10

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES
"ACATLAN"

ANALISIS ESTADISTICO DEL CANCER PEDIATRICO EN EL D.F. MEDIANTE TABLAS DE CONTINGENCIA Y SERIES DE TIEMPO

(Período 1981 - 1991)

T E S I S

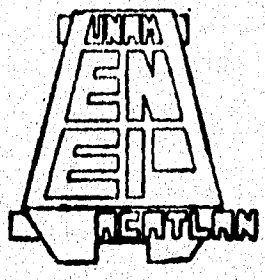
Que para obtener el Título de

LICENCIADA EN MATEMATICAS APLICADAS Y COMPUTACION

p r e s e n t a

NORMA CORREA NAVARRETE

ASESOR DE TESIS DRA. SILVIA RUIZ VELASCO ACOSTA



ACATLAN, EDO. DE MEXICO 1996

TESIS CON
FALLA DE ORIGEN

TESIS CON
FALLA DE ORIGEN

Universidad Nacional Autónoma de México
UNAM
Biblioteca Central
Dirección General de Bibliotecas de la UNAM

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.

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES "ACATLAN"

DIVISION DE MATEMATICAS E INGENIERIA
PROGRAMA DE ACTUARIA Y M.A.C.

VNIVERSIDAD NACIONAL
AVFNOMA DE
MEXICO

SRITA. NORMA CORREA NAVARRETE
Alumna de la carrera de M.A.C.
P r e s e n t e .

Por acuerdo a su solicitud presentada con fecha 9 de junio de 1994, me complace notificarle que esta Jefatura tuvo a bien asignarle el siguiente tema de tesis: "ANALISIS ESTADISTICO DEL CANCER PEDIATRICO EN EL D.F. MEDIANTE TABLAS DE CONTINGENCIA Y SERIES DE TIEMPO. (Período 1981-1991)", el cual se desarrollará como sigue:

INTRODUCCION
CAP. I Antecedentes.
CAP. II Tablas de Contingencia.
CAP. III Series de Tiempo.
CAP. IV Análisis de las variables de estudio.
CAP. V Análisis de las series de tiempo.
CONCLUSIONES.
ANEXO.
GLOSARIO.
BIBLIOGRAFIA.

Asimismo, fue asignado como Asesor de Tesis: DRA. SILVIA RUIZ VELASCO ACOSTA.

Ruego a usted tomar nota que en cumplimiento de lo especificado en la Ley de Profesiones, deberá presentar servicio social durante un tiempo mínimo de seis meses como requisito básico para sustentar Examen Profesional, así como de la disposición de la Coordinación de la Administración Escolar en el sentido de que se imprima en lugar visible de los ejemplares de tesis el título del trabajo realizado. Esta comunicación deberá imprimirse en el interior de la misma.

A T E N T A M E N T E
"POR MI RAZA HABLARA EL ESPIRITU"
Acatlán, Edo. Méx. febrero 6 de 1996.

ACT. LAURA MA. RIVERA BECERRA
Jefe del Programa de Actuaría
y M.A.C.

cg'

TESIS

COMPLETA

ANALISIS ESTADISTICO DEL CANCER PEDIATRICO EN EL D.F. MEDIANTE TABLAS DE CONTINGENCIA Y SERIES DE TIEMPO. (período 1981-1991)

Con todo el amor, toda la paciencia, toda la dedicación, toda la esperanza y toda la ilusión de poder concluir este proyecto que, tantas trabas tuvo que enfrentar, pero que al final se vio concluido, y hoy me proporciona todo lo entregado multiplicado por mil, para darme íntegramente una satisfacción inigualable.

CONTENIDO

INTRODUCCION 1

CAPITULO I. ANTECEDENTES 3

I) DEFINICIONES BASICAS 4

II) EL PROCESO DEL CANCER 5

III) EL CANCER INFANTIL (Una clasificación general de las neoplasias malignas del niño) 6

IV) ANTECEDENTES PREVIOS A LA INVESTIGACION 11

CAPITULO II. TABLAS DE CONTINGENCIA 13

I) DEFINICIONES BASICAS 13

II) ASOCIACION E INDEPENDENCIA EN TABLAS DE 2x2 14

III) ASOCIACION E INDEPENDENCIA EN TABLAS DE IxJ 22

CAPITULO III. SERIES DE TIEMPO 29

I) DEFINICIONES BASICAS 29

II) ANTECEDENTES 30

III) IDENTIFICACION DE MODELOS 35

IV) ESTIMACION DE MODELOS 44

V) EXAMEN DEL MODELO 48

VI) MODELOS ESTACIONALES 51

CAPITULO IV. ANALISIS DE LAS VARIABLES DE ESTUDIO 55

I) SEXO DEL PACIENTE Y TIPO DE CANCER 56

II) DOMICILIO (DELEGACION) DEL PACIENTE Y TIPO DE CANCER 57

III) INSTITUCION DE SALUD DEL PACIENTEY TIPO DE CANCER 58

IV) PESO DEL NIÑO AL NACER Y TIPO DE CANCER 60

V) TIPO DE EMBARAZO Y TIPO DE CANCER 62

VI) ESCOLARIDAD DEL PADRE DEL NIÑO Y TIPO DE CANCER 62

VII) ESCOLARIDAD DE LA MADRE DEL NIÑO Y TIPO DE CANCER 63

VIII) EDAD DE LA MADRE AL NACIMIENTO DEL HIJO Y TIPO DE CANCER 64

IX) TIEMPO TRANSCURRIDO ENTRE EL INICIO DE SIGNOS Y SINTOMAS AL DIAGNOSTICO Y TIPO DE CANCER 65

X) EDAD DEL NIÑO AL DIAGNOSTICO Y TIPO DE CANCER 67

XI) FAMILIARES (C/S) CANCER Y TIPO DE CANCER 68

XII) RAMA DE LA CUAL PROCEDE EL CANCER Y TIPO DE CANCER 69

XIII) INTERNAMIENTO DURANTE SU PRIMER AÑO DE VIDA POR ALGUN PROCESO INFECCIOSO Y TIPO DE CANCER 71

XIV) MALFORMACION EN EL PACIENTE Y TIPO DE CANCER 72

XV) ESCOLARIDAD DE LA MADRE Y TIEMPO TRANSCURRIDO ENTRE INICIO DE SIGNOS Y SINTOMAS AL DIAGNOSTICO 72

CAPITULO V. ANALISIS DE LAS SERIES DE TIEMPO 75

I) SERIE GENERAL 76

II) SEXO 87

III) TIPO DE CANCER 95

IV) FAMILIARES DEL PACIENTE (C/S CANCER) 102

V) EDAD DEL PACIENTE AL DIAGNOSTICO DEL CANCER 108

VI) ESCOLARIDAD DEL PADRE 114

VII) ESCOLARIDAD DE LA MADRE 120

VIII) EDAD DE LA MADRE AL NACIMIENTO DEL HIJO 126

CONCLUSIONES 133

ANEXO 137

GLOSARIO 145

BIBLIOGRAFIA 149

A la Dra. Silvia Ruiz, en quien encontré los elementos fundamentales para poder realizar la tesis:

- conocimientos,
- paciencia,
- dedicación y
- un gran apoyo.

Muchas gracias, Silvia.

Agradezco a la Unidad de Epidemiología del Hospital de Pediatría del Centro Médico Nacional siglo XXI, por las facilidades prestadas al proporcionar la base de datos de pacientes con cáncer pediátrico en el D.F. del período 1981-1991, así como artículos referentes al tema.

Y muy en especial, quiero agradecer

al Dr. Arturo Fajardo, por su entusiasmo y apoyo invaluable en la realización de la presente investigación.

Al gran equipo que tuvo tan importante colaboración en el desarrollo de la tesis, es decir, todos mis amigos y compañeros que contribuyeron con su granito de arena, para poder hacer de este trabajo la gran montaña que significa para mí.

A mis padres.

Pero sobre todas las cosas

al ser todo poderoso, que me ha permitido vivir hasta estos momentos y culminar una de mis metas más preciadas.

INTRODUCCION

La vida, sin duda alguna, es el regalo más valioso que se nos ha dado, y la salud es un factor fundamental para lograr que ésta sea de una buena calidad. Sin embargo, el mundo que nos rodea está lleno de diversas enfermedades y muchas de éstas carecen de una completa cura. Una de ellas es el cáncer, el cual es difícil detectar en sus primeras etapas, y por lo tanto, es casi imposible asegurar la cura total del paciente. El trabajo que aquí se presenta trata precisamente de esta enfermedad, pero bajo el siguiente contexto: se estudia exclusivamente al cáncer pediátrico, y el estudio se realiza en el D.F. con los pacientes que se les detectó neoplasias malignas en los principales hospitales que se dedican a la atención del niño. La idea fundamental es colocar una pequeñísima parte del gigantesco rompecabezas que representa esta enfermedad.

La investigación fue posible gracias a que se proporcionó la base de datos con casos de neoplasias malignas en el D.F. El instrumento de recolección de datos, el cual se presenta en el anexo, tiene una gran cantidad de variables sociodemográficas y otras que dan más bien información sobre los antecedentes relacionados con la enfermedad de cada paciente. A partir de todas ellas se realizó una selección cuidadosa de las que contienen información médica y que además fue posible obtener frecuencias adecuadas para su estudio, es decir, que se tuvieron suficientes datos para analizarlos.

De acuerdo a todo lo anterior, se decide aplicar dos metodologías: Tablas de Contingencia y Series de Tiempo. La primera nos ayuda a determinar si existe asociación entre las variables de estudio y los doce grupos de cáncer, y en caso de haberlo, en qué grado está. Esto nos proporciona en una forma real y práctica pistas para determinar posibles factores de riesgo en el desarrollo de la enfermedad. Las series de tiempo se utilizan con el propósito de tener un aspecto descriptivo en el proceso del cáncer pediátrico. Por este lado, los factores de interés son la tendencia, la existencia de variación estacional y los modelos que mejor se ajustan a cada serie. Por lo tanto, el análisis va direccionado hacia las cuestiones señaladas. Esto es debido al carácter fundamentalmente médico que tiene el objetivo de realizar la investigación.

Antes de comenzar con el análisis, se presentan los antecedentes y la teoría concerniente a las dos metodologías señaladas. En los antecedentes se mencionan las principales razones para hacer una investigación sobre cáncer infantil y se explican a grandes rasgos los doce grupos de neoplasias malignas. Después de obtener los resultados se presentan las conclusiones y al final se tiene el apéndice, donde se encuentran datos sobre la población proyectada para el período de estudio, el instrumento de recolección de datos y un pequeño programa que se utilizó para obtener las frecuencias.

El trabajo en conjunto es el que nos da las razones, el análisis y las conclusiones a las que se llegan sobre el fenómeno del cáncer pediátrico en el D.F. Ahora bien, resulta importante señalar que el objetivo que se pretende alcanzar es detectar asociaciones entre las variables sociodemográficas y los diferentes tipos de neoplasias malignas, así como describir el comportamiento del proceso del cáncer pediátrico en base a los datos recolectados en el D.F. (período 1981-1991).

I. ANTECEDENTES

El cáncer infantil, según la literatura revisada, ha sido poco estudiado, sobre todo en México, país en vías de desarrollo que no cuenta con datos estadísticos a nivel nacional. Además, un estudio de tal magnitud implica altos costos, agregándole a esto que no hay un material lo suficientemente confiable para llevar a cabo un análisis serio como lo amerita. Sin embargo, es importante realizar su estudio, ya que el impacto que causa sobre los menores es muy grande, tanto en el aspecto físico como emocional, además de que ocupa un lugar considerable como causa de muerte en la infancia. Por todo lo anterior y por las razones que pudiesen escapar a su mención, pero que existen y son palpables por los pacientes que sufren de cáncer así como por sus familiares, es que se realiza el presente trabajo basado en los datos recolectados en el D.F. También se trata de colaborar en la investigación sobre cáncer infantil realizada en el Hospital de Pediatría del Centro Médico Nacional Siglo XXI.

Por último, y como punto de partida hacia el largo camino que se vislumbra ahora, que promete tener un destino esclarecedor para todos los que, perseverantes, no den como fallidas las pistas que parezcan decepcionar, sino como experiencias para indagar por otros caminos; aún quizá, tomar las mismas con diferente enfoque, o con otras características, como sea, es que se mencionan algunas de las razones más importantes para llevar a cabo el estudio del cáncer infantil según la literatura revisada sobre lo que nos atañe:

— El cáncer es una de las principales causas de morbilidad y mortalidad de la infancia.

— El impacto que recibe un menor al sufrir de cáncer es relativamente mayor que en las personas adultas.

— En los casos más extremos, hay una mayor pérdida en años de vida en comparación con las personas adultas que padecen de alguna enfermedad maligna.

— En el caso de niños sobrevivientes, no pueden acudir a la escuela, y esto, lógicamente, afecta sus planes como profesionistas, además de que los tratamientos son sumamente agresivos, pudiendo ocasionarles una pérdida de fertilidad y otros efectos adversos al crecimiento del niño.

— Algunos tratamientos pueden ser carcinogénicos y el riesgo de que se desarrolle una segunda neoplasia en un paciente curado de un primer tumor es mayor en el niño que en las personas de mayor edad que han sufrido de algún cáncer.

I) DEFINICIONES BASICAS

En este apartado se presentan algunas definiciones básicas generales relacionadas con el tema.

NEOPLASIA:

La neoplasia se define como una masa anormal de tejido que prolifera sin relación con el tipo o la velocidad de crecimiento de los tejidos normales, y que es capaz de crecimiento independiente después de que cesa el estímulo que la ha producido. Las características de las neoplasias son: crecimiento celular excesivo, capacidad de invadir las estructuras vecinas y producir metástasis, además de que crecen por expansión y por infiltración, penetrando en los intersticios de los tejidos circunvecinos.

CANCER:

La palabra cáncer ha venido a significar una neoplasia maligna.

ANAPLASIA:

La anaplasia es el proceso mediante el cual se desarrolla una neoplasia maligna. Consiste en la transformación de células bien diferenciadas y funcionalmente normales en otras que presentan proliferación anormal, estructura irregular y metabolismo a tal grado alterado que pone en peligro la supervivencia del huésped.

TUMOR:

Por tumor se entiende toda masa o abultamiento anormal, cualquiera que sea su estructura o sitio. Los tumores se dividen en dos grandes categorías: tumor no neoplásico y tumor neoplásico.

TUMOR NEOPLASICO:

Son aquellos tumores que están formados por tejidos cuyas células muestran exagerada hiperplasia, es decir, multiplicación aumentada. A esto puede añadirse el aumento del tamaño de las células. Los tumores neoplásicos se dividen en dos grandes categorías:

TUMOR NEOPLASICO BENIGNO O NO CANCEROSO:

Está constituido por una proliferación de células que conservan las características del tejido que les da origen. Al crecer desplazan simplemente los tejidos vecinos.

TUMOR MALIGNO O CANCER:

Se caracteriza por un alto grado de reproducción celular y de tener un carácter invasor y aún destructivo de los tejidos o estructuras cercanas. Además, tiene tendencia a propagarse a distancia, lo que constituye las metástasis.

TUMOR EMBRIONARIO:

Se considera tumor embrionario a la neoplasia que se origina durante el desarrollo embrionario, fetal o postnatal temprano, a partir de un rudimento orgánico particular o de tejido aún inmaduro.

II) EL PROCESO DEL CANCER

Para empezar de alguna manera todo el recorrido que sigue el proceso del cáncer, de una manera totalmente sencilla y basada en el manual de Cancerología Básica, se tiene lo siguiente: la célula es la unidad fundamental de todo el organismo, es el sitio donde se realizan procesos fisiológicos que condicionan la vida del individuo. Esta tiene un periodo de duración variable, las que tienen más largo ciclo vital son las del sistema nervioso y muscular, y con una vida más corta, la de los epitelios de revestimiento. Ocupan un sitio, que aunque variable o fijo, está bien determinado, y llevan a cabo sus funciones en interrelación unas con otras, dentro de los sistemas y aparatos respectivos. El proceso de regeneración y de crecimiento de las células es armonioso en tiempo y en espacio, e indudablemente repercute en el estado de todo el conjunto celular del organismo.

Cuando el ritmo de crecimiento celular y de regeneración de las células no es normal sino que está acelerado, cuando se rompe el equilibrio, la relación entre estas células con las de una parte o el todo del organismo, esto es, cuando las células se multiplican en una proporción mayor tanto en el lugar como en el momento, sin satisfacer una necesidad orgánica, se constituye una neoplasia. Por lo anterior, la neoplasia queda constituida por células que se han reproducido en mayor proporción que lo normal, fenómeno que puede ocurrir en uno o varios sitios.

Las neoplasias se caracterizan por un crecimiento celular persistente y con tendencia a aumentar, que no se asemeja a los procesos inflamatorios conocidos, que no tiene finalidad defensiva ni utilidad evidente y que es en cierta forma autónomo en cuanto a los mecanismos del organismo huésped. El crecimiento es invasor y en ocasiones destructor de los tejidos circundantes y es capaz de determinar la muerte del enfermo.

Las neoplasias se dividen en dos categorías: las histológicamente benignas y las histológicamente malignas, éstas últimas constituyen el gran grupo de los llamados cánceres.

Las neoplasias benignas se caracterizan por su analogía con el tejido que les dio origen, por estar bien limitadas, por no presentar deformaciones celulares, por no afectar el estado general ni comprometer la vida del enfermo, en tanto que el cáncer, es decir la neoplasia maligna, se caracteriza por lo siguiente:

1) No son limitadas, ya que las células invaden tejidos y estructuras vecinas y aún emigran a sitios distantes siguiendo los intersticios celulares, la vía linfática, sanguínea, etc.

2) Presentan diferencias con las células del tejido que les dio origen, aún cuando tengan algún parecido con ellas, produciéndose atipias y monstruosidades.

3) El crecimiento y la multiplicación de las células es exagerado y atípico.

4) En su crecimiento invaden tejidos u órganos por la llegada y la proliferación de las células malignas y no por la transformación de las células normales de dichas estructuras en malignas.

Por lo tanto, las características fundamentales de los cánceres es que son proliferaciones celulares exageradas, autónomas, inútiles, nocivas y de etiología desconocida. Además, cabe hacer notar que si bien los cánceres son autónomos en cuanto a su crecimiento, no dejan de existir relaciones entre las neoplasias y el organismo, que aunque limitadas, son indudables.

III) EL CANCER INFANTIL (Una clasificación general de las neoplasias malignas del niño)

Dentro de todos los cánceres que sufre la población a nivel mundial, es importante destacar que existen diferencias claras en los cánceres sufridos por adultos y los cánceres que padecen los niños. La población infantil padece diversos tipos de cánceres o neoplasias malignas, los que, si se expusieran en su totalidad en esta parte del presente trabajo nos llevaría a mencionar demasiados tipos de cánceres, lo que incluso puede resultar tedioso y poco útil o práctico para los fines que persigue la presente. La clasificación que se toma es la realizada por Birch y Marsden. De acuerdo a ésta, sólo se explican doce grupos de cáncer o neoplasias malignas con sus rasgos generales sin ir más a fondo sobre las subdivisiones que pudieran presentar cada una de estas neoplasias. Además, no debemos perder de vista que el objetivo del estudio no es realizar un análisis exhaustivo sobre las diferentes neoplasias que atacan al infante, sino más bien analizar el comportamiento del proceso del cáncer pediátrico en su forma general, más que particular, ya que no se cuenta con los datos suficientes como para realizar un estudio sobre alguna neoplasia específica. Por lo tanto, se presentan las doce neoplasias con la intención de que se tenga el conocimiento a grandes rasgos sobre estas enfermedades, puesto que es el objetivo de esta sección, para comprender posteriormente la forma en que se realiza el análisis.

I. LEUCEMIA

La leucemia, llamada literalmente como "sangre blanca", es una enfermedad causada por los tejidos formadores de la sangre como son la médula ósea, las glándulas y el bazo. El problema radica principalmente cuando las células son puestas en libertad en la corriente sanguínea y existe una hiperproducción de células blancas, las cuales, en su mayoría no llegan a adquirir caracteres de madurez. Uno de los síntomas que presentan los pacientes de leucemia es la propensión a sangrar, debido a que la ebullición de las células blancas impide la formación de las plaquetas de la sangre, las cuales son necesarias para su coagulación. Otro aspecto que resulta interesante, según estudios realizados, es la conclusión de que la leucemia no es un proceso al azar, puesto que se han observado claramente grupos de riesgo, como es el caso de los gemelos monocigotos, donde, al padecer leucemia uno de ellos, la probabilidad de que el otro contraiga la enfermedad es muy alto. Otro grupo de alto riego son los pacientes con el síndrome de Down.

En la cuestión etiológica, el desarrollo de la leucemia ha tomado cuatro direcciones fundamentales:

–Susceptibilidad genética, ya que se ha encontrado una frecuencia relativamente elevada en la leucemia familiar. Existe una gran incidencia de leucemia en gemelos monocigotos y en pacientes con ciertos síndromes cromosómicos.

—En los factores ambientales, el potencial leucemógeno de las radiaciones ionizantes es bien conocido. Por ejemplo, se ha informado el aumento de la leucemia en individuos expuestos profesionalmente a radiación. Otro hecho es la alta incidencia de leucemia en supervivientes de las explosiones atómicas de Hiroshima y Nagasaki. Algunas personas expuestas a radioterapia presentan mayor probabilidad de sufrir leucemia aguda que la población general. Análisis recientes sugieren que ocurre un efecto leucemógeno significativo en niños expuestos in útero a dosis de 0.8 rads. También en individuos expuestos en forma crónica a sustancias químicas como el benceno se les ha desarrollado la enfermedad.

—Infecciones virales, pues existen pruebas de que ciertos virus causan leucemia en varias familias animales, como aves, roedores, bovinos y felinos. Sin embargo, hasta ahora no hay datos que vinculen en definitiva la infección viral con el desarrollo de leucemia en el hombre. La base de la etiología viral en la leucemia aguda en el ser humano se encuentra en dos líneas principales: la capacidad de los virus de producir leucemia en otras especies de mamíferos y la presencia de marcadores bioquímicos de infección viral en células leucémicas humanas. Pero el hecho de que un virus esté presente en células leucémicas humanas no necesariamente implica su participación etiológica, ya que también es posible que sea sólo un "pasajero". En todo caso, a pesar de la falta de datos concluyentes, existe la prueba circunstancial de esta relación así como el estímulo para investigar dicha posibilidad.

—Las alteraciones inmunitarias, las cuales se basan en el concepto de que al disminuir la inmunovigilancia normal del huésped se favorece el desarrollo de clonas malignas. Así, se ha observado que pacientes con enfermedades por inmunodeficiencia congénita presentan mayor riesgo de sufrir neoplasias malignas linforreticulares. Sin embargo, personas con inmunodeficiencia congénita también tienen anormalidades cromosómicas subyacentes, por lo que en estos casos el origen de la leucemia podría ser multifactorial.

II. LINFOMAS

Los linfomas son una categoría incluida dentro de las neoplasias malignas de la infancia. Este tipo de enfermedad resulta de especial interés debido a que se originan de células que por sí mismas son esenciales para mantener la integridad del organismo, ya que poseen la capacidad de reconocer y participar en la eliminación de materiales extraños. Las neoplasias linfoides se pueden originar en cualquier órgano o tejido, incluso en la circulación de las células del sistema inmunitario, tanto en el sistema vascular como a través de casi todos los tejidos del organismo, pero se desarrollan principalmente en los lugares donde el tejido linfoide está concentrado. Las manifestaciones clínicas son muy variadas y por lo general van relacionadas con los sitios anatómicos en donde se desarrollan y el tipo de estructuras que invaden. Generalmente el cuadro clínico se caracteriza por dolor, distensión abdominal y masa abdominal palpable. También es posible que presenten vómitos, diarrea y sangrado gastrointestinal.

Ahora bien, en cuestión etiológica, no existen evidencias claras que puedan dar pistas sobre las causas de esta enfermedad, puesto que se cree que existe una gran variedad de mecanismos que dan como resultado el surgimiento de los linfomas. Los principales sospechosos son, por un lado, los virus, y por otro la inmunodeficiencia, ya que se ha observado que individuos con inmunodeficiencia heredada presentan un alto riesgo de cáncer. Los linfomas son clasificados en Linfoma Hodgkin o Enfermedad de Hodgkin y Linfoma no Hodgkin.

—Enfermedad de Hodgkin

Es un proceso de tipo tumoral que afecta primariamente los ganglios linfáticos, son sumamente pequeños y se presentan en mayor número en ciertos sitios bien definidos, tales como los lados del cuello, las axilas, la ingle, la parte central del tórax y en la parte posterior y dentro de la cavidad abdominal. Los ganglios linfáticos sirven como receptores o depósitos a las células blancas de la sangre del tipo linfoide, las cuales se encuentran allí en gran número. Estas células actúan como barricadas contra la propagación de bacterias infecciosas a través del cuerpo, esto es por lo que en presencia de la infección los grupos de

ganglios linfáticos de una parte determinada del cuerpo se hinchan y a menudo se reblandecen. Se ha insistido en el carácter neoplásico de la enfermedad de Hodgkin tomando como base las siguientes consideraciones:

1) si no se trata se disemina por todo el sistema linfático y con el tiempo afecta al tejido no linfoide.

2) reacciona al tratamiento con agentes anticancerosos.

3) las células gigantes características de la imagen microscópica presentan derivación aneuploide y clona.

–Linfoma no Hodgkin

Las principales características de estos linfomas es que tienen un patrón de crecimiento difuso, son propensos a la invasión más allá del tejido ganglionar y tienen una amplia diseminación.

Por la parte etiológica podemos mencionar que se consideró que ciertos determinantes climatológicos, como la temperatura ambiental, la precipitación pluvial en determinado tiempo y la altitud de la región crearían las condiciones adecuadas para la aparición del linfoma en toda el Africa Tropical, además, también se sugirió un virus como factor causal.

III. TUMORES DEL SISTEMA NERVIOSO CENTRAL

Los tumores del sistema nervioso central son un grupo de neoplasmas, muchos de los cuales son únicos en niños. Estos pueden desarrollarse en cualquier edad, pero es más probable que ocurran entre los cinco y diez años. Los síntomas pueden ser efectos generales, vómitos, daños en la vista, alargamiento del cráneo, convulsiones, perturbaciones mentales. Por la parte etiológica no existen datos definitivos de la causa de los tumores del sistema nervioso central en los niños.

IV. NEUROBLASTOMA

El neuroblastoma es un tipo de tumor que se presenta en la niñez con mayor acentuación que en la edad adulta. Este tumor se deriva del sistema nervioso simpático y médula suprarrenal. Tiene la característica de crecer rápidamente y producir metástasis, además, que si a la enfermedad se le permite evolucionar de acuerdo a su naturaleza, es capaz de terminar la vida del paciente en pocas semanas, o a lo más en varios meses. Se dice que este tipo de tumor es embrionario porque se deriva de tejido en desarrollo que no ha alcanzado su diferenciación completa. Se tiene conocimiento de que muchos neuroblastomas están presentes al nacimiento o poco después, y la mayor parte de estas neoplasias se presentan en los primeros años de vida.

En lo que respecta a su etiología, se puede pensar en circunstancias externas que influyen en su aparición. Por ejemplo, de acuerdo a la recopilación de datos sobre este tema, se ha encontrado que hay una incidencia significativa menor del neuroblastoma de niños del Africa Tropical comparada con la observada en niños negros americanos. También existe la posibilidad de que el neuroblastoma tenga un carácter hereditario, puesto que se ha encontrado casos de gemelos idénticos y en hermanos con esta enfermedad.

V. RETINOBLASTOMA

Esta enfermedad es la más común de las enfermedades malignas de los ojos en el niño. Se deriva de la retina y tiene un carácter embrionario, ya que proviene de las capas nucleares de la retina en distintos periodos de diferenciación temprana. Esta neoplasia es una enfermedad esencialmente de la niñez. Se ha observado desde el nacimiento e incluso en el feto, pero resulta menos frecuente en niños mayores y es extremadamente rara después de la pubertad. La enfermedad puede atacar a uno o los dos ojos, pero por lo regular se ha encontrado que los tumores en ambos ojos son distintos e independientes y no presentan propagaciones o metástasis del uno al otro.

Los síntomas que presentan son pérdida de visión, pupilas dilatadas, masas alrededor del ojo, además de que la mayoría de estos niños parecen tener una inteligencia abajo de lo normal.

VI. TUMOR DE WILMS

Al tumor de Wilms también se le llama tumor mixto, ya que se encuentra formado por diferentes tipos de células. Esta enfermedad se presenta en la lactancia y en la edad temprana, rara vez va más allá de los primeros diez años de vida. El síntoma más precoz es la aparición de una masa o tumor en el abdomen o en el costado. En algunas ocasiones, el niño se puede quejar de dolor antes de que aparezcan signos poco corrientes. El dolor es de tipo cólico y puede estar unido al estreñimiento, además de que se pueden sentir cansados y débiles antes de que se haya descubierto el tumor, o incluso puede haber orina sanguinolenta.

En la parte etiológica, se ha encontrado en algunas familias anomalías congénitas que han sido relacionadas con el tumor de Wilms y en otras ocasiones se han encontrado anormalidades cromosómicas en niños con esta enfermedad.

VII. TUMORES HEPATICOS

El carcinoma hepático constituye la enfermedad maligna más frecuente, aunque en general las neoplasias malignas del hígado son raras. Los tumores hepáticos declinan en frecuencia después de un máximo durante el primer año de vida.

La exposición industrial al cloruro de vinilo recientemente ha sido comunicada como causa productora de angiosarcoma hepático. La noretandronola y otras hormonas andrógenas usadas en el tratamiento de algunas anemias han sido asociadas recientemente con tumores benignos y tumores malignos.

Los síntomas y signos suelen ser una masa indolora, firme, en el cuadrante superior derecho. La falta de apetito, pérdida de peso, fiebre o rara vez la ictericia pueden ser el síntoma inicial.

VIII. TUMORES OSEOS

Los tumores óseos malignos de la infancia no son numerosos, se pueden señalar los siguientes: a) osteosarcoma y condrosarcoma, b) Tumor de Ewing y c) Linfoma óseo (sarcoma de células reticulares). La lesión cística solitaria de huesos largos más común es el quiste óseo solitario, éste se sitúa centralmente por debajo de la epífisis y en la metáfisis. A medida que el hueso crece, el tumor se extiende y lo debilita, terminando por fracturarse. Los síntomas principales de los tumores óseos, ya sea primarios o metastásicos, son dolor y tumefacción. Una fractura puede ser la primera en llamar la atención a una zona de destrucción cortical por cáncer. Cabe señalar que la principal dificultad de las lesiones císticas de huesos largos en la infancia es la identificación de su naturaleza y su clasificación.

IX. RABDOMIOSARCOMA

Este tipo de tumores es originado por tejidos de tipo embrionario. Existe una mayor frecuencia en los primeros cinco años de vida, y aún es posible observarlo en niños con más edad, también es posible verlo en la adolescencia e incluso en la edad adulta joven. En cuanto al sexo, no existe una tendencia clara. Los sitios donde ocurre con mayor frecuencia, sobre todo durante la niñez, son partes determinadas como la órbita, la nasofaringe, el paladar, el oído medio, la vejiga, la vagina, el útero, la próstata, la zona paratesticular, la pelvis y los tejidos blandos del tronco y las extremidades.

En lo que respecta a su etiología, la mayor parte de los casos no presenta una tendencia familiar, sin embargo, ha habido ocasiones en que más de un hermano ha presentado el tumor. También existe la posibilidad de que haya una mayor frecuencia de rabdomiosarcoma en niños de condición baja y que han estado expuestos a contaminantes químicos.

X. TUMORES DE CÉLULAS GERMINALES

Este tipo de tumores son poco frecuentes en los niños. La edad de mayor incidencia tanto para tumores ováricos como para los testículos son los dos años, y hay una segunda elevación en la incidencia que comienza pasados los seis años en los tumores ováricos y pasados los catorce años en los tumores testiculares.

En el primer año de vida, el signo inicial de un tumor testicular suele ser una masa escrotal que a veces se descubre al nacer. No suele haber dolor ni signos inflamatorios al principio. Es raro que el primer signo sea una metástasis del proceso. En los niños mayores suele notarse una hinchazón paulatina del testículo afectado durante algunas semanas, junto con dolor y sensibilidad local exagerada en más de la mitad de los casos. En algunos enfermos, las primeras manifestaciones clínicas pueden deberse a metástasis en los ganglios linfáticos peritoneales o en los pulmones. En los tumores ováricos, los síntomas de comienzo más frecuentes son dolor, náuseas y vómitos. Algunos enfermos no tienen síntomas y casualmente se les encuentra una masa abdominal o una hinchazón de todo el abdomen.

De todos los tumores de células germinales del ovario, el 95% son teratomas quísticos benignos, pero hay una relación inversa entre la edad y la posibilidad de un proceso maligno, de tal modo que el 84% de los tumores de células germinales son malignos en las niñas menores de 10 años.

XI. CARCINOMAS

Los carcinomas son tumores o neoplasias malignas de origen epitelial, con anaplasia en mayor o menor grado y con capacidad de provocar metástasis a distancia en cualquier momento de su evolución. Puede ocurrir en cualquier parte del organismo donde haya epitelio.

XII. MAL DEFINIDOS

Tumores que por mala técnica, material insuficiente o por su naturaleza misma no es posible clasificar en un tipo específico.

IV) ANTECEDENTES PREVIOS A LA INVESTIGACION

Como antecedente a la presente investigación se tiene el trabajo realizado en el Hospital de Pediatría de centro Médico Nacional Siglo XXI sobre el cáncer pediátrico. Este se inicia con la recopilación de los casos diagnosticados con cáncer pediátrico en el D.F. para el período 1981-1991 en los principales hospitales que dedican su atención al infante como son:

1.- Hospital Infantil de México " Federico Gómez".

2.-Hospital de Pediatría del Centro Médico Nacional Siglo XXI y el Servicio de Hematología y Oncología Pediátrica del hospital general Centro Médico " La Raza" del Instituto Mexicano del Seguro Social (IMSS).

3.-Servicio de Hematología y Oncología Pediátrica del Hospital "20 de Noviembre" del Instituto de Seguridad Social al Servicio de los trabajadores del Estado (ISSSTE).

4.-Servicio de Hematología y Oncología Pediátrica del Hospital General de México de la Secretaría de Salud.

5.-Servicio de Hematología y Oncología del Instituto Nacional de Pediatría.

Con anterioridad a la revisión de los expedientes, fue creada una cédula con las siguientes variables de interés:

- Edad al diagnóstico del cáncer.
- Sexo.
- Tipo de neoplasia detectada.
- Fecha de inicio de síntomas.
- Peso al nacer.
- Tipo de parto.
- Antecedentes de haber padecido algún proceso infeccioso.
- Existencias de malformaciones en el niño.
- Antecedentes de familiares que hayan padecido algún tipo de cáncer.
- Edad de los padres al concebir a su hijo.
- Escolaridad de los padres y salario de los mismos.
- Lugar de nacimiento.
- Lugar de residencia.
- Fecha de diagnóstico de la enfermedad.
- Tiempo transcurrido entre inicio de síntomas y el diagnóstico de cáncer.
- Fecha de concepción, de nacimiento y de inicio de sintomatología.

La clasificación para las diferentes neoplasias está basada en la clasificación de Birch y Marsden, la ocupación de los padres tiene la clasificación de acuerdo a la Organización Internacional Uniforme de Ocupaciones. También cabe aclarar que según información hecha en el anterior trabajo, se hace hincapié a que los datos tienen cierta subestimación debido a la pérdida de información ocurrida por los sismos de 1985, el cual afectó a una de las principales instituciones (IMSS).

II. TABLAS DE CONTINGENCIA

En las ciencias sociales y médicas, dentro de una investigación, los datos son generalmente recolectados por medio de entrevistas o cuestionarios, o ambos. Después de este paso, la información obtenida se codifica e introduce en los archivos que se encuentran en la computadora, para posteriormente realizar su respectivo análisis. Como un primer punto, muchas veces se desea saber el grado de asociación que algunas variables guardan. Las tablas de contingencia son sumamente útiles en este caso, ya que implican toda una metodología, la cual involucra diversos estadísticos, que nos dan una medida sobre el grado de asociación entre las variables de estudio. Por todo esto, es que sin duda alguna es necesario hacer una revisión de la metodología señalada, puesto que ésta nos ayudará en el estudio del cáncer pediátrico en el D.F. El objetivo de utilizarlas es estudiar la asociación entre las variables relacionadas al cáncer pediátrico, así como el grado de asociación si es el caso. Otro aspecto que se debe aclarar es que no se desea ahondar demasiado en la teoría, ya que fundamentalmente, la tesis tiene un carácter más bien práctico y por eso es que se encuentra delimitada la información que aquí se presenta.

I) DEFINICIONES BASICAS

Para comprender de una forma más clara la metodología de las tablas de contingencia comenzaremos con las definiciones básicas:

POBLACION:

Una población o universo es la totalidad de unidades elementales acerca de las cuales se desea información.

OBSERVACION:

Indica cualquier clase de medida obtenida en la investigación en una unidad elemental.

VARIABLE CATEGORICA:

Es la variable con ciertas características de naturaleza cualitativa, además de que tiene separaciones bien marcadas entre sus posibles valores. Ejemplos de este tipo de variables son el sexo, lugar de nacimiento, tipo de cáncer del paciente, etc.

CATEGORIA DE UNA VARIABLE:

Son las posibles denominaciones en que puede estar la variable.

VARIABLE CATEGORIZADA:

Si consideramos a una variable continua, por ejemplo, la edad, es posible categorizar a esta variable reduciendo el número de valores que se pueden tener. Luego entonces, es posible definir una nueva variable, digamos, "generación", con los siguientes valores o categorías:

niños:	0	≤	edad <	15
adolescentes:	15	≤	edad <	18
jóvenes:	18	≤	edad <	45

TABLA DE CONTINGENCIA DE DOS DIMENSIONES:

Es cuando la tabla sólo se refiere a dos variables.

TABLA DE CONTINGENCIA MULTIDIMENSIONAL:

Cuando la tabla se refiere a tres o más variables.

Las tablas de contingencia son bastante útiles en la investigación cuando se desea saber si las variables son o no independientes. En caso de no ser independientes, nos ayudan a determinar el grado de asociación. Para llevar a cabo el estudio, la población debe estar clasificada en base a dos o más variables cualitativas, donde cada elemento de la población pertenece a sólo una categoría. La dimensión de estas tablas puede ser de dos o más. Ahora bien, dentro de las tablas de dos dimensiones, la tabla de 2x2 es la más sencilla, por lo que se comenzará la exposición con este tipo de tablas.

II) ASOCIACION E INDEPENDENCIA EN TABLAS DE 2 X 2

Supongamos que en nuestra población de estudio se han definido 2 variables (ya sean categóricas o categorizadas): A y B, cuyas categorías son A_1 y A_2, B_1 y B_2 respectivamente. Es decir, se tienen las siguientes combinaciones:

$$(A_1,B_1),\ (A_1,B_2),\ (A_2,B_1),\ (A_2,B_2).$$

A cada una de estas combinaciones se le llama celda, y se denota como f_{ij} a la frecuencia observada en la celda (A_i, B_j). Lo anterior se ilustra en la tabla 2.1 que se muestra enseguida.

	B_1	B_2	Total
A_1	f_{11}	f_{12}	f_{1o}
A_2	f_{21}	f_{22}	f_{2o}
	f_{o1}	f_{o2}	f_{oo}

Tabla 2.1
FRECUENCIAS DE UNA TABLA DE 2X2.

donde:

f_{io} = Total de las frecuencias que están dentro de la categoría A_i.
f_{oj} = Total de las frecuencias que están dentro de la categoría B_j.
f_{oo} = Total de todas las frecuencias de la tabla.

Matemáticamente, lo anterior se representa de la siguiente forma:

$$\begin{aligned} f_{i0} &= \Sigma_i f_{ij} \\ f_{0j} &= \Sigma_i f_{ij} \\ f_{00} &= \Sigma_j f_{i0} \\ f_{00} &= \Sigma_i f_{0j} \\ f_{00} &= \Sigma_i \Sigma_j f_{ij}, \end{aligned} \qquad \text{.....(2.1)}$$

Las principales preguntas que surgen a partir de este tipo de tablas son las siguientes:

1.-¿Existe o no independencia entre las variables A y B ?.

2.-Si no existe independencia, entonces ¿cuál es el grado de su relación o asociación?.

Por otro lado, las tablas pueden formarse de distintas maneras de acuerdo a la forma de clasificar a la población de estudio, lo que quiere decir que existe una estructura de la tabla. Esta se explica enseguida:

ESTRUCTURA DE LA TABLA

Una tabla de frecuencias puede provenir de distintos tipos de muestreo, por ejemplo:

I. Entrevistar f_{00} gentes y clasificarlas de acuerdo a sus respuestas a dos preguntas A y B. **Clasificación cruzada.**

II. Entrevistar f_{10} hombres y f_{20} mujeres y clasificarlas de acuerdo a sus respuestas a una pregunta (B). **Clasificación Prospectiva.**

II. Tomar f_{01} respuestas B_1, y f_{02} respuestas positivas a B_2 y clasificarlas de acuerdo a su categoría de la variable A, digamos, a su sexo. **Clasificación retrospectiva.**

PRUEBA DE INDEPENDENCIA DE A Y B

La idea de independencia es muy importante teóricamente en este tipo de análisis, ya que la no independencia es la principal razón que nos conduce a una investigación más profunda sobre la existencia de asociación entre las variables de estudio. En términos matemáticos, si A y B son independientes, se infiere que f_{11}/f_{01} será aproximadamente igual a f_{12}/f_{02}, y f_{11}/f_{10} será aproximadamente igual a f_{21}/f_{20}. A su vez, f_{21}/f_{01} será aproximadamente igual a f_{22}/f_{02} y f_{12}/f_{10} será aproximadamente igual a f_{22}/f_{20}. Si A y B no tienen relación alguna, se necesita una medida de comprobación de independencia.

Consideremos la siguiente tabla:

	B_1	B_2	Total
A_1	p_{11}	p_{12}	p_{10}
A_2	p_{21}	p_{22}	p_{20}
	p_{01}	p_{02}	p_{00}

Tabla 2.2
PROBABILIDADES DE UNA TABLA DE 2X2

donde:

p_{ij} = probabilidad teórica que corresponde a la celda (i,j), que está dentro de las categorías A_i y B_j.

$p_{ij} = p_{i0}p_{0j}$ y $f_{ij} = f_{00}\, p_{i0}\, p_{0j}$

$$\begin{aligned} p_{0j} &= \Sigma p_{ij} = p_{1j} + p_{2j} \\ p_{i0} &= \Sigma p_{ij} = p_{i1} + p_{i2} \\ p_{00} &= \Sigma p_{ij} = p_{11} + p_{12} + p_{21} + p_{22} = 1. \end{aligned} \quad \text{.....(2.2)}$$

Notemos que las ecuaciones (2.2) son análogas a las ecuaciones (2.1). Si A y B son independientes, entonces podemos esperar que:

$$\frac{p_{11}}{p_{01}} = \frac{p_{12}}{p_{02}} = \frac{p_{10}}{p_{00}} = \frac{p_{10}}{1} = p_{10}.$$

Al despejar p_{11} se llega a la igualdad 2.3

$$p_{11} = p_{10}\, p_{01}. \quad \text{.....(2.3)}$$

Además, se tiene que:

$$\frac{p_{11}}{p_{10}} = \frac{p_{21}}{p_{20}} = \frac{p_{01}}{p_{00}} = \frac{p_{01}}{1} = p_{01},$$

y al despejar p_{11} de la expresión anterior nos queda:

$$p_{11} = p_{01}\, p_{10}. \quad \text{.....(2.4)}$$

Generalizando:

Si A y B son independientes, entonces

$p_{ij} = p_{i0}\, p_{0j}$ con i=1,2; j=1,2. .....(2.5)

Además, como

$$\frac{p_{11}}{p_{01}} = \frac{p_{12}}{p_{02}}$$

entonces

$$p_{11}(p_{02}) = p_{12}(p_{01}), \quad \text{.....(2.6)}$$

y como $p_{0j} = \Sigma_i p_{ij}$, se obtiene lo siguiente:

$$p_{02} = \Sigma_i \; p_{i2} = p_{12} + p_{22},$$
$$p_{01} = \Sigma_i \; p_{i1} = p_{11} + p_{21}. \qquad \text{.....(2.7)}$$

Sustituyendo (2.7) en (2.6) nos queda que:

$$p_{11}\,(p_{12}+p_{22}) = p_{12}\,(p_{11}+p_{21}),$$
$$p_{11}p_{12} + p_{22}p_{11} = p_{12}p_{11} + p_{21}p_{12},$$
$$p_{22}p_{11} = p_{21}p_{12}.$$

Ahora, si se divide la igualdad anterior por $p_{21}p_{12}$ se tiene que

$$\frac{p_{22}p_{11}}{p_{21}p_{12}} = 1. \qquad \text{.....(2.8)}$$

La relación (2.8) involucra a las cuatro celdas de la tabla, cuyo término se le refiere con el nombre de razón de diferencia o razón de momios. Convencionalmente se le designa con el símbolo griego ψ (psi) y es la unidad únicamente si A y B son independientes, el estimador de p_{ij} es:

$$\hat{p}_{ij} = \frac{f_{ij}}{f_{00}}. \qquad \text{.....(2.9)}$$

Si despejamos f_{ij} de (2.9), nos queda la siguiente expresión:

$$f_{ij} = f_{00}\hat{p}_{i0}\hat{p}_{0j}.$$

Al sustituir $\hat{p}_{i0} = \frac{f_{i0}}{f_{00}}$ y $\hat{p}_{0j} = \frac{f_{0j}}{f_{00}}$ en la expresión anterior, nos queda lo siguiente:

$$f_{ij} = f_{00}\left(\frac{f_{i0}}{f_{00}}\right)\left(\frac{f_{0j}}{f_{00}}\right) = \frac{f_{i0}f_{0j}}{f_{00}},$$

por lo que se llega a la siguiente igualdad:

$$\frac{f_{ij}}{f_{00}} = \frac{f_{i0}f_{0j}}{f^2_{00}}. \qquad \text{.....(2.10)}$$

Si A y B son independientes y se tiene un total de f_{00} observaciones en la tabla, entonces la frecuencia más probable en la celda (i,j) será :

$$c_{ij} = f_{00}\hat{p}_{ij} = \frac{f_{i0}f_{0j}}{f_{00}}.$$

La prueba de bondad de ajuste que usualmente se hace está basada en la x^2, donde se hace una comparación entre las frecuencias observadas y las esperadas. Se necesita probar la hipótesis de que $p_{ij} = p_{i0}p_{0j}$. Por lo tanto, se tiene la hipótesis nula :

$$H0 : \quad p_{ij} = p_{i0}\, p_{0j}.$$

Para realizar la prueba de bondad de ajuste, se utiliza el estadístico:

$$X^2 = \sum \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \quad i,j = 1,2.$$

el cual, si las variables son independientes, se distribuye aproximadamente como una X^2con 1 grado de libertad. Para el caso de la tabla de 2x2 se tiene el siguiente procedimiento:

a) Calcular $X^2 = \dfrac{f_{00}\,(f_{11} f_{22} - f_{12} f_{21})^2}{f_{10} f_{20} f_{01} f_{02}}$.

b) Comparar el valor de X^2 contra el valor en tablas de la distribución X^2_1.

Si el valor de X^2 es menor que el de X^2_1, entonces no se puede rechazar la hipótesis nula de independencia.

MEDIDAS DE ASOCIACION EN TABLAS DE 2X2

Si después de realizar la prueba de independencia ésta se rechaza, podemos pensar que existe alguna asociación entre las variables, por lo que es conveniente tener una medida que nos indique su magnitud. Entre las medidas propuestas se encuentran las siguientes:

Q DE YULE

Yule (1900) sugiere el coeficiente de asociación:

$$Q = \frac{f_{11} f_{22} - f_{12} f_{21}}{f_{11} f_{22} + f_{12} f_{21}}. \qquad \text{.....(2,11)}$$

Suponiendo que la frecuencia total f_{00} es razonablemente grande, entonces la distribución se comporta como una normal con varianza:

$$\frac{1}{4}(1-Q^2)^2\left(\frac{1}{f_{11}} + \frac{1}{f_{12}} + \frac{1}{f_{21}} + \frac{1}{f_{22}}\right),$$

por lo que es posible obtener un intervalo de confianza. El rango de Q es (-1, 1), y los valores extremos (negativo o positivo) corresponden a una asociación completa. El cero a su vez, corresponde a una no asociación (independencia).

EL ODDS RATIO (RAZON DE MOMIOS)

La razón de momios está basada en la idea de que si A y B están asociados, entonces el valor de la medida de asociación debe ser inalterado por un incremento en el número de observaciones; si cada una de las frecuencias de las celdas fueran repentinamente dobladas, no implica que la medida de asociación será afectada. Simpson (1951) y Edwards (1963) defendieron la idea de que la magnitud de una medida de asociación no es afectada por el orden en el cual las categorías de la variable son escritas. Los dos argumentos anteriores inducen a la estadística:

$$C = \frac{f_{11}\, f_{22}}{f_{12}\, f_{21}}.$$

La función C es análoga a la definida por la ecuación (2.8) y posee algunas propiedades. Q está en función de la medida C, ya que podemos reescribir la definición (2.11) de la siguiente manera:

partiendo de la expresión:

$$Q = \frac{f_{11}\, f_{22} - f_{12}\, f_{21}}{f_{11}\, f_{22} + f_{12}\, f_{21}}.$$

y al dividir al numerador y denominador por f12 f21 nos queda:

$$Q = \frac{\frac{f_{11}f_{22} - f_{12}f_{21}}{f_{12}f_{21}}}{\frac{f_{11}f_{22} + f_{12}f_{21}}{f_{12}f_{21}}} = \frac{\frac{f_{11}f_{22}}{f_{12}f_{21}} - \frac{f_{12}f_{21}}{f_{12}f_{21}}}{\frac{f_{11}f_{22}}{f_{12}f_{21}} + \frac{f_{12}f_{21}}{f_{12}f_{21}}},$$

$$Q = \frac{\frac{f_{11}f_{22}}{f_{12}f_{21}} - 1}{\frac{f_{11}f_{22}}{f_{12}f_{21}} + 1}.$$

Por lo tanto:

$$Q = \frac{C - 1}{C + 1}.$$

Es posible encontrarnos con el inconveniente de que alguna frecuencia sea 0, por lo que se sugiere la medida ψ:

$$\psi = \frac{\left(f_{11} + \frac{1}{2}\right)\left(f_{22} + \frac{1}{2}\right)}{\left(f_{12} + \frac{1}{2}\right)\left(f_{21} + \frac{1}{2}\right)}. \qquad \text{.....(2.12)}$$

El rango de ψ es de $(0, \infty)$. La asociación corresponde al valor 1, por lo tanto es difícil explicar por asimetría. Con $\text{Log}(\psi)$ se tiene un rango de $(-\infty, \infty)$. El valor 0 se refiere a una no asociación o independencia.

MEDIDAS λ (lambda) DE GOODMAN Y KRUSKAL

Goodman y Kruskal (1954) sugieren varias medidas con una interpretación probabilística, éstas son λ_a, λ_b y λ, las cuales están basadas en el trabajo de Guttman (1941). Dichas medidas también serán expuestas más adelante para el caso de tablas de IxJ.

El estadístico λ_b involucra una comparación de las dos situaciones siguientes:

-Si se escoge aleatoriamente un individuo de la población de estudio, y suponemos la categoría de B a la cual pertenece, ya sea (a) sin ninguna información adicional, o (b) conociendo su categoría de A.

-Si las categorías de A y B son totalmente incorrelacionadas, entonces no se tendrá una mejora en la segunda situación mayor que en la primera.

Si al conocer la categoría de B se nos facilita el camino para dar una correspondencia o categoría correcta de A, entonces A y B deben estar correlacionadas. Si nuestra probabilidad de tener correcta la suposición no es mejorada, entonces A y B están incorrelacionadas. La medida de Guttman es una simple función de un relativo mejoramiento en nuestra suposición de la correspondencia de cierta variable. La medida λ_b cuantifica la mejora en nuestra suposición de la categoría B como el decremento relativo en la probabilidad de error entre las dos suposiciones mencionadas anteriormente, bajo el supuesto de que nuestra elección es la más grande de las categorías B en cada ocasión. El estadístico λ_b está dado por la siguiente expresión:

$$\lambda_b = \frac{\sum_{i=1}^{I} f_{im} - f_{0m}}{f_{00} - f_{0m}}$$

donde:

f_{im} = a la frecuencia mayor de la fila i,
f_{0m} = al total mayor de las columnas,
f_{00} = al total de todas las frecuencias de la tabla.

Cuando estamos interesados en el incremento de la probabilidad en la predicción de la categoría A, dado que la categoría B es conocida, se utiliza la medida

$$\lambda_a = \frac{\left(\sum_{j=1}^{J} f_{mj} - f_{m0} \right)}{f_{00} - f_{m0}},$$

donde:

f_{mj} = a la frecuencia mayor de la columna j,
f_{m0} = al total mayor de las filas,
f_{00} = al total de todas las frecuencias de la tabla.

En algunas ocasiones no tiene sentido enfocar nuestra atención sobre alguna variable como dependiente de la otra. Si nosotros asumimos que A es tan importante para B, como B para A, entonces, se tiene la siguiente medida:

$$\lambda = \frac{\left(\sum_{i=1}^{I} f_{im} - f_{0m}\right) + \left(\sum_{j=1}^{J} f_{mj} - f_{m0}\right)}{2f_{00} - f_{m0} - f_{0m}}.$$

MEDIDAS τ DE GOODMAN Y KRUSKAL

Además de las medidas predictivas lambda, Goodman y Kruskal (1954) sugieren la medidas τ. Para el caso de tablas de 2X2, Las tres medidas τ (τ_a, τ_b, τ_{ab}) son igual a:

$$\frac{(f_{11} f_{22} - f_{12} f_{21})^2}{f_{10} f_{20} f_{01} f_{02}}. \qquad \text{.....(2.13)}$$

Notemos que esta medida es la misma que la de la X^2, lo único que se hizo fue dividir por f_{00}.

MEDIDAS BASADAS DIRECTAMENTE SOBRE X^2

Porque la prueba básica de independencia es la de la X^2, se ha usado el estadístico como una medida de asociación. Sin embargo, su rango es de 0 a α, el cual no está acorde con el rango tradicional de -1 a 1. Varias transformaciones se han sugerido, las cuales incluyen las siguientes:

$$\phi = \left(\frac{X^2}{f_{00}}\right)^{1/2},$$

$$C = \left(\frac{X^2}{X^2 + f_{00}}\right)^{1/2}.$$

Al estadístico ϕ en algún tiempo se le refirió como la raíz cuadrada del coeficiente de contingencia y el cuadrado de ϕ, ϕ^2 es igual al valor τ dado en (2.13). Al estadístico ϕ se le puede interpretar como un coeficiente de correlación. C es llamado el coeficiente de contingencia.

III) ASOCIACION E INDEPENDENCIA EN TABLAS DE I X J

Ahora ya consideremos la situación general de A y B, donde se tienen I categorías posibles para la variable A, las cuales se denotarán por A_1, A_2,...A_I y J posibles categorías de la variable B dadas por $B_1,B_2,...B_J$. Los datos son las frecuencias observadas en un total de I x J celdas. La pregunta sigue siendo la misma, existe o no independencia entre las variables A y B ?. Lo anterior se ilustra con la tabla 2.3.

	B_1	B_2		B_j	TOTAL
A_1	f_{11}	f_{12}	.....	f_{1j}	f_{1o}
A_2	f_{21}	f_{22}		f_{2j}	f_{2o}
:				:	:
:				:	:
A_i	f_{i1}	f_{i2}	.....	f_{ij}	f_{io}
TOTAL	f_{o1}	f_{o2}	.....	f_{oj}	f_{oo}

Tabla 2.3
FRECUENCIAS OBSERVADAS EN UNA TABLA DE IXJ

Si A y B son independientes, se espera que :

$$\frac{f_{ij}}{f_{oj}} \cong \frac{f_{io}}{f_{oo}} \quad \forall \; i,j$$

Al reescribir a las f_{ij} por p_{ij} (sus probabilidades) se tiene que:

$$\frac{p_{ij}}{p_{oj}} \cong \frac{p_{io}}{p_{oo}} = \frac{p_{io}}{1} = p_{io}$$

$$p_{ij} = p_{io}\, p_{oj} \quad \text{para } i = 1,2,..I; \quad j = 1, 2,...J. \qquad (2.14)$$

Como el valor de p_{ij} se desconoce, la proporción f_{ij}/f_{oo} estima las probabilidades. Estos estimadores nos pueden ayudar a deducir las frecuencias esperadas de las celdas. Si A y B son independientes se tiene la siguiente expresión:

$$e_{ij} = \frac{f_{io}\, f_{oj}}{f_{oo}}, \qquad (2.15)$$

De acuerdo a la expresión 2.15, la suma: $e_{1j} + e_{2j} + ... e_{ij}$ puede ser escrita como:

$$\frac{f_{1o}\, f_{oj}}{f_{oo}} + \frac{f_{2o}\, f_{oj}}{f_{oo}} + ...+ \frac{f_{io}\, f_{oj}}{f_{oo}} = \frac{f_{1o}f_{oj} + f_{2o}f_{oj} + ...f_{io}f_{oj}}{f_{oo}} = \frac{(f_{1o} + f_{2o} + ...f_{io})f_{oj}}{f_{oo}},$$

Notemos que lo que se encuentra dentro de los paréntesis es f_{oo}, por lo que la expresión anterior se reduce de la siguiente forma:

$$\frac{f_{oo}}{f_{oo}}\, f_{oj} = f_{oj}.$$

Si se conocen I-1 de las entradas de las primeras J-1 columnas de los valores esperados y el total de los marginales, es posible deducir las restantes (I+J-1) entradas. Esto significa que se tendrán (I-1)(J-1) grados de libertad para la prueba de bondad de ajuste. El estadístico X^2 para una tabla de I X J es:

$$X^2 = \Sigma_i \Sigma_j \frac{(f_{ij}-e_{ij})2}{e_{ij}} \quad \text{.....(2.16)}$$

Las distribución de X^2 es aproximadamente una X^2 con (I-1)(J-1) grados de libertad.

MEDIDAS DE ASOCIACION PARA TABLAS DE I X J

De forma alternativa, después de realizar la prueba de independencia entre dos variables, y encontrar dependencia, es posible obtener una simple y rápida cuantificación de ésta. Para ello se cuenta con algunas medidas de asociación, aunque muchas no tienen una interpretación probabilística directa.

Existen dos grandes clases de variables que son utilizadas en los estudios sobre tablas de contingencia: variables nominales, las cuales no cambian en sus categorías; por ejemplo, el tipo de cáncer tendría sus categorías en leucemia, retinoblastoma, etc. La segunda clase de variable es la ordinal y se refiere a una variable cuyas categorías deben ser ordenadas, por ejemplo, la edad con categorías 0-4, 5-9, 10-14. Al considerar el orden tenemos más información, y por lo general, una categoría ordenada es la categorización de una variable continua.

MEDIDAS DE ASOCIACION USANDO λ_b

Las tres medidas λ (λ_a, λ_b y λ) son muy similares en cuanto a su caracter. El estadístico λ_b involucra una comparación de las siguientes dos situaciones; un individuo se escoge aleatoriamente de la población de estudio y suponemos la categoría B a la cual peretenece, ya sea:

a) Sin que se nos proporcione una información adicional.

b) Proporcionándonos su categoría de A.

Si las correspondencias de A y B son independientes, no nos mejora la segunda situación más que la primera, caso contrario cuando si son dependientes, proporcionándonos una mejora la segunda situación. La medida λ_b cuantifica esta mejora como un decremento relativo en la probabilidad de error en la suposición de la categoría de B entre las dos situaciones, bajo el supuesto de que nuestra suposición consistirá en seleccionar las categorías de B más grandes en cada ocasión. Es decir, nos da una asociación predictiva.

El estadístico es :

$$\lambda_b = \frac{\sum_{i=1}^{I} f_{im} - f_{0m}}{f_{00} - f_{0m}}.$$

Donde:

f_{im} = al valor más grande de las celdas en la i-ésima fila de la tabla,
f_{0m} = al total mayor de los marginales por columna,
f_{00} = al total de las frecuencias de la tabla.

MEDIDAS DE ASOCIACION USANDO λ_a o λ

Si ahora se conoce el valor de la categoría B y estamos interesados en conocer la categoría de A, el incremento efectivo de predicción está dado por el estadístico:

$$\lambda_a = \frac{\left(\sum_{j=1}^{J} f_{mj} - f_{m0}\right)}{f_{00} - f_{m0}},$$

donde:

f_{mj} = al valor más grande de las celdas en la j-ésima columna de la tabla,
f_{m0} = al total mayor de los marginales por fila,
f_{00} = al total de las frecuencias de la tabla.

Si asumimos que es de igual interés A y B entonces la medida indicada es:

$$\lambda = \frac{\left(\sum_{i=1}^{I} f_{im} - f_{0m}\right) + \left(\sum_{j=1}^{J} f_{mj} - f_{m0}\right)}{2f_{00} - f_{m0} - f_{0m}}.$$

MEDIDAS DE ASOCIACION BASADAS SOBRE LA X^2

De las medidas de asociación utilizadas en tablas que contienen variables de tipo nominal u ordinal, o ambas, se tiene en primer lugar al estadístico:

$$\varphi = \left(\frac{X^2}{f_{00}}\right)^{1/2},$$

donde, desafortunadamente φ puede exceder de la unidad. Una modificación propuesta por Pearson (1904) es el coeficiente de contingencia:

$$C = \left(\frac{X^2}{X^2 + f_{00}}\right)^{1/2},$$

el cual tiene la limitación de que C<1. Esto fue resuelto por Cramer (1946) quien propone el coeficiente V:

$$V = \left(\frac{X^2}{f_{00}\,\min((I-1),(J-1))}\right)^{1/2}.$$

Otra medida más, basada en la X^2 es la T de Tschuprow:

$$T = \left(\frac{X^2/f_{00}}{\sqrt{(I-1)(J-1)}}\right)^{1/2}.$$

Todas estas medidas tienen como objeto encontrar una función de X^2 limitada al rango (0,1).

MEDIDAS τ DE GOODMAN Y KRUSKAL

En lugar de predecir siempre la categoría más probable, con las medidas τ las distintas categorías son predecibles en proporción al total de sus categorías observadas. El estadístico es :

$$\tau_b = \frac{\sum \sum_{ij} ((f_{00} f_{ij} - f_{i0} f_{0j})^2 / f_{i0})}{f_{00}(f_{00}^2 - \sum_j f^2_{0j})}.$$

De forma equivalente se tienen las las medidas τ_a y τ.

COEFICIENTE DE INCERTIDUMBRE

Un desarrollo para producir una medida interpretable de asociación basada en la proporción de la probabilidad de la ji- cuadrada es el coeficiente de incertidumbre, este nos indica una reducción en la incertidumbre de acerca de la categoría de la variable A dado que conocemos la categoría de la variable B o viceversa, por ejemplo, U_A nos indica la reducción de incertidumbre en la categoría A si conocemos la categoría de B, los estadísticos son los siguientes:

Con la variable A dependiente:

$$U_A = \frac{U(A) + U(B) - U(AB)}{U(A)}.$$

Con la variable B dependiente:

$$U_B = \frac{U(A) + U(B) - U(AB)}{U(B)},$$

y para el caso asimétrico:

$$U = \frac{U(A) + U(B) - U(AB)}{U(A)+U(B)},$$

donde:

$$U(A) = - \sum_{i=1}^{I} \frac{f_{i0}}{f_{00}} \log \frac{f_{i0}}{f_{00}},$$

$$U(B) = - \sum_{j=1}^{J} \frac{f_{0j}}{f_{00}} \log \frac{f_{0j}}{f_{00}},$$

$$U(AB) = - \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{f_{ij}}{f_{00}} \log \frac{f_{ij}}{f_{00}} \qquad \text{para } f_{ij} > 0.$$

MEDIDAS DE ASOCIACION PARA TABLAS ORDINALES

A continuación se describirán tres medidas de asociación. La base de estas medidas es que las I categorías ordenadas de la variable A representan I posibles rangos para el total de f_{oo} observaciones. Estas observaciones son designadas por categoría 1 a la variable con el rango más grande que la categoría 2 y así sucesivamente, con su correspondiente rango en la variable B. Con una asociación positiva entre A y B nosotros podemos esperar que una observación que tiene un rango alto en la variable A también lo tenga para la variable B, y uno que tenga un rango bajo para la variable A también lo tenga en la variable B.

Consideremos un par típico de observaciones, una perteneciente a la celda (i,j), con la categoría i de la variable A y la categoría j de la variable B. La otra observación perteneciente a la celda (i', j').

Las medidas de asociación de tipo ordinal son todas funciones de las siguientes cuatro funciones:

S= a la suma o total de los productos de cada par de observaciones para los cuales se cumple que $i>i'$ y $j>j'$ o $i<i'$ y $j<j'$.

D= a la suma o total de los productos de cada par de observaciones para los cuales se cumple que $i>i'$ y $j<j'$ o $i<i'$ y $j>j'$.

T_a = a la suma o total de los productos de cada par de observaciones para los cuales se cumple que $i=i'$.

T_b = a la suma o total de los productos de cada par de observaciones para los cuales se cumple que $j=j'$.

Cuando existe una fuerte asociación entre las variables A y B, el número S es muy grande y el número D es muy pequeño.

γ DE GOODMAN Y KRUSKAL

Goodman y Kruskal (1954) propusieron la siguiente medida:

$$\gamma = \frac{S - D}{S + D}.$$

Esta medida tiene una interpretación probabilística directa: Si las variables A y B son independientes, entonces, en promedio se tendrá al valor de 0 para γ. Sin embargo, si $\gamma = 0$ no necesariamente implica que A y B sean independientes. El rango de valores para γ es de -1 a 1. La distribución γ es aproximadamente una normal.

ESTADISTICO τ DE KENDALL

Es una medida de correlación entre dos conjuntos de rangos. Esto se puede adaptar para la forma general de las tablas de contingencia de IxJ, con categorías ordenadas. Las frecuencias marginales en la tabla son el número de observaciones vinculadas en los diferentes rangos. La medida Tau de Kendall está basada sobre la diferencia de S-D. Para obtener una medida de dicha diferencia, ésta puede ser estandarizada en el rango de -1 a 1. Diferentes métodos de estandarización han dado origen a los estadísticos Tau:

$$\tau_a = \frac{2(S-D)}{f_{00}(f_{00}-1)},$$

$$\tau_b = \frac{2(S-D)}{\sqrt{(S+D+T_a)(S+D+T_b)}},$$

$$\tau_c = \frac{2m(S-D)}{f^2_{00}(m-1)}.$$

Donde:

$$m = \text{mínimo } (I,J).$$

La estadística τ_a es usada comúnmente como una medida de correlación de rangos. Esto no es aplicable para tablas de contingencia donde se asume que no hay observaciones relacionadas o vinculadas. Los otros dos estadísticos son usados como medidas de asociación en tablas ordenadas. τ_b sólo puede tener el valor de +1 para una tabla cuadrada. Por su parte, τ_c puede obtener estos valores extremos aparte de un pequeño efecto que se produce cuando f_{00} no es múltiplo de m. El principal problema de estas dos últimas medidas es que no tienen una interpretación probabilística clara.

D DE SOMERS

Somers (1962) sugiere una medida de asociación para tablas de contingencia con categorías ordenadas conforme al caso asimétrico, donde se tiene una variable explicatoria y una dependiente. Es decir, cuando una de las variables B, puede ser vista como dependiente de la otra variable A, o viceversa. Los estadísticos son los siguientes:

$$D_A = \frac{2(S-D)}{f^2_{00} - \sum_{j=1}^{J} f^2_{0j}},$$

$$D_B = \frac{2(S-D)}{f^2_{00} - \sum_{i=1}^{I} f^2_{i0}},$$

$$D = \frac{(S-D)}{f^2_{00} - \sum_{j=1}^{J} f^2_{0j} + f^2_{00} - \sum_{i=1}^{I} f^2_{i0}}.$$

COMPARACION DE LAS MEDIDAS DE ASOCIACION

El propósito de una medida de asociación es que ésta nos provea de una respuesta numérica sobre el grado de asociación que se dá en las variables de estudio. Cabe notar que diferentes medidas de asociación enfocan su atención sobre diferentes aspectos de la relación entre las variables. Kruskal (1958) remarca que no hay razón para usar sólo una medida de asociación, ya que diversas medidas nos darán información correspondiente a los diferentes tipos de asociación.

La elección para las clasificaciones de tipo nominal está en tres tipos de estadísticos: los basados en la X^2, los que indican una reducción proporcional en el error de predicción de cierta categoría cuando se conoce la otra, y las medidas de incertidumbre. El primer grupo no tiene en realidad una interpretación inmediata. El segundo si la tiene, pero con una desventaja; a saber, si el valor del modelo cae únicamente en una columna o fila las medidas resultantes serán cero. Por su parte, el coeficiente de incertidumbre está basado en la probabilidad de la proporción de la X^2, y tiene una interpretación probabilística.

Cada una de las medidas expuestas son importantes en los diferentes aspectos de la asociación entre las variables. Por lo tanto, se puede elegir λ de Goodman y Kruskal (para datos nominales), o γ (para datos ordinales) si las dos variables son de igual importancia. De la misma forma, se puede elegir λ_b de Goodman y Kruskal o d_{ba} de Somers si la variable B depende de la variable A. Sin embargo, todas estas medidas no son más que un simple cálculo preliminar para un análisis más detallado.

En realidad, el investigador es el que debe tomar la decisión sobre qué medida de asociación utilizar. Para ello ha de tenerse en cuenta el buen juicio y el conocimiento sobre el fenómeno del que se trate. Además, es importante mencionar que cualquiera que sea la medida elegida, ésta nos dirá lo mismo con respecto al grado de asociación que guardan las variables en estudio.

III. SERIES DE TIEMPO

En diversas ramas de la ciencia existen fenómenos que es necesario analizar y que se adecúan a la metodología de las series de tiempo. Entre ellas encontramos a la economía, física, demografía, mercadotecnia, etc. En este trabajo se analiza a la rama de la ciencia más generosa y dedicada al servicio del ser humano, es decir, a la medicina, en específico al cáncer pediátrico del D. F.

Como se verá, una serie de tiempo es simplemente una colección de datos secuenciales a través del tiempo, y una característica sumamente importante de estas series, es el hecho de que las observaciones sucesivas son por lo regular no independientes. A partir de estos datos, se busca predecir valores futuros en base a observaciones pasadas. El objetivo de hacer una revisión de esta metodología es lograr una aplicación adecuada al fenómeno de interés, el cáncer pediátrico, por lo que se toma en consideración a las series de tiempo discretas. La metodología en sí es una parte de la Estadística, la cual, en su análisis utiliza modelos probabilísticos llamados promedios móviles y modelos autorregresivos. Para comenzar el análisis como primer punto se considera a la gráfica de las observaciones contra el tiempo, ya que de esta forma es posible visualizar factores importantes en la serie como son la media, la varianza, la existencia de estacionalidad, o tal vez, alguna discontinuidad o algún dato que sobresalga bastante en nuestra serie.

El análisis de datos basados en series de tiempo, tiene toda una historia, según la cual fue criticada por ser sumamente laboriosa. Gracias al avance computacional hoy en día existen diversos paquetes estadísticos que hacen de esto una labor relativamente sencilla. Por tal motivo, es que resulta conveniente utilizar la metodología de series de tiempo para el estudio de fenómenos que así lo requieran, en este caso, el cáncer pediátrico.

I) DEFINICIONES BASICAS

En esta parte se expondrán los principales conceptos relacionados con la metodología de series de tiempo. La idea es comprender de una forma clara y sencilla la teoría concerniente, por lo cual se tienen las definiciones:

FUNCION DE AUTOCORRELACION (ACF):

La función de autocorrelación mide la correlación entre un valor observado y_t y otro y_{t+k}, que están separados por un intervalo de longitud k.

FUNCION DE AUTOCORRELACION PARCIAL (PACF):

Es la función que mide el efecto parcial de y_{t-k} para explicar el comportamiento de y_t en un modelo que ya incluye $y_{t-1},...y_{t-(k-1)}$.

PRINCIPIO DE PARSIMONIA:

Este principio pretende encontrar el modelo más representativo con el menor número posible de parámetros.

REALIZACION:

A partir de un fenómeno, es posible tomar una secuencia de observaciones de éste, siendo el conjunto de observaciones una realización de dicho fenómeno.

SERIE DE TIEMPO:

Es una colección de datos generados secuencialmente a través del tiempo, cuyas observaciones sucesivas son generalmente dependientes entre sí.

SERIE DE TIEMPO DISCRETA:

Una serie de tiempo discreta en el parámetro tiempo es aquélla que se obtiene con registros instantáneos o promedios, o valores acumulados, es decir, las observaciones se toman únicamente de un tiempo específico, generalmente con intervalos iguales.

II) ANTECEDENTES

OBJETIVOS DEL ANALISIS DE SERIES DE TIEMPO

Cualquier teoría lleva consigo, de una forma implícita razones importantes que la justifican, pués de alguna manera la teoría nos servirá para hacer una aplicación práctica. A las series de tiempo se les puede asociar una gran variedad de aplicaciones, pero en cada una de éstas llevará por lo general los principales objetivos de su análisis, como son:

1.-DESCRIPCION:

Enunciar el comportamiento del proceso para obtener una visualización clara de éste. Para ello se utilizan las principales características de la gráfica, como son su variación estacional, tendencia, varianza, etc., lo cual ayuda a comprender mejor al proceso a analizar.

2.-EXPLICACION:

La explicación tiene como objeto realizar un análisis de tipo causa-efecto, ya sea al relacionar un fenómeno con el medio en que se desarrrolla, o cuando la variación de una serie de tiempo explica la variación de otra serie.

3.-PREDICCION O PRONOSTICO:

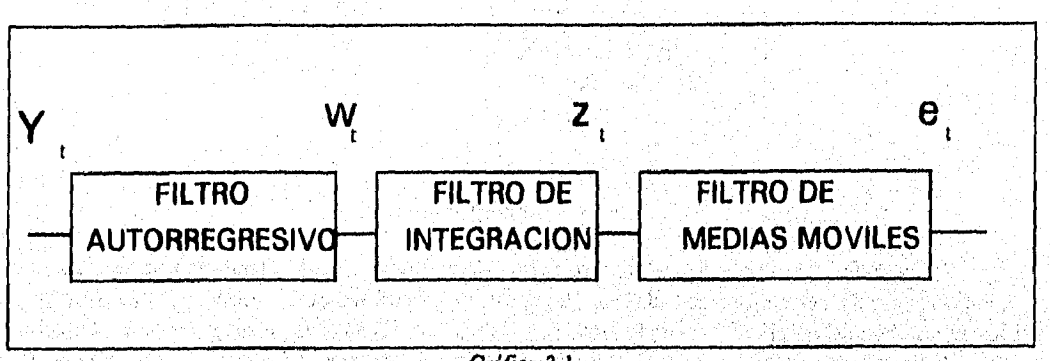
A partir de una serie de tiempo observada se pueden predecir los valores futuros de la serie con el fin de tomar decisiones o prever comportamientos futuros.

4.-CONTROL:

Obtener pronósticos muchas veces tiene como objeto realizar acciones de control, es decir, modificar el comportamiento del proceso con el fin de obtener algún beneficio.

METODOLOGIA DE BOX - JENKINS

Uno de los métodos para ajustar series de tiempo es el de Box- Jenkins, el cual consiste en extraer los movimientos predecibles de los datos observados. La serie de tiempo se descompone en varios componentes, llamados "filtros", hasta obtener residuales no predecibles cuyo comportamiento tiene poca influencia en el resultado final. Esta metodología hace uso principalmente de tres filtros lineales: el autorregresivo, el de integración y el de medias móviles, como se ilustra enseguida:



Gráfica 3.1

El proceso iterativo de Box- Jenkis para construir los modelos lineales de series de tiempo consiste de los siguientes pasos:

1. Identificación de las especificaciones preliminares del modelo, es decir, examinar los datos para de esta forma identificar al que parezca adecuado.

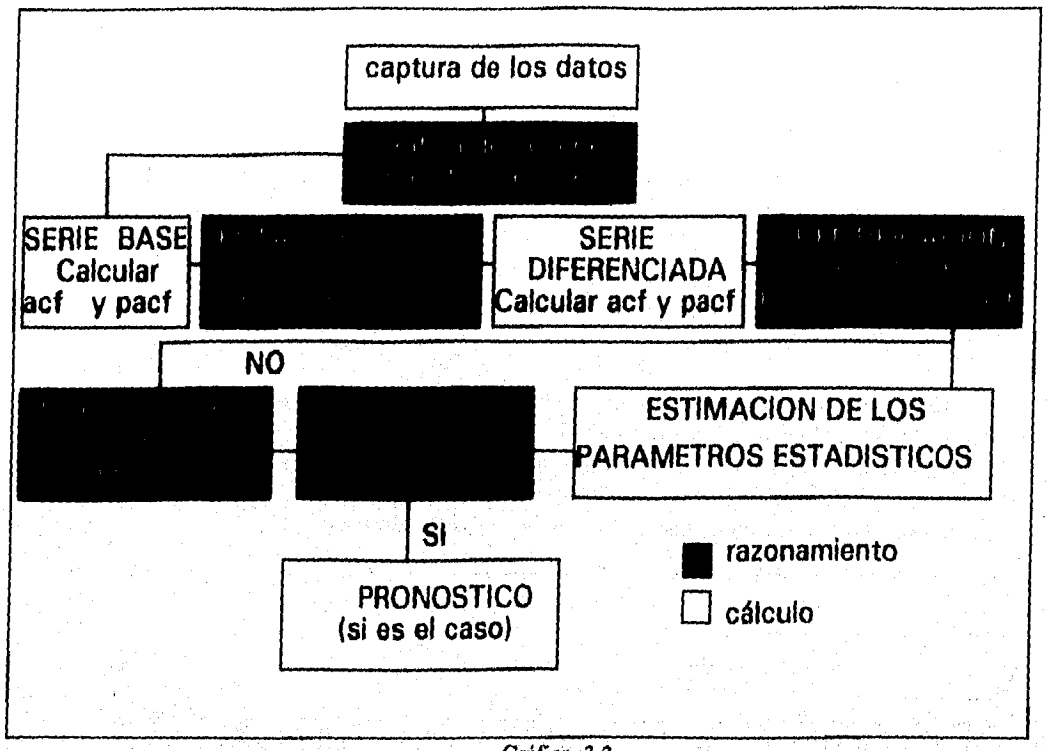
2. Estimación de los parámetros del modelo elegido.

3. Diagnóstico de la adecuación del modelo. Examinar los residuales del modelo filtrado para ver si es el modelo adecuado.

4. Considerar modelos alternativos si es necesario. Si el primer modelo no resulta adecuado por alguna razón, se debe probar otro modelo alternativo hasta encontrar un modelo adecuado o satisfactorio.

5. Pronóstico de realizaciones futuras del modelo elegido, si es el caso.

El diagrama del método de Box- Jenkins es el siguiente:



Gráfica 3.2

En la gráfica 3.2, se encuentran conceptos que aún no se han definido, como son transformaciones y diferencias, los que se explican más adelante. El análisis, como está indicado, necesita por una parte, que se hagan los cálculos requeridos, y por otra, de un razonamiento y juicio de la persona que debe interpretar el procedimiento. Por ejemplo, en la identificación se requiere de una amplia experiencia, así como de un buen razonamiento, claro, además del conocimiento del fenómeno del que se trate, ya que siempre se debe tener presente que la base de todo esto es elegir un modelo sencillo.

En el proceso del diagnóstico es muy importante el estudio de los residuales, ya que cuando se pruebe un modelo, es necesario averiguar si éstos tienen un comportamiento similar al ruido blanco. Por otro lado, cabe aclarar que existen modelos que se ajustan a los datos, pero dan como resultado pronósticos inadecuados, siendo conveniente en estos casos regresar al proceso de identificación, o incluso, revisar si el método es el adecuado. Ahora, cabe hacer notar que de acuerdo al objetivo final del método se puede sacrificar el ajuste por la precisión de los pronósticos.

ESTACIONARIDAD

Si consideramos una serie de tiempo como una realización de todas las series que pueden ser generadas por el mismo mecanismo, entonces se tiene una muestra de tamaño 1. Con esta muestra es posible estimar la media, la cual está en función del tiempo. Si no hay tendencia, entonces se asume que la media es constante y que el valor observado para cada período se puede representar por esta media, es decir, se tendrá que:

$$E(Y_t) = E(Y_{(t)}) = \mu. \quad \ldots\ldots (3.1)$$

$$\bar{Y} = E(Y_t) \text{ (estimada)} = (1/n) \sum_{t=1,n} Y_t. \quad \text{..... (3.2)}$$

El que la media de la serie sea constante es una primera condición de estacionaridad. Otra condición es que la varianza del proceso sea también constante, es decir:

$$\text{Var } Y_t = \text{var } Y_{t+h} = \sigma^2_y. \quad \text{..... (3.3)}$$

Por sencillez, sobre todo en el análisis teórico, se puede usar la variable transformada que representa a las desviaciones con respecto a la media:

$$y_t = Y_t - \bar{Y}.$$

Existe otra condición que se refiere a la autocorrelación de los datos.

$$\rho_k = \text{corr}(y_t, y_{t+k}),$$

$$\rho_k = \frac{E(y_t y_{t+k})}{\sigma^2_y}, \qquad -1<\rho_k<1. \quad \text{..... (3.4)}$$

Si la serie es estacionaria el valor de ρ_k sólo debe depender de k, lo cual implica que la relación entre dos variables sólo depende del intervalo que las separa.

DATOS NO ESTACIONARIOS

El primer paso en el análisis de las series de tiempo es graficar las observaciones contra el tiempo. Una vez hecha la gráfica se puede visualizar rápidamente características específicas de la gráfica como son la tendencia, la varianza, estacionalidad, etc. Cuando la gráfica tiene una varianza no constante, es preciso realizar una estabilización de la varianza.

ESTABILIZACION DE LA VARIANZA:

Para lograr una estabilización de la varianza, es decir, hacerla constante, se puede recurrir a varias transformaciones, dentro de las cuales, las más usadas son las logarítmicas y las raíces cuadradas. Estas transformaciones resultan útiles cuando:

-La varianza es proporcional al crecimiento de la serie.

-El nivel medio de la serie se incrementa o decrementa a una tasa constante, generalmente esta suposición es válida a corto y mediano plazo.

También es posible utilizar otras transformaciones de tipo exponencial, como el recíproco, recíproco de la raíz cuadrada, etc.

Una vez estabilizada la varianza, si la gráfica presenta tendencia se procede a eliminarla.

ELIMINACION DE LA TENDENCIA:

Existen varios métodos para la eliminación de la tendencia. Uno de ellos es usar un modelo de regresión, el cual resulta útil cuando la tendencia es fija y determinística, pero el problema surge a partir de la dificultad de saber cuándo un cambio en el nivel de la serie se debe a que la tendencia sea determinística o estocástica. Es ésta una de las razones por las que se usan las diferencias, las cuales resultan generalmente adecuadas para series con un número grande de observaciones. El método de las diferencias consiste en restar los valores de las observaciones uno de otro en un orden pre-establecido. Un polinomio de orden uno se vuelve constante al aplicarle una diferencia, uno de orden 2 se vuelve constante al aplicar dos diferencias y así sucesivamente.

Por ejemplo, para las primeras diferencias se tendrá:

$$\Delta y_t = y_t - y_{t-1}, \qquad \text{..... (3.5)}$$

y para las segundas diferencias:

$$\begin{aligned} \Delta^2 y_t &= \Delta y_t - \Delta y_{t-1}, \\ \Delta^2 y_t &= y_t - y_{t-1} - (y_{t-1} - y_{t-2}), \\ \Delta^2 y_t &= y_{t-2}\, y_{t-1} + y_{t-2}. \end{aligned} \qquad \text{..... (3.6)}$$

Podemos notar de las igualdades (3.5) y (3.6), que cada vez que se diferencia la serie se pierde una observación. El objetivo de eliminar la tendencia es obtener una nueva serie, la cual sea analizada más fácilmente, para después introducir nuevamente la tendencia. Lo anterior puede explicarse de una manera sencilla, supongamos ahora las primeras diferencias:

$$z_t = y_t - y_{t-1}, \qquad \text{.....(3.7)}$$

A partir de (3.7) es posible reconstruir la serie original y_t, en forma recursiva, por la siguiente ecuación:

$$\begin{aligned} y_t &= z_t + y_{t-1}, \\ &= z_t + z_{t-1} + y_{t-2}, \\ &\vdots \\ &= z_t + z_{t-1} + z_{t-2} + \ldots\ldots \end{aligned}$$

FLUCTUACIONES ESTACIONALES

La estacionalidad en una serie resulta muy evidente cuando se ha eliminado la tendencia, y en algunos casos es posible visualizarla aún cuando está presente. Como es preciso eliminar dicha estacionalidad se recurre a procedimientos de ajuste, como son las diferencias estacionales. Dichas diferencias son similares a las diferencias ordinarias presentadas anteriormente, esto se expresa de la siguiente forma:

$$\Delta_s y_t = y_t - y_{t-s},$$

donde s es la longitud del periodo de la fluctuación estacional, y se pierden s observaciones cada vez que se realiza una diferencia.

III) IDENTIFICACION DE MODELOS

En la etapa de identificación se debe elegir un modelo de la clase general de los ARIMA, donde se elige el orden de las diferencias ordinarias y estacionales, así como el orden de los polinomios autorregresivos y de medias móviles. Los procedimientos pueden resultar hasta cierto punto inexactos y por ende requerir del juicio personal del investigador, para apoyarse éste puede utilizar de una forma eficiente los paquetes de cómputo estadísticos que incluyan series de tiempo.

MODELOS AUTORREGRESIVOS

MODELO AUTORREGRESIVO DE PRIMER ORDEN.

Una serie de tiempo estacionaria se determina por un proceso autorregresivo de primer orden si los valores actuales de la serie z_t son expresados como una función lineal del primer valor anterior y un choque aleatorio e_t, es decir:

$$z_t = \phi_1 z_{t-1} + e_t \qquad \text{..... (3.8)}$$

donde:

ϕ_1 es el parámetro autorregresivo, el cual describe el efecto de un cambio de z_{t-1} en z_t.

e_t son los errores o ruido blanco,

y las propiedades de e_t son las siguientes:

$E(e_t) = 0$

$$E(e_t e_k) = \begin{cases} \sigma^2 & \text{si } t = k \\ 0 & \text{otro caso} \end{cases},$$

$(e_t z_{t-1}) = 0,$

$e_t \sim N(0, \sigma^2)$.

Restricciones en los parámetros/ Condiciones de estacionaridad.

Como ya se ha mencionado anteriormente, para que el proceso sea estacionario, en primer lugar la media debe ser constante, es decir:

$E(z_t) = E(z_{t-1}) = \mu,$
$E(z_t) = E(\phi_1 z_{t-1} + e_t),$
$E(z_t) = \phi_1 E(z_{t-1}),$
$E(z_t) = \phi_1 E(z_t).$

De aquí se deduce que $\mu = 0$ sin importar el valor de ϕ_1.

En segundo lugar, la varianza debe ser también constante:

$$\text{Var } z_t = \text{Var } (\phi_1 z_{t-1} + e_t),$$

$$\text{Var } z_t = \phi^2_1 \text{ var } (z_{t-1}) + \text{Var } (e_t).$$

Si $\text{Var } (z_t) = \text{Var } (z_{t-1}) = \gamma_0,$

entonces :

$$\gamma_0 = \phi^2_1 \gamma_0 + \sigma^2_e.$$

Al despejar a γ_0 nos queda:

$$\gamma_0 = \frac{\sigma^2}{1 - \phi_1^2}.$$

En la expresión anterior se observa que $-1 < \phi_1 < 1$ la cual es llamada condición de estacionaridad. En cuanto a su función de autocorrelación tenemos que:

$$\gamma_1 = \text{Cov } (z_t z_{t-1}),$$
$$\gamma_1 = E(z_t z_{t-1}),$$
$$\gamma_1 = E (\phi_1 z_{t-1} + e_t) z_{t-1}),$$
$$\gamma_1 = \phi_1 E (z_{t-1})^2 + E (e_t z_{t-1}),$$
$$\gamma_1 = \phi_1 \text{ Var } (z_{t-1}),$$ pero como $\text{Var } (z_{t-1}) = \gamma_0$ entonces nos queda:

$$\gamma_1 = \phi_1 \gamma_0.$$

$$\gamma_2 = \text{Cov}(z_t, z_{t-2}),$$
$$\gamma_2 = E (z_t z_{t-2}),$$
$$\gamma_2 = \phi_1 E(z_{t-1} z_{t-2}) + E (e_t z_{t-2}),$$
$$\gamma_2 = \phi_1 \gamma_1,$$
$$\gamma_2 = \phi^2_1 \gamma_0,$$
$$\gamma_3 = \phi^3_1 \gamma_0.$$

En general:

$$\gamma_k = \phi^k_1 \gamma_0,$$ si se divide por γ_0 nos queda lo siguiente:

$$\rho_k = \phi^k_1, \qquad \text{con } k > 0.$$

A partir de lo anterior es fácil visualizar que la función de autocorrelación sólo depende de k, y por la condición de estacionaridad sus valores son infinitos y decrecientes en forma geométrica.

Función de memoria

Si utilizamos la ecuación de golpe o choque aleatorio de un proceso AR(1):

$$z_t = e_t + \phi_1 e_{t-1} + \phi^2_1 e_{t-2} + \phi^3_1 e_{t-3} + \ldots\ldots\ldots,$$

donde el modelo AR(1) tiene memoria infinita, pero si el proceso es estacionario, es decir, $-1 < \phi_1 < 1$, el efecto se disipa al alejarse de t. El coeficiente de memoria con retraso k es el coeficiente de e_{t-k} y la gráfica de dichos coeficientes es la función de memoria.

MODELOS AUTORREGRESIVOS DE ORDEN MAYOR

Cuando el valor de z_t depende de dos o más valores anteriores, el modelo será un AR(p), como se indica a continuación de manera simbólica:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \phi_3 z_{t-3} + \ldots\ldots\ldots + \phi_p z_{t-p} + a_t \qquad \ldots\ldots\ldots (3.9)$$

Este modelo tiene "p" condiciones de estacionaridad, y su función de autocorrelación (ACF) es decreciente infinita.

MODELOS DE MEDIAS MOVILES

Otra forma de representar el proceso es en base a los choques aleatorios ponderados en el tiempo y acumulados.

MODELOS DE MEDIAS MOVILES DE PRIMER ORDEN

Por ejemplo, para un modelo MA(1) :

$$z_t = a_t - \theta_1 a_{t-1} \qquad \ldots\ldots\ldots (3.10)$$

Donde:

θ_1 describe el efecto del choque aleatorio anterior a z_t y
a_t tiene las mismas suposiciones anteriores, es decir:
a_t ($a_t \sim N(0, \sigma^2)$, $E(a_t a_s) = 0$.

Restricciones en los parámetros/Condición de invertibilidad:

La media y la varianza no imponen condiciones de estacionaridad, como se observa enseguida:

$$E(z_t) = E(a_t - \theta_1 a_{t-1}) = 0,$$

$$\gamma_0 = \text{Var } z_t = \text{Var}(a_t - \theta_1 a_{t-1}),$$
$$\gamma_0 = \text{var } z_t = \sigma^2 + \theta^2_1 \sigma^2 = (1 + \theta^2_1)\sigma^2,$$
$$\gamma_1 = \text{Cov}(z_t, z_{t-1}) = -\theta_1 \sigma^2,$$
$$\gamma_2 = \text{Cov}(z_t, z_{t-2}) = E[(a_t - \theta_1 a_{t-1})(a_{t-2} - \theta_1 a_{t-3})],$$
$$= 0,$$

$$\gamma_k = 0 \text{ para } k > 1.$$

En forma general se tendrá lo siguiente:

$$\rho_1 = -\theta_1 / (1 + \theta^2_1),$$

$$\rho_k = 0, \quad k>1.$$

En base a lo anterior se observa que θ_1 no tiene condiciones de estacionaridad. Pero al transformar la ecuación de golpe aleatorio en una ecuación de forma invertida, es decir de la forma autorregresiva, se obtiene un AR( ), el cual, para que sus coeficientes sean decrecientes se requiere que $|\theta_1| < 1$. A esto se le llama condición de invertibilidad.

Función de memoria

La función de memoria se puede obtener directamente de la definición del proceso, puesto que los coeficientes de los errores son 1 y $-\theta_1$ para a_t y a_{t-1} respectivamente.

MODELOS DE MEDIAS MOVILES DE ORDEN MAYOR

El proceso general de medias móviles denotado por MA(q) es el siguiente:

$$z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \qquad (3.11)$$

La función de memoria es la gráfica de los valores $1, -\theta_1, -\theta_2, -\theta_3, \ldots -\theta_q$ y el efecto del choque persistirá por "q" períodos. La función de autocorrelación se truncará después de q intervalos.

MODELOS MEZCLADOS AUTORREGRESIVOS Y DE MEDIAS MOVILES

Supongamos ahora un modelo que tiene tanto parte de medias móviles, como parte autorregresiva, digamos el modelo ARMA (1, 1) :

$$z_t = \phi_1 z_{t-1} + a_t - \theta_1 a_{t-1} \qquad (3.12)$$

el cual tiene condiciones de estacionaridad, es decir $|\phi_1| < 1$ y de invertibilidad $|\theta_1| < 1$, como se ve enseguida:

$E(z_t) = 0$, por lo que no hay restricción.

$$\gamma_0 = \frac{1 + \theta_1^2 - 2\theta_1\phi_1}{1 - \phi^2}\sigma^2,$$

$$\gamma_1 = \frac{(1 - \theta_1\phi_1)(\phi_1 - \theta_1)}{1 - \phi_1^2}\sigma^2,$$

$$\gamma_2 = \phi_1\gamma_1$$

Para $k \geq 2$:

$$\gamma_k = \phi_1\gamma_{k-1}$$
$$= \phi_1^{k-1}\gamma_1$$

De las igualdades anteriores se tiene que:

$$\rho_k = \phi_1^{k-1}\rho_1$$

La condición de estacionaridad será $|\phi_1| < 1$.

Función de memoria

La función de memoria se puede obtener directamente de la definición del proceso, puesto que los coeficientes de los errores son 1 y - θ_1 para a_t y a_{t-1} respectivamente.

MODELOS DE MEDIAS MOVILES DE ORDEN MAYOR

El proceso general de medias móviles denotado por MA(q) es el siguiente:

$$z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \quad \ldots\ldots (3.11)$$

La función de memoria es la gráfica de los valores 1, $-\theta_1$, $-\theta_2$, $-\theta_3$,...$-\theta_q$, y el efecto del choque persistirá por "q" períodos. La función de autocorrelación se truncará después de q intervalos.

MODELOS MEZCLADOS AUTORREGRESIVOS Y DE MEDIAS MOVILES

Supongamos ahora un modelo que tiene tanto parte de medias móviles, como parte autorregresiva, digamos el modelo ARMA (1, 1) :

$$z_t = \phi_1 z_{t-1} + a_t - \theta_1 a_{t-1} \quad \ldots\ldots (3.12)$$

el cual tiene condiciones de estacionaridad, es decir $|\phi_1| < 1$ y de invertibilidad $|\theta_1| < 1$, como se ve enseguida:

$E(z_t) = 0$, por lo que no hay restricción,

$$\gamma_0 = \frac{1 + \theta_1^2 - 2\theta_1\phi_1}{1 - \phi^2}\sigma^2,$$

$$\gamma_1 = \frac{(1 - \theta_1\phi_1)(\phi_1 - \theta_1)}{1 - \phi_1^2}\sigma^2,$$

$$\gamma_2 = \phi_1 \gamma_1,$$

Para $k \geq 2$:

$$\gamma_k = \phi_1 \gamma_{k-1},$$
$$= \phi_1^{k-1}\gamma_1.$$

De las igualdades anteriores se tiene que:

$$\rho_k = \phi_1^{k-1}\rho_1.$$

La condición de estacionaridad será: $|\phi_1| < 1$.

Si se sustituyen los e_{t-1}, e_{t-2}, e_{t-3}, etc en forma sucesiva, hasta obtener una ecuación de la forma invertida:

$$z_t = \phi_1 z_{t-1} + e_t - \theta_1 e_{t-1},$$

$$e_{t-1} = z_{t-1} - \phi_1 z_{t-2} + \theta_1 e_{t-2},$$

$$z_t = \phi_1 z_{t-1} + e_t - \theta_1 (z_{t-1} - \phi_1 z_{t-2} + \theta_1 e_{t-2}),$$

$$z_t = (\phi_1 z_{t-1} - \theta_1 z_{t-2}) + e_t + \theta_1 \phi_1 z_{t-2} - \theta^2_1 + e_{t-2},$$

$$\vdots$$

para que la invertibilidad tenga sentido, es decir, que sus coeficientes sean decrecientes se requiere que $|\theta_1| < 1$.

Función de memoria

Al sustituir los z_{t-1}, z_{t-2},.....del ARMA(1,1) hasta obtener una ecuación de golpe aleatorio:

$$z_t = e_t + (\phi_1 - \theta_1)e_{t-1} + \phi_1(\phi_1 - \theta_1)e_{t-2} + \phi^2_1(\phi_1 - \theta_1)e_{t-3} + \ldots\ldots\ldots$$

se observa que los coeficientes de memoria van decreciendo en forma exponencial.

Función de autocorrelación

$$\rho_1 = \frac{(1 - \theta_1\phi_1)(\phi_1 - \theta_1)}{1 + \theta^2_1 - 2\theta_1\phi_1},$$

$$\rho_K = \phi_1 \rho_{K-1},$$

$$= \phi^{k-1}_1 \rho_1, \; k > 1.$$

La ACF del ARMA(1,1) decrece a partir del valor inicial 1. El valor de ρ_1 depende de ambos parámetros y su signo depende de la diferencia $(\phi_1 - \theta_1)$.

MODELOS MEZCLADOS AUTORREGRESIVOS Y DE MEDIAS MOVILES DE ORDEN SUPERIOR

El modelo general ARMA (p,q) tiene la siguiente forma:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3.13)$$

Este modelo tiene condiciones tanto de estacionaridad como de invertibilidad. Su función de autocorrelación será decreciente y el efecto de las medias móviles desaparece después de q períodos.

MODELOS NO ESTACIONARIOS

MODELOS AUTORREGRESIVOS INTEGRADOS Y DE MEDIAS MOVILES

Un modelo ARIMA (p,d,q) tiene d diferencias ordinarias, las necesarias para que el modelo sea estacionario, y se dice que es integrado ya que al realizar el pronóstico será necesario realizar un proceso opuesto a las diferencias, es decir, se integrará o se sumarán los valores para reconstruir a la variable original, ya que el valor z_t se obtiene mediante la siguiente expresión: ..... (3.14)

$$z_t = w_t + w_{t-1} + w_{t-2} +$$

NOTACION

Operador diferencia

El operador diferencia se denota con la siguiente expresión: ..... (3.15)

$$\Delta z_t = z_t - z_{t-1},$$

y la diferencia consecutiva de orden d será:

$$\Delta^d z_t$$

A la diferencia estacional de primer orden con cierta longitud de periodo "s" se denota de la siguiente manera: ..... (3.16)

$$\Delta_s z_t = z_t - z_{t-s}$$

En forma general, las combinaciones de diferencias estacionales y ordinarias son

$$\Delta^D_s \Delta^d z_t$$

Operador de salto hacia atrás

El operador de salto hacia atrás se expresa de la siguiente forma: ..... (3.17)

$$B z_t = z_{t-1}$$

En forma general, el k-ésimo salto hacia atrás se expresa de la siguiente manera:

$$B^k z_t = z_{t-k}$$

MODELOS NO ESTACIONARIOS

MODELOS AUTORREGRESIVOS INTEGRADOS Y DE MEDIAS MOVILES

Un modelo ARIMA (p,d,q) tiene d diferencias ordinarias, las necesarias para que el modelo sea estacionario, y se dice que es integrado ya que al realizar el pronóstico será necesario realizar un proceso opuesto a las diferencias, es decir, se integrará o se sumarán los valores para reconstruir a la variable original, ya que el valor z_t se obtiene mediante la siguiente expresión:

$$z_t = w_t + w_{t-1} + w_{t-2} + \quad \text{..... (3.14)}$$

NOTACION

Operador diferencia

El operador diferencia se denota con la siguiente expresión:

$$\Delta z_t = z_t - z_{t-1}, \quad \text{..... (3.15)}$$

y la diferencia consecutiva de orden d será:

$$\Delta^d z_t$$

A la diferencia estacional de primer orden con cierta longitud de período "s" se denota de la siguiente manera:

$$\Delta_s z_t = z_t - z_{t-s}, \quad \text{..... (3.16)}$$

En forma general, las combinaciones de diferencias estacionales y ordinarias son :

$$\Delta^D{}_s \Delta^d z_t$$

Operador de salto hacia atrás

El operador de salto hacia atrás se expresa de la siguiente forma:

$$B z_t = z_{t-1}, \quad \text{..... (3.17)}$$

En forma general, el k-ésimo salto hacia atrás se expresa de la siguiente manera:

$$B^k z_t = z_{t-k}$$

AUTOCORRELACION PARCIAL

MODELOS AUTORREGRESIVOS

Cuando se construye un modelo autorregresivo se debe checar si al incluir una nueva z_{t-k} en el modelo los datos están mejor representados. Es decir, cuando se ha ajustado un modelo AR(k-1) y se desea ver si lo adecuado sería un modelo AR (k), se debe incluir a z_{t-k}. Si el valor de $|\phi_k|$ es grande, quiere decir que se debe incluir el z_{t-k}, en otro caso no es conveniente agregarlo. Este coeficiente mide la correlación que no es tomada por el modelo AR(k-1); es decir, mide el efecto parcial de z_{t-k} para explicar el comportamiento de z_t en un modelo que ya incluye $z_{t-1},...z_{t-(k-1)}$.

El coeficiente autorregresivo en el modelo de mayor orden, ϕ_k, se define como la autocorrelación parcial al intervalo k y se denota como ρ_{kk}. La gráfica de las ρ_{kk} será la función de autocorrelación parcial o PACF.

Luego entonces, si el modelo adecuado es un AR(1), ρ_{22} debe ser cero, o tomando el error muestral prácticamente cero, así como las siguientes autocorrelaciones parciales. En un modelo AR(p), las PACF's nos sirven para determinar el valor de p, y es este precisamente el punto clave para llevar a cabo la identificación del modelo.

MODELOS DE MEDIAS MOVILES

Partiendo de un modelo MA(1), la ecuación equivalente de forma invertida es:

$$z_t = -\rho_1 z_{t-1} - \theta^2_1 z_{t-2} - ..... + c_t$$

Las PACF del modelo anterior no se truncan, ya que su comportamiento consiste en decrecer muy rápidamente por sus condiciones de invertibilidad, además de que tienden hacia el infinito.

MODELOS MEZCLADOS

Supongamos un modelo mezclado ARMA(1,1):

$$z_t = \phi_1 z_{t-1} + c_t - \theta_1 c_{t-1},$$

su forma invertida equivalente es:

$$z_t = c_t + (\phi_1 - \theta_1) z_{t-1} + \theta_1(\phi_1 - \theta_1) z_{t-2} + \theta^2_1 (\phi_1 - \theta_1) z_{t-3} + ... ,$$

Si el modelo anterior es estacionario e invertible $|\phi_1| < 1$ y $|\theta_1| < 1$, la PACF desaparecerá gradualmente.

DUALIDAD ENTRE LOS MODELOS AR Y MODELOS MA

Después de analizar los modelos anteriores es posible concluir que existe cierta dualidad entre los procesos AR y MA como sigue:

-La ACF de un modelo MA se comporta como la PACF de un modelo AR; y
-La PACF de un modelo MA se comporta como la ACF de un modelo AR.
-Los procesos AR deben cumplir condiciones de estacionaridad, pero no de invertibilidad.
-Los procesos MA tienen condiciones de invertivilidad pero no de estacionaridad.

MODELOS NO ESTACIONALES

En cuanto a los modelos no estacionales, a continuación se presenta un cuadro con las características de las ACF y PACF de dichos modelos:

MODELO	ACF	PACF
AR(p)	decae	se trunca después de p
MA(q)	se trunca después de q	decae
ARMA(p,q)	decae	decae

AUTOCORRELACIONES MUESTRALES Y AUTOCORRELACIONES PARCIALES MUESTRALES

Debido a que es desconocido el proceso que rige a cierta serie de tiempo, es necesario estimar un modelo a partir de sus ACF y PACF, las cuales son estimadas a su vez, a partir de una muestra de tamaño n, siendo n el número de datos disponibles y tomando en cuenta que se trata de procesos estacionarios.

AUTOCORRELACIONES MUESTRALES

De acuerdo a Box y Jenkins, el estimador más adecuado por sus propiedades es el siguiente:

$$\hat{\rho}_k = r_k = c_k/c_0, \qquad \text{..... (3.18)}$$

donde :

$$\hat{r}_k = c_k = (1/n) \sum_{t=1,n-k} z_t z_{t+k}, \quad k \geq 0. \qquad \text{..... (3.19)}$$

La varianza aproximada de este estimador es, de acuerdo a Bartlett (1946) :

$$\mathrm{Var}(r_k) \approx \frac{1}{n}\left(1 + 2\left(\rho_1^2 + \rho_2^2 + \ldots + \rho_q^2\right)\right), \qquad k > q \qquad \text{..... (3.20)}$$

pero como las ACF teóricas son desconocidas :

$$\hat{\mathrm{Var}}(r_k) \approx \frac{1}{n}\left(1 + 2\sum_{1}^{q} r_i^2\right), \qquad k > q \qquad \text{..... (3.21)}$$

La raíz cuadrada de (3.21) es el error estándar para muestras grandes. Estas medidas tienen como objetivo determinar cuándo una autocorrelación se considera estadísticamente significativa. Es decir, probar si una autocorrelación se puede considerar igual a cero con cierto nivel de confianza.

Para lo anterior se construye un estadístico, (distribuido asintóticamente como t de Student), para probar la siguiente hipótesis:

$$H_0: \rho_k = 0$$
$$H_a: \rho_k \neq 0$$

$$t_{rk} = \frac{r_k}{Sr_k}. \qquad \text{.....(3.22)}$$

Considerando que el valor del estadístico t de Student para muestras grandes (valor de tablas para una normal) es 1.96 ($\alpha = 0.05$) se tiene la siguiente regla de decisión:

$$\rho_k = 0 \qquad \text{si} \quad |t_{rk}| \leq 1.96 \approx 2.$$

Visto de otra forma, puede probarse la hipótesis utilizando los intervalos de confianza (del 95% en general), de modo que los valores de r_k que queden fuera del intervalo formado por $[-2sr_k, 2sr_k]$ pueden considerarse como significativos.

Cuando se desea saber si el grupo de las primeras k autocorrelaciones de los residuales muestra evidencia de falta de adecuación al modelo se puede utilizar el estadístico de Box y Pierce donde se muestra que para un proceso sólo aleatorio, es decir, ruido blanco, donde todas las $\rho_k = 0$ se tiene el estadístico:

$$Q(k) = n(n+2) \sum_{k=1}^{K} \frac{1}{n-k} (r_k)^2, \qquad \text{.....(3.23)}$$

el cual se distribuye aproximadamente como una χ^2 con k grados de libertad. Se probará la hipótesis:

$H_0: \rho_1 = \rho_2 = \ldots\ldots = \rho_k = 0,$

H_a: al menos una distinta de cero.

Si el valor de Q es menor que el valor en tablas de la ji cuadrada con k grados de libertad, dado un nivel de significancia, entonces las autocorrelaciones usadas para calcular la prueba se pueden asumir no distintas de cero. Es decir, los datos que generaron las autocorrelaciones son ruido blanco. Si $Q > \chi^2$ las autocorrelaciones son significativamente diferentes de cero e indica la existencia de algún patrón.

AUTOCORRELACIONES PARCIALES MUESTRALES

Las PACF's muestrales se pueden estimar utilizando las ecuaciones de Yule- Walker:

$$\hat{\rho}_{kk} = r_{kk} = \begin{cases} r_1 & k=1 \\ \dfrac{r_k - \sum_{j=1,k-1} r_{k-1,j}\, r_{k-j}}{1 - \sum_{j=1,k-1} r_{k-1,j}\, r_j} & k= 2,3,\ldots \end{cases}$$

donde:

$$r_{kj} = r_{k-1,j} - r_{kk}\, r_{k-1,k-j} \qquad j= 1,2,3\ldots k-1. \qquad \text{.....(3.24)}$$

El método consiste en sustituir las ACF's estimadas en las ecuaciones de Yule- Walker y despejar el k-ésimo coeficiente del conjunto de ecuaciones correspondientes al modelo con k parámetros autorregresivos.

Según Quenoville (1949):

$$\text{Var}(r_{kk}) \approx \frac{1}{n}, \qquad k > p$$

donde:

n es el número de observaciones disponibles,
p son las primeras autocorrelaciones parciales que se suponen diferentes de cero.

Puede obtenerse el siguiente estadístico:

$$tr_{kk} = \frac{r_{kk}}{1/(n)^{1/2}}. \qquad \text{..... (3.25)}$$

La prueba de hipótesis es la siguiente:

H_0: $\rho_{kk} = 0$,

H_1: $|\rho_{kk}| > 0$,

entonces $\rho_{kk} = 0$ si $|tr_{kk}| \leq 1.96 \approx 2$.

De igual forma se puede tomar la decisión en base al intervalo de confianza, el cual será $\left[-2/\sqrt{n}, 2/\sqrt{n}\right]$.

IV) ESTIMACION DE MODELOS

METODOS DE MINIMOS CUADRADOS Y MAXIMA VEROSIMILITUD

Cuando ya se han elegido los valores p, d, q para el modelo $\phi(B)\, z_t = \theta(B) a_t$ es preciso obtener los estimadores para los p parámetros autorregresivos $\phi_1, \ldots, \phi_p$ y los q parámetros de medias móviles $\theta_1 .. \theta_q$. De igual forma que en un modelo de regresión lineal, se eligen los valores de los parámetros que minimicen la suma de los cuadrados de las diferencias entre la serie real $z_t = \Delta^d y_t$ y la serie ajustada z_t.

La ecuación $a_t = \theta^{-1}(B)\, \phi(B)\, z_t$ ..... (3.26)

representará el valor de los residuales, cuando se han sustituido los valores estimados de los parámetros. El objetivo es minimizar la suma de los errores al cuadrado, es decir, minimizar la expresión:

$$S(\phi, \theta) = \sum_t a_t^2. \qquad \text{..... (3.27)}$$

Se denotará a los vectores que minimizan dicha expresión de la siguiente forma :

$$\hat{\phi} = (\hat{\phi}_1, \ldots \hat{\phi}_p) \quad \text{y}$$

$$\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_q),$$

y a los residuales respectivos a estos parámetros como:

$$\hat{c}_t = \hat{\theta}(B)^{-1} \hat{\phi}(B) z_t.$$

Entonces se tendrá la siguiente expresión:

$$S(\hat{\phi}, \hat{\theta}) = \sum \hat{c}_t^2.$$

El proceso de estimación puede resultar difícil cuando en el modelo se tiene la parte de medias móviles, ya que la ecuación no es lineal y por consiguiente es preciso hacer uso de un proceso iterativo para minimizar nuestra expresión. Además el primer error de la serie de tiempo, c_1, depende de valores no observados $z_0, z_{-1}, \ldots, z_{-p+1}$ y $c_0, c_{-1}, \ldots c_{-q+1}$. Ahora, para esto se deben tener valores de $z_0, z_{-1}, \ldots z_{-p+1}$, no observados, por lo que es preciso utilizar un método para inicializar la serie.

EL PROCEDIMIENTO DE ESTIMACION

Teniendo T + d observaciones disponibles para una serie de tiempo homogénea no estacionaria, la cual requiere "d" diferencias ordinarias para que sea estacionaria y se denotan estas observaciones como $y_{-d+1}, \ldots y_0, y_1, \ldots y_T$.

Al diferenciar la serie se obtiene una serie estacionaria z_t con $z_1, \ldots z_t$ observaciones. Es necesario estimar los parámetros ϕ, θ para el modelo ARMA (p,q).

Haciendo uso de la suposición de que los errores c_t se distribuyen como una normal con media cero y varianza σ^2 se tiene la siguiente función de verosimilitud:

$$F(\phi, \theta, \sigma) = | (2 \pi \sigma^2)^{-1/2} |^T \exp(-s(\phi, \theta) / 2 \sigma^2)$$

$$= (2 \pi \sigma^2)^{-t/2} \exp(-s(\phi, \theta) / 2 \sigma^2),$$

obteniendo el logaritmo nos queda:

$$L(\phi, \theta, \sigma) = - \frac{T(\log 2\pi \sigma^2)}{2} - \frac{s(\phi, \theta)}{2 \sigma^2}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.28)$$

De lo anterior se dice que $L(\phi, \theta, \sigma)$ es la función de verosimilitud logarítmica condicional, ya que la suma de errores cuadrados $S(\phi, \theta)$ depende de valores pasados no observables $z_0, z_{-1}, \ldots z_{-p+1}, c_0, c_{-1}, \ldots, c_{-q+1}$.

Los estimadores de máxima verosimilitud de ϕ y θ estarán dados por la minimización de la suma de los residuales al cuadrado. Entonces, bajo la suposición de normalidad, el estimador de máxima verosimilitud es igual al de mínimos cuadrados.

INICIALIZACION DE LA SERIE

Debido a que $S(\phi, \theta)$ y $L(\phi, \theta, \sigma)$ son condicionantes con respecto a valores pasados no observables de z_t y a_t, los estimadores de mínimos cuadrados que se obtengan dependen de la elección de z_0, z_{-1},...etc.

Una alternativa es tomar sus valores incondicionales, es decir, cero, esto es recomendable si los valores reales de $\phi_1,..., \phi_p$ no son muy cercanos a 1 (a la no estacionaridad) y si el número de observaciones T es grande en relación con p y q.

Otra opción es hacer pronósticos hacia atrás, usando el operador de salto hacia adelante:

$$Fz_t = z_{t+1},$$

$$z_t = \phi^{-1}(F)\,\theta(F)\,a_t. \quad \text{.....(3.29)}$$

Se realiza la estimación usando los valores esperados incondicionales, se estiman los parámetros, se hace el pronóstico hacia atrás, se reestima y así sucesivamente hasta encontrar convergencia.

ESTIMACION NO LINEAL DE LOS PARAMETROS DEL MODELO

Se desea encontrar valores de ϕ y θ que minimicen la suma de errores cuadrados:

$$S(\phi, \theta) = \Sigma_t\ a^2_t = \Sigma_t\ (a_t/\phi, \theta, z)^2, \quad \text{.....(3.30)}$$

donde:

z es el vector de las observaciones z_t
y a_t son los errores, además a_t está condicionado a ϕ, θ, z.

Si el modelo es sólo autorregresivo, o sea que tiene la forma

$$z_t = \phi_1 z_{t-1} +...+ \phi_p z_{t-p} + a_t, \quad \text{..... (3.31)}$$

entonces la ecuación anterior tiene la forma general:

$$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} +...+ a_t, \quad \text{..... (3.32)}$$

la igualdad 3.32 es posible estimar con regresión lineal múltiple de la siguiente forma:

$$\phi = (X^T\ X)^{-1}\ X^T\ Y.$$

Si el proceso tiene una parte de medias móviles tendrá la forma:

$$\theta^{-1}(B)\ \phi(B)\ z_t = a_t, \quad \text{..... (3.33)}$$

donde se observa que esta última ecuación no es lineal en los parámetros y no se podrá realizar la estimación por mínimos cuadrados, pero sí por una rutina general iterativa no lineal.

Si β representa los p+ q parámetros (ϕ, θ) que se desean estimar entonces se deben elegir valores para β que minimicen

$$S(\beta) = \sum_{t=1}^{T} (a_t/ z, \beta)^2 = \sum_{t=1}^{T} (a_t)^2, \quad \text{..... (3.34)}$$

es posible expander e_t en una serie de Taylor alrrededor de un valor inicial β_0 para los valores de los parámetros

$$[e_t] = [e_t / z, \beta_0] + \sum_{i=1}^{p+q} (\beta_i - \beta_{i,0}) \frac{\partial (e_t)}{\partial \beta_i} \Big|_{\beta = \beta_0} + \frac{1}{2} \sum_{i=1}^{p+q} (\beta_i - \beta_{i,0})^2 \frac{\partial^2 (e_t)}{\partial \beta^2_i} \Big|_{\beta = \beta_0} + \ldots \quad \ldots\ldots\ldots\ldots (3.35)$$

donde $\beta_{i,0}$ es el valor inicial para el i-ésimo componente del vector β_0. $[e_t]$ se aproximará con los dos primeros términos de la expansión de Taylor, entonces, sea

$$X_{i,t} = - \frac{\partial [e_t]}{\partial \beta_i} \Big|_{\beta = \beta_0} \quad \ldots\ldots\ldots\ldots (3.36)$$

y

$$[e_{t,0}] = [e_t / z, \beta_0] \quad \ldots\ldots\ldots\ldots (3.37)$$

$$[e_t] \approx [e_{t,0}] - \Sigma_{i=1,p+q} (\beta_i - \beta_{i,0}) X_{i,t} \quad \ldots\ldots\ldots\ldots (3.38)$$

y

$$[e_{t,0}] + \Sigma_{i=1,p+q} \beta_{i,0} X_{i,t} = \Sigma_{i=1,p+q} \beta_i X_{i,t} + [e_t]. \quad \ldots\ldots\ldots\ldots (3.39)$$

Los parámetros β_i pueden estimarse de la ecuación (3,39) por regresión lineal :

$$\mathbf{Y} = \begin{vmatrix} [e_{1,0}] + \Sigma_{i=1,p+q} \beta_{i,0} X_{i,1} \\ \vdots \\ [e_{T,0}] + \Sigma_{i=1,p+q} \beta_{i,0} X_{i,t} \end{vmatrix}$$

$$\mathbf{X} = \begin{vmatrix} X_{1,1} \; X_{2,1} \ldots\ldots\ldots\ldots X_{p+q,1} \\ \vdots \\ X_{1,T} \; X_{2,T} \ldots\ldots\ldots\ldots X_{p+q,T} \end{vmatrix}$$

$$\beta = \begin{vmatrix} \beta_1 \\ \vdots \\ \beta_1 \end{vmatrix} \qquad \mathbf{Y} = \mathbf{X}\,\beta + [e_t]$$

Se hace la regresión ordinaria por mínimos cuadrados y se llega a un nuevo estimador de β, llamado β_1. Después, con una nueva expansión de taylor de $[e_t]$ alrededor de β_1 se obtiene otra ecuación equivalente a (3.39) y se genera un nuevo estimador β_2 repitiéndose el proceso hasta que $\beta_k - \beta_{k-1} \approx 0$ ó $|\beta_k - \beta_{k-1}| \leq \xi$. β_k será el estimador final de $\phi_1, \phi_2, \ldots \phi_p, \theta_1, \theta_2, \ldots \theta_q$.

Cabe aclarar que es posible que el proceso sea divergente o que existan soluciones múltiples. Si el proceso resulta divergente se puede reestimar el modelo una o más veces con distintos valores iniciales, y si no se encuentra una solución convergente se debe buscar una nueva especificación del modelo. Si ocurre la convergencia en el primer intento es conveniente analizar si hay soluciones múltiples reestimando el

modelo con distintos valores iniciales. Si se tienen soluciones múltiples el estimador final será el que minimice la suma de errores al cuadrado.

OBTENCION DE UN VALOR INICIAL PARA LOS PARAMETROS

Cuando se realiza la estimación no lineal es preciso hacer una inicialización tentativa de los parámetros $\beta_0 = (\phi_0, \theta_0)$; además de que la convergencia del proceso es más rápida si los valores iniciales son buenos. Para ello es posible usar la función de autocorrelación, la cual resulta adecuada para modelos sencillos, pero no para complejos.

Si el modelo contiene una parte de medias móviles, las ecuaciones que relacionan los parámetros y la función de autocorrelación no son lineales. Luego entonces, las soluciones se tornan complicadas, ya que para obtener soluciones para el modelo MA(q) se necesitan resolver q ecuaciones simultáneas no lineales.

Existen también tablas para obtener aproximaciones iniciales de los parámetros, pero algunos autores (Bowerman y O' Conell) consideran que es posible utilizar el valor 0.1 para cualquier parámetro desconocido. Otros autores toman el cero, ambos casos con resultados favorables.

V) EXAMEN DEL MODELO

Se explicarán en esta parte cuatro pruebas de diagnóstico para evaluar el ajuste del modelo. En caso de haber una falta de ajuste se debe buscar diferentes especificaciones.

ANALISIS DE ESTACIONARIDAD

Cuando el modelo ya ha sido identificado y estimado sus parámetros, se debe verificar si este modelo puede mejorarse, por lo que es importante verificar que el modelo sea estacionario, ya sea en su forma original o con ayuda de las diferentes transformaciones o mediante diferencias o ambas. Aunque cabe notar que la serie no debe ser sobrediferenciada, ya que nos puede generar un modelo más complicado, lo que está en desacuerdo con el principio de parsimonia. Por lo tanto, se debe elegir la opción que dé el mejor resultado en cuanto a ajuste y pronósticos en el proceso estacionario.

ANALISIS DE RESIDUALES

El análisis de residuales nos sirve para verificar que el modelo sea muy cercano al proceso que rige a la serie. Para que esto sea cierto se supone que los errores se comportarán como ruido blanco. Para verificarlo se observa la gráfica de su ACF, la cual no debe tener picos y sus autocorrelaciones muestrales deben ser todas pequeñas.

Un estimador del error estándar de las autocorrelaciones es el estadístico Q:

$$Q(k) = n(n+2) \sum_{k=1}^{K} \frac{1}{n-k} r^2_k \quad (c).$$

Si los errores son ruido blanco, Q se distribuye aproximadamente como una ji cuadrada con k - p -q grados de libertad. Esta hipótesis es rechazada cuando los valores de Q son grandes comparados con los valores en tablas.

Otra forma de saber si los errores son ruido blanco es evaluar las autocorrelaciones de las diferencias de los residuales, si son ruido blanco entonces $e_t - e_{t-1} = a_t - \theta_1 e_{t-1}$; $\theta_1=1$. La primera autocorrelación de un proceso MA(1) con $\theta_1=1$ es -0.5, $\rho_1 = -1/2$, por lo que, si los residuales son ruido blanco, sus primeras diferencias deben mostrar una autocorrelación cercana a -0.05.

En ciertas ocasiones, sobre todo cuando se trata de modelos estacionales, puede ser que no se hayan filtrado todas las características de la serie, entonces, para detectar una periodicidad en los residuales es preciso analizar al periodograma integral, el cual está definido como:

$$C(f_j) = \sum_{i=1}^{j} \frac{R^2(f_i)}{ns^2}, \qquad \text{.....(3.40)}$$

$$R^2(f_i) = \frac{2}{n}\left(\sum_{t=1}^{n} e_t \cos 2\pi f_i t \right)^2 + \left(\sum_{t=1}^{n} e_t \operatorname{sen} 2\pi f_i t \right)^2$$

$f_i = \frac{i}{n}$ = frecuencia; $\frac{1}{f_i}$ = período; e_t = errores; $s^2 = \hat{\sigma}^2$.

El valor de $R^2(f_i)$ es una medida que indica el ajuste de una función trigonométrica (seno, coseno) con frecuencia f_i con respecto a los residuales, es decir, es una herramienta que nos sirve para correlacionar las e's con las funciones seno y coseno de distintas frecuencias.

Un residual con patrón estacional de período 12 (s=12) tendrá valores grandes en $R^2(1/2)$, $R^2(1/24)$, $R^2(1/36)$, etc. y por consiguiente el periodograma acumulativo c(fj) tendrá saltos en estas frecuencias.

En la mayoría de los paquetes estadísticos esto es manejado por medio de la gráfica del periodograma integral contra las frecuencias o el período. Para una serie de ruido blanco su gráfica tendrá la forma aproximada de una línea recta que une los puntos (0,0) y (0.5, 1) puesto que todas las frecuencias tienen la misma importancia dentro del proceso.

MODELO SOBREESPECIFICADO/Omitir parámetros.

Una manera de localizar a los parámetros redundantes es usando el estadístico t de los coeficientes y el estimador de las correlaciones entre los estimadores.

Se puede decir que un coeficiente es significativamente distinto de cero si el valor absoluto del estimador puntual es al menos dos veces el error estándar, también es posible usar el estadístico T del estimador

$$T = \frac{\beta_j - \beta_j}{\sigma \beta_j}.$$

Se debe probar la hipótesis:

H_0: $\beta_j = 0$
H_a: $\beta_j \neq 0$

Si el valor en tablas de t con un nivel de significancia α y (n-p-q) grados de libertad es mayor que T, existe evidencia de que el parámetro puede ser cero.

Si algún parámetro es insignificante, esto nos indica que el modelo puede estar sobreespecificado y es necesario simplificarlo, si dicho parámetro es el de mayor orden, entonces por lo general el modelo se simplificará eliminando este parámetro del modelo.

Si el parámetro insignificante no es el de mayor orden, se debe examinar las correlaciones entre los estimadores para saber cuál parámetro eliminar, por ejemplo, una correlación alta entre dos estimadores nos indica que hay posibilidad de que uno de los dos parámetros se omita sin que cambie el ajuste del modelo.

Si el parámetro insignificante no es el de mayor orden, pero está fuertemente correlacionado con el de mayor orden se debe evaluar el modelo sin el parámetro de mayor orden; si esta correlación no existe, se reestimará el modelo suprimiendo el parámetro insignificante.

Si las correlaciones entre estimadores son altas para coeficientes de polinomios, por ejemplo ϕ_2 y θ_3, siendo o no alguno de ellos insignificante, la situación requiere de un análisis más cuidadoso.

Unicamente podrán evaluarse modelos suprimiendo uno de los parámetros, además cuando el parámetro insignificante no es el de mayor orden y no exista correlación, probablemente se estará reflejando un efecto de fluctuación estacional.

MODELOS SUBESPECIFICADOS/ Ajustar parámetros extra.

Una forma de verificar que el modelo tiene el número de parámetros adecuado es incluir otro parámetro en forma adicional y verificar si con esta adición se logra un mejor modelo.

Cuando se ajustan parámetros extra al modelo se corre el riesgo de caer en la llamada redundancia de parámetros, y esto ocurre cuando se agrega al mismo tiempo un parámetro autorregresivo y otro de medias móviles, por lo que jamás se debe agregar un parámetro AR y uno MA en forma simultánea.

PRONOSTICOS

Una vez encontrado el modelo adecuado para nuestra serie de tiempo éste se puede utilizar para generar pronósticos .

Al período actual se le llama fecha de origen y si se desea pronosticar "h" períodos de tiempo hacia adelante, es decir, hasta el período n+h, al intervalo de tiempo h se le llama horizonte, otra forma de denotar al valor futuro estimado de la variable aleatoria es $z_n(h)$.

la expresión

$$e_n(h) = z_{n+h} - z_n(h), \quad \text{..... (3.42)}$$

es el error al hacer un pronóstico en el momento n para un futuro n=h. El pronóstico óptimo será el valor de $z_n(h)$ que minimice el costo esperado $C(e_n(h))$ y por lo general se puede usar una función de costo cuadrática $C(e)= \alpha e^2$, donde α es una constante positiva.

Esta función define el costo en forma proporcional al cuadrado del error y tiene la propiedad de que la solución para el pronóstico óptimo es la media de la distribución del pronóstico.

La función de costo implica que los costos sean simétricos, si esto no se apega a la realidad se puede tomar los valores:

$$C(e) = \begin{cases} \alpha a & e > 0 \\ 0 & e = 0 \\ -\beta e & e < 0 \end{cases}$$

Con α y B > 0, siendo los costos unitarios de subestimar y sobreestimar z_{n+h} respectivamente. La media de la distribución del pronóstico minimiza el valor esperado de la función de costo cuadrática y es posible calcularla de la siguiente forma:

Sea z_t un proceso ARMA(p,q), estacionario e invertible para t= n+h.

$$z_{n+h} = \phi_1 z_{n+h-1} + \phi_2 z_{n+h-2} + \ldots + \phi_p z_{n+h-p} + e_t - \theta_1 e_{n+h-1} - \ldots - \theta_q e_{n+h-q}$$

1. Reemplazar el error actual y los pasados e_{n+j}, $j \leq 0$ con los residuales.
2. Reemplazar cada error futuro e_{n+j}, $0<j \leq h$ por su valor esperado, cero.
3. Reemplazar las observaciones z_{n+j}, $j \leq 0$ con los valores observados.
4. Reemplazar cada valor futuro de z_{n+j}, $0<j<h$, con el pronóstico adecuado $z_n(j)$; de modo que hay que pronosticar $z_{n+1},\ldots,z_{n+h-1}$ para pronosticar z_{n+h}, además, los parámetros del modelo se reemplazan por sus estimadores.

ACTUALIZACION DE LOS PRONOSTICOS

En cuanto se conozca el punto z_{n+1} se pueden ajustar los pronósticos de dos formas:

1. Pronósticos actualizados en forma secuencial:
   usando el punto adicional se reestiman los parámetros.

2. Pronóstico adaptativo:
   se dejan los valores de los parámetros, pero se cambia el origen.

VI) MODELOS ESTACIONALES

En las series de tiempo observadas con intervalos mensuales o trimestrales, las cimas o valles por lo regular ocurren alrededor de los mismos trimestres o meses de años sucesivos, por lo que resulta lógico tratar de explotar la correlación entre los mismos meses o trimestres de años sucesivos, así como la correlación entre trimestres o meses sucesivos.

MODELOS PURAMENTE ESTACIONALES

Este modelo se aplica si el valor actual de la serie z_t se puede expresar como una función lineal del valor de la serie s períodos atrás, z_{t-s} y el choque aleatorio c_t, es decir:

$$z_t - \Phi_1 z_{t-s} = c_t, \quad \text{.....} (3.43)$$

$$(1 - \Phi B^S) z_t = c_t, \quad \text{.....} (3.44)$$

donde:

Φ_1 es el parámetro autorregresivo estacional. Este modelo se puede denotar como AR(1)s o SARI(1).

Un modelo SAR(P) queda expresado como:

$$\Phi(B)^S z_t = c_t, \quad \text{.....} (3.45)$$

donde:

$$\Phi(B)^s = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \text{.....} - \Phi_p B^{Ps}, \quad \text{.....} (3.46)$$

$$z_t = \Delta^D \Delta^d W_t. \quad \text{.....} (3.47)$$

La función de autocorrelación del modelo estacional autorregresivo es similar en características generales a uno ordinario autorregresivo excepto que los valores de las autocorrelaciones aparecen en múltiplos de "s".

MODELOS ESTACIONALES AUTORREGRESIVOS

Un modelo SAR, es decir, que se encuentra gobernado por un proceso autorregresivo estacional de orden P tendrá la forma:

$$\Phi(B)^s Z_t = c_t, \quad \text{.....} (3.48)$$

donde

$$\Phi(B)^s = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \text{.....} - \Phi_p B^{ps}, \quad \text{.....} (3.49)$$

$$z_t = \Delta^D{}_s \Delta^d W_t, \quad \text{.....} (3.50)$$

Los valores de las autocorrelaciones aparecen en múltiplos de "s".

MODELOS ESTACIONALES DE MEDIAS MOVILES

Un modelo SMA, es decir, que se encuentra gobernado por un proceso de medias móviles estacionales y de orden Q tendrá la forma :

$$z_t = \Theta(B)^s c_t, \quad \text{.....} (3.51)$$

donde:

$$\Theta(B)^s = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \text{....} - \Theta_Q B^{Qs}, \quad \text{.....} (3.52)$$

y

$$z_t = \Delta^D{}_s \Delta^d W_t. \quad \text{.....} (3.53)$$

La función de autocorrelación es semejante a la de medias móviles, sólo que los valores de las autocorrelaciones aparacen en los intervalos múltiplos de "s". Para un modelo estacional de medias móviles de orden 1, la función de autocorrelación tendrá sólo un valor diferente de cero en el intervalo "s ".
$\rho_s = - \Theta_1 / (1+ \Theta_1)$ y los demás son ceros. La función de autocorrelación parcial será decreciente infinita, con valores significativos en múltiplos de s.

MODELOS ESTACIONALES MEZCLADOS

Un modelo estacional mezclado autorregresivo y de medias móviles, de orden (p,q) se denota de la siguiente manera:

$$\Phi (B)^s Z_t = \Theta (B)^s e_t.$$

Las autocorrelaciones distintas de cero aparecen en los intervalos múltiplos de s.

En general , puede denotarse ARIMA(P,D,Q)s donde:

P = Orden del proceso AR estacional.
Q = Orden del proceso MA estacional.
D = Orden de las diferencias estacionales.
S = Longitud del período.

MODELOS GENERALES MULTIPLICATIVOS ESTACIONALES

Es posible combinar todos los modelos en una clase general llamada modelos ARIMA multiplicativos y se expresa de la siguiente forma:

$$\phi(B)\Phi(B^s) z_t = \theta(B) \Theta (B^s)e_t \quad \text{..... (3.54)}$$

ó

ARIMA(p,d,q)x (P,D,Q)$_s$.

donde $z_t = \Delta_s^D \ \Delta^d w_t$. .....(3.55)

Este modelo da como resultado una gran cantidad de modelos con pocos parámetros, y es casi seguro que se encontrará alguno adecuado al fenómeno que se esté manejando.

FUNCIONES DE AUTOCORRELACION Y AUTOCORRELACION PARCIAL DE MODELOS ESTACIONALES.

Se puede resumir las características de los modelos multiplicativos de sus ACF y PACF de la siguiente forma:

ACF y PACF de modelos estacionales con longitud de período s.

MODELO	ACF	PACF
AR(P)XSAR(P)	decae	se trunca después de p+sP
MA(q)xSMA(Q)	se trunca después de q + sQ	decae
mezclados	decae	decae

IV. ANALISIS DE LAS VARIABLES DE ESTUDIO

El objetivo del presente capítulo es detectar la relación que existe entre las distintas variables de estudio, incluyendo los tipos de cáncer que presentan los pacientes. Se requiere tener un conocimiento más amplio sobre las condiciones socioeconómicas del paciente para ayudar a esclarecer de alguna forma el problema en cuestión.

En primer lugar, a la tabla en turno que le corresponda ser analizada se le aplicará la prueba de independencia X^2, la cual se explica en el capítulo II. Esta prueba se realiza por medio del paquete estadístico STATGRAPHICS versión 7.0 en la parte correspondiente a tablas de contingencia. En caso de existir dependencia entre la variable que se estudia y el tipo de cáncer, se procede a realizar un análisis más detallado, como es el grado de asociación existente.

Antes de comenzar con el análisis de las variables, se presentan las categorías correspondientes al tipo de cáncer. Todas las demás serán estudiadas con dicha variable, excepto la última tabla que está formada por las variables escolaridad de la madre, y tiempo transcurrido entre el inicio de signos y síntomas al diagnóstico.

VARIABLE A: Tipo de cáncer

CATEGORIAS:

A1: Leucemias.
A2: Linfomas.
A3: Tumores del S.N.C.*
A4: Neuroblastomas.
A5: Retinoblastomas.
A6: Tumor de Wilms.
A7: Tumores hepáticos.
A8: Tumores óseos.
A9: Rabdomiosarcoma.
A10: Tumores de células germinales.
A11: Carcinomas.
A12: Mal definidos.

*: Sistema Nervioso Central.

I) SEXO DEL PACIENTE Y TIPO DE CANCER

Se considera la posibilidad de que el sexo del paciente tenga cierta influencia o represente una característica en el tipo de cáncer desarrollado. Para tener una base sólida en aceptar o rechazar la idea, en primera instancia se construye una tabla con los datos que indican el número de pacientes que pertenecen a uno de los dos sexos y además desarrollaron un tipo de cáncer en específico. Las frecuencias encontradas se muestran en la tabla 4.1.

VARIABLE B: SEXO

CATEGORIAS:

B1: Masculino.
B2: Femenino.

		TIPO DE CANCER												
		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	TOTAL
SEXO	B1	238	152	64	15	44	20	10	42	34	37	11	13	680
	B2	215	62	53	14	34	32	7	30	16	23	11	10	507
	TOTAL	453	214	117	29	78	52	17	72	50	60	22	23	1187

TABLA 4.1

Se desea probar la hipótesis de que el tipo de cáncer es independiente del sexo, contra la alternativa de que las dos variables son dependientes. De manera más formal se tiene lo siguiente:

H0: El tipo de cáncer es independiente del sexo.
Ha: El tipo de cáncer no es independiente del sexo.

El resultado obtenido de la prueba de bondad de ajuste es $X^2 = 32.2772$ con 11 grados de libertad y un nivel de significancia de 0.00068. Si consideramos que el valor elegido de α es 0.05, se rechaza la hipótesis nula de independencia. Esto implica un estudio sobre el grado de asociación, por lo que se obtienen los estadísticos que se muestran en la tabla 4.2.

ESTADISTICO	SIMETRICO	RENGLONES DEPENDIENTES	COLUMNAS DEPENDIENTES
Lambda	0.00967	0.02367	0.00000
Coeficiente de Incertidumbre	0.01043	0.02032	0.00702
D de Somers	-0.02684	-0.02167	-0.03526
	VALOR	SIGNIFICANCIA	
Coeficiente de Contingencia	0.16270		
V de Cramers	0.16490		
Gama Condicional	-0.04413		
τ_b de Kendall	-0.02764	0.28242	
τ_c de Kendall	-0.03451		

TABLA 4.2

Conforme a los datos que presenta la tabla 4.2, se observa en el estadístico lambda y el coeficiente de incertidumbre un valor mayor en las renglones dependientes con respecto al valor simétrico y el de columnas dependientes. Esto indica que si se conoce el tipo de cáncer al cual pertenece el paciente, la reducción del error al predecir el sexo es de un 2 por ciento aproximadamente. Otro aspecto necesario de subrayar, es que el valor del estadístico lambda, tomando en consideración a las columnas dependientes, cae en una sola categoría, la leucemia. Esto no es nada extraño si se considera que es el tipo de cáncer con mayor frecuencia. Por su parte, el coeficiente de contingencia y la V de Cramers presentan un valor muy similar, pero no demasiado grande, por lo que se puede decir que el grado de asociación no es muy fuerte.

De acuerdo a la tabla 4.1 se observa en la celda (B2, A6) que la independencia se rompe. Esto se deduce debido a que la frecuencia de niñas con tumor de Wilms es 44.1 % mayor a la frecuencia esperada. En términos reales, esto implica que el tumor de Wilms presenta una tendencia en el sexo femenino.

II) DOMICILIO (DELEGACION) DEL PACIENTE Y TIPO DE CANCER

Otra de las variables consideradas para el análisis de tablas de contingencia es la delegación a la que pertenece el paciente. Esto nos proporciona algunos indicadores como el nivel socioeconómico, el grado de contaminación existente en determinada zona, u otro factor relevante que se pueda tomar como riesgoso. Se consideran a las 16 delegaciones del D.F. y se obtienen las frecuencias de cada una de ellas de acuerdo a los doce diferentes tipos de cáncer. En la tabla 4.3 se encuentran las frecuencias obtenidas.

VARIABLE C: DOMICILIO (DELEGACION)

CATEGORIAS:

C1: Alvaro Obregón.	C5: Cuauhtémoc.	C9: Miguel Hidalgo.	C13: Magdalena Contreras.
C2: Azcapotzalco.	C6: Gustavo A. Madero.	C10:Tlalpan.	C14: Cuajimalpa.
C3: Benito Juárez.	C7: Iztacalco.	C11:V. Carranza.	C15: Tláhuac.
C4: Coyoacán.	C8: Iztapalapa.	C12:Xochimilco.	C16: Milpa Alta.

		TIPO DE CANCER												
		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	TOTAL
	C1	46	15	12	2	5	0	0	7	4	3	0	1	95
	C2	19	13	5	1	2	2	0	4	4	5	2	1	58
	C3	24	12	4	3	7	4	1	6	1	4	2	1	69
D	C4	36	16	5	2	6	4	1	5	6	2	4	4	91
E	C5	37	16	7	3	9	6	2	5	1	3	1	3	93
L	C6	61	30	14	4	8	4	4	10	7	9	4	2	157
E	C7	23	7	12	3	8	4	1	7	4	4	0	1	74
G	C8	76	46	19	2	16	11	3	16	12	13	4	4	222
A	C9	13	13	4	3	4	3	1	1	2	2	0	1	47
C	C10	28	11	9	2	1	4	1	4	1	2	2	1	66
I	C11	30	13	11	0	5	2	1	2	5	6	2	1	78
O	C12	18	7	6	0	3	3	0	2	1	3	0	0	43
N	C13	11	4	1	2	3	1	0	0	0	1	0	1	24
	C14	9	3	0	0	0	1	0	0	1	0	1	1	16
	C15	15	3	3	1	0	2	1	2	0	2	0	1	30
	C16	3	3	3	0	0	1	0	0	1	0	0	0	11
	TOTAL	449	212	115	28	77	52	16	71	50	59	22	23	1174

TABLA 4.3

Conforme a los datos que se tienen en la tabla 4.3, se desea probar la hipótesis de que el tipo de cáncer es independiente de la delegación a la que pertenece el paciente, contra la alternativa de que no es independiente, es decir:

H0: El tipo de cáncer es independiente de la delegación a la que pertenece el paciente.
Ha: El tipo de cáncer no es independiente de la delegación a la que pertenece el paciente.

El resultado de la prueba de bondad de ajuste es como sigue: el valor del estadístico X^2 es de **133.376** con **165** grados de libertad y una significancia de **0.966300**. Como vemos, el nivel de significancia de la prueba es bastante considerable y nos lleva sin lugar a dudas a aceptar la hipótesis nula de que los datos son independientes.

III) INSTITUCION DE SALUD DEL PACIENTE Y TIPO DE CANCER

La institución de salud también es considerada dentro del estudio, ya que por medio de ellas es como se recolectó la información obtenida. Podemos inferir que de acuerdo al hospital al que se incurre está la clase social a la que pertenece el paciente. Los hospitales considerados son 6, los cuales se presentan dentro de las categorías correspondientes a la variable institución de salud.

VARIABLE D : INSTITUCION DE SALUD

CATEGORIAS:

D1: Centro Hospitalario "20 de Noviembre".
D2: Hospital de Pediatría del centro Médico Nacional Siglo XXI.
D3: Hospital General del centro Médico "La Raza" y
Hospital general de la Secretaría de Salud.
D4: Hospital Infantil de México "Dr. Federico Santos".
D5: Instituto Nacional de Pediatría.

Observemos que la categoría D3 considera a dos hospitales, al Centro Médico "La Raza" y al Hospital General de la Secretaría de Salud. Esto se hizo para efectos de consistencia y potencia de la prueba, pues al obtener las frecuencias de estos dos hospitales por separado se observan frecuencias bajas en comparación con las otras categorías. Debido a esto, se decide unir a los dos hospitales anteriormente mencionados, quedando como resultado sólo cinco categorías. Las frecuencias obtenidas se encuentran en la tabla 4.4 que se muestra enseguida:

		TIPO DE CANCER												
		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	TOTAL
INSTITUCION DE SALUD	D1	68	17	11	5	4	2	2	19	7	12	2	3	152
	D2	86	43	9	6	9	9	2	10	1	5	0	1	181
	D3	24	38	15	4	8	12	1	7	16	5	4	12	146
	D4	67	27	30	1	8	7	2	4	5	11	1	3	166
	D5	210	88	52	13	49	22	9	33	21	26	15	4	542
	TOTAL	455	213	117	29	78	52	16	73	50	59	22	23	1187

TABLA 4.4

La hipótesis nula y la alternativa, para la prueba de bondad de ajuste son:

H0: El tipo de cáncer es independiente de la institución de salud a la que pertenece el paciente.

Ha: El tipo de cáncer no es independiente de la institución de salud a la que pertenece el paciente.

El resultado obtenido es un valor $\mathbf{X^2=164.093}$ con **44** grados de libertad y un nivel de significancia de **9.99201E-16**. De acuerdo a éste último dato, y tomando en cuenta que el valor α elegido es de **0.05**, la **hipótesis nula se rechaza. Por lo tanto, es necesario analizar la asociación que se da entre las variables institución de salud y tipo de cáncer.**

ESTADISTICO	SIMETRICO	RENGLONES DEPENDIENTES	COLUMNAS DEPENDIENTES
Lambda	0.01598	0.01240	0.01913
Coeficiente de Incertidumbre	0.03915	0.04638	0.03386
D de Somers	0.01633	0.01552	0.01722
	VALOR	SIGNIFICANCIA	
Coeficiente de Contingencia	0.34850		
V de Cramers	0.18590		
Gama Condicional	0.02156		
τ_b de Kendall	0.01635	0.48584	
τ_c de Kendall	0.01543		

TABLA 4.5

El coeficiente lambda nos indica un valor mayor para las columnas dependientes. Esto implica un incremento en la probabilidad de acertar en la predicción de la categoría correspondiente al tipo de cáncer, dado que se conoce la institución de salud. Para el caso del coeficiente de incertidumbre, el valor más significativo se encuentra para las renglones dependientes. De acuerdo con el coeficiente de contingencia se tiene una asociación considerable.

En la tabla 4.4 existen varias celdas donde se observa la ruptura de la independencia, pero se pone atención sobre todo en aquéllas con un porcentaje de casos mayor al esperado en caso de independencia. Los pacientes que presentaron linfomas, tumor de Wilms, rabdomiosarcoma y carcinomas tienen una frecuencia en la categoría D3 del 45, 87.6, 60.2, y 47.8 % respectivamente mayor a la esperada. La categoría D3 corresponde al Centro médico La Raza y al Hospital General de la Secretaría de Salud. Lo anterior describe como los cánceres señalados tienen una incidencia mayor en los dos hospitales.

IV) PESO DEL NIÑO AL NACER Y TIPO DE CANCER

Sin duda alguna, el peso del niño al nacer se ha tomado en cuenta en varios estudios médicos y por ende se considera en la presente. Lo que interesa es saber si para alguno o algunos tipos de cáncer, el peso extremo (bajo o alto) resulta un factor de riesgo.

VARIABLE E: PESO DEL NIÑO AL NACER (gramos)

CATEGORIAS:

E1: < 2500.
E2: 2500 - 2999.
E3: 3000 - 3499.
E4: ≥ 3500.

		TIPO DE CANCER												
		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	TOTAL
PESO DEL NIÑO AL NACER	E1	103	65	24	5	14	10	3	15	2	2	1	1	245
	E2	94	32	35	9	20	11	2	13	8	11	3	3	241
	E3	125	52	35	9	23	19	3	26	17	13	10	5	337
	E4	133	65	23	6	21	12	9	19	14	22	5	6	335
	TOTAL	455	214	117	29	78	52	17	73	41	48	19	15	1158

TABLA 4.6

Las hipótesis que se establecen son las siguientes:

H0: El tipo de cáncer es independiente del peso del niño al nacer.
Ha: El tipo de cáncer no es independiente del peso del niño al nacer.

El valor de X^2 es de **62.5853** con **33** grados de libertad. El nivel de significancia alcanzado es de **1.41004E-3**, lo que nos lleva a rechazar la hipótesis nula. El grado de asociación se estudia por medio de los estadísticos que se presentan en la tabla 4.7.

ESTADISTICO	SIMETRICO	RENGLONES DEPENDIENTES	COLUMNAS DEPENDIENTES
Lambda	0.02428	0.04507	0.00000
Coeficiente de Incertidumbre	0.01741	0.02097	0.01489
D de Somers	0.05612	0.05458	0.05774
	VALOR	**SIGNIFICANCIA**	
Coeficiente de Contingencia	0.22644		
V de Cramers	0.13422		
Gama Condicional	0.07335		
τ_b de Kendall	0.05614	0.01876	
τ_c de Kendall	0.05724		

TABLA 4.7

El estadístico lambda nos indica que para el caso de columnas dependientes, éstas recaen en una sola categoría, la leucemia. Por otra parte, los estadísticos lambda y el coeficiente de incertidumbre tienen el valor más grande en las renglones dependientes. Esto indica que si se conoce el tipo de cáncer, la probabilidad de acertar en el peso del niño es mayor que si no se conociera. En la otra dirección, columnas dependientes, tendría más sentido. Es decir, si se conoce el peso del niño, se mejora la probabilidad de acertar en el tipo de cáncer desarrollado.

Para el caso de los neuroblastomas, con un peso del niño al nacer de 2500 a 2999 gramos, se tiene una frecuencia del 49.1 % mayor a la esperada. La frecuencia de casos con tumores hepáticos y un peso al nacer mayor o igual a 3500 gramos es 83% mayor a la esperada. La frecuencia de los casos con tumores de células germinales y un peso mayor o igual a los 3500 gramos fue 58.4 % mayor a la esperada. Por su parte, la frecuencia de los pacientes con carcinomas y un peso al nacer mayor o igual a 3500 gramos supera en un 80.9% a la frecuencia esperada. Las frecuencias esperadas son en caso de independencia. De acuerdo a estos datos, y considerando como un posible factor de riesgo el peso del niño al nacer, observamos que resulta más significativo un sobrepeso, como se indica en las enfermedades señaladas anteriormente, exceptuando a los pacientes que padecieron neuroblastomas.

V) TIPO DE EMBARAZO Y TIPO DE CANCER

Se trata de estudiar también el tipo de embarazo de la madre, es decir, si tuvo el término correspondiente o no. Esto nos puede dar algún indicador sobre el posterior desarrollo de cáncer del paciente.

VARIABLE F: TIPO DE EMBARAZO

CATEGORIAS:

F1: Término.

F2: No término.

		TIPO DE CANCER											
		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	TOTAL
EMBARAZO	F1	428	203	108	28	70	48	15	68	46	55	21	21	1111
	F2	14	6	5	0	5	3	1	1	0	0	1	1	37
	TOTAL	442	209	113	28	75	51	16	69	46	55	22	22	1148

TABLA 4.8

Con los datos que se tienen en la anterior tabla se desea probar la hipótesis de que el tipo de cáncer es independiente del tipo de embarazo de la madre, contra la alternativa de que no es independiente. Se tiene de manera más formal lo siguiente:

H0: El tipo de cáncer es independiente del tipo de embarazo de la madre.
Ha: El tipo de cáncer no es independiente del tipo de embarazo de la madre.

El resultado de la prueba de bondad de ajuste es un valor X^2 de **10.3274** con **11** grados de libertad y un nivel de significancia alcanzado de **0.501201**, lo cual implica la aceptación de que las variables tipo de cáncer y tipo de embarazo no presentan relación alguna.

VI) ESCOLARIDAD DEL PADRE DEL NIÑO Y TIPO DE CANCER

La escolaridad del padre nos revela factores de tipo social y económico en el desarrollo del cáncer pediátrico. La variable se divide en tres categorías: primaria o menos, secundaria y bachillerato o más. Las frecuencias obtenidas con respecto a la variable escolaridad del padre y tipo de cáncer se muestran en la tabla 4.9.

VARIABLE G: ESCOLARIDAD DEL PADRE

CATEGORIAS:

G1: Primaria o menos.
G2: Secundaria.
G3: Bachillerato o más.

		TIPO DE CANCER												
		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	TOTAL
ESCOLARIDAD	G1	182	81	38	5	28	18	6	24	17	26	10	5	440
DEL	G2	88	31	22	8	18	10	5	10	7	9	3	5	216
PADRE	G3	97	35	26	9	14	12	3	24	12	16	4	3	255
	TOTAL	367	147	86	22	60	40	14	58	36	51	17	13	911

TABLA 4.9

Con los datos que se tienen en la anterior tabla se desea probar la hipótesis de que el tipo de cáncer es independiente de la escolaridad del padre, contra la alternativa de que no son independientes. Se tiene de manera más formal lo siguiente:

H_0: El tipo de cáncer es independiente de la escolaridad del padre.
H_a: El tipo de cáncer no es independiente de la escolaridad del padre.

El valor encontrado para la X^2 es de **21.7757** con **22** grados de libertad y un nivel de significancia alcanzado de **0.473346**. Si se toma en consideración que el valor α elegido es de **0.05** se acepta la hipótesis nula. Por lo tanto, no existe razón para pensar en la existencia de asociación entre la escolaridad del padre y el tipo de cáncer del niño.

VII) ESCOLARIDAD DE LA MADRE DEL NIÑO Y TIPO DE CANCER

La escolaridad de la madre también es considerada. Al igual que la escolaridad del padre y en una medida aún mayor, esta variable nos presenta indicadores bastante interesantes sobre el nivel social y cultural de la madre. Esto nos indica los hábitos o creencias que influyen en el comportamiento de la madre ante las enfermedades de los hijos.

VARIABLE H: ESCOLARIDAD DE LA MADRE

CATEGORIAS:

H1: Primaria o menos.
H2: Secundaria o más

		TIPO DE CANCER												
		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	TOTAL
ESCOLARIDAD	H1	244	115	57	9	39	22	4	29	16	26	13	9	583
DE LA	H2	158	58	41	14	27	22	10	36	23	26	4	6	425
MADRE	TOTAL	402	173	98	23	66	44	14	65	39	52	17	15	1008

TABLA 4.10

De acuerdo a las frecuencias obtenidas, se desea probar la hipótesis de que el tipo de cáncer es independiente de la escolaridad de la madre, contra la alternativa de que no son independientes. Es decir, se tiene de manera más formal lo siguiente:

H0: El tipo de cáncer es independiente de la escolaridad de la madre.
Ha: El tipo de cáncer no es independiente de la escolaridad de la madre.

El resultado obtenido es una valor de X^2=28.9512 con 11 grados de libertad. El nivel de significancia de la prueba es de 2.30987E-3, lo que nos conduce a rechazar la hipótesis nula de independencia. El grado de asociación entre las variables se estudia mediante los siguientes estadísticos:

ESTADISTICO	SIMETRICO	RENGLONES DEPENDIENTES	COLUMNAS DEPENDIENTES
Lambda	0.02425	0.05882	0.00000
Coeficiente de Incertidumbre	0.01097	0.02114	0.00741
D de Somers	0.07573	0.06135	0.09891
	VALOR	SIGNIFICANCIA	
Coeficiente de Contingencia	0.16709		
V de Cramers	0.16947		
Gama Condicional	0.12483		
τ_b de Kendall	0.07790	0.00537	
τ_c de Kendall	0.09648		

TABLA 4.11

El estadístico lambda, en cuanto a columnas dependientes, indica que el valor cae en una sóla categoría, la correspondiente a la leucemia. El valor para las renglones dependientes es un poco considerable; éste presenta el mayor índice de asociación en comparación a los otros dos casos. Esto también lo demuestra el coeficiente de incertidumbre. Lo anterior nos indica que si se conoce el tipo de cáncer al que pertenece el paciente, se tiene mayor certeza en la predicción de la escolaridad de la madre. Esto no tiene significado en términos prácticos, pero si se toma en cuenta la otra lambda, la de columnas dependientes, si tiene significado. Lo anterior nos describe la mejora que se tiene en la predicción del tipo de cáncer al conocer la escolaridad de la madre. Esto resulta bastante interesante, puesto que se ha dado gran énfasis al hecho de que la madre es quien se encarga directamente del niño, desde la gestación, la alimentación, educación y otra serie de factores que pueden influir en el desarrollo de las enfermedades del niño. El coeficiente de contingencia y la V de Cramers presentan una asociación más o menos considerable.

En la tabla 4.10 se tiene que la frecuencia de casos con tumores hepáticos cuya madre tenía una escolaridad de secundaria o más es 69.4 % mayor a la esperada en caso de independencia.

VIII) EDAD DE LA MADRE AL NACIMIENTO DEL HIJO Y TIPO DE CANCER

La edad de la madre al nacimiento del hijo tiene una importancia relevante, ya que nos indica si existe una edad de riesgo en procrear un hijo que posteriormente desarrolle algún tipo de cáncer a temprana edad.

VARIABLE I: EDAD DE LA MADRE AL NACIMIENTO DE SU HIJO (años).

CATEGORIAS:

I1: 15-19.
I2: 20-24.
I3: 25-29.
I4: 30-34.
I5: 35 y más.

		TIPO DE CANCER												
		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	TOTAL
EDAD DE LA	I1	55	26	18	4	8	6	4	4	7	7	0	3	142
MADRE	I2	142	71	28	11	30	19	6	26	18	21	7	4	383
AL	I3	110	52	32	8	23	12	4	22	11	11	4	6	295
NACIMIENTO	I4	88	26	16	3	11	5	0	13	6	12	5	7	192
DEL HIJO	I5	49	26	15	1	4	7	3	6	4	3	4	0	122
	TOTAL	444	201	109	27	76	49	17	71	46	54	20	20	1134

TABLA 4.12

Con los datos que se tienen en la tabla 4.12 se desea probar la hipótesis de que el tipo de cáncer es independiente de la edad de la madre al nacimiento del niño, contra la alternativa de que no es independiente. De manera más formal se tiene que:

H0: El tipo de cáncer es independiente de la edad de la madre al nacimiento del niño.
Ha: El tipo de cáncer no es independiente de la edad de la madre al nacimiento del niño.

El valor que resulta para la X^2 es de **46.1454** con **44** grados de libertad y un nivel de significancia de **0.383638**. Este último valor implica la aceptación de la hipótesis nula de independencia.

IX) TIEMPO TRANSCURRIDO ENTRE EL INICIO DE SIGNOS Y SINTOMAS AL DIAGNOSTICO Y TIPO DE CANCER

Resulta interesante estudiar la relación existente entre el tiempo que transcurre de los síntomas al diagnóstico de la enfermedad, con el tipo de cáncer. Esto nos habla de la dificultad en realizar el diagnóstico de acuerdo a la enfermedad del paciente, aunque también tiene que ver el momento en que los padres comienzan a tratar al niño. De cualquier forma, resulta bastante adecuado analizar estas dos variables. Las frecuencias obtenidas se encuentran en la tabla 4.13.

VARIABLE J: TIEMPO TRANSCURRIDO ENTRE EL INICIO DE SIGNOS Y SINTOMAS Y EL DIAGNOSTICO

CATEGORIAS:

J1: Menor o igual a un mes.
J2: Más de un mes.

		TIPO DE CANCER												
		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	TOTAL
TIEMPO	J1	263	61	40	10	16	29	7	14	17	26	6	9	498
	J2	187	149	73	19	62	22	10	58	32	33	15	13	673
	TOTAL	450	210	113	29	78	51	17	72	49	59	21	22	1171

TABLA 4.13

Se desea probar la hipótesis de que el tipo de cáncer es independiente del tiempo transcurrido entre el inicio de signos y síntomas al diagnóstico, contra la alternativa de que no son independientes. Por lo tanto, se tienen las siguientes hipótesis:

Ho: El tipo de cáncer es independiente del tiempo transcurrido entre el inicio de signos y síntomas y el diagnóstico.

Ha: El tipo de cáncer no es independiente del tiempo transcurrido entre el inicio de signos y síntomas y el diagnóstico.

El resultado de la prueba de bondad de ajuste es un valor de **103.822** para la X^2 con **11** grados de libertad. El nivel de significancia alcanzado es 0, por lo tanto, la hipótesis nula se rechaza. Ahora corresponde analizar la asociación entre las variables de estudio.

ESTADISTICO	SIMETRICO	RENGLONES DEPENDIENTES	COLUMNAS DEPENDIENTES
Lambda	0.06809	0.16667	0.00000
Coeficiente de Incertidumbre	0.03437	0.06691	0.02313
D de Somers	0.17432	0.14075	0.22889
	VALOR	**SIGNIFICANCIA**	
Coeficiente de Contingencia	0.28538		
V de Cramers	0.29776		
Gama Condicional	0.28450		
τ_b de Kendall	0.17949	0.00000	
τ_c de Kendall	0.22378		

TABLA 4.14

Se observa en la tabla 4.12 que el valor del estadístico lambda cae en una sola categoría del tipo de cáncer, la leucemia. Esto es claro, ya que se observa un valor de 0 para el caso de columnas dependientes, además de que es fácil visualizar en la tabla de contingencia como la columna con las frecuencias mayores es la correspondiente a la leucemia. Por otra parte, tanto en lambda como en el coeficiente de incertidumbre se observa un valor mayor para el caso de renglones dependientes. Lo anterior indica que si se conoce el tipo de cáncer al que pertenece un paciente, existe una cierta mejora en la probabilidad de acertar en el tiempo transcurrido entre el inicio de signos y síntomas. Es claro que resulta mucho más útil considerar el caso contrario, es decir la mejora al predecir el tipo de cáncer dado que se conoce el tiempo transcurrido entre el inicio de signos y síntomas. El coeficiente de contingencia y la V de Cramers nos indican una asociación considerable.

En la tabla 4.13 se tiene que la frecuencia de casos con tumores óseos y un tiempo transcurrido de más de un mes es un 40.2% mayor a la esperada en caso de independencia.

X) EDAD DEL NIÑO AL DIAGNOSTICO Y TIPO DE CANCER

Es importante determinar si existe relación entre el tipo de cáncer y la edad del niño. Para obtener la tabla de contingencia con respecto a estas variables se divide la edad del paciente en tres categorías: de 0 a 4, 5 a 9 y de 10 a 14 años. Las frecuencias obtenidas se presentan en la tabla 4.15.

VARIABLE K: EDAD DEL NIÑO AL DIAGNOSTICO DE CANCER (AÑOS)

CATEGORIAS:

K1: 0-4.
K2: 5-9.
K3: 10-14.

		TIPO DE CANCER												
EDAD		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	TOTAL
DEL	K1	210	60	47	19	70	44	14	3	25	33	6	9	540
NIÑO	K2	133	88	36	7	4	8	2	27	9	14	7	6	341
AL	K3	86	53	31	2	2	0	0	41	9	9	7	3	243
DIAGNOSTICO	TOTAL	429	201	114	28	76	52	16	71	43	56	20	18	1124

TABLA 4.15

Se requiere probar la hipótesis de que el tipo de cáncer es independiente de la edad del niño al diagnóstico contra la alternativa de que no es independiente. De acuerdo a esto, se tiene lo siguiente:

H0: El tipo de cáncer es independiente de la edad del niño al diagnóstico.
Ha: El tipo de cáncer no es independiente de la edad del niño al diagnóstico.

El resultado de la prueba de bondad de ajuste nos indica un valor para la X^2 de **216.548** con **22** grados de libertad y un nivel de significancia de **0**. Por lo tanto, la hipótesis nula de independencia es rechazada. Queda ahora por estudiar el grado de su asociación, para esto se tiene la tabla 4.16.

ESTADISTICO	**SIMETRICO**	**RENGLONES DEPENDIENTES**	**COLUMNAS DEPENDIENTES**
Lambda	0.05238	0.11473	0.00000
Coeficiente de Incertidumbre	0.07235	0.10439	0.05536
D de Somers	-0.01951	-0.01748	-0.02207
	VALOR	**SIGNIFICANCIA**	
Coeficiente de Contingencia	0.40192		
V de Cramers	0.31037		
Gama Condicional	-0.02752		
τ_b de Kendall	-0.01964	0.43342	
τ_c de Kendall	-0.02087		

TABLA 4.16

El estadístico lambda indica que para el caso de columnas dependientes el valor cae en una sola categoría, la leucemia. Al igual que en las otras tablas que presentan a lambda con un valor de 0 en las columnas dependientes, se debe a las frecuencias mayores en esta categoría con respecto a las otras once

restantes. De acuerdo a los estadísticos lambda y el coeficiente de incertidumbre, el decremento de la probabilidad de error, conociendo el tipo de cáncer al predecir la edad del paciente, es del 11.473 y 10.439 % respectivamente. El coeficiente de contingencia y la V de Cramers tienen un valor considerable.

En la tabla 4.15 se observa que para los casos del retinoblastoma, tumor de Wilms y tumores hepáticos con una edad de 0 a 4 años al diagnóstico se tiene una frecuencia de un 91.7, 76.1, 82.1 %, respectivamente, mayor a la esperada en caso de independencia. Por otro lado, los casos con tumores óseos y los carcinomas con una edad al diagnóstico de 10 a 14 años, tienen una frecuencia de un 67.1 y 61.9 %, respectivamente, mayor a la esperada en caso de independencia.

XI) FAMILIARES (C/S) CANCER Y TIPO DE CANCER

En muchas ocasiones las enfermedades tienen un carácter hereditario. En base a esto, se decide hacer la clasificación de los pacientes de acuerdo a dos categorías; los que tuvieron familiares con cáncer y los que no. Las frecuencias obtenidas se encuentran en la tabla 4.17.

VARIABLE L: FAMILIARES CON CANCER

CATEGORIAS:

L1: Si.

L2: No.

		TIPO DE CANCER												
FAMILIARES		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	TOTAL
(C/S)	L1	68	38	13	5	9	14	3	24	11	6	3	1	195
CANCER	L2	385	172	102	24	68	38	14	48	38	52	19	22	982
	TOTAL	453	210	115	29	77	52	17	72	49	58	22	23	1177

TABLA 4.17

Con los datos que se tienen en la anterior tabla, se desea probar la hipótesis de que el tipo de cáncer es independiente de la existencia de familiares con cáncer, contra la alternativa de que no son independientes, es decir, se tiene de manera más formal lo siguiente:

H0: El tipo de cáncer es independiente de la existencia de familiares con o sin cáncer.
Ha: El tipo de cáncer no es independiente de la existencia de familiares con o sin cáncer.

El resultado de la prueba de bondad de ajuste es un valor para la X^2 de **28.9513** con **11** grados de libertad. El nivel de significancia alcanzado es de **2.30978E-3**, por lo que la hipótesis nula es rechazada. Los valores de los estadísticos obtenidos sobre la asociación de las variables son los que se muestran en la tabla 4.18.

ESTADISTICO	SIMETRICO	RENGLONES DEPENDIENTES	COLUMNAS DEPENDIENTES
Lambda	0.00000	0.00000	0.00000
Coeficiente de Incertidumbre	0.00952	0.02570	0.00585
D de Somers	-0.02616	-0.01763	-0.05069
	VALOR	**SIGNIFICANCIA**	
Coeficiente de Contingencia	0.15494		
V de Cramers	0.15684		
Gama Condicional	-0.06296		
τ_b de Kendall	-0.02990	0.24713	
τ_c de Kendall	-0.02803		

TABLA 4.18

En la tabla de 4.18 se observa que el valor del estadístico lambda cae en una sola columna, la leucemia, y una sola fila, los pacientes que no tuvieron familiares con cáncer. El coeficiente de incertidumbre indica un valor mayor para el caso de las renglones dependientes. Por su parte, el coeficiente de contingecia y la V de Cramers presentan el índice de asociación muy similar, aunque no muy significativo.

En la tabla 4.17 se observa que para los casos que presentaron el tumor de Wilms o tumores óseos, y tuvieron familiares con cáncer, tienen una frecuencia mayor con respecto a la esperada en caso de independencia, de un 62.5 y 101.2 % respectivamente.

XII) RAMA DE LA CUAL PROCEDE EL CANCER Y TIPO DE CANCER

De acuerdo a los pacientes que dijeron tener familiares con cáncer, se obtuvo una nueva división. El objetivo es averiguar la rama de la cual procede el cáncer, ya sea paterna o materna.

VARIABLE M: RAMA DE LA CUAL PROCEDE EL CANCER
(paterna o materna)

CATEGORIAS:

M1: Paterna.

M2: Materna.

		TIPO DE CANCER												
RAMA		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	TOTAL
DE LA CUAL	M1	24	8	4	4	5	3	3	8	5	3	3	1	71
PROCEDE	M2	37	24	9	1	3	11	0	14	3	2	0	0	104
EL CANCER	TOTAL	61	32	13	5	8	14	3	22	8	5	3	1	185

TABLA 4.19

Las hipótesis que se plantean son:

H0: El tipo de cáncer es independiente de la rama de la cual procede el cáncer.
Ha: El tipo de cáncer no es independiente de la rama de la cual procede el cáncer.

El valor que se obtiene para la X^2 es de **23.5147** con **11** grados de libertad. El nivel de significancia alcanzado es de **0.0149420**. De acuerdo a un $\alpha = 0.05$, la hipótesis nula es rechazada. Por lo tanto, no son independientes las variables estudiadas. Los índices de asociación se muestran en la tabla 4.20.

ESTADISTICO	SIMETRICO	RENGLONES DEPENDIENTES	COLUMNAS DEPENDIENTES
Lambda	0.08108	0.21127	0.00000
Coeficiente de Incertidumbre	0.05629	0.11102	0.03771
D de Somers	-0.10252	-0.08174	-0.13746
	VALOR	**SIGNIFICANCIA**	
Coeficiente de Contingencia	0.34417		
V de Cramers	0.36656		
Gama Condicional	-0.16686		
τ_b de Kendall	-0.10600	0.11330	
τ_c de Kendall	-0.13257		

TABLA 4.20

De acuerdo a los dos primeros índices de asociación que presenta la tabla 4.20 se observa que el valor mayor lo tiene el caso de renglones dependientes. Esto nos muestra que al conocer el tipo de cáncer se tiene una mejora en la predicción de la rama de la cual procede el cáncer. Para el caso de las columnas dependientes se observa en el estadístico lambda que el valor cae en una sola categoría, la leucemia. El coeficiente de contingencia y la V de Cramers muestran un valor considerable.

Los pacientes que refirieron tener familiares con cáncer por parte de la madre y además presentaron enfermedades tales como neuroblastomas, retinoblastomas, tumores hepáticos, rabdomiosarcoma, tumores de células germinales, carcinomas y mal definidos tienen una frecuencia del 97.2, 54, 46.5, 54, 47.9, 146.5 y 46.5 %, respectivamente, mayor a las esperadas en caso de independencia.

XIII) INTERNAMIENTO DURANTE SU PRIMER AÑO DE VIDA POR ALGUN PROCESO INFECCIOSO Y TIPO DE CANCER

Se considera si el paciente tuvo internamiento por algún proceso infeccioso, ya que esto nos revela de alguna forma el estado de salud del niño en sus primeros meses de vida. Lo que interesa es ver si existe alguna relación con lo anterior y el desarrollo de un posterior cáncer. Las frecuencias las podemos observar en la tabla 4.21.

VARIABLE N: INTERNAMIENTO POR ALGUN PROCESO INFECCIOSO DURANTE SU PRIMER AÑO DE VIDA

CATEGORIAS:

N1: Si.

N2: No.

		TIPO DE CANCER												
INTERNAMIENTO		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	TOTAL
EN SU	N1	33	18	9	4	6	7	3	6	5	1	3	1	96
PRIMER AÑO	N2	410	191	106	23	69	42	13	63	45	56	18	21	1057
DE VIDA	TOTAL	455	214	117	29	78	52	17	73	50	60	22	23	1153

TABLA 4.21

Con los datos que se tienen en la anterior tabla, se desea probar la hipótesis de que el tipo de cáncer es independiente del internamiento del niño por algún proceso infeccioso durante su primer año de vida, contra la alternativa de que no son independientes. Es decir, se tiene de manera más formal lo siguiente:

H0: El tipo de cáncer es independiente del internamiento del paciente por algún proceso infeccioso durante su primer año de vida.

Ha: El tipo de cáncer no es independiente del internamiento del paciente por algún proceso infeccioso durante su primer año de vida.

El resultado de la prueba de bondad de ajuste de la X^2 es **11.3741** con **11** grados de libertad y un nivel de siginificación de **0.412479**. La decisión que se toma es, razonablemente, desechar la hipótesis alternativa y tomar como cierta la nula, es decir, que las variables estudiadas son independientes.

XIV) MALFORMACION EN EL PACIENTE Y TIPO DE CANCER

Otro aspecto que se considera es una posible malformación en el paciente. Las frecuencias obtenidas se encuentran en la tabla 4.22.

VARIABLE O: MALFORMACION EN EL PACIENTE

CATEGORIAS:

O1: Si.

O2: No.

		TIPO DE CANCER												
MALFORMACION		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	TOTAL
EN EL PACIENTE	O1	8	2	1	1	0	2	1	2	1	1	0	0	19
	O2	440	209	116	27	77	50	16	69	49	58	22	22	1155
	TOTAL	448	211	117	28	77	52	17	71	50	59	22	22	1174

TABLA 4.22

Las hipótesis propuestas son:

H_0: El tipo de cáncer es independiente de la existencia o no de una malformación en el paciente.
H_a: El tipo de cáncer no es independiente de la existencia o no de una malformación en el paciente.

El resultado de la prueba de bondad de ajuste de la X^2 es 8.01490 con 11 grados de libertad y un nivel de significancia alcanzado de 0.711968, por lo que se acepta la hipótesis nula. Es decir, las variables son independientes.

XV) ESCOLARIDAD DE LA MADRE Y TIEMPO TRANSCURRIDO ENTRE INICIO DE SIGNOS Y SINTOMAS AL DIAGNOSTICO

En la mayoría de los casos, la forma de actuar ante las enfermedades de los hijos va ligado con el nivel cultural de los padres, en forma específica, el de la madre, ya que es ella la que atiende al niño la mayor parte del tiempo.

VARIABLE H: ESCOLARIDAD DE LA MADRE

CATEGORIAS:

H1: Primaria o menos.
H2: Secundaria o más.

VARIABLE J: TIEMPO TRANSCURRIDO ENTRE EL INICIO DE SIGNOS Y SINTOMAS AL DIAGNOSTICO

CATEGORIAS:

J1: Un mes o menos.
J2: Más de un mes.

		TIEMPO		
ESCOLARIDAD DE LA MADRE		J1	J2	TOTAL
	H1	244	432	676
	H2	175	328	503
	TOTAL	419	760	1179

TABLA 4.23

El resutado de la prueba de bondad de ajuste es un valor de **0.213880** con **1** grado de libertad. El nivel de significancia es **0.643743**. Con la corrección de Yates se tiene un valor de **0.160768** con **1** grado de libertad y el nivel de significancia es de **0.688451**. Por lo tanto, la hipótesis nula es aceptada, y se concluye que no hay asociación entre las variables.

V. ANALISIS DE LAS SERIES DE TIEMPO

El análisis que se realiza sobre series de tiempo para el cáncer pediátrico en el D. F. tiene un carácter marcado en cuanto a su descripción, y un poco a la explicación. Esto es debido al problema que nos ocupa, es decir, una enfermedad mortal que ocurre entre los infantes. También resulta interesante la forma de la gráfica, ya que ésta nos proporciona información de la existencia de tendencia, variación estacional, varianza creciente, etc., lo que nos da un panorama descriptivo sobre el comportamiento del proceso del cáncer pediátrico. Si se encuentra en la gráfica tendencia creciente nos indica que hay un aumento de la población que sufre de esta enfermedad conforme pasa el tiempo, quedando claro que no se omite el importante hecho de que al estar considerando el crecimiento de la serie, también se toma en cuenta el crecimiento de la población de interés para esos años. En el anexo se presenta información sobre la población media proyectada durante los años 1981-1991 bajo la hipótesis de fecundidad programática y fecundidad alternativa. De acuerdo a esta información, se concluye que la población sufre un decremento para el período considerado, por lo que al realizar el análisis de las series de tiempo, el factor de crecimiento de la población no afecta nuestro análisis, en todo caso, se está haciendo una subestimación de la tasa de crecimiento al no considerar el decremento de la población del D.F. Por lo tanto, no existe ningún inconveniente en utilizar los datos en su forma original.

La manera de utilizar la información proporcionada para los propósitos de la presente, consiste en retomar la mayor parte de las variables contempladas en el instrumento de recopilación de datos utilizada en Centro Médico y dar origen a diversas series bajo los siguientes criterios:

En primer lugar se considera el número de casos detectados con cáncer pediátrico de forma mensual, "serie general". Esta será el punto de partida para el análisis de las subsecuentes series, ya que a partir de ella, se subdividirá de acuerdo a las variables de estudio marcadas en el análisis previo. Resulta importante aclarar que para hacer el desglose de diversas series de acuerdo a las variables de estudio, éstas tomarán su división de acuerdo a la naturaleza de ellas mismas, así como de sus características. Por ejemplo, al dividir los datos que se tienen por sexo, resulta que el número de frecuencias tanto para el sexo femenino, como para el masculino son convenientes en un análisis de series de tiempo, pues no presentan demasiados ceros, además de que siempre será interesante el comportamiento de éstas, dado que en los estudios ya realizados en diferentes enfermedades se han encontrado diferencias significativas entre los dos sexos. De igual forma se intentará realizar diferentes divisiones con las demás variables, pero, se insiste, no se debe olvidar que es primordial tener datos adecuados para realizar un análisis objetivo, porque de no ser así, probablemente éstas no representen más que algo semejante al ruido blanco, o que en realidad no nos diga nada en términos médicos. Luego entonces, en cada división se explicará la razón del por qué está hecha de esa forma y se tratará de hacer una comparación dentro de la variable correspondiente.

La concepción del niño, el inicio de síntomas y la fecha de nacimiento se dividen en las cuatro estaciones del año. Sin embargo, los datos no se prestan para llegar a obtener series adecuadas, por lo que se omiten en el análisis. La variable peso del niño al nacer resulta interesante en cuanto a información médica, pero las frecuencias que se ajustaron fueron de menos de 3000 gramos y de 3000 gramos o más, lo cual no dice mucho en términos médicos.

En cuanto a la forma de presentar los resultados del análisis de cada una de las series aceptadas, en primer lugar se explican los modelos encontrados de acuerdo a sus transformaciones, haciendo referencia a las gráficas que ilustran los razonamientos que condujeron a la aceptación de dichos modelos. Después se presenta una tabla comparativa con los diferentes modelos, donde se indica el número de diferencias ordinarias y estacionales, el valor estimado de cada uno de los parámetros, el valor de la varianza y la desviación estándar de los residuales, el número de iteraciones y otros estadísticos, los cuales nos ayudan a llevar a cabo la discusión sobre la elección del modelo más adecuado. Es importante que tengamos en cuenta que los modelos que se presentan en el capítulo son los que finalmente se consideraron con un mejor ajuste, y que no fueron los únicos en ser probados. Para el caso de la serie general, es la única donde se muestra con más detalle la forma en que se llega a los modelos que mejor representan al proceso, ya que se ilustra el procedimiento con las gráficas correspondientes a los modelos que se acercaron al elegido.

I) SERIE GENERAL

Resulta obvia la razón del por qué se considera en forma prioritaria a la serie de tiempo que se ha denominado "General", ésta nos da la principal pauta para identificar características especiales e importantes en un estudio como éste. Para ello, como primer punto se toman las frecuencias obtenidas de los once años considerados y se representan gráficamente, como se observa en la gráfica 5.1. En ella existen características que a simple vista es posible identificar, por ejemplo, se observa una clara evidencia de tendencia creciente, pues así lo indican los picos irregulares que se forman con los datos, los cuales van alcanzando mayor altura, conforme el tiempo, para este caso los meses, van transcurriendo. La variación estacional no se hace presente, o al menos no se visualiza en primera instancia de una forma clara, aunque es posible pensar en que los registros o casos detectados tuvieran alguna relación con los meses del año.

Respecto a la varianza de la gráfica en cuestión es un poco aventurado afirmar que es constante, pues bien, puede ser un poco creciente. Debido a esto último, se decide hacer el análisis de la serie, por un lado en su forma original y por otro aplicarle la transformación raíz cuadrada. Luego entonces, el análisis para la serie de tiempo en estudio, la general, se realiza en dos partes.

SERIE GENERAL SIN NINGUNA TRANSFORMACION

Como se mencionó anteriormente, la serie general manifiesta un claro crecimiento, por lo que es necesario aplicarle una diferencia ordinaria. La gráfica resultante junto con su ACF (Función de Autocorrelación) y PACF (Función de Autocorrelación Parcial) se observan en la gráfica 5.2. Si ponemos atención a la serie general con la diferencia ordinaria, ya tiene media constante, a su vez no se observa varianza creciente ni una posible variación estacional. Ya no es necesario aplicarle otra diferencia. En la ACF se hace bastante notoria la primera función de autocorrelación, además parece truncarse en 1. En cuanto a la PACF, son sobresalientes la 1,2 y 3, es decreciente infinita. Estos datos nos ayudarán a elegir el modelo o varios modelos que sea posible ajustar. El primer modelo tentativo para esta serie es el ARIMA(0,1,1), el cual parece ser adecuado. Observemos la gráfica 5.3, donde la ACF y PACF de los residuales no presentan valores significativos, es decir, tienen un comportamiento similar al ruido blanco. En lo que respecta al periodograma integral de los residuales se tiene un ajuste, que si no es excelente, es bastante aceptable. El siguiente modelo que presenta un buen ajuste es el ARIMA(3,1,0). Para dicho modelo podemos apreciar su ACF, PACF y periodograma integral de los residuales en la gráfica 5.4, donde las funciones de autocorrelación y autocorrelación parcial son muy pequeñas. Sólo en el valor 18 hay un valor un poquito sobresaliente, valor que en realidad no tiene gran importancia, puesto que la función de autocorrelación está muy alejada, además de que el valor no es muy significativo.

Hasta aquí se han analizado los datos sin ninguna transformación y con tan sólo la diferencia ordinaria. Ahora pongamos atención a la gráfica 5.5, en donde se muestra de nuevo a la serie general, pero

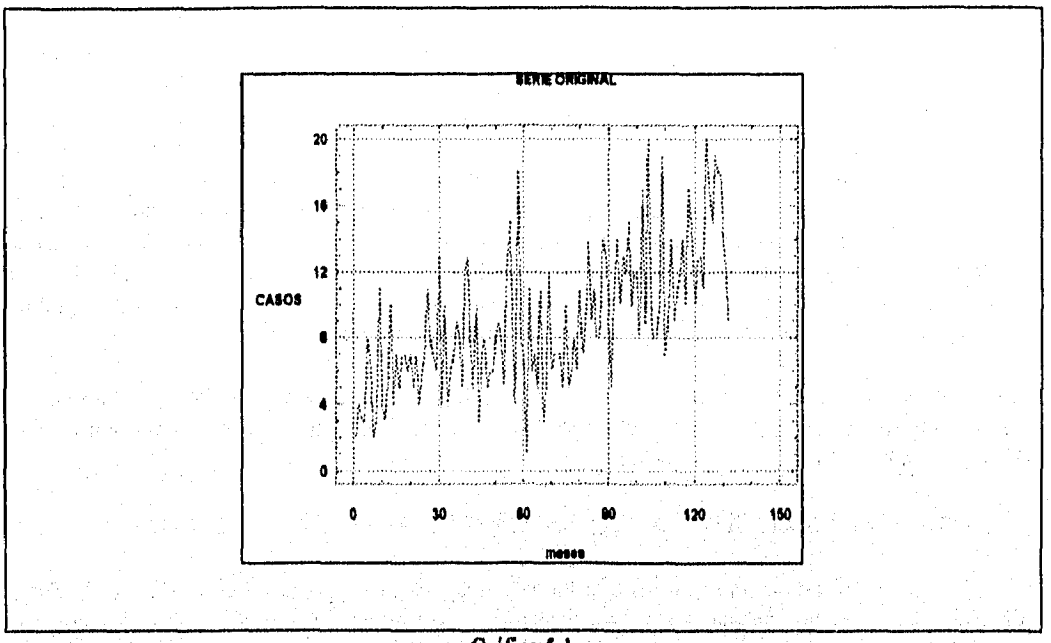
con una diferencia ordinaria y una estacional. El agregarle la diferencia estacional tiene como propósito observar el efecto que le causa a los datos, ya que es de esperarse un valor significativo en 12, tanto en la ACF como en la PACF, en forma similar a lo que ocurre cuando al ruido blanco se le aplica una diferencia ordinaria y se llega al modelo ARIMA(0,1,1). En esta gráfica, la 5.5, observamos primero a la serie ya estacionaria, pues es evidente la media y la varianza constante, además de que no hay fluctuación estacional. En su ACF correspondiente vemos valores significativos en 1 y 12, en la PACF sobresalen el 1,2,3 y 12. En base a esto, es posible intentar con varios modelos alternativos, como es natural, se comienza con modelos bastante sencillos de acuerdo al principio de parsimonia, hasta encontrar un modelo adecuado como es el ARIMA(3,1,0). Para este modelo se observa en las funciones de autocorrelación y autocorrelación parcial de los residuales un valor significativo en 12 (ver gráfica 5.6), motivo por el cual se ajusta a un ARIMA(3,1,0)x(0,1,1). Para dicho modelo podemos visualizar su ACF, PACF y periodograma integral de los residuales en la gráfica 5.7, los cuales presentan un ajuste bastante bueno.

Ahora que se han encontrado los modelos para la serie general sin ninguna transformación, se aplica la transformación raíz cuadrada, y se procede de forma semejante al análisis hecho anteriormente con la serie general sin ninguna transformación.

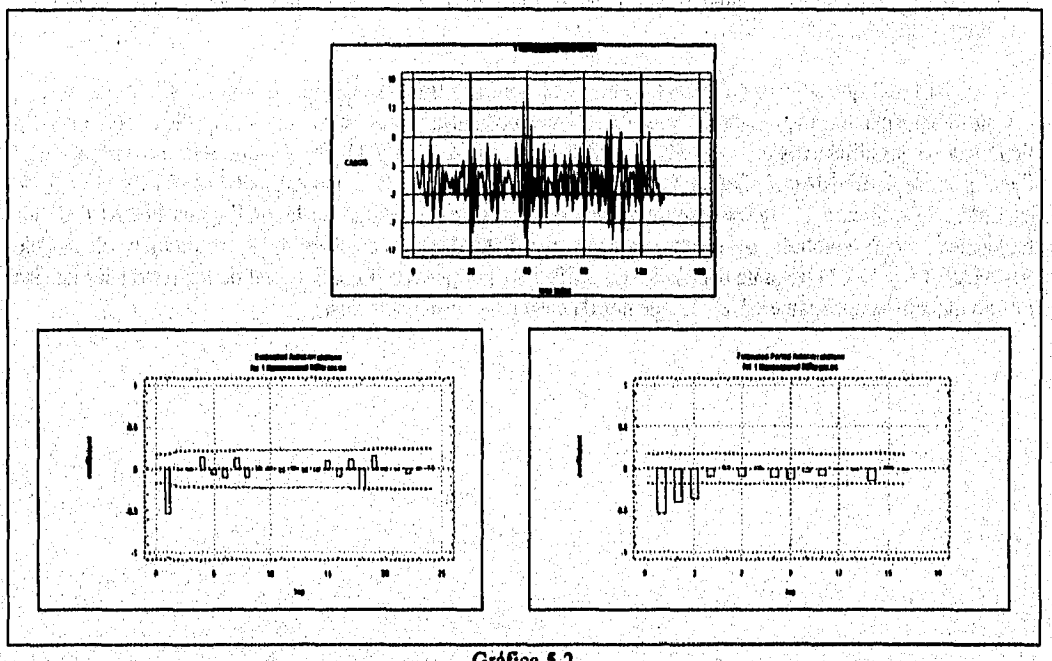
SERIE GENERAL CON LA TRANSFORMACION RAIZ CUADRADA

La serie general que ya incluye la transformación raíz cuadrada se observa en la gráfica 5.8. Como podemos ver, el único aspecto que falta por atender para transformar la serie en estacionaria, es la tendencia, la cual es creciente, por lo que se le aplica una diferencia ordinaria. La serie resultante, al igual que su ACF y PACF las vemos en la gráfica 5.9. Ahora los datos presentan estacionaridad, la ACF tiene un valor sobresaliente en 1, la PACF tiene a su vez valores significativos en 1,2 y 3. El modelo apropiado para esta serie es el ARIMA(3,1,0). En la gráfica 5.10 lo podemos ver, puesto que la ACF y PACF de los residuales no presentan valores significativos. Además, el periodograma integral de los residuales manifiesta un buen ajuste, por lo que se tiene un modelo adecuado para la serie general con la transformación raíz cuadrada y una diferencia ordinaria.

Al igual que se hizo en la serie general sin ninguna transformación, se hará el estudio de la serie con la transformación raíz cuadrada y una diferencia ordinaria y una estacional, ver gráfica 5.11. La ACF tiene valores significativos en 1,4 y 12, la PACF los tiene en 1,2,3 y 11. De acuerdo a esta información se llega al modelo ARIMA(3,1,0)x(0,1,1). Para ello, primero se prueba el modelo ARIMA(3,1,0), el cual, de acuerdo a la gráfica 5.12, se observa un valor significativo en 12, tanto en la ACF como la PACF de los residuales, luego entonces se decide agregar al ARIMA(3,1,0) el $SMA(1)_{12}$, resultando el modelo ARIMA(3,1,0)x(0,1,1) bastante adecuado. La ACF, PACF y periodograma integral de los residuales de este último modelo se aprecian en 5.13, lo que nos da evidencia de su buen ajuste.

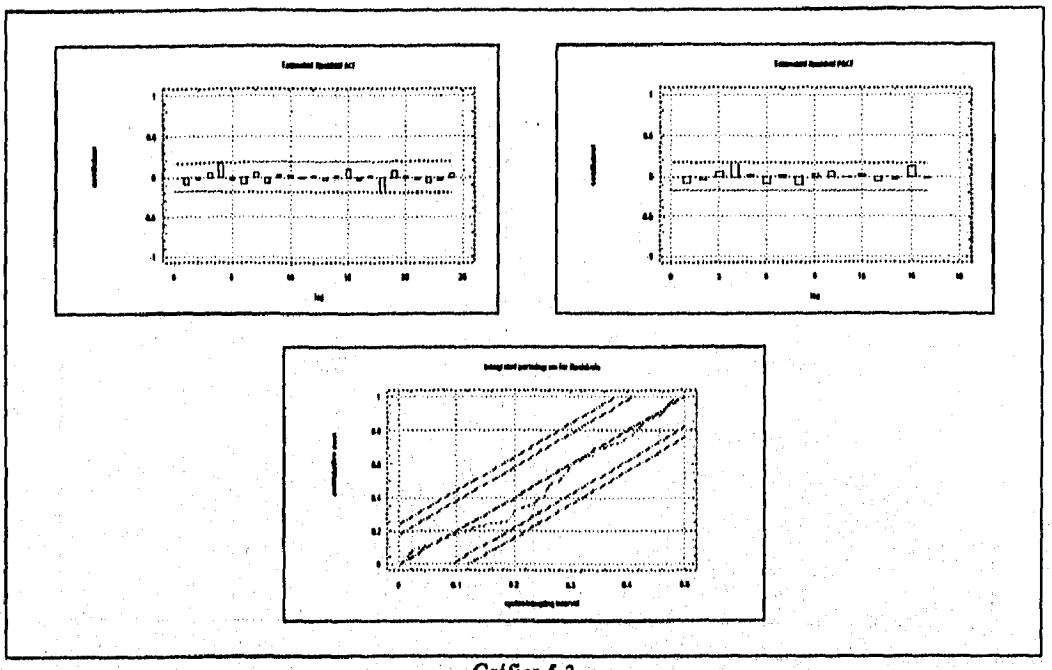


Gráfica 5.1
(SERIE GENERAL)
NUMERO DE CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991.

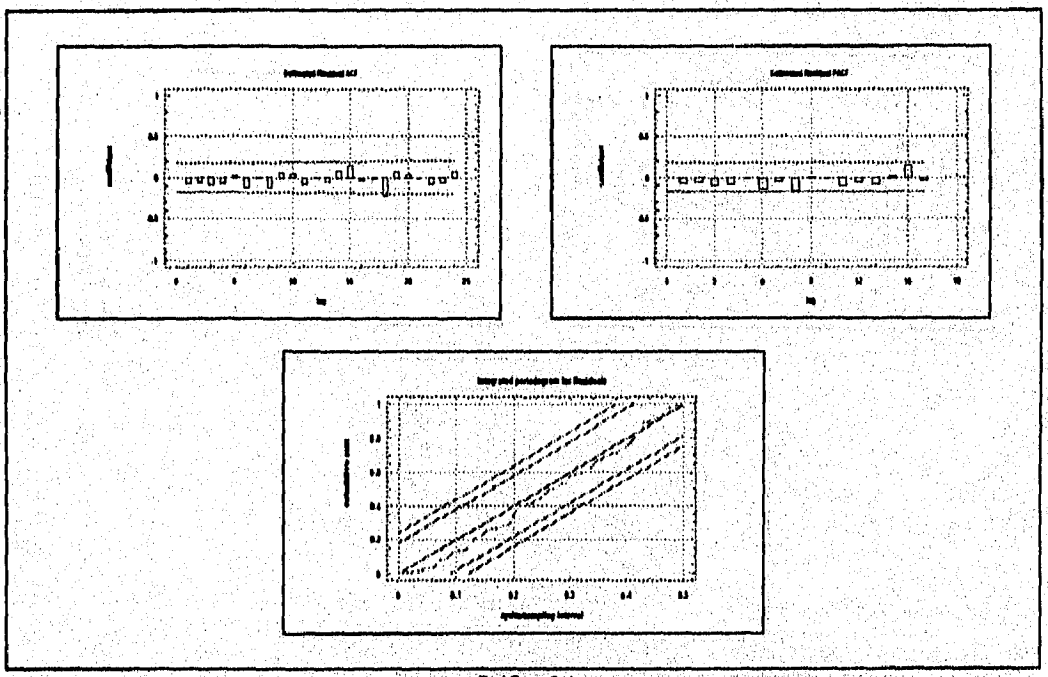


Gráfica 5.2
SERIE GENERAL
CON UNA DIFERENCIA ORDINARIA, ACF y PACF.

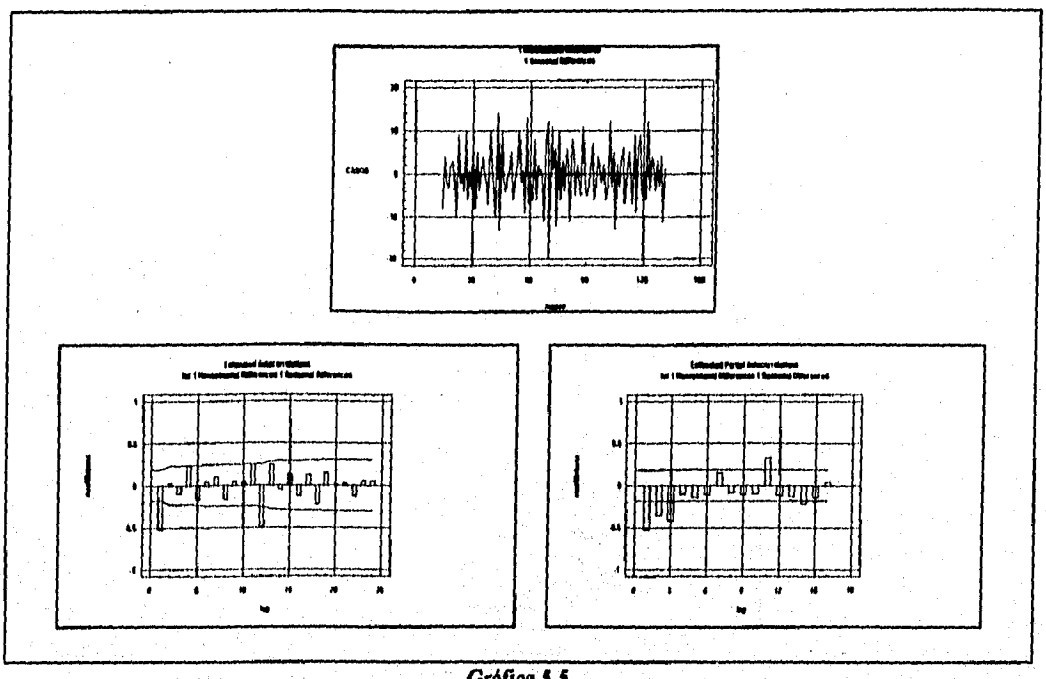
ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA



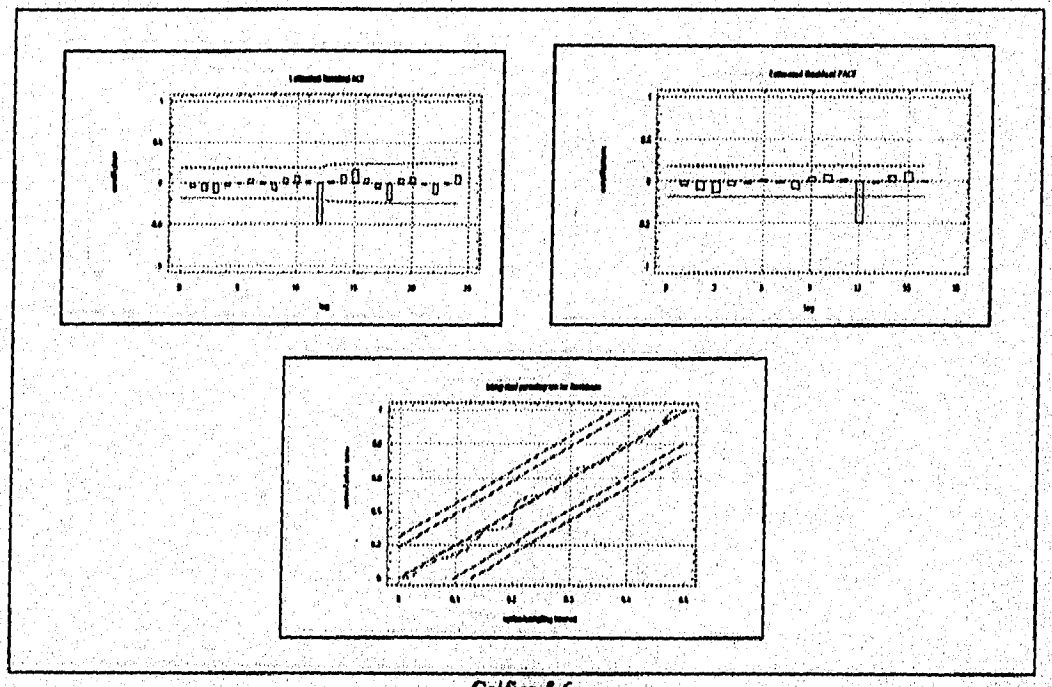
Gráfica 5.3
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(SERIE GENERAL, MODELO: ARIMA(0,1,1)).



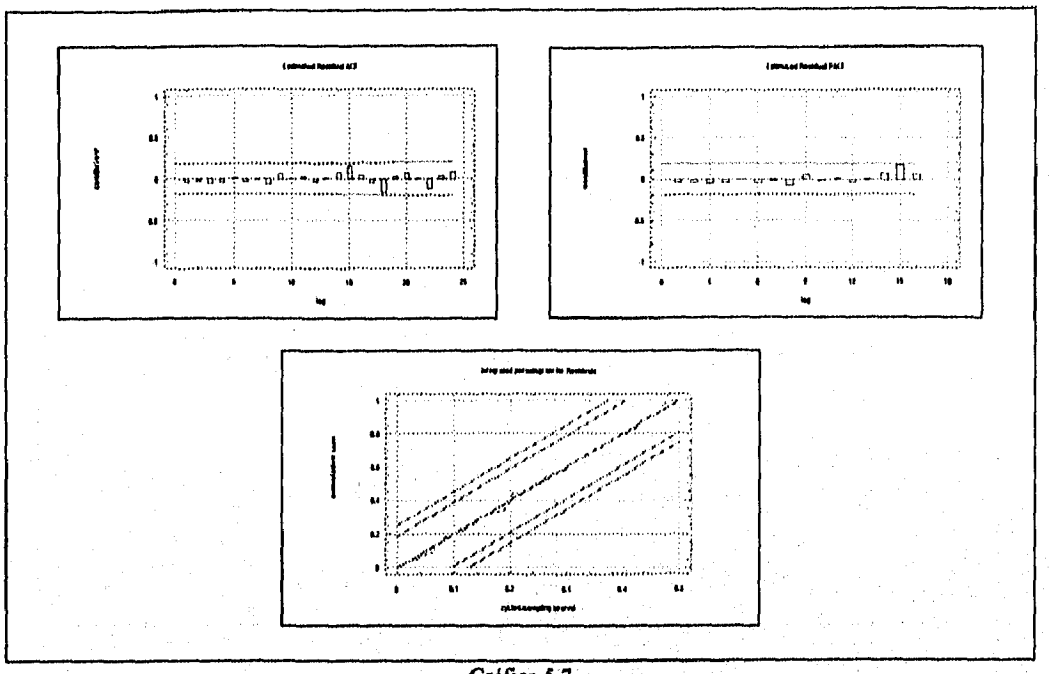
Gráfica 5.4
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(SERIE GENERAL, MODELO: ARIMA(3,1,0)).



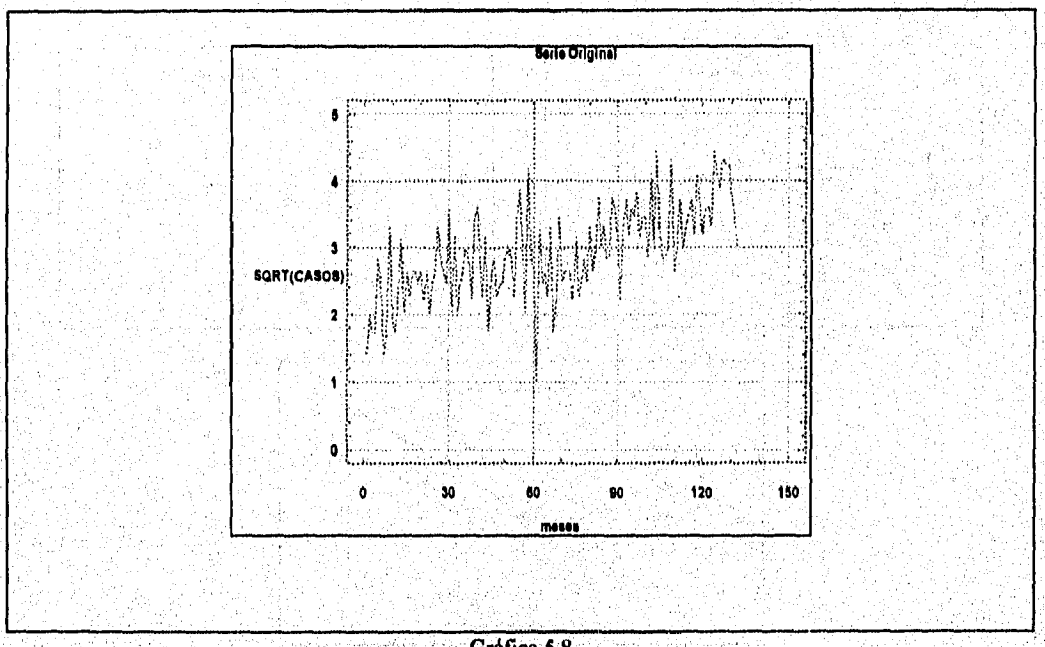
Gráfica 5.5
SERIE GENERAL
CON UNA DIFERENCIA ORDINARIA Y UNA ESTACIONAL, ACF Y PACF.



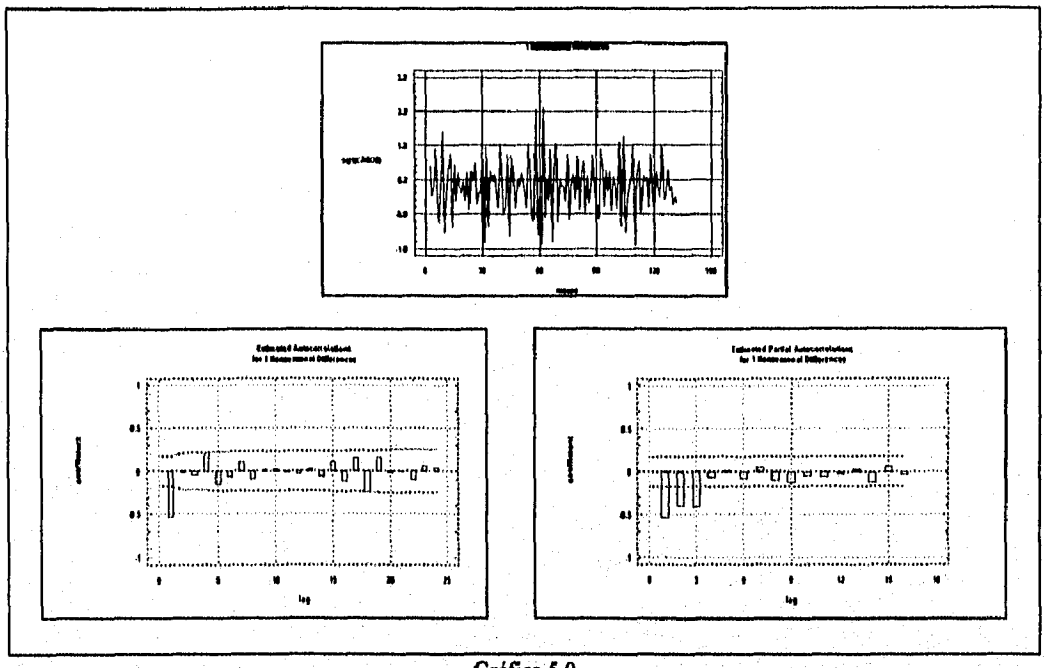
Gráfica 5.6
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(SERIE GENERAL, MODELO: ARIMA(3,1,0)x(0,1,0)).



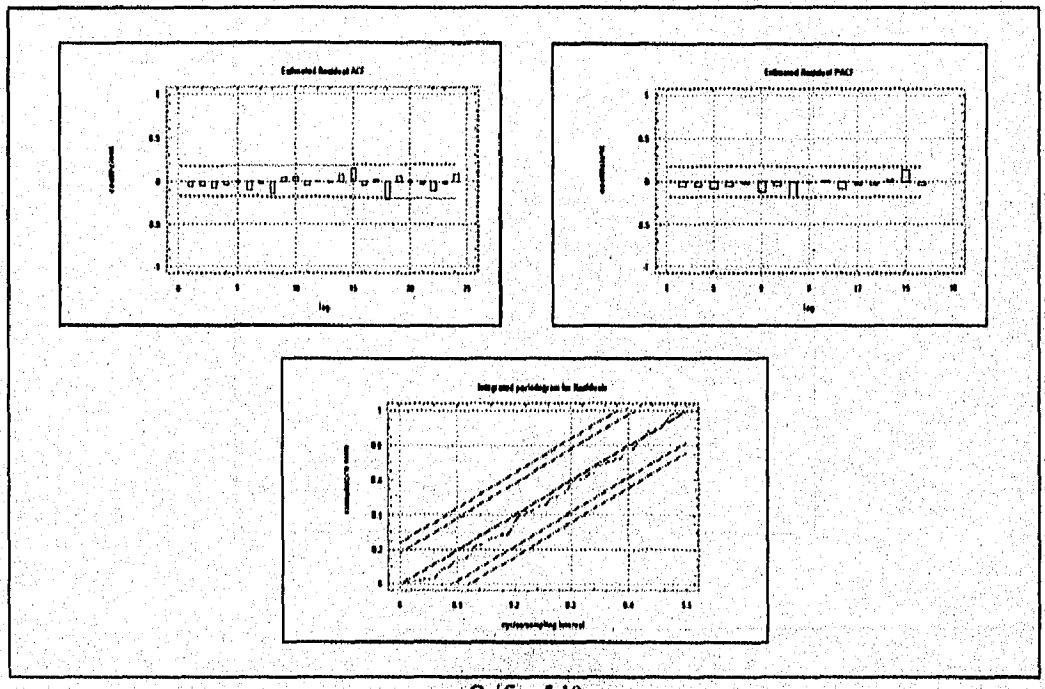
Gráfica 5.7
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(SERIE GENERAL, MODELO: ARIMA(3,1,0)x(0,1,1)).



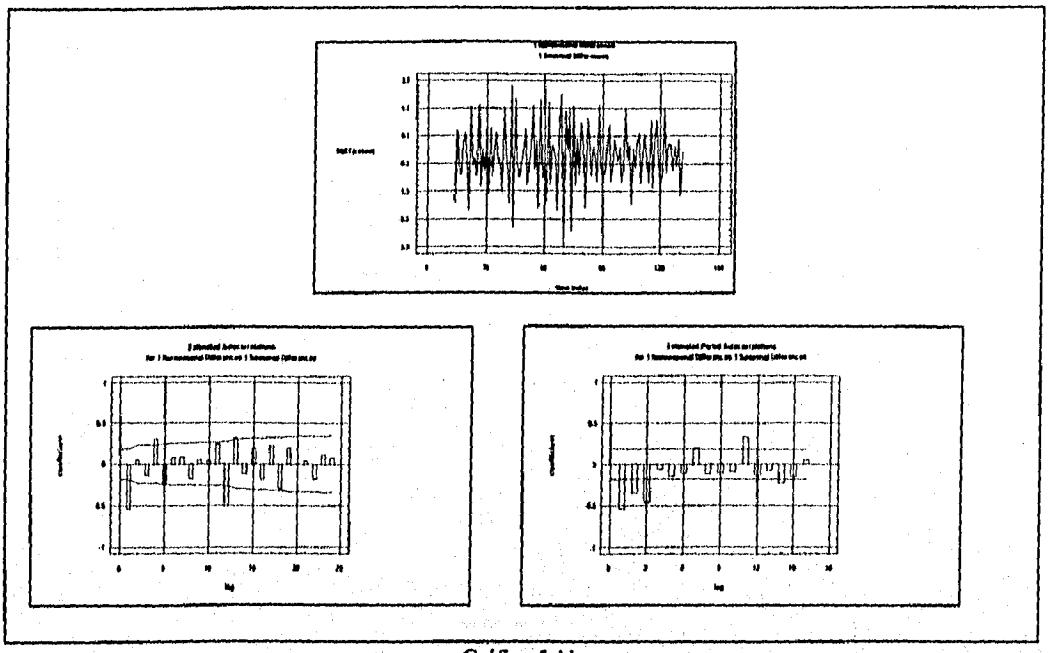
Gráfica 5.8
SERIE GENERAL
CON LA TRANSFORMACION RAIZ CUADRADA.



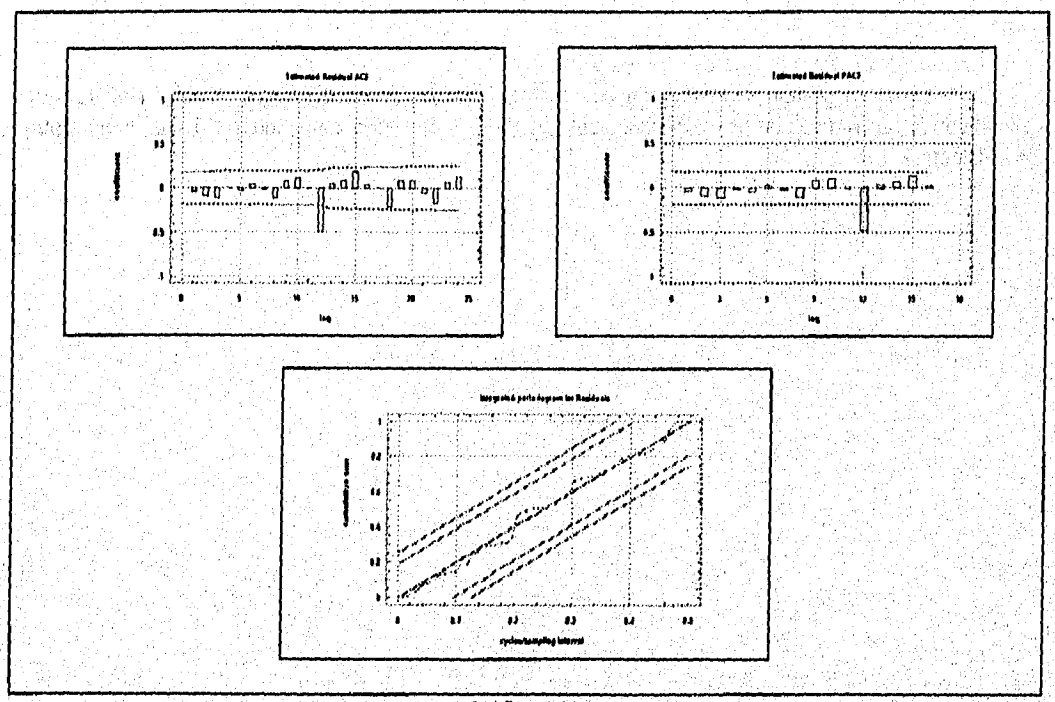
Gráfica 5.9
SERIE GENERAL CON LA TRANSFORMACION RAIZ CUADRADA
Y UNA DIF. ORDINARIA, ACF Y PACF.



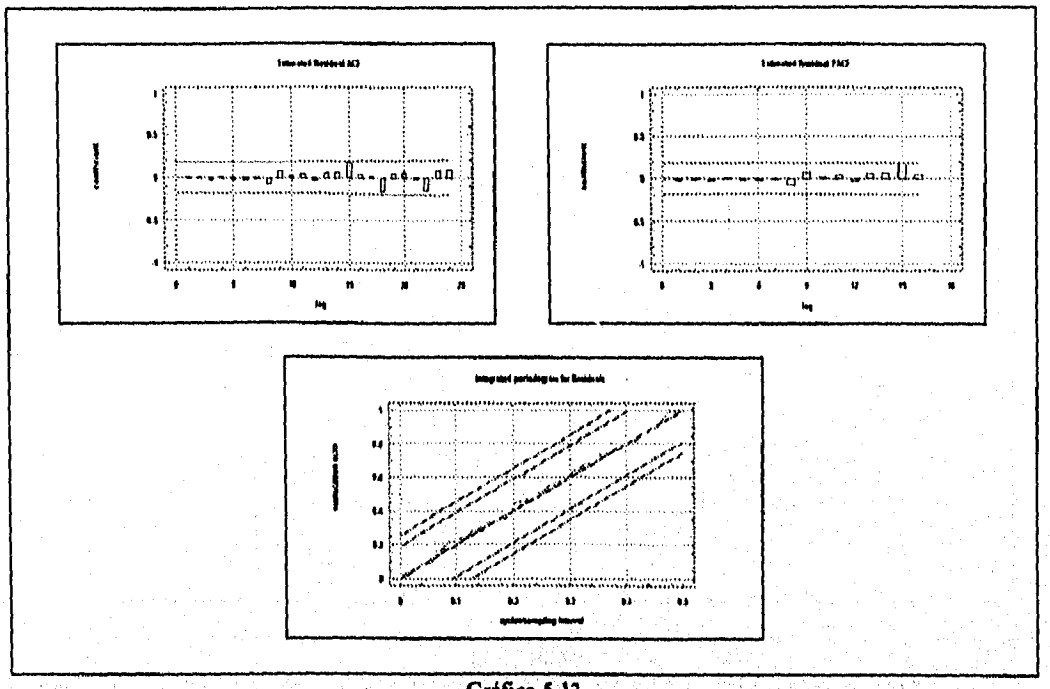
Gráfica 5.10
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(SERIE GENERAL, MODELO: ARIMA(3,1,0)).



Gráfica 5.11
RAIZ CUADRADA DE LA SERIE GENERAL CON UNA DIFERENCIA ORDINARIA Y UNA ESTACIONAL, ACF Y PACF.



Gráfica 5.12
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(SERIE GENERAL, MODELO: ARIMA(3,1,0)x(0,1,0)).



Gráfica 5.13
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(SERIE GENERAL, MODELO: ARIMA(3,1,0)x(0,1,1)).

Ahora se presentan los principales parámetros estimados de los diferentes modelos para la serie general con la transformación raíz cuadrada, como sin ella, en el cuadro comparativo 5.1 que se encuentra más adelante.

SERIE GENERAL					
Modelo	ARIMA(0,1,1)	ARIMA(3,1,0)	ARIMA (3,1,0)x(0,1,1)	ARIMA(3,1,0)	ARIMA (3,1,0)x(0,1,1)
Transformación	--	--	--	SQRT (Raíz cuadrada)	SQRT (Raíz cuadrada)
# de diferencias ordinarias	1	1	1	1	1
# de diferencias estacionales	--	--	1	--	1
# de parámetros	1	3	4	3	
ϕ(1) (est. t)*	--	-.88530 -10.63083	-.84818 -9.86140	-.90366 -11.18453	-.86593 -10.23934
ϕ(2) (est. t)	--	-.66913 -6.77161	-.67147 -6.76985	-.70382 -7.35355	-.69669 -7.15791
ϕ(3) (est. t)	--	-.37146 -4.42602	-.41211 -4.76070	-.42035 -5.17876	-.44717 -5.28797
θ(1) (est. t)	.90634 25.73031	--	--	--	--
Φ(1) (est. t)	--	--	--	--	--
Θ(1) (est. t)	--	--	.70551 9.60901	--	.63996 8.25997
Varianza	10.9535	11.3921	15.7331	0.311557	0.461047
Desviación estándar	3.3096	3.37521	3.96649	0.558173	0.679004
Iteraciones	3	2	4	1	4

Cuadro comparativo 5.1

*: Estadístico t= al cociente del estimador entre su error estándar.

De acuerdo a los modelos que se presentan en el cuadro comparativo 5.1 para el cáncer pediátrico en el período 1981-1991, podemos observar que el modelo que predomina es el ARIMA(3,1,0). Este modelo se encuentra tanto al aplicarle la transformación raíz cuadrada como cuando no. Por otro lado, cuando se desea averiguar qué es lo que pasa cuando se hace una diferencia ordinaria y una estacional, se observa que está nuevamente incluido el modelo ARIMA(3,1,0), claro, ahora también se encuentra afectado por la diferencia estacional que se le ha incluido y el cual está representado por el modelo $SMA(1)_{12}$. Por todo esto, se puede concluir que la serie de tiempo que representa el cáncer pediátrico para el período de 1981-1991 en el D.F. se ajusta a un modelo ARIMA(3,1,0). Además, esta afirmación se apoya si vemos en la tabla comparativa que el modelo con la transformación raíz cuadrada tiene varianza y desviación estándar mínima. Además se tiene una sola iteración, lo cual indica que los valores iniciales de los parámetros convergen rápidamente.

Es de importancia señalar que la serie sí tiene tendencia creciente, además, al aplicarle una diferencia estacional se observa que no tiene variación estacional. Esto indica que el cáncer pediátrico no va ligado estrictamente con las estaciones del año, pero que representa una grave amenaza para los infantes, pues los resultados nos indican que a pesar de que la población de interés va en decremento para el período de estudio, el número de casos refleja un claro crecimiento. Aunque no se debe olvidar que el incremento de los registros con cáncer pediátrico puede tener cierta influencia por las mejoras en el servicio de salud, el cual ha ido creciendo de acuerdo al transcurso de los años.

Otro detalle que se desea comentar en cuanto a la forma de la gráfica, es como la serie de tiempo tiene un, digamos, ritmo de crecimiento, el cual decae a partir del dato 60, aproximadamente, el correspondiente al año de 1985, para después volver a presentar el mismo comportamiento, es decir, una tendencia evidentemente creciente. La causa probablemente la encontremos en el terremoto de 1985, donde se tuvo pérdida de información. Esto resulta desagradable, puesto que demuestra como el comportamiento del proceso del cáncer pediátrico tiene un patrón indiscutible a la tendencia creciente, aún tomando en cuenta lo que ya se ha mencionado sobre las mejoras en servicio de salud que se han venido dando.

Después de todo el análisis realizado con la serie de tiempo general y encontrar varios modelos de acuerdo a la transformación raíz cuadrada y aún en su ausencia. Cada una de éstas con una diferencia ordinaria y una ordinaria junto a una estacional, se tienen hechos importantes que es conveniente recalcar:

1. Hay tendencia creciente.
2. No existe variación estacional.
3. El modelo que mejor se ajusta es el ARIMA(3,1,0) con la transformación raíz cuadrada.
4. La varianza es creciente, de acuerdo al ajuste del modelo.

Por todo lo anterior, se concluye que la problemática del cáncer infantil no es estacional. Tampoco hay evidencia de que el proceso sea repetitivo en el transcurso de los años. Esto se observa tanto en la gráfica de la serie general como en el modelo ajustado. Además, los casos han ido en un aumento considerable conforme al paso del tiempo.

II) SEXO

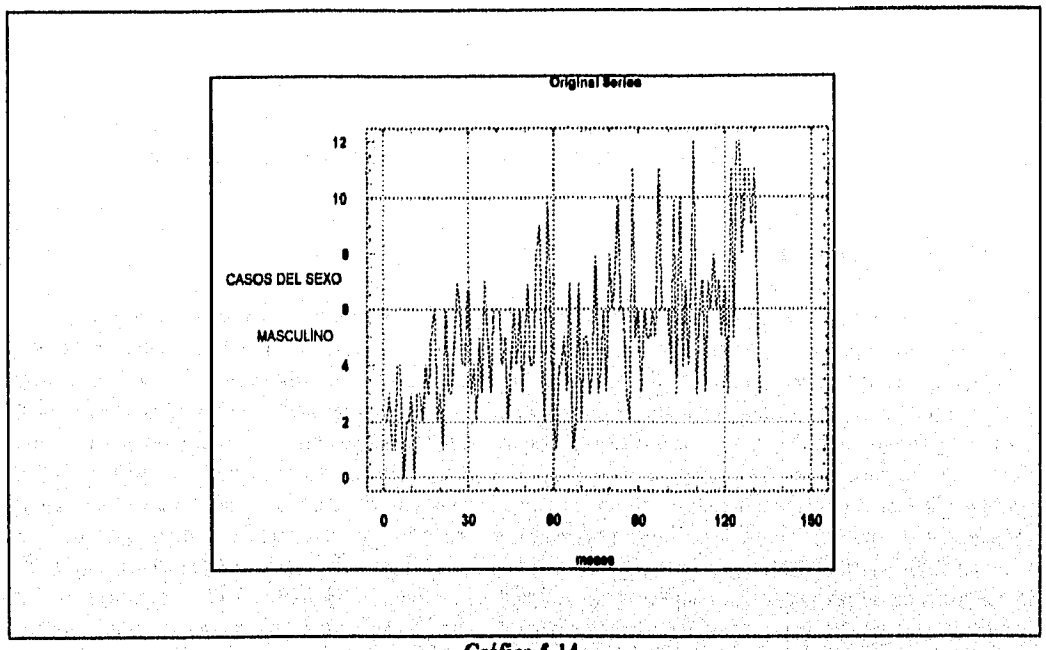
Al ser divididos los datos por sexo masculino y femenino se encuentran dos series, que por la forma de la gráfica, resultan ambas interesantes, además de que son adecuados para realizar su análisis.

SEXO MASCULINO

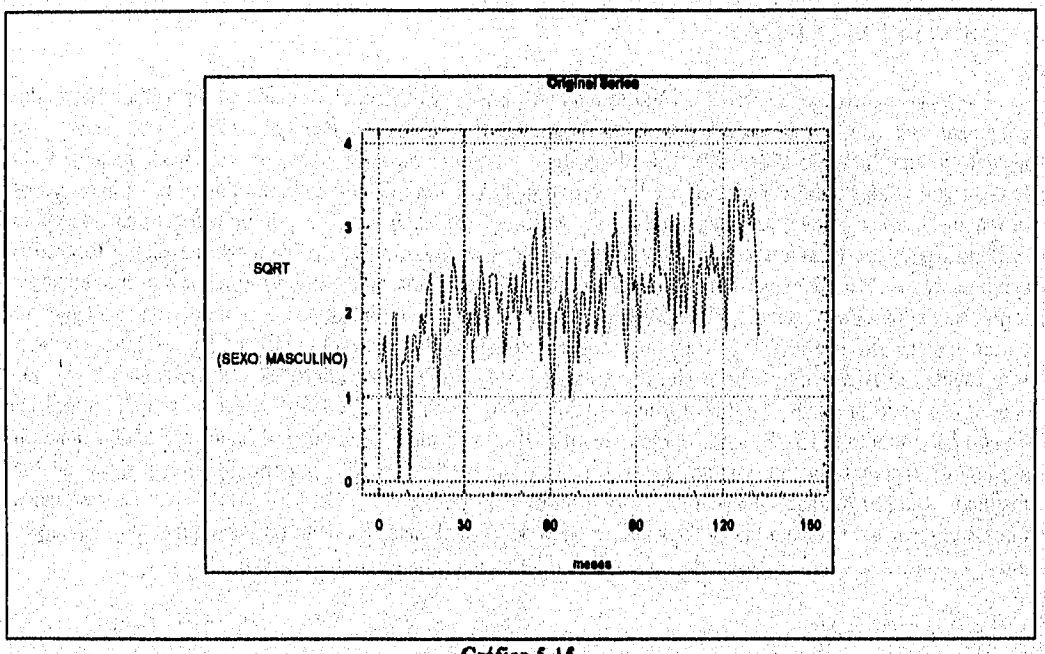
Los datos para el sexo masculino se presentan en la gráfica 5.14, en donde observamos las siguientes características: varianza creciente, tendencia creciente, no se observa fluctuación estacional. Debido a que la varianza resulta creciente es necesario aplicarle alguna transformación, siendo la adecuada para este tipo de datos la transformación raíz cuadrada. La serie que resulta la observamos en la gráfica 5.15. Esta serie de datos sigue presentado una tendencia creciente, luego entonces, es preciso aplicarle una diferencia ordinaria como lo vemos en la gráfica 5.16, donde además se encuentran la ACF y PACF correspondientes. Como podemos observar, en la gráfica del número de casos cuyo sexo es masculino, con la transformación raíz cuadrada y la diferencia ordinaria, los datos ya son estacionarios. Por su parte, la ACF tiene un valor significativo en 1 y la PACF sobresale en 1,2 y 3. Con el modelo ARIMA(0,1,1) la ACF, PACF así como el periodograma integral de los residuales los vemos en la gráfica 5.17. Las gráficas de las funciones de autocorrelación y de autocorrelación parcial de los residuales indican un comportamiento típico del ruido blanco. Por su parte, el periodograma integral de los residuales presenta un buen ajuste. Con un modelo ARIMA(3,1,0) tenemos la ACF y PACF, así como el periodograma integral de los residuales en la gráfica 5.18. De igual forma al modelo anteriormente propuesto, se observa que la ACF, PACF y el periodograma integral de los residuales muestran un ajuste adecuado.

SEXO FEMENINO

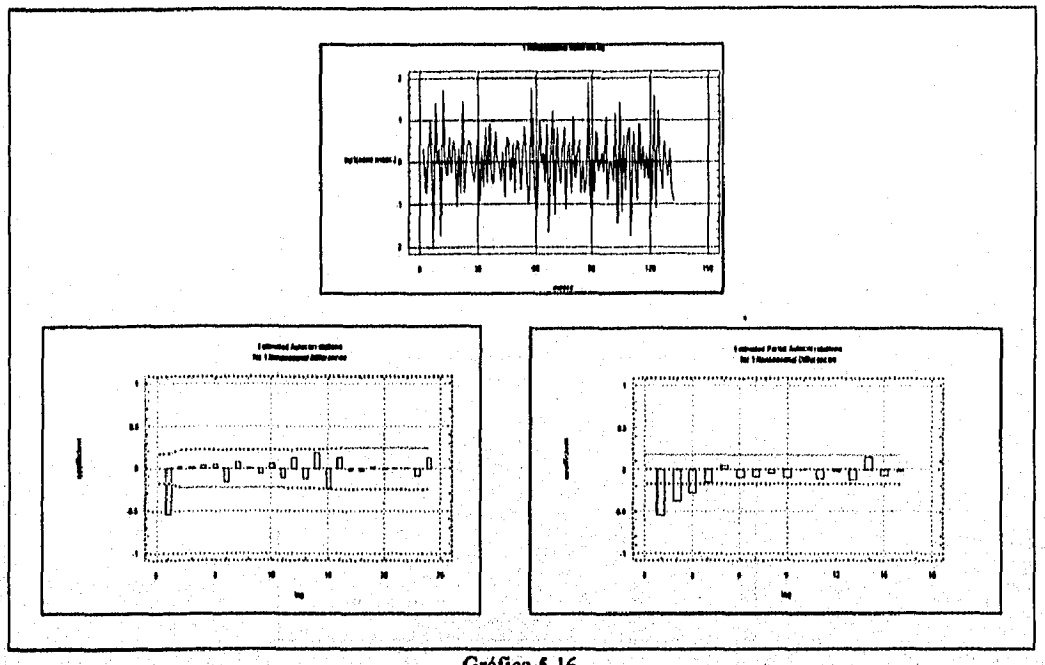
Ahora corresponde llevar a cabo el análisis para el sexo femenino. Como se ha venido haciendo, comenzaremos con la gráfica de los datos que contienen a los casos de niñas que padecieron de cáncer y se les detectó en el período 1981-1991. En esta gráfica se observa varianza creciente, tendencia creciente y no hay variación estacional, ver gráfica 5.19. Con la transformación raíz cuadrada se elimina la varianza y nos queda la gráfica con la varianza estable, ver gráfica 5.20. Ahora queda por aplicarle una diferencia ordinaria pues así lo amerita la condición de los datos en estudio, ya que como se ha dicho, la gráfica original de los casos del sexo femenino tienen una tendencia claramente creciente. Con esta diferencia se logra hacer los datos estacionarios, porque a simple vista ya no se observa ni tendencia creciente, ni varianza creciente, ni fluctuación estacional. Esto lo podemos observar en la gráfica 5.21, en donde también se presentan la ACF, con un valor significativo en 1, y la PACF, con valores significativos en 1,2 y 3. De acuerdo a esta información, se toma un modelo ARIMA(0,1,1). La ACF y PACF de los residuales muestran un comportamiento típico del ruido blanco, mientras que el periodograma integral de los residuales tiene un ajuste más o menos bueno, veamos la gráfica 5.22. La ACF, PACF y el periodograma integral de los residuales para el modelo ARIMA(3,1,0) los vemos en la gráfica 5.23. La ACF y PACF de los residuales muestran ser ruido blanco, el periodograma integral de los residuales presenta un buen ajuste. Sin embargo, ahora se estudia los datos incluyendo aparte de la diferencia ordinaria una estacional.



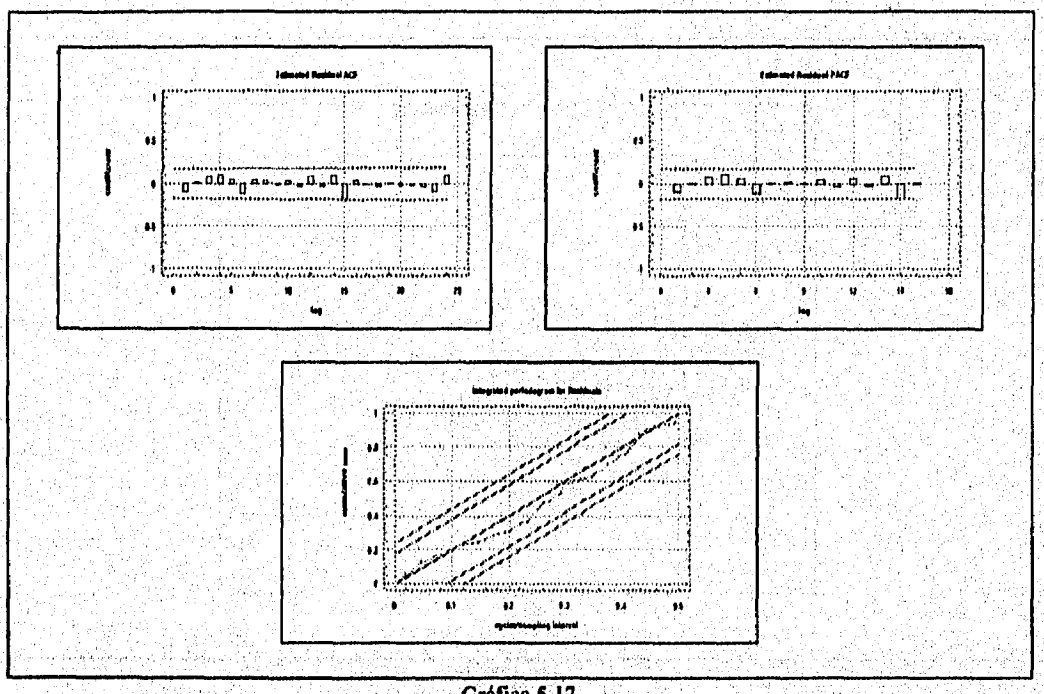
Gráfica 5.14
NUMERO DE CASOS DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYO SEXO ES MASCULINO.



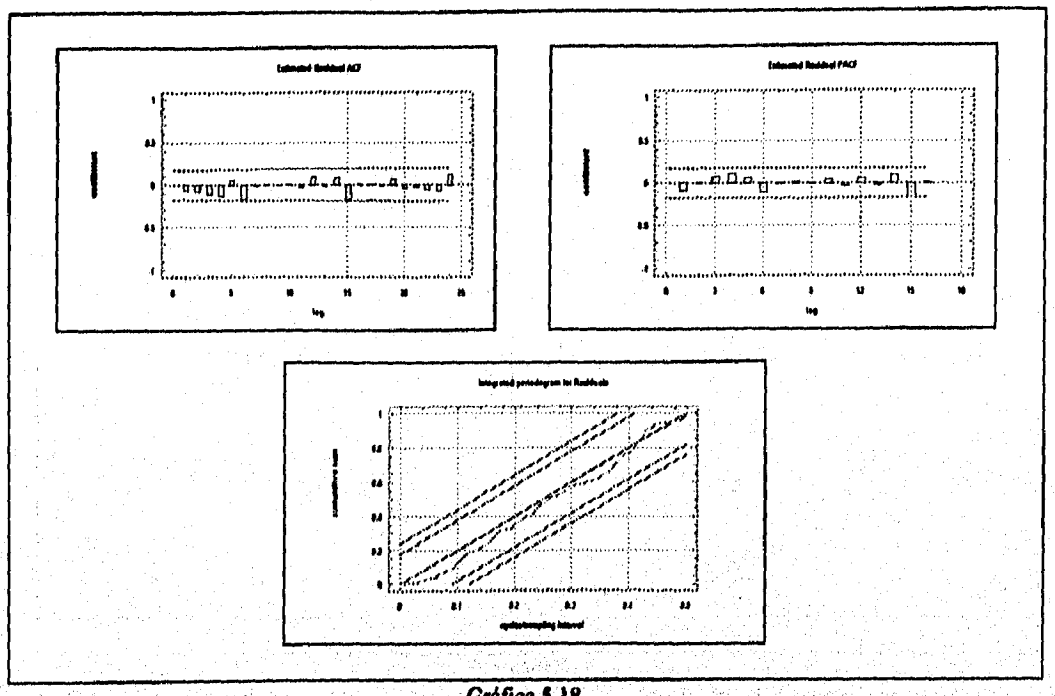
Gráfica 5.15
CASOS MENSUALES CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYO SEXO ES MASCULINO CON LA TRANSFORMACION RAIZ CUADRADA.



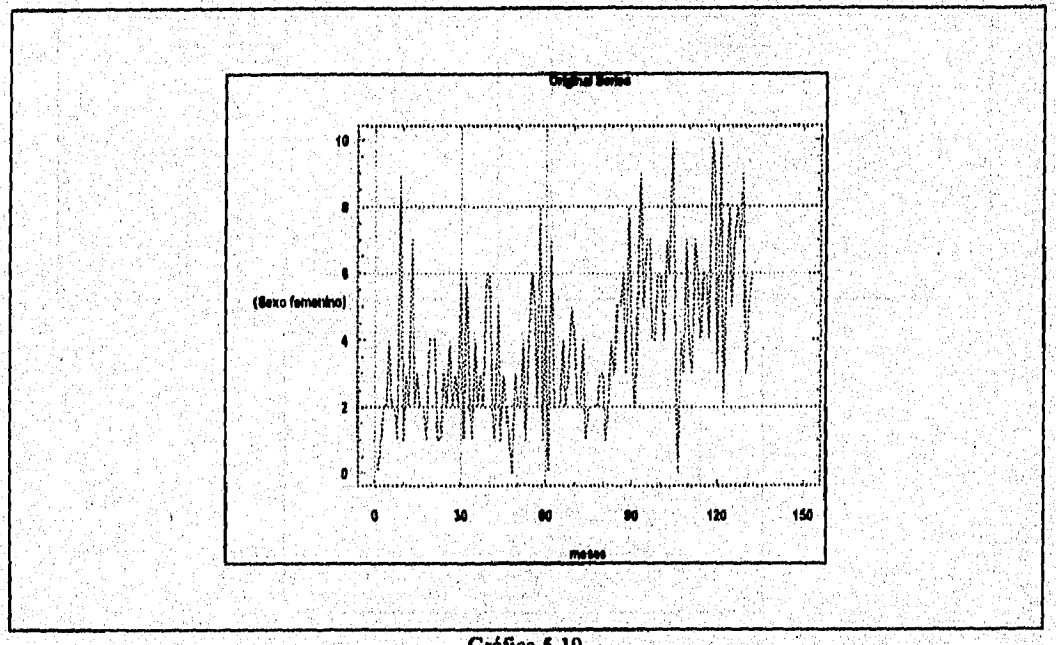
Gráfica 5.16
RAIZ CUADRADA DE LA SERIE DEL SEXO MASCULINO CON UNA DIFERENCIA ORDINARIA, ACF Y PACF.



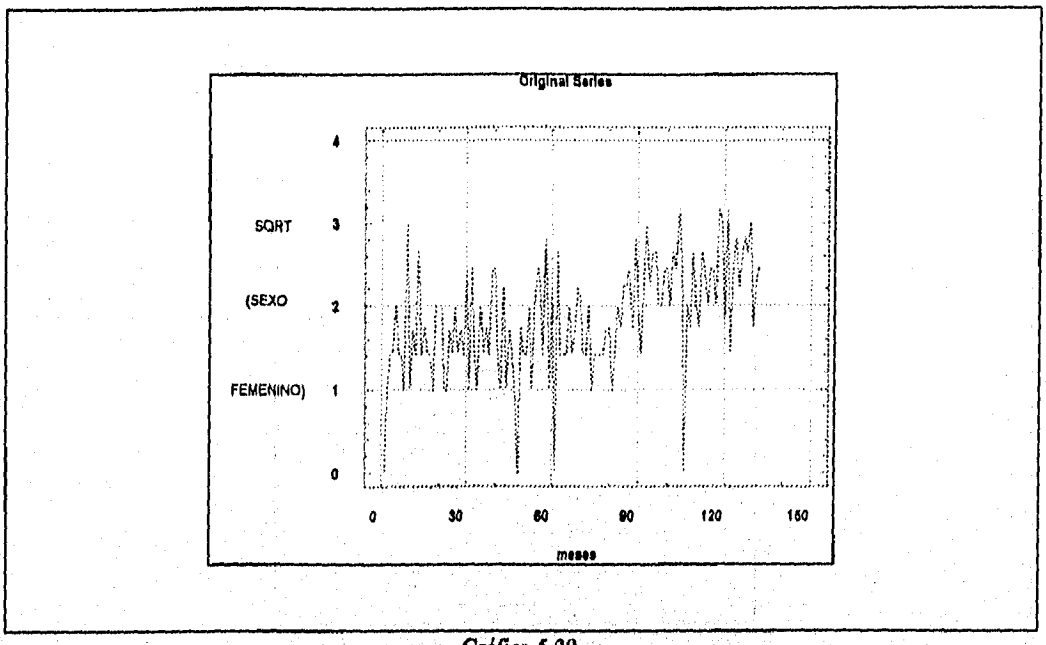
Gráfica 5.17
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(SEXO MASCULINO, MODELO ARIMA(0,1,1).



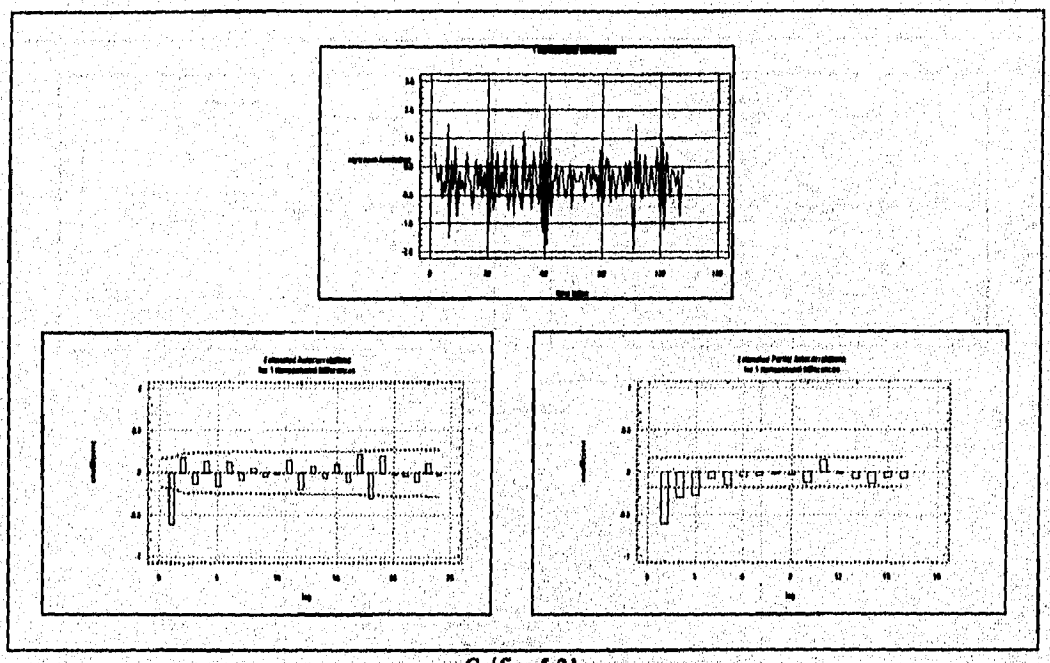
Gráfica 5.18
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(SEXO MASCULINO, MODELO ARIMA(3,1,0)).



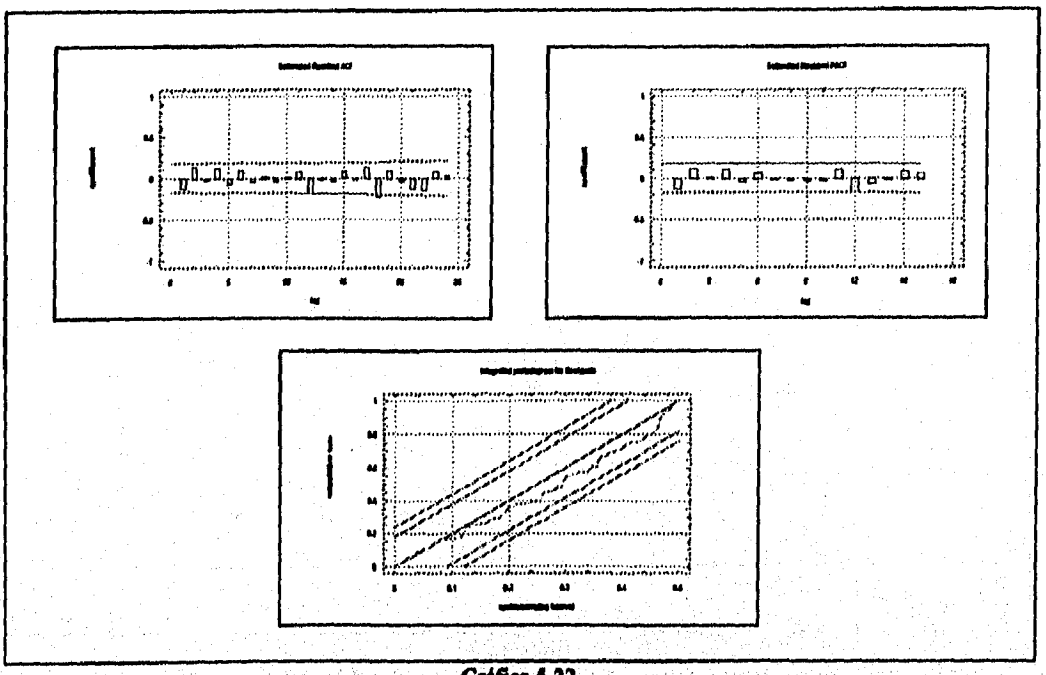
Gráfica 5.19
NUMERO DE CASOS MENSUALES CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYO SEXO ES FEMENINO.



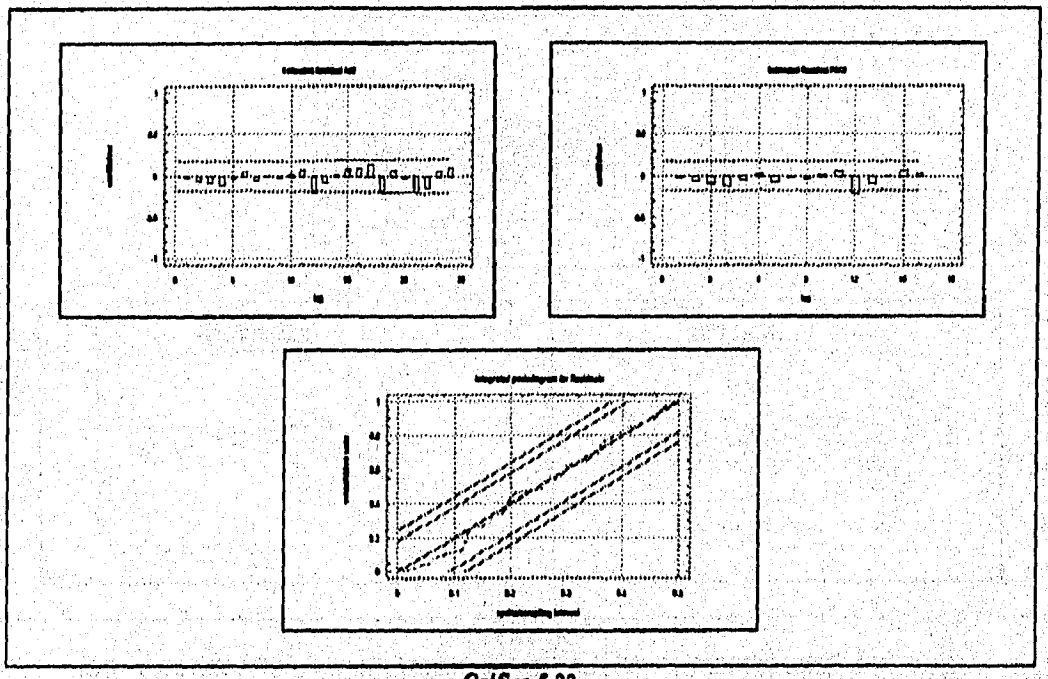
Gráfica 5.20
NUMERO DE CASOS MENSUALES CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYO SEXO ES FEMENINO CON LA TRANSFORMACION RAIZ CUADRADA.



Gráfica 5.21
SERIE DEL SEXO FEMENINO CON LA TRANSFORMACION RAIZ CUADRADA Y UNA DIFERENCIA ORDINARIA, ACF Y PACF.



Gráfica 5.22
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES (SEXO FEMENINO, MODELO ARIMA(0,1,1)).



Gráfica 5.23
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES (SEXO FEMENINO, MODELO ARIMA(3,1,0)).

Después de haber encontrado varias alternativas en cuanto a los modelos adecuados a los datos correspondientes al sexo masculino y femenino se resume lo anterior por medio de la tabla comparativa 5.2.

SEXO				
	MASCULINO		**FEMENINO**	
Modelo	ARIMA(0,1,1)	ARIMA(3,1,0)	ARIMA(0,1,1)	ARIMA(3,1,0)
Transformación	SQRT (Raíz cuadrada)	SQRT (Raíz cuadrada)	SQRT (Raíz cuadrada)	SQRT (Raíz cuadrada)
# de diferencias ordinarias	1	1	1	1
# de diferencias estacionales	--	--	--	--
# de parámetros	1	3	1	3
ϕ(1) (est. t)	--	-.89721 -10.44706	--	-.85996 -10.07107
ϕ(2) (est. t)	--	-.64655 -6.23274	--	-.50533 -4.78589
ϕ(3) (est. t)	--	-.31477 -3.65408	--	-.27442 -3.18363
θ(1) (est. t)	.95286 34.74522	--	.79198 14.68017	--
Φ(1) (est. t)	--	--	--	--
Θ(1) (est. t)	--	--	--	--
Varianza	0.30132	0.334331	.404202	0.405312
Desviación estándar	0.548926	0.578213	.635769	0.636641
Iteraciones	4	3	2	1

Cuadro comparativo 5.2

De acuerdo a los modelos que presentan un mejor ajuste se tiene que, básicamente puede ser un ARIMA(0,1,1) con la transformación raíz cuadrada y una diferencia ordinaria o el modelo ARIMA(3,1,0) con la misma transformación y la misma diferencia, pero podemos observar una varianza y desviación estándar del ruido blanco un poco menor para el modelo ARIMA(0,1,1). Por tal motivo, además de que es el modelo más sencillo, se toma como el óptimo para representar la serie de casos cuyo sexo es masculino.

Otros hechos interesantes, los cuales ya se han mencionado al comenzar el análisis, es que de igual forma que la serie general, se sigue presentando una tendencia creciente, así como una varianza creciente, la cual parece ser proporcional al crecimiento de la serie, lo que significa que el aumento de casos detectados con cáncer es mayor con respecto al mes anterior en porcentaje, conforme el tiempo transcurre. Esto puede tener relación con el hecho, de como ya se ha mencionado, el servicio de salud ha mejorado a través de los años, pero la realidad en concreto sólo se podría ver con mayor exactitud realizando otro análisis exhaustivo sobre el caso.

En cuanto a los casos detectados con cáncer del sexo femenino tenemos una tendencia creciente, la varianza parece ser un poco creciente, y los datos no presentan estacionalidad. De igual forma que la gráfica correspondiente al sexo masculino. No hay evidencia de fluctuación estacional. Esto se observa primero en la gráfica original y después en la gráfica con la transformación raíz cuadrada y la diferencia ordinaria. El modelo que se puede tomar es el ARIMA(0,1,1) con la transformación raíz cuadrada, aunque también puede resultar un modelo bastante bueno el modelo ARIMA(3,1,0) con la transformación raíz cuadrada, ya que el ajuste para ambos modelos es bastante aceptable.

En los modelos encontrados se puede observar que, básicamente, tanto para el sexo masculino como para el femenino siguen una misma trayectoria. Es decir, al aplicarles una diferencia ordinaria, los modelos que es posible ajustar son por un lado, un ARIMA(0,1,1) o un ARIMA(3,1,0), y cuando se les aplica, además de la diferencia ordinaria la estacional, tan sólo se le añade a los modelos la parte estacional con un $SMA(1)_{12}$. Con lo anterior se puede concluir que, como era de esperarse, las series de tiempo para el sexo masculino y para el femenino no tienen estacionalidad al igual que la serie general, pero sí existe tendencia para ambas series. Luego entonces, estas series tienen modelos semejantes y características semejantes, por lo que se deduce que deben tener comportamientos muy similares.

III) TIPO DE CANCER

En cuanto al tipo de cáncer se tomaron doce principales grupos, los mismos que se explican en el capítulo de Antecedentes. Cada uno de ellos contiene a diversos cánceres específicos. Para cada uno de los doce cánceres mencionados se obtienen sus respectivas frecuencias, dando como resultado un peso considerable en el cáncer tipo I, la leucemia, con un 38% del total de casos, por lo que se decide hacer la división en dos partes, por una lado será la leucemia y por otro se considera a los once cánceres restantes. Lo que se pretende es ver si el proceso del cáncer tipo I (leucemia) es parecido al de los demás cánceres, o existen diferencias significativas.

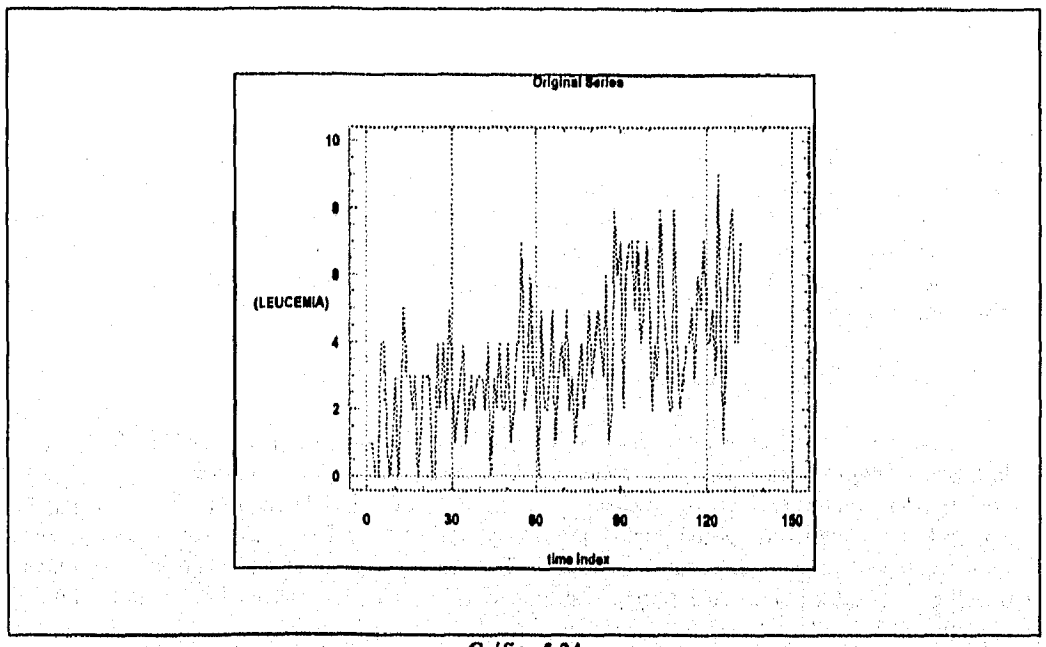
CANCER TIPO I (LEUCEMIA)

Los datos correspondientes al cáncer tipo I, la leucemia, están representados en la gráfica 5.24. En ella podemos observar varianza un poco creciente, tendencia creciente y no se identifica variación estacional a simple vista. Al aplicar la transformación raíz cuadrada para eliminar la varianza, la gráfica que nos queda es la 5.25. Ahora corresponde eliminar la tendencia, que como se dijo, se visualiza creciente. Una vez aplicada la transformación raíz cuadrada (estabilización de la varianza creciente) y la diferencia ordinaria para eliminar la tendencia. La ACF tiene un valor significativo en 1 y la PACF en 1,2,3,4, 6, e incluso 14, ver gráfica 5.26.

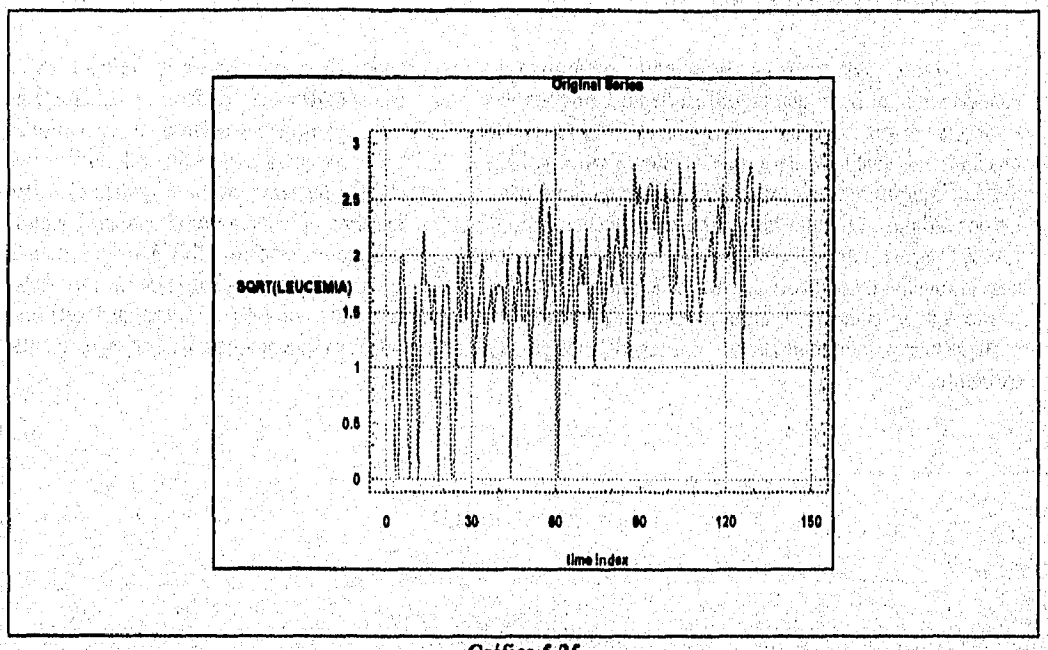
Después de haber probado varios modelos, comenzando desde luego del más sencillo, se llega al modelo ARIMA(4,1,0). En éste la ACF, PACF y el periodograma integral de los residuales muestran ser ruido blanco, por lo que parece ser un buen modelo para representar los datos, ver gráfica 5.27.

CANCERES TIPO II-XII

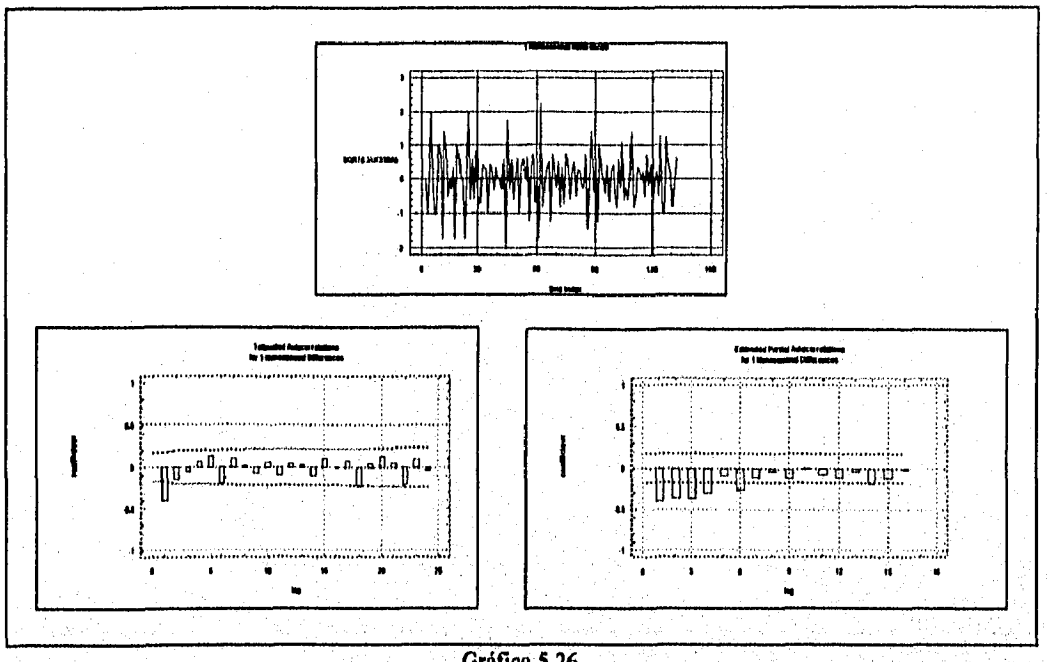
La serie de casos que presentaron algún tipo de cáncer del II-XII se muestra en la gráfica 5.28, la cual presenta, al igual que la gráfica de los casos con leucemia, varianza creciente, tendencia creciente, y no se observa variación estacional aparente. Con la transformación raíz cuadrada para eliminar la varianza creciente nos queda la serie que se ilustra en la gráfica 5.29. Para eliminar la tendencia se le aplica una diferencia ordinaria, la gráfica que nos queda junto con su ACF y PACF las tenemos en la gráfica 5.30. En las dos últimas gráficas a las que hacemos mención, en la ACF se observa un valor sobresaliente en 1 y en la PACF se tienen valores sobresalientes en 1 y 2. En primer lugar se elige un modelo ARIMA(0,1,1), para el cual la función de autocorrelación, la de autocorrelación parcial y el periodograma integral de los residuales son los que se encuentran en la gráfica 5.31, en donde nos podemos dar cuenta que el ajuste al modelo es bastante bueno, y en realidad no es necesario indagar sobre otro modelo ya que el ajuste de éste es realmente excelente.



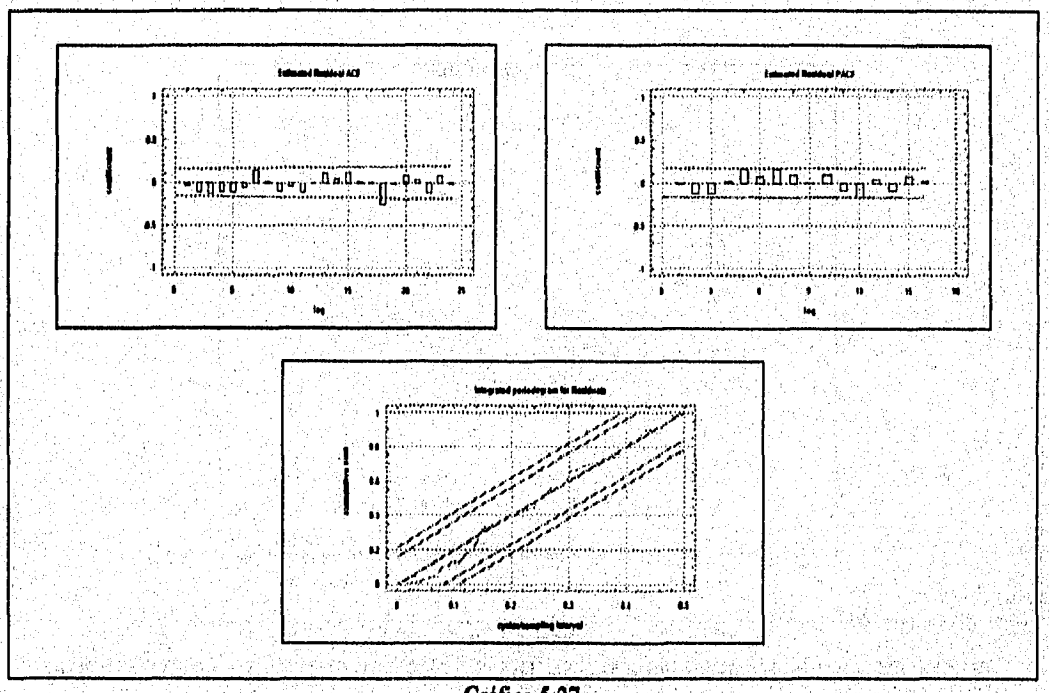
Gráfica 5.24
NUMERO DE CASOS MENSUALES DETECTADOS CON LEUCEMIA EN EL D.F. PARA EL PERIODO 1981-1991.



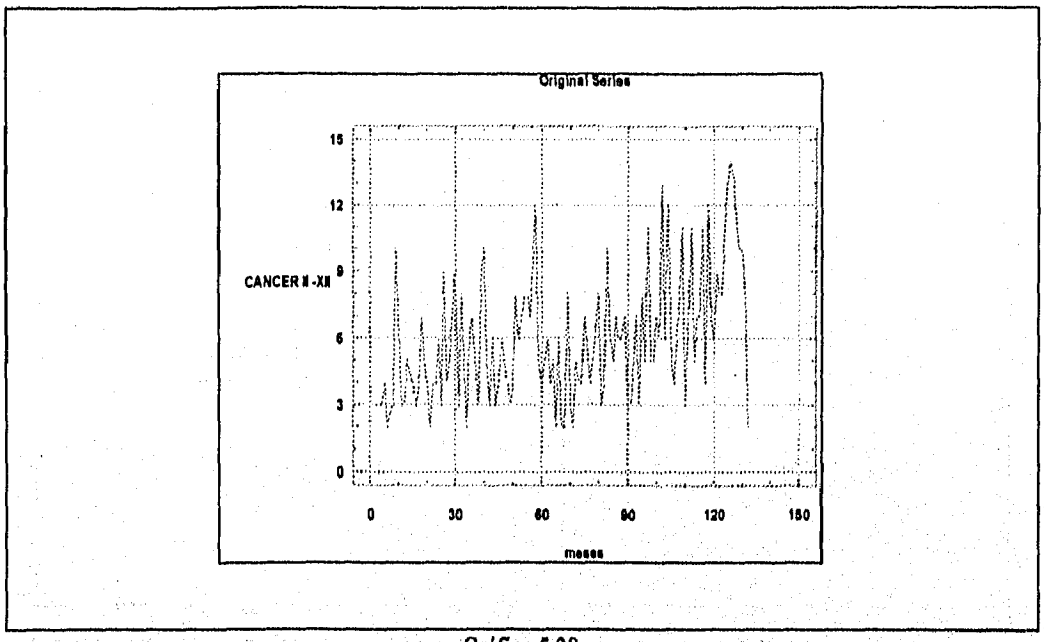
Gráfica 5.25
RAIZ CUADRADA DEL NUMERO DE CASOS MENSUALES DETECTADOS CON LEUCEMIA EN EL D.F. PARA EL PERIODO 1981-1991.



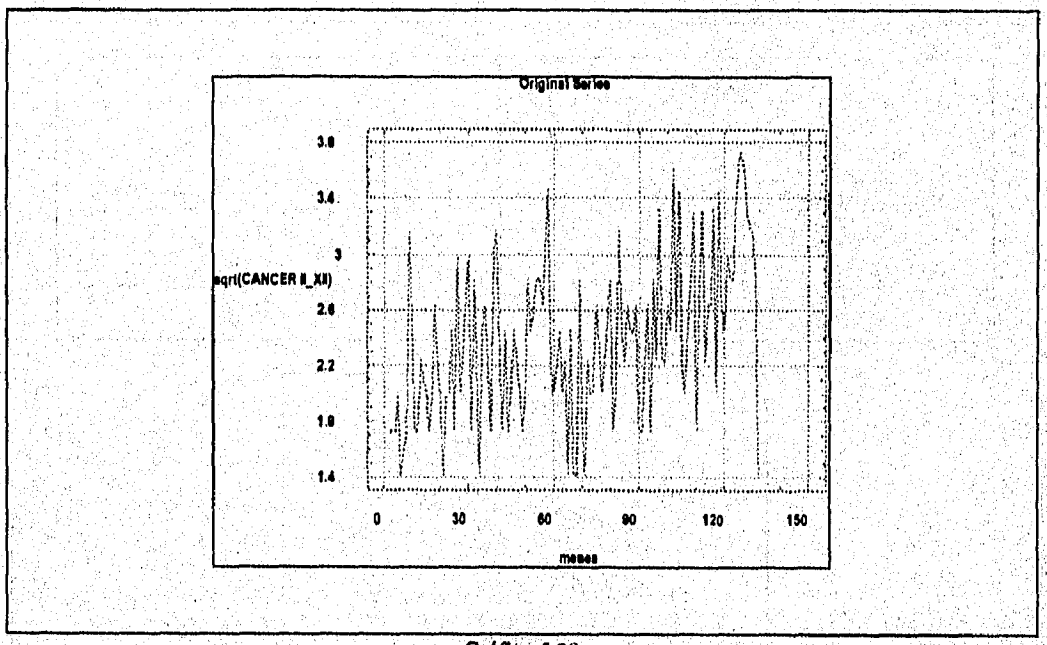
Gráfica 5.26
SERIE DE CASOS CON LEUCEMIA CON LA TRANSFORMACION RAIZ CUADRADA Y UNA DIFERENCIA ORDINARIA.



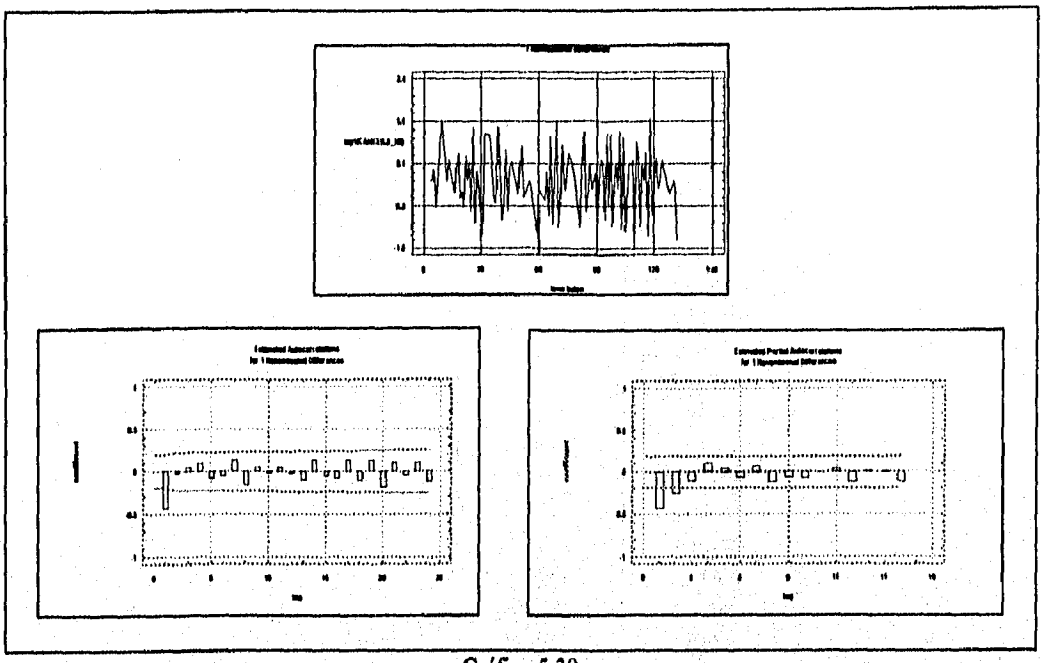
Gráfica 5.27
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(TIPO DE CANCER: LEUCEMIA, MODELO ARIMA(4,1,0)).



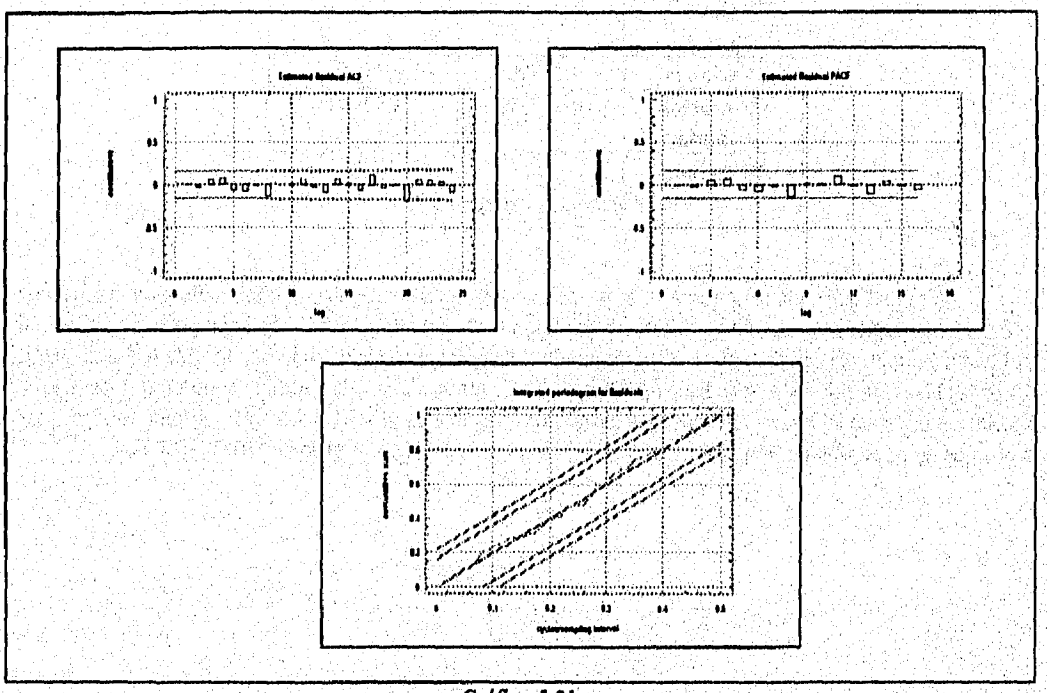
Gráfica 5.28
NUMERO DE CASOS MENSUALES DIAGNOSTICADOS CON ALGUN TIPO DE CANCER DEL II-XII, EN EL D.F., PARA EL PERIODO 1981-1991.



Gráfica 5.29
SERIE DE CASOS MENSUALES CON ALGUN TIPO DE CANCER CON LA TRANSFORMACION RAIZ CUADRADA.



Gráfica 5.30
NUMERO DE CASOS MENSUALES DETECTADOS CON ALGUN TIPO DE CANCER DEL II-XII, CON UNA DIFERENCIA ORDINARIA, ACF Y PACF.



Gráfica 5.31
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(TIPO DE CANCER II-XII, MODELO ARIMA(0,1,1)).

TIPO DE CANCER		
	CANCER TIPO I	**CANCER TIPO II-XII**
Modelo	ARIMA(4,1,0)	ARIMA(0,1,1)
Transformación	SQRT (Raíz cuadrada)	SQRT (Raíz cuadrada)
# de diferencias ordinarias	1	1
# de diferencias estacionales	--	--
# de parámetros	4	1
ϕ(1) (est. t)	-.79001 -9.30104	--
ϕ(2) (est. t)	-.73531 -7.51309	--
ϕ(3) (est. t)	-.57613 -5.85528	--
ϕ(4) (est. t)	-.30225 -3.53713	--
θ(1) (est. t)	--	0.76217 9.28958
Φ(1) (est. t)	--	--
Θ(1) (est. t)	--	--
Varianza	0.368906	0.352641
Desviación estándar	0.607376	0.593836
Iteraciones	1	3

Cuadro comparativo 5.3

Al igual que en la serie general, se puede apreciar que las series de tiempo, tanto para la leucemia como la que representa los once cánceres restantes, no contienen la parte estacional. Esto se denota al aplicarles la diferencia estacional, y como resultado lo que se obtiene son modelos a los que hay que incluir un $SMA(1)_{12}$. Por otro lado, si se hace una comparación entre los casos de leucemia contra todos los demás, se tiene que el modelo que representa a los casos de leucemia es el ARIMA(4,1,0), en tanto que para los casos que manifestaron un tipo de cáncer del II-XII tienen como mejor modelo al ARIMA(0,1,1).

Como podemos ver, de acuerdo a los casos que se les detectó leucemia, éstos siguen conservando una estructura básicamente de tendencia creciente. Es decir, se tiene al igual que en la serie general, un crecimiento de los casos detectados con leucemia conforme pasa el tiempo, la varianza revela ser un poco creciente, lo que nos indica que su crecimiento es proporcional al de la serie. Por otro lado, no hay variación estacional, puesto que al incluir la diferencia estacional a la serie, sólo se encuentra la inclusión de un $SMA(1)_{12}$ como respuesta a esto, lo que indica que sólo refleja el efecto de agregar dicha diferencia.

En cuanto a la serie de tiempo que contiene a los casos de cáncer tipo II-XII, nos damos cuenta que tienen una tendencia creciente, aunque no muy marcada, la varianza también es ligeramente creciente y el modelo que se adecúa a dichos datos en realidad es bastante sencillo, el ARIMA(0,1,1).

Ahora bien, si nosotros consideramos a las dos gráficas, las correspondientes a la serie original del cáncer tipo I (leucemia) y la correspondiente a los cánceres tipo II-XII vemos que a simple vista se tiene un mismo patrón, pues las gráficas son bastante parecidas y conservan a su vez las mismas características, como son tendencia creciente, varianza creciente, sin presencia de fluctuación estacional. Sin embargo, es importante recalcar que el modelo que mejor representa a los datos para el cáncer tipo I, la leucemia, es el ARIMA(4,1,0), mientras que para los cánceres restantes en conjunto, es el ARIMA(0,1,1), lo cual nos dice que probablemente haya diferencias significativas entre los casos de leucemia contra los demás, y tal vez valdría la pena realizar una investigación sobre los casos de leucemia en específico. Por lo tanto, queda el caso de la leucemia como una posible línea de investigación.

IV) FAMILIARES DEL PACIENTE (C/S CANCER)

De acuerdo a los casos estudiados el 83% negó tener familiares con cáncer y el 17% restante contestó afirmativamente. Conforme a lo anterior, se forman dos series de tiempo. Por el número de casos que presenta la primera serie, es de esperarse un modelo semejante al de la general.

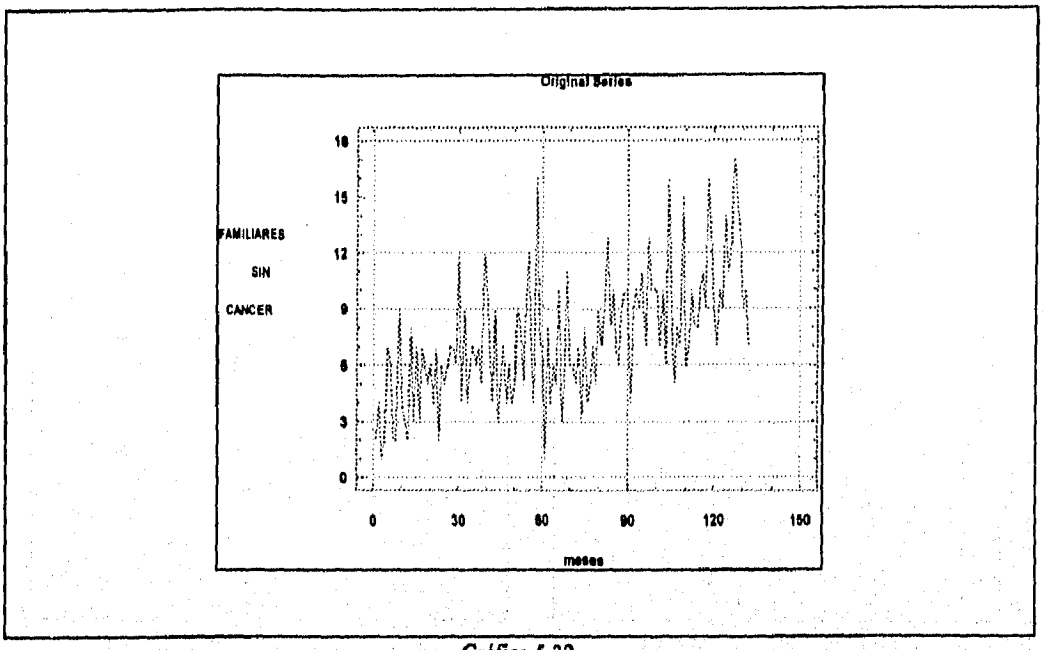
CASOS QUE MANIFESTARON NO TENER FAMILIARES CON CANCER

Los casos que dijeron no tener familiares con cáncer se muestran en la gráfica 5.32. Esta presenta varianza un poco creciente, tendencia creciente, sin variación estacional a simple vista. Como se ha venido haciendo con las gráficas anteriores, en primer lugar se estabiliza la varianza con la transformación raíz cuadrada, ver la gráfica 5.33. Para eliminar la tendencia que aún sigue presente se le aplica una diferencia ordinaria. La serie que resulta al igual que su ACF y PACF se observan en la gráfica 5.34. De acuerdo a ésta, la ACF parece truncarse en 1 y la PACF es decreciente infinita, por lo que se sugiere un modelo ARIMA(0,1,1). Dicho modelo nos revela la ACF, PACF y periodograma integral de los residuales más o menos adecuados ya que se tiene un comportamiento similar al del ruido blanco y el periodograma integral de los residuales tiene un buen ajuste, observar la gráfica 5.35. Por lo tanto, parece ser un buen modelo para los datos que se analizan.

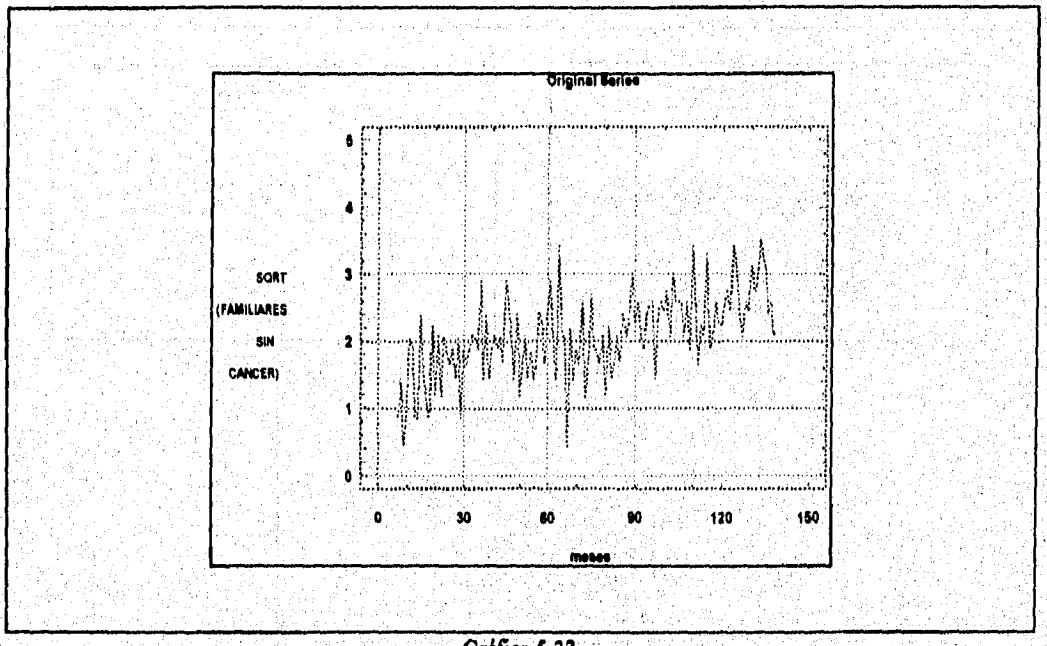
Ahora corresponde analizar los casos que dijeron tener familiares con cáncer.

CASOS QUE MANIFESTARON TENER FAMILIARES CON CANCER

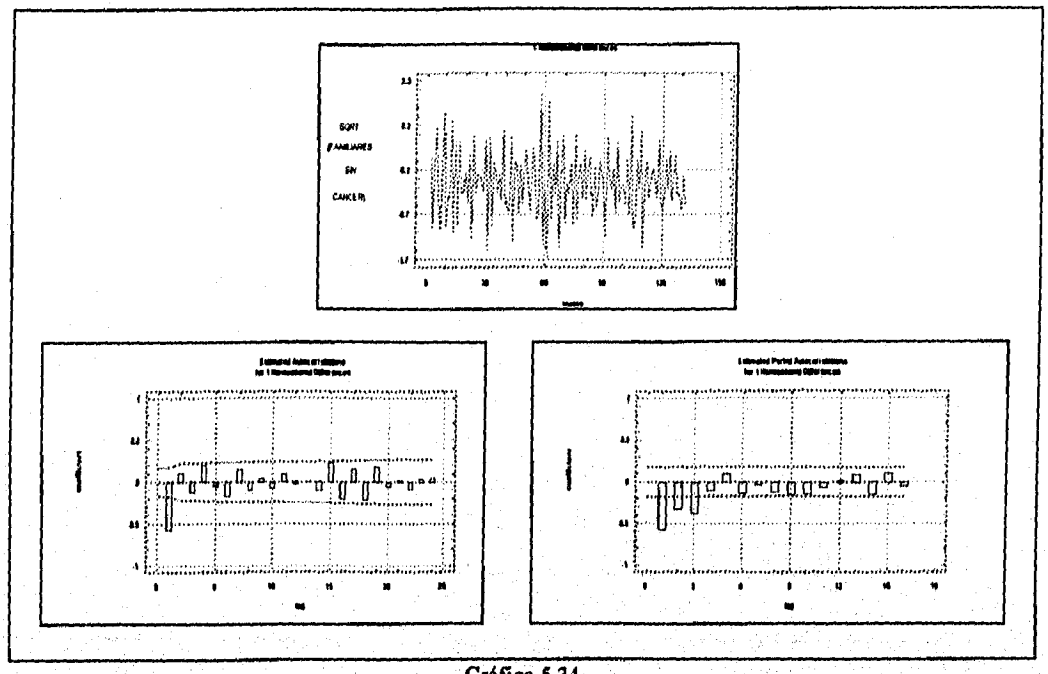
La gráfica del número de casos que dijeron tener familiares con cáncer, ver gráfica 5.36, presenta lo siguiente: en primer lugar, la varianza se observa creciente, hay tendencia creciente, no se hace evidente la presencia de fluctuación estacional. Después de aplicarle la transformación raíz cuadrada, ver gráfica 5.37, se le hace una diferencia ordinaria. La serie con esta diferencia así como su ACF y PACF son las que se encuentran en la gráfica 5.38. Como podemos observar, la ACF tiene un valor significativo en 1, 2 y 13. En cuanto a la PACF se observan valores sobresalientes en 1,2,3,4,8 y12. El modelo que se encuentra más adecuado para estos datos es el ARIMA(0,1,1). La ACF, PACF y el periodograma integral de los residuales para el modelo anteriormente propuesto se encuentran en la gráfica 5.39, donde se muestra la ACF, PACF y periodograma integral de los residuales, los cuales nos indican un buen ajuste, ya que como podemos observar, se tiene un comportamiento similar al del ruido blanco por parte de la ACF y PACF. En cuanto al periodograma integral de los residuales se tiene un ajuste aceptable.



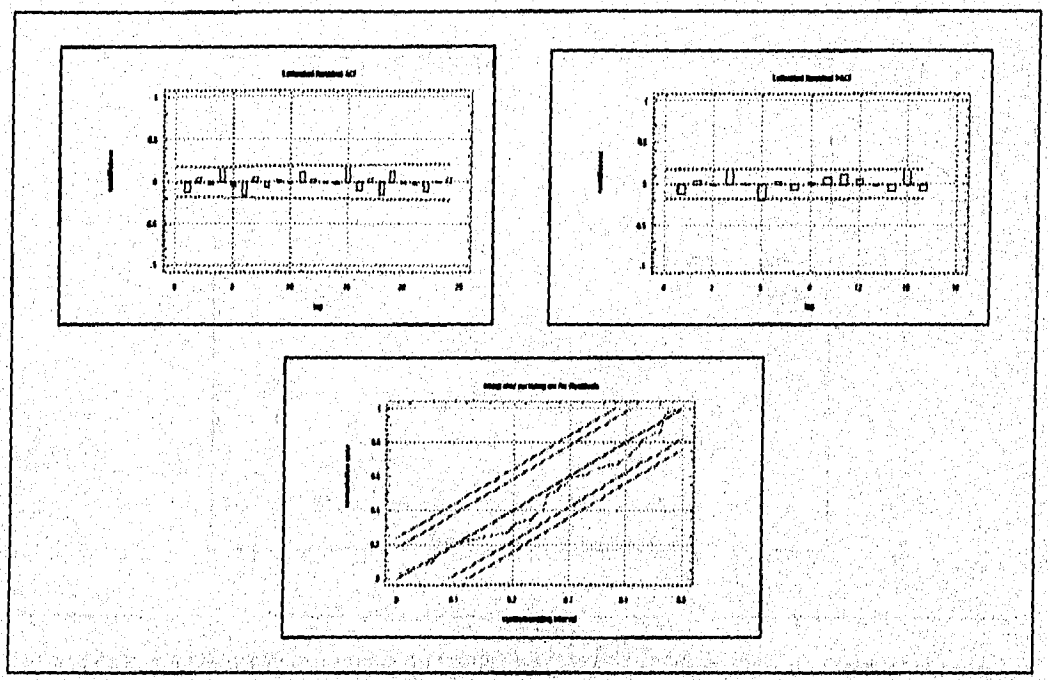
Gráfica 5.32
SERIE DE TIEMPO QUE REPRESENTA EL NUMERO DE CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO Y QUE ADEMAS NO TIENE FAMILIARES QUE HAYAN PADECIDO DE ALGUN TIPO DE CANCER.



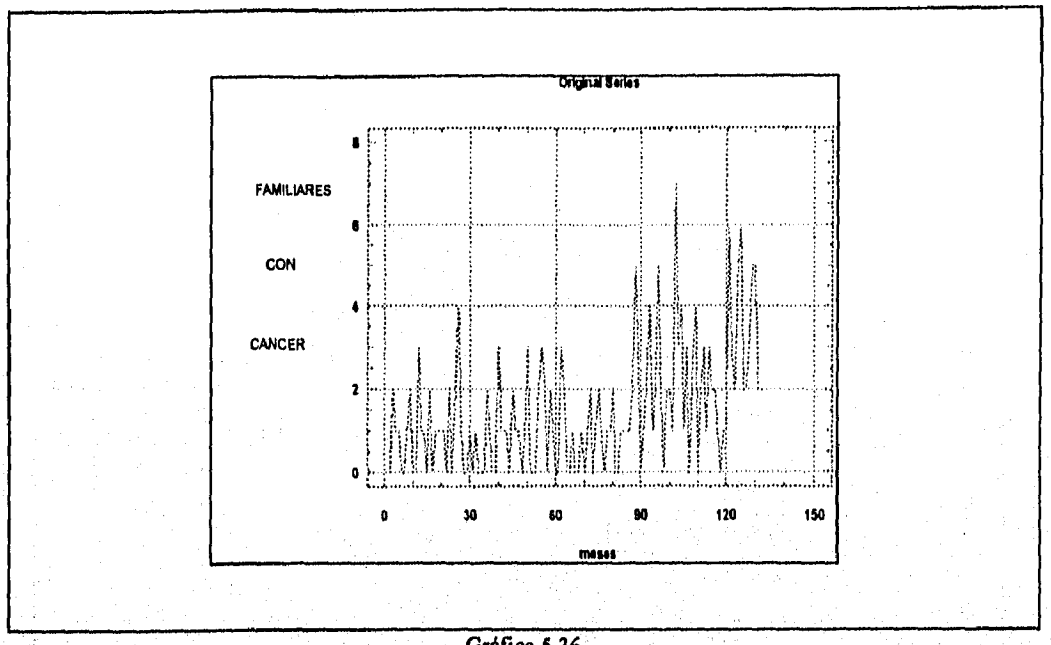
Gráfica 5.33
RAIZ CUADRADA DEL NUMERO DE CASOS MENSUALES SOBRE CANCER PEDIATRICO QUE DIJERON NO TENER FAMILIARES CON CANCER.



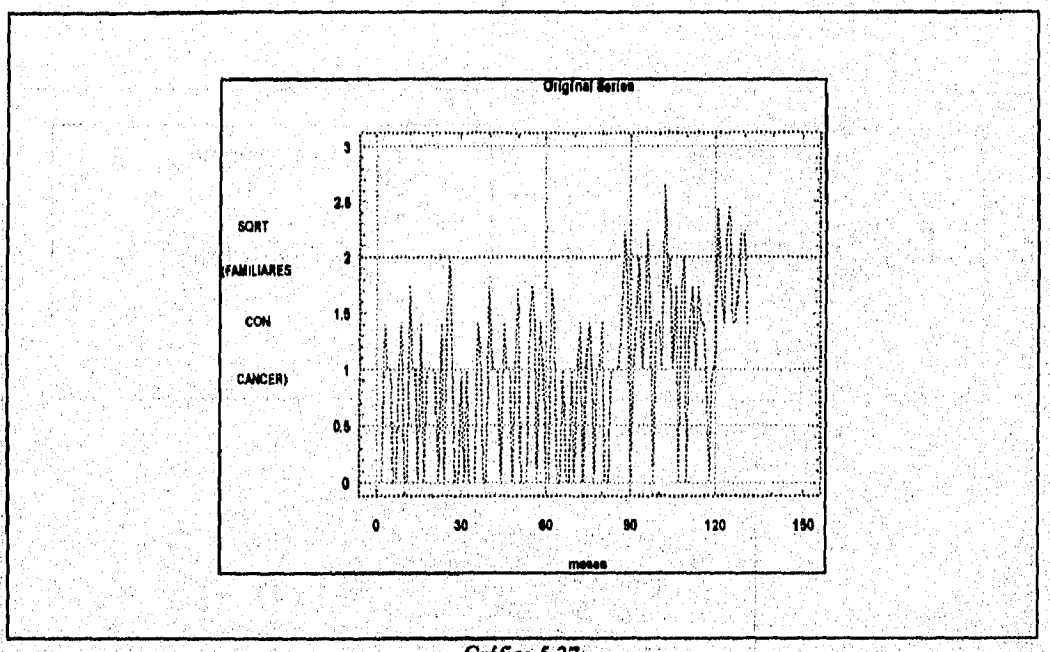
Gráfica 5.34
CASOS MENSUALES DE CANCER PEDIATRICO SIN FAMILIARES CON CANCER, CON UNA TRANSFORMACION RAIZ CUADRADA Y UNA DIFERENCIA ORDINARIA, ACF Y PACF.



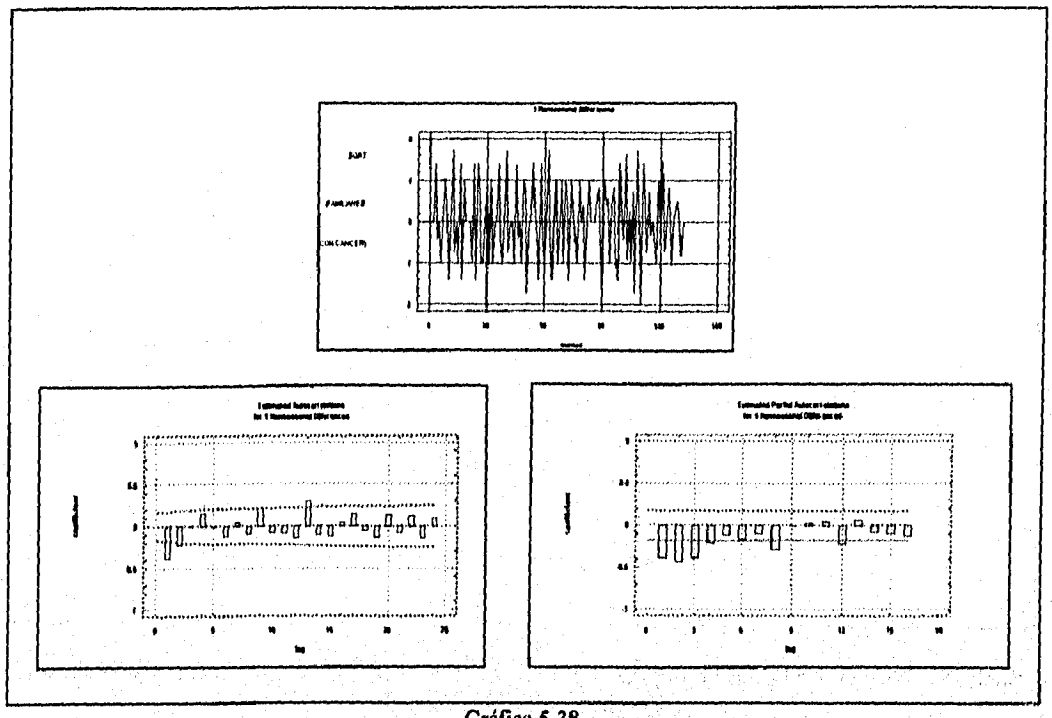
Gráfica 5.35
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(CASOS QUE REFIRIERON NO TENER FAMILIARES CON CANCER, MODELO: ARIMA(0,1,1)).



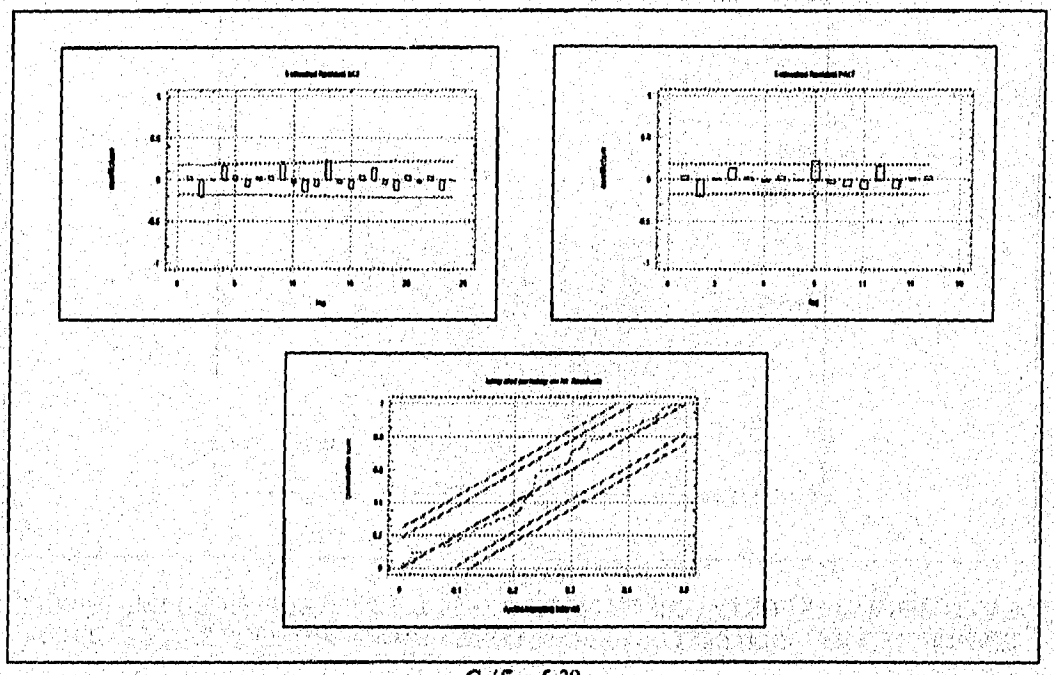
Gráfica 5.36
CASOS MENSUALES PARA EL PERIODO 1981-1991 EN EL D.F. Y QUE MANIFESTARON TENER FAMILIARES CON CANCER.



Gráfica 5.37
CASOS MENSUALES CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 QUE DIJERON TENER FAMILIARES CON CANCER, CON LA TRANSFORMACION RAIZ CUADRADA.



Gráfica 5.38
RAIZ CUADRADA DEL NUMERO DE CASOS MENSUALES CON CANCER PEDIATRICO QUE DIJERON TENER FAMILIARES CON CANCER, CON UNA DIFERENCIA ORDINARIA, ACF Y PACF.



Gráfica 5.39
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(CASOS QUE DIJERON TENER FAMILIARES CON CANCER, MODELO ARIMA(0,1,1)).

Ahora corresponde mostrar los modelos encontrados en conjunto, como lo vemos en el cuadro comparativo 5.4 que se muestra a continuación.

FAMILIARES DEL PACIENTE

	SIN CANCER	CON CANCER
	ARIMA(0,1,1)	ARIMA(0,1,1)
Transformación	SQRT (Raíz cuadrada)	SQRT (Raíz cuadrada)
# de diferencias ordinarias	1	1
# de diferencias estacionales	--	--
# de parámetros	1	1
ϕ(1) (est. t)	--	--
ϕ(2) (est. t)	--	--
ϕ(3) (est. t)	--	--
θ(1) (est. t)	0.89377 24.90527	0.93116 28.58433
Φ(1) (est. t)	--	--
Θ(1) (est. t)	--	--
Varianza	0.284763	0.473131
Desviación estándar	0.533632	0.687845
Iteraciones	3	4

Cuadro comparativo 5.4

Realmente encontrar un modelo adecuado para estas dos series, la de los casos que tuvieron familiares con cáncer y los que no, fue sencillo, si se considera que el modelo elegido al final de cuentas es el ARIMA(0,1,1) en ambos casos, pero no se tiene un ajuste exacto, ya que hay valores en la ACF y PACF de los residuales un poquito sobresalientes. Pero, si tomamos en cuenta que se intentó con varios modelos alternativos y el que presentó mejor ajuste, además de que es el más sencillo fue dicho modelo, no existe objeción para afirmar que es el adecuado. Ahora bien, si comparamos las dos series originales, es decir, la 5.45 contra la 5.49 podemos darnos cuenta que la primera tiene, digamos, una mejor estructura, la cual es parecida a la serie general, donde se observa más claramente la tendencia creciente y la varianza es ligeramente creciente. Por otro lado, la serie de los casos que manifestaron tener familiares con cáncer tiene un patrón casi estacionario hasta el dato 60, y después de éste es cuando comienza a crecer la serie. Esto ocurre precisamente después de 1985, ya que como se mencionó, hubo una pérdida de información, debido al terremoto de ese año. Ahora bien, estos datos tienen un comportamiento en cierto grado, similar al del ruido blanco, puesto que su gráfica no indica un patrón realmente consistente en cuanto tendencia creciente, por ejemplo. Esto nos lleva a que los casos con cáncer pediátrico están básicamente conformados por pacientes cuyos familiares no tuvieron cáncer, y como era de esperarse tienen un comportamiento similar a

la serie general, aunque el modelo que mejor se ajusta es el ARIMA(0,1,1), y en la serie general se tienen como modelo adecuado al ARIMA(3,1,0).

V) EDAD DEL PACIENTE AL DIAGNOSTICO DEL CANCER

La edad del paciente al diagnóstico de cáncer está comprendida hasta los catorce años. Luego entonces, la división que se hace es por un lado de 0 a 4 , y por otro de 5 a 14 años. La razón por la que se hace de esta forma es debido a las frecuencias obtenidas, quedando adecuadas de dicha forma para llevar a cabo su análisis.

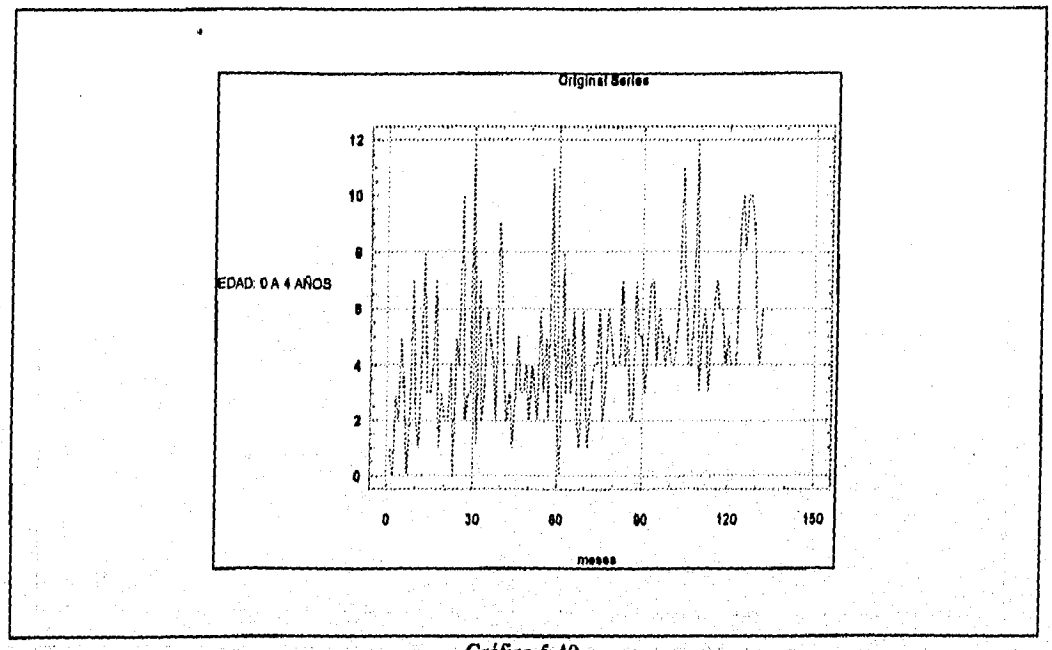
EDAD DE 0 A 4 AÑOS

La serie que contiene los datos de los pacientes de 0 a 4 años la podemos observar en la gráfica 5.40. Esta muestra varianza un poco creciente, tendencia creciente, sin indicios de variación estacional. Por consiguiente, en primer término se trata de eliminar la varianza mediante la transformación raíz cuadrada. La gráfica con la variable transformada la vemos en la gráfica 5.41, sin embargo, la tendencia aún está presente, por lo que se necesita aplicar una diferencia ordinaria. La gráfica con la transformación raíz cuadrada y la diferencia ordinaria, así como su ACF y PACF las vemos en la gráfica 5.42, donde la ACF muestra valores sobresalientes en 1y 4, la PACF a su vez, los tiene en 1,2 y 3. El primer modelo tentativo que parece adecuarse al fenómeno es el ARIMA(0,1,1). Observemos sus respectivas gráficas de la ACF, PACF y periodograma integral de los residuales en 5.43, donde se aprecia un buen ajuste.

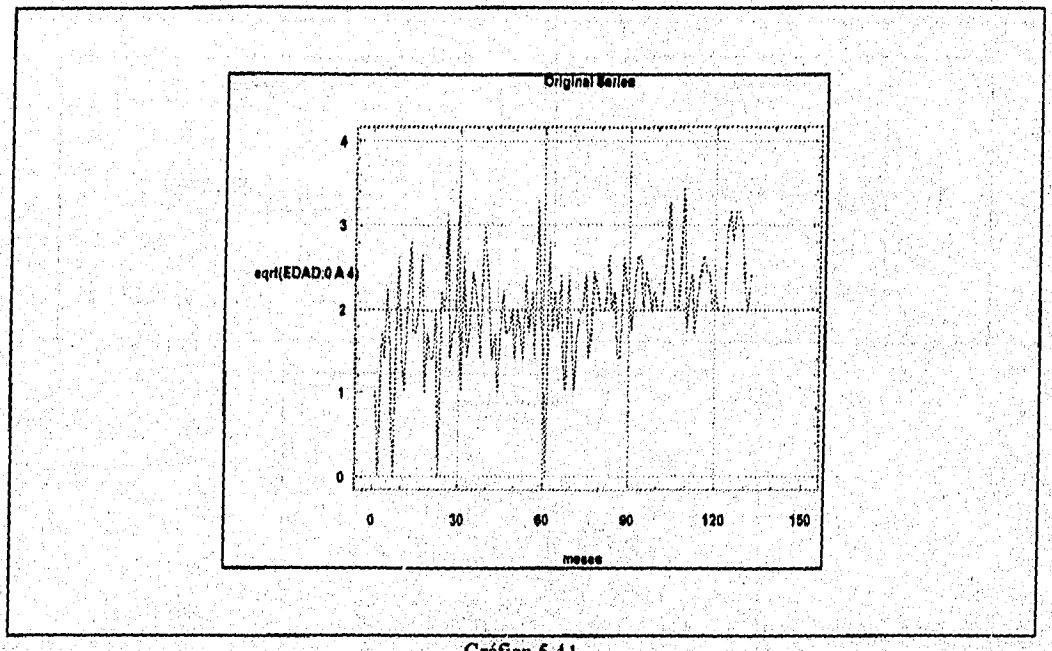
Ahora corresponde analizar lo que sucede para los pacientes que se les diagnosticó cáncer a la edad de 5 a 14 años.

EDAD DE 5 A 14 AÑOS

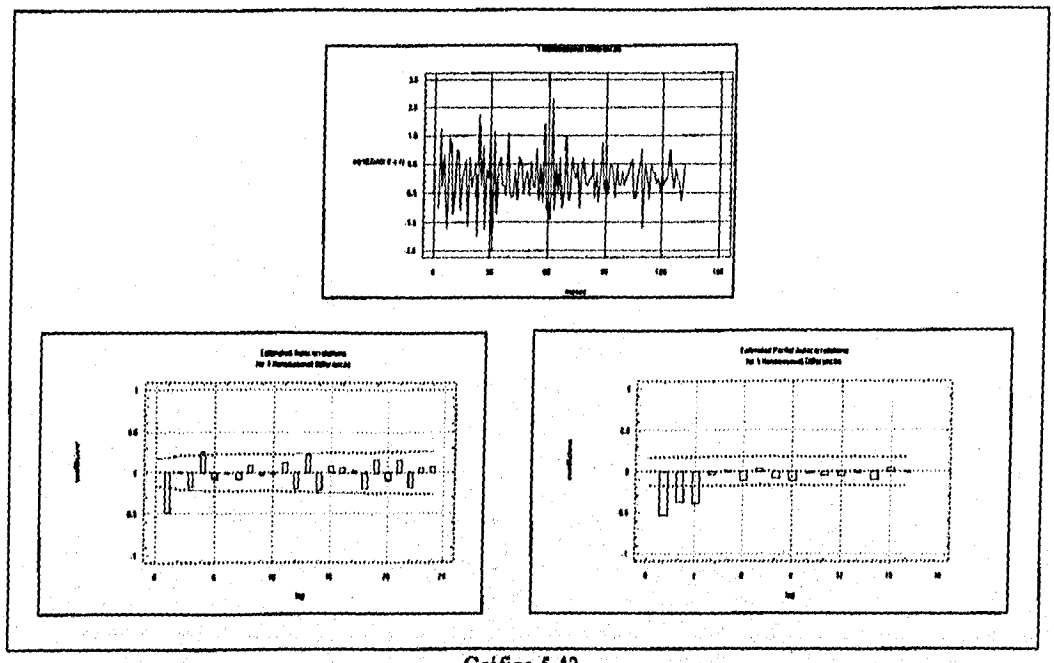
En la gráfica 5.44 observamos que corresponde a los casos de pacientes diagnosticados con cáncer pediátrico cuya edad es de 5 a 14 años, esta gráfica muestra varianza creciente, tendencia creciente y no se observa variación estacional. Para corregir el efecto de la varianza creciente se le aplica a los datos la transformación raíz cuadrada, el resultado de esto lo vemos en la gráfica 5.45, donde se observa fácilmente que la varianza se muestra constante, sin embargo, la tendencia es aún creciente, por lo que es necesario aplicar una diferencia ordinaria. Los datos ya estacionarios se observan en la gráfica 5.46 junto con su ACF y PACF. Lo que se observa en la ACF es un valor significativo en 1, la PACF tiene valores sobresalientes en 1,2,3,8. El primer modelo tentativo será un ARIMA(0,1,1), La ACF, PACF y el periodograma integral de los residuales los vemos en 5.47. En la ACF se visualiza un valor un poco significativo en 8, así como en su PACF, y en lo que respecta al periodograma integral de los residuales se tiene un buen ajuste.



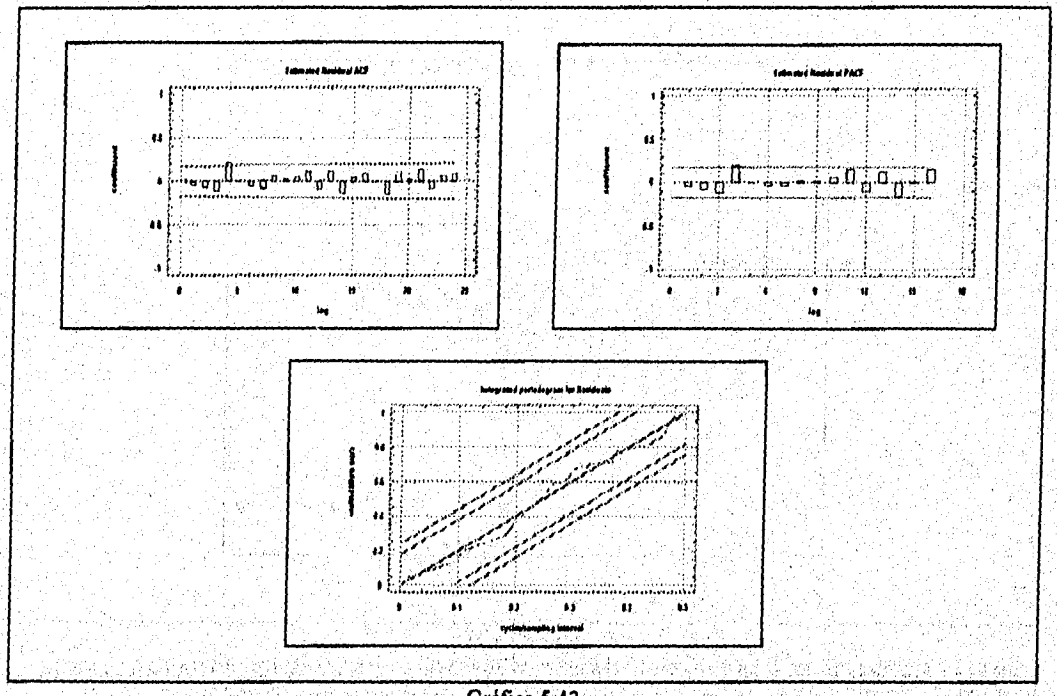
Gráfica 5.40
CASOS DIAGNOSTICADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA EDAD ES DE 0 A 4 AÑOS.



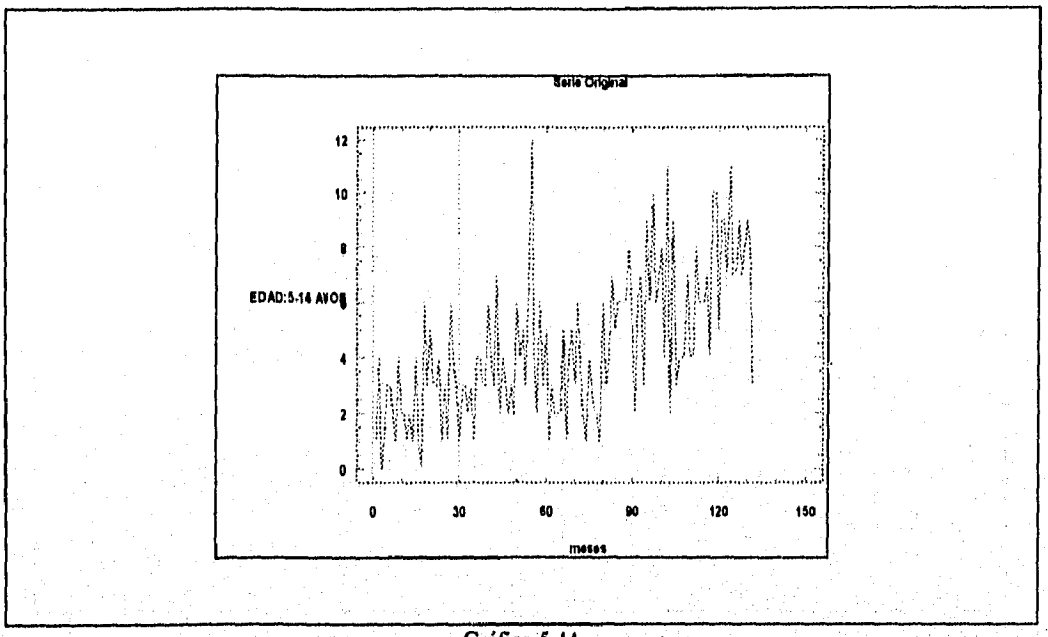
Gráfica 5.41
RAIZ CUADRADA DEL NUMERO DE CASOS MENSUALES CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA EDAD ES DE 0 A 4 AÑOS.



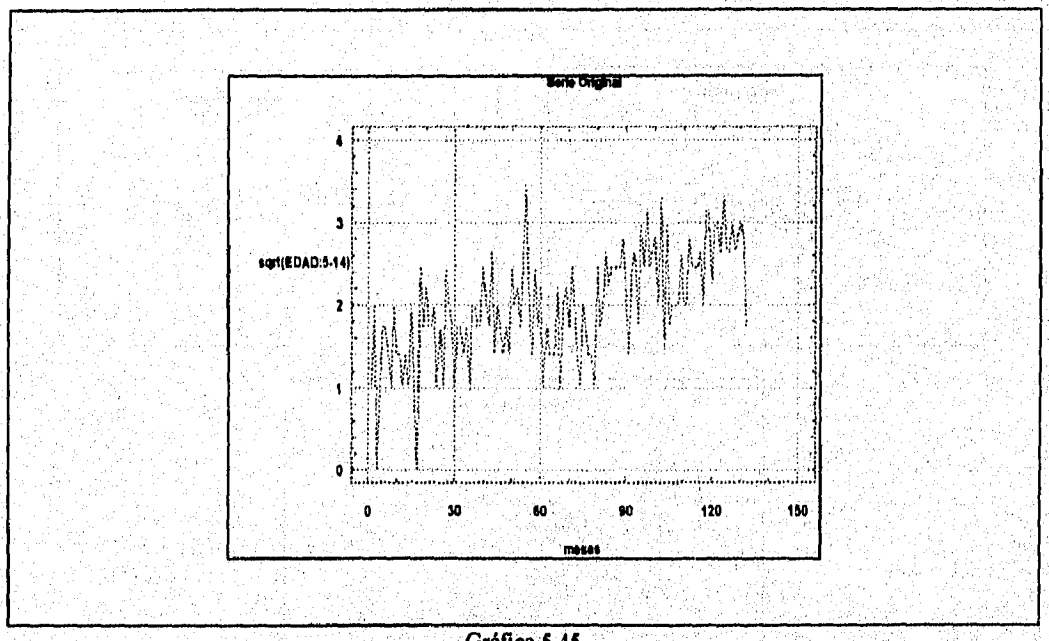
Gráfica 5.42
CASOS MENSUALES DE CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA EDAD ES DE 0 A 4 AÑOS CON LA TRANSFORMACION RAIZ CUADRADA Y UNA DIFERENCIA ORDINARIA, ACF Y PACF.



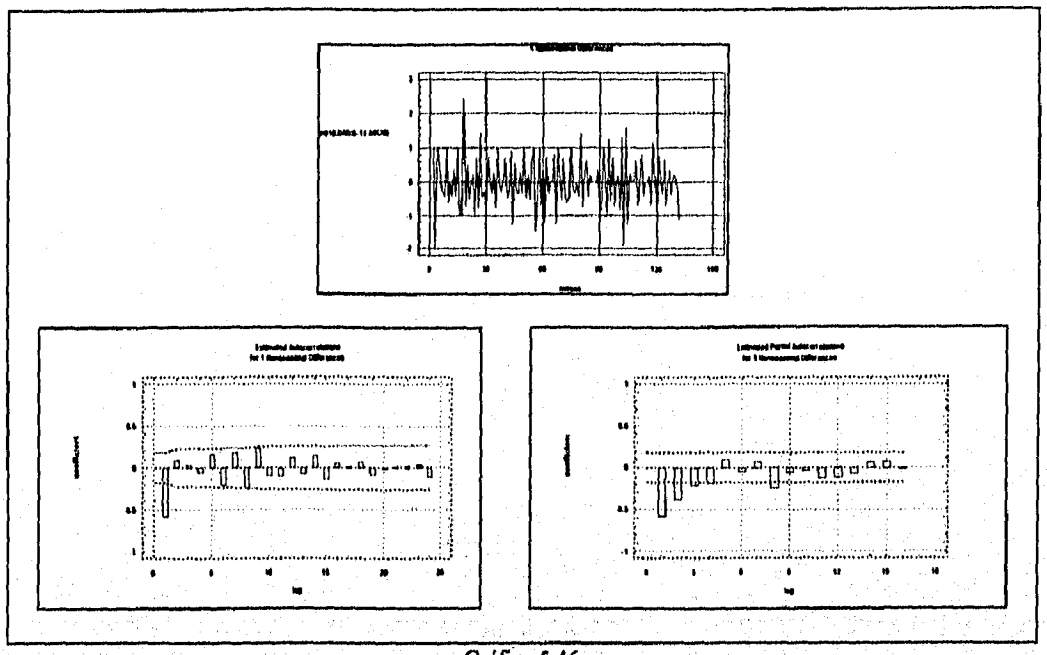
Gráfica 5.43
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(EDAD DEL PACIENTE AL DIAGNOSTICO: 0-4 AÑOS, MODELO: ARIMA(0,1,1)).



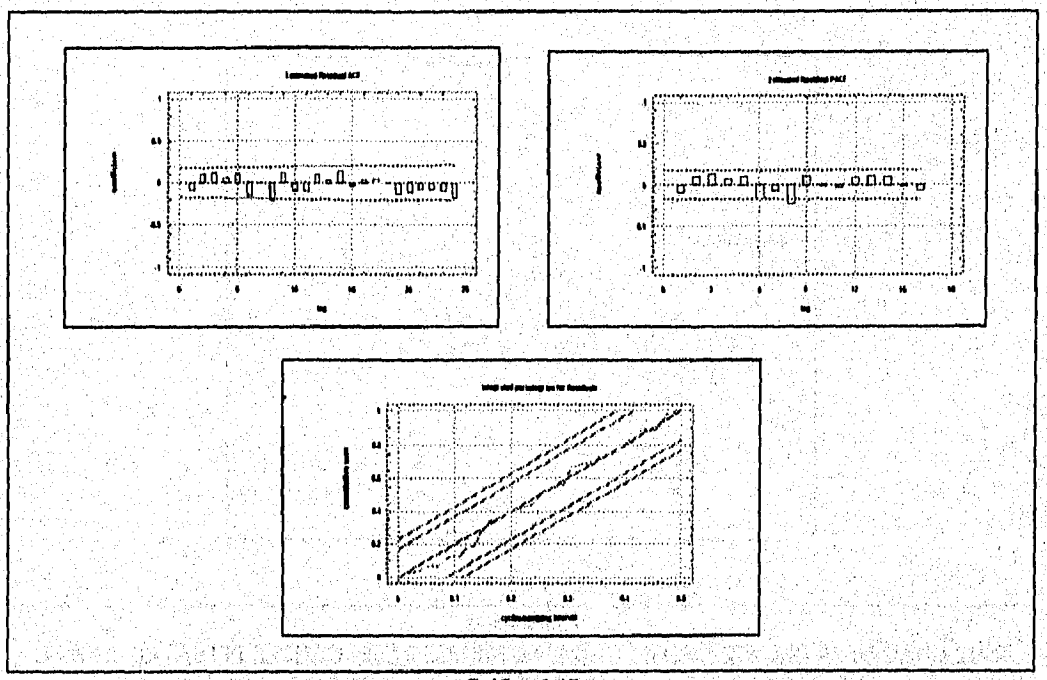
Gráfica 5.44
CASOS MENSUALES DIAGNOSTICADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA EDAD ES DE 5 A 14 AÑOS.



Gráfica 5.45
RAIZ CUADRADA DEL NUMERO DE CASOS MENSUALES CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA EDAD ES DE 5 a 14 AÑOS.



Gráfica 5.46
CASOS MENSUALES CON CANCER PEDIATRICO PARA EL D.F. EN EL PERIODO 1981-1991 CUYA EDAD ES DE 5 A 14 AÑOS CON LA TRANSFORMACION RAIZ CUADRADA Y UNA DIFERENCIA ORDINARIA ASI COMO SU ACF Y PACF.



Gráfica 5.47
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(EDAD AL DIAGNOSTICO: 5 A 14 AÑOS, MODELO: ARIMA(0,1,1)).

El cuadro comparativo de los modelos elegidos se muestra en el siguiente cuadro.

EDAD DEL NIÑO AL DIAGNOSTICO DE CANCER

	0 A 4 AÑOS	5 A 15 AÑOS
Modelo	ARIMA(0,1,1)	ARIMA(0,1,1)
Transformación	SQRT (Raíz cuadrada)	SQRT (Raíz cuadrada)
# de diferencias ordinarias	1	1
# de diferencias estacionales	--	0
# de parámetros	1	1
ϕ(1) (est. t)	--	--
ϕ(2) (est. t)	--	--
ϕ(3) (est. t)	--	--
θ(1) (est. t)	.92293 29.96235	0.92362 25.53358
Φ(1) (est. t)	--	--
Θ(1) (est. t)	--	--
Varianza	0.385204	0.28465
Desviación estándar	0.620648	0.533526
Iteraciones	6	4

Cuadro comparativo 5.5

Si ponemos atención a las gráficas originales de acuerdo a los dos grupos establecidos, es decir, la de 0 a 4 y la de 5 a 14 años, podemos ver que la tendencia de la gráfica 5.44 es más consistente que la de la gráfica 5.40. La varianza de las dos gráficas nos muestra algo semejante. En cuanto a la variación estacional, ninguna de las dos la presenta. Ahora, el modelo que mejor se ajusta a la primera gráfica es el ARIMA(0,1,1) y para la segunda gráfica también, lo que nos indica que, probablemente se tiene un mismo comportamiento para ambos grupos de edades.

Otro factor que podemos detectar es que lo primeros sesenta datos que corresponden aproximadamente de 1981 a 1985 no tiene un cierto comportamiento, sino más bien, parecen aleatorios. Esto puede deberse al sismo ocurrido en 1985, el cual, como ya se ha puesto de manifiesto, afectó a los principales hospitales del D.F. del IMSS, pero después de esta fecha se observa la tendencia (creciente) y la varianza (creciente) con mayor claridad. Luego entonces, es importante indicar que el fenómeno que se presenta en los datos puede estar afectando el análisis, ya que prácticamente se están subestimando 60 datos. En lo que respecta a los modelos encontrados para los dos grupos de edades, podemos afirmar que para ambas series, el modelo más adecuado es el ARIMA(0,1,1).

VI) ESCOLARIDAD DEL PADRE

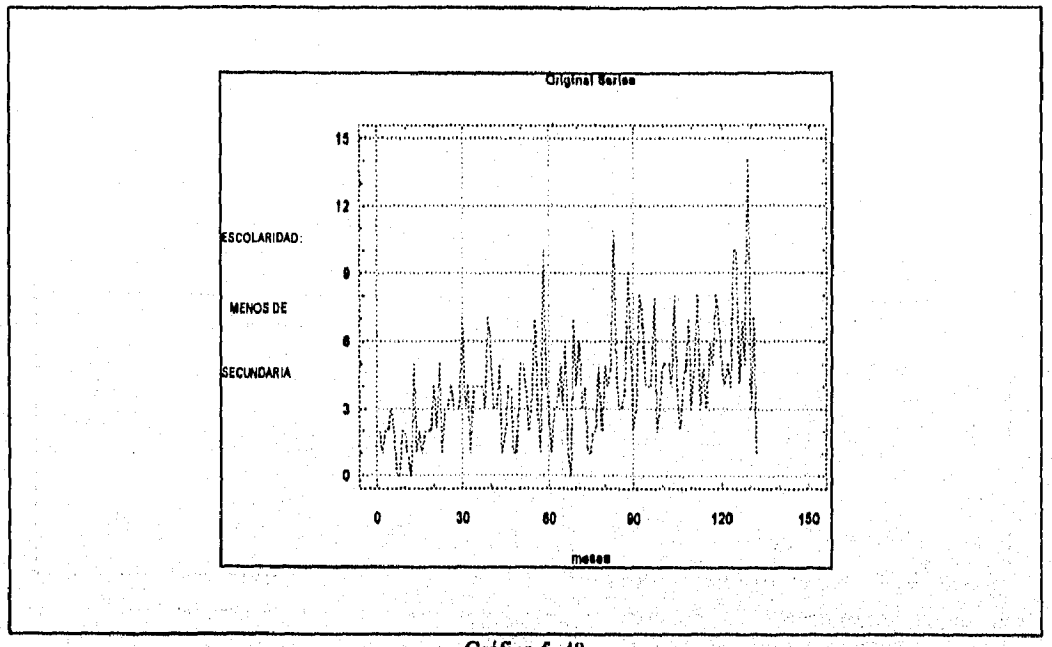
La escolaridad del padre se divide por un lado, desde aquéllos que no saben leer ni escribir hasta menos de secundaria, y por otro lado, los que tienen secundaria o más. La razón por la cual se hizo la clasificación de esta manera es de acuerdo a las frecuencias obtenidas, además de que tal división nos proporciona información en el sentido de educación.

MENOS DE SECUNDARIA

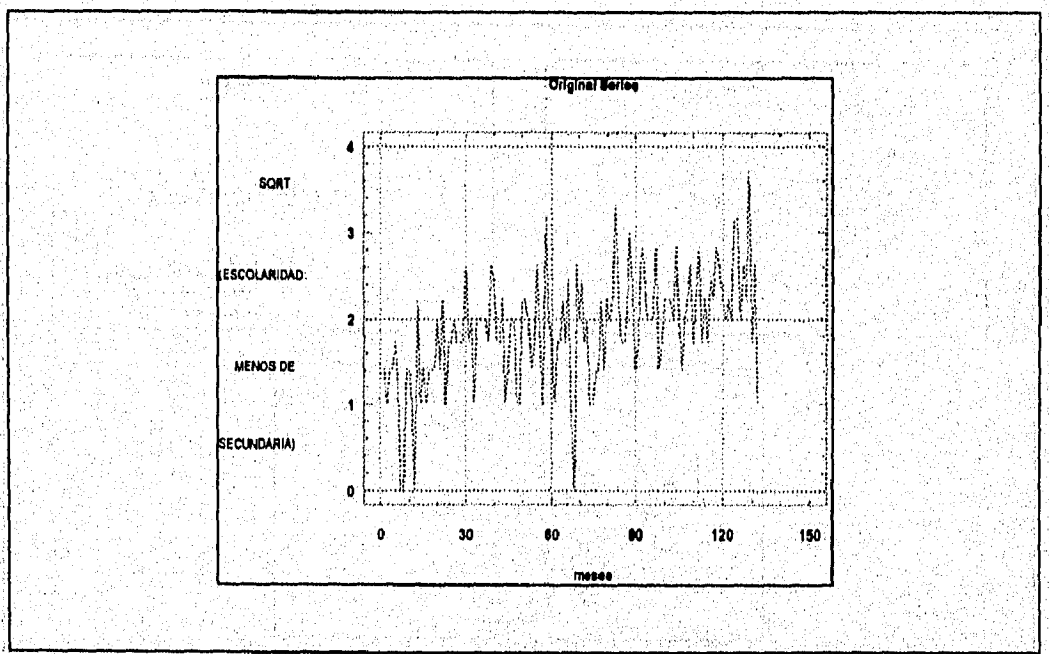
Podemos ver que la gráfica 5.48, que representa a los casos cuyos padres tienen escolaridad de menos de secundaria, tiene varianza creciente, tendencia creciente, sin rasgos claros de una variación estacional. Una vez hecha la transformación raíz cuadrada para atacar de esa forma a la varianza creciente, nos queda la gráfica 5.49. Lo que resta por hacer, debido a la tendencia que sigue presente, es una diferencia ordinaria, como lo vemos en la gráfica 5.50, también se encuentra contemplada la ACF con un valor significativo en 1, y la PACF con valores sobresalientes en 1,2,3. El modelo que se encuentra adecuado es el ARIMA(4,1,0), para el cual, ver gráfica 5.51, la ACF y PACF de los residuales denotan el comportamiento típico del ruido blanco. En el periodograma integral de los residuales se tiene un ajuste bastante mejorado con respecto al del modelo anterior.

SECUNDARIA O MAS

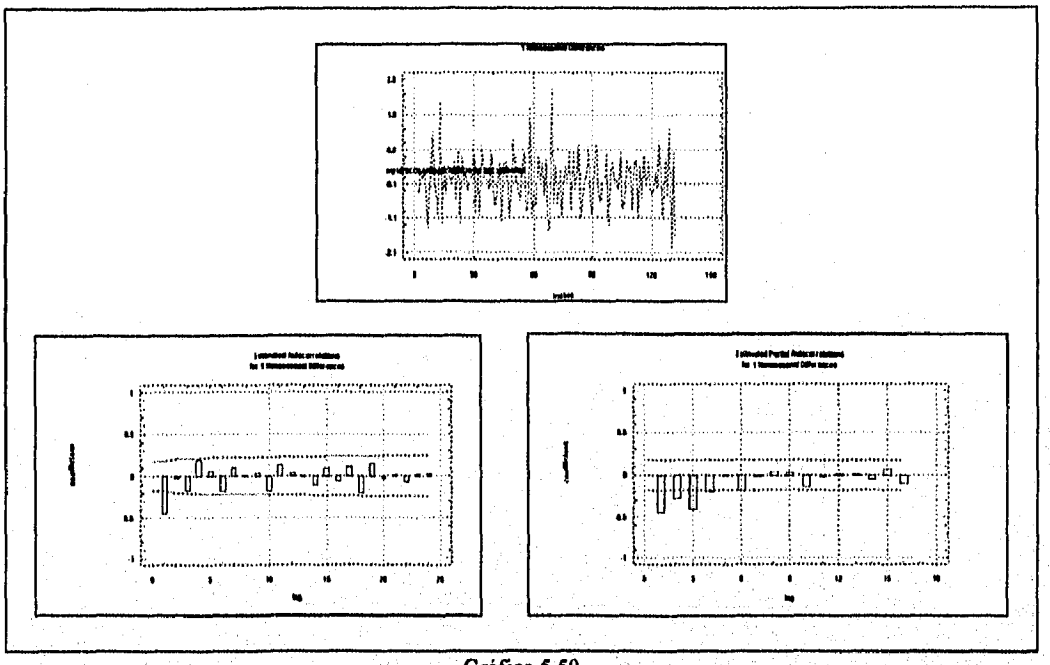
Empecemos el estudio de estos datos de igual forma a las anteriores, es decir, con su respectiva serie original. Para ello observemos con atención a la gráfica 5.52, donde es evidente que hay tendencia creciente, la varianza se muestra también un poco creciente, la variación estacional no se pone de manifiesto. Después de nombrar las características de la serie, resulta preciso eliminar la varianza mediante la transformación raíz cuadrada. La serie que nos resulta es la que se ve en la gráfica 5.53. Después de aplicar la diferencia ordinaria para eliminar la tendencia, así como la ACF y PACF las tenemos en la gráfica 5.54. La serie diferenciada nos indica que ya es estacionaria, la ACF tiene un valor significativo en 1 y la PACF los tiene en 1, 2, 3 y 6. El modelo que se toma es un ARIMA(0,1,1), para el que se obtiene su respectivo análisis de los residuales como lo indica la gráfica 5.55 en donde se observa un comportamiento similar al ruido blanco por parte de la función de autocorrelación y la función de autocorrelación parcial de los residuales, a su vez el periodograma integral de los residuales deja ver que se tiene un buen ajuste.



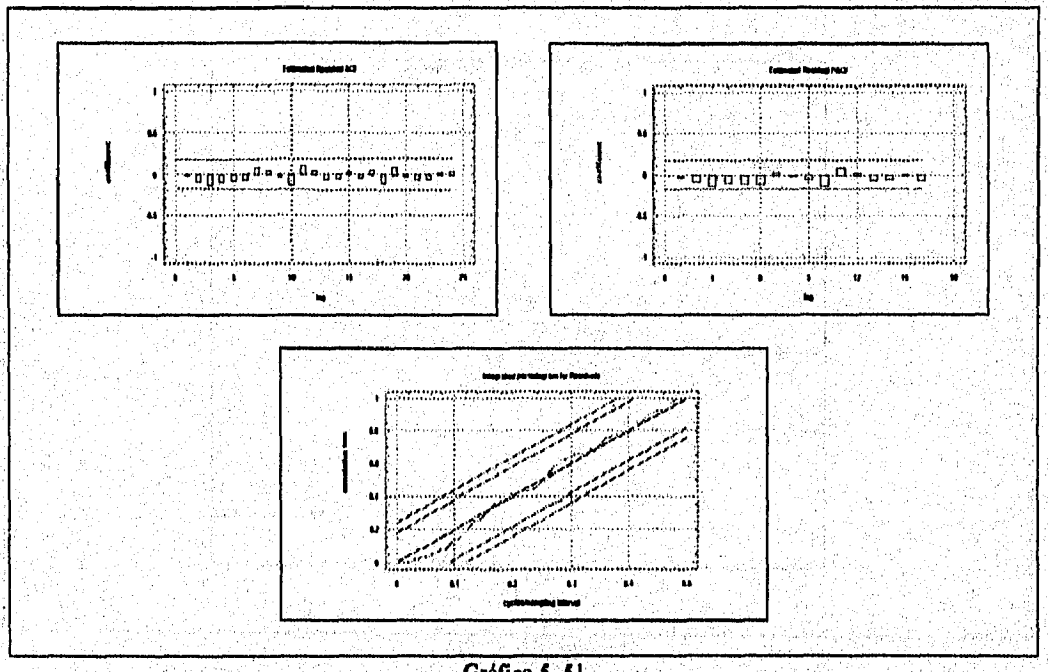
Gráfica 5. 48
CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA ESCOLARIDAD DEL PADRE ES MENOS DE SECUNDARIA.



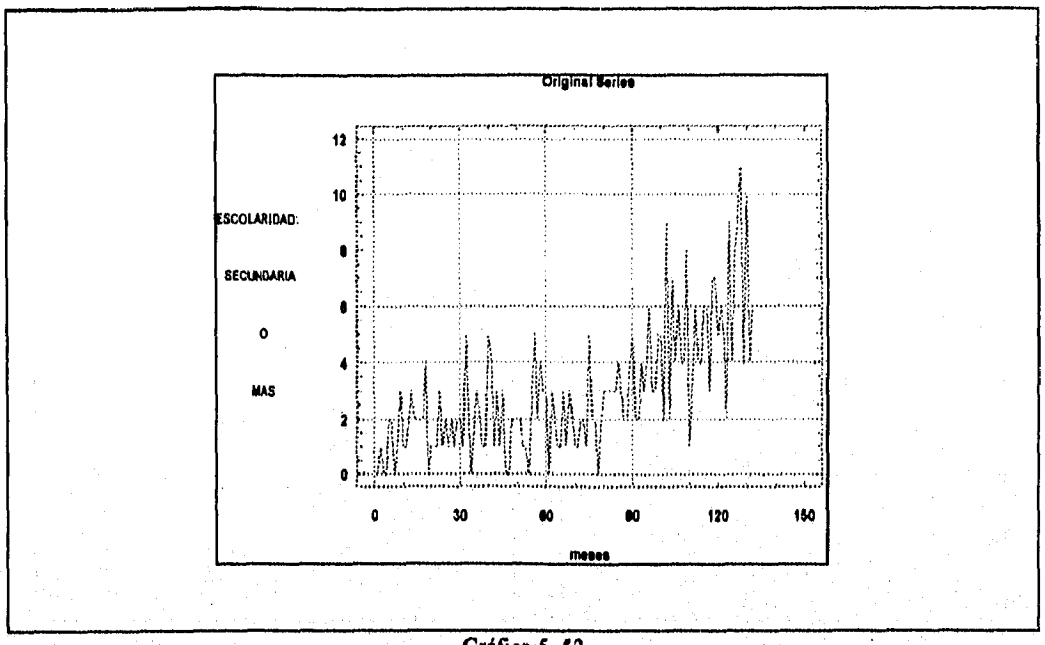
Gráfica 5. 49
RAIZ CUADRADA DEL NUMERO DE CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYO PADRE TUVO ESCOLARIDAD DE MENOS DE SECUNDARIA.



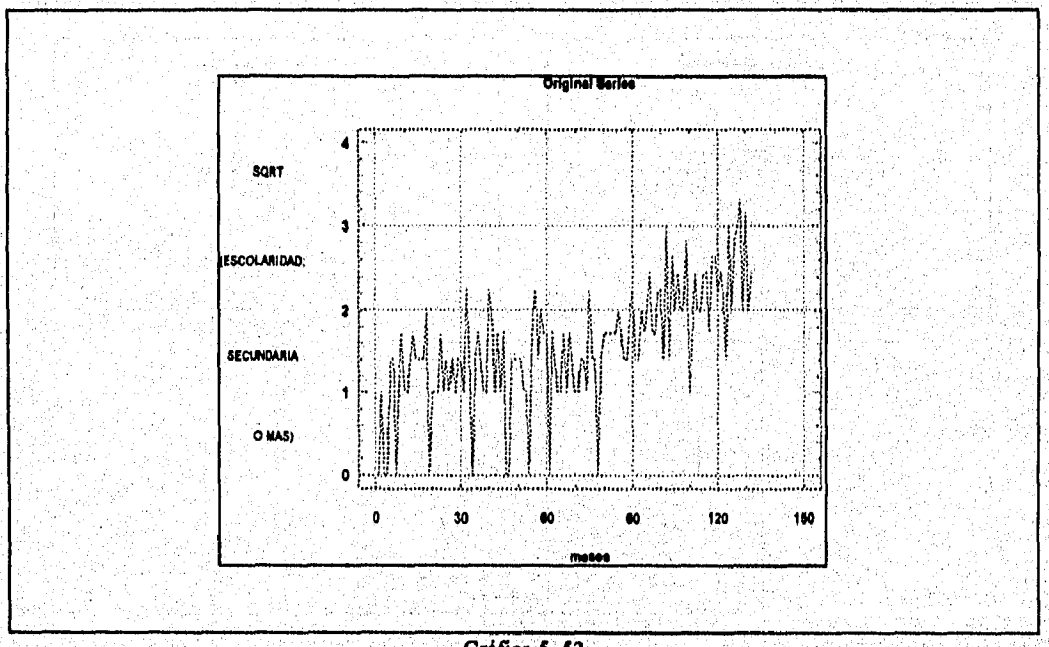
Gráfica 5.50
CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYO PADRE TUVO ESCOLARIDAD DE MENOS DE SECUNDARIA, CON LA TRANSFORMACION RAIZ CUADRADA Y UNA DIFERENCIA ORDINARIA, ACF Y PACF.



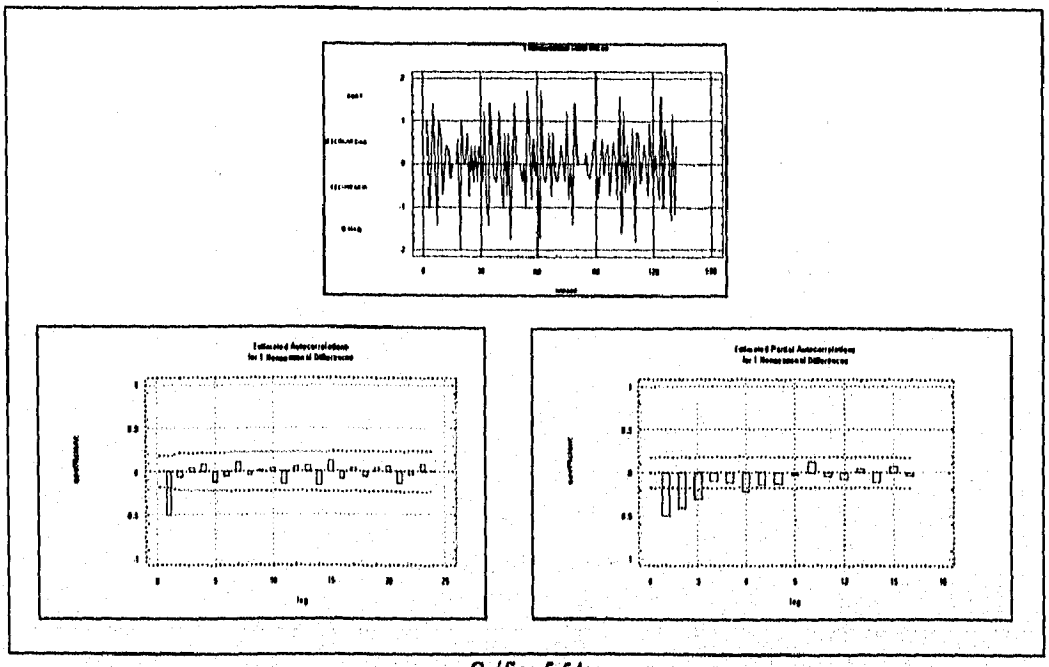
Gráfica 5. 51
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(ESCOLARIDAD DEL PADRE: MENOS DE SECUNDARIA, MODELO: ARIMA(4,1,0)).



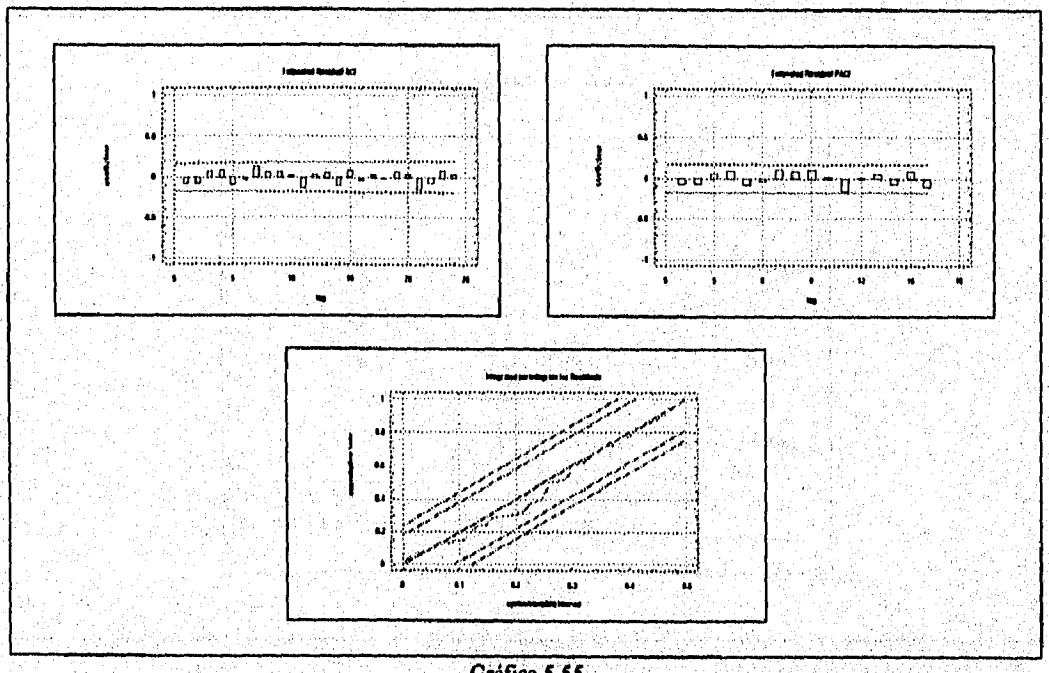
Gráfica 5. 52
CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. EN EL PERIODO 1981-1991 CUYA ESCOLARIDAD DEL PADRE ES SECUNDARIA O MAS.



Gráfica 5. 53
RAIZ CUADRADA DEL NUMERO DE CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYO PADRE TUVO ESCOLARIDAD DE SECUNDARIA O MAS.



Gráfica 5.54
CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA ESCOLARIDAD DEL PADRE ES DE SECUNDARIA, CON LA TRANSFORMACION RAIZ CUADRADA Y UNA DIFERENCIA ORDINARIA, ACF Y PACF.



Gráfica 5.55
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(ESCOLARIDAD DEL PADRE: SECUNDARIA O MAS, MODELO: ARIMA(0,1,1)).

ESCOLARIDAD DEL PADRE

	MENOS DE SECUNDARIA	SECUNDARIA O MAS
Modelo	ARIMA(4,1,0)	ARIMA(0,1,1)
Transformación	(SQRT Raíz cuadrada)	(SQRT Raíz cuadrada)
# de diferencias ordinarias	1	1
# de diferencias estacionales	--	--
# de parámetros	4	1
ϕ(1) (est. t)	-.83319 -9.43017	--
ϕ(2) (est. t)	-.68392 -6.73686	--
ϕ(3) (est. t)	-.64669 -6.33778	--
ϕ(4) (est. t)	-.25051 -2.74874	--
θ(1) (est. t)	--	.85854 19.71256
Φ(1) (est. t)	--	--
Θ(1) (est. t)	--	--
Varianza	0.357509	0.331295
Desviación estándar	0.59792	0.575583
Iteraciones	4	4

Cuadro comparativo 5.6

De acuerdo al cuadro comparativo 5.6 nos podemos dar cuenta que los modelos ajustados son diferentes para los dos casos, lo que indica que el comportamiento de la serie de datos para los casos cuyo padre tuvo escolaridad de menos de secundaria es diferente a los de secundaria o más. En cuanto a las gráficas originales, se vuelve a observar un cambio significativo a partir del 85 aproximadamente para ambas series, también parece tener una tendencia más pronunciada en los casos cuyos padres tuvieron secundaria o más, esto se puede deber a que la educación de las personas ha ido en aumento. Otro punto que ya se ha comentado con otras series es el hecho de que no se presenta variación estacional, como era de esperarse, pues la serie general tampoco la presenta.

VII) ESCOLARIDAD DE LA MADRE

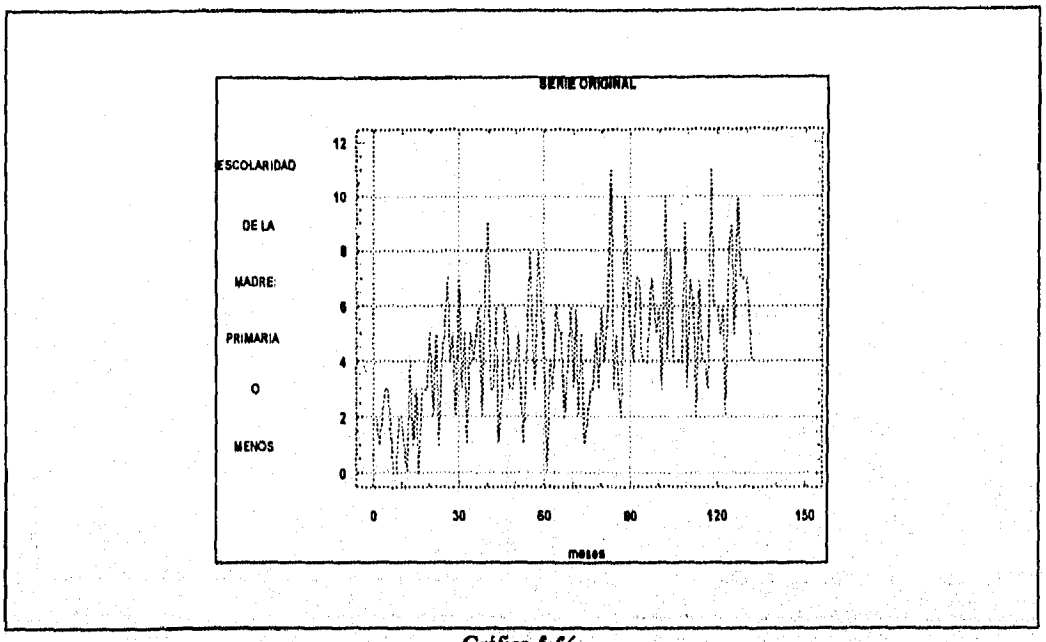
La escolaridad de la madre está dividida de acuerdo a dos categorías; de primaria o menos y de más de primaria. El hacer esta división tiene fundamentos de tipo cultural y social, ya que realmente es significativa la separación que se hace en cuanto a éstas dos categorías. Además, los datos se prestaron para hacer tal división, es decir, se encontraron series adecuadas para llevar a cabo el análisis de series de tiempo.

PRIMARIA O MENOS

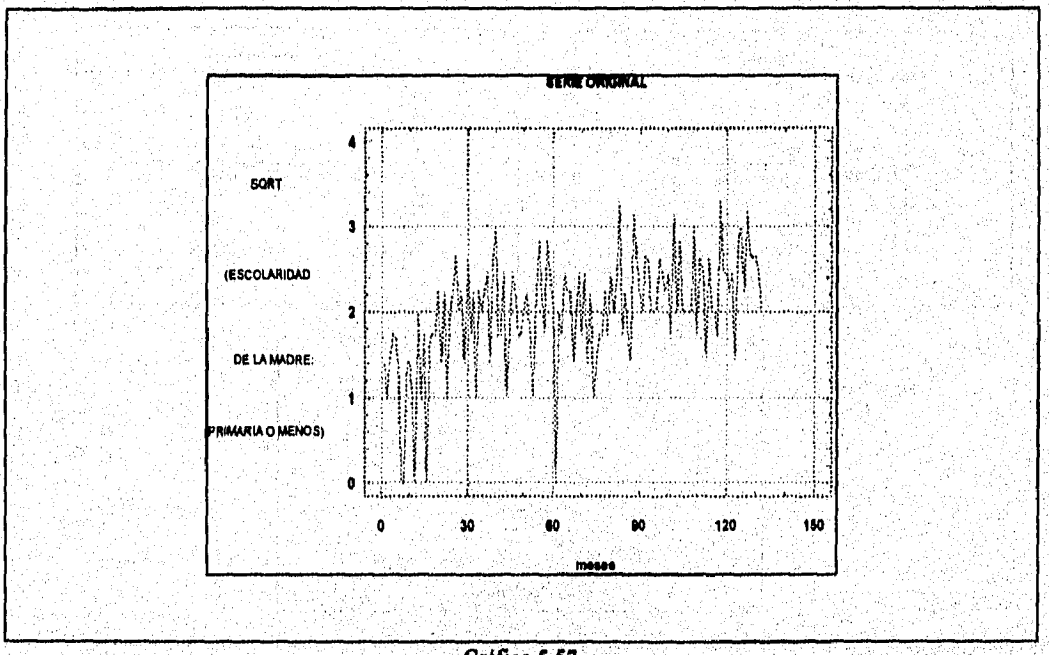
Comenzaremos esta parte revisando la gráfica de los casos cuya madre alcanzó como nivel máximo de educación hasta la primaria, es decir, incluye también las madres que tuvieron menos de primaria. Los datos a los que se refiere se encuentran en la gráfica 5.56, donde observamos que se tiene una varianza creciente, tendencia creciente, sin evidencias de variación estacional. Como siguiente paso se le aplica a los datos la transformación raíz cuadrada para eliminar la varianza creciente, ver gráfica 5.57. Después de esto, es necesario aplicar una diferencia ordinaria para eliminar la tendencia, la gráfica resultante, así como su ACF y PACF se ilustran en la gráfica 5.58. En la ACF se observa un valor significativo en 1, en la PACF encontramos valores sobresalientes en 1,2,3,4 y 8. El modelo que se considera adecuado es el ARIMA(4,1,0). Para dicho modelo se encuentra la ACF, PACF y periodograma integral de los residuales en la gráfica 5.59.

MAS DE PRIMARIA

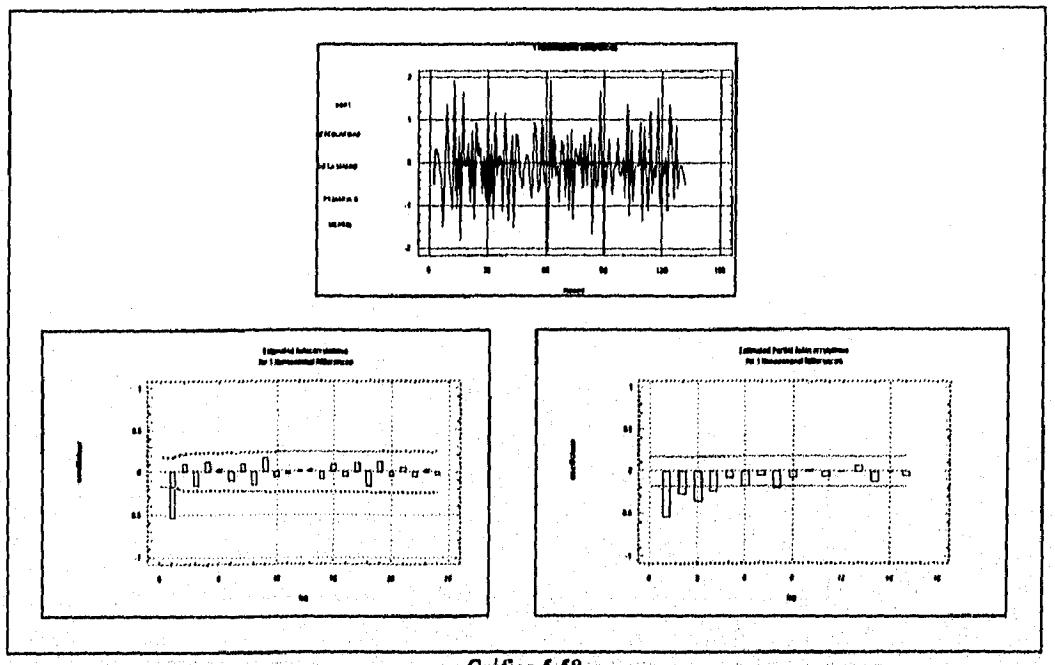
Para los casos cuyas madres tenían un nivel de escolaridad de más de primaria se presenta la gráfica 5.60. En ella podemos ver que se tiene varianza creciente, tendencia creciente y variación estacional. Al eliminar la varianza creciente por medio de la transformación raíz cuadrada nos queda la gráfica 5.61 y con la diferencia ordinaria tenemos a la gráfica 5.62, donde se aprecia que los datos ya presentan estacionaridad. En esta misma gráfica también observamos la ACF con un valor significativo en 1, y la PACF con valores sobresalientes en 1,2,3 y 6. El modelo que se adecúa a los datos es un ARIMA(0,1,1), para el cual tenemos el estudio de sus residuales por medio de la ACF, PACF y periodograma integral de los residuales, donde observamos que se tiene un buen ajuste, ver gráfica 5.63.



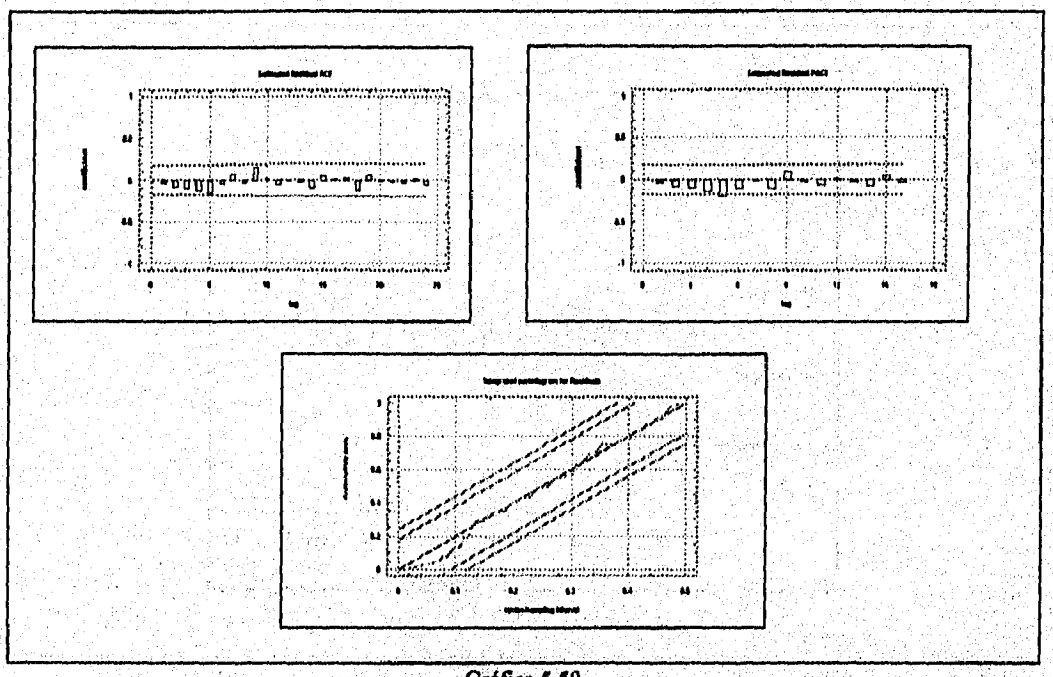
Gráfica 5.56
CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA ESCOLARIDAD DE LA MADRE ES DE PRIMARIA O MENOS.



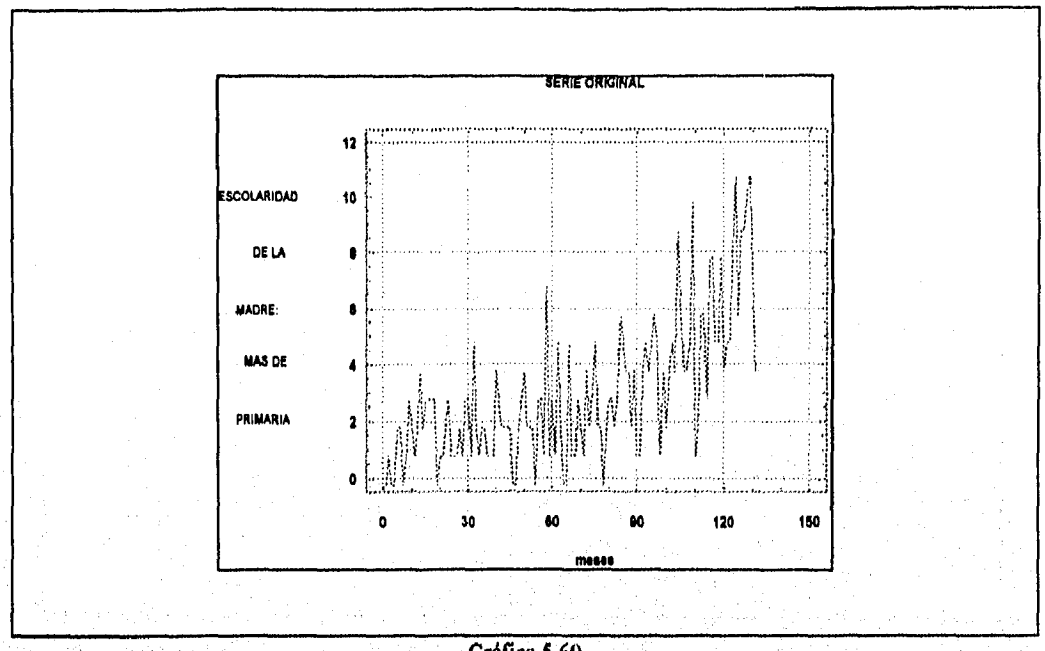
Gráfica 5.57
RAIZ CUADRADA DEL NUMERO DE CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA MADRE TUVO ESCOLARIDAD DE PRIMARIA O MENOS.



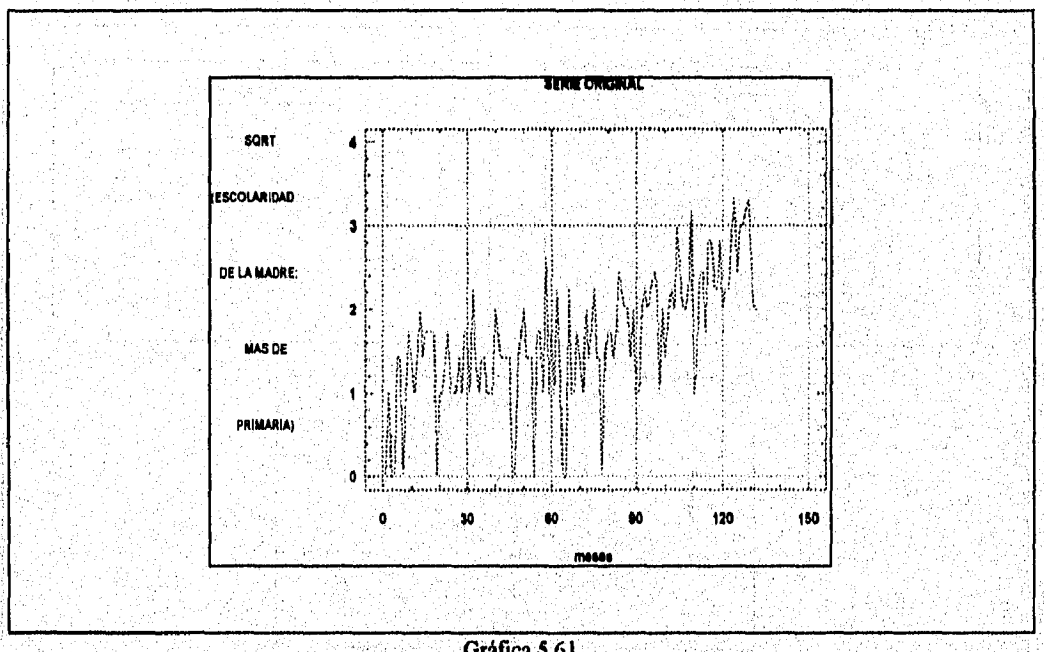
Gráfica 5.58
NUMERO DE CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA MADRE TUVO ESCOLARIDAD DE PRIMARIA O MENOS CON LA TRANSFORMACION RAIZ CUADRADA Y UNA DIFERENCIA ORDINARIA, ACF Y PACF.



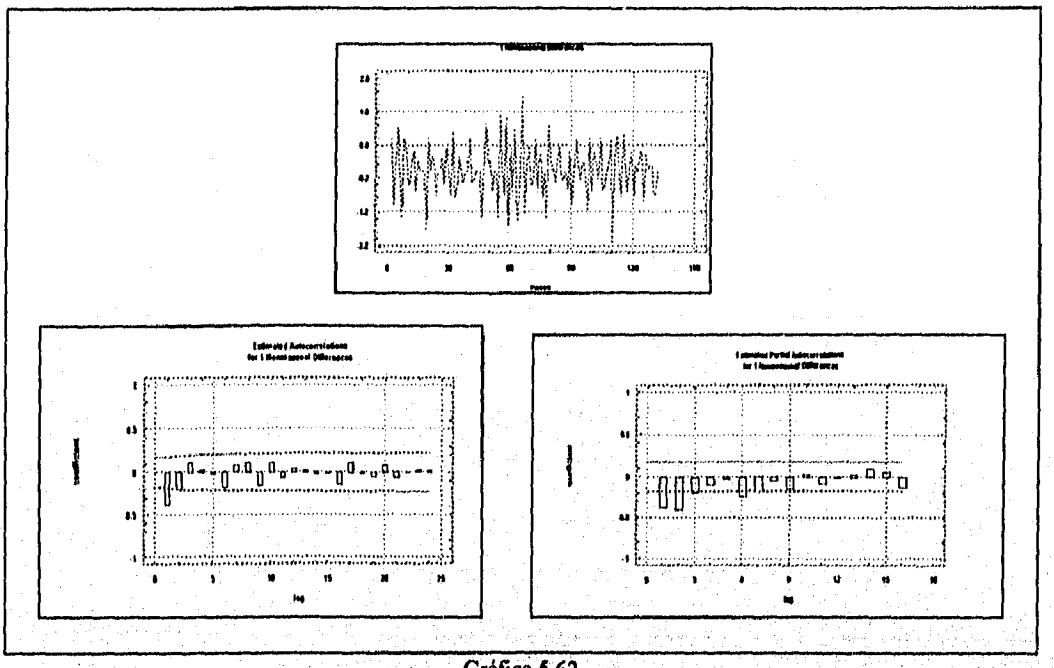
Gráfica 5.59
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(ESCOLARIDAD DE LA MADRE: PRIMARIA O MENOS, MODELO: ARIMA(4,1,0)).



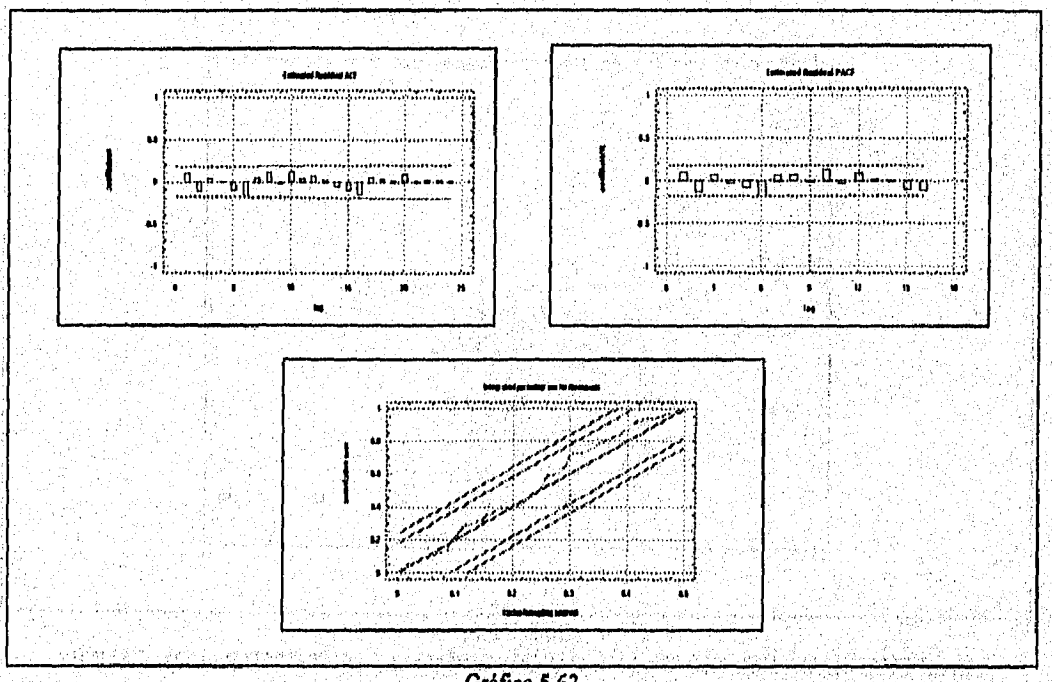
Gráfica 5.60
CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA ESCOLARIDAD DE LA MADRE ES MAS DE PRIMARIA.



Gráfica 5.61
RAIZ CUADRADA DEL NUMERO DE CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA ESCOLARIDAD DE LA MADRE ES DE MAS DE PRIMARIA.



Gráfica 5.62
CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA ESCOLARIDAD DE LA MADRE ES DE MAS DE PRIMARIA, CON LA TRANSFORMACION RAIZ CUADRADA Y UNA DIFERENCIA ORDINARIA, ACF Y PACF.



Gráfica 5.63
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(ESCOLARIDAD DE LA MADRE: MAS DE PRIMARIA, MODELO: ARIMA(0,1,1)).

ESCOLARIDAD DE LA MADRE AL NACIMIENTO DEL HIJO

	PRIMARIA O MENOS	MAS DE PRIMARIA
Modelo	ARIMA(4,1,0)	ARIMA(0,1,1)
Transformación	SQRT (Raíz cuadrada)	SQRT (Raíz cuadrada)
# de diferencias ordinarias	1	1
# de diferencias estacionales	--	--
# de parámetros	4	1
ϕ(1) (est. t)	-.85729 -9.88996	--
ϕ(2) (est. t)	-.62359 -5.90302	--
ϕ(3) (est. t)	-.53416 -5.05591	--
ϕ(4) (est. t)	-.23953 -2.76220	--
θ(1) (est. t)	--	0.84240 16.51845
Φ(1) (est. t)	--	--
Θ(1) (est. t)	--	--
Varianza	0.372673	0.378462
Desviación estándar	0.61047	0.615193
Iteraciones	1	4

Cuadro comparativo 5.7

De acuerdo a los datos que presenta el cuadro comparativo 5.7 podemos ver que el mejor modelo para la serie correspondiente a los casos cuya madre tuvo como nivel escolar de primaria o menos corresponde al modelo ARIMA(4,1,0). Por su parte, el modelo que mejor se ajusta a los datos que corresponden a los casos con escolaridad de la madre de más de primaria es un ARIMA(0,1,1), el cual es un modelo bastante sencillo, además de que fue el que presentó un mejor ajuste. Luego entonces, sí hay diferencias en cuanto a los comportamientos de las dos series de tiempo analizadas, ya que los modelos son muy diferentes.

De la misma forma que en la serie general fue posible apoyar la idea de que los datos para ambas series carecen de variación estacional, por lo que queda descartada la relación que pudiese haber entre las enfermedades que están presentando los niños contra las diferentes estaciones del año. Otro dato que no se debe pasar por alto es el hecho de que las gráficas originales de estas dos series muestran un cierto comportamiento, con una tendencia definida en lo que respecta a los aproximadamente primeros 60 datos, y después de éstos, se aprecia una caída, para enseguida empezar de nuevo la tendencia. Esto se debe probablemente, de la misma manera que en todas las series anteriores, a lo ocurrido en 1985, es decir, el terremoto que conmocionó a la ciudad de México.

También es importante que veamos como la tendencia creciente es más pronunciada en los casos cuya escolaridad de la madre es de más de primaria.

VIII) EDAD DE LA MADRE AL NACIMIENTO DEL HIJO

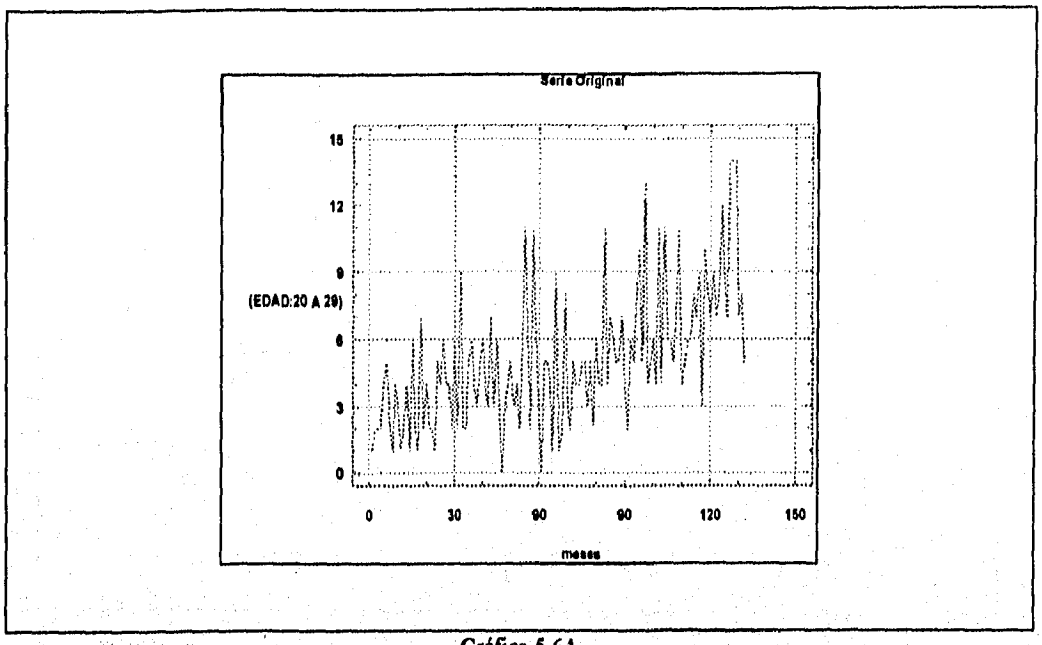
Para hacer la división de la edad de la madre al nacimiento del hijo se obtienen las frecuencias respectivas, tratando de agruparlas de tal forma que nos diga algo en cuestión médica, y que además sea posible hacer un estudio de series de tiempo. La división que se hace para tal efecto es de 20 a 29 años, considerando a estas mujeres con un bajo riesgo, el otro grupo que podemos considerar con cierto riesgo son las que tienen de 15 a 19 años junto a las de 30 años y más.

20 A 29 AÑOS

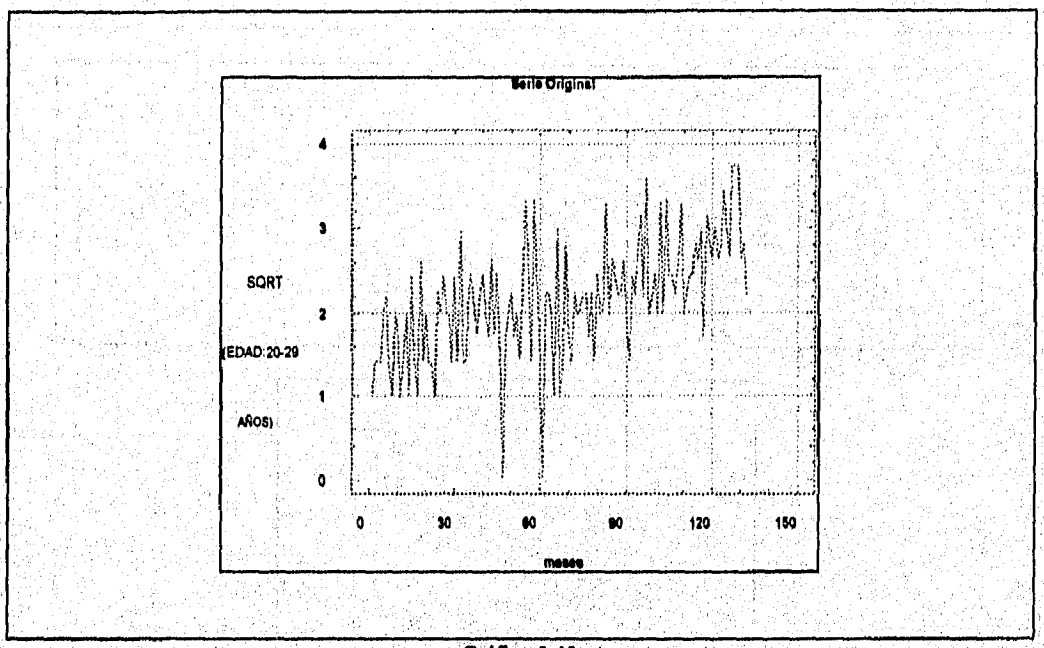
La gráfica que corresponde a los casos cuya edad de la madre es de 20 a 29 años la encontramos en la gráfica 5.64. En ésta se aprecia varianza creciente, tendencia creciente, sin fluctuación estacional. De igual forma que a las demás series de tiempo analizadas, se procede a estabilizar la varianza con la transformación raíz cuadrada, ver gráfica 5.65. Como siguiente punto, se elimina la tendencia con una diferencia ordinaria, tal como lo muestra la gráfica 5.66. En esta misma gráfica se encuentra incluida la ACF y PACF, con 1,14 y 1,2,3,13 valores sobresalientes respectivamente. El modelo que se encuentra adecuado, además de sencillo es el ARIMA(0,1,1). En la gráfica 5.67 vemos la ACF y PACF de los residuales para dicho modelo, los cuales nos indican un comportamiento típico del ruido blanco, el periodograma integral de los residuales muestra un buen ajuste, por lo que se llega a la conclusión de que es un buen modelo para los datos en cuestión.

15 A 19 AÑOS Y 30 AÑOS O MAS

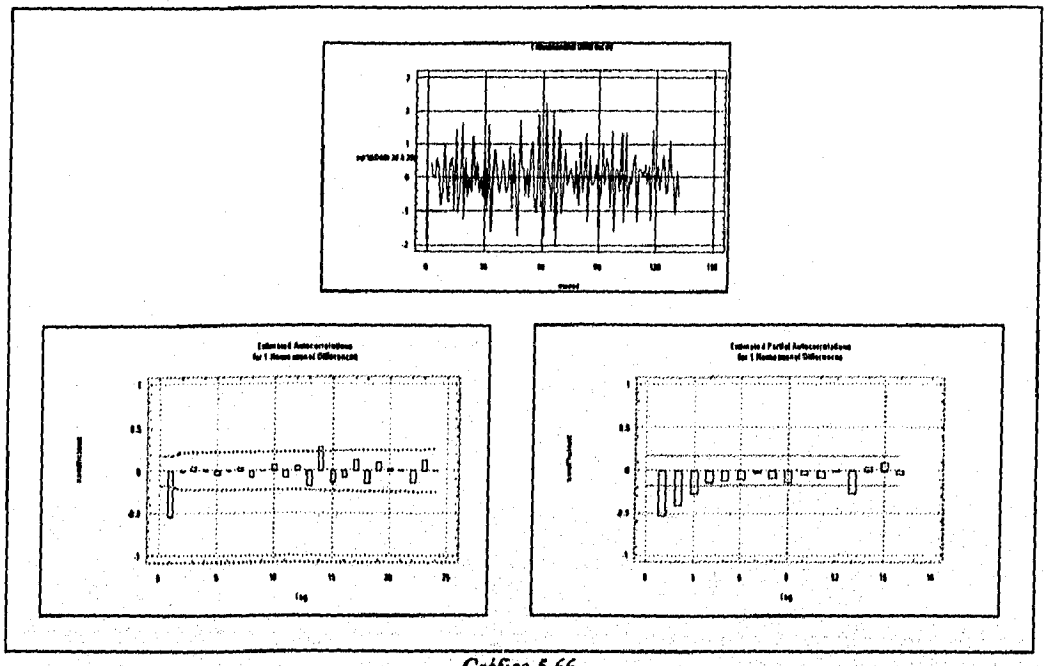
Para estos datos se encuentra, al igual que en la gráfica anterior, varianza creciente, tendencia creciente, sin presencia de variación estacional, ver gráfica 5.68. Con la transformación raíz cuadrada se logra eliminar la varianza creciente como lo demuestra la gráfica 5.69, y después de aplicarle una diferencia ordinaria se logra una serie de tiempo estacionaria, como lo vemos en la gráfica 5.70, donde además visualizamos la ACF y PACF. El modelo propuesto que también resulta adecuado es el ARIMA(0,1,1). Observemos en la gráfica 5.71 a la ACF, PACF y periodograma integral de los residuales para dicho modelo, los cuales apoyan la decisión de tomarlo como el óptimo.



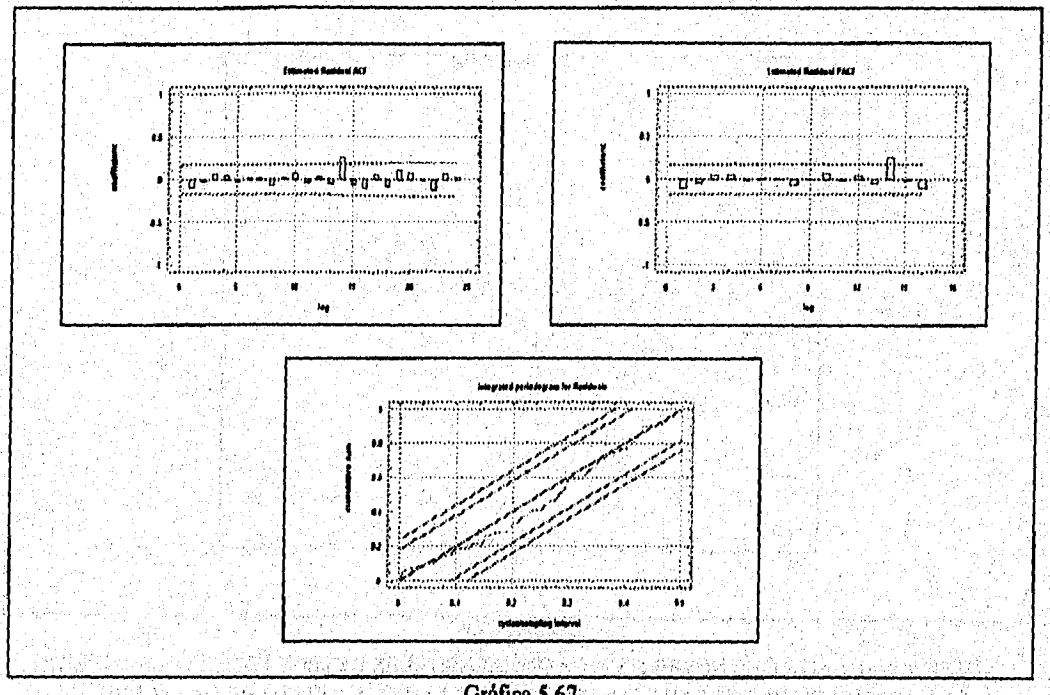
Gráfica 5.64
CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA EDAD DE LA MADRE AL NACIMIENTO DEL HIJO FUE DE 20-29 AÑOS.



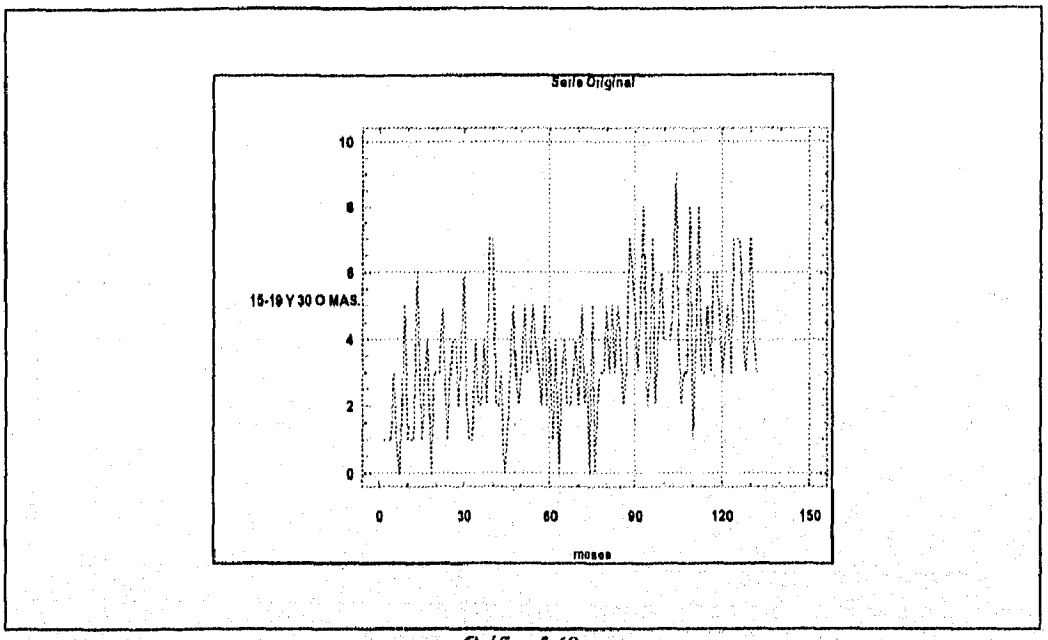
Gráfica 5.65
RAIZ CUADRADA DEL NUMERO DE CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA EDAD DE LA MADRE AL NACIMIENTO DEL HIJO FUE DE 20-29 AÑOS.



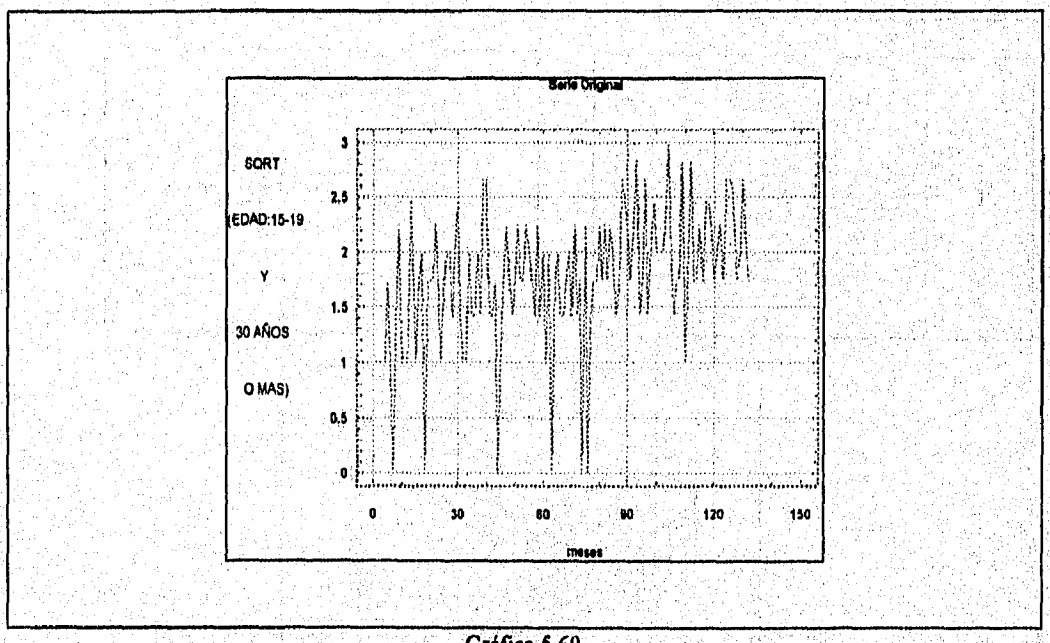
Gráfica 5.66
CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA EDAD DE LA MADRE AL NACIMIENTO DEL HIJO FUE DE 20-29 AÑOS, CON LA TRANSFORMACION RAIZ CUADRADA Y UNA DIFERENCIA ORDINARIA, ACF Y PACF.



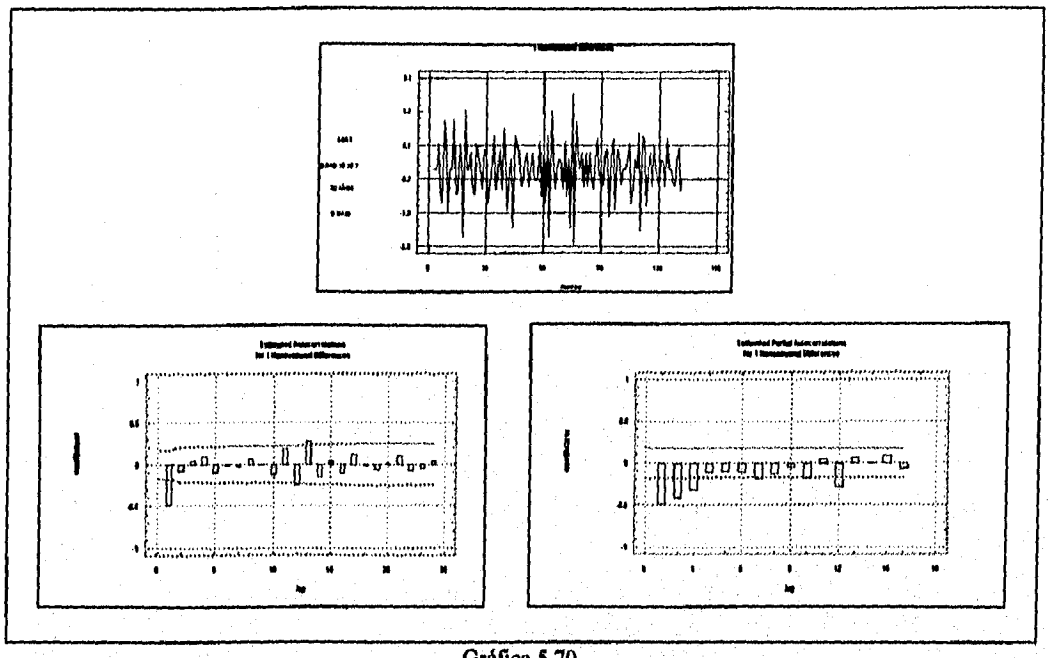
Gráfica 5.67
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
EDAD DE LA MADRE AL NACIMIENTO DEL HIJO: 20-29 AÑOS, MODELO(ARIMA(0,1,1)).



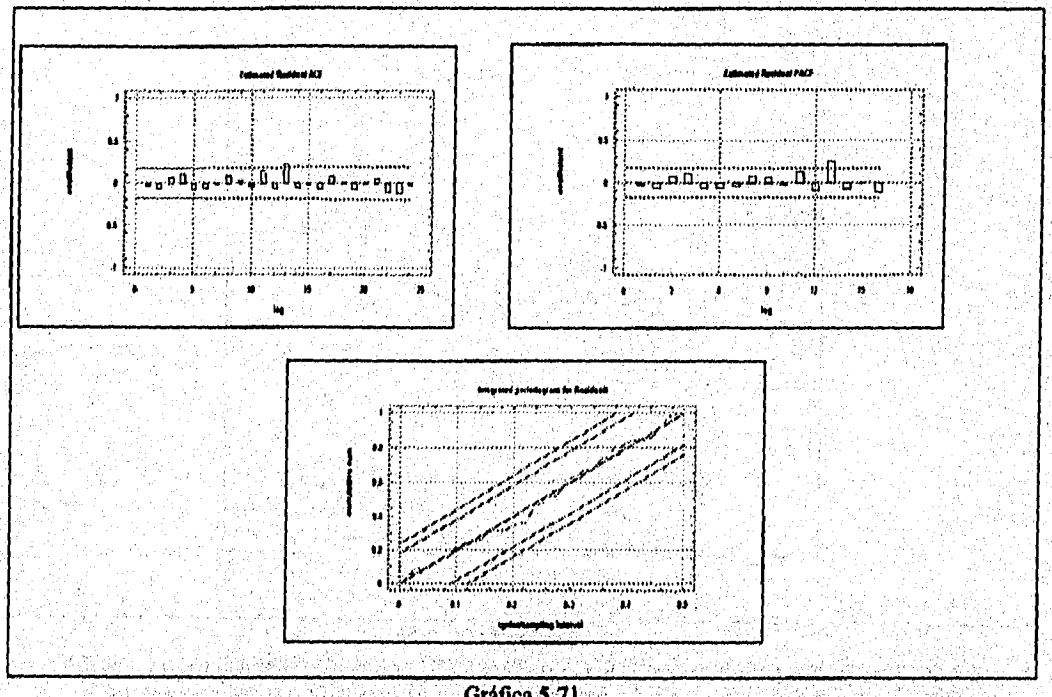
Gráfica 5.68
CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA EDAD DE LA MADRE AL NACIMIENTO DEL HIJO FUE DE 15-19 Y 30 AÑOS O MAS.



Gráfica 5.69
RAIZ CUADRADA DEL NUMERO DE CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA EDAD DE LA MADRE AL NACIMIENTO DEL HIJO FUE DE 20-29 AÑOS.



Gráfica 5.70
CASOS MENSUALES DETECTADOS CON CANCER PEDIATRICO EN EL D.F. PARA EL PERIODO 1981-1991 CUYA EDAD DE LA MADRE AL NACIMIENTO DEL HIJO FUE DE 15-19 Y 30 AÑOS O MAS, CON LA TRANSFORMACION RAIZ CUADRADA Y UNA DIFERENCIA ORDINARIA, ACF Y PACF.



Gráfica 5.71
ACF, PACF Y PERIODOGRAMA INTEGRAL DE LOS RESIDUALES
(EDAD DE LA MADRE AL NACIMIENTO DEL HIJO: 15-19 Y 30 AÑOS O MAS, MODELO: ARIMA(0,1,1)).

EDAD DE LA MADRE AL NACIMIENTO DEL HIJO

	15-19 Y 30 AÑOS O MAS	20-29 AÑOS
Modelo	ARIMA(0,1,1)	ARIMA(0,1,1)
Transformación	SQRT (Raíz cuadrada)	SQRT (Raíz cuadrada)
# de diferencias ordinarias	1	1
# de diferencias estacionales	--	--
# de parámetros	1	1
ϕ(1) (est. t)	--	--
ϕ(2) (est. t)	--	--
ϕ(3) (est. t)	--	--
ϕ(4) (est. t)	--	--
θ(1) (est. t)	0.95334 40.28766	0.93323 .35.48543
Φ(1) (est. t)	--	--
Θ(1) (est. t)	--	--
Varianza	0.32736	0.358606
Desviación estándar	0.572154	0.598837
Iteraciones	3	5

Cuadro comparativo 5.8

De acuerdo a lo encontrado en las series de tiempo analizadas, es posible darnos cuenta que fundamentalmente tienen un comportamiento similar. Esto es posible deducirlo por el simple hecho de que las dos series tienen características semejantes en su serie original, como lo es la media no constante, varianza creciente, y sin fluctuación estacional. Además, el modelo elegido para ambos casos es el ARIMA(0,1,1), lo cual viene a apoyar lo anterior. Por otro lado la notoria pérdida de información que se da en el 85 también se encuentra reflejado en estas dos últimas series.

Los datos que corresponden a la edad de la madre de 20 a 29 años tienen una tendencia más marcada, lo cual indica que el proceso se encuentra más determinado en cuanto a los casos de madres con dicha edad. Esto no es extraño, ya que existen los medios para dar información a las parejas del mejor momento para tener hijos, y se ha dado apoyo a las campañas de planificación familiar. Es de esperarse que si se siguiera recolectando datos para años posteriores al estudio, la tendencia estaría de acuerdo a las madres que pertenecieran a este intervalo de edad. Otro hecho que se desea recalcar es el factor variación estacional, el cual no está presente, lo cual indica y refuerza aún más la conclusión de que no existe relación alguna entre la época del año y los casos de pacientes que se les detecta cáncer, independientemente de la edad de la madre al nacimiento del niño.

CONCLUSIONES

Para ilustrar las conclusiones se hará uso de las gráficas originales, las cuales se muestran en su conjunto en el cuadro A, y de dos cuadros comparativos: el B y el C. En el primer cuadro se resumen todas las series analizadas con elementos tales como la existencia o no de variación estacional, tendencia, el tipo de transformación y el modelo elegido. En el segundo cuadro se muestran las variables que, de acuerdo a la prueba de bondad de ajuste X^2, presentan asociación con algún tipo de cáncer y que además tienen una frecuencia mayor a la esperada en caso de independencia.

Si comenzamos con la simple inspección visual de las gráficas, con el objeto de hacer una comparación de la serie general contra todas las demás variables, observamos que a simple vista, las que más se asemejan a ésta son la serie de casos cuya edad de la madre al nacimiento del hijo fue de 20 a 29 años, así como la de los pacientes que negaron tener familiares con cáncer. En realidad, estas dos series presentan un 59.8% y un 83.4% del total de casos que contestaron a estas cuestiones, respectivamente.

En el cuadro B, para cada una de las series, el modelo que se tiene como óptimo presenta la transformación raíz cuadrada. Esto nos indica que los datos tienen varianza creciente, la cual es proporcional al crecimiento de la serie. Es decir, conforme el tiempo transcurre, el aumento, en porcentaje, de casos notificados con cáncer pediátrico es mayor. De la misma forma, las variables presentan tendencia creciente. Esto indica que el número de casos está aumentando, tanto en la serie general, como en las catorce divisiones establecidas. Respecto a una posible asociación del cáncer pediátrico con las estaciones del año, podemos ver que ninguna de las series presenta la parte estacional y esto se puede ver también con los modelos elegidos, ya que ninguno considera la estacionalidad.

Todas las series, sin incluir a la general, presentan como mejor modelo al ARIMA(0,1,1), excepto la correspondiente al tipo de cáncer leucemia y a la de escolaridad del padre con menos de secundaria, las cuales tienen como mejor modelo al ARIMA(4,1,0). Esto indica que la mayoría de los procesos son semejantes, pero también es importante hacer la observación de que el porcentaje de casos cuyo padre tiene escolaridad de menos de secundaria es del 59.7% del total de casos con leucemia.

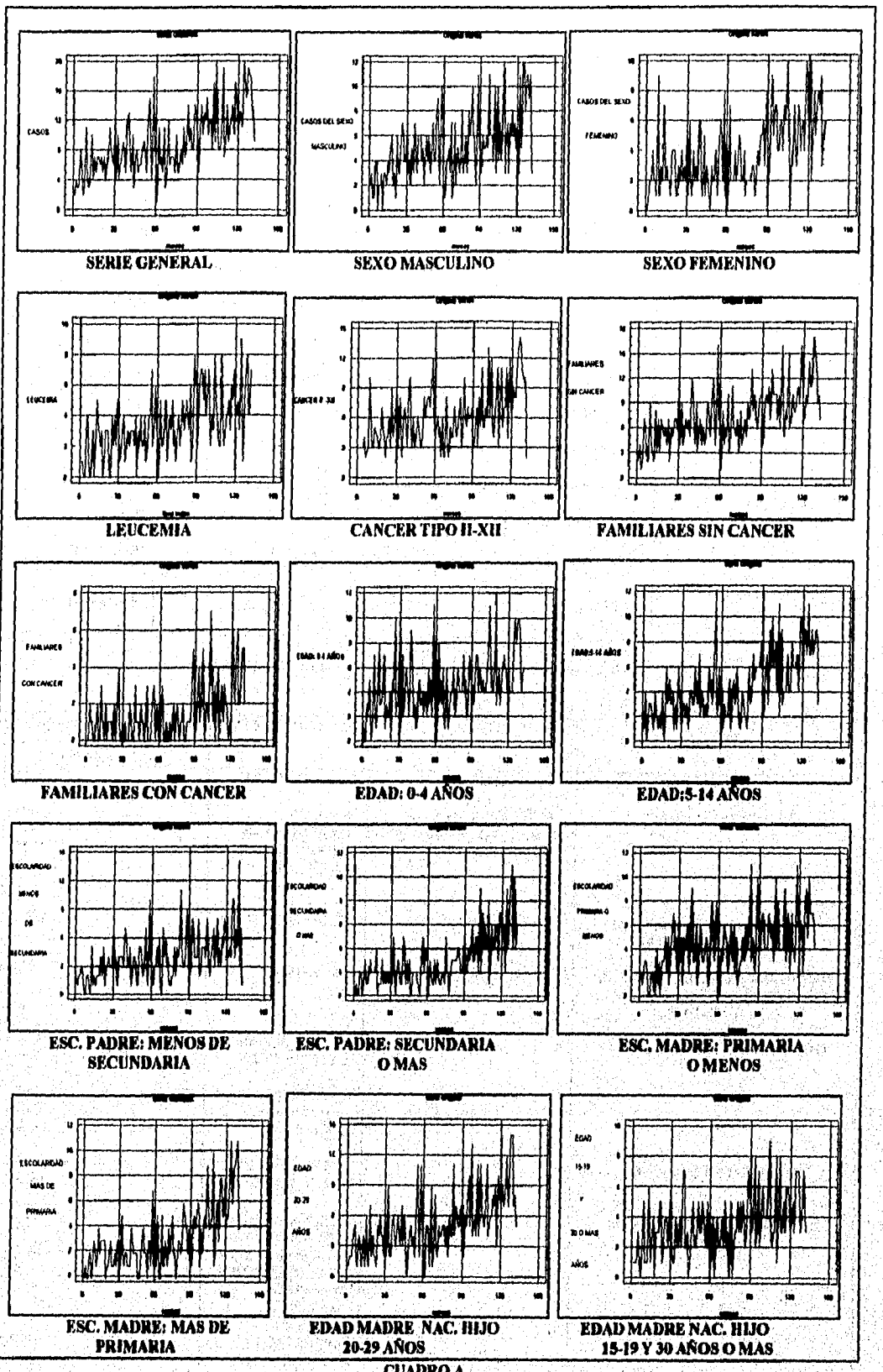
Del cuadro B es preciso subrayar tres aspectos: primero, la serie general presenta como modelo óptimo al ARIMA(3,1,0) y ninguna de las otras subseries lo tiene como mejor modelo. Esto implica que ninguna de las variables analizadas presenta un comportamiento completamente semejante al de la serie general. Lo anterior lo podemos observar en el cuadro A, donde todas las series presentan un comportamiento similar. Sin embargo, deben existir diferencias, puesto que se intentó ajustar a la serie general un modelo ARIMA(0,1,1) y aunque se encontró que la varianza y la desviación estándar eran menores al modelo elegido, la ACF y PACF de los residuales mostraron un valor significativo en 4; además el periodograma integral de los residuales no tenía un buen ajuste. Segundo, la serie de casos cuyo padre tuvo escolaridad de menos de secundaria y los casos con leucemia presentan un mismo modelo óptimo, el ARIMA(4,1,0). Esto nos dice que ambas series tienen un comportamiento similar, pero no hay que olvidar que del total de casos con leucemia, los casos cuyo padre tuvo escolaridad de menos de secundaria representan un 59.7%, el cual es mayoría. Tercero, la comparación entre cada una de las variables nos indica que la leucemia tiene un comportamiento diferente a la serie de los demás cánceres. A su vez, los casos cuya escolaridad del padre es de menos de secundaria es diferente a los de secundaria o más. Para todos los demás casos los modelos elegidos como óptimos son los mismos. Esto indica que los comportamientos de sus series siguen trayectorias análogas, que si se unen en una sola, forman un proceso un poco diferente, al cual se le ajusta el modelo ARIMA(3,1,0), estamos hablando de la serie general.

La información que presenta el cuadro comparativo C es también relevante, puesto que nos da elementos sumamente interesantes en materia de salud. El tumor de Wilms presenta un porcentaje mayor a la frecuencia esperada en caso de independencia en un 44.1%. Lo anterior indica que esta enfermedad guarda relación con el sexo femenino. En lo que respecta a los cuatro tipos de cáncer que se señalan en las instituciones del Centro Médico La Raza y el Hospital General de la Secretaría de Salud, se tendría que estudiar con mayor detenimiento cuestiones de tipo social o incluso ambiental, para llegar a plantear hipótesis sobre la relación existente entre los cánceres mencionados y las instituciones que presentan una frecuencia mayor a la esperada en caso de independencia. También existe la posibilidad de que halla una especialización del hospital en cuanto algún tipo de enfermedad.

El tumor hepático, por su parte, se relaciona con los casos que tuvieron madres con escolaridad de secundaria o más. De acuerdo al marco de referencia, este tipo de tumores presenta como factor de riesgo la exposición a sustancias industriales y a las hormonas usadas en el tratamiento de algunas anemias, por lo que se puede investigar el vínculo entre una escolaridad de la madre de más de secundaria y los aspectos señalados. El tiempo transcurrido entre el inicio de signos y síntomas al diagnóstico de más de un mes tiene asociación con el tumor óseo. Esto nos indica de la probable dificultad que existe al diagnosticar una enfermedad de los huesos. En realidad esto parece tener una lógica bastante clara, puesto que para los niños expresar sus síntomas es más difícil que para el adulto, además de que no es nada visible una enfermedad de esta índole. La literatura revisada también señala la dificultad en identificar enfermedades de esta naturaleza.

El retinoblastoma y el tumor de Wilms presentan una asociación en la edad de 0-4 años. Esto coincide con lo señalado al respecto en el marco de referencia. Los tumores hepáticos, tumores óseos y carcinomas tienen una frecuencia mayor a la esperada en porcentajes del más del 60% en la edad de 10-14 años. Otro hecho que apoya la teoría de que el cáncer tiene un aspecto hereditario es que el tumor de Wilms y los tumores óseos tienen antecedentes familiares con algún tipo de cáncer. Esto también refuerza la teoría existente sobre dichas enfermedades. Ahora bien, de los casos que presentaron una asociación con los antecedentes familiares se tiene que el neuroblastoma, el retinoblastoma, los tumores hepáticos, el rabdomiosarcoma, los tumores de las células germinales, los carcinomas y los mal definidos tienen este antecedente por el lado materno. Lo anterior habla de la importancia de la madre en la gestación de un ser humano, así como el carácter hereditario, el cual es un factor fundamental en las posibles enfermedades que se padecen en el transcurso de la vida. Pero es importante señalar que en casi la totalidad de casos, es de esperarse que la madre sea quien lleve a consulta al niño, y además tenga un mejor conocimiento de las enfermedades padecidas por sus familiares que por las del padre del paciente.

Por último, se quiere enfatizar en lo relevante de las dos metodologías utilizadas, así como del avance computacional, que hicieron que la presente fuera posible de realizarse.



CUADRO A
GRAFICAS ORIGINALES DE LAS SERIES DE TIEMPO ANALIZADAS.

VARIABLE	CATEGORIA	VAR. EST.	TENDENCIA	TRANSFORMACION	MODELO
SERIE GENERAL		NO	CRECIENTE	RAIZ CUADRADA	ARIMA(3,1,0)
SEXO	MASCULINO	NO	CRECIENTE	RAIZ CUADRADA	ARIMA(0,1,1)
	FEMENINO	NO	CRECIENTE	RAIZ CUADRADA	ARIMA(0,1,1)
TIPO DE CANCER	LEUCEMIA	NO	CRECIENTE	RAIZ CUADRADA	ARIMA(4,1,0)
	CANCER II-XII	NO	CRECIENTE	RAIZ CUADRADA	ARIMA(0,1,1)
FAMILIARES	SIN CANCER	NO	CRECIENTE	RAIZ CUADRADA	ARIMA(0,1,1)
	CON CANCER	NO	CRECIENTE	RAIZ CUADRADA	ARIMA(0,1,1)
EDAD DEL PACIENTE	0-4 AÑOS	NO	CRECIENTE	RAIZ CUADRADA	ARIMA(0,1,1)
	5-14 AÑOS	NO	CRECIENTE	RAIZ CUADRADA	ARIMA(0,1,1)
ESC. DEL PADRE	MENOS DE SEC.	NO	CRECIENTE	RAIZ CUADRADA	ARIMA(4,1,0)
	SEC. O MAS	NO	CRECIENTE	RAIZ CUADRADA	ARIMA(0,1,1)
ESC. DE LA MADRE	PRIMARIA O MENOS	NO	CRECIENTE	RAIZ CUADRADA	ARIMA(0,1,1)
	MAS DE PRIMARIA	NO	CRECIENTE	RAIZ CUADRADA	ARIMA(0,1,1)
EDAD MADRE NAC. HIJO	15-19 Y 30 AÑOS O MAS	NO	CRECIENTE	RAIZ CUADRADA	ARIMA(0,1,1)
	20-29 AÑOS	NO	CRECIENTE	RAIZ CUADRADA	ARIMA(0,1,1)

CUADRO B
MODELOS CORRESPONDIENTES A LAS SERIES DE TIEMPO ANALIZADAS.

VARIABLE	CATEGORIA	TIPO DE CANCER	PORCENTAJE EXCEDIDO EN EN CASO DE INDEPENDENCIA
SEXO	FEMENINO	TUMOR DE WILMS	44.1%
INSTITUCION DE SALUD	CENTRO MEDICO LA RAZA Y HOSPITAL GENERAL DE LA SECRETARIA DE SALUD	LINFOMAS	45%
		TUMOR DE WILMS	87.6%
		RABDOMIOSARCOMA	60.2%
		CARCINOMAS	47.8%
ESCOLARIDAD DE LA MADRE AL NACIMIENTO DEL HIJO	SECUNDARIA O MAS	TUMOR HEPATICO	69.40%
TIEMPO TRANSCURRIDO ENTRE EL INICIO DE SIGNOS Y SINTOMAS AL DIAGNOSTICO	MAS DE UN MES	TUMOR OSEO	40.2%
EDAD DEL NIÑO AL DIAGNOSTICO	0-4 AÑOS	RETINOBLASTOMA	91.7%
		TUMOR DE WILMS	76.1%
	10-14 AÑOS	TUMORES HEPATICOS	82.1%
		TUMORES OSEOS	67.1%
		CARCINOMAS	61.9%
FAMILIARES (C/S) CANCER	SI	TUMOR DE WILMS	62.5%
		TUMORES OSEOS	101.2%
RAMA DE LA CUAL PROCEDE EL CANCER	MATERNA	NEUROBLASTOMAS	97.2%
		RETINOBLASTOMAS	54%
		TUMORES HEPATICOS	146.5%
		RABDOMIOSARCOMA	54%
		TUMORES DE CELULAS GERMINALES	47.9%
		CARCINOMAS	146.5%
		MAL DEFINIDOS	46.5%

CUADRO C: VARIABLES QUE PRESENTARON ASOCIACION CON EL TIPO DE CANCER.

A N E X O

ANEXO

PROYECTO:

"FRECUENCIA DEL CANCER EN NIÑOS RESIDENTES EN EL D.F. DE 1981-1991".

INSTRUMENTO DE RECOLECCION DE DATOS FOLIO

NOMBRE DEL PACIENTE:

NUMERO DE AFILIACION:

EDAD (meses) : SEXO 1) MASC. 2)FEM.

DOMICILIO (Delegación):
(Sólo para pacientes del D.F.)

LUGAR DE RESIDENCIA:

INSTITUCION DE SALUD:

ANTECEDENTES PERSONALES NO PATOLOGICOS

LUGAR DE NACIMIENTO:

FECHA DE NACIMIENTO: MES AÑO

MES DE CONCEPCION:

PESO DEL NIÑO AL NACER:

ORDEN DEL NACIMIENTO:

EMBARAZO TERMINO 1) SI 2) NO

OCUPACION DEL PADRE:

OCUPACION DE LA MADRE:

ESCOLARIDAD DEL PADRE: ESCOLARIDAD DE LA MADRE:

1) ANALFABETA 2)SABE LEER Y ESCRIBIR 3) PRIMARIA INCOMPLETA
4)PRIMARIA COMPLETA 5)SECUNDARIA INCOMPLETA 6)SECUNDARIA COMPLETA
7)BACHILLERATO 8)TECNICO 9) OTROS

NIVEL SOCIOECONOMICO: 0)BAJO 1)MEDIO 2)ALTO

SALARIO DEL PADRE:

SALARIO DE LA MADRE:

SALARIO CONJUNTO (pesos):

EDAD DEL PADRE AL NACIMIENTO DEL HIJO:

EDAD DE LA MADRE AL NACIMIENTO DEL HIJO:

ANTECEDENTES RELACIONADOS CON EL PADECIMIENTO

FECHA DE INICIO DE SIGNOS Y SINTOMAS MES AÑO
FECHA DEL DIAGNOSTICO DE CANCER MES AÑO

TIEMPO TRANSCURRIDO ENTRE EL INICIO DE DE SINTOMAS Y LA FECHA DEL DIAGNOSTICO EN EL PACIENTE:

SIGNOS Y SINTOMAS INICIALES

SNC_____
01) CONVULSIONES
02) CEFALEA
04) VOMINTOS
08) DIPLOPIA
16) OTROS

CVS_____
01) INSUFICIENCIA CARDIACA
2)HT ARTERIAL.

RESP_____
1) DIFICULTAD RESPIRATORIA
2) ODINOFAGIA

URI_____
01) SANGRADO
02)NEFROMEGALIAS DISURIA
08) POLAQUIURIA

DGT_____
01) VOMITO
02) OCLUSION INTESTINAL
04) DOLOR ABDOMINAL
08) STD
16) TUMORACION ABDOMINAL
32) VISCEROMEGALLAS
64) OTRAS

ME_____
01) DIFICULTAD DEAMBULAR
02) MIALGIAS
04)ARTRALGIAS
08) CAIDA PROPIA ALTURA
16) AUMENTO DE VOLUMEN
32) MS:
64) MI

TI_____
01) SANGRADO PIEL Y MUCOSAS
02)EPISTAXIS
04) PALIDEZ DE TEGUMENTOS
08 OTROS

SI Y SG_____
01) FIEBRE
02) ATAQUE EDO. GENERAL
04) PERDIDA DE PESO
08) HIPOREXIA

16) ASTENIA Y ADINAMIA

OCULAR_____
01) EDEMA PALPEBRAL
02) LAGRIMEO
04)PROPTOSIS
08)SECRECION OCULAR

16) DIS. AGUD. VISUAL
32) NISTAGMUS
64) MASA OCUPATIVA

ENDO_____
01) CAMBIOS DE VOZ
02) AUMENTO DE VELLO CORPORAL
04)CRECIMIENTO DEL PENE
08) FALTA CARAC. SEX.

TIPO DE CANCER____

ESTUDIO HISTOPATOLOGICO _____ 1) SI 0) NO

CLASIFICACION INTERNACIONAL DE LAS ENFERMEDADES_____

CLASIFICACION DETALLADA_____

EDAD DEL PACIENTE AL DIAGNOSTICO DE CANCER:_____

ANTECEDENTES HEREDOFAMILIARES.

FAMILIARES CON ACNCER: _____ 1) SI 0) NO

PARENTESCO_____

TIPO DE CANCER EN EL FAMILIAR_____

ANTECEDENTES PERSONALES PATOLOGICOS _____
(Enfermedades exantemáticas)

01) sarampión 02) varicela 04) rubeola 08)hepatitis
16)escarlatina 32) otra

EL PACIENTE TUVO INTERNAMIENTO POR ALGUN PROCESO INFECCIOSO DURANTE SU PRIMER AÑO_____
1) SI 0) NO

VIVE EL PACIENTE _____ 1)SI 0)NO

EL PACIENTE TIENE ALGUNA MALFORMACION_____ 1)SI 0)NO

DONDE LA MALFORMACION_____

EL PACIENTE TIENE OTRO CANCER_____ 1)SI 0)NO

DIAGNOSTICO DEL SEGUNDO CANCER_____

PROGRAMA PARA OBTENER LAS FRECUENCIAS DE LAS SERIES.

```
selec a
        use c:\tesis\datos\df81_91.dbf
selec b
        use c:\tesis\series.dbf
selec c
        use c:\tesis\series2.dbf
k=202
i=0
do while i<=8
        i=i+1
                k=k+1
                        if k<=9
                                vari="VAR0"+str(k,1,0)
                        else
                                vari="VAR"+str(k,3,0)
                        endif

                n=0
                no1=81
                cond0=120
                cond1="ESCOLARPA"
                cond2=180

                if i<=9
                        cond2=str(i,1,0)
                else
                        cond2=str(i,2,0)
                endif

        do while ano1<=91
                año=str(ano1,2,0)
                mes="0"
                mes1=0

                do while mes1<=11
                        mes1=mes1+1
                        if mes1 <=9
                                mes="0"+ str(mes1,1,0)
                        else
                                mes=str(mes1,2,0)
                        endif
selec a

        go top

        count to a for ading=ano .and. fechadedia=mes .and. &cond1=cond2
                fecha="F"+ano1+"_"+mes
                if ano1=91
                        select c
                else
                        select b
                endif

                replace &fecha with a for variable=vari

                        enddo
                        ano1=ano1+1
        enddo
enddo
close all
```

CASOS MENSUALES DE CANCER PEDIATRICO EN EL D.F.

	AÑOS										
MESES	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991
ENERO	2	10	7	8	6	1	7	11	15	19	13
FEBRERO	4	4	11	5	8	11	5	8	10	7	13
MARZO	3	7	8	12	9	6	10	8	12	9	11
ABRIL	3	5	7	13	8	7	5	14	12	14	20
MAYO	8	7	6	8	5	5	6	13	8	9	17
JUNIO	6	7	13	5	12	11	8	10	17	11	15
JULIO	2	6	4	10	15	3	6	5	9	12	19
AGOSTO	3	7	10	3	9	5	11	11	20	14	18
SEPTIEMBRE	11	5	4	7	4	12	7	14	10	10	18
OCTUBRE	4	7	6	8	18	6	9	10	8	17	14
NOVIEMBRE	3	4	7	5	8	7	14	13	8	14	12
DICIEMBRE	5	6	9	6	7	7	9	12	10	10	9

POBLACION MEDIA PROYECTADA PARA DISTRITO FEDERAL
HIPOTESIS DE FECUNDIDAD PROGRAMATICA

	HOMBRES	MUJERES	AMBOS SEXOS
AÑO	0-11	0-11	0-11
1980	1455706	1415639	2871346
1981	1457195	1412874	2870069
1982	1453113	1405689	2858802
1983	1444030	1394481	2838511
1984	1430511	137950	2810161
1985	1413112	1361603	2774715
1986	1392382	1340755	2733136
1987	1368863	1317523	2686387
1988	1343092	1292335	2635427
1989	1315597	1265622	2581220
1990	1286900	1237823	2524723
1991	1257425	1209285	2466710

POBLACION MEDIA MASCULINA PROYECTADA PARA DISTRITO FEDERAL
HIPOTESIS DE FECUNDIDAD PROGRAMATICA

EDAD	1980	1985	1990	1995
0-4	616443	578696	509223	454645
5-9	617250	588779	550715	479559
10-14	534960	606202	575485	536898

POBLACION MEDIA FEMENINA PROYECTADA PARA DISTRITO FEDERAL
HIPOTESIS DE FECUNDIDAD PROGRAMATICA

EDAD	1980	1985	1990	1995
0-4	592149	557433	489746	436609
5-9	602547	565867	530334	460558
10-14	538823	592273	553117	516829

POBLACION MEDIA TOTAL PROYECTADA PARA DISTRITO FEDERAL
HIPOTESIS DE FECUNDIDAD PROGRAMATICA

EDAD	1980	1985	1990	1995
0-4	1208592	1136129	998969	891254
5-9	1219797	1154646	1081049	940117
10-14	1073783	1198475	1128602	1053727

POBLACION MEDIA PROYECTADA PARA DISTRITO FEDERAL
HIPOTESIS DE FECUNDIDAD ALTERNATIVA

AÑO	HOMBRES 0-11	MUJERES 0-11	AMBOS SEXOS 0-11
1980	1455706	1415639	2871346
1981	1454448	1410288	2864736
1982	1448456	1401287	2849742
1983	1438890	1389608	2828498
1984	1426799	1376122	2802921
1985	1413112	1361603	2774715
1986	1398646	1346726	2745373
1987	1384103	1332067	2716170
1988	1370067	1318104	2688171
1989	1357010	1305215	2662225
1990	1345288	1293681	2638969
1991	1335236	1283769	2619005

POBLACION MEDIA MASCULINA PROYECTADA PARA DISTRITO FEDERAL
HIPOTESIS DE FECUNDIDAD ALTERNATIVA

EDAD	1980	1985	1990	1995
0-4	616443	578696	569291	562879
5-9	617250	588779	550135	535488
10-14	534960	606202	574913	536194

POBLACION MEDIA FEMENINA PROYECTADA PARA DISTRITO FEDERAL
HIPOTESIS DE FECUNDIDAD ALTERNATIVA

EDAD	1980	1985	1990	1995
0-4	592149	557433	547385	540521
5-9	602547	56867	529658	514049
10-14	538823	592273	552447	516007

POBLACION MEDIA TOTAL PROYECTADA PARA DISTRITO FEDERAL
HIPOTESIS DE FECUNDIDAD ALTERNATIVA

EDAD	1980	1985	1990	1995
0-4	1208592	1136129	1116676	1103400
5-9	1219797	1154646	1079793	1049537
10-14	1073783	1198475	1127360	1052201

GLOSARIO

Andrógeno: Sustancia que induce la masculinización.

Bazo: Organo intraabdominal situado en la parte más elevada de la porción izquierda de la cavidad abdominal.

Benceno: la estructura básica de los compuestos aromáticos, hidrocarburo muy tóxico derivado del aceite liviano de alquitrán; usado como solvente.

Carcinogénico: cancerígeno; que causa cáncer.

Cigoto: Individuo resultante de la unión de dos gametos.

Clona: Uno o un grupo de células, organismos o plantas, genéticamente idénticas derivadas por reproducción vegetativa a partir de un único progenitor.

Cromosomas: Son cuerpos microscópicos existentes en el núcleo de las células, constituidos por una sustancia especial de naturaleza proteica llamada cromatina. Son individualidades morfológicas celulares bien definidas, cuyo número es fijo y constante para todas las células de una especie animal (y vegetal) y, en cambio, varía de una especie a otra.

Diferenciación: Proceso de adquisición de caracteres individuales distintos.

Distensión: estiramiento violento de los tejidos y partes ligamentosas de una articulación.

Epífisis: Extremo de un hueso largo, desarrollado a partir de un centro secundario de osificación.

Epitelio: Revestimiento de las superficies internas y externas del cuerpo que incluye el revestimiento de los vasos y otras pequeñas cavidades, consiste de células unidas por sustancias intercelulares que actúan de cemento y se clasifica según el tipo de células y número de capas que lo formen.

Etiología: parte de la medicina que tiene por objeto el estudio de las causas de las enfermedades.

Fisiológico: Con este adjetivo (que procede de fisiología) se indica todo lo que es normal en el funcionamiento específico de los diversos órganos y tejidos que constituyen el organismo de cada ser viviente (humano, animal, vegetal). Por eso fisiológico (normal) se usa habitualmente en medicina en contraposición de patológico que quiere decir morboso, anormal.

Ganglios linfáticos: son la fuente principal de linfocitos de la sangre periférica, y como parte del sistema reticuloendolelial, sirven como mecanismos de defensa para eliminar los agentes nocivos, como bacterias y toxinas. Probablemente desempeñen una función en la producción de anticuerpos.

Gemelos monocigotos: los dos descendientes desarrollados a partir de un cigoto u óvulo fecundado, y que por lo tanto tienen genomas idénticos.

Glándulas: Son órganos especiales que tienen la función específica de elaborar con material "tosco" procedente de la sangre unas sustancias determinadas, algunas de las cuales son nocivas (tóxicas) y otras útiles para el organismo.

Hormona andrógena: Hormona masculina.

Ictericia: Coloración amarilla de la piel, mucosas y secreciones, debido a la presencia de pigmentos biliares en la sangre.

Infiltración: Acción de pasar a una sustancia, célula o tejido. El gas, líquido o materia disuelta que ha entrado en cualquier sustancia, célula o tejido.

Inmunidad: Insensibilidad relativa de una persona o animal para una infección por microorganismos patológicos o para los efectos nocivos de ciertas sustancias antigénicas.

Inmunodeficiencia: Proceso en el que existe una deficiencia de la respuesta inmunitaria.

Intersticial: situado en los interespacios de un tejido.

Intersticios: tejido intersticial.

Linfoide: que es semejante al tejido del sistema linfático.

Linforreticular: Perteneciente o relativo a las células reticuloendoteliales de los ganglios linfáticos.

Médula ósea: Es una sustancia blanda, gelatinosa, amarilla o rojiza, contenida en el interior del conducto medular situada en la zona central de los huesos largos o en la cavidad interna de los otros tipos de huesos. Posee la importantísima función específica de la reproducción de glóbulos rojos y de ciertas formas de glóbulos blancos de la sangre.

Médula suprarrenal: parte blanda más interna, de color pardo rojizo de la glándula suprarrenal; sintetiza, almacena y libera catecolaminas.

Metáfisis: parte más amplia a nivel de la extremidad del tallo o diáfasis de un hueso largo, adyacente al disco epifisario. Durante el desarrollo contiene la zona de crecimiento y está constituida por hueso esponjoso.

Metástasis: La palabra "metástasis" quiere decir "hacia, estar después, más allá". Tiene raíces griegas, y el prefijo meta indica un "cambio" (como en metabolismo y metamorfosis). Por lo tanto, su significado es separación de origen. Es la propagación del tumor maligno desde un sitio primario a puntos secundarios situados a distancia.

Monocigoto: Relativo o derivado de un sólo cigoto.

Morbilidad: situación o estado de enfermedad; incidencia o prevalencia de una enfermedad o de todas las enfermedades de una población.

Nasofaringe: Porción de la faringe encima del velo del paladar.

Orbita: Cada una de las cavidades óseas debajo de la frente, que contienen el globo ocular y los tejidos blandos que lo rodean, formadas por el frontal, esfenoides, etmoides, lagrimal, maxilar, palatino y cigomático.

Proliferar: crecer por reproducción de células semejantes.

Radiación: Emisión de rayos luminosos, calorías, químicos, etc.

Radioterapia: Tratamiento de las enfermedades por toda clase de rayos.

Retina: Membrana, la más interna de las tres que forman el globo ocular.

Síndrome cromosómico: Patología determinada por alteraciones de un sólo cigoto.

Sistema nervioso simpático: es el sistema nervioso de la vida orgánica construido por dos cordones nerviosos extendidos a cada lado de la columna vertebral. Este sistema inerva los músculos de fibra lisa y el corazón, y tiene una relación íntima con las secreciones, movimientos vasculares y procesos tróficos.

Tejido linfoide: redecilla de tejido reticular cuyos interespacios contienen linfocitos; el tejido linfoide puede ser difuso o densamente agregado, como sucede en los nódulos y los ganglios linfáticos.

Teratoma: neoplasia verdadera formada por tipos distintos de tejido, ninguno de los cuales es nativo del área en la cual aparece.

Tórax: Pecho, Parte del cuerpo ubicado entre el cuello y el diafragma respiratorio encerrado por las costillas.

BIBLIOGRAFIA

1. Chatfield, C. 1989. THE ANALYSIS OF TIME SERIES. (4a. ed.) London: Chapman and Hall.

2. DICCIONARIO DE CIENCIAS MEDICAS ILUSTRADO. 1993. Traducido por Editorial Médica Panamericana S.A. Aires: Editorial Médica Panamericana.

3. DICCIONARIO ENCICLOPEDICO ILUSTRADO DE MEDICINA. 1992. (27a. ed.) España: Interamericana McGraw-Hill.

4. EPIDEMIOLOGY OF CANCER IN CHILDREN Vol. 7. 1985. The Johns Hopkins University School of Hygiene and Public Healt .

5. Everitt, B. S. 1992.THE ANALYSIS OF CONTINGENCY TABLES. London: Chapman and Hall.

6. González, María del Carmen.1990. MODELOS DE DECISION CON PROCESOS ESTOCASTICOS II. México: ENEP Acatlán.

7. ONCOLOGIA PEDIATRICA, Vol. 48 No.3, 1990. México: Anales Nestlé.

8. Proyecciones de la Población de México y de las Entidades Federativas:1980-2010. 1990. Aguas Calientes: Instituto Nacional de Estadística, Geografía e Informática y el Consejo Nacional de Población.

9. Ruiz, Silvia. 1982. MODELOS LINEALES GENERALIZADOS Y SU RELACION CON GLIM. Tesis inédita para licenciatura. Universidad Nacional Autónoma de México.

10. Segatore, I. 1975. DICCIONARIO MEDICO. (5a. ed.) Barcelona: Terde.

11. Upton, Graham J. G. 1978. THE ANALYSIS OF CROSS-TABULATE DATA. Chichester: Jhon Wiley and Sons.

12. Wei,Willian W. S. 1990. TIME SERIES ANALYSIS. Univariate and Multivariate Methods. California: Addison - Wesley.