

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

Tesis presentada como requisito para obtener el título de Licenciatura en Actuaría



**“EL IMPACTO DE LA CONTAMINACIÓN EN LA ECONOMÍA DE LA
SECRETARÍA DE SALUD POR MEDIO DE PROCESOS DE
RAMIFICACIÓN TIPO GALTON-WATSON”**

Presenta:

RUBÉN FUENTES GONZÁLEZ

Director de Tesis:

DR. JOSÉ RAÚL CASTRO ESPARZA

PUEBLA, PUE.

31 DE AGOSTO DE 2015

Dedicatoria

A mi madre, mi padre, mi tía y mi prima.

Agradecimientos

A los profesores, profesoras, amigos y amigas de la FCFM BUAP con los que tuve la oportunidad de compartir el salón de clases.

En especial al Dr. José Raúl Castro Esparza, Mtro. Manuel Ignacio Trujillo Mazorra, Mtro. José Asunción Hernández y al Mtro. Ángel Tejeda Moreno, por su incondicional apoyo y enseñanza.

Gracias a todos los antes mencionados por permitirme aprender de ellos en innumerables ocasiones.

Introducción

Dentro de la probabilidad y la estadística inferencial destacan importantes áreas como son la estimación de parámetros, procesos estocásticos y métodos computacionales aplicados.

En este trabajo se hará uso de esta teoría enfocándose de manera particular en un proceso de ramificación tipo Galton-Watson a través de la simulación estocástica.

La contaminación es un problema global que va en aumento y descontrol día con día y el alcance de los daños va más allá de un aire impuro. Se ha observado que los altos niveles de contaminación deterioran el medio ambiente, los recursos naturales y muy en particular la salud de la población a través de impurezas en el aire, bacterias y la mutación de las mismas. Estas bacterias generadas por la contaminación ocasionan alrededor del 100% de las distintas enfermedades en las vías respiratorias.

Dada esta situación, la Secretaría de Salud (SSA) debe costear y ejecutar un plan médico para combatir la propagación de infecciones en las personas. El problema es que estas epidemias pueden no estar previstas y mucho menos contempladas financieramente por la SSA.

Se pretende dar respuesta a la pregunta: ¿Es posible estimar el total de personas que requerirán un tratamiento médico por parte de la SSA? La hipótesis es que se puede construir un algoritmo que simule el comportamiento de la propagación de infecciones en individuos y estimar el total de personas enfermas.

Índice general

Introducción

1. Capítulo descriptivo

- a. Planteamiento del problema
- b. Justificación de la investigación
- c. Variables de estudio
- d. Alcance y limitaciones de la investigación

2. Capítulo teórico

- a. Marco teórico

3. Definición y ejecución de la solución algorítmica

- a. Construcción del algoritmo
- b. Ejecución en Microsoft Excel VBA
- c. Resultados de la investigación

4. Conclusiones generales

5. Referencias bibliográficas

**“El impacto de la contaminación en la economía de la Secretaría de Salud
por medio de procesos de ramificación tipo Galton-Watson”**

Rubén Fuentes González

31 de Agosto de 2015

Capítulo 1

Planteamiento del problema.

Considérese un escenario en el cual se comienza con un individuo que contrajo la enfermedad E y en el transcurso de 24 horas ese individuo estuvo en contacto con algunos otros individuos, de los cuales n resultaron infectados igualmente con la enfermedad E . Así mismo, estos n individuos infectados, en el transcurso de otras 24 hrs, estarán en contacto con algunos otros individuos de los cuales m también resultarán infectados.

Suponiendo que cada persona infectada sólo tiene una etapa de posible contagio y además comienza un tratamiento médico, se espera entonces que después de varios periodos de 24 horas se observe que la enfermedad E tuvo un inicio, propagación y extinción.

Con base en lo anterior, se implementará un algoritmo capaz de simular ésta situación para así estimar el número total de individuos que contrajeron la enfermedad E y tuvieron que recibir un tratamiento médico costado por la SSA.

Por medio de la observación y el registro de datos se puede generar un histograma de frecuencias en el que se podrá ajustar una curva, la cual nos dirá muy cercanamente cuál es la distribución probabilística que rige este fenómeno epidemiológico.

A la frecuencia de este tipo de fenómenos que ocurren 0, 1, 2, 3,... veces durante un periodo definido de tiempo, se le puede adaptar una distribución Poisson.

Una vez que se conoce la distribución se pueden generar variables aleatorias independientes e idénticamente distribuidas por medio del método de la transformada inversa, estimar el parámetro λ con los datos epidemiológicos observados y registrados previamente. Estas variables aleatorias van a simular el número de individuos contagiados entre cada periodo de 24 horas por cada uno de los individuos enfermos en ese instante. El número de contagios que produce un individuo es independiente al producido por otro, es por eso que se consideran variables aleatorias independientes y como se rigen por una misma distribución probabilística se les denomina idénticamente distribuidas.

El número de individuos contagiados por periodo es desconocido pero se sabe que cambia conforme al tiempo y es así como se asocia con un proceso estocástico discreto ya que se trata de individuos enteros.

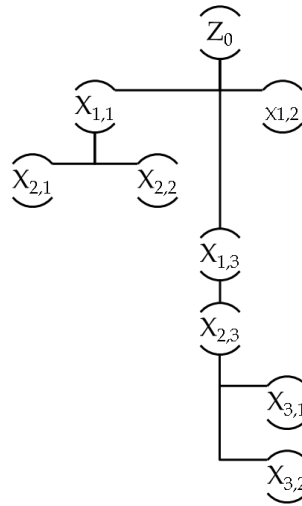
En este caso se estaría considerando a la primera persona infectada como la semilla del proceso tipo Galton-Watson. $\{Z_n\}$ es un proceso de ramificación, donde $\{Z_n\}$ = número de individuos infectados en la n -ésima generación, $n = 0, 1, 2, \dots$ y se supone que $Z_0 = 1$.

Cada individuo infectado de cada generación se expresa como sigue:

$X_{n,j}$ donde n es el número de generación y j es el número de individuo de dicha generación. La distribución de los individuos es representada por:

$$X_{n,j} \sim \text{Poisson}(\lambda)$$

Gráficamente se observaría así:



Cada vez que el proceso termina se guarda el número total de personas que fueron infectadas y se vuelve a simular el proceso bajo los mismos parámetros.

Después de un número suficientemente grande de iteraciones, por el teorema de límite central y la ley de los grandes números se puede estimar de manera casi puntual el valor esperado del total de individuos infectados.

Justificación de la investigación

Hoy en día, la contaminación ambiental es uno de los principales problemas a nivel global y es de suma importancia analizar la situación, particularmente, en un lugar como la capital de México.

El Distrito Federal es una de las principales ciudades del país, se caracteriza por la industria, el comercio, la tecnología y sobre todo por la gran población que lo habita.

Tan solo el 99.5% de su población se encuentra urbanizada, lo que significa que cuenta con servicios públicos como agua, luz, gas, internet, teléfono, bancos, centros comerciales, fábricas, etcétera. Y sólo el 0.5% es de tipo rural, así que posiblemente carecen de algunos servicios pero se practica la agricultura, la silvicultura, ganadería o algunas otras formas de producción.

El producto interno bruto (PIB) es la suma de los valores de todos los bienes o servicios finales producidos en un país en un periodo determinado y originalmente se calcula por año. Según el Instituto Nacional de Estadística y Geografía (INEGI), el Distrito Federal contribuye en un 17.7% sobre el PIB nacional, y del total de unidades económicas en México, que son establecimientos donde se genera o comercializa algún bien o servicio con la finalidad de promover el flujo de dinero, el 11.4% del total en el país, están en el DF.

De acuerdo a otros datos del INEGI, actualmente habitan alrededor de 8 851 080 personas en el DF. Respecto a la densidad poblacional se puede decir que hay aproximadamente 6 personas por metro cuadrado en el DF, que si bien se puede visualizar estarían un poco reducidos en espacio.

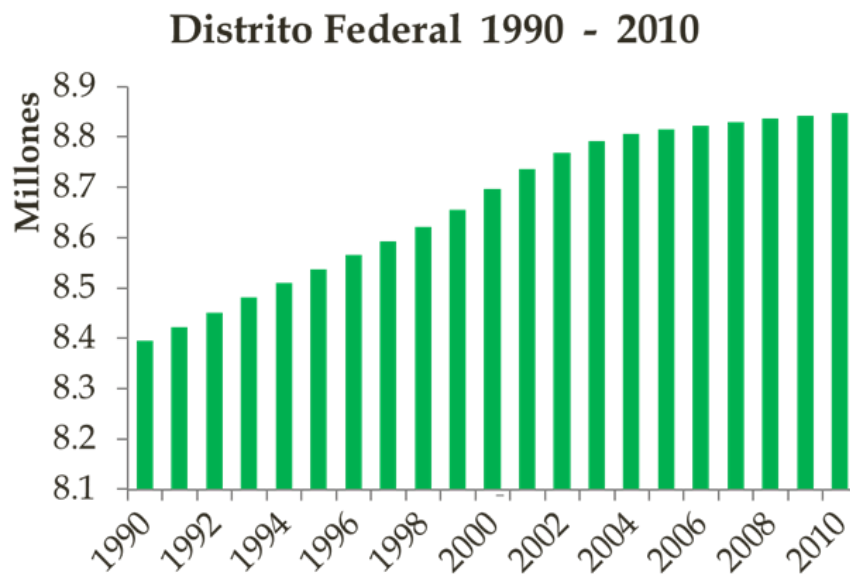


Figura 1.1 Población Distrito Federal 1990-2010

Fuente: INEGI

La población del DF representa el 7.9% de la población total del país, sin embargo la superficie territorial es apenas el 0.1% de la total del país.

Actualmente, en el D.F. existe un aumento excesivo de natalidad o población, además de escasez de recursos así como deterioro del medio ambiente, ríos contaminados, generación desmesurada de basura, muchos autos contaminando el aire, y en la mayoría de los casos la ciudad se expande talando zonas verdes.

Variables de estudio

La contaminación ambiental es la presencia de partículas en lugares, formas y concentraciones de manera que sean nocivas para la salud, seguridad o bienestar de la población o que simplemente sean perjudiciales para la vida vegetal o animal. Se genera a partir de la industria, el comercio, agricultura, hogares, vehículos, etc.

Aproximadamente en el año 1975, la Secretaría del Medio Ambiente del Gobierno del Distrito Federal creó el Sistema de Monitoreo Atmosférico de la Ciudad de México, mejor conocido por sus siglas como SIMAT.

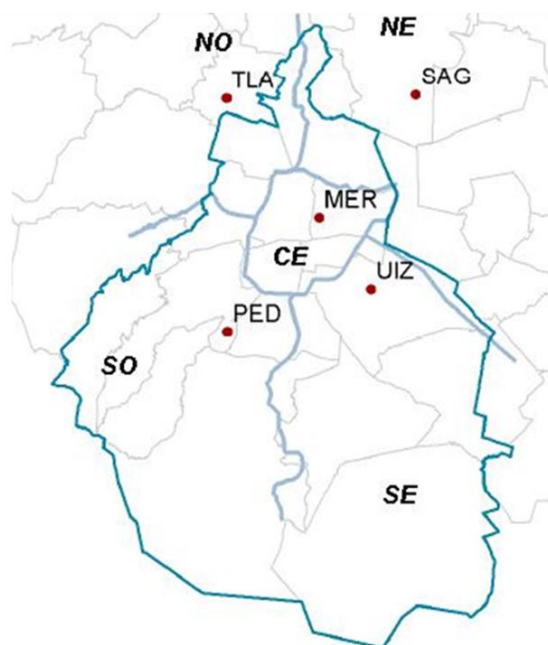
El objetivo de dicho sistema es vigilar y monitorear la calidad del aire en la zona metropolitana del Valle de México como medida de prevención para protección de la salud de sus habitantes.

Cuenta con un subsistema llamado RAMA que es la Red Automática de Monitoreo Atmosférico, realiza mediciones continuas y permanentes sobre los niveles de concentración del ozono, dióxido de azufre, dióxido de nitrógeno, monóxido de carbono, y partículas menores a 10 micrómetros.

Existen algunos otros subsistemas más que estudian la meteorología, la radiación solar, partículas húmedas y secas, partículas menores a 10 micrómetros que contengan plomo, entre otras cosas que puedan favorecer las lluvias ácidas.

RAMA cuenta aproximadamente con 34 estaciones distribuidas en el Distrito Federal y Estado de México, Las estaciones de monitoreo se van adaptando según la densidad poblacional de la zona y las direcciones del viento. Así es como se van dando de alta y baja algunas estaciones, según mantenimiento y funcionamiento.

Sin embargo, hubo 5 estaciones que funcionaron correctamente del 1° de enero del 2000 al 31 de diciembre del 2009, funcionaron para los 5 gases anteriormente mencionados, las 24 horas, los 365 días de cada año.



☞ Tlalpan

☞ La Merced

☞ Pedregal

☞ Universidad de Iztapalapa

☞ San Agustín

Figura 1.2 Estaciones de Monitoreo Atmosférico, Ciudad de México.

Fuente: RAMA.

Los diferentes niveles de concentración de los gases contaminantes generan ciertos daños a la salud pero estas dependen del tiempo al que la persona esté expuesta y la concentración del contaminante.

Existen dos tipos de exposiciones, la aguda que se refiere a periodos cortos (horas) pero niveles altos de concentración, y la crónica se refiere a periodos largos (meses o años) y bajas concentraciones.

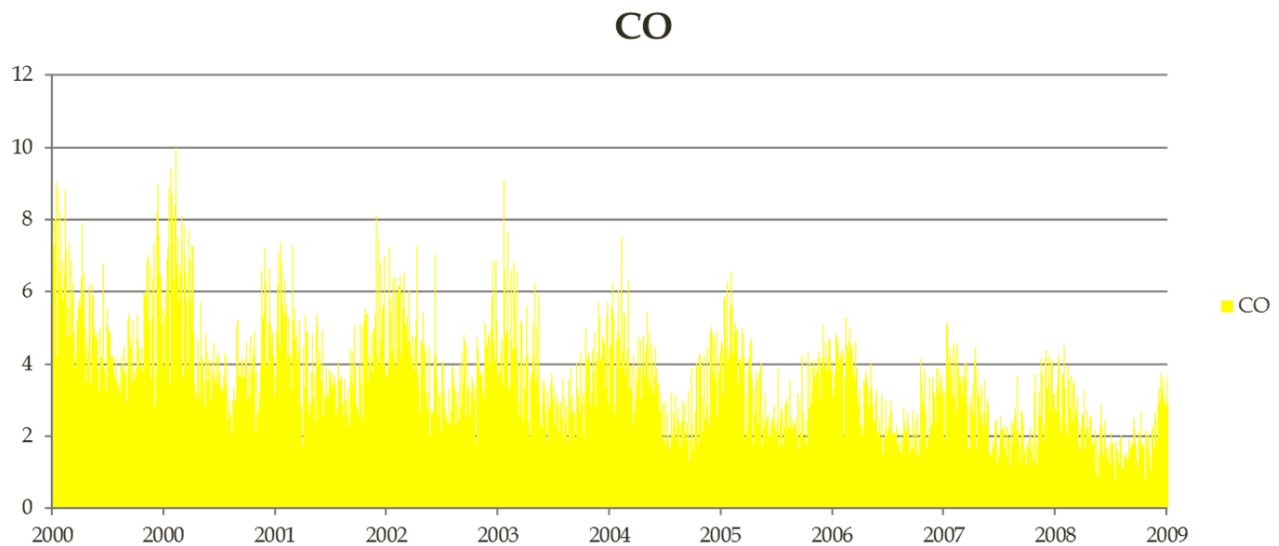


Figura 1.3 Concentración de CO

Fuente: RAMA.

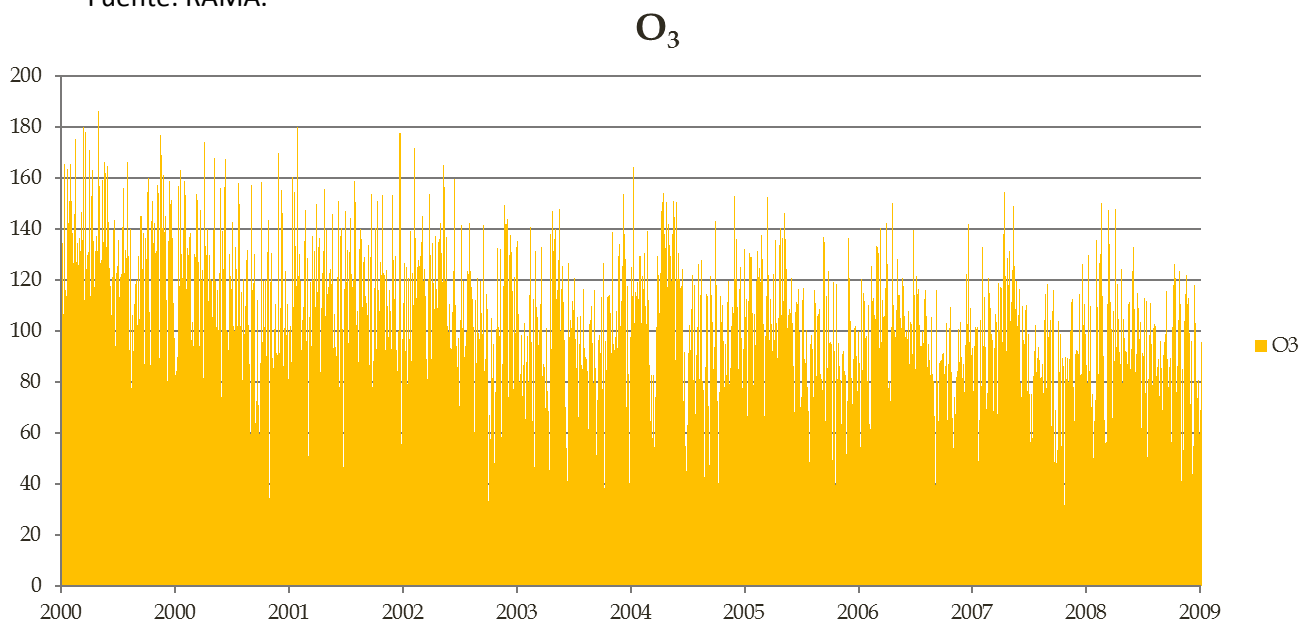


Figura 1.4 Concentración de O₃

Fuente: RAMA.

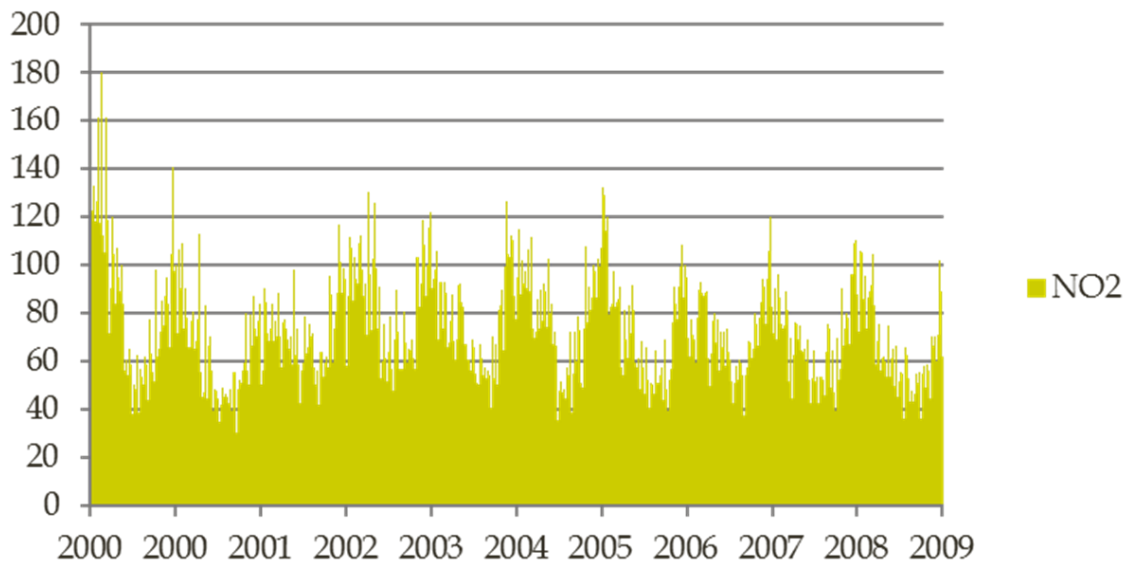


Figura 1.5 Concentración de NO₂

Fuente: RAMA.

PM₁₀

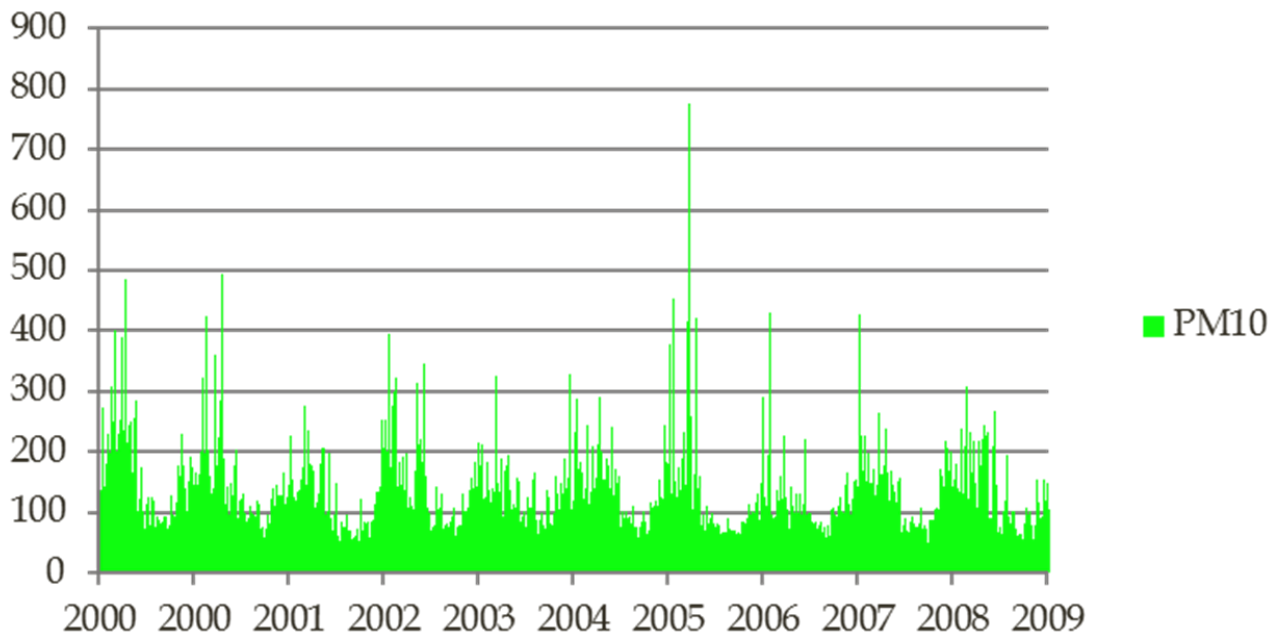


Figura 1.6 Concentración de PM₁₀

Fuente: RAMA.

SO₂

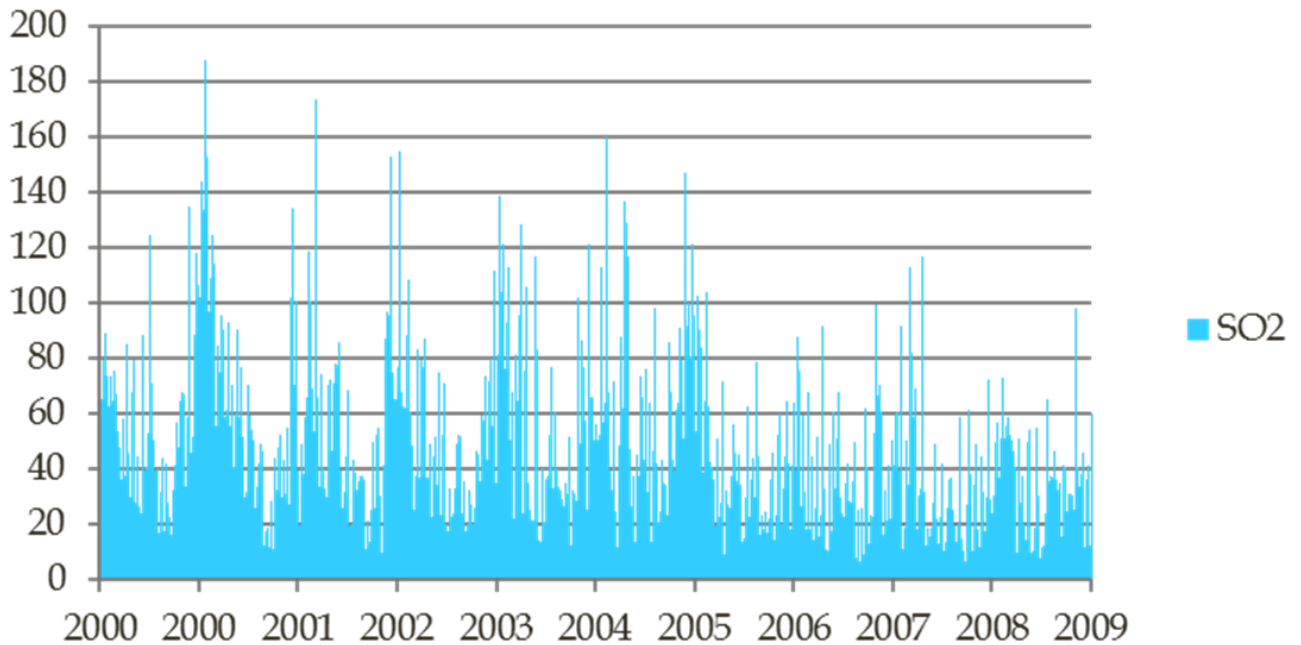


Figura 1.7 Concentración de SO₁₀

Fuente: RAMA.

El Sistema Nacional de Información en Salud (SINAIS) explica lo siguiente:

La exposición aguda produce en general tos, asma, bronquitis, afecta vías respiratorias, ojos, sistema inmunológico, problemas visuales, reduce capacidad mental, altera metabolismo, lesiona los pulmones.

La exposición crónica genera: sensibilidad a la neumonía y bronquitis, daña y lesiona pulmones, produce enfermedades cardiacas y agrava las respiratorias, en algunos casos requiere hospitalización y puede ocasionar muerte prematura.

Una enfermedad respiratoria es un proceso en los seres vivos de manera que su estado de salud se ve afectado o alterado por microorganismos, partículas o sustancias tóxicas en el medio ambiente. Las más comunes son: tos, gripa, neumonía, pulmonía, bronquitis, asma e influenza AH1N1.

La Secretaría de Salud se creó en el año 1938 y se encarga de crear y administrar establecimientos de salubridad de tal forma que las personas afiliadas a alguno de sus distintos institutos puedan hacer uso de los servicios médicos.

La cantidad de derechohabientes ha disminuido en los últimos años, sin embargo es curioso cómo la población cada vez aumenta más y cada vez menos gente está afiliada a algún programa o instituto de la SSA por lo tanto deben pagar un médico particular si quieren ser atendidos. Según el SINAIS, actualmente el 55.9% de la población es derechohabiente.

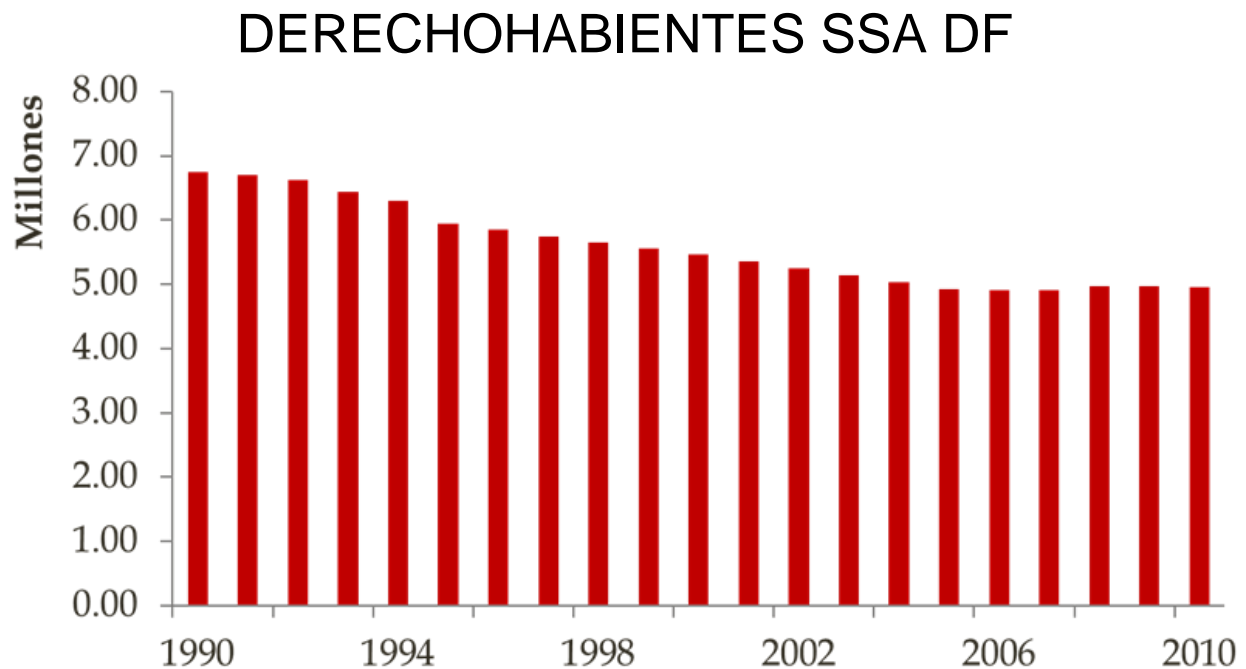


Figura 1.8 Población Derechohabiente del Distrito Federal, 1990-2010

Fuente: SINAIS.

De acuerdo a la Secretaría de Finanzas del Distrito Federal, el gobierno federal proporciona cada año a la SSA aproximadamente \$5,432 millones de pesos los cuales son utilizados para los servicios que ofrecen los institutos de la SSA.

RECURSOS FINANCIEROS D.F.

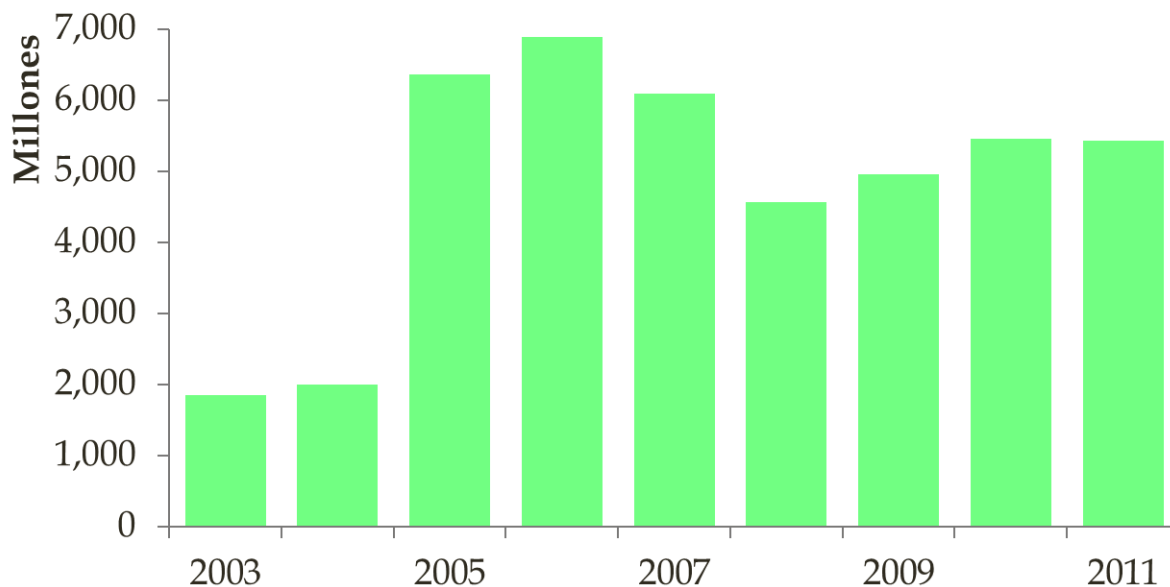


Figura 1.9 Recursos Financieros para SSA del Distrito Federal, 2003-2011

Fuente: SEFIN



Figura 2.1 Casos Reportados por Vías Respiratorias, Distrito Federal, 2000-2009

Fuente: SINAIS

De acuerdo a los datos antes mencionados, se supone la existencia de una estrecha relación entre los niveles de contaminación ambiental, el número de habitantes y el número de casos reportados por concepto de enfermedades del tipo respiratorio. Todo respecto a la Ciudad de México, Distrito Federal.

Inicialmente el modelo se propone y plantea en Microsoft Excel.

$$Q = [A(P)^T + B(CO)^U + C(NO_2)^V + D(O_3)^W + E(SO_2)^X + F(PM_{10})^Y]^Z$$

Dónde:

Q : número de casos reportados por concepto de enfermedades respiratorias.

P : número de habitantes totales en el DF.

CO : nivel de concentración del monóxido de carbono.

NO_2 : nivel de concentración del dióxido de nitrógeno.

O_3 : nivel de concentración del ozono.

SO_2 : nivel de concentración del dióxido de azufre.

PM_{10} : nivel de concentración de partículas menores a 10 micrómetros.

Los parámetros: A, B, C, D, E, F, T, U, V, W, X, Y, Z se inicializan arbitrariamente en 1.

Posteriormente el complemento Palisade Evolver 5.5 for Excel, se encargará de iterar estos parámetros por medio de algoritmos genéticos con el objetivo de maximizar el nivel de ajuste entre el modelo y los datos reales ya conocidos.

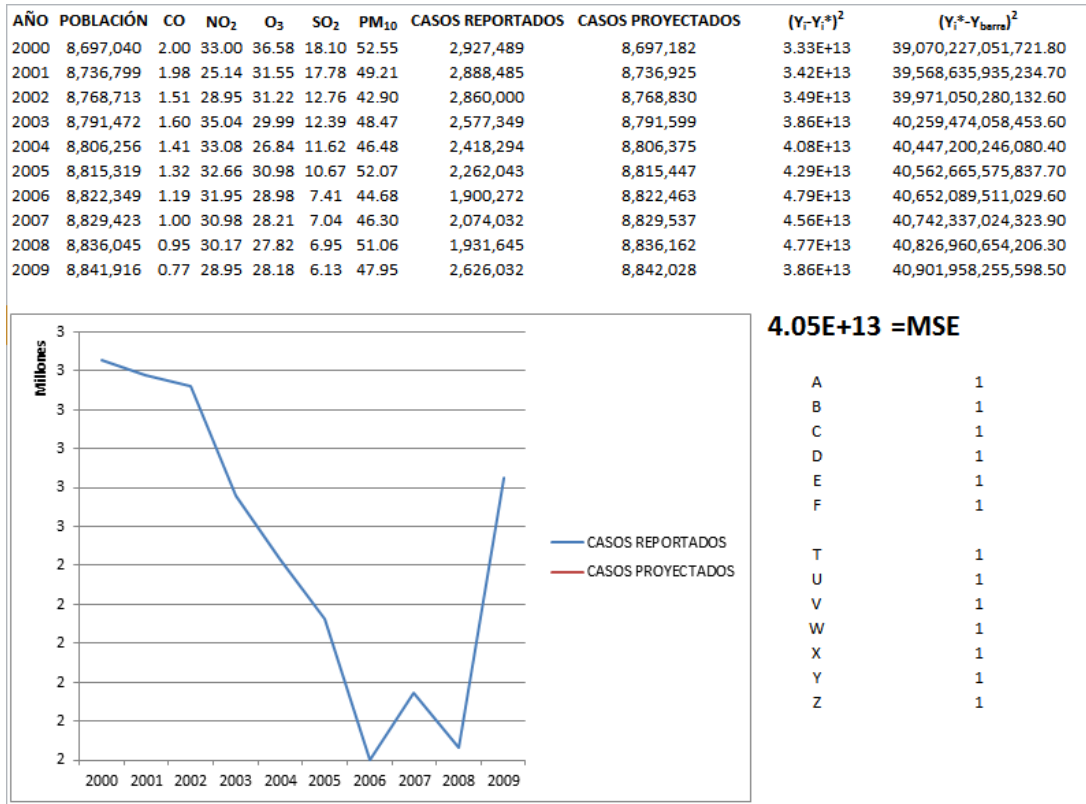


Figura 2.2 Regresión Múltiple Evolver 5.5

Fuente: Elaboración Propia.

Hasta este momento el modelo propuesto presenta un Error Cuadrático Medio (ECM) o Mean Squared Error (MSE) bastante grande.

Ahora se procede a ingresar las variables y objetivos por medio de Evolver 5.5.

3.42E+13	39,568,635,935,234.70
3.49E+13	39,971,050,280,132.60
3.86E+13	40,259,474,058,453.60
4.08E+13	40,447,200,246,080.40
4.29E+13	40,562,665,575,837.70
4.79E+13	40,652,089,511,029.60
4.56E+13	40,742,337,024,323.90
4.77E+13	40,826,960,654,206.30
3.86E+13	40,901,958,255,598.50

4.05E+13 =MSE

A	1
B	1
C	1
D	1
E	1
F	1
T	1
U	1
V	1
W	1
X	1
Y	1
Z	1

Figura 2.2 Inicialización de parámetros Evolver 5.5

Fuente: Elaboración Propia.

Por cuestiones de convergencia, se acotan los valores que pueden tomar los parámetros A, B, C, D, E, F, T, U, V, W, X, Y, Z a un valor mínimo de -500 y un valor máximo de 500.

Una vez ejecutado el software, comenzará a iterar los parámetros e irá mostrando el incremento en el nivel de ajuste del modelo.

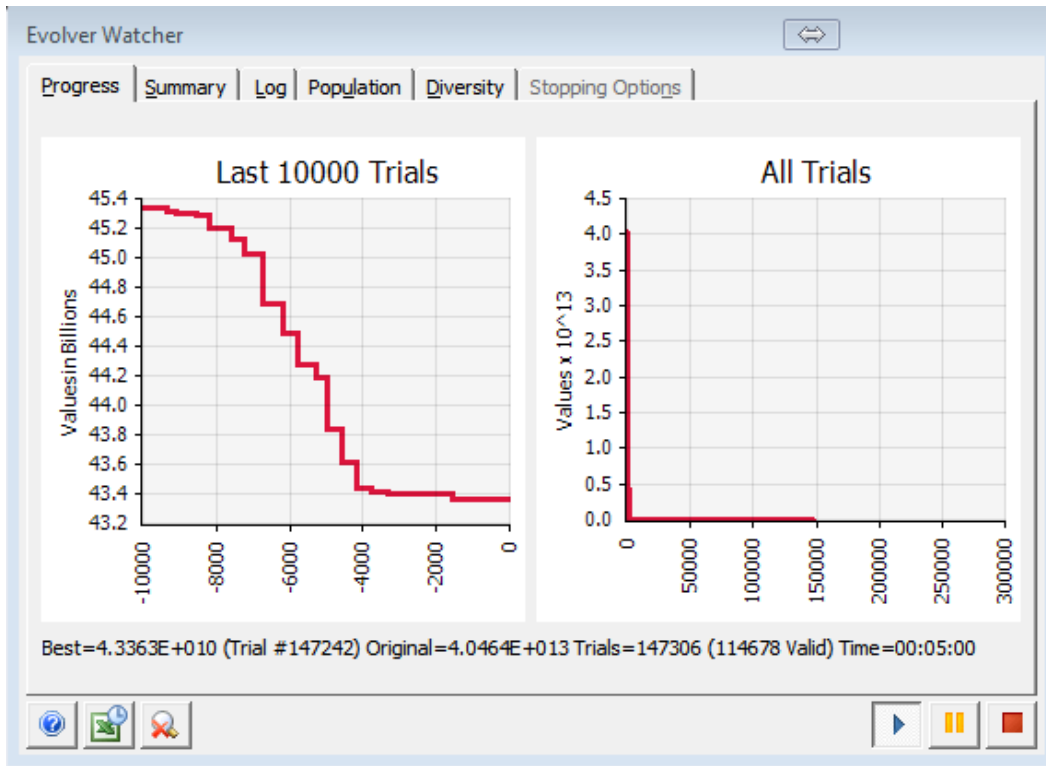
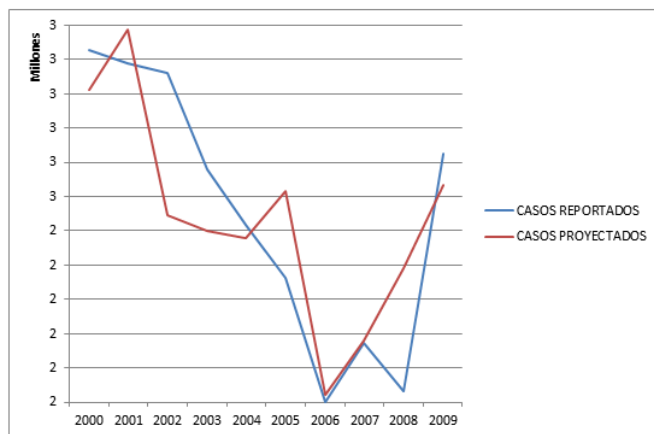


Figura 2.3 Iteraciones Evolver 5.5

Fuente: Elaboración Propia.

AÑO	POBLACIÓN	CO	NO ₂	O ₃	SO ₂	PM ₁₀	CASOS REPORTADOS	CASOS PROYECTADOS	$(Y_i - Y_i^*)^2$	$(Y_i^* - Y_{\text{barras}})^2$
2000	8,697,040	2.00	33.00	36.58	18.10	52.55	2,927,489	2,809,487	1.39E+10	131,712,775,080.17
2001	8,736,799	1.98	25.14	31.55	17.78	49.21	2,888,485	2,986,375	9.58E+09	291,395,568,929.59
2002	8,768,713	1.51	28.95	31.22	12.76	42.90	2,860,000	2,446,892	1.71E+11	107,815.67
2003	8,791,472	1.60	35.04	29.99	12.39	48.47	2,577,349	2,400,762	3.12E+10	2,097,795,972.51
2004	8,806,256	1.41	33.08	26.84	11.62	46.48	2,418,294	2,377,573	1.66E+09	4,759,741,763.74
2005	8,815,319	1.32	32.66	30.98	10.67	52.07	2,262,043	2,514,570	6.38E+10	4,624,858,857.03
2006	8,822,349	1.19	31.95	28.98	7.41	44.68	1,900,272	1,922,726	5.04E+08	274,405,986,179.26
2007	8,829,423	1.00	30.98	28.21	7.04	46.30	2,074,032	2,078,277	1.80E+07	135,635,703,963.39
2008	8,836,045	0.95	30.17	27.82	6.95	51.06	1,931,645	2,291,998	1.30E+11	23,890,818,764.07
2009	8,841,916	0.77	28.95	28.18	6.13	47.95	2,626,032	2,534,206	8.43E+09	7,681,029,675.23



4.30E+10 =MSE

A	-500
B	0.194570721
C	6.528322901
D	-499.829574
E	-10.76558257
F	1.71666332
T	-500
U	-3.683605801
V	-0.095794378
W	-3.397675325
X	-1.469070328
Y	0.193406666
Z	7.047130392

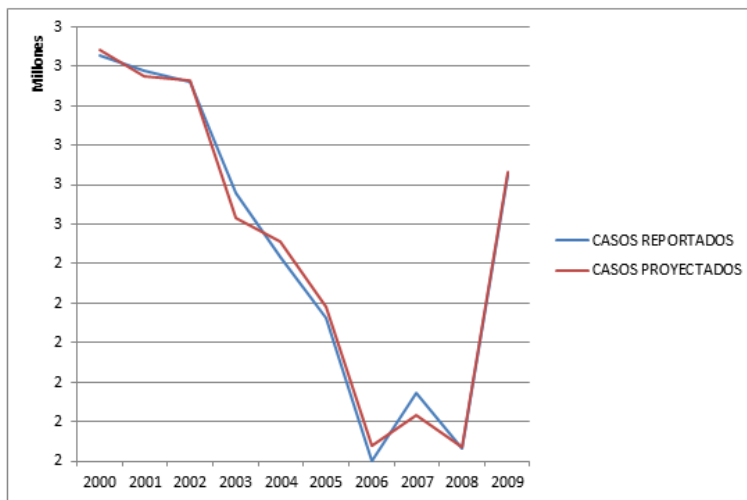
Figura 2.4 Evolución de la Regresión Múltiple

Fuente: Elaboración Propia.

Una vez que Evolver ya no muestra cambios significativos en el valor del MSE, es posible observar gráficamente un resultado bastante aceptable.

Los registros del campo “CASOS REPORTADOS” son datos reales, obtenidos del SINAIS, y posteriormente son comparados con los registros del campo “CASOS PROYECTADOS”, los cuales son estimados por medio del modelo propuesto.

AÑO	POBLACIÓN	CO	NO ₂	O ₃	SO ₂	PM ₁₀	CASOS REPORTADOS	CASOS PROYECTADOS	$(Y_i - Y_i^*)^2$	$(Y_i^* - Y_{\text{error}})^2$
2000	8,697,040	2.00	33.00	36.58	18.10	52.55	2,927,489	2,941,727	2.03E+08	245,186,369,566.97
2001	8,736,799	1.98	25.14	31.55	17.78	49.21	2,888,485	2,874,882	1.85E+08	183,456,302,710.82
2002	8,768,713	1.51	28.95	31.22	12.76	42.90	2,860,000	2,863,223	1.04E+07	173,604,567,309.12
2003	8,791,472	1.60	35.04	29.99	12.39	48.47	2,577,349	2,516,468	3.71E+09	4,886,506,651.78
2004	8,806,256	1.41	33.08	26.84	11.62	46.48	2,418,294	2,455,525	1.39E+09	80,296,714.31
2005	8,815,319	1.32	32.66	30.98	10.67	52.07	2,262,043	2,288,851	7.19E+08	24,873,572,549.23
2006	8,822,349	1.19	31.95	28.98	7.41	44.68	1,900,272	1,939,346	1.53E+09	257,269,741,068.78
2007	8,829,423	1.00	30.98	28.21	7.04	46.30	2,074,032	2,017,665	3.18E+09	183,954,365,179.83
2008	8,836,045	0.95	30.17	27.82	6.95	51.06	1,931,645	1,936,627	2.48E+07	260,035,964,877.67
2009	8,841,916	0.77	28.95	28.18	6.13	47.95	2,626,032	2,631,983	3.54E+07	34,380,244,226.67



1.10E+09 =MSE

A	-2.731434804
B	0.38712779
C	6.556738532
D	-12.35431841
E	-10.91821433
F	12.05310389
T	-0.066162396
U	-3.796097553
V	0.013988993
W	-0.880214618
X	-1.039696329
Y	-0.238595
Z	6.588255436

Figura 2.5 Evolución de la Regresión Múltiple

Fuente: Elaboración Propia.

Este óptimo alcanzado por el software es representado gráficamente e indica que el modelo propuesto se adapta bastante bien. En otras palabras, quiere decir que prácticamente se puede explicar Q a través de P, CO, NO₂, O₃, SO₂ y PM₁₀.

Alcance y limitaciones de la investigación

Los alcances que presenta esta investigación son los siguientes:

1. Refinar los parámetros del modelo propuesto con base en una mejor recolección de datos.
2. Cuantificar monetariamente el impacto que representa la calidad del aire de la Ciudad de México en el costo que la SSA debe considerar para atender los casos de enfermedades por vías respiratorias.

Las limitaciones que presenta esta investigación son las siguientes:

1. A pesar de recurrir a anuarios estadísticos de los cuales se supone veracidad de la información, los datos son vulnerables a alteraciones como errores de captura, modificación intencional de estos por parte de instituciones gubernamentales, entre otros factores humanos.
2. El refinamiento de la estimación, tanto de los parámetros del modelo explicativo como de los resultados finales obtenidos a través de la solución algorítmica, depende de la calidad y potencia de los algoritmos de utilizados por Evolver.
3. El algoritmo contempla únicamente una etapa de contagio y recuperación.
4. Existen algunos otros contaminantes que pudieran ser significativos y no están siendo monitoreados, aunque hipotéticamente están siendo considerados dentro de los errores de las estimaciones.

5. No se cuenta con sistemas de detección de partículas en muchos lugares donde pudiera requerirse la implementación de un sistema de simulación como el que se está desarrollando.
6. Se supone la existencia de uniformidad en los casos presentados de enfermos, por lo que descargo cualquier clase de efecto provocado por la estacionalidad en los reportes de enfermedades.

Capítulo 2

Marco teórico

Las variables aleatorias presentan una distribución que permanece constante a través del tiempo.

Uno de los objetivos de este capítulo es construir un modelo que nos permita explicar la estructura y prever la evolución, al menos a corto plazo, de una variable que observamos a lo largo del tiempo. La variable observada puede ser económica, física o social. Supondremos a lo largo del tema que los datos se obtienen en intervalos regulares de tiempo (horas, días, años) y el objetivo es utilizar el posible *momentum* en el comportamiento de la serie con el fin prever su evolución futura. Así, una serie temporal será una sucesión de valores de una variable obtenidos de manera secuencial durante el tiempo.

Concepto de proceso estocástico

Definición: Un proceso estocástico es una colección o familia de variables aleatorias $\{ X_t, \text{ con } t \in T \}$, ordenadas según el subíndice t que en general se suele identificar con el tiempo.

Por tanto, para cada instante t se tiene una variable aleatoria distinta representada por X_t , con lo que un proceso estocástico puede interpretarse como una sucesión de variables aleatorias cuyas características pueden variar a lo largo del tiempo. Por ejemplo, si observamos sólo unos pocos valores de t , tendríamos una imagen similar a la de la figura siguiente:

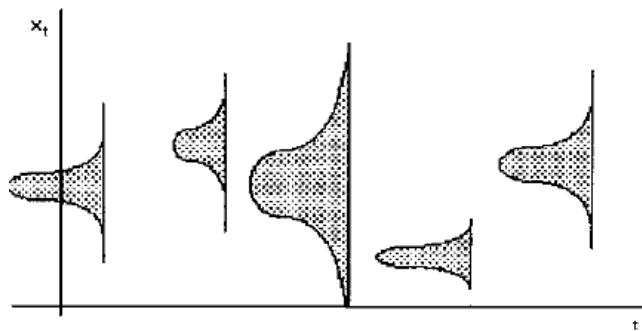


Figura 2.6 Concepto de Proceso Estocástico

Fuente: Sociedad Matemática Mexicana (SMM)

En la que se representa para cada t la función de densidad correspondiente a X_t . Aunque en la figura se han representado unas funciones de densidad variables, un proceso estocástico no tiene por qué presentar esas diferencias en la función de densidad a lo largo del tiempo.

A los posibles valores que puede tomar la variable aleatoria se le denominaran estados, por lo que se puede tener un espacio de estados discreto y un espacio de estados continuo. Por otro lado, la variable tiempo siempre es de tipo continua, aunque en ocasiones suele truncarse y aproximarse por una de tipo discreto. En el caso del tiempo discreto se podría tomar como ejemplo que los cambios de estado ocurran cada día, cada mes, cada año, etcétera. En el caso del tiempo continuo, los cambios de estado se podrían realizar en cualquier instante.

Por tanto, dependiendo de cómo sea el conjunto de subíndices T y el tipo de variable aleatoria dado por X_t se puede establecer una clasificación en particular a cada proceso estocástico.

En esta investigación se estará trabajando con un proceso de estado discreto y tiempo discreto, llamado cadena.

En el caso de procesos estocásticos con espacio de estados discreto, una secuencia de variables que indique el valor del proceso en intervalos sucesivos suele representarse de la siguiente manera:

$$\{X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = x_n\}$$

En la que cada variable $X_i, i = 0, \dots, n$. tiene una distribución de probabilidades que, en general, es distinta de las otras variables aunque podrían tener características comunes.

El principal interés del estudio a realizar en el caso discreto es el cálculo de probabilidades de ocupación de cada estado a partir de las probabilidades de cambio de estado.

Si en el tiempo $n - 1$ se está en el estado X_{n-1} , ¿con qué probabilidad se estará en el estado X_n en el instante siguiente n ? Esta probabilidad se denotará como:

$$P(X_n = x_n | X_{n-1} = x_{n-1})$$

A este tipo de probabilidad condicionada se le denomina probabilidad de transición o de cambio de estado. A las probabilidades del tipo $P(X_n = x_n)$ se les denomina probabilidades de ocupación de estado.

Otro tipo de probabilidad de interés es la de ocupar un cierto estado en un instante n , dado que en todos los instantes anteriores, desde $n = 0$ hasta $n-1$, se conoce en qué estados estuvo el proceso. Esto se puede escribir como:

$$P(X_n = x_n | X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1})$$

Nótese que esta probabilidad depende de toda la historia pasada del proceso, mientras que la probabilidad de transición depende únicamente del estado actual que ocupe el proceso.

Propiedad de Markov: Se dice que un proceso cumple la propiedad de Markov cuando toda la historia pasada del proceso se puede resumir en la posición actual que ocupa el proceso para poder calcular la probabilidad de cambiar a otro estado, es decir, se cumple la propiedad siguiente:

$$P(X_n = x_n | X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}) = P(X_n = x_n | X_{n-1} = x_{n-1})$$

Aquellas Cadenas que cumplen la propiedad de Markov se llaman Cadenas de Markov.

Otra manera de denotar a las probabilidades de transición es de la forma siguiente:

$$P(X_n = j \mid X_{n-1} = i) = p_{ij}(n)$$

Una propiedad interesante que puede tener una Cadena es que los valores $p_{ij}(n)$ no dependan del valor de n . Es decir, las probabilidades de cambiar de estado son las mismas en cualquier instante.

Procesos de ramificación

Existen algunos modelos matemáticos denominados Procesos de Ramificación. Estos modelos fueron olvidados por varios años, pero fueron retomados extensivamente entre los años veinte y treinta por su gran interés matemático y como una base teórica para estudios de poblaciones de individuos como serían: genes, apellidos, géneros, etcétera.

El proceso de ramificación más simple, es el de Galton-Watson, el cual tiene un ingrediente particular, y es el uso de la teoría de probabilidad y los procesos estocásticos.

Los primeros indicios sobre el uso de los procesos de ramificación se encuentran en el siglo XVIII en el famoso libro de tres volúmenes del investigador Thomas Malthus, titulado "An Essay on the Principles of Populations", en el cual se plantea que una población no controlada deberá crecer exponencialmente. Malthus relata

que en un pueblo de Berna, de las 487 familias burguesas existentes, 379 se extinguieron en el lapso de dos siglos (1583-1783).

Fue en 1874 cuando Francis Galton y H. W. Watson trabajaron en el problema de extinción de familias, mostrando como la probabilidad puede ser aplicada al estudio de los efectos del azar en el desarrollo de familias o poblaciones. Lamentablemente este primer trabajo tenía una hipótesis que no podía ser aceptada, ya que asumía que las familias distinguidas son más dadas a desaparecer que las ordinarias. Es así como Galton vuelve a plantear el problema de la siguiente manera:

Sean p_0, p_1, p_2, \dots las respectivas probabilidades de que un hombre tenga 0, 1, 2, ..., hijos varones; donde cada hijo varón tiene las mismas probabilidades que sus padres de tener hijos varones; y así sucesivamente. ¿Cuál es la probabilidad de que la línea masculina se extinga después de r -generaciones? O de una manera más general ¿Cuál es dicha probabilidad para un número de descendientes en la línea masculina de una generación dada?

La primera correcta y completa determinación de la probabilidad de extinción para dicho proceso fue dada por J. F. Steffenson (1930-1932). Después el modelo fue retomado por Kolmogorov y Dimitriev (1938), los cuales aumentaron notablemente su estudio y le dieron el nombre de los Procesos de Ramificación.

Para empezar con el estudio de este modelo, es necesario imaginar individuos que puedan generar nuevos individuos del mismo tipo, por ejemplo: el hombre, las bacterias reproduciéndose o neutrones en una reacción en cadena.

Primero se comienza con un conjunto inicial de individuos, al cual se le llamará generación cero, si estos individuos tienen descendientes, estos nuevos integrantes van a conformar a la primera generación y los descendientes de esta primera generación van a conformar a la segunda generación y así sucesivamente.

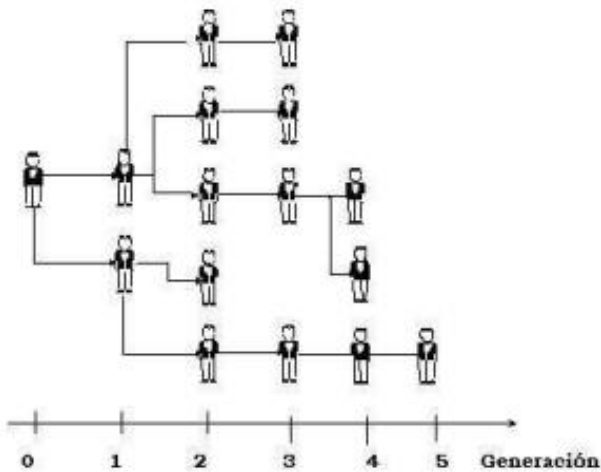


Figura 2.7 Ramificación

Fuente: Centro de Investigación en Matemáticas (CIMAT)

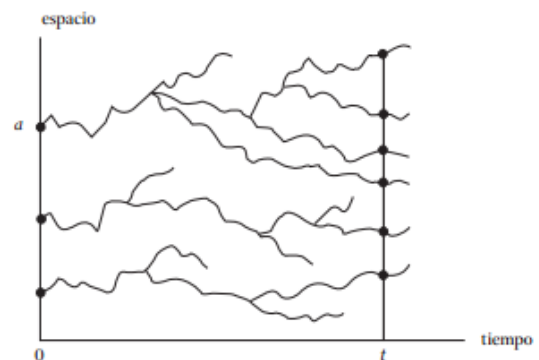


Figura 2.8 Ramificación

Fuente: CIMAT

Hay que hacer notar que se seguirá con atención el tamaño de las generaciones sucesivas, no el tiempo al cual cada individuo nace; así como tampoco la relación familiar de cada individuo.

Se denotará por Z_i como el número de individuos de la generación i . Además:

1. Si el tamaño de la n -ésima generación es conocido, entonces la ley de probabilidades que gobierna a las generaciones siguientes, no depende del tamaño de las generaciones previas a la n -ésima; en otras palabras Z_i forma una cadena de Markov.

2. La cadena de Markov considerada en esta sección, tiene una propiedad especial, la cual consiste en que diferentes individuos no interfieren con los otros; en el sentido de que el número de descendientes de un individuo no depende en cuántos otros individuos están presentes.

Referencia: "Relación entre los Procesos de Ramificación y los Procesos de Lévy, CIMAT"

Descripción matemática.

Se interpretará a Z_n como el número de individuos en la n -ésima generación; durante todo este capítulo se supondrá que el proceso inicia con un solo individuo, o de otra manera que $Z_0 = 1$.

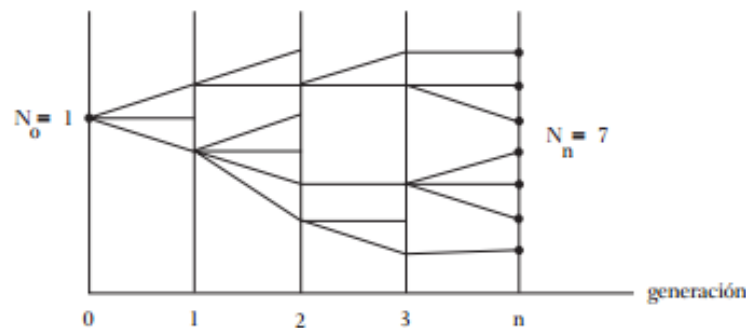


Figura 2.9 Ramificación

Fuente: CIMAT

Además se define $\{X_{i,n} : n \geq 1, i \geq 1\}$ como una sucesión de variables aleatorias independientes e idénticamente distribuidas, las cuales representan el número de descendientes que puede tener el i -ésimo individuo de la n -ésima generación.

De esta manera se escribirá a Z_{n+1} como: $Z_{n+1} = \sum_{i=1}^{Z_n} X_{i,n+1}$

Por esta última relación se observa que si $Z_n = 0$, la suma siempre va a ser cero, en consecuencia se afirmará que el cero es un estado absorbente.

Se denota P como la medida de probabilidad del proceso. Entonces la función masa de probabilidad de X_i está dada por:

$$P(X_{i,n} = k) = p_k \quad \text{para toda } k, n \in \{0, 1, 2, 3, \dots\} \quad \text{y} \quad \sum_{k=0}^{\infty} p_k = 1$$

Donde p_k es la probabilidad de que un individuo de la n -ésima generación tenga k hijos.

Con toda esta información se puede obtener una probabilidad de transición para el proceso que se denota por $P_{i,j}$ la cual representa la probabilidad de que en la $(n+1)$ -ésima generación haya j individuos dado que en la generación pasada n -ésima la población total era de i individuos.

La probabilidad de transición para toda $i, j, n \in \{0, 1, 2, 3, \dots\}$ está dada por:

$$P_{ij} = P(Z_{n+1} = j | Z_n = i) = P\left(\sum_{k=1}^i X_{k,n+1} = j\right)$$

Generación de variables aleatorias.

El método de la transformada inversa se presentará y utilizará para la generación de variables aleatorias tipo discreta.

Se considerará una variable aleatoria continua con función de distribución F . Este método en general se basa en la siguiente proposición:

Sea U una variable aleatoria uniforme $(0,1)$. Para cualquier función de distribución continua F , la variable aleatoria X definida como:

$$X = F^{-1}(U), \text{ tiene distribución } F.$$

$F^{-1}(U)$, es definida como el valor de x tal que $F(x) = u$.

Demostración: Sea F_x la función de distribución de $X = F^{-1}(U)$. Entonces:

$$F_x(x) = P\{X \leq x\} = P\{F^{-1}(U) \leq x\}$$

Como F es una función de distribución, resulta que $F(x)$ es una función monótona creciente de X y por lo que la desigualdad " $a \leq b$ " es equivalente a la desigualdad " $F(a) \leq F(b)$ ".

Entonces se puede ver lo siguiente:

$$\begin{aligned} F_x(x) &= P\{F(F^{-1}(U)) \leq F(x)\} && \text{como } F(F^{-1}(U)) = U \\ &= P\{U \leq F(x)\} = F(x) && \text{como } U \text{ es uniforme } (0,1) \end{aligned}$$

La proposición anterior muestra que podemos generar una variable aleatoria X a partir de la función de distribución continua F mediante la generación de un número aleatorio U y luego haciendo $X = F^{-1}(U)$.

Simulación estocástica.

La simulación es una técnica numérica para conducir experimentos en una computadora digital. Estos experimentos comprenden ciertos tipos de relaciones matemáticas y lógicas, las cuales son necesarias para describir el comportamiento y la estructura de sistemas complejos del mundo real a través de largos períodos.

La simulación estocástica es entonces la representación de fenómenos reales por medio de una computadora, a través de un modelo matemático compuesto por variables aleatorias. En este caso particular, las variables aleatorias a manejar son independientes e idénticamente distribuidas (*idd*).

Ley de los grandes números

Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas con media μ . Entonces, para cualquier $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{X_1 + \dots + X_n}{n} - \mu \right| > \epsilon \right) = 0$$

Teorema del Límite Central

Sea X_1, X_2, \dots, X_n un conjunto de variables aleatorias, independientes e idénticamente distribuidas con media μ y varianza $0 < \sigma^2 < \infty$.

Sea $S_n = X_1 + \dots + X_n$, entonces $\lim_{n \rightarrow \infty} P \left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z \right) = \varphi(z)$

Definición y ejecución de la solución algorítmica

Construcción del algoritmo

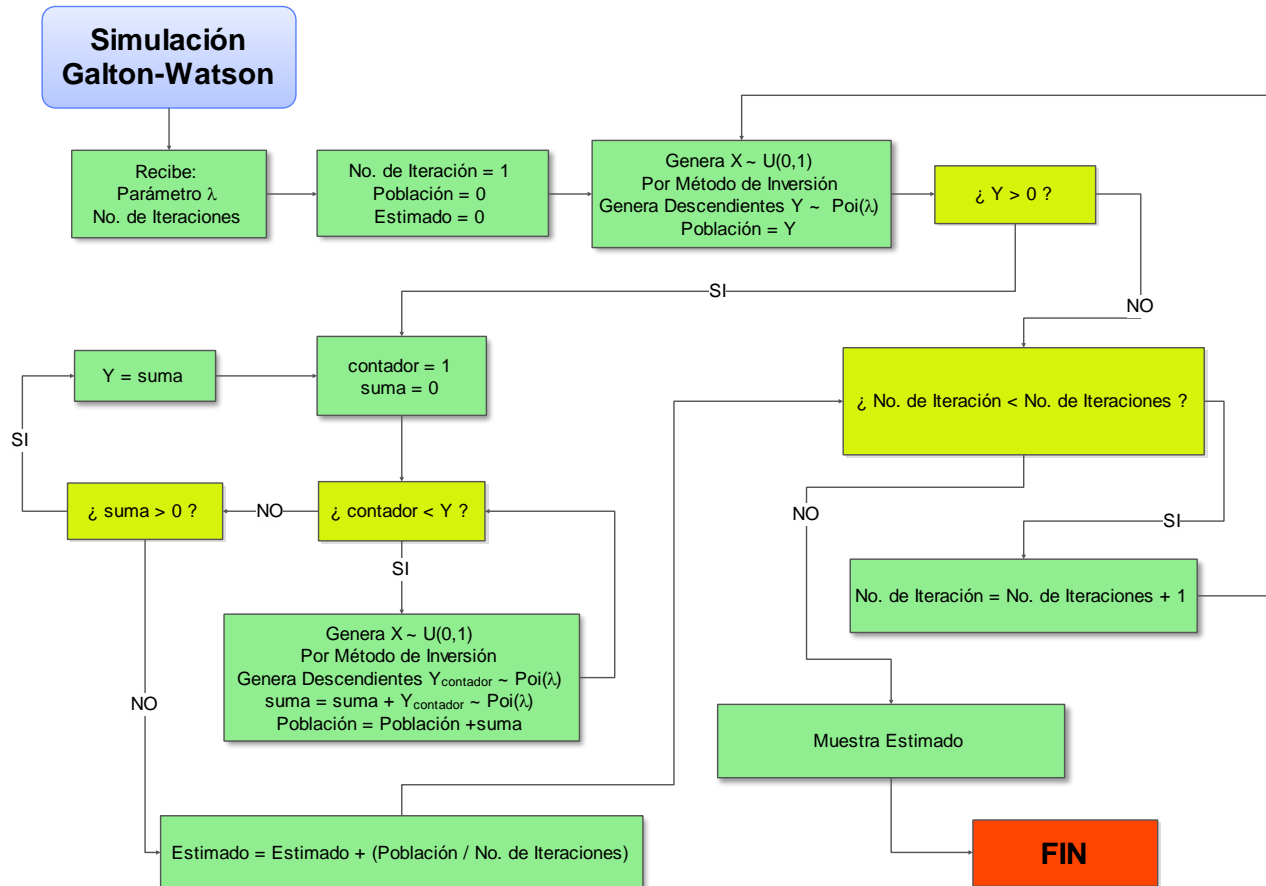


Figura 3.1 Algoritmo Proceso de Ramificación Galton-Watson

Fuente: Elaboración Propia.

Ejecución en Microsoft Excel VBA

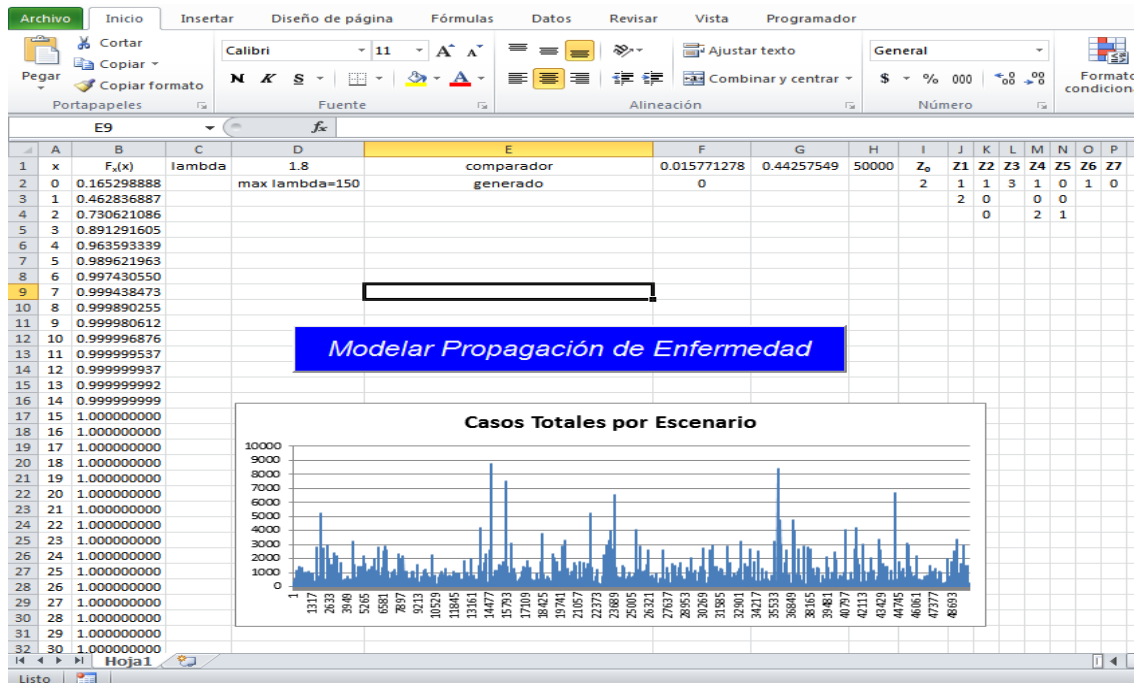


Figura 3.2 Interfaz inicial del software.

Fuente: Elaboración Propia.

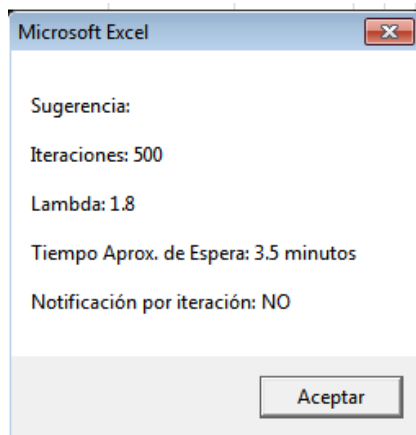


Figura 3.3 Sugerencias del software.

Fuente: Elaboración Propia.

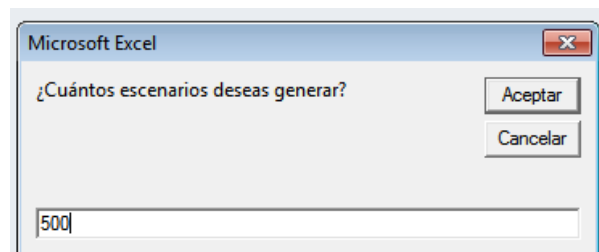


Figura 3.4 Número de escenarios a simular.

Fuente: Elaboración Propia.

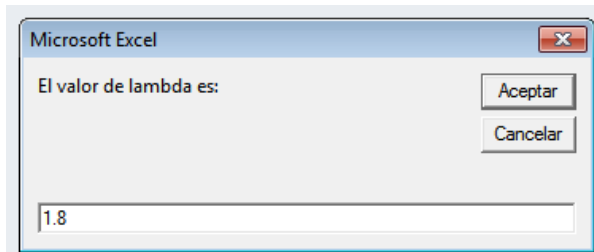


Figura 3.5 Número de descendientes promedio.

Fuente: Elaboración Propia.

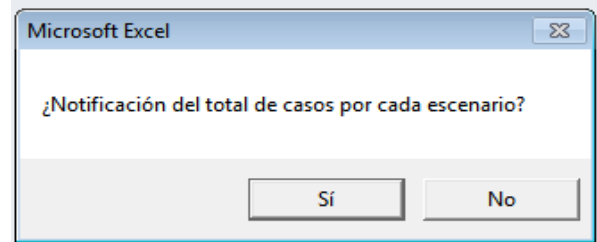


Figura 3.6 Opción de notificación por escenario.

Fuente: Elaboración Propia.

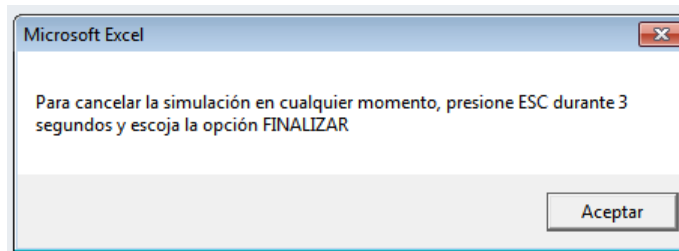


Figura 3.7 Aviso de cancelación del software.

Fuente: Elaboración Propia.

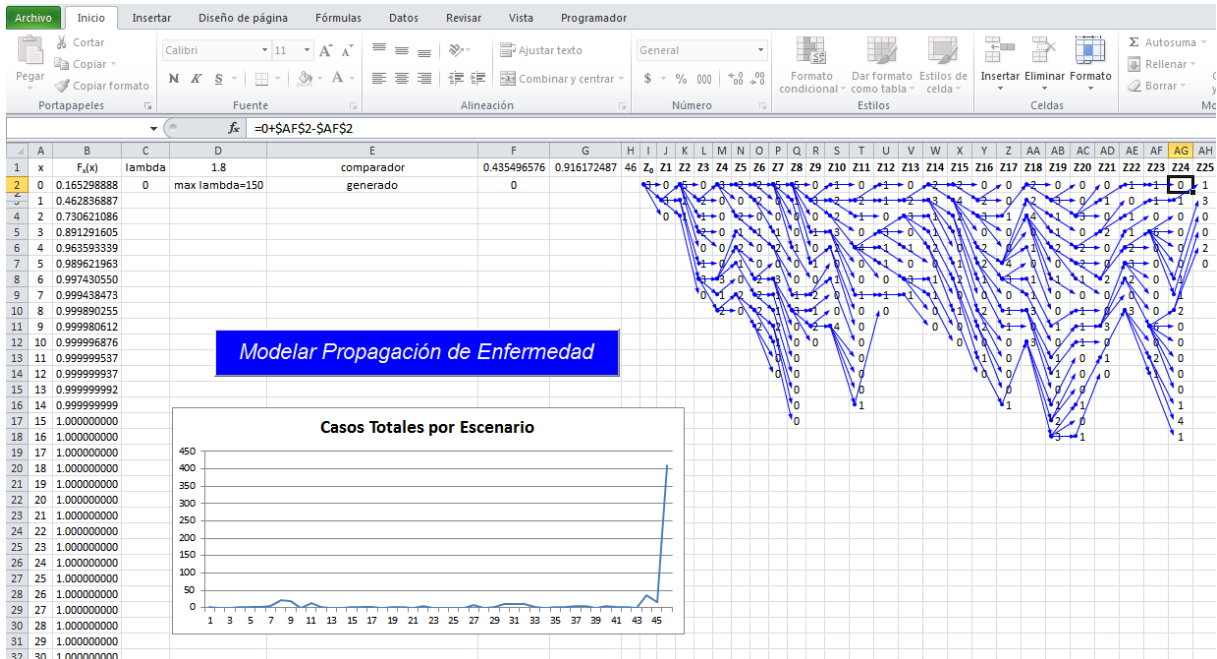


Figura 3.8 Simulación gráfica.

Fuente: Elaboración Propia.

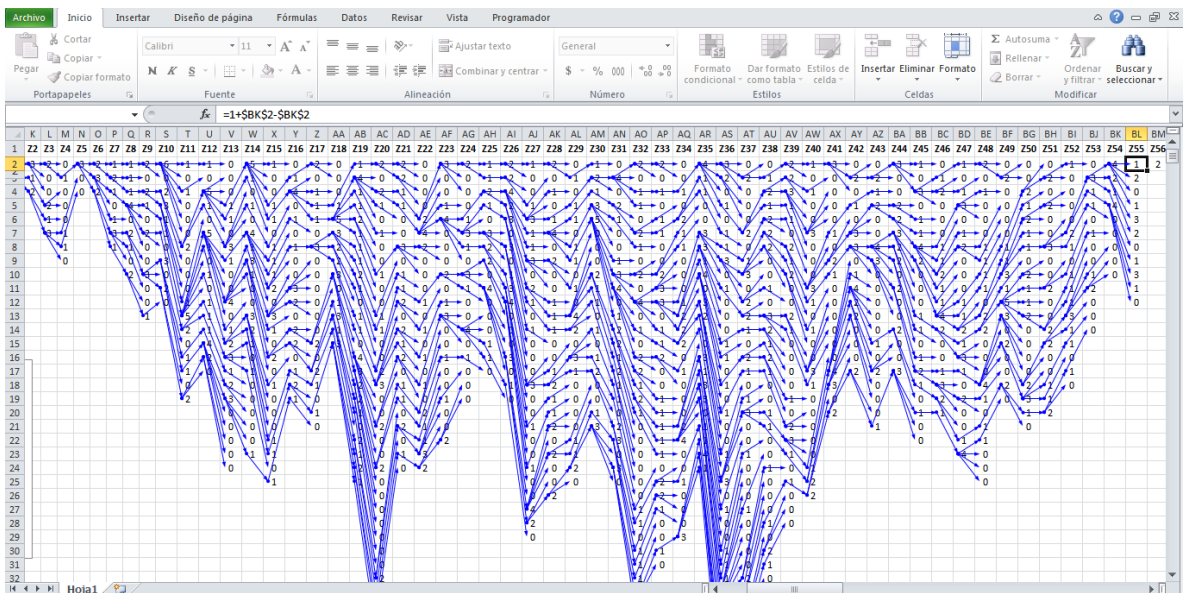


Figura 3.9 Simulación gráfica.

Fuente: Elaboración Propia.

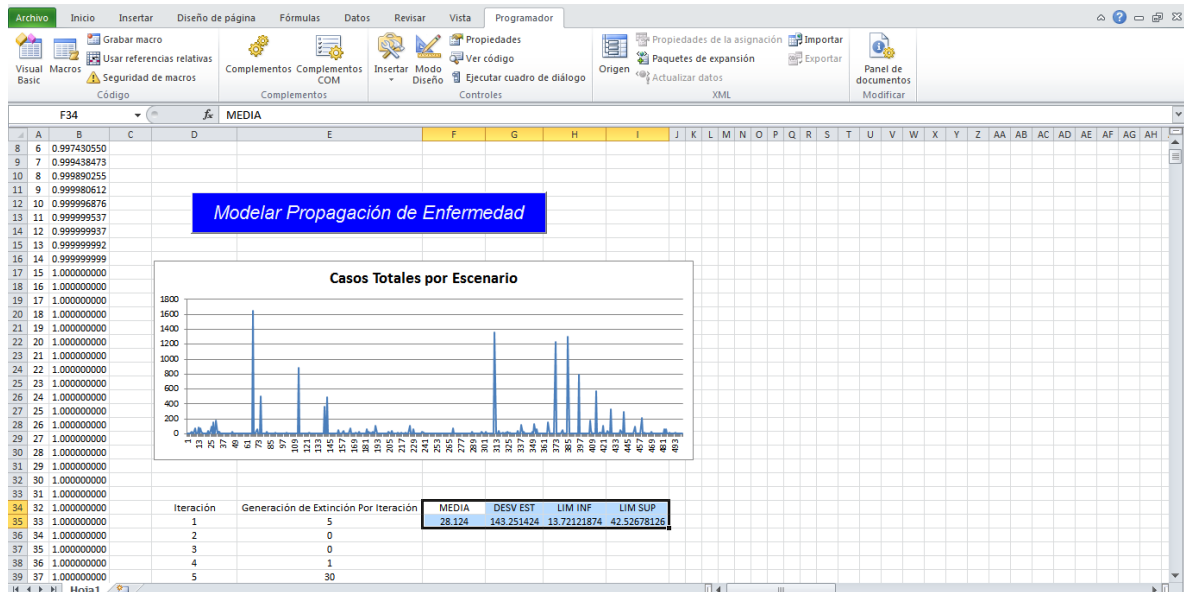


Figura 4.1 Simulación gráfica.

Fuente: Elaboración Propia.

Las imágenes previas muestran cómo se utiliza el programa y para fines demostrativos se seleccionaron como parámetros las sugerencias establecidas.

Resultados de la Investigación

500 escenarios VS 50,000 escenarios

ESCENARIOS	MEDIA MUESTRAL	DESV. ESTANDAR MUESTRAL	LIMITE INFERIOR	LIMITE SUPERIOR
500	28.124	143.251	15.568	40.680
50,000	27.188	172.565	25.675	28.701

Tabla 1.1 Comparativa de resultados.

Fuente: Elaboración Propia.

En cada caso se calculó la media muestral, desviación estándar muestral y el intervalo de confianza al 95%, cuyos cálculos matemáticos son las siguientes:

$$\text{Media Muestral: } \sum_{i=1}^{500} \frac{y_i}{500} = 28.124$$

$$\text{Desviación Estándar Muestral: } \sqrt{\frac{\sum_{i=1}^{500} (y_i - 28.124)^2}{499}} = 143.251$$

$$\text{Límite Inferior al 95\%: } 28.124 - 1.96 * \frac{143.251}{\sqrt{500}} = 15.568$$

$$\text{Límite Superior al 95\%: } 28.124 + 1.96 * \frac{143.251}{\sqrt{500}} = 40.680$$

De acuerdo a la Ley de los Grandes Números, para un número de datos suficientemente grande, estaremos más cerca del valor de los parámetros estimados, en este caso se hace se refiere a la media muestral.

Según el Teorema del Límite Central, la media muestral, cuyo valor es la suma de variables aleatorias y_i , se aproxima a una distribución normal para un número de y_i 's suficientemente grande. Es por eso que se asume normalidad y los intervalos de confianza al 95% fueron calculados de la siguiente manera:

$$\text{Límite Inferior: } \bar{y} - Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$$

$$\text{Límite Superior: } \bar{y} + Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$$

Respecto a la convergencia del algoritmo es bastante aceptable a pesar de la diferencia significativa entre el número de escenarios generados. Sin embargo, la desviación estándar muestral es considerablemente mayor que la media muestral y esto podría significar mucha variabilidad en las observaciones o que algunos datos atípicos son los que están impactando directamente en esta.

Lo que sigue es truncar los datos atípicos que rebasen los límites extremos, los cuales son definidos a continuación:

ESCENARIOS	LIMITE INFERIOR	LIMITE SUPERIOR	PERCENTIL 1%	PERCENTIL 99%
500	15.568	40.680	16.350	36.650
50,000	25.675	28.701	26	28

Tabla 1.2 Truncamiento de datos atípicos.

Fuente: Elaboración Propia.

ESCENARIOS	MEDIA MUESTRAL	DESV ESTANDAR MUESTRAL	LIMITE INFERIOR	LIMITE SUPERIOR
500	18.640	6.061	18.109	19.171
50,000	26.214	0.0616	26.209	26.220

Tabla 1.3 Comparativa de resultados truncados.

Fuente: Elaboración Propia.

Conclusiones generales

La alta variabilidad en los resultados observados no impide poder trabajar con ellos, ya que encontrando los límites extremos y truncando de acuerdo a ellos, se puede definir una nueva muestra y recalculando la media muestral, desviación estándar muestral y el intervalo de confianza al 95%. Luego entonces se podrá apreciar una disminución significativa en cuanto a la variabilidad de la nueva muestra.

Será viable estimar un total de individuos siempre y cuando se consideren los siguientes puntos cruciales:

- 1) El parámetro λ debe ser el adecuado para asegurar la extinción de la población infectada en alguna generación Z_t .
- 2) El equipo de cómputo donde se realice la simulación deberá contar con altas o suficientes especificaciones técnicas para poder soportar el λ dado por el usuario.

Anexos.

Código implementado en Visual Basic for Applications.

```

Sub ramificacion()
Range("F34") = ""
Range("G34") = ""
Range("H34") = ""
Range("I34") = ""
Range("F35") = ""
Range("G35") = ""
Range("H35") = ""
Range("I35") = ""
Application.ScreenUpdating = False
Range("D35").Select
Range(Selection, Selection.End(xlDown)).Select
Selection.ClearContents
Range("E35").Select
Range(Selection, Selection.End(xlDown)).Select
Selection.ClearContents
Range("H1").Select
MsgBox ("Sugerencia:" & Chr(13) & Chr(13) & "Iteraciones: 500" & Chr(13) &
Chr(13) & "Lambda: 1.8" & Chr(13) & Chr(13) & "Tiempo Aprox. de Espera: 3.5
minutos" & Chr(13) & Chr(13) & "Notificación por iteración: NO")
iteraciones = Val(InputBox("¿Cuántos escenarios deseas generar?"))
Total = 0
lambda = Val(InputBox("El valor de lambda es:"))
respuesta = MsgBox("¿Notificación del total de casos por cada escenario?",
vbYesNo)
MsgBox ("Para cancelar la simulación en cualquier momento, presione ESC
durante 3 segundos y escoja la opción FINALIZAR")
For contador = 1 To iteraciones
truco = 0
Worksheets("Hoja1").Cells(1, 8) = contador
limpieza
Worksheets("Hoja1").Cells(1, 4) = lambda
u = aleatorio()
Worksheets("Hoja1").Cells(1, 6) = u
Worksheets("Hoja1").Cells(2, 9) = Worksheets("Hoja1").Cells(2, 6)
truco = truco + Worksheets("Hoja1").Cells(2, 6)
ndpgen = Worksheets("Hoja1").Cells(2, 6)
generacion = 0
If ndpgen <> 0 Then
generacion = 1
For i = 1 To ndpgen

```

```

u = aleatorio()
Worksheets("Hoja1").Cells(1, 6) = u
Worksheets("Hoja1").Cells(i + 1, 9 + generacion).Font.ThemeColor =
xlThemeColorDark1
Worksheets("Hoja1").Cells(i + 1, 9 + generacion) = Worksheets("Hoja1").Cells(2, 6)
truco = truco + Worksheets("Hoja1").Cells(2, 6)
Next
While revsuma(generacion) <> 0
Worksheets("Hoja1").Cells(1, generacion + 9) = "Z" & generacion
ciclo = ndreng(generacion)
r = 2
For i = 1 To ciclo
For j = 1 To Worksheets("Hoja1").Cells(i + 1, generacion + 9)
u = aleatorio()
Worksheets("Hoja1").Cells(1, 6) = u
Worksheets("Hoja1").Cells(r, generacion + 10).Font.ThemeColor =
xlThemeColorDark1
Worksheets("Hoja1").Cells(r, generacion + 10) = Worksheets("Hoja1").Cells(2, 6)
truco = truco + Worksheets("Hoja1").Cells(2, 6)
r = r + 1
Next
Next
generacion = generacion + 1
Wend
Worksheets("Hoja1").Cells(1, generacion + 9) = "Z" & generacion
Application.ScreenUpdating = False
Cells.Select
Cells.EntireColumn.AutoFit
Range("I2").Select
Application.ScreenUpdating = True
If respuesta = vbYes Then MsgBox "El escenario No. " & contador & " , resultó con
un total de " & truco & " casos."
Else
End If
Total = truco + Total
Worksheets("Hoja1").Cells(34 + contador, 4) = contador
Worksheets("Hoja1").Cells(34 + contador, 5) = truco
cic = Now
ActiveSheet.ChartObjects("1 Gráfico").Activate
ActiveChart.SeriesCollection(1).XValues = "=Hoja1!$D$35:$D$" & 34 + contador
ActiveChart.SetSourceData Source:=Range("E34:E" & 34 + contador)
Worksheets("Hoja1").Range("H1").Select
columnas123 = Range("I1").CurrentRegion.Columns.Count
For i123 = 9 To columnas123 + 7
renlon123 = 2
inicio123 = 2
tope123 = 0

```

```

Cells(renglon123, i123).Select
While Cells(renglon123, i123) <> ""
If Cells(renglon123, i123) <> 0 Then
dependiente123 = Cells(renglon123, i123).Address
If renglon123 = 2 Then tope123 = tope123 + Cells(renglon123, i123) + 1
If renglon123 > 2 Then tope123 = tope123 + Cells(renglon123, i123)
For j123 = inicio123 To tope123
Cells(j123, i123 + 1).Font.ColorIndex = xlAutomatic
Cells(j123, i123 + 1) = "=" & Cells(j123, i123 + 1) & "+" & dependiente123 & "-" &
dependiente123
Next
inicio123 = tope123 + 1
Else
If renglon123 = 2 Then tope123 = 1
End If
If Cells(renglon123, i123) > 0 Then Cells(renglon123, i123).ShowDependents
renglon123 = renglon123 + 1
Wend
Next
Next
Range("F34") = "MEDIA"
Range("G34") = "DESV EST"
Range("H34") = "LIM INF"
Range("I34") = "LIM SUP"
Range("F35") = "=AVERAGE(E35:E" & iteraciones + 34 & ")"
Range("G35") = "=STDEV.S(E35:E" & iteraciones + 34 & ")"
Range("H35") = "=F35-TINV(0.025," & iteraciones - 1 & ")*G35/SQRT(" &
iteraciones & ")"
Range("I35") = "=F35+TINV(0.025," & iteraciones - 1 & ")*G35/SQRT(" &
iteraciones & ")"
Columns("H:H").AutoFit
Columns("I:I").AutoFit
MsgBox "Se reportan " & Round(Total / iteraciones, 1) & " casos en promedio."
End Sub

```

```

Function revsuma(numero_de_generacion)
renglon = 0
suma = 0
blanco = False
While blanco = False
If Worksheets("Hoja1").Cells(renglon + 2, numero_de_generacion + 9) <> "" Then
suma = suma + Worksheets("Hoja1").Cells(renglon + 2, numero_de_generacion +
9)
renglon = renglon + 1
Else
blanco = True
End If

```

```
Wend
revsuma = suma
End Function
```

```
Function ndreng(numero_de_generacion)
    renglon = 0
    blanco = False
    While blanco = False
    If Worksheets("Hoja1").Cells(renglon + 2, numero_de_generacion + 9) <> "" Then
        renglon = renglon + 1
    Else
        blanco = True
    End If
    Wend
    ndreng = renglon
End Function
```

```
Sub limpieza()
    Application.ScreenUpdating = False
    Range("I2:XFD1048576").Select
    Selection.ClearContents
    Range("J1").Select
    Range(Selection, Selection.End(xlToRight)).Select
    Selection.ClearContents
    Range("I1").Select
    Application.ScreenUpdating = True
End Sub
```

```
Function aleatorio()
    Application.Calculate
    aleatorio = Worksheets("Hoja1").Cells(1, 7)
End Function
```

Referencias bibliográficas

- 1) Reuven Y. Rubinstein. (2007). Simulation and the Monte Carlo Method. USA. Wiley.
- 2) Sheldon M. Ross. (2012). Simulation. USA. Academic Press.
- 3) Palisade Corporation. (2009). Evolver. USA. Palisade Corporation.
- 4) Andrew F. Seila. Vlatok Ceric. (2003). Applied Simulation Modeling. USA. Cengage Learning.
- 5) George Canavos. (1988). Probabilidad y Estadística. México. Mc. Graw-Hill.
- 6) DeGroot, M.H. (1988). Probabilidad y Estadística. México. Addison Wesley Iberoamericana.
- 7) Kalbfleisch J.G. (1979). Probability and Statistical Inference. USA. Spriger Verlag.
- 8) Hoel Port Stone. (1972). Introduction to Stochastic Processes. USA. Houghton Mifflin Company.
- 9) Luis Rincón. (2012). Introducción a los procesos estocásticos. México.
- 10) Kai Lai Chung. (2002). Elementary Probability Theory With Stochastic Processes and an Introduction to Mathematical Finance. USA. Springer.
- 11) Wayne L. Winston. (2008). Financial Models Using Simulation and Optimization. USA. Palisade Corporation.
- 12) Luis G. Gorostiza. (2009). Ramificación. México. CINVESTAV.