_i - \sum_{i=1}^n \ln y_i!$$

para el modelo reducido se tiene que:

$$L[\beta; y] = - \sum_{i=1}^n \mu_i + \sum_{i=1}^n y_i \ln \mu_i - \sum_{i=1}^n \ln y_i!$$

donde μ_i es el cálculo de los estimadores de máxima verosimilitud de los coeficientes de regresión. Así, la devianza de β esta definida en la ecuación (2.9), $D(\beta)$ para el modelo Poisson esta dado por:

$$2 \left[- \sum_{i=1}^n (y_i - \hat{\mu}_i) + \sum_{i=1}^n y_i (\ln y_i - \ln \hat{\mu}_i) \right] = 2 \sum_{i=1}^n \left[y_i \ln \frac{y_i}{\hat{\mu}_i} - (y_i - \hat{\mu}_i) \right] \quad (2.14)$$

de esta ecuación se nota que cuando y_i se acerca a $\hat{\mu}_i$, este toma el ajuste de la regresión Poisson y la expresión para $D(\beta)$ se acerca a cero. Hay que recordar que la función de puntaje esta definida en la ecuación (2.4) y se puede ver que si un término constante aparece en el predictor lineal, entonces $\sum_{i=1}^n (y_i - \hat{\mu}_i) = 0$. Así, la expresión de la devianza en (2.14) simplemente se reduce a:

$$D(\beta) = 2 \sum_{i=1}^n y_i \ln \frac{y_i}{\hat{\mu}_i}$$

2.6.3 Sobredispersión

La igualdad de la media con la varianza se conoce como equidispersión, propiedad de la distribución Poisson que es frecuentemente violada en la vida real, la sobredispersión (o bajodispersión) significa que la varianza excede (o es menor que) la media. La falla de la Poisson con el supuesto de equidispersión tiene consecuencias cuantitativas similares al romperse el supuesto de homoscedasticidad en el modelo de regresión lineal [Frome(1983)], [Consul(1992)], [Frome(1973)].

Una indicación de las magnitudes de la sobredispersión o bajodispersión se puede obtener simplemente comparando la media y la varianza muestral de la variable de conteo dependiente. Este fenómeno puede surgir de diferentes maneras, por ejemplo, se observa un proceso Poisson fuera de un intervalo en el cual el largo es aleatorio más que fijo; alternativamente los datos podrían producirse por un proceso Poisson agrupado donde cada evento contribuye con una cantidad aleatoria al total. En otras palabras se observa que $Y = Z_1 + Z_2 + \dots + Z_N$ donde las Z 's son independientes e idénticamente distribuidas y N tiene una distribución Poisson independiente de Z . Se encuentra que:

$$E[Y] = E[N]E[Z]$$

y

$$Var[Y] = E[N]Var[Z] + Var[N]E^2[Z] = E[N]E^2[Z]$$

entonces hay sobredispersión si $E^2[Z] > E[Z]$.

Otra manera de sobredispersión puede surgir en estudios del comportamiento y de propensión de accidentes donde hay variabilidad entre cada sujeto, el número de incidentes Y para cada individuo dado sería Poisson con media Z , esta media se puede apreciar como una variable aleatoria que puede suponerse en la población y tiene una distribución gamma con media μ y escala $\phi\mu$; en otras palabras, $E[Z] = \mu$ y $Var[Z] = \mu/\phi$ imitando la distribución Poisson [McCullagh(1983)].

En las pruebas para sobredispersión, contempla la igualdad de la media y la varianza, impuesta por la distribución Poisson, como la hipótesis nula contra la alternativa de que la varianza excede al valor esperado. Estas son implementadas por pruebas de la Poisson con media y varianza μ_i , contra la binomial negativa con $E[y_i|\mathbf{x}_i] = \mu_i = \mu(\mathbf{x}_i, \beta)$ y $Var[y_i|\mathbf{x}_i] = \mu_i + \alpha g(\mu_i)$, para una función específica $g(\mu_i)$. Usualmente $g(\mu_i) = \mu_i$ ó $g(\mu_i) = \mu_i^2$. La hipótesis nula es:

$$H_0 : \alpha = 0 \quad vs. \quad H_a : \alpha > 0$$

Las pruebas de bajodispersión se construyen de manera similar:

$$H_0 : \alpha = 0 \quad vs. \quad H_a : \alpha < 0$$

2.7 Medidas Pseudo R^2

No hay una definición general de la R^2 en los modelos no lineales o un número de medida global que pueda ser propuesto, por eso se usa "pseudo" como calificativo. La pseudo R^2 usualmente tiene la propiedad de que coincide con la interpretación de la R^2 de una regresión estándar, sin embargo las características atractivas de R^2 en un modelo lineal desaparecen en los no lineales.

El punto de partida para obtener R^2 es la descomposición de la suma total de cuadrados como en el caso normal. Se tiene que:

$$\begin{aligned} \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 \\ &+ \sum_{i=1}^n (\hat{\mu}_i - \bar{y})^2 \\ &+ 2 \sum_{i=1}^n (y_i - \hat{\mu}_i)(\hat{\mu}_i - \bar{y}) \end{aligned}$$

El término izquierdo de la igualdad es conocido como la suma total de cuadrados (TSS), el primer sumando es la suma residual de cuadrados (RSS), el segundo es la suma explicada de cuadrados (ESS) y el tercero es cero para las estimaciones de mínimos cuadrados ordinarios para un modelo de regresión lineal.

Si el término constante en un modelo es cero, esto no implica que lo sea para todos los estimadores y modelos, incluyendo el Poisson; lo que conduce a las siguientes

2.8 Interpretación de los coeficientes

medidas: $R^2 = 1 - RSS/TSS$ ó $R^2 = ESS/TSS$, estas R^2 no necesariamente aumentan conforme se agreguen regresores e incluso si una intercepción se incluye, el primero puede ser negativo y el segundo puede exceder la unidad. Además, los estimadores Poisson calculados vía máxima verosimilitud no minimizan a RSS .

La devianza es la generalización de la suma de cuadrados en los modelos lineales generalizados, Cameron y Windmeijer ([Cameron(1996)], [Cameron(1997)]) proponen una pseudo R^2 basada en la descomposición de la devianza como:

$$D(y, \bar{y}) = D(y, \hat{\mu}) + D(\hat{\mu}, \bar{y}) \quad (2.15)$$

donde $D(y, \bar{y})$ es la devianza del modelo solamente en el intercepto, $D(y, \hat{\mu})$ corresponde a los residuales y $D(\hat{\mu}, \bar{y})$ a la explicada.

Si se divide la ecuación (2.15) entre $D(y, \bar{y})$ se obtiene:

$$R_{DEV}^2 = 1 - \frac{D(y, \hat{\mu})}{D(y, \bar{y})} \quad (2.16)$$

$$= \frac{D(\hat{\mu}, \bar{y})}{D(y, \bar{y})} \quad (2.17)$$

se observa una reducción en (2.16) con la inclusión de regresores, también la R^2 basada en términos de la devianza explicada (2.17) se sitúa entre 0 y 1 y se incrementa cuando se agregan regresores al modelo.

Para el modelo de regresión lineal Poisson la devianza esta dada por (2.14), y la pseudo R^2 queda como:

$$R_{DEV, Poisson}^2 = \frac{\sum_{i=1}^n \left[y_i \ln \frac{y_i}{\bar{y}} - (y_i - \hat{\mu}_i) \right]}{\sum_{i=1}^n y_i \ln \frac{y_i}{\bar{y}}}$$

donde $y \ln y = 0$ si $y = 0$.

2.8 Interpretación de los coeficientes

Es importante la interpretación de los coeficientes de regresión, se mencionan los diferentes sentidos según las características propias de las variables independientes.

En un modelo de regresión lineal

$$\hat{E}[y|\mathbf{x}] = \mathbf{x}'\hat{\beta} = \hat{\beta}_0 + \sum_{i=1}^n \hat{\beta}_i x_i$$

indica la media condicional de la respuesta que depende de los regresores, el coeficiente $\hat{\beta}_j = k$ significa una unidad de cambio en el j -ésimo regresor que incrementa la media de la respuesta k unidades, lo anterior se puede visualizar derivando con

respecto a la variable que corresponde al parámetro β_j , entendiendo que ese operador expresa una razón de cambio en la relación lineal $\frac{\partial}{\partial x_j} \hat{E}[y|\mathbf{x}] = \beta_j$.

En la regresión Poisson el sentido de los parámetros depende de la función liga (el logaritmo natural) y de la naturaleza de los regresores.

El caso más común es considerar la exponencial del valor medio de la respuesta:

$$E[y|\mathbf{x}] = e^{\mathbf{x}'\boldsymbol{\beta}} \quad (2.18)$$

diferenciando con respecto a x_j que denota el j -ésimo parámetro se tiene que

$$\frac{\partial}{\partial x_j} E[y|\mathbf{x}] = \beta_j e^{\mathbf{x}'\boldsymbol{\beta}} \quad (2.19)$$

esto expresa que si $\hat{\beta}_j = k$ y $e^{\mathbf{x}'\hat{\boldsymbol{\beta}}} = Q$ entonces una unidad de cambio en el j -ésimo regresor incrementa el valor esperado de y por $k \times Q$ unidades. Los valores difieren en cada individuo según los valores que toman las \mathbf{x}' s, como consecuencia esto hace que el análisis sea más complicado.

Una solución en la interpretación es agregar a todos los individuos y calcular el promedio de las respuestas:

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial x_{ij}} E[y_i|\mathbf{x}_i] = \frac{1}{n} \sum_{i=1}^n \beta_j e^{\mathbf{x}'_i \boldsymbol{\beta}}$$

la anterior expresión se simplifica en

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial x_{ij}} E[y_i|\mathbf{x}_i] = \beta_j \bar{y}$$

Una segunda opción es calcular la respuesta para un individuo con el promedio de las características:

$$\left. \frac{\partial}{\partial x_j} E[y|\mathbf{x}] \right|_{\mathbf{x}_j = \bar{\mathbf{x}}} = \beta_j e^{\bar{\mathbf{x}}'\boldsymbol{\beta}}$$

El promedio de la función exponencial evaluada en varios puntos excede al valor esperado de los mismos¹. Por lo tanto el promedio de las características proporciona respuestas más pequeñas que el promedio de las respuestas. Debido a la necesidad de menores cálculos, es común en la regresión no lineal reportar las respuestas a la media muestral de los regresores; esto es conceptualmente mejor, que reportar el promedio de la respuesta sobre todos los individuos.

Otro punto de vista para interpretar los parámetros, es primero calcular la ecuación (2.19) para valores seleccionados de x_j , este es quizás el mejor método, los valores reales de x_j de particular interés varía de aplicación en aplicación.

¹La desigualdad de Jensen expresa que el valor esperado de una función convexa es mayor o igual que la función evaluada en el valor esperado.

Es útil notar que la interpretación directa de los coeficientes es posible sin realizar cálculos adicionales de la siguiente manera:

El coeficiente β_j es igual al cambio proporcional sobre la media si el j -ésimo regresor cambia por una unidad. Lo que sigue es reescribir la ecuación (2.19) como $\partial E[y|\mathbf{x}]/\partial x_j = \beta_j E[y|\mathbf{x}]$ o como $\ln(E[y|\mathbf{x}]) = \mathbf{x}'\boldsymbol{\beta}$, usando la ecuación (2.18) y diferenciando con respecto a x_j se tiene:

$$\beta_j = \frac{\partial E[y|\mathbf{x}]}{\partial x_j} \times \frac{1}{E[y|\mathbf{x}]}$$

el signo de $\partial E[y|\mathbf{x}]/\partial x_j$ esta dado por el del escalar β_j .

Si un coeficiente de regresión es k veces más grande que otros, entonces el efecto de una unidad de cambio del regresor asociado es k múltiplo de otro, este resultado es:

$$\frac{\frac{\partial E[y|\mathbf{x}]}{\partial x_j}}{\frac{\partial E[y|\mathbf{x}]}{\partial x_i}} = \frac{\beta_j e^{\mathbf{x}'\boldsymbol{\beta}}}{\beta_i e^{\mathbf{x}'\boldsymbol{\beta}}} = \frac{\beta_j}{\beta_i} = k$$

En algunas ocasiones los regresores son transformados usando el logaritmo natural en la ecuación (2.18). Por ejemplo:

$$E[y|\mathbf{x}] = e^{\beta_1 \ln(x_1) + \mathbf{x}'_2 \boldsymbol{\beta}} = x_1^{\beta_1} e^{\mathbf{x}'_2 \boldsymbol{\beta}}$$

entonces β_1 es una elasticidad, dado por el porcentaje de cambio en $E[y|\mathbf{x}]$ para un $k\%$ de cambio en x_1 . Particularmente esta formulación es apropiada si x_1 es una medida de exposición.

Otro de los sentidos de los coeficientes del modelo de regresión Poisson es cuando se incluyen términos de interacción. Para el caso de una interacción se tiene que:

$$E[y|\mathbf{x}] = e^{\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_2 x_3}$$

entonces la variación proporcional en la media por unidad de cambio en x_3 es igual a $\beta_3 + \beta_4 x_2$, porque

$$\frac{\partial E[y|\mathbf{x}]}{\partial x_3} \times \frac{1}{E[y|\mathbf{x}]} = \beta_3 + \beta_4 x_2$$

esto es semejante a la medida de semielasticidad, porque el efecto de cambio en x_3 varía de acuerdo al valor de los regresores, en este caso de x_2 .

La interpretación de los parámetros anteriormente comentados son apropiados para regresores continuos.

Para una variable indicadora d empleada como regresor, que toma valores 0 ó 1; se tiene el siguiente modelo:

$$E[y|d, \mathbf{x}_2] = e^{\beta_1 d + \mathbf{x}'_2 \boldsymbol{\beta}}$$

entonces

$$\frac{E[y|d = 1, \mathbf{x}_2]}{E[y|d = 0, \mathbf{x}_2]} = \frac{e^{\beta_1 + \mathbf{x}_2' \beta}}{e^{\mathbf{x}_2' \beta}} = e^{\beta_1}$$

esta media condicional es e^{β_1} veces más grande que la variable indicadora cuando toma el valor cero.

2.9 Anexo: Identificación de parámetros de la familia exponencial

Para las distribuciones normal, Poisson y binomial se identifican los parámetros de la familia exponencial a través de los resultados de las ecuaciones (2.1), (2.2) y (2.3).

2.9.1 Distribución normal

La función de densidad de probabilidad de una variable aleatoria normal con parámetros μ y σ^2 esta dada por:

$$\begin{aligned} f(y; \mu, \sigma) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y - \mu)^2}{2\sigma^2}\right\} \\ &= \exp\left\{\frac{\left(y\mu - \frac{\mu^2}{2}\right)}{\sigma^2} - \frac{1}{2}\left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right]\right\} \end{aligned}$$

esta función de densidad es de la forma que se muestra en la ecuación (2.1) con:

$$\begin{aligned} \theta &= \mu \\ b(\theta) &= \frac{\mu^2}{2} \\ a(\phi) &= \phi \\ \phi &= \sigma^2 \\ c(y, \phi) &= -\frac{1}{2}\left\{\frac{y^2}{2} + \ln(2\pi\sigma^2)\right\} \end{aligned}$$

nótese que el parámetro de localización es μ y el de escala natural es σ^2 como se esperaba.

Ahora se puede calcular fácilmente la media y la varianza, considerando que $\theta = \mu$, $b(\theta) = \mu^2/2$, $a(\phi) = \sigma^2$, por lo tanto

$$\begin{aligned} E[y] &= b'(\theta) = \mu \\ Var[y] &= \frac{d^2b(\theta)}{d\theta^2} a(\phi) = \sigma^2 \end{aligned}$$

2.9.2 Distribución Poisson

La función de densidad de probabilidad Poisson con media y varianza μ esta dada por:

$$\begin{aligned} f(y; \mu) &= \frac{e^{-\mu} \mu^y}{y!} \\ &= \exp \{y \ln \mu - \mu - \ln y!\} \end{aligned}$$

nuevamente homologando con la ecuación (2.1):

$$\begin{aligned} \theta &= \ln \mu \\ b(\theta) &= e^\theta \\ \phi &= a(\phi) = 1 \\ c(y, \phi) &= -\ln y! \end{aligned}$$

los parámetros de localización y escala son μ y $\phi = 1$ respectivamente. Para identificar la varianza se usa la regla de la cadena:

$$\begin{aligned} E[y] &= \frac{db(\theta)}{d\theta} \\ &= \frac{db(\theta)}{d\mu} \frac{d\mu}{d\theta} \end{aligned}$$

además se sabe que $\frac{d\mu}{d\theta} = e^\theta = \mu$ es la media, $E[y] = 1 \times \mu = \mu$, y

$$\begin{aligned} Var[y] &= \frac{d\mu}{d\theta} \\ &= \frac{dE[y]}{d\theta} \\ &= \mu \end{aligned}$$

por lo tanto se tiene que la media y la varianza de la distribución Poisson coinciden.

2.9.3 Distribución binomial

La función de densidad de probabilidad es:

$$\begin{aligned} f(y; n, p) &= \binom{n}{y} p^y (1-p)^{n-y} \\ &= \exp \left\{ \ln \binom{n}{y} + y \ln \left(\frac{p}{1-p} \right) + n \ln (1-p) \right\} \end{aligned}$$

nuevamente se identifican las funciones que componen la ecuación general de la familia exponencial

$$\begin{aligned}\theta &= \ln\left(\frac{p}{1-p}\right) \\ b(\theta) &= n \ln(1-p) = n \ln(1 + e^\theta) \\ \phi &= a(\phi) = 1 \\ c(y, \phi) &= \ln\left(\frac{n}{y}\right)\end{aligned}$$

entonces la media y la varianza son

$$\begin{aligned}E[y] &= \frac{db(\theta)}{d\theta} = np \\ Var[y] &= \frac{d^2b(\theta)}{d\theta^2} = np(1-p)\end{aligned}$$

Capítulo 3

Análisis Estadístico

Los estados isquémicos de los principales órganos blandos del cuerpo, como el corazón y los pulmones, suelen presentarse en el hombre en los últimos años de su existencia; los factores meteorológicos como el frío y la humedad (solos o en combinación) son elementos naturales que influyen en la agudeza de este síntoma, sin embargo, en la actualidad los contaminantes del aire en las zonas urbanas son factores de influencia para que estas afecciones se presenten con mayor frecuencia y en un número mayor de personas. El estudio analiza la asociación entre la mortalidad diaria por problemas pulmonares y cardiovasculares, provocada por trastornos isquémicos en personas de 65 años y más en la Ciudad de México y área metropolitana, y los efectos que pueden causar los niveles diarios de contaminación y factores meteorológicos.

3.1 Análisis exploratorio de datos

Este análisis estima la asociación de los principales contaminantes del aire y la mortalidad diaria por isquemia (pulmonar y cardiovascular) vía regresión Poisson. La población en consideración son las personas fallecidas en 1994 en la Ciudad de México, cuya edad es 65 años o más; como información adicional se consideran otros factores de tipo meteorológico.

Las series de los conteos diarios de muertes por isquemia y datos meteorológicos fueron obtenidos del Instituto Nacional de Salud Pública, cuya sede se encuentra en Cuernavaca, Morelos. Los datos de contaminación del aire se obtuvieron de la Red Automática de Monitoreo Ambiental del Distrito Federal.

Las mediciones diarias de las concentraciones de los contaminantes y factores meteorológicos se obtienen para:

1. El ozono (O_3) y el bióxido de nitrógeno (NO_2), la medición máxima de una hora al día.
2. El bióxido de azufre (SO_2), el promedio móvil en 24 horas.

3. Las partículas fracción respirable (PM_{10}), el promedio de 24 horas.
4. El monóxido de carbono (CO), se realizan tres mediciones: el promedio y el promedio móvil de 8 horas al día y el máximo de una hora al día.
5. La temperatura, el valor mínimo diario (TMP) en grados centígrados y para la humedad relativa, el mínimo porcentaje del día (HRM); estos datos meteorológicos se consideran como factores de confusión en el modelo.

El resumen descriptivo de las variables del estudio se detalla en la Tabla 3.1

Tabla 3.1: Distribuciones de muertes por isquemia, contaminantes del aire y variables meteorológicas (Cd. de México, 1994)

Variable	Media	Min	Percentiles					Max
			10	25	50	75	90	
Isquemia	16.7	5	10	25	50	75	90	32
$O_3^{(PPM)}$	0.134	0.039	0.091	0.117	0.136	0.157	0.176	0.217
$NO_2^{(PPM)}$	0.038	0.019	0.026	0.031	0.037	0.045	0.052	0.073
$SO_2^{(PPM)}$	0.020	0.009	0.015	0.017	0.019	0.021	0.025	0.047
$CO^{(PPM)}$	0.340	0.17	0.245	0.278	0.327	0.390	0.454	0.618
CO Móvil	0.396	0.201	0.294	0.324	0.378	0.456	0.520	0.803
CO Max	0.634	0.272	0.425	0.503	0.600	0.740	0.890	13.620
$PM_{10}^{(\mu g/m^3)}$	75.01	23.36	42.98	58.58	72.42	88.31	106.18	175.29
$TPM^{(^\circ c)}$	10.79	1.45	7.88	9.38	11.05	12.38	13.21	15.76
$HRM^{(\%)}$	79.79	45.79	61.49	71.86	82.79	89.75	92.95	96.84

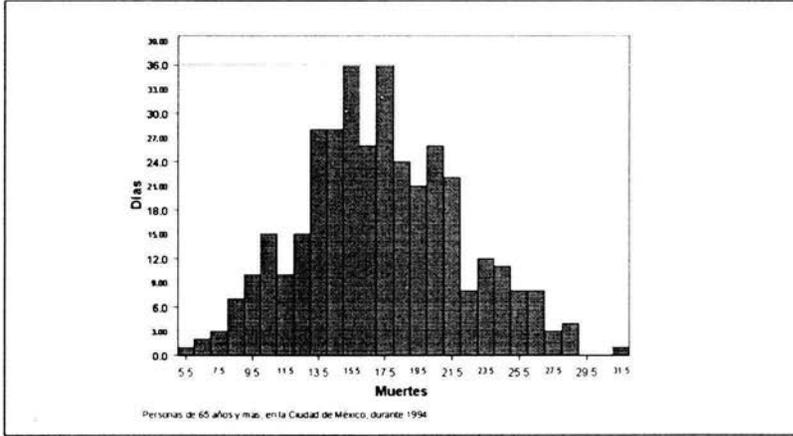
Los efectos de las variables explicativas pueden ser inmediatos o pueden tener algún patrón de retraso, es decir, que la variable respuesta se ve afectada por las variables explicativas de días anteriores. Los patrones de retraso pueden tener alguna forma funcional, la más simple es el promedio móvil, donde la contribución en la explicación de la variable respuesta depende de los efectos de varios días; para el análisis se calculan los retrasos simples de uno a siete días y las medias móviles de tres, cinco y siete días de los registros de contaminación ambiental y factores meteorológicos. Además se modificaron las variables independientes, en cada uno de los casos de las transformaciones anteriormente mencionadas en "dummies" por cuartiles en cada variable.

Tabla 3.2: Estadísticas de muertes por isquemia (Cd. de México, 1994).

Total	Media	Mediana	Moda	Rango	Varianza	Desviación estándar
6,110	16.7	17	15 y 17	27	22.64	4.76

Las muertes causadas por isquemia en 1994, en individuos de 65 años o más, sumaron 6,110, ocurriendo 5 casos como mínimo y 32 como máximo por día.

3.1 Análisis exploratorio de datos



Gráfica 3.1: Distribución de muertes por isquemia por día en personas de 65 años o más (Cd. de México, 1994).

La gráfica 3.1 se asemeja a una distribución normal, con un promedio cercano a 17 y una desviación estándar de 4.76, sin embargo, se presentan dos modas 15 y 17, en 36 días del año.

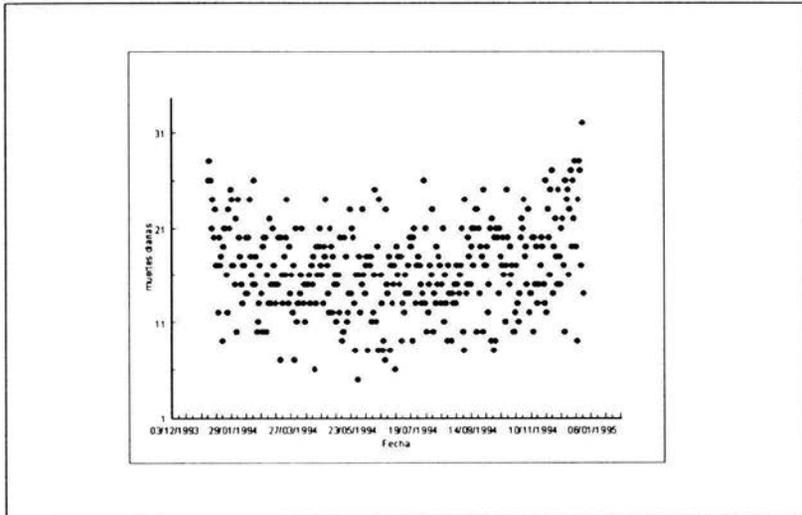
Las correlaciones de la variable respuesta con respecto a la contaminación y factores meteorológicos, se puntualizan en la tabla 3.3, al considerar los efectos inmediatos, no hay testimonios que alguna de estas variables sea un factor de riesgo, sin embargo, se podría mencionar que una sola variable o una combinación de ellas sea la causa directa de muerte por isquemia, aunque se busca presentar un modelo que dependa de los niveles diarios de contaminación y muestre una variación positiva sobre estas incidencias.

Tabla 3.3: Correlaciones de muertes por isquemia vs. niveles de contaminación del aire y variables meteorológicas (Cd. de México, 1994)

Variable	Correlación
O_3	0.028
NO_2	0.039
SO_2	0.064
CO	0.159
CO Max	0.138
CO Móvil	0.152
PM_{10}	0.179
TPM	-0.245
HRM	-0.096

El monóxido de carbono y las partículas fracción respirable son los contaminantes que presentan las correlaciones más altas de manera directa, los factores meteorológicos dan indicios de una correlación inversa con la respuesta.

La serie de incidencia diaria (ver gráfica 3.2) muestra las muertes por isquemia y presenta máximos en los primeros y últimos meses del año, que es en particular la temporada fría.



Gráfica 3.2: Serie: muertes por isquemia, personas de 65 años o más (Cd. de México, 1994).

Se tiene interés de conocer cómo influyen las variables independientes con la respuesta, para esto se revisaron las correlaciones de los contaminantes con la causa de muerte por isquemia como efecto inmediato.

Tabla 3.4: Estadísticas de contaminantes y factores meteorológicos

Contaminante	Media	Mediana	Varianza	Desviación estándar	Coficiente de variación
O_3	0.134	0.136	0.001	0.033	24.63%
NO_2	0.038	0.037	0.000	0.01	26.32%
SO_2	0.02	0.019	0.000	0.005	25.0%
CO	0.34	0.327	0.007	0.083	24.41%
PM_{10}	75.01	74.42	560.26	23.67	31.56%
TMP	10.79	11.05	4.77	2.18	20.2%
HRM	79.78	82.79	141.68	11.9	14.92%

3.1 Análisis exploratorio de datos

Tabla 3.5: Correlaciones de muertes por isquemia vs. variables de contaminación del aire y factores meteorológicos con retrasos simples y promedios móviles (Cd. de México, 1994).

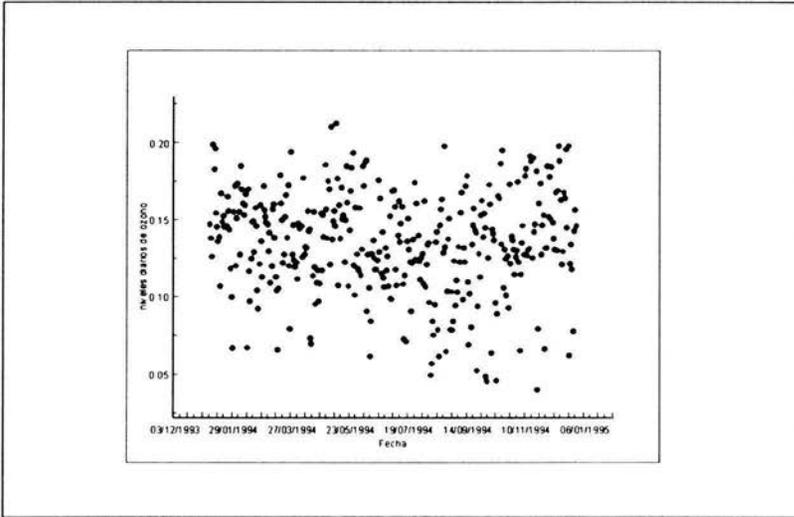
	Retrasos simples							Promedios móviles		
	1	2	3	4	5	6	7	3	5	7
Contaminación del aire										
<i>CO</i>	0.151	0.245	0.183	0.153	0.161	0.132	0.138	0.234	0.237	0.237
<i>CO_{mv}</i>	0.142	0.248	0.189	0.163	0.168	0.122	0.131	0.234	0.241	0.241
<i>CO_{mx}</i>	0.12	0.245	0.208	0.139	0.17	0.122	0.129	0.238	0.243	0.243
<i>NO₂</i>	0.041	0.081	0.037	0.044	0.038	-0.02	-0.023	0.061	0.06	0.06
<i>O₃</i>	0.083	0.049	0.08	0.075	0.042	-0.003	-0.003	0.101	0.115	0.115
<i>PM₁₀</i>	0.185	0.156	0.145	0.102	0.123	0.075	0.094	0.184	0.17	0.17
<i>SO₂</i>	0.005	-0.012	0.002	-0.031	-0.013	-0.046	-0.015	-0.002	-0.015	-0.015
Factores Meteorológicos										
<i>TMP</i>	-0.219	-0.232	-0.273	-0.284	-0.254	-0.262	-0.199	-0.269	-0.296	-0.296
<i>HRM</i>	-0.069	-0.061	-0.044	-0.014	-0.006	0.01	-0.005	-0.062	-0.043	-0.043

A continuación se describen cada una de los contaminantes y factores meteorológicos y se analizan las correlaciones con las muertes por isquemia utilizando retrasos simples y promedios móviles.

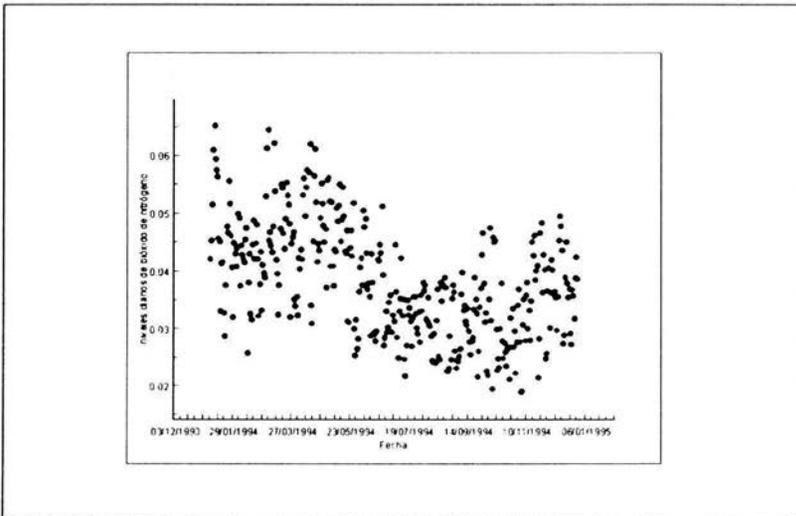
1. **Ozono (O_3).** La correlación que aporta más información sobre la respuesta es el promedio móvil de 5 días, aproximadamente 12% (ver tabla 3.5). Sus niveles diarios describen una alta variación con respecto a su promedio (0.135 *PPM*). El nivel de tolerancia para este contaminante en la Ciudad de México es 0.11 *PPM* una hora promedio máximo; sin embargo, el promedio de los datos observados en 1994 fue mayor a la norma establecida; en los registros se observó que 288 días (78.9%) no cumplieron con lo que se establece.
2. **Bióxido de nitrógeno (NO_2).** La correlación con la isquemia es casi nula, mostrándose inexistente una relación lineal con la variable respuesta. El promedio diario es de 0.038 *PPM* con una variación de 0.0001. La norma que se establece en el Distrito Federal es de 0.21 *PPM*, en 1994 no hubo días que la superaran. El nivel máximo registrado fue 0.07 *PPM* que ocurrió en los primeros días del año, en temporada fría. Hay que notar que en el segundo semestre los niveles decaen, esto puede explicarse por los periodos vacacionales (escolares y de fin de año) lo que supone un menor flujo vehicular. Se observa en la gráfica 3.4 que tiene un comportamiento contrario a lo largo del año con la humedad relativa mínima.
3. **Bióxido de azufre (SO_2).** La correlación con la isquemia es casi nula. Los niveles más altos se registraron a finales del mes de marzo, el promedio fue de 0.02 *PPM* muy por debajo de la norma que se establece en la Ciudad de México (0.13 *PPM*). La gráfica 3.5, muestra algunas observaciones, muy por

arriba del promedio y hacen que la dependencia en retrasos con ella misma no pase del 50% en alguna transformación de retraso.

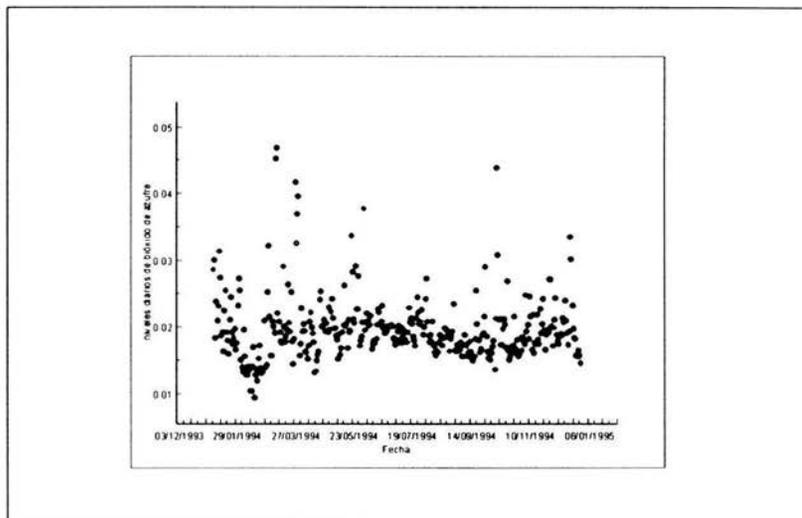
4. **Monóxido de carbono (CO)**. Tiene un patrón (gráfica 3.6) similar al de las muertes ocasionadas por isquemia: en el principio y el final del año se ubican los niveles más altos como en la variable dependiente. Sus correlaciones con respecto a las muertes por isquemia son mucho más altas que para los demás contaminantes (ver tabla 3.5), pero no indican una relación lineal alta con la respuesta; sin embargo, el factor de retraso de dos días sugiere (0.249) un efecto que puede ser decisivo en el desarrollo de los modelos. Sus concentraciones diarias son elevadas en comparación con el resto de los contaminantes. La norma establecida en el Distrito Federal es 0.11 PPM 8 horas promedio, mientras que la media se ubicó muy por arriba (0.34 PPM).
5. **Partículas fracción respirable (PM_{10})**. Sus niveles son altos a principios y a finales del año, mientras que los niveles más bajos son en agosto; el promedio anual es $75.01 \mu g/m^3$, la mitad del límite establecido en la Ciudad de México, en tres días de enero se presentaron registros mayores a la norma. La varianza de estos datos es alta, lo cual muestra la proporción relativa de la desviación estándar con respecto a la media con 31.56% de coeficiente de variación (tabla 3.4). Es el contaminante con la segunda correlación más alta (0.21) tomando un día de retraso.
6. **Temperatura mínima (TMP)**. Existe una relación lineal inversa con la respuesta, así también se observa en la tabla 3.5 que los efectos acumulativos son factores importantes de causa de muerte por isquemia, las correlaciones de esta variable transformada más relevantes son las que involucran los cuatro días de retraso y el efecto acumulado de cinco días de promedio móvil. Tiene un patrón inverso a los conteos de las muertes diarias (gráfica 3.8), a lo largo del año se reporta un promedio de $10.79 ^\circ C$, mostrando desviaciones de $2.2 ^\circ C$ y un coeficiente de variación de más del 20%.
7. **Humedad relativa mínima (HRM)**. Las correlaciones son prácticamente nulas con la respuesta, aunque la poca relación lineal que muestran operan de manera inversa, a juzgar este factor de confusión no aporta una evidencia lineal con la isquemia, distinto a la temperatura mínima (3.9). Tiene un patrón similar al de la temperatura mínima sólo que desfasado por tres meses, los resultados que describen esta serie están contenidos en las tablas 3.1 y 3.5, destaca que esta variable presenta un coeficiente de variación de 14.90%.



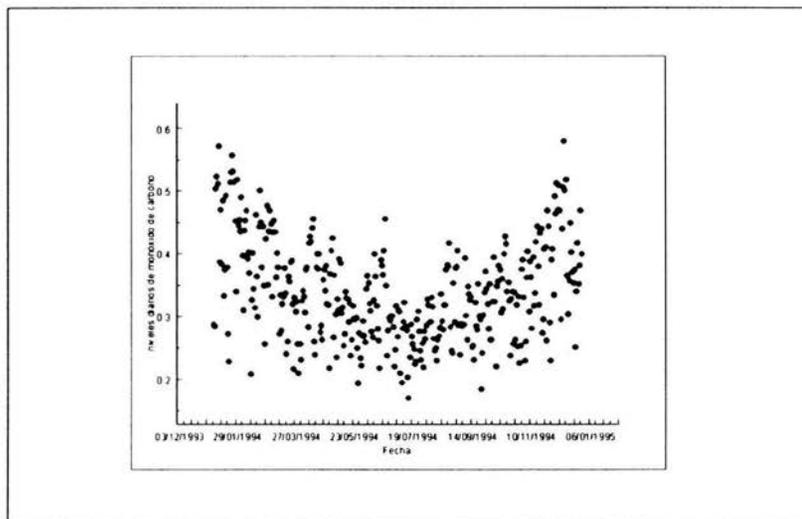
Gráfica 3.3: Ozono: serie promedio máximo una hora (Cd. de México, 1994).



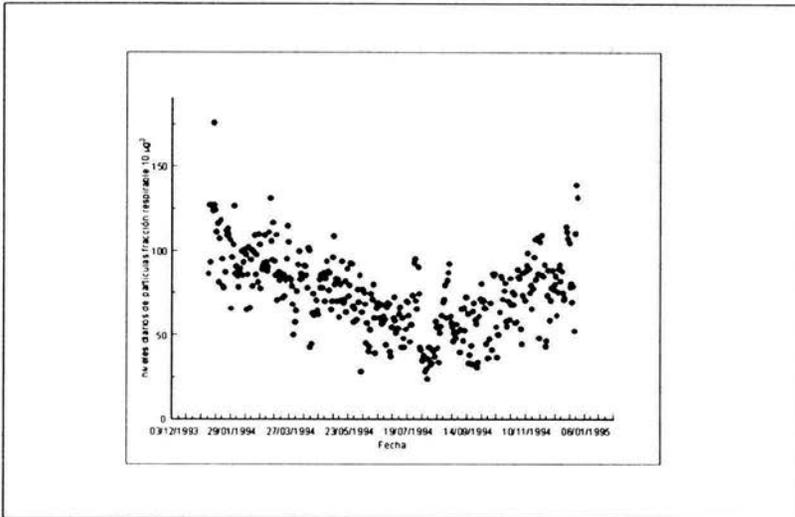
Gráfica 3.4: Bióxido de nitrógeno: serie promedio máximo una hora (Cd. de México, 1994).



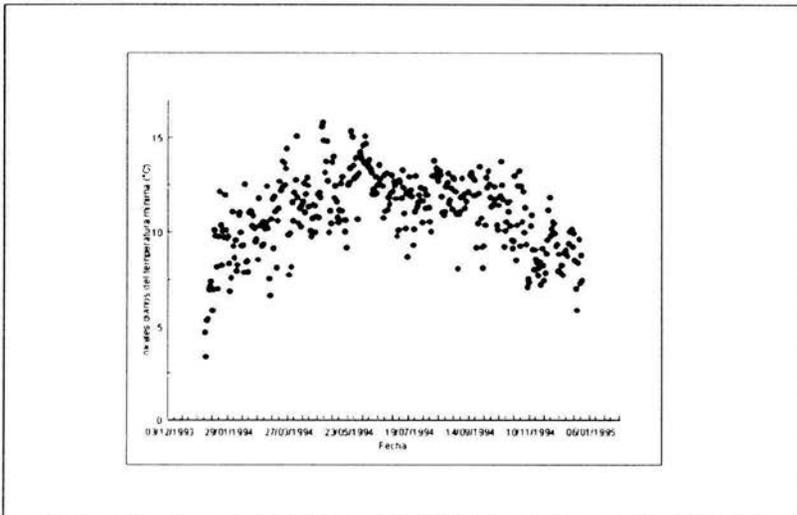
Gráfica 3.5: Bióxido de azufre: serie promedio móvil 24 horas (Cd. de México, 1994).



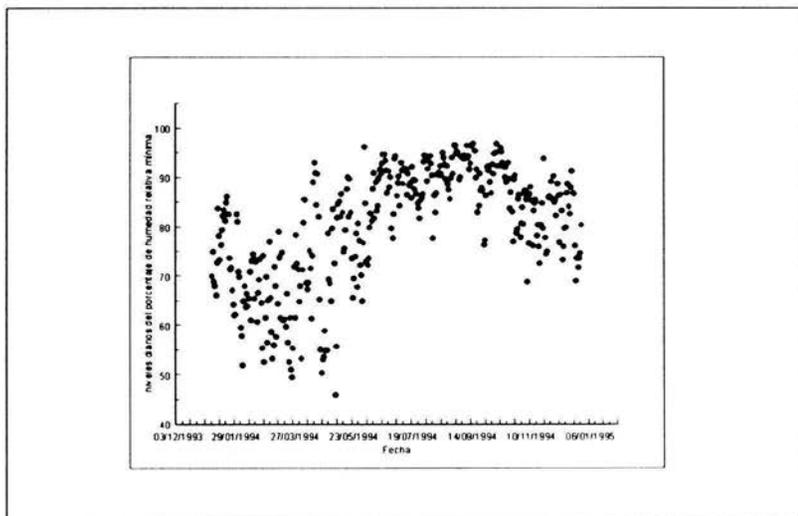
Gráfica 3.6: Monóxido de carbono: serie promedio móvil 8 horas (Cd. de México, 1994).



Gráfica 3.7: Partículas fracción respirable: serie promedio móvil 24 horas (Cd. de México, 1994).



Gráfica 3.8: Temperatura: serie valor mínimo diario (Cd. de México, 1994).



Gráfica 3.9: Humedad relativa: serie porcentaje mínimo diario (Cd. de México, 1994).

3.2 Construcción de modelos

El objetivo del estudio es modelar la variabilidad diaria del número de muertes causadas por isquemia como consecuencia de los cambios diarios en los niveles de contaminación ambiental y factores meteorológicos. Se pretende explicar esta relación, con el empleo de un modelo estadístico lineal que considera los datos de conteo tomando sólo valores enteros no negativos, y que además son pequeños con relación a la población en estudio.

Para el análisis, la regresión Poisson es una aproximación que cumple con la naturaleza de las características de la variable dependiente. Las variables independientes que buscan explicar su efecto sobre las muertes diarias por isquemia, con las diferentes combinaciones de los niveles de contaminación diaria (O_3 , SO_2 , NO_2 , PM_{10} y CO) y factores meteorológicos (temperatura mínima y humedad mínima relativa).

Se definen, la variable dependiente y como el número diario de muertes por isquemia y las covariables que la explican como:

$$\mathbf{x} = (x_{O_3}, x_{SO_2}, x_{NO_2}, x_{PM_{10}}, x_{CO}, x_{TMP}, x_{HRM})$$

El modelo supone que el logaritmo natural del número de muertes causadas por isquemia está relacionado linealmente con un conjunto de variables independientes.

el modelo básico se especifica de la siguiente forma:

$$\begin{aligned} \log(E[y|\mathbf{x}]) &= \beta_0 + \beta_1 x_{CO} + \beta_2 x_{NO_2} + \beta_3 x_{O_3} + \beta_4 x_{PM_{10}} \\ &\quad + \beta_5 x_{SO_2} + \beta_6 x_{HRM} + \beta_7 x_{TMP} \\ E[y|\mathbf{x}] &= e^{(\beta_0 + \beta_1 x_{CO} + \beta_2 x_{NO_2} + \beta_3 x_{O_3} + \beta_4 x_{PM_{10}} + \beta_5 x_{SO_2} + \beta_6 x_{HRM} + \beta_7 x_{TMP})} \end{aligned}$$

la segunda expresión puede ser vista como la media de la respuesta dadas las co-variables, donde β_0 representa la media general del modelo, x_i el nivel diario del contaminante o factor meteorológico i ; además las covariables son transformadas en indicadoras por cuartiles de los niveles de cada uno de los contaminantes incluyendo las estructuras de retraso simples y acumulativas de cada una. La elección del modelo se basa en el valor máximo de la log-verosimilitud o en la devianza más pequeña con respecto a la devianza del modelo saturado.

3.2.1 Modelos con efectos principales

La idea es incluir al principio los niveles de contaminación como un efecto directo de las muertes ocasionadas por isquemia y posteriormente añadir las variables meteorológicas como factores de confusión que inciden de manera constante sobre esta causa.

Para la selección de las variables candidatas a ser incluidas en el modelo, se utilizó un procedimiento "stepwise", que consiste en agregar o discriminar variables que disminuyen la devianza o incrementan la log-verosimilitud en el modelo.

La devianza nula del conjunto de datos es 482.8341 y a partir de este valor se compara el correspondiente al modelo propuesto. Como se menciona en la sección 2.6.2, el valor de la devianza debe disminuir con respecto a la nula, además se recurre a un porcentaje de ajuste que lo determina la pseudo R^2 de la sección 2.7.

A continuación se mencionan los modelos y pasos que se realizaron para lograr proporcionar el mejor ajuste de los datos:

Modelo 1 Para los cinco contaminantes se consideran las mediciones de los datos del día; de las mediciones de CO, se toma el máximo diario como la variable que proporciona la menor devianza en el modelo con 462.4978 y explica 4.21% de ajuste que es demasiado pobre (ver tabla 3.19).

Modelo 2 Agregando los factores meteorológicos la devianza es 453.9946 y se explica 5.97% (ver tabla 3.20).

- Ahora para determinar cuál es la variable que se relaciona más con la causa de muerte, se realiza un análisis de regresión Poisson para cada variable independiente y se observa su aportación en el modelo multivariado; de los

contaminantes, las partículas suspendidas (PM_{10}) son las que generan menor devianza pero el ajuste es muy bajo; considerando los factores meteorológicos la temperatura explica 4.21%, con la log-verosimilitud más alta (ver tabla 3.6).

Tabla 3.6: Regresión Poisson bivariada. Estadísticas de ajuste de contaminantes y factores meteorológicos.

Contaminante	Devianza	Log-Verosimilitud	Ajuste
CO	473.50	10783.4385	1.93%
CO_{mv}	475.55	10782.4122	1.51%
CO_{mx}	476.92	10781.7315	1.23%
NO_2	482.67	10778.8548	0.03%
O_3	482.30	10779.0401	0.11%
PM_{10}	472.38	10784.0013	2.17%
SO_2	480.00	10780.1914	0.59%
Factores meteorológicos			
TMP	462.51	10788.9328	4.21%
HRM	480.18	10780.1007	0.55%

- Se observa que modelar empleando los datos del día no es buena estrategia ya que los efectos ocasionados por los contaminantes no son inmediatos, por lo que se recurre a las estructuras de retraso, es decir, mediciones de días anteriores (retrasos simples) o promedios de las mismas durante un determinado número de días (promedios móviles).

Tabla 3.7: Regresión Poisson bivariada. Porcentaje de ajuste: retrasos simples y promedios móviles

	Retrasos simples								Promedios móviles		
	0	1	2	3	4	5	6	7	3	5	7
Contaminantes											
CO	1.9	2.2	5.8	3.2	2.3	2.5	1.7	1.9	5.3	5.4	5.3
CO_{mv}	1.5	2.0	5.9	3.4	2.6	2.7	1.4	1.7	5.3	5.6	5.3
CO_{mx}	1.2	1.4	5.8	4.2	1.9	2.5	1.5	1.6	5.5	5.7	5.5
NO_2	0.0	0.2	0.6	0.1	0.2	0.1	0.0	0.0	0.4	0.3	0.1
O_3	0.1	0.7	0.2	0.6	0.5	0.2	0.0	0.0	1.0	1.3	0.8
PM_{10}	2.2	3.3	2.4	2.0	1.0	1.5	0.5	0.9	3.3	2.8	2.4
SO_2	0.6	0.0	0.0	0.0	0.1	0.0	0.2	0.0	0.0	0.0	0.1
Factores Meteorológicos											
TMP	4.2	4.6	5.2	7.2	7.7	6.2	6.6	3.8	7.0	8.4	8.5
HRM	0.6	0.5	0.4	0.2	0.0	0.0	0.0	0.0	0.4	0.2	0.1

- Se realizan los ajustes bivariados (tabla 3.7), considerando las estructuras de retraso para encontrar la más adecuada a incluirse en el modelo, los porcentajes

de ajuste indican las siguientes variables:

- Medición máxima de monóxido de carbono: dos días de retraso.
- Bióxido de nitrógeno: dos días de retraso.
- Ozono: promedio móvil de cinco días.
- Partículas suspendidas: un día de retraso.
- Bióxido de azufre: sin retraso.
- Humedad relativa mínima: sin retraso.
- Temperatura mínima: promedio móvil de siete días.

Modelo 3 el ajuste considerando solamente contaminantes es de 8.9%, con una devianza de 440.0063;

Modelo 4 incluyendo los factores meteorológicos se tiene un ajuste del 11.7% y una devianza de 426.3935.

Tabla 3.8: Regresión Poisson bivariada. Porcentaje de ajuste: variables indicadoras de retrasos simples y promedios móviles.

	Retrasos simples							Promedios móviles			
	0	1	2	3	4	5	6	7	3	5	7
Contaminantes											
<i>CO</i>	2.0	1.4	5.0	4.0	1.9	1.6	0.9	2.4	3.8	4.1	3.6
<i>CO_{mv}</i>	1.6	1.1	4.0	3.8	1.9	1.5	1.3	1.2	3.4	3.8	3.3
<i>CO_{mx}</i>	1.6	0.6	5.0	6.0	1.2	2.2	1.4	1.6	3.6	3.4	3.9
<i>NO₂</i>	0.3	0.3	0.6	1.0	0.4	0.3	0.6	0.2	0.3	0.1	0.4
<i>O₃</i>	0.2	1.5	1.2	0.7	1.1	1.3	0.1	1.0	1.2	1.7	1.2
<i>PM₁₀</i>	2.1	2.4	2.4	1.0	0.8	1.6	0.5	0.8	3.1	2.8	1.8
<i>SO₂</i>	1.6	0.1	0.4	0.7	0.6	0.3	0.1	0.3	0.3	0.5	1.6
Factores Meteorológicos											
<i>TMP</i>	4.7	4.4	3.4	5.5	7.3	5.2	7.0	3.0	4.5	5.8	7.2
<i>HRM</i>	1.1	1.7	0.9	0.6	0.5	0.8	0.7	1.2	1.5	0.6	1.5

- Otra alternativa es formar categorías o clases de las variables independientes que incrementen la flexibilidad del modelo. Esto es, no tomarlas como un continuo, sino agruparlas por intervalos, respetando el orden de estas. Se forman categorías o clases dividiendo los rangos de las variables independientes (contaminantes) por cuartiles, considerando los retrasos simples y promedios móviles, para este caso se construyen variables indicadoras en donde para el primer cuartil corresponde el arreglo (1,0,0), el segundo (0,1,0) y el tercero (0,0,1), para la última categoría, el cuartil que indica los valores que caen entre

75 y 100 por ciento, por construcción queda determinado como (0, 0, 0), que se identifica como el grupo de control, en la tabla 3.8 se muestran los ajustes por contaminante de las variables indicadoras, estos sugieren las variables a incluir en el modelo:

- Medición máxima de monóxido de carbono: tres días de retraso.
 - Bióxido de nitrógeno: tres días de retraso.
 - Ozono: cinco días de retraso.
 - Partículas suspendidas: promedio móvil de cinco días.
 - Bióxido de azufre: sin retraso.
 - Humedad relativa mínima: seis días de retraso.
 - Temperatura mínima: seis días de retraso.
- En la tabla 3.9 se muestra el orden de inclusión de las variables y las ganancias en devianza y ajuste, esto es conocido como el método "forward", que esencialmente es ir agregando las variables que vayan reduciendo la mayor cantidad de devianza con relación a la nula; en consecuencia se obtienen:

Modelo 5 considerando solamente contaminantes el ajuste es de 12.95%, con una devianza de 420.3211;

Modelo 6 al incluir los factores meteorológicos se tiene un ajuste del 18.7% y una devianza de 392.5519.

Tabla 3.9: Regresión Poisson, método "forward", ajuste global, con variables indicadoras por cuartiles

Variable	Transformación	1. Devianza	Diferencia	2. Ajuste	Diferencia
<i>COmx</i>	3 días previos	454.0112	28.8229	5.97%	5.97%
<i>NO₂</i>	3 días previos	444.1745	9.8367	8.01%	2.04%
<i>O₃</i>	5 días previos	434.4773	9.6972	10.02%	2.01%
<i>PM₁₀</i>	media móvil 5 días	427.4387	7.0386	11.47%	1.46%
<i>SO₂</i>	día de la medición	420.3211	7.1176	12.95%	1.47%
<i>HRM</i>	6 días previos	408.5168	11.8043	15.39%	2.44%
<i>TMP</i>	6 días previos	392.5519	15.9649	18.70%	3.31%

Modelo 7 Finalmente, se encontró una variante que aumenta el ajuste al 19.11% y su devianza se reduce a 390.5677; empleando una transformación sobre la temperatura mínima, sin embargo, cambian tres variables del modelo 6. Las variaciones entre estos modelos y las variables a incluir se describen en la tabla 3.10.

Tabla 3.10: Comparación modelos 6 y 7. Regresión Poisson: transformación de corrección en la temperatura

Variable	Modelo 6	Modelo 7
$CO_{m,x}$		3 días previos
NO_2		3 días previos
O_3	5 días previos	2 días previos
PM_{10}		media móvil 5 días
SO_2	día de la medición	media móvil 7 días
HRM		6 días previos
TMP	6 días previos	$ln(\text{media móvil 5 días})$
Devianza	392.5519	390.5677
Ajuste	18.70%	19.11%

Las variables son indicadoras por cuartiles, con excepción del ln .

El **modelo 7** es el mejor ajuste que se obtiene utilizando solamente efectos principales.

3.2.2 Modelos con interacciones

Un modelo con interacciones, considera la inclusión de los efectos combinados de dos o más variables independientes. Se tomará el **modelo 7** incluyendo sólo interacciones dobles, sin considerar la temperatura ya que se presenta en el modelo como un porcentaje variable que afecta la estimación del valor esperado del resto. El modelo que incluye todos los pares de interacciones ajusta en un 47.8%; sin embargo resulta muy complicado en su formulación matemática por contar con 155 parámetros con poca significancia en su mayoría, se ejemplificarán algunos casos de interacciones dobles en la tabla 3.11.

Tabla 3.11: Regresión Poisson, diez mejores ajustes con una sola interacción doble

Interacción		Devianza	Parámetros	Ajuste
Variable 1	Variable 2			
$CO_{m,x}$	SO_2	370.4113	30	23.28%
NO_2	SO_2	375.0353	30	22.33%
NO_2	HRM	375.5277	30	22.22%
$CO_{m,x}$	PM_{10}	378.7825	30	21.55%
PM_{10}	SO_2	379.0201	30	21.50%
$CO_{m,x}$	O_3	379.8228	30	21.33%
$CO_{m,x}$	HRM	380.0808	30	21.28%
PM_{10}	HRM	381.1398	29	21.06%
Devianza Modelo 7		390.5677	17	19.11%
Devianza todas las interacciones		251.8230	155	47.8%

Lo interesante de incluir en los modelos combinaciones de pares de variables.

es mostrar cuales reducen más devianza, además, dan indicios de cuales son los contaminantes y niveles que en conjunto aportan mayor información para la causa de muerte. El resumen de la tabla 3.11 proporciona los mejores ajustes involucrando solamente un par de variables, por ejemplo, cuando interactúan el monóxido de carbono y el bióxido de azufre, la devianza es 370.4113 y predice linealmente 23.28%.

Desafortunadamente conforme se vayan agregando las interacciones de dos variables, el modelo pierde parsimonia, es decir, que el número de parámetros aumenta y el modelo se complica. En el anterior ejemplo la relación lineal tiene 30 parámetros, así agregando una interacción más, por ejemplo, *CO* con *HMR* el modelo se extiende a 36 parámetros. A juzgar de la parsimonia de los ajustes, es conveniente mencionar, que el análisis sólo proporciona una manera de describir las asociaciones combinadas de las variables independientes, como un efecto directo a la respuesta.

La idea es conservar las interacciones que son relevantes, para esto, se recurre al uso de las significancias de los parámetros en el modelo que contiene todas las interacciones de dos variables, utilizando el método de “backward” se discriminan los parámetros con mayor probabilidad anularse en el ajuste.

Modelo 8 El mejor modelo con interacciones que incorpora todos los efectos principales reduce la devianza a 313.8529 y tiene un ajuste de 34.87%, además 25 de los 54 parámetros son significativos al 5%.

Modelo 9 Se encontró un modelo que explica cerca del doble del modelo 7 y contempla sólo los efectos principales significativos, con una devianza de 315.5184 y ajusta 34.65%, mucho mejor en parsimonia (reduce 12 parámetros) y confiabilidad de los parámetros que el modelo 8, con una reducción marginal de ajuste y ligero aumento en la devianza. El modelo cuenta con 42 parámetros con una probabilidad de ser anulados no mayor al 11% y 36 de estos tienen una confianza mayor al 95%; el detalle de este modelo se ubica en las tablas 3.27 y 3.36.

3.3 Resultados

En esta sección se abordará la interpretación de los parámetros de regresión, qué cambia en cada modelo de acuerdo a las transformaciones que se aplicaron a las variables independientes, además, la liga que se emplea para una respuesta Poisson da un sentido distinto a los coeficientes, dado en los ajustes donde los errores son de tipo normal. Los primeros cinco modelos tienen interpretaciones similares, sin embargo, debido a el bajo ajuste que presentaron estos resultados se omiten, abordando de manera general su significado en las conclusiones de este trabajo.

En los **modelos 6 y 7**, se empleo una categorización de las variables independientes por cuartiles que mejora el porcentaje de ajuste en lugar del empleo de los

datos en forma continua, distinguiendo en el **modelo 7** la variable de la temperatura de manera continua y transformada en logaritmo, su significado es diferente al modelo 6. En los **modelos 8 y 9** se contempla el contraste de dos variables independientes que incrementan la relación lineal con la respuesta, la diferencia entre estos dos últimos modelos radica en la exclusión de las variables cuyos parámetros no cumplieron un cierto nivel de significancia.

El resultado se expresa en la escala de la respuesta con relación a la media del grupo de control, en los casos cuando el valor del coeficiente es mayor que cero la exponencial de este afecta a la media global como un porcentaje que la incrementa o disminuye cuando los parámetros son menores que cero, ahora, por la manera de cómo se construyeron la variables indicadoras:

- Se define que una relación es directa cuando la media general es la mayor y al afectarse por los coeficientes disminuye gradualmente, cuando disminuye el cuartil; viendo esto retrospectivamente, es decir, cuando aumenta el cuartil se van incrementando el número de muertes hasta llegar al promedio global y la relación es inversa cuando ocurre lo contrario.
- La relación es casi directa cuando la media general es la mayor y las muertes estimadas para los demás cuartiles son menores o iguales; y casi indirecta o inversa si ocurre lo contrario.

Notación Para referirse a las categorías de las variables se utiliza: *Variable seguido por Li* donde i indica los días de retraso o por *Pi* promedio móvil de i días **seguido por Ci** donde i indica el cuartil, por ejemplo: NO_2L3C2 indica el bióxido de nitrógeno con tres días de retraso en el cuartil 2.

Dos categorías separadas por “ * ” señala interacción entre ellas.

Dos o más categorías separadas por “ , ” indica la suma de efectos de las mismas respecto a otra.

Tabla 3.12: Muertes por Isquemia. Comparativo de estadísticas descriptivas: Modelos 6, 7 y 9

Modelo	Predicción			Estimaciones			Observados
	6	7	9	6	7	9	1994
Media	16.7	14.7	15.9	16.6	16.6	16.4	16.7
Mediana	16	16	16	17	17	17	17
Moda	16	16	20	17	17	19	15 y 17
Desv. Estándar	2.3	1.5	3.37	2.1	2.1	3.55	4.76
Coef. de variación	0.141	0.103	0.212	0.125	0.124	0.216	0.285
Máximo	28	35	34	23	25	25	32
Mínimo	11	11	9	12	12	9	5
Rango	17	24	25	11	13	14	27

La explicación de los modelos 6, 7 y 9 se detallarán a continuación.

3.3.1 Modelo 6: variables indicadoras como efectos principales y con transformaciones de retraso

Se revisan las estadísticas descriptivas de la predicción de todos los casos y las que corresponden a las estimaciones de los datos que se analizan; se aprecian diferencias mínimas en las medidas de tendencia central, no así en las de variación y los valores extremos de las estimaciones en los datos observados, principalmente esto se debe al pobre ajuste que existió en el modelo 6 (tabla 3.12).

La interpretación de los coeficientes es a partir del grupo de control, en el ajuste el promedio de muertes fue de 20 personas de 65 años o más en el año de 1994, que corresponde al último cuartil de los niveles de los contaminantes y factores meteorológicos, lo anterior se deduce a partir de la constante del modelo lineal de la regresión Poisson como $e^{3.0198} \cong 20$, así, los demás coeficientes funcionan como un factor de cambio porcentual del conteo de las muertes en cada cambio de categoría.

Los resultados que se pueden observar en el modelo (ver tabla 3.33):

- Se tiene en el grupo de variables indicadoras del monóxido de carbono con retrasos de tres días (CO_{mxL3}), una significancia estadística del 5% y tienen una relación casi directa, es decir, el ajuste pone en claro que cuando los registros de este contaminante son altos se incrementa la causa de muerte, la variación porcentual entre los valor más alto y el más bajo es de 19.1% que en términos de muertes es de 3 personas.
- El bióxido de nitrógeno con retrasos de tres días (NO_{2L3}), al contrario del CO muestra que a niveles bajos de NO_2 hay una ocurrencia mayor de muertes, sin embargo, hay coeficientes que en conjunto resultan tener poca significancia y la variación porcentual máxima es de 9.6%, alcanzando un máximo de dos individuos muertos con relación a la media general.
- Para los coeficientes del Ozono (O_3) con retraso de 5 días la relación es casi inversa, pero carecen de confiabilidad; la variación porcentual máxima es de 6.2%, esto significa \pm una muerte alrededor del promedio del grupo de control.
- Para el contaminante PM_{10} , partículas suspendidas, sólo los coeficientes de los dos primeros cuartiles presentan una confiabilidad máxima del 85%. Se puede considerar que existe una relación directa, ya que el coeficiente del tercer cuartil es teóricamente cero. La variación para este conjunto de variables indicadoras es 11.6%, que significa 2 muertes como máximo.
- Las mediciones de bióxido de azufre (SO_2) es el único contaminante que tiene efectos inmediatos sobre la causa de muerte, los parámetros tienen un nivel de

significancia de aproximadamente 7%, se observa una variación porcentual de 10.9%. Con una relación casi directa, con dos muertes de diferencia entre el mínimo y el máximo.

- El efecto que tiene la humedad relativa mínima sobre la muerte por isquemia, resulta tener una relación casi directa; cuando el porcentaje es bajo la causa de muerte se mantiene por debajo del promedio. La variación porcentual es cerca del 15%.
- El grupo de variables indicadoras de la temperatura mínima, es el que tiene el patrón mejor definido y muestra ser un factor importante sobre este tipo de muerte. Presenta una relación inversa, es decir, se observa que el valor de los parámetros decrece conforme se avanza en el cuartil, lo que provoca que a temperaturas muy bajas se incrementa la predicción de la respuesta.

Lo anterior, permite formar criterios sobre los días en donde la ocurrencia de muertes es alta, 28 casos; sujeto a los resultados de la regresión Poisson, este modelo indica que los niveles donde se alcanza el máximo de eventos es cuando:

- El CO rebasa los niveles de 0.740 PPM , límite inferior del cuartil del grupo de control de la medición máxima tres días previos.
- Las mediciones de NO_2 menores a 0.031 PPM que es el primer cuartil, con retrasos de tres días.
- Los niveles de O_3 son menores a 0.117 PPM que es el primer cuartil, con retrasos de cinco días.
- Cuando el promedio móvil de cinco días anteriores del PM_{10} es mayor a 72.43 $\mu g/m^3$, valor comprendido en el tercer y cuarto cuartil.
- La medición del día de SO_2 mayor a 0.021, que es el último cuartil.
- En los factores que no se pueden controlar, como la temperatura mínima y el porcentaje de humedad relativa, los niveles críticos son menores a 9.38 $^{\circ}C$ (primer cuartil) y 89.75% (grupo de control) respectivamente. Para estos factores de confusión se toman las mediciones seis días anteriores.

3.3.2 Modelo 7: variables indicadoras como efectos principales, con transformaciones de retraso y la temperatura como medida de elasticidad

Este modelo presenta algunas modificaciones al seis, esto se debe en esencia, a la modificación en las mediciones diarias de temperatura: se conservan los grupos de

variables indicadoras de CO , NO_2 y la humedad relativa con la configuración anterior de retrasos, cambiando el ozono, el bióxido de azufre y PM_{10} por las estructuras de dos días previos, el promedio móvil de siete días y la media móvil de tres días anteriores respectivamente. En el ajuste la ganancia es marginal (0.4%), pero el modelo 7 reduce el número de parámetros.

La temperatura ahora se incluye transformada en logaritmo natural de la media móvil de cinco días previos, el factor que la involucra es del siguiente estilo:

$$e^{(\beta_{TMP} \ln(TMPP5_i))} = TMPP5_i^{\beta_{TMP}}$$

así, $TMPP5_i$ es la temperatura del día i y β_{TMP} es el valor del coeficiente asociado a la variable en cuestión, además, este parámetro funciona como una medida de elasticidad en la temperatura.

El concepto de elasticidad tiene un origen puramente económico, pero se puede trasladar al problema y busca medir el impacto o el grado de variación sobre el número de muertes dadas las mediciones diarias de temperatura mínima, además, esta variable opera como una medida de exposición.

Para el cálculo de las estadísticas descriptivas de las predicciones se recurre al uso de algunas medidas básicas de la temperatura como el promedio ($10.8581^\circ C$) y el mínimo ($3.5782^\circ C$), el máximo de muertes se cálculo a partir del valor mínimo y para el resto de las estimaciones se empleo el promedio.

En la tabla 3.12 se muestran las estadísticas descriptivas de las estimaciones y predicciones del modelo 7; comparando los modelos 7 y 6: la predicción de la media es menor a los datos observados y estimados, que en el ajuste del modelo 6 eran muy similares; la variación se reduce a 1.5 contra 2.3, esto trae como consecuencia una disminución del coeficiente de variación; la amplitud del rango de muertes se incrementa 7 en las predicciones (24 contra 17) y 2 en las estimaciones utilizando los datos de 1994 (13 contra 11).

La media general del modelo se observa muy inflada, 46 ($\cong e^{3.8299}$) muertes; hay que considerar que la estimación siempre es acompañada por el factor que genera la temperatura mínima y representa una proporción de esta, por lo tanto la media variará de forma proporcional dependiendo del valor que tome el promedio móvil de la temperatura, esto se debe a que oscila en un rango continuo. Para el análisis se puede tomar el mínimo o el promedio de la temperatura, en este caso la mejor opción fue el promedio anual de las medias móviles de cinco días de la temperatura mínima ($10.8581^\circ C$).

La variación de temperatura mínima se refleja en los resultados de manera inversa, es decir, que incrementa las muertes cuando toma valores pequeños y las disminuye si los valores se acercan a la medición máxima. De las 46 muertes propuestas en la primera aproximación del modelo 7; la media general utilizando el mínimo de la temperatura estima 28 muertes (61.29%) y estima 16 (35.74%) utilizando el

valor máximo, si se utiliza el promedio de la temperatura se estima que 18 (40.03%) personas murieron por esta afección.

El análisis de los coeficientes de regresión de los resultados de las variables afectadas por la temperatura como un factor de exposición es el siguiente:

- El máximo del monóxido de carbono con tres días de retraso (CO_{mxL3}), tiene la mayor significancia estadística en el coeficiente del primer cuartil (8.1%), se observa que la relación es casi directa ya que a niveles altos se incrementa el valor esperado. La variación porcentual entre el valor máximo y el mínimo de muertes es 13.7% (aproximadamente 2 muertes).
- En el bióxido de nitrógeno con tres días de retraso (NO_2L3), sólo el primer cuartil presenta significancia estadística (7.1%), los coeficientes de los demás cuartiles presentan una alta probabilidad de ser anulados por lo que no sobresalen en el modelo; se observa una relación casi inversa. El valor máximo de muertes (20) se obtiene en el primer cuartil y el mínimo (18) para valores que se sitúan por arriba de la mediana, presentando una variación del 11.1%.
- Para el ozono con dos días de retraso (O_3L2), los coeficientes del segundo y cuarto cuartil presentan significancia estadística máxima de 9.7%, considerando este hecho y tomando en cuenta las anteriores categorías se puede presumir que esta variable tiene una relación directa, con una diferencia de un individuo entre los valores extremos (11.8%).
- Para las partículas suspendidas con el promedio móvil de tres días (PM_{10P3}), los coeficientes de las categorías situadas por debajo de la mediana carecen de ajuste, los dos últimos cuartiles proporcionan información relevante en el modelo, presentando una relación inversa, con 20 y 18 muertes respectivamente, variando en un 11.1%.
- En el bióxido de azufre con el promedio móvil de siete días (SO_2P7), la aportación es pobre debido a la confiabilidad de los coeficientes, la más alta es 82.1%, estimando 17 muertes, una menos que la media general, la variación es de aproximadamente de 6%.
- Para la humedad relativa mínima con seis días de retraso ($HRL6$), las estimaciones del segundo cuartil tienen una confianza de 75.5%, el resto presentan una significancia menor al 5%, este factor de confusión no muestra un patrón definido, mostrando una variación de 12.5%. Las muertes se incrementan cuando la humedad se ubica entre 45.79% y 71.48% o es mayor a 89.65%, estimando 18 defunciones.

En los resultados se distinguen los **rangos críticos** de los contaminantes y humedad relativa mínima que pronostican una mayor mortalidad por causa de isquemia:

- *CO*: valores mayores a 0.740 *PPM*, último cuartil, media general 18 muertes.
- *NO₂*: valores entre 0.019 y 0.031 *PPM*, primer cuartil, incrementan el número de decesos de la media general 9.53%.
- Ozono: valores entre 0.136 y 0.157 *PPM*, tercer cuartil, hay que aclarar que el coeficiente tiene una confianza de 55.6%, por lo que la aportación a la media general es aproximadamente 3% más de muertes.
- Partículas suspendidas: valores entre 74.97 y 87.55 $\mu\text{g}/\text{m}^3$, tercer cuartil, incrementan cerca de un 8% los desenlaces.
- Bióxido de azufre: valores entre 0.0196 y 0.0210 *PPM*, tercer cuartil, incrementan 2.11% las muertes. Este contaminante no es determinante para los fallecimientos debido a la carencia de significancia de los parámetros.
- Humedad relativa: valores mayores a 89.65%, último cuartil.

La combinación anterior predice un total de 57 acontecimientos fatales sin considerar el factor de la temperatura, al tomar en cuenta el valor mínimo se estiman cerca de 35 sucesos (61.29%) mientras que con el promedio anual se predicen 23 (40.03%).

A partir del modelo, observando los datos de 1994, los rangos de las variables explicativas donde se produjeron más muertes fueron: *CO_{ML}*, *NO₂*, *PM₁₀* y *SO₂* en el último cuartil; *O₃* y la humedad relativa en el primero. Sin embargo, la variable que provocó la mayor influencia fue la temperatura mínima, considerando el valor mínimo del promedio móvil de cinco días (3.58°C) se estimaron 25 muertes.

3.3.3 Modelo 9: variables indicadoras como efectos principales, con transformaciones de retraso, interacciones y la temperatura como medida de elasticidad

Los modelos 8 y 9 tienen un razonamiento muy similar, se considera en esta sección sólo el modelo 9 que incluye los efectos principales e interacciones con significancia máxima en los coeficientes (11%), el porcentaje de ajuste es 34.65%, observándose una diferencia de 0.21% con el modelo 8 que ajusta 34.87%.

La temperatura mínima de nuevo oscila en un intervalo continuo, que opera como un parámetro de variación proporcional, se emplea para el análisis el promedio (10.8581°C) y para estimar el máximo de muertes por isquemia se utiliza el mínimo (3.5782°C).

Con los modelos 6 y 7 existen diferencias en las predicciones. La varianza se acerca a la observada, al igual que el coeficiente de variación (6.9% menor al observado); los valores extremos revelan una diferencia notable entre el máximo y mínimo real, el rango aumenta uno en comparación al modelo 7; las estimaciones en los datos

observados aumentan el rango de 13 a 16, comparados con los datos observados aún hay una diferencia de 9 muertes (ver tabla 3.12).

El modelo tiene un promedio mayor que los ajustes anteriores $e^{3.9233} \simeq 51$ individuos muertos sin considerar el factor de la temperatura mínima, al incluirlo el número de muertes disminuye 39.19%, obteniendo cerca de 20 defunciones promedio y como máximo cerca de 31 (60.61%) decesos, nuevamente como en el modelo 7 este factor reduce el valor esperado general y afecta de manera directa al valor obtenido por las variables que fueron transformadas en categorías.

Este modelo contempla 42 parámetros:

- Cuatro de quince posibles efectos principales de los contaminantes.
 - El bióxido de azufre sugiere un incremento de la causa de muerte (10.46%).
 - Muestran una reducción en la respuesta: el ozono (6.68%), el bióxido de nitrógeno (10.66%) y monóxido de carbono (36.04%).
- Dos de la humedad relativa mínima.
- Treinta y cuatro que incluyen a los términos con interacción:
 - 18 que presentan incrementos en las muertes, entre 12.01% y 54.64%.
 - 16 que muestran reducciones de muertes, entre 10.07% y 35.30%.
- Media general (constante).
- El coeficiente asociado a la temperatura mínima.

En el modelo de interacciones, los coeficientes no tienen una interpretación práctica; esto se debe en esencia a que una interacción se basa en el efecto acumulado que ocasiona una categoría de algún contaminante con la acción conjunta de las otras variables. Dicho efecto se calcula derivando la estimación del modelo con respecto a la variable en cuestión, se suman los coeficientes que la involucran y la exponencial de la suma se multiplica por la constante de variación promedio que corresponde a la temperatura mínima; para obtener el total de muertes ocasionadas por los efectos acumulados dada una categoría de algún contaminante se multiplica el efecto por la media general.

El análisis del modelo contempla las seis variables categorizadas, cada una influye sobre el promedio general que estima el número de muertes en los niveles más altos, es decir, los valores del cuarto cuartil de las variables.

Evaluar por contaminante considerando sus interacciones con los otros, proporciona una idea de su aportación a la causa de muerte.

Monóxido de carbono

A continuación se presentan las interacciones de este contaminante por cuartil (ver tabla 3.13):

1. El primer cuartil ($CO_{mx}L3C1$) tiene interacciones con las siguientes variables¹ y se indican sus incrementos al efecto acumulado:

- (a) $HRL6C1$: 45.16%.
- (b) $PM_{10}P3C1$: 23.74%.
- (c) O_3L2C2 : 15.37%.
- (d) SO_2P7C1 : -15.3%.
- (e) NO_2L3C3 : -19.69%.

el cambio proporcional con estas interacciones tiene un incremento global² de 40.96%, que multiplicado por la media general y la variación del porcentaje promedio de la temperatura mínima estima 28 decesos.

2. El segundo cuartil ($CO_{mx}L3C2$)³ tiene interacciones con tres contaminantes cada uno con dos niveles, lo que genera 8 posibles efectos derivados, los casos críticos son los siguientes:

Caso uno: El cambio proporcional con las siguientes interacciones tiene un incremento global de 64.2%, es decir 33 muertos por isquemia.

- (a) NO_2L3C2 : 22.65%.
- (b) $PM_{10}P3C1$: 54.64%.
- (c) $HRL6C1$: 35.36%.

Caso dos: El cambio proporcional con las siguientes interacciones tiene un incremento global de 62.4%, es decir 32 muertos.

- (a) NO_2L3C1 : 21.34%.
- (b) $PM_{10}P3C1$: 54.64%.
- (c) $HRL6C1$: 35.36%.

3. El tercer cuartil ($CO_{mx}L3C3$) tiene interacciones con dos variables con dos niveles cada una, todas incrementan la causa de muerte con un individuo.

¹Ver notación en la sección 3.3

²Se obtiene multiplicando los incrementos, es decir $1.4516 \times 1.2374 \times 1.1537 \times (1 - .153) \times (1 - .1969)$

³En todos los casos de interacción se incluye el factor del efecto principal $CO_{mx}L3C2$: -36.04%.

Tabla 3.13: Muertes por isquemia, modelo 9: Interacciones del monóxido de carbono por cuartil

Grupos de variables	Suma de coeficientes	Variación porcentual	Muertes estimadas
<i>COmxL3C1</i>			
<i>NO₂L3C3, O₃L2C2, PM₁₀P3C1, SO₂P7C1, HRL6C1</i>	0.3433	40.96%	28
<i>COmxL3C2</i>			
<i>NO₂L3C1, PM₁₀P3C1, HRL6C1</i>	0.4852	62.4%	32
<i>NO₂L3C1, PM₁₀P3C1, HRL6C2</i>	0.3560	42.8%	28
<i>NO₂L3C1, PM₁₀P3C2, HRL6C1</i>	0.3046	35.6%	27
<i>NO₂L3C1, PM₁₀P3C2, HRL6C2</i>	0.1754	19.2%	24
<i>NO₂L3C2, PM₁₀P3C1, HRL6C1</i>	0.4960	64.2%	33
<i>NO₂L3C2, PM₁₀P3C1, HRL6C2</i>	0.3668	44.3%	29
<i>NO₂L3C2, PM₁₀P3C2, HRL6C1</i>	0.3154	37.1%	27
<i>NO₂L3C2, PM₁₀P3C2, HRL6C2</i>	0.1862	20.5%	24
<i>COmxL3C3</i>			
<i>NO₂L3C1, PM₁₀P3C2</i>	0.0439	4.5%	21
<i>NO₂L3C1, PM₁₀P3C3</i>	0.0733	7.6%	21
<i>NO₂L3C2, PM₁₀P3C2</i>	0.0504	5.2%	21
<i>NO₂L3C2, PM₁₀P3C3</i>	0.0798	8.3%	21

Bióxido de nitrógeno

A continuación se presentan las interacciones de este contaminante por cuartil (ver tabla 3.14):

1. El primer cuartil (*NO₂L3C1*), interactúa con dos contaminantes en dos niveles y con el tercer cuartil de la humedad relativa, generando cuatro combinaciones de efectos, sobresaliendo:

- (a) *COmxL3C3* : 33.71%.
- (b) *HRL6C3* : 12.01%.
- (c) *PM₁₀P3C2* : -11.61%.

que incrementa en 32.4% el promedio de muertes, prediciendo 26.

2. El segundo cuartil (*NO₂L3C2*)⁴ considera tres contaminantes con dos niveles cada uno, generando 8 (2³) interacciones, 3 son críticas:

Caso uno: El cambio proporcional con las siguientes interacciones tiene un incremento global de 38.20%, es decir 27 muertes.

- (a) *COmxL3C3* : 29.84%.

⁴En todos los casos de interacción se incluye el factor del efecto principal *NO₂L3C2* : -14.15%.

- (b) $PM_{10}P3C3$: 28.29%.
- (c) O_3L2C1 : 14.15%.
- (d) $HRL6C2$: -15.33%.

Caso dos: El cambio proporcional con las siguientes interacciones tiene un incremento global de 31.14%, es decir 26 muertes.

- (a) $COmxL3C3$: 29.84%.
- (b) $PM_{10}P3C3$: 28.29%.
- (c) O_3L2C1 : 14.15%.
- (d) $HRL6C1$: -19.65%.

Caso tres: El cambio proporcional con las siguientes interacciones tiene un incremento global de 30.55%, es decir 26 muertes.

- (a) $PM_{10}P3C3$: 28.29%.
- (b) $COmxL3C2$: 22.65%.
- (c) O_3L2C1 : 14.15%.
- (d) $HRL6C2$: -15.33%.

3. El tercer cuartil (NO_2L3C3) muestra una disminución al número de muertes de -44.85%.

Tabla 3.14: Muertes por isquemia, modelo 9: Interacciones del bióxido de nitrógeno por cuartil

Grupos de variables	Suma de coeficientes	Variación porcentual	Muertes estimadas
NO_2L3C1			
$COmxL3C2, PM_{10}P3C1, HRL6C3$	-0.1286	-12.1%	17
$COmxL3C2, PM_{10}P3C2, HRL6C3$	0.1834	20.1%	24
$COmxL3C3, PM_{10}P3C1, HRL6C3$	-0.0315	-3.1%	19
$COmxL3C3, PM_{10}P3C2, HRL6C3$	0.2805	32.4%	26
NO_2L3C2			
$COmxL3C2, HRL6C1, PM_{10}P3C1, O_3L2C1$	-0.3719	-31.06%	14
$COmxL3C2, HRL6C1, PM_{10}P3C3, O_3L2C1$	0.2142	23.89%	25
$COmxL3C2, HRL6C2, PM_{10}P3C1, O_3L2C1$	-0.3195	-27.35%	14
$COmxL3C2, HRL6C2, PM_{10}P3C3, O_3L2C1$	0.2666	30.55%	26
$COmxL3C3, HRL6C1, PM_{10}P3C1, O_3L2C1$	-0.3150	-27.02%	14
$COmxL3C3, HRL6C1, PM_{10}P3C3, O_3L2C1$	0.2711	31.14%	26
$COmxL3C3, HRL6C2, PM_{10}P3C1, O_3L2C1$	-0.2626	-23.20%	15
$COmxL3C3, HRL6C2, PM_{10}P3C3, O_3L2C1$	0.3235	38.20%	27
NO_2L3C3			
$COmxL3C1, PM_{10}P3C1, HRL6C1$	-0.5952	-44.85%	11

Ozono

Este contaminante presenta dos interacciones, a continuación se presentan por cuartil (ver tabla 3.15):

1. El primer cuartil (O_3L2C1) tiene interacciones con las siguientes variables:

- (a) NO_2L3C2 : 14.15%.
- (b) $PM_{10}P3C2$: -13.65%.

el cambio proporcional con estas interacciones tiene una disminución en la estimación de las muertes de -1.44%.

2. El segundo cuartil (O_3L2C2) tiene interacciones con las siguientes variables:

- (a) $COmxL3C1$: 15.37%.
- (b) $PM_{10}P3C2$: -13.65%.

el cambio proporcional con estas interacciones considerando el efecto principal O_3L2C1 : -6.68%, muestra una ligera baja de 0.67%, estimando 20 muertes.

Estas interacciones no modifican la media general por lo que se consideran irrelevantes.

Tabla 3.15: Muertes por isquemia, modelo 9: Interacciones del ozono por cuartil

Grupos de variables	Suma de coeficientes	Variación porcentual	Muertes estimadas
	O_3L2C1		
$NO_2L3C2, PM_{10}P3C2$	-0.0145	-1.44%	20
	O_3L2C2		
$COmxL3C1, PM_{10}P3C2$	-0.0067	-0.67%	20

Partículas suspendidas

A continuación se presentan las interacciones de este contaminante por cuartil (ver tabla 3.16):

1. El primer cuartil ($PM_{10}P3C1$), interactúa con tres contaminantes y con el segundo cuartil de la humedad relativa, generando seis combinaciones de efectos, sobresaliendo dos casos:

Caso uno: Con un incremento de 63.28% en el promedio de muertes, estimando 32:

- (a) NO_2L3C1 : -35.30%.

(b) $COmxL3C2$: 54.64%.

(c) SO_2P7C1 : 16.66%.

(d) $HRL6C2$: 18.54%.

Caso dos: Con un incremento de 52.67% en el promedio de muertes, estimando 30:

(a) NO_2L3C3 : -23.65%.

(b) $COmxL3C1$: 23.74%.

(c) SO_2P7C1 : 16.66%.

(d) $HRL6C2$: 18.54%.

2. El segundo cuartil ($PM_{10}P3C2$), interactúa con cuatro contaminantes, generando cuatro combinaciones de efectos, sobresaliendo dos casos:

Caso uno: Con un incremento de 30.06% en el promedio de muertes:

(a) $COmxL3C2$: 29.08%.

(b) O_3L2C1 : -13.65%.

(c) NO_2L3C1 : -11.61%.

(d) SO_2P7C1 : 32.01%.

Caso dos: Con un incremento de 29.68% en el promedio de muertes:

(a) $COmxL3C2$: 29.08%.

(b) O_3L2C2 : -13.90%.

(c) NO_2L3C1 : -11.61%.

(d) SO_2P7C1 : 32.01%.

En ambos casos estiman 26 muertes.

3. El tercer cuartil ($PM_{10}P3C3$), interactúa con dos contaminantes y la humedad relativa, cada uno con un cuartil, generando una combinación de efectos:

(a) $COmxL3C3$: -19.00%.

(b) NO_2L3C2 : 28.29%.

(c) $HRL6C1$: 16.33%.

el cambio proporcional de esta interacción tiene un incremento en la estimación de las muertes de 20.88%, estimando 24 defunciones.

Tabla 3.16: Muertes por isquemia, modelo 9: Interacciones de las partículas suspendidas por cuartil

Grupos de variables	Suma de coeficientes	Variación porcentual	Muertes estimadas
<i>PM₁₀P3C1</i>			
<i>NO₂L3C1, CO_mxL3C1, SO₂P7C1, HRL6C2</i>	0.1018	10.72%	22
<i>NO₂L3C1, CO_mxL3C2, SO₂P7C1, HRL6C2</i>	0.3247	38.36%	27
<i>NO₂L3C2, CO_mxL3C1, SO₂P7C1, HRL6C2</i>	0.2002	22.16%	24
<i>NO₂L3C2, CO_mxL3C2, SO₂P7C1, HRL6C2</i>	0.4231	52.67%	30
<i>NO₂L3C3, CO_mxL3C1, SO₂P7C1, HRL6C2</i>	0.2674	30.66%	26
<i>NO₂L3C3, CO_mxL3C2, SO₂P7C1, HRL6C2</i>	0.4903	63.28%	32
<i>PM₁₀P3C2</i>			
<i>CO_mxL3C2, O₃L2C1, NO₂L3C1, SO₂P7C1</i>	0.2628	30.06%	26
<i>CO_mxL3C2, O₃L2C2, NO₂L3C1, SO₂P7C1</i>	0.2599	29.68%	26
<i>CO_mxL3C3, O₃L2C1, NO₂L3C1, SO₂P7C1</i>	-0.2097	-18.92%	16
<i>CO_mxL3C3, O₃L2C2, NO₂L3C1, SO₂P7C1</i>	-0.2126	-19.15%	16
<i>PM₁₀P3C3</i>			
<i>CO_mxL3C3, NO₂L3C2, HRL6C1</i>	0.1897	20.88%	24

Bióxido de azufre

A continuación se presentan las interacciones de este contaminante por cuartil (ver tabla 3.17):

1. El primer cuartil (*SO₂P7C1*), interactúa con dos contaminantes y con el tercer cuartil de la humedad relativa, sobresaliendo:
 - (a) *CO_mxL3C1* : -15.30%.
 - (b) *PM₁₀P3C2* : 32.01%.
 - (c) *HRL6C3* : -17.47%.

se observa un decremento de -7.73% en el promedio de muertes, estimando 18, dos menos que la media general:

2. El segundo cuartil (*SO₂P7C2*), interactúa sólo con *HRL6C3*, se observa un descenso de -16.96% en el promedio de muertes, estimando 16.
3. El tercer cuartil (*SO₂P7C3*), interactúa sólo con *HRL6C3*, la aportación del efecto principal de 10.46% se observa un descenso global de -14.15% en el promedio de muertes, estimando 17.

Tabla 3.17: Muertes por isquemia, modelo 9: Interacciones del bióxido de azufre por cuartil

Grupos de variables	Suma de coeficientes	Variación porcentual	Muertes estimadas
<i>SO₂P7C1</i>			
<i>COmxL3C1, PM₁₀P3C1, HRL6C3</i>	-0.2040	-18.45%	16
<i>COmxL3C1, PM₁₀P3C2, HRL6C3</i>	-0.0804	-7.73%	18
<i>SO₂P7C2</i>			
<i>HRL6C3</i>	-0.1859	-16.96%	16
<i>SO₂P7C3</i>			
<i>HRL6C3</i>	-0.1677	-14.15%	17

Humedad relativa

A continuación se presentan las interacciones de este contaminante por cuartil (ver tabla 3.18):

1. El primer cuartil (*HRL6C1*)⁵, interactúa con tres contaminantes dos de ellos con dos niveles, generando cuatro combinaciones de efectos, sobresaliendo dos casos:

Caso uno: Con un incremento de 22.85% en el promedio de muertes, estimando 24:

- (a) *COmxL3C1* : 45.16%.
- (b) *NO₂L3C3* : -10.07%.
- (c) *PM₁₀P3C3* : 16.33%.

Caso dos: Con un incremento de 14.56% en el promedio de muertes, estimando 23:

- (a) *COmxL3C2* : 35.36%.
- (b) *NO₂L3C3* : -10.07%.
- (c) *PM₁₀P3C3* : 16.33%.

2. El segundo cuartil (*HRL6C2*)⁶, tiene interacciones con las siguientes variables:

- (a) *NO₂L3C2* : -15.33%.
- (b) *PM₁₀P3C1* : 18.64%.

el cambio proporcional con estas interacciones tiene una disminución en la estimación de las muertes de -10.33%, estimando 18.

⁵En todos los casos de interacción se incluye el factor del efecto principal *HRL6C1* : -19.11%.

⁶En todos los casos de interacción se incluye el factor del efecto principal *HRL6C2* : -10.66%.

3. El tercer cuartil (*HRL6C3*), tiene interacción con los tres cuartiles de la variable del SO_2 , y con el primer cuartil del NO_2 , estas interacciones se sitúa por debajo de la media general observando se como la más alta:

(a) SO_2P7C2 : -16.69%.

(b) NO_2L3C1 : 12.01%.

el cambio proporcional con estas interacciones tiene una disminución en la estimación de las muertes de -6.99%, estimando 18.

Tabla 3.18: Muertes por isquemia, modelo 9: Interacciones de la humedad relativa mínima por cuartil

Grupos de variables	Suma de coeficientes	Variación porcentual	Muertes estimadas
<i>HRL6C1</i>			
$CO_{mx}L3C1, NO_2L3C2, PM_{10}P3C3$	0.0931	9.76%	22
$CO_{mx}L3C1, NO_2L3C3, PM_{10}P3C3$	0.2058	22.85%	24
$CO_{mx}L3C2, NO_2L3C2, PM_{10}P3C3$	0.0232	2.35%	20
$CO_{mx}L3C2, NO_2L3C3, PM_{10}P3C3$	0.1359	14.56%	23
<i>HRL6C2</i>			
$NO_2L3C2, PM_{10}P3C1$	-0.109	-10.33%	18
<i>HRL6C3</i>			
SO_2P7C1, NO_2L3C1	-0.0786	-7.56%	18
SO_2P7C2, NO_2L3C1	-0.0725	-6.99%	18
SO_2P7C3, NO_2L3C1	-0.1538	-14.26%	17

El anterior análisis de efectos solamente constituye un punto de vista individual de contaminante y categoría, que describe la la gravedad de la combinación con uno o varios contaminantes y niveles sobre la estimación de las muertes por isquemia.

Criterios de Ajuste

Tabla 3.19: Modelo 1: Bondad del ajuste de efectos inmediatos, contaminantes.

Criterio	gl	Valor
Devianza	352	462.4978
Devianza Escalada	352	462.4978
Ji cuadrada de Pearson	352	451.7643
Ji cuadrada de Pearson Escalada	352	451.7643
Log verosimilitud		10788.9401
1-Pseudo R ²		0.04212

Tabla 3.20: Modelo 2: Bondad del ajuste de efectos inmediatos, contaminantes y factores meteorológicos.

Criterio	gl	Valor
Devianza	350	453.9946
Devianza Escalada	350	453.9946
Ji cuadrada de Pearson	350	441.7956
Ji cuadrada de Pearson Escalada	350	441.7956
Log verosimilitud		10793.1917
1-Pseudo R ²		0.05973

Tabla 3.21: Modelo 3: Bondad del ajuste de contaminantes transformados respecto al tiempo.

Criterio	gl	Valor
Devianza	352	440.0063
Devianza Escalada	352	440.0063
Ji cuadrada de Pearson	352	429.7836
Ji cuadrada de Pearson Escalada	352	429.7836
Log verosimilitud		10800.1859
1-Pseudo R ²		0.088701

Tabla 3.22: Modelo 4: Bondad del ajuste de contaminantes y factores meteorológicos transformados respecto al tiempo.

Criterio	gl	Valor
Devianza	350	426.3935
Devianza Escalada	350	426.3935
Ji cuadrada de Pearson	350	414.4962
Ji cuadrada de Pearson Escalada	350	414.4962
Log verosimilitud		10806.9922
1-Pseudo R ²		0.11689

Tabla 3.23: Modelo 5: Bondad del ajuste de variables indicadoras de contaminantes transformados respecto al tiempo.

Criterio	gl	Valor
Devianza	342	420.3211
Devianza Escalada	342	420.3211
Ji cuadrada de Pearson	342	409.2631
Ji cuadrada de Pearson Escalada	342	409.2631
Log verosimilitud		10810.0284
1-Pseudo R ²		0.1247

Tabla 3.24: Modelo 6: Bondad del ajuste de variables indicadoras de contaminantes y factores meteorológicos transformados respecto al tiempo.

Criterio	gl	Valor
Devianza	336	392.5519
Devianza Escalada	336	392.5519
Ji cuadrada de Pearson	336	381.7297
Ji cuadrada de Pearson Escalada	336	381.7297
Log verosimilitud		10823.9131
1-Pseudo R ²		0.1869

Tabla 3.25: Modelo 7: Bondad del ajuste de variables indicadoras de contaminantes y factores meteorológicos transformados respecto al tiempo: temperatura como factor de variación proporcional.

Criterio	gl	Valor
Devianza	338	390.5677
Devianza Escalada	338	390.5677
Ji cuadrada de Pearson	338	380.4231
Ji cuadrada de Pearson Escalada	338	380.4231
Log verosimilitud		10824.9051
1-Pseudo R ²		0.19109

Tabla 3.26: Modelo 8: Bondad del ajuste con todos los efectos principales e interacciones significativas.

Criterio	gl	Valor
Devianza	304	313.8529
Devianza Escalada	304	313.8529
Ji cuadrada de Pearson	304	310.4995
Ji cuadrada de Pearson Escalada	304	310.4995
Log verosimilitud		10863.2626
1-Pseudo R ²		0.34866

Tabla 3.27: Modelo 9: Bondad del ajuste de efectos principales e interacciones significativas.

Criterio	gl	Valor
Devianza	316	315.5184
Devianza Escalada	316	315.5184
Ji cuadrada de Pearson	316	311.6285
Ji cuadrada de Pearson Escalada	316	311.6285
Log verosimilitud		10862.4298
1-Pseudo R ²		0.34653

Análisis de los estimadores de los parámetros

Tabla 3.28: Modelo 1: Regresión Poisson, efectos inmediatos de contaminantes.

Parámetro	gl	Estimador β	Error Est.	χ^2	$Pr < \chi^2$	$\exp(\beta)$
Constante	1	2.6324	0.0759	1202.4	0.00	13.907
<i>COmx</i>	1	0.2275	0.1011	5.0656	0.02	1.2555
<i>NO₂</i>	1	-5.0240	1.9950	6.3418	0.01	0.0065
<i>O₃</i>	1	-0.1803	0.4765	0.1432	0.71	0.8350
<i>PM₁₀</i>	1	0.0020	0.0007	7.3900	0.01	1.0020
<i>SO₂</i>	1	5.1608	2.7777	3.452	0.06	174.3

Tabla 3.29: Modelo 2: Regresión Poisson, efectos inmediatos de contaminantes y factores meteorológicos.

Parámetro	gl	Estimador β	Error Est.	χ^2	$Pr < \chi^2$	$\exp(\beta)$
Constante	1	3.0853	0.1921	257.97	0.00	21.874
<i>COmx</i>	1	0.1726	0.1057	2.6672	0.10	1.1884
<i>NO₂</i>	1	-4.7070	2.2328	4.4442	0.04	0.0090
<i>O₃</i>	1	-0.1083	0.4893	0.0490	0.82	0.8974
<i>PM₁₀</i>	1	0.0009	0.0008	1.1759	0.28	1.0009
<i>SO₂</i>	1	4.4856	2.8404	2.4939	0.11	88.73
<i>TMP</i>	1	-0.0199	0.0073	7.4037	0.01	0.9803
<i>HR</i>	1	-0.0016	0.0014	1.3247	0.25	0.9984

Tabla 3.30: Modelo 3: Regresión Poisson, contaminantes transformados respecto al tiempo.

Parámetro	gl	Estimador β	Error Est.	χ^2	$Pr < \chi^2$	$\exp(\beta)$
Constante	1	2.4152	0.1061	518.44	0.00	11.192
<i>COmxL2</i>	1	0.9619	0.1917	25.181	0.00	2.6167
<i>NO₂L2</i>	1	-6.0357	1.8989	10.103	0.02	0.0024
<i>O₃P5</i>	1	0.4536	0.8182	0.3074	0.58	1.574
<i>PM₁₀L1</i>	1	0.0013	0.0007	3.2895	0.07	1.0013
<i>SO₂</i>	1	4.3075	2.6251	2.6925	0.10	74.255

Tabla 3.31: Modelo 4: Regresión Poisson, contaminantes y factores meteorológicos transformados respecto al tiempo.

Parámetro	gl	Estimador β	Error Est.	χ^2	$Pr < \chi^2$	$\exp(\beta)$
Constante	1	3.0396	0.2113	206.92	0.00	20.897
<i>COmxL2</i>	1	0.7806	0.2117	13.601	0.00	2.1828
<i>NO₂L2</i>	1	-4.9837	2.1540	5.3535	0.02	0.0068
<i>O₃P5</i>	1	0.1498	0.8273	0.0328	0.86	1.1616
<i>PM₁₀L1</i>	1	-0.0002	0.0008	0.0455	0.83	0.9998
<i>SO₂</i>	1	4.8243	2.6463	3.3236	0.07	124.5
<i>TMPP7</i>	1	-0.0013	0.0014	0.8841	0.35	0.9987
<i>HR</i>	1	-0.0322	0.0094	11.836	0.00	0.9683

Tabla 3.32: Modelo 5: Regresión Poisson, variables indicadoras de contaminantes transformadas respecto al tiempo.

Parámetro	gl	Estimador β	Error Est.	χ^2	$Pr < \chi^2$	$\exp(\beta)$
Constante	1	3.0259	0.0456	4397.2	0.00	20.613
<i>COmxL3C1</i>	1	-0.1382	0.0452	9.3290	0.00	0.87092
<i>COmxL3C2</i>	1	-0.1981	0.0409	23.483	0.00	0.8203
<i>COmxL3C3</i>	1	-0.1276	0.0380	11.283	0.00	0.8802
<i>NO₂L3C1</i>	1	0.1343	0.0464	8.3820	0.00	1.1437
<i>NO₂L3C2</i>	1	0.0695	0.0424	2.6915	0.10	1.072
<i>NO₂L3C3</i>	1	-0.0280	0.0384	0.5323	0.47	0.9724
<i>O₃L5C1</i>	1	0.0079	0.0387	0.0420	0.84	1.0079
<i>O₃L5C2</i>	1	-0.0501	0.0376	1.7765	0.18	0.9511
<i>O₃L5C3</i>	1	-0.0777	0.0368	4.4469	0.04	0.9252
<i>PM₁₀P5C1</i>	1	-0.1107	0.0471	5.5197	0.02	0.8952
<i>PM₁₀P5C2</i>	1	-0.0992	0.0407	5.9447	0.01	0.9056
<i>PM₁₀P5C3</i>	1	-0.0187	0.0377	0.2449	0.62	0.9815
<i>SO₂C1</i>	1	-0.0638	0.0372	2.9497	0.09	0.9382
<i>SO₂C2</i>	1	-0.0739	0.0382	3.7392	0.05	0.9288
<i>SO₂C3</i>	1	-0.1017	0.0372	7.4871	0.01	0.9033

Tabla 3.33: Modelo 6: Regresión Poisson, variables indicadoras de contaminantes y factores meteorológicos transformados respecto al tiempo.

Parámetro	gl	Estimador β	Error Est.	χ^2	$Pr < \chi^2$	$\exp(\beta)$
Constante	1	3.0198	0.0729	1715.1	0.00	20.487
<i>COmxL3C1</i>	1	-0.1089	0.0478	5.2017	0.02	0.8968
<i>COmxL3C2</i>	1	-0.1751	0.0419	17.462	0.00	0.8394
<i>COmxL3C3</i>	1	-0.1089	0.0387	7.9108	0.01	0.8968
<i>NO₂L3C1</i>	1	0.0919	0.0495	3.4507	0.06	1.0963
<i>NO₂L3C2</i>	1	0.0498	0.0444	1.2572	0.26	1.0511
<i>NO₂L3C3</i>	1	-0.0539	0.0390	1.9099	0.17	0.9475
<i>O₃L5C1</i>	1	0.0394	0.0396	0.9909	0.32	1.0402
<i>O₃L5C2</i>	1	-0.0289	0.0385	0.5636	0.45	0.9715
<i>O₃L5C3</i>	1	-0.0604	0.0373	2.6142	0.11	0.9414
<i>PM₁₀P5C1</i>	1	-0.1101	0.0616	3.1917	0.07	0.8957
<i>PM₁₀P5C2</i>	1	-0.0706	0.0485	2.1187	0.15	0.9318
<i>PM₁₀P5C3</i>	1	0.0010	0.0405	0.0006	0.98	1.001
<i>SO₂C1</i>	1	-0.0744	0.0376	3.9179	0.05	0.9283
<i>SO₂C2</i>	1	-0.0722	0.0387	3.4886	0.06	0.9303
<i>SO₂C3</i>	1	-0.1036	0.0374	7.6927	0.01	0.9016
<i>HRL6C1</i>	1	-0.1240	0.0522	5.6383	0.02	0.8834
<i>HRL6C2</i>	1	-0.0881	0.0462	3.6359	0.06	0.9157
<i>HRL6C3</i>	1	-0.1407	0.0430	10.723	0.00	0.8688
<i>TMPL6C1</i>	1	0.1660	0.0432	14.791	0.00	1.1806
<i>TMPL6C2</i>	1	0.0776	0.0397	3.8116	0.05	1.0807
<i>TMPL6C3</i>	1	0.0315	0.0393	0.6407	0.42	1.032

Tabla 3.34: Modelo 7: Regresión Poisson, variables indicadoras de contaminantes y factores meteorológicos transformados respecto al tiempo, temperatura como factor de variación proporcional.

Parámetro	gl	Estimador β	Error Est.	χ^2	$Pr < \chi^2$	$\exp(\beta)$
Constante	1	3.8299	0.1714	499.01	0.00	46.058
<i>COmxL3C1</i>	1	-0.0867	0.0497	3.0509	0.08	0.9170
<i>COmxL3C2</i>	1	-0.1477	0.0424	12.115	0.00	0.8627
<i>COmxL3C3</i>	1	-0.0986	0.0397	6.1756	0.01	0.9061
<i>NO₂L3C1</i>	1	0.0910	0.0506	3.2400	0.07	1.0953
<i>NO₂L3C2</i>	1	0.0360	0.0444	0.6572	0.42	1.0367
<i>NO₂L3C3</i>	1	-0.0274	0.0382	0.5159	0.47	0.9730
<i>O₃L2C1</i>	1	0.0054	0.0400	0.0180	0.89	1.0054
<i>O₃L2C2</i>	1	-0.0629	0.0379	2.7558	0.10	0.9390
<i>O₃L2C3</i>	1	0.0281	0.0368	0.5848	0.44	1.0285
<i>PM₁₀P3C1</i>	1	-0.0110	0.0584	0.0352	0.85	0.9891
<i>PM₁₀P3C2</i>	1	0.0077	0.0489	0.0251	0.87	1.0077
<i>PM₁₀P3C3</i>	1	0.0774	0.0411	3.5426	0.06	1.0805
<i>SO₂P7C1</i>	1	0.0119	0.0435	0.0748	0.78	1.012
<i>SO₂P7C2</i>	1	-0.0559	0.0416	1.8040	0.18	0.9456
<i>SO₂P7C3</i>	1	0.0209	0.0370	0.3184	0.57	1.0211
<i>HRL6C1</i>	1	-0.1007	0.0481	4.3767	0.04	0.9042
<i>HRL6C2</i>	1	-0.0491	0.0423	1.3516	0.25	0.9521
<i>HRL6C3</i>	1	-0.1165	0.0404	8.3027	0.00	0.89
$\ln(TMPP5)$	1	-0.3839	0.0757	25.691	0.00	0.6812

Tabla 3.35: Modelo 8: Regresión Poisson, con todos los efectos principales e interacciones significativas.

Parámetro	gl	Estimador β	Error Est.	χ^2	$Pr < \chi^2$	$\exp(\beta)$
Constante	1	3.9352	0.1851	451.74	0.00	51.172
<i>COmxL3C1</i>	1	0.0603	0.0867	0.4837	0.49	1.0622
<i>COmxL3C2</i>	1	-0.4599	0.0985	21.803	0.00	0.6314
<i>COmxL3C3</i>	1	-0.0156	0.0691	0.0508	0.82	0.9845
<i>NO₂L3C1</i>	1	-0.039	0.1033	0.1426	0.71	0.9618
<i>NO₂L3C2</i>	1	-0.1568	0.1012	2.4011	0.12	0.8549
<i>NO₂L3C3</i>	1	0.0192	0.0542	0.1259	0.72	1.0194
<i>O₃L2C1</i>	1	0.0046	0.0511	0.008	0.93	1.0046
<i>O₃L2C2</i>	1	-0.0639	0.0492	1.6866	0.19	0.9381
<i>O₃L2C3</i>	1	0.006	0.039	0.0237	0.88	1.006
<i>PM₁₀P3C1</i>	1	0.0391	0.1804	0.0469	0.83	1.0399
<i>PM₁₀P3C2</i>	1	-0.0599	0.0834	0.5146	0.47	0.9419
<i>PM₁₀P3C3</i>	1	-0.0183	0.0667	0.0754	0.78	0.9819
<i>SO₂P7C1</i>	1	0.032	0.0684	0.2193	0.64	1.0325
<i>SO₂P7C2</i>	1	0.0201	0.0522	0.1486	0.70	1.0203

continúa en la siguiente página

3.3 Resultados

continuación de la página anterior

Parámetro	gl	Estimador β	Error Est.	χ^2	$Pr < \chi^2$	$\exp(\beta)$
SO_2P7C3	1	0.1146	0.046	6.2129	0.01	1.1214
$HRL6C1$	1	-0.2247	0.0751	8.9427	0.00	0.7988
$HRL6C2$	1	-0.1284	0.0585	4.8105	0.03	0.8795
$HRL6C3$	1	-0.0218	0.0844	0.0666	0.80	0.9784
$\ln(TMPP5)$	1	-0.3971	0.0786	25.513	0.00	0.6723
$COmxL3C1 \cdot NO_2L3C3$	1	-0.2703	0.1242	4.7342	0.03	0.7632
$COmxL3C1 \cdot O_3L2C2$	1	0.1417	0.0708	4.0023	0.05	1.1522
$COmxL3C1 \cdot PM_{10}P3C1$	1	0.172	0.1212	2.0139	0.16	1.1877
$COmxL3C1 \cdot SO_2P7C1$	1	-0.1795	0.0844	4.5277	0.03	0.8357
$COmxL3C1 \cdot HRL6C1$	1	0.3867	0.1004	14.841	0.00	1.4721
$COmxL3C2 \cdot NO_2L3C1$	1	0.2342	0.1205	3.7751	0.05	1.2639
$COmxL3C2 \cdot NO_2L3C2$	1	0.2282	0.1121	4.1427	0.04	1.2563
$COmxL3C2 \cdot PM_{10}P3C1$	1	0.423	0.1302	10.553	0.00	1.5265
$COmxL3C2 \cdot PM_{10}P3C2$	1	0.2901	0.0991	8.5688	0.00	1.3366
$COmxL3C2 \cdot HRL6C1$	1	0.3259	0.1171	7.7436	0.01	1.3853
$COmxL3C2 \cdot HRL6C2$	1	0.1816	0.0833	4.7532	0.03	1.1991
$COmxL3C3 \cdot NO_2L3C1$	1	0.3219	0.1203	7.1598	0.01	1.3797
$COmxL3C3 \cdot NO_2L3C2$	1	0.2773	0.0997	7.7438	0.01	1.3196
$COmxL3C3 \cdot PM_{10}P3C2$	1	-0.1657	0.1033	2.573	0.11	0.8473
$COmxL3C3 \cdot PM_{10}P3C3$	1	-0.1861	0.0906	4.2211	0.04	0.8302
$NO_2L3C1 \cdot PM_{10}P3C1$	1	-0.4725	0.2084	5.1425	0.02	0.6234
$NO_2L3C1 \cdot PM_{10}P3C2$	1	-0.1097	0.0992	1.2209	0.27	0.8961
$NO_2L3C1 \cdot HRL6C3$	1	0.1297	0.08	2.6289	0.11	1.1385
$NO_2L3C2 \cdot O_3L2C1$	1	0.1309	0.0778	2.8261	0.09	1.1399
$NO_2L3C2 \cdot PM_{10}P3C1$	1	-0.3922	0.2021	3.7643	0.05	0.6756
$NO_2L3C2 \cdot PM_{10}P3C3$	1	0.2464	0.0916	7.2292	0.00	1.2794
$NO_2L3C2 \cdot HRL6C1$	1	-0.2463	0.1124	4.7988	0.03	0.7817
$NO_2L3C2 \cdot HRL6C2$	1	-0.1731	0.0852	4.1271	0.04	0.8411
$NO_2L3C3 \cdot PM_{10}P3C1$	1	-0.3151	0.2141	2.1665	0.14	0.7297
$NO_2L3C3 \cdot HRL6C1$	1	-0.134	0.0795	2.8426	0.09	0.8746
$O_3L2C1 \cdot PM_{10}P3C2$	1	-0.1378	0.086	2.5634	0.11	0.8713
$O_3L2C2 \cdot PM_{10}P3C2$	1	-0.1426	0.0777	3.3671	0.07	0.8671
$PM_{10}P3C1 \cdot SO_2P7C1$	1	0.1403	0.0904	2.4102	0.12	1.1506
$PM_{10}P3C1 \cdot HRL6C2$	1	0.1816	0.0852	4.5429	0.03	1.1991
$PM_{10}P3C2 \cdot SO_2P7C1$	1	0.272	0.0937	8.4282	0.00	1.3126
$PM_{10}P3C3 \cdot HRL6C1$	1	0.1543	0.0801	3.7118	0.05	1.1668
$SO_2P7C1 \cdot HRL6C3$	1	-0.1821	0.1201	2.3012	0.13	0.8335
$SO_2P7C2 \cdot HRL6C3$	1	-0.1849	0.0984	3.5356	0.06	0.8312
$SO_2P7C3 \cdot HRL6C3$	1	-0.2552	0.0903	7.983	0.00	0.7748

Tabla 3.36: Modelo 9: Regresión Poisson, con efectos principales e interacciones significativas.

Parámetro	gl	Estimador β	Error Est.	χ^2	$Pr < \chi^2$	$\exp(\beta)$
Constante	1	3.9233	0.1769	491.60	0	50.567
<i>COMxL3C2</i>	1	-0.4469	0.0951	22.105	0	0.6396
<i>NO₂L3C2</i>	1	-0.1526	0.0825	3.4195	0.06	0.8585
<i>O₃L2C2</i>	1	-0.0691	0.0429	2.595	0.11	0.9332
<i>SO₂P7C3</i>	1	0.0995	0.0341	8.5327	0	1.1046
<i>HRL6C1</i>	1	-0.2121	0.058	13.38	0	0.8089
<i>HRL6C2</i>	1	-0.1127	0.0521	4.6772	0.03	0.8934
$\ln(TMPP5)$	1	-0.3928	0.0748	27.601	0	0.6752
<i>COMxL3C1</i> • <i>NO₂L3C3</i>	1	-0.2193	0.0984	4.9632	0.03	0.8031
<i>COMxL3C1</i> • <i>O₃L2C2</i>	1	0.143	0.0661	4.6765	0.03	1.1537
<i>COMxL3C1</i> • <i>PM₁₀P3C1</i>	1	0.213	0.1044	4.1631	0.04	1.2374
<i>COMxL3C1</i> • <i>SO₂P7C1</i>	1	-0.1661	0.079	4.4208	0.04	0.8470
<i>COMxL3C1</i> • <i>HRL6C1</i>	1	0.3727	0.0905	16.9542	0	1.4516
<i>COMxL3C2</i> • <i>NO₂L3C1</i>	1	0.1934	0.0939	4.2401	0.04	1.2134
<i>COMxL3C2</i> • <i>NO₂L3C2</i>	1	0.2042	0.0989	4.2637	0.04	1.2265
<i>COMxL3C2</i> • <i>PM₁₀P3C1</i>	1	0.4359	0.1214	12.8986	0	1.5464
<i>COMxL3C2</i> • <i>PM₁₀P3C2</i>	1	0.2553	0.084	9.2504	0	1.2908
<i>COMxL3C2</i> • <i>HRL6C1</i>	1	0.3028	0.1134	7.1243	0.01	1.3536
<i>COMxL3C2</i> • <i>HRL6C2</i>	1	0.1736	0.0822	4.4639	0.03	1.1896
<i>COMxL3C3</i> • <i>NO₂L3C1</i>	1	0.2905	0.0951	9.327	0	1.3371
<i>COMxL3C3</i> • <i>NO₂L3C2</i>	1	0.2611	0.0838	9.6977	0	1.2984
<i>COMxL3C3</i> • <i>PM₁₀P3C2</i>	1	-0.2172	0.0634	11.7491	0	0.8048
<i>COMxL3C3</i> • <i>PM₁₀P3C3</i>	1	-0.2107	0.0654	10.395	0	0.81
<i>NO₂L3C1</i> • <i>PM₁₀P3C1</i>	1	-0.4354	0.1165	13.9638	0	0.647
<i>NO₂L3C1</i> • <i>PM₁₀P3C2</i>	1	-0.1234	0.0753	2.6839	0.1	0.8839
<i>NO₂L3C1</i> • <i>HRL6C3</i>	1	0.1134	0.0698	2.6411	0.1	1.1201
<i>NO₂L3C2</i> • <i>O₃L2C1</i>	1	0.1323	0.0699	3.5792	0.06	1.1415
<i>NO₂L3C2</i> • <i>PM₁₀P3C1</i>	1	-0.337	0.1054	10.22	0	0.7139
<i>NO₂L3C2</i> • <i>PM₁₀P3C3</i>	1	0.2491	0.0778	10.2389	0	1.2829
<i>NO₂L3C2</i> • <i>HRL6C1</i>	1	-0.2188	0.1058	4.2777	0.04	0.8035
<i>NO₂L3C2</i> • <i>HRL6C2</i>	1	-0.1664	0.0832	4.0055	0.05	0.8467
<i>NO₂L3C3</i> • <i>PM₁₀P3C1</i>	1	-0.2698	0.1249	4.6664	0.03	0.7635
<i>NO₂L3C3</i> • <i>HRL6C1</i>	1	-0.1061	0.0597	3.1597	0.08	0.8993
<i>O₃L2C1</i> • <i>PM₁₀P3C2</i>	1	-0.1468	0.0721	4.1453	0.04	0.8635
<i>O₃L2C2</i> • <i>PM₁₀P3C2</i>	1	-0.1497	0.0724	4.2752	0.04	0.8610
<i>PM₁₀P3C1</i> • <i>SO₂P7C1</i>	1	0.1541	0.0734	4.4069	0.04	1.1666
<i>PM₁₀P3C1</i> • <i>HRL6C2</i>	1	0.1701	0.0828	4.2259	0.04	1.1854
<i>PM₁₀P3C2</i> • <i>SO₂P7C1</i>	1	0.2777	0.0787	12.4648	0	1.3201
<i>PM₁₀P3C3</i> • <i>HRL6C1</i>	1	0.1513	0.0627	5.8181	0.02	1.1633
<i>SO₂P7C1</i> • <i>HRL6C3</i>	1	-0.192	0.0919	4.3663	0.04	0.8253
<i>SO₂P7C2</i> • <i>HRL6C3</i>	1	-0.1859	0.065	8.1937	0	0.8304
<i>SO₂P7C3</i> • <i>HRL6C3</i>	1	-0.2672	0.0595	20.1521	0	0.7655

Conclusiones

Se consideran dos hechos sobresalientes: los resultados de muerte que los modelos ajustan y las limitaciones que presentaron los datos.

Sobre los modelos

El estudio aborda la variación a través de los cuartiles de cada contaminante y factor meteorológico. Los estimadores de los parámetros, tanto en los modelos de efectos principales y de interacciones se obtuvieron por métodos de mínimos cuadrados iterativos ponderados. Se realizaron diferentes modelos considerando la inclusión o exclusión de distintas variables, dando interpretaciones marginales de los coeficientes de regresión por contaminante-cuartil, con la finalidad de establecer situaciones críticas que aporten evidencias para la determinación de un incremento sustancial en el número de muertes por isquemia en personas de 65 años o más.

Los resultados se basan en la construcción de cuartiles para cada una de las variables, tomando como grupo de control el último cuartil. Además, se parte del hecho de que los niveles máximos de contaminación y factores meteorológicos, pueden ser los topes para la exacerbación de muertes por isquemia.

Los resultados más importantes corresponden a los modelos seis, siete y nueve, donde se obtuvieron los mejores ajustes con distintos enfoques. A continuación se resumen tales ajustes:

Modelo seis Los datos ajustan 18.69%, con 14 parámetros significativos al 10% de los 22 posibles, incluyendo las variables: monóxido de carbono y bióxido de nitrógeno, ambos con tres días de retraso; ozono y partículas suspendidas con retrasos y promedio móvil de cinco días respectivamente; bióxido de azufre medición del día y los factores de confusión como la humedad relativa y temperatura mínimas con seis días de retraso. Los rangos críticos son: valores superiores a 0.740 *PPM* en *CO*, las mediciones menores a 0.031 *PPM* en *NO₂*, mediciones menores a 0.117 *PPM* en *O₃*, en las *PM₁₀* los datos que rebasan las 72.43 $\mu\text{g}/\text{m}^3$, para el *SO₂* mediciones mayores a 0.021 *PPM*, temperaturas menores a 9.38 °C y un porcentaje de humedad relativa mayor

a 89.75. Con esta combinación se propicia un incremento en las muertes de 34.62%, sobre un total de 20 en promedio y un estimado máximo de 28.

Modelo siete Presento mayor parsimonia que el seis, el principal cambio fue transformar la temperatura mínima con el logaritmo natural, que reduce el número de parámetros a 20, la mitad de ellos tienen un nivel de significancia máximo del 10%, aumentando el ajuste a 19.11%. Aunque el aumento de ajuste es marginal, la transformación proporciona más información que al incluirla por cuartiles, este coeficiente está dado como un porcentaje de cambio en la media $E[y]$, resultando viable debido a que la temperatura mínima es una medida de exposición para la población en general. En el modelo se conservaron los retrasos de las variables CO , NO_2 , HR y se incluyeron las mediciones de dos días previos del O_3 , el promedio móvil de tres y siete días para las PM_{10} y SO_2 respectivamente, y el logaritmo del promedio móvil de 5 días de la temperatura mínima.

Lo más relevante del ajuste se puede distinguir en los rangos donde se pronostica una mayor mortalidad: para el CO los valores mayores a 0.740 PPM , NO_2 valores entre 0.019 y 0.031 PPM , para el O_3 mediciones entre 0.136 y 0.157 PPM , en las PM_{10} valores entre 74.97 y 87.55 $\mu g/m^3$, el SO_2 no es una variable determinante para el incremento de muertes, en la humedad relativa valores mayores a 89.65%. El coeficiente -0.3839 asociado a la temperatura mínima funciona como una medida de elasticidad, que con el promedio anual de la variable reduce 40.03% la media general (57) que no considera la temperatura en 23 muertos y empleando el valor mínimo de la temperatura estima 35 decesos como máximo (61.29%).

Modelo nueve Se genera a partir del modelo siete, incluyendo interacciones y nuevamente la temperatura como factor de variación proporcional sobre la media global. Los datos ajustan 34.65%; sin embargo pierde parsimonia al considerar 42 parámetros con una confianza máxima de 89%. Las derivadas condicionales de contaminante-cuartil dan información sobre las aportaciones sustantivas en la causa de muerte.

Se tienen tres efectos combinados críticos: el primero y el segundo, los que genera el segundo cuartil del CO con NO_2L3C2 , $PM_{10}P3C1$, $HRL6C1$; y con NO_2L3C1 , $PM_{10}P3C1$, $HRL6C1$ con un efecto del 64.2% y 62.4% respectivamente; el tercero, las partículas suspendidas en el primer cuartil con $COMxL3C1$, NO_2L3C1 , $HRL6C2$ con un efecto del 63,28%. La primera combinación estima 33 muertes y las dos últimas 32. Esto se debe a que la mayoría de las interacciones tienen aportaciones positivas y pocas inhiben el efecto. Es útil realizar paso por paso esta composición de combinaciones que permite realizar juicios para determinar las interacciones con mayor peso.

Limitaciones y sugerencias

Los ajustes a la respuesta consideran que la exposición de la población a los niveles de contaminación y factores meteorológicos permanece constante, y no contemplan las siguientes situaciones:

- Las mediciones diarias no afectan a los individuos de la misma manera.
- No se considera el tiempo de exposición de la población, o tampoco si se encuentra al aire libre o en lugares cerrados.
- La variación de los niveles de contaminación dependiendo de la zona geográfica de la Ciudad de México.
- Se generalizan los efectos acumulativos y de retrasos, por lo que se robustece el tiempo de latencia en días; ya que en realidad cada variable no tiene la misma permanencia en el ambiente y sus efectos en la salud son distintos.
- El acceso a la atención médica y otras condiciones en general de los individuos.

Sin embargo, la estandarización de las variables flexibiliza el procedimiento de estimación del modelo, relajando los supuestos del clima, niveles de contaminación y sus interacciones así como las características socio-demográficas de la población.

Estudios realizados sobre salud y niveles diarios de contaminación, elaboran sus ajustes de series de datos con períodos de tiempo superiores a un año; esto se debe principalmente a que la mayoría de las variables muestran una variación sistemática con respecto al tiempo y puede ser asociada con simples análisis de correlaciones para cualesquiera dos variables, sin embargo esto no garantiza que se establezca una relación lineal ya que la tendencia en períodos de tiempo mayores a un año, por lo general no lo es, ya que la variación puede crecer o disminuir año con año, esto identifica como una tendencia.

Otra necesidad de emplear series de datos mayores a un año, es la búsqueda de componentes temporales, ya que la mayoría de las variables de salud, clima y contaminación tienen frecuentemente asociado un factor de temporalidad que proporciona una idea de cómo debe ser el ajuste del modelo; estas variaciones temporales se presentan a través de los años y determinan un patrón, lo que genera la posibilidad de asociaciones que agudizan el efecto en la respuesta; es necesario remover los patrones (largo y/o número de ondas) porque en algunas ocasiones se describen a intervalos determinados de tiempo y no pueden ser interpretados como senoidales, ya que cuando el período muestra un comportamiento singular, no es necesariamente senoidal o una función par simétrica, por ejemplo la existencia de períodos fríos largos seguidos por períodos cálidos cortos o períodos con ascensos bruscos y descensos suaves, o viceversa.

Conclusiones

Otro componente sistemático que puede sesgar las series de regresión es el que involucra categorías como días calendario específicos, días de la semana o períodos de vacaciones o de descanso, aunque generalmente estos patrones no son realmente necesarios, sin embargo tienen que revisarse, aunque no influyen directamente en la extensión del periodo de referencia.

Cabe hacer mención que existen varios métodos de adecuación para controlar el clima y la temporalidad, mientras se examinan las asociaciones entre la contaminación del aire y los conteos de mortalidad y morbilidad. En cada caso se requiere un manejo cuidadoso y buen juicio.

El análisis de datos de este trabajo está sujeto única y exclusivamente a 1994, sin considerar las limitaciones en los supuestos mencionados, ni los patrones de tendencia y temporalidad, por lo tanto se supone que pueden existir mejores ajustes al considerar series más grandes.

Apéndice A

Mínimos cuadrados iterativos ponderados

La mayoría de las estimaciones en los modelos lineales generalizados, requieren métodos iterativos para el cálculo de los parámetros. En general a la s -ésima iteración, el método de Newton-Rapson actualiza el último parámetro estimador $\hat{\beta}_s$, por la fórmula:

$$\hat{\beta}_{s+1} = \hat{\beta}_s + \hat{\mathbf{H}}_s^{-1} \hat{\mathbf{g}}_s$$

donde

$$\mathbf{g} = \frac{\partial \ln L}{\partial \beta}$$

y

$$\mathbf{H} = \frac{\partial^2 \ln L}{\partial \beta \partial \beta'}$$

siendo evaluadas en $\hat{\beta}_s$. Lo anterior lo convertimos en:

$$\hat{\beta}_{s+1} = \hat{\beta}_s + \left[\sum_{i=1}^n \hat{\mu}_{is} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \sum_{i=1}^n \mathbf{x}_i (y_i - \hat{\mu}_{is})$$

considere la función de la media como $\hat{\mu}_{is} = e^{\mathbf{x}_i' \hat{\beta}_s}$. Esto puede reescribirse como:

$$\hat{\beta}_{s+1} = \left[\sum_{i=1}^n \left(\sqrt{\hat{\mu}_{is}} \mathbf{x}_i \right) \left(\sqrt{\hat{\mu}_{is}} \mathbf{x}_i \right)' \right]^{-1} \sum_{i=1}^n \left(\sqrt{\hat{\mu}_{is}} \mathbf{x}_i \right) \left\{ \sqrt{\hat{\mu}_{is}} \frac{(y_i - \hat{\mu}_{is})}{\hat{\mu}_{is}} + \sqrt{\hat{\mu}_{is}} \mathbf{x}_i' \hat{\beta}_s \right\}$$

que es la expresión para el estimador de mínimos cuadrados ordinarios de la regresión:

$$\sqrt{\hat{\mu}_{is}} \left(\frac{(y_i - \hat{\mu}_{is})}{\hat{\mu}_{is}} + \mathbf{x}_i' \hat{\beta}_s \right) = \left(\sqrt{\hat{\mu}_{is}} \mathbf{x}_i \right)' \hat{\beta}_{s+1} + u_i$$

donde u_i es el error.

Así el estimador de máxima verosimilitud Poisson puede ser calculado por regresión de mínimos cuadrados ordinarios iterativos. Esto es equivalente y puede ser calculada por regresión de mínimos cuadrados ponderados de:

$$\frac{y_i - \hat{\mu}_{is}}{\hat{\mu}_{is}} + \mathbf{x}_i' \hat{\beta}_s \quad \text{en} \quad \mathbf{x}_i$$

donde los pesos $\sqrt{\hat{\mu}_{is}}$ cambian en cada iteración para la función de la media condicional general $\hat{\mu}_i = \mu(\mathbf{x}_i, \beta)$.

En el método de puntaje ("score"), se reemplaza \mathbf{H} por $E[\mathbf{H}]$, produciendo una regresión similar con variable dependiente:

$$\frac{(y_i - \hat{\mu}_{is}) + \left. \frac{\partial \mu_i}{\partial \beta'} \right|_{\hat{\beta}_s}}{\sqrt{\hat{\mu}_{is}}}$$

y el regresor:

$$\frac{\left. \frac{\partial \mu_i}{\partial \beta'} \right|_{\hat{\beta}_s}}{\sqrt{\hat{\mu}_{is}}}$$

Índice de Tablas

1.1	Criterios de la calidad del aire.	12
1.2	Índice Metropolitano de la Calidad del Aire (IMECA)	13
2.1	Ligas canónicas para los modelos lineales generalizados	20
2.2	Transformaciones a información constante	28
3.1	Distribuciones de muertes por isquemia, contaminantes del aire y variables meteorológicas	40
3.2	Estadísticas de muertes por isquemia	40
3.3	Correlaciones de muertes por isquemia vs. niveles de contaminación del aire y variables meteorológicas	41
3.4	Estadísticas de contaminantes y factores meteorológicos	42
3.5	Correlaciones de muertes por isquemia vs. variables con retrasos simples y promedios móviles	43
3.6	Regresión Poisson bivariada. Estadísticas de ajuste de contaminantes y factores meteorológicos.	50
3.7	Regresión Poisson bivariada. Porcentaje de ajuste: retrasos simples y promedios móviles	50
3.8	Regresión Poisson bivariada. Porcentaje de ajuste: variables indicadoras de retrasos simples y promedios móviles.	51
3.9	Regresión Poisson, método "forward", ajuste global, con variables indicadoras por cuartiles	52
3.10	Comparación modelos 6 y 7. Regresión Poisson: transformación de corrección en la temperatura	53
3.11	Regresión Poisson, diez mejores ajustes con una sola interacción doble	53
3.12	Muertes por Isquemia. Comparativo de estadísticas descriptivas: Modelos 6, 7 y 9	55
3.13	Modelo 9: Interacciones del monóxido de carbono por cuartil	63
3.14	Modelo 9: Interacciones del bióxido de nitrógeno por cuartil	64
3.15	Modelo 9: Interacciones del ozono por cuartil	65
3.16	Modelo 9: Interacciones de las partículas suspendidas por cuartil	67
3.17	Modelo 9: Interacciones del bióxido de azufre por cuartil	68

ÍNDICE DE TABLAS

3.18	Modelo 9: Interacciones de la humedad relativa mínima por cuartil	69
3.19	Modelo 1: Bondad del ajuste de efectos inmediatos, contaminantes.	70
3.20	Modelo 2: Bondad del ajuste de efectos inmediatos, contaminantes y factores meteorológicos.	70
3.21	Modelo 3: Bondad del ajuste de contaminantes transformados respecto al tiempo.	70
3.22	Modelo 4: Bondad del ajuste de contaminantes y factores meteorológicos transformados respecto al tiempo.	71
3.23	Modelo 5: Bondad del ajuste de variables indicadoras de contaminantes transformados respecto al tiempo.	71
3.24	Modelo 6: Bondad del ajuste de variables indicadoras de contaminantes y factores meteorológicos transformados respecto al tiempo.	71
3.25	Modelo 7: Bondad del ajuste de variables indicadoras de contaminantes y factores meteorológicos transformados respecto al tiempo; temperatura como factor de variación proporcional.	72
3.26	Modelo 8: Bondad del ajuste con todos los efectos principales e interacciones significativas.	72
3.27	Modelo 9: Bondad del ajuste de efectos principales e interacciones significativas.	72
3.28	Modelo 1: Regresión Poisson, efectos inmediatos de contaminantes.	73
3.29	Modelo 2: Regresión Poisson, efectos inmediatos de contaminantes y factores meteorológicos.	73
3.30	Modelo 3: Regresión Poisson, contaminantes transformados respecto al tiempo.	73
3.31	Modelo 4: Regresión Poisson, contaminantes y factores meteorológicos transformados respecto al tiempo.	74
3.32	Modelo 5: Regresión Poisson, variables indicadoras de contaminantes transformadas respecto al tiempo.	74
3.33	Modelo 6: Regresión Poisson, variables indicadoras de contaminantes y factores meteorológicos transformados respecto al tiempo.	75
3.34	Modelo 7: Regresión Poisson, variables indicadoras de contaminantes y factores meteorológicos transformados respecto al tiempo, temperatura como factor de variación proporcional.	76
3.35	Modelo 8: Regresión Poisson, con todos los efectos principales e interacciones significativas.	76
3.36	Modelo 9: Regresión Poisson, con efectos principales e interacciones significativas.	78

Índice de Figuras

3.1	Distribución de muertes por isquemia por día	41
3.2	Serie: muertes por isquemia	42
3.3	Ozono: serie promedio máximo una hora	45
3.4	Bióxido de nitrógeno: serie promedio máximo una hora	45
3.5	Bióxido de azufre: serie promedio móvil 24 horas	46
3.6	Monóxido de carbono: serie promedio móvil 8 horas	46
3.7	Partículas fracción respirable: serie promedio móvil 24 horas	47
3.8	Temperatura: serie valor mínimo diario	47
3.9	Humedad relativa: serie porcentaje mínimo diario	48

ÍNDICE DE FIGURAS

Bibliografía

- [Albert(1988)] Albert, James H. Computational Methods Using a Bayesian Hierarchical Generalized Linear Model. *Journal of the American Statistical Association*, 83.Theory and Methods(404):1037–1044, 1988.
- [Bachárová(1996)] Bachárová, L., Fandácová K., Bratinka M., et al. The Association between Air Pollution and Dially Number of Deaths: Findings from the Slovak Republic Contribution to the APHEA. *J Epidemiol Community Health*, 50 (Suppl 1):S19–S21, 1996.
- [Brant(1987)] Brant, Rollin. Residual Components in Generalized Linear Models *Can. J. Stat.*, 15(2):115–126, 1987.
- [Bravo(1987)] Bravo, H. *La Contaminación del Aire en México*. Fundación Universo XXI, México, D.F., 1987.
- [Breslow(1984)] Breslow, N. E. Extra-Poisson Variation in Log-Linear Models. *Appl. Statist.*, 33(1):38–44, 1984.
- [Burnett(1994)] Burnett, R. T., Krewski, D. Air Pollution Effects on Hospital Admission Rates: A Random Effects Modelling Approach *Can. J. Stat.*, 22:441–458, 1994.
- [Calderón(1987)] Calderón, E., Hernández, B. Crecimiento Actual de la Ciudad de México. *Ciencia y Desarrollo*, 76:49–66, 1987.
- [Cameron(1996)] Cameron, A. C., Windmeijer, F. A. R-Squared Measures for Count Data Regression Models with Applications to Health Care Utilization *Journal of Business and Economic Statistics*, 14:209–220, 1996.
- [Cameron(1997)] Cameron, A. C., Windmeijer, F. A. An R-Squared Measures of Goodness of Fit for some Common Nonlinear Regression Models *Journal of Econometrics*, 77:329–342, 1997.
- [Cameron(1998)] Cameron, Colin ; Trivedi, A.; Pravin, K *Regression Analysis of Count Data*. Econometric Society Monographs.Cambridge University Press. 1998.

BIBLIOGRAFÍA

- [Ciencias(1991a)] ¿Qué mide el IMECA? *Revista Ciencias*, 22:41-44, 1991a.
- [Consul(1992)] Consul, P. C., Famoye, Felix Generalized Poisson Regression Model. *Commun. Statist. -Theory Meth.*, 21(1):89-109, 1992.
- [Corey(1994)] Corey, P. N., Lou, W.Y., Broder, I. Analysis of the Relationship between Air Pollution and Asthma. *Can. J. Stat.*, 22:449-470, 1994.
- [Christiansen(Jun.,1997)] Christiansen, Cindy L., Morris, Carl N. Hierarchical Poisson Regression Modeling. *Journal of the American Statistical Association(Theory and Methods)*, Vol. 92:618-632, Jun., 1997.
- [Dab(1996)] Dab, W., Medina, S., Quénel, P., et al. Short Term Respiratory Health Effects of Ambient Air Pollution: Results of the APHEA Project in Paris *J Epidemiol Community Health*, 50 (Suppl 1):S42-S46, 1996.
- [DDF(1996)] Departamento del Distrito Federal, Gobierno del Estado de México, Secretaría del Medio Ambiente, Recursos Naturales y Pesca, Secretaría de Salud Programa para Mejorar la Calidad del Aire en el Valle de México 1995- 2000. 1996.
- [Dean(1992)] Dean, C. B. Testing for Overdispersion in Poisson and Binomial Models. *Journal of the American Statistical Association*, 87. Theory and Method (418):451-457, 1992.
- [Dobson(1991)] Dobson, J.A. *An Introduction to Generalized Linear Models*. Chapman and Hall, University of Newcastle New South Wales, Australia, 1991.
- [Estrada(1996)] Estrada Hernández, Mónica Aplicación de Modelos ARIMA al Ozono en la Zona Suroeste de la Ciudad de México. Tesina para obtener la Especialidad en Estadística, UACPyP, IIMAS, UNAM, 1996.
- [Everitt(1977)] Everitt, B. S. *The Analysis of Contingency Tables*. Chapman and Hall, Great Britain, 1977.
- [Ezcurra(1990)] Ezcurra, E. *De las Chinampas a la Megalopolis. El Medio Ambiente en la Cuenca de México*, volumen num 91 of *Colección la Ciencia desde México*. SEP/FCE/CONACYT, México, 1990.
- [Fahrnier(1985)] Fahrnier, Ludwig, Kaufman, Heinz. Consistency and Asymptotic Normality of the Maximum Likelihood in Generalized Linear Models. *The Annals of Statistics*, 13:342-368, 1985.
- [Fahrnier(1994)] Fahrnier, Ludwig; Tutz, Gerhard. *Multivariate Statistical Modelling based on Generalized Linear Models*. Springer-Verlag, New York, 1994.

- [Famoye(1993)] Famoye, Felix. Restricted Generalized Poisson Regression Model. *Commun. Statist. -Theory Meth.*, 22(5):1335-1354, 1993.
- [Freeman(1987)] Freeman, Daniel H. Jr. *Applied Categorical Data Analysis*, Vol. 79 of *STATISTICS: Textbooks and Monographs*. Marcel Dekker, Inc., New York, 1987.
- [Frome(1973)] Frome, Edward L.; Kutner, Michael H.; Beauchamp, John J. Regression Analysis of Poisson-Distributed Data. *Journal of the American Statistical Association*, 68:935-942, 1973.
- [Frome(1983)] Frome, E.L. The Analysis of Rates using Poisson Regression Models. *Biometrics*, 39:665-674, 1983.
- [Frome(Aug., 1989)] Frome, Edward L.; Morris, Max D. Evaluating Goodness of Fit of Poisson Regression Models in Cohort Studies. *The American Statistician*, Vol. 43. N°3:144-147, Aug., 1989.
- [Fuentes(1991)] Fuentes, G. V. La Contaminación por Partículas Suspendidas en la Atmósfera del Valle de México. *Revista Ciencias*, 22:45-49, 1991.
- [Ilabaca(1996)] Ilabaca Marileo, Mauricio Alfonso Relación entre la Contaminación Atmosférica y las Consultas por Emergencias Respiratorias Pediátricas, en el Servicio de Urgencia del Hospital Luis Calvo Mackenna, en Santiago de Chile. Tesis de Maestría, Escuela de Salud Pública de México, Instituto de Salud Pública, 1996.
- [Jáuregui(1988)] Jáuregui E. Efectos del Clima Urbano sobre los Niveles de Contaminantes en la Ciudad de México. *Geográfica y Desarrollo*, 1:37-34, 1988.
- [Katsouyanni(1996)] Katsouyanni K, Schwartz J, Spix C, Touloumi G, et al. Short Term Effects of Air Pollution on Health: A European Approach using Epidemiologic Time Series Data: The APHEA Protocol. *J Epidemiol Community Health*, 50 (Suppl 1):S12-S18, 1996.
- [Lawless(1987)] Lawless, Jerarld F. Negative Binomial and Mixed Poisson Regression. *Can. J. Stat.*, 15(2):209-225, 1987.
- [Legerreta(1991)] Legerreta, J. Contaminación Atmosférica en la Ciudad de México. *Revista Ciencias*, 22:53-62, 1991.
- [Linsey(1998)] Linsey, James K. *Applying Generalized Linear Models*. Springer-Verlag, New York, 1998.
- [López(1999)] López García, Adriana. Modelos Aditivos Generalizados Aplicados a Problemas de Contaminación y Salud Tesis de Maestría, UACPyP, IIMAS, UNAM, 1999.

BIBLIOGRAFÍA

- [Lovett(1988)] Lovett, Andrew A; Gatrell, Anthony C. The Geography of Spina Bifida in England and Wales. *Transactions of the Institute of British Geographers*, Vol. 13 (3):288-302, 1988.
- [McCullagh(1980)] McCullagh, Peter. Regression Models for Ordinal Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 42 (2):109-142. 1980.
- [McCullagh(1983)] McCullagh, P. Nelder, J.A. *Generalized Linear Models*. Chapman and Hall. London, 2nd edition, 1983.
- [McNeney(1994)] McNeney, B., Petkau, A. J. Overdispersed Poisson Regression Models for Studies of Air Pollution and Human Health. *Can. J. Stat.*, 22: 421-440, 1994.
- [Mood(1974)] Mood, Alexander M.; Graybill, Franklin A.; Boes, Duane C. *Introduction to the Theory of Statistics*. Probability and Statistics. McGraw - Hill International Editions, Ney York, third edition, 1974.
- [Morris(1995)] Morris Robert D, Naumova Elena N, et al. Ambient Air Pollution and Hospitalization for Congestive Heart Failure among elderly People in seven large US Cities *Am J Public Health*, 85:1361-1365, 1995.
- [Myers(2001)] Myers, Raymond H; Montgomery, Douglas C; Vining, G. Geoffrey. *Generalized Linear Models: with Applications in Engeriering and the Scienes*. Wiley Series in Probability and Statistics, 2001.
- [Nelder(1972)] Nelder, J. A.; Wedderburn, R. W. M. Generalized Linear Models. *J. R. Statist Soc. A.*, 135:370-384, 1972.
- [PEMEX(1991)] *Programa Integral contra la Contaminación Atmosférica. Un Compromiso Común*. PEMEX, México, 1991.
- [Peralta(1989)] Peralta, J. A. Ciudad y Transporte Urbano. *La Jornada Semanal*, p. 11, mayo 1989. citado en Ezcurra.
- [Pierce(1986)] Pierce, Donald A., Schafer, Daniel W. Residuals in Generalized Linear Models. *Journal of the American Statistical Association*, 81. Theory and Methods(396):977-986, 1986.
- [Plummer(1996)] Plummer, Martyn; Clayton, David Estimation of Population Exposure in Ecological Studies. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 58(1).(396):113-126, 1996.

- [Ponce de León(1996)] Ponce de León Antonio, Ross Anderson H, Bland J Martin, et al. Effects of Air Pollution on Daily Hospital Admissions for Respiratory Disease in London Between 1987-88 and 1991-92 *J Epidemiol Community Health*, 50 (Suppl 1):S63-S70, 1996.
- [Pönkä(1996)] Pönkä, Antti; Virtanen, Mikko. Asthma and Ambient Air Pollution in Helsinki. *J Epidemiol Community Health*, 50 (Suppl 1):S59-S62, 1996.
- [Pope III(1996)] Pope III, C. Arden; Kalkstein, Laurence S. Synoptic Weather Modeling and Estimates of the Exposure-Response Relationship between Daily Mortality and Particulate Air Pollution. *Environ Health Perspect*, 104:414-420, 1996.
- [Pregibon(1980)] Pregibon, Daryl Goodness of Link Test for Generalized Linear Models. *Appl. Statist.*, 29(1):15-24, 1980.
- [Pyne(1979)] Pyne, David. Single-Variable Poisson: A Goodness-of-Fit Test and the Comparison of Regression Coefficients. *Journal of the American Statistical Association*, 79:489-193, 1979.
- [Reade-Christopher(1996)] Reade-Christopher, Susan J.; Kupper, Lawrence L. Effects of Exposure Misclassification on Regression Analyses of Epidemiologic Follow-Up Study Data. *Biometrics*. Vol. 47.(2):535-548, 1996.
- [Restrepo(1985)] Restrepo I., Phillips D. *La basura: consumo y desperdicio en el D.F.*. Centro de Ecodesarrollo. México, 1985.
- [Reyes(1990)] Reyes, L.S. Salud y contaminación atmosférica. *Neumol. Cir. Torax.*, 49:22-38, 1990.
- [Rivero(1983)] Rivero Serrano, Octavio. *Contaminación Atmosférica y Enfermedad Respiratoria*. Fondo de Cultura Económica, México, 1983.
- [Riveros(1990)] Riveros, R.H. La contaminación atmosférica en la Ciudad de México. *Ciencia y Desarrollo*, p ginas 73-80, 1990.
- [Romieu(1995)] Romieu, Isabelle; Meneses, Fernando; Sienna-Monge, J. José; et al. Effects of urban air pollutants on emergency visits for childhood asthma in Mexico city. *Am J Epidemiol*, 141:546-53, 1995.
- [Romieu(1995)] Romieu, Isabelle. Editorial: Is carbon monoxide a risk factor for hospital admission for heart failure? *Am J Public Health*, 85:1343-1344, 1995.
- [Schwartz(1996)] Schwartz, J.; Spix, C.; Touloumi, G., et al. Methodological Issues in Studies of Air Pollution and Daily Counts Deaths or Hospital Admissions. *J Epidemiol Community Health*, 50 (Suppl 1):S3-S11, 1996.

BIBLIOGRAFÍA

- [Shouten(1996)] Shouten J.P, Vonk J.M, de Graaf A Short Terms Effects of Air Pollution on Emergency Hospital Admissions Fo Respiratory Disease: Results Ofthe APHEA Project in Two Major Cities in Netherlands 1977-89. *J Epidemiol Community Health*, 50:S22-S29, 1996.
- [Spix(1996)] Spix Claudia, Wichmann H Erich Daily Mortality and Air Pollutants: Findings from Köln, Germany. *J Epidemiol Community Health*, 50 (Suppl 1): S52-S58, 1996.
- [Suárez(1991)] Suárez, B.G. Análisis de la Calidad Atmosférica en la Ciudad de México. *Información Científica y Tecnológica*, 13:36-40, 1991.
- [Sunyer(1996)] Sunyer Jordi, Castellsagué Jordi, Sáez Marc, et al. Air Pollution and Mortality in Barcelona. *J Epidemiol Community Health*, 50 (Suppl 1):S76-S80, 1996.
- [Tellez(1996)] Tellez Rojo Solis, Martha María. Tesina Relación entre Enfermedades Respiratorias y Niveles de Contaminación en tres Hospitales del IMSS. Tesina para obtener la Especialidad en Estadística, UACPyP, IIMAS, UNAM, 1996.
- [Thall(Mar. 1988)] Thall, Peter F. Mixed Poisson Likelihood Regression Models for Longitudinal Interval Count Data. *Biometrics*, Vol. 44 (1):197-209, Mar. 1988.
- [Touloumi(1996)] Touloumi G, Samoli E, Katsouyanni K. Daily Mortality and "winter Type" Air Pollution an Athens, Greece - a Time Series Analysis within the APHEA Project. *J Epidemiol Community Health*, 50 (Suppl 1):S47-S51, 1996.
- [Vigotti(1996)] Vigotti M. A, Rossi G, Bisanti L, et al. Short Terms Effects of Urban Air Pollution on Respiratory Health in Milan, Italy, 1980-89. *J Epidemiol Community Health*, 50 (Suppl 1):S71-S75, 1996.
- [Wang(1985)] Wang P. C. Adding a Variable in Generalized Linear Models. *Technometrics*, 27(3):273-276, 1985.
- [Wang(1987)] Wang P. C. Residual Plots for Detecting Nonlinearity in Generalized Linear Models. *Technometrics*, 29(4):435-438, 1987.
- [Wedderburn(1972)] Wedderburn, R. W. M. On the Existence and Uniqueness of the Maximum Likelihood for Certain Generalized Linear Models. *Biometrika*, 63(1):27-32, 1972.
- [Wedderburn(1974)] Wedderburn, R. W. M. Quasi-likelihood Functions, Generalized Linear Models, and Gauss-Newton Method. *Biometrika*, 61(3):439-447, 1974.
- [Williams(1987)] Williams, D. A. Generalized Linear Models Diagnostics using the Deviance and Single case Deletions. *Appl Statist.*, 36(2):181-191, 1987.

- [Wojtyniak(1996)] Wojtyniak, Bogdan; Piekarski, Tomasz. Short Terms Effects of Air Pollution on Mortality in Polish Urban Populations -What is Different? *J Epidemiol Community Health*, 50 (Suppl 1):S36- S41, 1996.
- [Zmirou(1996)] Zmirou D, Barunandzadeh T, Balducci F, et al. Short Terms Effects of Air Pollution on Mortality in the City of Lyon France, 1985-90. *J Epidemiol Community Health*, 50 (Suppl 1):S30- S35, 1996.