

7



# UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Escuela Nacional de Estudios Profesionales

"ACATLAN"

## MATEMATICAS APLICADAS Y COMPUTACION

NEOPLASIAS MALIGNAS EN NIÑOS RESIDENTES EN LA REPUBLICA MEXICANA, ASOCIADOS CON LA ESCOLARIDAD Y OCUPACION DE SUS PADRES POR MEDIO DE ANALISIS DE CORRESPONDENCIAS (1980 - 1992)

T E S I S

QUE PARA OBTENER EL TITULO DE LICENCIADO EN MATEMATICAS APLICADAS Y COMPUTACION

PRESENTA

JOSE JUAN CASTILLO PEREZ

299424



Acatlán, Estado de México



ASESOR DR. JUAN MANUEL JUNCO CARRERA



Universidad Nacional  
Autónoma de México



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

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

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

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

A todos aquellos párvulos que transitan por un centro de enseñanza, que comparten el mismo espacio cada día con sus compañeros y profesores.

*“Sé ignorante para que puedas seguir aprendiendo”.*

*“El límite de la ignorancia cuando el conocimiento tiende a infinito es uno mismo”*

*Jonás de Rabel*

Y a todas aquellas personas quienes confiaron en mí, no cabría en una hoja tantos y tantos nombres, dado que un agradecimiento debe ser pequeño, les dedicó este trabajo y ...

*Amo el canto del zenzontle  
Pájaro de cuatrocientas voces.  
Amo el color del jade y el enervante  
Perfume de las flores, pero  
Amo más, a mi hermano el hombre.*

Nezahualcoyotl.

## **AGRADECIMIENTOS.**

En especial al Mto. Arturo Fajardo Gutiérrez quien proporciono la base de datos de niños con cáncer de la población derechohabiente del IMSS para poder realizar este trabajo, además de ser un vínculo importante en mis pasos. A la Dra. en Ciencias Araceli Pérez Díaz quien me demostró la solución diversos problemas; sean complejos o sencillos, así como al Dr. Juan Manuel Junco Carrera quien me dirigió en el proceso de investigación.

INTRODUCCIÓN .....	ix-xiv
1. PROCESO, EVOLUCIÓN Y TRATAMIENTOS .....	1-11
1.1 Estudio descriptivo de neoplasias malignas en niños residentes en la República Mexicana asociados con escolaridad y ocupación de sus Padres .....	1-3
1.2 Metástasis del cáncer .....	4-7
1.2.1 Factores de riesgo para desarrollar una neoplasia maligna.....	5-7
1.3 Tratamientos para el cáncer pediátrico .....	8-11
2. ANÁLISIS DE DATOS CATEGÓRICOS .....	13-18
2.1 Tablas de contingencia .....	13-14
2.2 Análisis discriminante .....	15-16
2.3 Análisis canónico .....	16-18
3. ANÁLISIS DE CORRESPONDENCIAS .....	19-36
3.1 Fundamentos .....	19-22
3.2 Distancia ji-cuadrado .....	22-24
3.3 Representación de las poblaciones .....	24-26
3.4 Representación simultánea .....	27-28
3.5 Interpretación del análisis .....	29-33
3.6 Prueba de independencia total .....	34
3.7 Descomposición aditiva de la ji-cuadrado .....	34
4. RESULTADOS .....	37-43
5. DISCUSIÓN .....	45
5.1 Direcciones Futuras .....	46
CONCLUSIONES .....	47-48
REFERENCIAS .....	49-51
ANEXOS .....	53-89

## INTRODUCCIÓN.

En todas las ciencias experimentales una tarea fundamental es la exploración, descripción, clasificación y análisis de los objetos y fenómenos para tratar de comprender su comportamiento y explicarlo si es posible<sup>1</sup>.

La estadística, ha sido en los últimos años la herramienta básica de los investigadores de diversos campos, haciendo uso tanto de los conocimientos matemáticos como de los conocimientos informáticos. Hay dos tipos de estadística, la estadística descriptiva, que sólo describe los datos y los presenta en forma gráfica y la estadística inferencial, ésta proporciona inferencias con respecto a una población a partir de la información contenida en una muestra<sup>2</sup>.

Existen diversos métodos de análisis y representación de datos que son una buena herramienta para alcanzar el objetivo de los investigadores, tales métodos se localizan en una rama de las matemáticas y la estadística conocida como análisis multivariado<sup>3</sup>.

El análisis multivariado es la rama de la estadística y del análisis de datos, que estudia, interpreta y elabora el material estadístico que involucra más de dos variables dependientes o independientes; que pueden ser de tipo cualitativo o cuantitativo. El análisis multivariado posee una gran gama de métodos de análisis y exploración de datos<sup>4, 1</sup>.

Dentro de esa gama de métodos se encuentra el análisis de correspondencias, que se ha empleado en análisis estadísticos desde que se introdujo al contexto multivariado por J. P. Benzècri en 1965. Es útil en los trabajos exploratorios donde no existen hipótesis previas del comportamiento de la población, tanto en las vertientes correlacionales como experimentales<sup>31</sup>.

El Análisis de Correspondencias (AC) es un análisis sobre una distancia ponderada con respecto a la puntuación real, llamada ji-cuadrado, cuya principal propiedad es la posibilidad de una representación geométrica simultánea de las variables (variable 1, variable 2). Esta metodología es apropiada para representar datos cualitativos organizados en una tabla de contingencia y de datos de presencia-ausencia<sup>5</sup>. En los estudios médicos y epidemiológicos se emplean frecuentemente datos de este tipo, donde la respuesta es una variable nominal organizada en una tabla de frecuencias. En los últimos años los clínicos han empleado métodos multivariados en sus análisis estadísticos, básicamente: regresión lineal múltiple, regresión logística múltiple, análisis discriminante múltiple y regresión de Cox<sup>14</sup>.

El análisis de correspondencias, técnica que se emplea ampliamente en las ciencias psicosociales lleva a considerar el estudio de la interacción como estudio de las estructuras de relación entre los elementos de un sistema. La principal ventaja del AC es la generalidad y versatilidad de sus aplicaciones como técnica de representación de la estructura geométrica simultánea del multiespacio de relación entre individuos y variables y su capacidad de tratar todo tipo de datos, tanto cualitativos como cuantitativos<sup>32</sup>.

El reciente uso de la computadora ha tenido un gran impacto sobre las ciencias de la salud en general y en el análisis bioestadístico, lo que ha llevado a ser una herramienta indispensable para el análisis de las grandes bases de datos de los clínicos<sup>6</sup>. Las recientes computadoras tienen una

capacidad computacional considerablemente alta y una capacidad gráfica de alta calidad. Los software con capacidades estadísticas tienen la habilidad de manejar grandes cantidades de datos y producir gráficas sofisticadas<sup>15</sup>.

En la actualidad existen muchos paquetes estadísticos que realizan análisis multivariado, dentro de esa gran gama, se encuentran GLIM, BMDP, SAS, STATA, SYSTAT, UNISTAT, STATGRAPHICS, STATISTICA, SPSS por mencionar solo algunos. El SPSS es un sistema que cubre las características anteriores, fue desarrollado para los científicos sociales, de ahí su nombre Statistical Package for Social Science<sup>33</sup>.

El cáncer es una de las enfermedades cuyas causas son parcialmente desconocidas, por consiguiente no hay medicamentos y tratamientos apropiados hasta este momento que pueda detener su crecimiento. El cáncer o neoplasia maligna, es una enfermedad rara y multifactorial, en la cual las células del huésped se transforman de normales en malignas como resultado de interacciones complejas de estímulos externos (medio ambiente) y genes que predisponen o suprimen el cáncer<sup>16</sup>.

El niño puede ser más susceptible a la exposición al medio ambiente que los adultos, porque sus sistemas están en desarrollo. El daño que se les puede causar puede ser muy importante, permanente o absoluto<sup>17</sup>. Las neoplasias malignas en pediatría han venido representando un capítulo de suma importancia en éstos últimos años, jugando un papel importante en la epidemiología de morbi-mortalidad en México<sup>18</sup>.

Se han realizado estudios no muy profundos acerca de neoplasias malignas que atacan a niños menores de 15 años de edad<sup>19</sup>. En México, se han realizado análisis de mortalidad con la intención de dar al cáncer el lugar que le corresponde como causa de muerte en la población pediátrica<sup>20</sup>. Se piensa que las neoplasias malignas en los niños se presentan en forma diferente según el estrato social como factor de asociación con diferentes tipos de patologías<sup>21</sup>. Para establecer estos estratos se emplean variables como escolaridad, ocupación de los padres, salario y condiciones de la vivienda. Por otro lado, la escolaridad y ocupación de los padres juega un papel importante en la epidemiología del cáncer pediátrico<sup>22</sup>, por lo que en este trabajo se enfoca éstas dos variables para establecer una asociación con el cáncer en niños.

Aún, no se ha establecido con claridad si las neoplasias malignas en el niño son diferentes según el nivel de escolaridad y ocupación de sus padres, clase social, factores genéticos, exposición al medio ambiente o si es un aspecto no aleatorio. Sin embargo, existen interrogantes que deben ser contestadas. ¿Las neoplasias malignas en los niños de la República Mexicana están relacionadas con el grado escolar y ocupación de sus padres?. Sólo en el futuro se podrá saber con certeza el papel que juega la escolaridad y ocupación de los padres con los diferentes tipos de neoplasias en pediatría que se desarrollan actualmente. Al conocer esta relación sería de punto de partida para programas educativos encaminados a identificar estos factores de riesgo, su control y prevención.

El cáncer en niños, patología aún poco estudiada en México, representa actualmente un problema serio de morbi-mortalidad en este grupo, sin existir hasta el momento publicaciones disponibles o estudios serios que determinen el nivel de educación de los padres prevalente en ésta población y su labor ocupacional de los mismos. Por lo tanto, se considera de interés el investigar el tipo de neoplasia maligna en niños asociada con la escolaridad y ocupación de sus padres.

Lo que se pretende realizar en este estudio descriptivo retrospectivo es determinar la relación existente de neoplasias malignas en niños según la escolaridad y ocupación de sus padres como factores de riesgo. Además, introducir una metodología que se ha empleado ampliamente en otras áreas del conocimiento y que provee información más detallada que las tablas habituales empleadas en Epidemiología para poder explorar asociaciones complejas entre variables.



## **Reseña histórica de las matemáticas**

Desde que el hombre concibió la idea del número como la relación de magnitud y forma de los fenómenos u objetos que le rodean, se desarrolla el pensamiento abstracto y por ende de las matemáticas. Según la historia, es en Egipto y en Babilonia donde se desarrolla el concepto de álgebra, estas civilizaciones expresaban sus conceptos e ideas mediante números y ecuaciones. Posteriormente, otras civilizaciones desarrollan más a profundidad este concepto, es en Alejandria donde se encuentran soluciones a sistemas de ecuaciones indeterminadas, con los matemáticos Herón y Diofante. En Arabia, el matemático Al-Jwarizmi escribió uno de los primeros libros árabes de álgebra con la teoría fundamental de ecuaciones<sup>29</sup>. Muchos otros matemáticos árabes desarrollaron métodos capaces de describir cualquier potencia de la incógnita  $x$ . En el siglo XII, varios matemáticos italianos desarrollan métodos para dar solución a ecuaciones de tercer y cuarto grado.

Un avance importante para el álgebra fue la introducción, en el siglo XVI, de símbolos para las incógnitas y para las operaciones y potencias algebraicas, así como el fundamento de la teoría de ecuaciones. En 1799 el matemático alemán Carl Friedrich Gauss demostró que toda ecuación polinómica tiene al menos una raíz en el plano complejo.

Importantes contribuciones surgen del trabajo de Gauss, en el siglo XIX, los matemáticos franceses Galois y Augustin Cauchy, el británico Arthur Cayley y los noruegos Niels Abel y Sophus Lie realizan estudios acerca de las propiedades de los sistemas numéricos, como permutaciones y combinaciones de las raíces de polinomios. El matemático y astrónomo irlandés William Rowan Hamilton desarrolló la aritmética de los números complejos y surgen las cuaternas, de aquí surge el álgebra abstracta y su aplicabilidad a la física. En tal sentido, el fundamento de las matemáticas es el álgebra (reducción de la palabra árabe al-yabr), y ha seguido evolucionando y aplicado en todas las ramas de las matemáticas y en muchas otras ciencias.

Actualmente, gracias a las computadoras los cálculos de sistemas de ecuaciones complejas se realizan en segundos, lo que da una nueva formulación para procesos que antes no se podían resolver.

## **Cómo convertirse en un matemático aplicado**

Este trabajo va dirigido a aquellos lectores (matemáticos aplicados) que desconocen el rumbo de ésta área, tan antigua como el pensamiento humano, y que tienen la necesidad de saber que las matemáticas tienen aplicación. Para tener una visión muy amplia de las matemáticas aplicadas se les presenta este manuscrito para ejemplificar una experiencia y pueda ser un escalón para proyectar sus inquietudes.

Se puede considerar este escrito como un escalón hacia aplicaciones más complejas. Un buen matemático aplicado debe tener el deseo de aprender y no considerar o dar por hecho que lo sabe todo, para que de esa forma pueda aplicar sus conocimientos a áreas, que en nuestro país están "en pañales", desarrollar su intelecto y abrir sus mentes para tener una visión más amplia de los fenómenos u objetos que nos rodean. Además, emplear las herramientas informáticas para crear nuevas formas de aprendizaje o metodologías que hagan más comprensible un fenómeno,

recuerden que el pensamiento matemático no tiene fronteras.

*“La esencia de las matemáticas reside en su libertad”*  
George Cantor

Finalmente, este trabajo pretende sembrar esa inquietud y curiosidad a aquellos jóvenes que tienen ánimos de dirigir, modelar sus ideas en una ecuación matemática. Espero cumplir ese objetivo.

### **La organización de este trabajo**

Este escrito está compuesto por cinco capítulos, cada uno de ellos proporciona teoría y metodología para plantear científicamente un problema e interpretarlo mediante la descripción los resultados de un problema o fenómeno, en este caso; cáncer en niños.

El capítulo 1, *Proceso, Evolución y Tratamientos*, está dividido en tres secciones. En la primera sección el lector podrá tener una visión amplia de la estructura básica de las componentes que conforman un protocolo de investigación, cómo se describe la problemática, cómo se describen las variables de estudio y sus definiciones, así como los recursos que se emplearon para su elaboración. La segunda sección, describe de forma amplia qué es el cáncer o neoplasia maligna su clasificación y proceso de metastasión, factores de riesgo que pueden causar un cáncer y agentes bacteriológicos que pueden interactuar con el ambiente para el desarrollo de una neoplasia. La última sección trata de los diferentes tratamientos convencionales que son empleados para combatir esta enfermedad, sus ventajas y desventajas, también terapias consideradas complementarias o alternativas, que actualmente, hasta la edición de este trabajo, son empleadas a nivel mundial.

El capítulo 2, *Análisis de datos categóricos*, muestra el uso de diferentes técnicas estadísticas multivariadas para analizar datos cualitativos organizados en una tabla de contingencia, según el objetivo de la investigación a realizar. Tres son las técnicas que se describen brevemente, son diferentes puntos de vista para abordar un problema, pero se debe seleccionar uno que se ajuste más a los objetivos de una investigación o análisis.

En el capítulo 3, *Análisis de Correspondencias*, se estudia el fundamento matemático de esta técnica exploratoria-descriptiva multivariada, en un sentido básico para su comprensión, ya que existen, dentro de esta metodología, dos tipos de este análisis. Uno sobre dos dimensiones, 2-D, y el otro sobre  $n$  dimensiones, N-D. El análisis de correspondencias (AC) implica el conocimiento de álgebra lineal, aspectos de estadístico y probabilidad, para con ello dar una interpretación de los resultados en forma congruente. Se mencionan sus ventajas y propiedades para aplicarlas en diferentes áreas del conocimiento, así como una nota histórica.

El capítulo 4, *Resultados*, trata acerca de los resultados realizados mediante la aplicación del análisis de correspondencias mediante el SPSS 8.0. Se describe, primeramente la escolaridad materna, paso a paso, interpretando los resultados arrojados por este paquete estadístico, para que el lector pueda comprender los resultados finales de este trabajo. Posteriormente, se dan solo resultados de las variables analizadas y que se consideran de interés.

En el capítulo 5, *Discusión*, se pretende comparar los resultados de este trabajo con algunos resultados que se han obtenido de investigaciones previas, comentando el objetivo del presente escrito. Además, cuenta con una subsección, cuyo propósito es dar al lector la noción del rumbo que debe tomar la investigación y en especial de las matemáticas aplicadas.

Las conclusiones son una inferencia de los resultados obtenidos, un cuestionamiento de ciertas formulaciones y si se cumplió lo que se pretendía al realizar este trabajo de investigación. El manuscrito también contiene referencias bibliográficas empleadas para la realización de este trabajo, en ellas se encuentran libros cuya metodología corresponde al tipo de técnica estadística empleada y otras fuentes de información consultada para complementar el trabajo realizado. Además contiene dos anexos: el Anexo A contiene las gráficas realizadas en MATLAB 5.0 y las tablas, por cada variable analizada, generadas por el SPSS 8.0. En el Anexo B se presenta la tabla de clasificación de Birch y Marsden para el cáncer en niños. Y el Anexo C se encuentra el cuestionario aplicado a los padres de los pacientes pediátricos con cáncer así como el manual de codificación de los datos.

## **1. PROCESO, EVOLUCIÓN Y TRATAMIENTO**

El hombre ha tratado de encontrar la causa de ciertos fenómenos, las enfermedades son uno de ellos. Para su estudio, ha tratado de describir los síntomas y características de las enfermedades, llevando a cabo el registro de su incidencia sobre una población de enfermos.

Este capítulo está dividido en tres secciones, la primera " ESTUDIO DE NEOPLASIAS MALIGNAS EN NIÑOS RESIDENTES DE LA REP. MEX. ASOCIADOS CON ESCOLARIDAD Y OCUPACIÓN DE SUS PADRES", pretende dar al lector una noción del proceso de la investigación realizada. La siguiente sección "Metástasis del cáncer", da una idea de la evolución del cáncer en general y sus riesgos. La última sección " Tratamientos para el cáncer pediátrico" provee de información general de las terapias oncológicas ortodoxas y no ortodoxas aplicadas actualmente a nivel mundial.

### **1.1 Estudio de neoplasias malignas en niños residentes de la República Mexicana asociados con escolaridad y ocupación de sus padres.**

El Instituto Mexicano del Seguro Social (IMSS) ha enrolado anualmente en promedio 2000 pacientes menores de 15 años con cáncer de la República Mexicana, en coordinación con las principales unidades hospitalarias del D.F. (Centro Médico La Raza "CMR", Hospital 20 de Noviembre, Hospital General de Secretaría de Salud "HGSS", Hospital Infantil de México "InfMex"; Instituto Nacional de Pediatría "INP").

Se considero para su estudio los expedientes de 5017 niños y niñas menores de 15 años, todos con diagnóstico de cáncer confirmado por un estudio cito morfológico de médula ósea o histopatológico. Para su análisis se utilizo la clasificación internacional para cáncer en niños propuesta por Birch y Marsden (**Anexo B**). Las neoplasias malignas estudiadas son: leucemias, linfomas, retinoblastomas, tumores del sistema nervioso central (SNC), tumores óseos, tumores renales, tumores de células germinales, tumores del sistema nervioso simpático (SNS), Tumores mal definidos, tumores hepáticos, tumores tejidos blandos y carcinomas; en ese orden.

No se tomo en cuenta los expedientes de niños con diagnóstico de cáncer que no son derechohabientes del IMSS, niños con diagnóstico tratados inicialmente en otra institución o en años previos al periodo de estudio, niños que murieron sin especificar el tipo de tumor o cáncer, ni pacientes pediátricos en los que la información recopilada fue incompleta.

Para la recolección de datos se creo una cédula (cuestionario **Anexo C**), con la cual se entrevisto a los padres de los niños con cáncer y se reviso los expedientes clínicos de los pacientes. Previo a la recolección de datos se adiestró al personal y se evaluó su estandarización con una prueba piloto, sé probo en tres ocasiones para su validación.

Las variables que se estudiaron fueron: neoplasias malignas, escolaridad y ocupación de los padres, edad del paciente, sexo, lugar de residencia e institución de salud. La variable ocupación de los padres se clasifico acorde a la Organización Internacional Uniforme de Ocupaciones <sup>23</sup>.

### **Variables de estudio:**

**Dependiente:**           cáncer (neoplasias malignas).

**Independientes:**

Escolaridad.  
Ocupación.  
Edad pediátrica.  
Sexo.  
Lugar de residencia.  
Institución de salud.

### **Definición conceptual de variables.**

<i>Cáncer</i>	Células que se reproducen de manera continua y a una velocidad sin control, formando tejido excesivo <sup>7</sup> .
<i>Escolaridad</i>	Grado máximo de aprendizaje obtenido en una institución de enseñanza (escuelas primaria, secundaria, preparatoria, institutos, etc.) por los padres o tutor.
<i>Ocupación</i>	Acción de emplear, permanecer en un oficio, dignidad.
<i>Edad pediátrica</i>	Periodo de la vida del ser humano que comprende desde que nace hasta los 18 años de edad.
<i>Sexo</i>	Diferencia física y constitutiva del hombre y la mujer.
<i>Lugar de residencia</i>	Localidad geográfica en donde vive el individuo y su familia.

VARIABLE	TIPO	DEFINICIÓN OPERACIONAL	ESCALA DE MEDICIÓN
cáncer	cualitativa	Tejido excesivo corroborado histopatológicamente	nominal
escolaridad	cualitativa	Se interrogó el grado escolar máximo de uno o ambos padres o el tutor y se estratifico como: analfabeta, sabe leer y escribir, primaria completa o incompleta, secundaria completa o incompleta, bachillerato completo o incompleto, licenciatura completa o incompleta, técnico u otros.	ordinal
ocupación	cualitativa	Se interrogó a los padres de los pacientes y se estratificaron en nueve grupos <sup>23</sup> .	ordinal
edad pediátrica	cualitativa	Pacientes de 0 a 15 años y se estratifico según los siguientes grupos: <1, 1-4, 5-9, 10 – 15 años.	ordinal
sexo	cualitativa	Se estratificara como 1) hombre, 0) mujer.	nominal
lugar de residencia	cualitativa	Se investigó el lugar de nacimiento del niño y el lugar en donde ha vivido en el último año como lugar de residencia (32 estados de la República Mexicana)	nominal

La técnica estadística empleada fue análisis de correspondencias, dado que la asociación (o interacción) se considera como el estudio de las estructuras de relación entre los sujetos (o elementos) de la población pediátrica. A su vez, hablar de relación entre dos o más variables supone plantear el problema de causalidad. El objetivo de esta técnica es profundizar en las relaciones de dependencia que se establecen entre un grupo de variables categóricas (neoplasias malignas) observadas en una misma población y cómo sus diferentes valores o categorías se relacionan entre sí. Se emplea la distancia ji-cuadrado como la métrica experimental con un nivel de significancia de 0.05 (1- $\alpha$ )100%.

Para la elaboración del manuscrito se utilizó el procesador de palabras MicrosoftWord 2000 (Microsoft Corporation 1999). Para analizar el conjunto de datos empleados en este estudio se empleo el paquete estadístico SPSS (versión 8.0, SPSS Inc. USA). Para la realización de las gráficas se empleó el sistema MATLAB 5.0. Una impresora HP LaserJet 1100.

## 1.2 Metástasis del cáncer.

El cáncer, crecimiento tisular producido por la proliferación continua de células anormales con capacidad de invasión y destrucción de otros tejidos, es considerado una enfermedad multifactorial, en la cual las células del huésped se transforman de normales en malignas como resultado de interacciones complejas de estímulos externos y genes que predisponen o suprimen la neoplasia. Aunque, se conoce que ciertos carcinógenos químicos y radiación ionizante son causas específicas de alteraciones a nivel de los genes, otras correlaciones son menos claras. Las masas cancerosas reciben también el nombre de tumores malignos.

El cáncer, puede originarse a partir de cualquier tipo de célula en cualquier tejido corporal, no es una enfermedad única sino un conjunto de enfermedades que se clasifican en función del tejido y célula de origen. Existen varios cientos de formas distintas de neoplasias malignas, siendo tres los principales subtipos: los sarcomas proceden del tejido conectivo como huesos, cartílagos, nervios, vasos sanguíneos, músculos y tejido adiposo. Los carcinomas proceden de tejidos epiteliales como la piel o los epitelios que tapizan las cavidades y órganos corporales, y de los tejidos glandulares de la mama y próstata. Los carcinomas incluyen algunos de los cánceres más frecuentes. Los carcinomas de estructura similar a la piel se denominan carcinomas de células escamosas. Los que tienen una estructura glandular se denominan adenocarcinomas. En el tercer subtipo se encuentran las leucemias y los linfomas, que incluyen los cánceres de los tejidos formadores de las células sanguíneas. Producen inflamación de los ganglios linfáticos, invasión del bazo y médula ósea, y sobreproducción de células blancas inmaduras. Estos factores ayudan a su clasificación.

El crecimiento canceroso se define por cuatro características que describen cómo las células cancerosas actúan de un modo distinto a las células normales de las que proceden. En primer lugar, la autonomía, ya que estas células han escapado al control que, en condiciones normales, rige el crecimiento celular. La segunda característica es la clonalidad, ya que el cáncer se origina a partir de una única célula progenitora que prolifera y da lugar a un clon de células malignas. Las otras dos características restantes son la anaplasia, ausencia de diferenciación normal y coordinada, y la metástasis o capacidad de crecer y diseminarse a otras partes del cuerpo. Estas características pueden ser expresadas por las células normales no malignas durante determinadas etapas, por ejemplo, durante la embriogénesis; no obstante, en las células cancerosas estas características tienen un grado inapropiado o excesivo.

Casi todos los cánceres forman tumores, pero no todos los tumores son cancerosos o malignos; la mayor parte son benignos (no ponen en peligro la salud). Los tumores benignos se caracterizan por un crecimiento localizado y suelen estar separados de los tejidos vecinos por una cápsula. Los tumores benignos tienen un crecimiento lento y una estructura semejante al tejido del que proceden. En ocasiones pueden producir alteraciones por obstrucción, compresión o desplazamiento de las estructuras vecinas, como sucede a veces en el cerebro. Algunos tumores benignos como los pólipos del colon pueden ser precancerosos.

El principal atributo de los tumores malignos es su capacidad de diseminación fuera del lugar de origen. La invasión de los tejidos vecinos puede producirse por extensión o infiltración, o a distancia, produciendo crecimientos secundarios conocidos como metástasis. La localización y vía de propagación de las metástasis varía en función de los cánceres primarios:

1. Cuando un cáncer invade la superficie del órgano de origen, las células pueden propagarse desde esta superficie a la cavidad vecina y a órganos adyacentes, donde pueden implantarse.
2. Las células malignas pueden viajar en el interior de los vasos linfáticos hacia los ganglios linfáticos, o también en los vasos sanguíneos. En la corriente circulatoria, estas células se detienen en el punto en el que los vasos son demasiado estrechos para su diámetro. Las células procedentes de tumores del tracto gastrointestinal se detienen principalmente en el hígado. Posteriormente pueden propagarse a los pulmones. Las células de otros tumores tienden a invadir los pulmones antes de propagarse a otros órganos. Por tanto, los pulmones y el hígado son dos localizaciones frecuentes de metástasis.
3. Muchas neoplasias envían células a la corriente circulatoria de manera temprana, y mientras algunas de estas células mueren, otras pueden invadir y penetrar en el árbol vascular y en los tejidos. Si este tejido tiene condiciones favorables para la célula tumoral, ésta se multiplica produciendo una metástasis. En ocasiones, sólo se multiplica un pequeño número de veces produciendo un cúmulo de células que permanecen quiescentes en forma de micrometástasis. Este estadio latente puede perdurar varios años, y por razones desconocidas puede reactivarse y producir un cáncer recurrente.

Muchas veces las células malignas conservan las características físicas y biológicas del tejido del que proceden a pesar de estar ampliamente diseminadas. De este modo, un patólogo puede, a través del examen microscópico de estas células, determinar la procedencia de los tumores metastásicos. Los tumores de las glándulas pueden ser identificados porque en ocasiones producen de forma indiscriminada la misma hormona producida por el tejido del que proceden. A veces, también responden a las hormonas que controlan esos tejidos en condiciones normales.

Cuanto más agresivo y maligno es un cáncer, menos recuerda a la estructura del tejido del que procede, pero la tasa de crecimiento del cáncer depende no sólo del tipo celular y grado de diferenciación, sino también de factores dependientes del huésped. Una característica de malignidad es la heterogeneidad celular del tumor. Debido a las alteraciones en la proliferación celular, las células cancerosas son más susceptibles a las mutaciones. Con la evolución, el tumor es cada vez menos diferenciado y de crecimiento más rápido. También puede desarrollar resistencia a la quimioterapia o a la radiación.

### **1.2.1 Factores de riesgo para desarrollar una neoplasia maligna.**

El cáncer en los niños es poco común. A diferencia del cáncer de los adultos, que es la segunda causa de muerte y se clasifican topológicamente, en los niños es la principal y no se puede diagnosticar tempranamente porque sus sistemas están en desarrollo y se clasifican histológicamente. Los tratamientos médicos que se emplean en ocasiones son muy fuertes, como la quimioterapia, baja los glóbulos rojos y las defensas de los niños, por lo que quedan más expuestos a virus y bacterias del ambiente.



Ciertos factores son capaces de provocar un cáncer en un porcentaje de los individuos expuestos a ellos. Entre éstos se encuentran la herencia, las infecciones, las radiaciones ionizantes, los productos químicos y las alteraciones del sistema inmunológico. Los investigadores estudian como estos diferentes factores pueden interactuar de una manera multifactorial y secuencial para producir tumores malignos. El cáncer es, en esencia, un proceso genético. Las alteraciones genéticas pueden ser heredadas, o producidas en alguna célula por un virus o por una lesión provocada de manera externa. Probablemente una serie de mutaciones secuenciales conduce a la única célula maligna que se multiplique como un solo clon. En un principio se consideró que un clon maligno era completamente anormal, y que la única curación posible era la eliminación de todas las células anormales del organismo. En la actualidad, se sabe que el problema reside en la incapacidad de la célula de diferenciarse en su estado adulto y funcional, quizás por la ausencia de algún factor necesario para esa diferenciación.

**Factores hereditarios.** Algunas formas de neoplasias malignas son más frecuentes en algunas familias: el cáncer de mama es un ejemplo de ello. Una forma de retinoblastoma sólo aparece cuando está ausente un gen específico. Estos genes, denominados genes supresores tumorales o antioncogenes, previenen en condiciones normales la replicación celular. Su ausencia elimina el control normal de la multiplicación celular. En algunos trastornos hereditarios, los cromosomas tienen una fragilidad intrínseca; estos procesos conllevan un riesgo elevado de neoplasia.

Las **infecciones** pueden llegar a provocar cáncer y, en concreto, aquellas relacionadas con los cánceres de estómago, hígado, cérvix y con el sarcoma de Kaposi (un tipo especial de cáncer que aparece en enfermos de SIDA).

El **virus HTLV**, sin embargo, es del tipo ARN, o retrovirus, como la mayor parte de los virus asociados a tumores en animales. Produce una leucemia humana. En presencia de una enzima denominada transcriptasa inversa, induce a la célula infectada a producir copias en ADN de los genes del virus, que de esta manera se incorporan al genoma celular. Estos virus del tipo ARN contienen un gen denominado oncogén viral capaz de transformar las células normales en células malignas. Distintas investigaciones han demostrado que los oncogenes virales tienen una contrapartida en las células humanas normales: es el proto-oncogén, u oncogén celular. Los productos de los oncogenes (las proteínas que producen) son factores de crecimiento (o proteínas necesarias para la acción de tales factores de crecimiento), que estimulan el crecimiento de las células tumorales.

Las **radiaciones ionizantes** son uno de los factores causales más reconocidos. La radiación produce cambios en el ADN, como roturas o trasposiciones cromosómicas en las que los cabos rotos de dos cromosomas pueden intercambiarse. La radiación actúa como un iniciador de la carcinogénesis, induciendo alteraciones que progresan hasta convertirse en cáncer después de un periodo de latencia de varios años. En este intervalo puede producirse una exposición a otros factores.

**Productos químicos.** El proceso por el que los productos químicos producen cáncer ha sido ampliamente estudiado. Algunos actúan como iniciadores. Sólo se requiere una única exposición, pero el cáncer no aparece hasta pasado un largo periodo de latencia y tras la exposición a otro agente denominado promotor. Los iniciadores producen cambios irreversibles en el ADN. Los promotores no producen alteraciones en el ADN pero sí un incremento de su

síntesis y una estimulación de la expresión de los genes. Su acción sólo tiene efecto cuando ha actuado previamente un iniciador, y cuando actúan de forma repetida. Los carcinógenos químicos producen también roturas y translocaciones cromosómicas.

**Factores ambientales.** Se calcula que los factores ambientales son la causa del 80% de los cánceres. La relación causa efecto más demostrada es la del humo del tabaco, inhalado de forma activa o pasiva; es responsable de cerca del 30% de las muertes por cáncer. Los factores alimenticios pueden ser responsables de un 40%, pero la relación causal no está tan establecida, y no se conocen con exactitud los constituyentes de la dieta que son responsables.

**Oncogenes.** Sea cual fuere la causa inicial por la que una célula normal se transforma en cancerosa y transmite a su descendencia esta alteración, es necesario que esta modificación se encuentre impresa en su ADN. Esto supone que todos los posibles agentes promotores o causales del cáncer actúan sobre un sustrato genético común dentro de la célula. Actualmente, se acepta que todos los factores cancerígenos (sustancias químicas, radiaciones, infecciones, etc) actúan sobre porciones del ADN que deben ser fundamentales en la regulación del crecimiento y diferenciación celular. Estas porciones reciben el nombre de proto-oncogenes y bajo la acción de diversos factores sufren pequeñísimas alteraciones que ocasionan el comienzo del fenómeno canceroso, desarrollándose el oncogén. Se cree que el proceso de reproducción sin control es la consecuencia de una serie de alteraciones que comienzan con un gen alterado o una mutación somática (una mutación de una célula normal de un tejido corporal), seguida de la acción promotora de algún agente que estimula la expresión de uno o varios oncogenes, o inhibe los efectos de uno o varios anti-oncogenes; en consecuencia se liberan factores de crecimiento. Es posible que el primer evento sea la falta de producción de los metabolitos necesarios para la diferenciación celular normal. La estimulación por los factores de crecimiento produce la proliferación del clon de células indiferenciadas, y un defecto del sistema inmunológico permite que estas células alteradas escapen a la destrucción por el sistema de control del organismo.

El dolor relacionado con el cáncer surge cuando la masa tumoral ejerce presión sobre los nervios o bloquea un conducto, de modo que se acumula la presión por la presencia de secreciones, lo que conlleva al término del niño.

### 1.3 Tratamientos para el cáncer pediátrico.

La oncología pediátrica es la ciencia que estudia los tumores en niños y se encarga de suministrar tratamientos capaces de evitar, disminuir y eliminar las células cancerosas.

En la actualidad existen dos métodos de curación, los tratamientos oncológicos (ortodoxos) y tratamientos o terapias alternativas (no ortodoxas). En la primera se emplean los siguientes tratamientos:

**Quimioterapia:** Consiste en el uso de medicamentos para eliminar células cancerosas. La quimioterapia se puede tomar en forma oral, o puede ser administrada en el cuerpo con una aguja en una vena o músculo. La quimioterapia se considera un tratamiento sistémico ya que el medicamento se introduce al torrente sanguíneo, viaja a través del cuerpo y puede eliminar células cancerosas por todo el cuerpo, la quimioterapia es útil para aquellos tumores cuya diseminación los hace inaccesibles a la cirugía o a la radioterapia. La quimioterapia a veces se puede aplicar al líquido que rodea el cerebro y la médula espinal a través de una aguja en el cerebro a la espalda (quimioterapia intratecal).

El tejido canceroso tiene una mayor proporción de células en división que los tejidos normales (en los que las células de soporte o de relleno están en una fase quiescente y son por tanto resistentes a los efectos del fármaco). Dentro de los tejidos normales, los que tienen una tasa de proliferación más rápida son la médula ósea y las células de recubrimiento del tracto gastrointestinal. Son los dos tejidos más sensibles al efecto de la quimioterapia y del grado de lesión de éstos depende la toxicidad, que limitará la máxima dosis tolerable de los fármacos anticancerosos.

Para que el tratamiento sea efectivo, la sensibilidad del tumor debe ser superior a la del tejido normal más sensible. Mientras algunos tumores son varias veces más sensibles, otros sólo son ligeramente sensibles. Por fortuna, las células de la médula ósea pueden dividirse a mayor velocidad que las células malignas y por tanto se recuperan con mayor rapidez. Si se repite un ciclo del fármaco en este momento, el tumor no ha tenido tiempo de crecer demasiado. Los ciclos repetidos reducen de forma paulatina el tumor antes de la aparición de resistencias.

La sensibilidad de ciertos tumores a la quimioterapia es tal que es posible la curación en un alto porcentaje: esto sucede en el cáncer uterino; las leucemias agudas (sobre todo en los niños); la enfermedad de Hodgkin y los linfomas difusos de células grandes; el carcinoma de testículo; el carcinoma de ovario; los carcinomas de células pequeñas del pulmón, y gran parte de los cánceres infantiles. Muchas veces estos procesos cancerosos se han diseminado en el momento del diagnóstico y no existe otra opción terapéutica. Otros cánceres avanzados tienen buena respuesta a la quimioterapia y pueden ser controlados durante periodos prolongados, por lo que se utiliza con frecuencia como tratamiento paliativo.

Los dos principales problemas que limitan la utilización de la quimioterapia son la toxicidad y la resistencia. Las técnicas que evitan o controlan la toxicidad y disminuyen el riesgo de resistencias se han ido perfeccionando. Es importante la instauración precoz del tratamiento, la utilización de dosis óptimas del fármaco, la repetición de los ciclos con intervalos cortos si es posible, siempre que se permita la recuperación del paciente de los efectos tóxicos

Es eficaz la utilización de múltiples fármacos. Los protocolos de quimioterapia utilizan diferentes fármacos (a menudo entre 3 y 6 al mismo tiempo), cada uno de los cuales es eficaz de forma aislada. Se combinan fármacos con diferentes mecanismos de acción para evitar la aparición de resistencias cruzadas, y con diferentes tipos de toxicidad para poder emplear la dosis óptima de cada fármaco, sin producir toxicidad aditiva que puede resultar fatal.

Se puede utilizar la quimioterapia junto a la cirugía o a las radiaciones en la denominada terapia combinada. Muchas veces se utiliza como tratamiento coadyuvante cuando la cirugía es la principal modalidad terapéutica. Suele administrarse tras la cirugía. Esta terapia es muy eficaz en el cáncer de mama. El objetivo principal de la quimioterapia es contribuir a la eliminación de las micrometástasis que pudieran existir previamente a la cirugía. Es más reciente la utilización de la quimioterapia previa a la cirugía como tratamiento coadyuvante, que puede además reducir la masa del tumor y facilitar su operabilidad.

**Radioterapia:** Las radiaciones ionizantes pueden ser electromagnéticas o por partículas y producen una destrucción tisular. La radiación electromagnética incluye los rayos gamma, una forma de emisión radiactiva, y los rayos X, que se producen cuando un haz de electrones impacta en un metal pesado. La radiación de partículas incluye haces de electrones, protones, neutrones, partículas alfa (núcleos de helio) y piones.

La sensibilidad de los tumores a las radiaciones es muy variable. Son tumores sensibles aquellos cuya sensibilidad es superior a la de los tejidos vecinos normales. Cuando tales tumores son además accesibles —los tumores superficiales o los tumores en órganos como el útero en el que se puede introducir una fuente de radiación— pueden ser curados mediante radioterapia. La propiedad de la radiación de respetar hasta cierto punto los tejidos normales permite el tratamiento de tumores en localizaciones donde no es posible la cirugía por la proximidad de tejidos vitales o porque el tumor ha empezado a infiltrar estructuras adyacentes que no pueden ser sacrificadas. La radioterapia también se emplea con frecuencia como tratamiento paliativo, sobre todo en las metástasis.

La radioterapia puede ser útil como coadyuvante a la cirugía. La radiación preoperatoria puede esterilizar las células tumorales con rapidez, impidiendo su diseminación en el acto quirúrgico. También puede disminuir la masa tumoral facilitando la cirugía, o transformando un tumor inoperable en otro operable. En otros casos la radioterapia se emplea en el postoperatorio.

**Terapia biológica:** su propósito es el de tratar que el cuerpo combata el cáncer. En ella se utilizan materiales producidos por el cuerpo o fabricados en un laboratorio para impulsar, dirigir o restaurar las defensas naturales del cuerpo contra la enfermedad. También se conoce con el nombre de terapia modificadora de la respuesta biológica o inmunoterapia.

**Cirugía:** El propósito del médico será el de extraer la mayor cantidad de tumor posible, dependiendo del sitio y tipo de cáncer. Si el cáncer está en un lugar del que no se puede remover, la cirugía puede limitarse a una biopsia del cáncer. La principal estrategia para el tratamiento curativo del cáncer es la escisión de todas las células malignas mediante una intervención quirúrgica. En el pasado, esto implicaba la escisión de todo el tejido afectado y de la mayor cantidad posible de tejido potencialmente afectado, incluidos los tejidos vecinos y los ganglios linfáticos. Para algunos tumores, y en especial el cáncer de mama, no es precisa una cirugía tan

ablative (mastectomía) en la mayor parte de los casos. Las mejoras en las técnicas quirúrgicas, los conocimientos en fisiología, en anestesia y la disponibilidad de potentes antibióticos y hemoderivados, han permitido realizar cirugías más limitadas, con menos secuelas y más pronta recuperación. Sin embargo, muchos cánceres están demasiado extendidos en el momento del diagnóstico para que la cirugía curativa sea posible. Si la extensión local del tumor afecta a tejidos vecinos que no pueden ser resecaados, o si existen metástasis a distancia, la cirugía no será un tratamiento curativo. Sin embargo, puede ser beneficiosa para el alivio sintomático de ciertas situaciones como la obstrucción, o puede tener el objetivo de disminuir la masa tumoral para permitir una mejor respuesta al tratamiento quimioterapéutico o radioterapéutico sucesivo.

**Trasplante de médula ósea:** consiste en administrar dosis elevadas de quimioterapia con o sin radioterapia con el fin de destruir toda la médula ósea del cuerpo, en un primer paso. Luego se extrae médula ósea sana de otra persona (un donante) cuyo tejido sea igual o parecido al del paciente. El donante puede ser un gemelo (el donante ideal), o un hermano o hermana u otra persona que no tenga ningún parentesco con el paciente. Por último, la médula sana del donante se le pone al paciente con una aguja en una vena, reemplazando así la médula ósea destruida. Este tipo de trasplante se conoce como alogénico.

Efectos indeseables pueden resultar de los tratamientos mucho después de que estos han terminado, por lo que es importante que el niño continúe bajo control médico. La quimioterapia puede acarrear más adelante problemas cardíacos, así como también problemas del riñón y de audición. Algunos tipos de radioterapia también pueden ocasionar problemas de crecimiento, de desarrollo y de la vista.

La segunda, medicina complementaria y alternativa (no ortodoxas, CAM, por sus siglas en inglés) también conocida como medicina integral, consta de una serie extensa de filosofías, enfoques y terapias curativas. Una terapia se denomina complementaria cuando ésta se usa además de los tratamientos convencionales u ortodoxos; cuando se usa en vez de los tratamientos convencionales se le denomina alternativa. Los tratamientos convencionales u ortodoxos, son aquellos tratamientos ampliamente aceptados y puestos en práctica por la comunidad médica establecida. Dependiendo de cómo se utilicen algunas terapias, éstas pueden ser consideradas complementarias o alternativas.

Las terapias complementarias o alternativas se utilizan como un esfuerzo para prevenir la enfermedad, reducir el estrés, prevenir o reducir los efectos secundarios y los síntomas, o con el fin de controlar o curar la enfermedad. Entre algunos de los métodos complementarios o alternativos comúnmente utilizados se encuentran las intervenciones que tratan sobre el control mental/corporal como la visualización o relajación, curación manual incluyendo la acupresión y los masajes, homeopatía, vitaminas o productos herbáceos y acupuntura.

Aunque existen pocos estudios sobre el uso de las terapias complementarias y alternativas contra el cáncer, un estudio a gran escala encontró que el porcentaje de pacientes en los Estados Unidos que utiliza estas terapias era del nueve por ciento en general<sup>24</sup>.

La evaluación científica es importante para entender si las terapias alternativas a través del desarrollo de estudios científicos mejoran o coadyuvan la salud. Los enfoques convencionales de tratamiento contra el cáncer se han estudiado por lo general para determinar su seguridad y

eficacia a través de un proceso científico riguroso que incluye estudios clínicos con un gran número de pacientes. A menudo, es muy poco lo que se sabe acerca de la seguridad y eficacia de los métodos complementarios y alternativos. Algunas de estas terapias complementarias o alternativas no han sido sujetas a evaluaciones rigurosas. Hoy en día, las que se consideraron alguna vez poco ortodoxas, están encontrando un lugar en el tratamiento contra el cáncer, no como curas, sino como terapias complementarias que pueden ayudar a los pacientes a sentirse mejor y recuperarse más pronto. Un ejemplo es la acupuntura. De acuerdo a un grupo de expertos participantes en una Conferencia de Consenso de los Institutos Nacionales de la Salud (National Institutes of Health) en noviembre de 1997, se encontró que la acupuntura era eficaz en el manejo de las náuseas y el vómito asociados con la quimioterapia y en el control del dolor asociado con la cirugía. Algunos enfoques, como el de "laetrile", han sido estudiados y se ha encontrado que son ineficaces o potencialmente dañinos.

Actualmente se practican en el norte de México, especialmente en Tijuana, donde los tratamientos son dosis de laetrile, dietas de nutrición cien por ciento fibras y jugos, ejercicios que complementan la inmunoterapia. Sin embargo, en el área médica aún no son aceptados como tratamientos capaces de tratar el cáncer<sup>34</sup>.

## 2. ANÁLISIS DE DATOS CATEGÓRICOS.

Hay dos formas en las cuales los datos estadísticos son recopilados. En una forma, las mediciones son registradas para cada individuo en la muestra, en la otra, los individuos son clasificados como pertenecientes a diferentes categorías. En muchas ocasiones, las clasificaciones son empleadas para presentar datos originales sobre mediciones directas, por ejemplo: distribución de frecuencias. Los datos colectados en la forma de mediciones quizás después sean agrupados y presentados como una distribución de frecuencias. Una ventaja importante de agrupar es la reducción considerable de datos.

El análisis de datos categóricos cae dentro de una amplia estructura de análisis multivariado. Los diferentes métodos dependen de la respuesta, si es binaria o polítoma. Una forma es el análisis de tablas de contingencia, en el cual se busca asociación entre filas y columnas, empleando diferentes estadísticos de prueba. En este contexto existen métodos alternativos de representación de datos colocados en una tabla de contingencia, como son: análisis discriminante, análisis canónico y análisis de correspondencias.

En este capítulo se describe brevemente la notación de tablas de contingencia, el Análisis Discriminante y el Análisis Canónico. El siguiente capítulo aborda más ampliamente el análisis de correspondencias.

### 2.1 Tablas de contingencia.

Una gran variedad de datos en las ciencias sociales y biológicas se representan en tablas cruzadas de clasificación, de enumeración o de conteo; comúnmente referidas como tablas de contingencia. Las unidades de una población muestral en tales circunstancias son clasificadas de acuerdo a cada una de varias variables categóricas o un conjunto de categorías que identifican o describen un objeto (Feinberg E., 1980).

Una tabla de contingencia es una forma de presentación de datos agrupados. En el caso más simple un grupo de  $N$  elementos pueden ser clasificado en sólo dos grupos, de acuerdo a presencia o ausencia de una cierta característica. Los grupos diferentes de clasificación son llamados categorías y estas pueden ser cuantitativas o cualitativas.

El objetivo primordial de las tablas de contingencia o dependencia se refiere a la independencia de dos métodos de clasificación de eventos observados, es decir, establecer relaciones de dependencia o asociación entre las columnas (o filas) de la tabla.

Supóngase que los datos corresponden a dos criterios de clasificación, los que se determinan como variable 1 y variable 2, las cuales se disponen en una tabla de contingencia:

		Variable 2				
variable 1		1	2	. . .	J	
1		$n_{11}$	$n_{12}$	. . .	$n_{1j}$	$n_{1\cdot}$
2		$n_{21}$	$n_{22}$	. . .	$n_{2j}$	$n_{2\cdot}$
.		.	.	.	.	.
.		.	.	.	.	.
.		.	.	.	.	.
I		$n_{i1}$	$n_{i2}$	. . .	$n_{ij}$	$n_{i\cdot}$
		$n_{\cdot 1}$	$n_{\cdot 2}$	. . .	$n_{\cdot j}$	$n$

Tabla de contingencia.

Un análisis típico de esta tabla involucra el estadístico  $\chi^2$  de Pearson, el cual es empleado para determinar si hay evidencia de cualquier asociación o dependencia significativa entre renglones y columnas de la tabla N, el cual es el siguiente:

$$\chi^2 = \frac{\sum \sum (n_{ij} - E(n_{ij}))^2}{E(n_{ij})}$$

donde  $E(n_{ij}) = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$ , con  $(i-1)(j-1)$  grados de libertad.

Si  $P = p_{ij} = \frac{n_{ij}}{n}$ , entonces, el estadístico puede ser reescrito como

$$\chi^2 = n \sum_i \sum_j \frac{(p_{ij} - r_i c_j)^2}{r_i c_j}$$

donde  $r_i = \frac{n_{i\cdot}}{n}$  y  $c_j = \frac{n_{\cdot j}}{n}$ .

La cantidad  $t = \frac{\chi^2}{n}$  es el coeficiente de asociación conocida como coeficiente de contingencia de Pearson o la **inercia total** de la tabla de contingencia.



## 2.2 Análisis discriminante múltiple

El Análisis Discriminante Múltiple (MDA, siglas en inglés), tiene sus orígenes en el trabajo realizado por Fisher en 1936 (Cuadras CM,1991). El análisis discriminante, es útil en situaciones donde la muestra total puede ser dividida en grupos, basada sobre una variable dependiente que tiene varias clases conocidas o categorías. El objetivo del MDA es crear grupos diferentes y predecir la probabilidad de que una entidad, individuo u objeto, pertenecerá a una clase o grupo particular basado sobre varias covariables cuantitativas.

### Discriminación con variables binarias.

Suponga que las poblaciones  $H_1, \dots, H_k$  representan  $K$  posibles estados (enfermedades) y  $A_1, \dots, A_n$  representan características que pueden estar presentes o ausentes (síntomas). Se definen entonces las variables binarias:

$$X_j(w) = \begin{cases} 1 & \text{si } w \text{ tiene la característica } A_j \\ 0 & \text{en caso contrario} \end{cases}$$

Supóngase a continuación que la probabilidad de presencia de  $A_j$  en  $H_i$  es

$$P(X_j = 1 | H_i) = p_{ij}.$$

La probabilidad de ausencia será:

$$P(X_j = 0 | H_i) = 1 - p_{ij}.$$

Si se supone que las variables  $X_1, \dots, X_n$  son independientes, la verosimilitud de la observación  $(x_1, \dots, x_n)$  en  $H_i$  es

$$L_i(x_1, \dots, x_n) = \prod_j p_{ij}^{x_j} (1 - p_{ij})^{1-x_j}$$

Entonces, si  $q_i$  es la probabilidad a priori de  $H_i$ , la Regla de Bayes consiste en asignar  $w$  a  $H_i$  si

$$q_i L_i(x_1, \dots, x_n) = \max_j (q_j L_j(x_1, \dots, x_n)).$$

La aplicación práctica de este procedimiento exige una adecuada estimación de las probabilidades  $p_{ij}$  y  $q_i$ . Si se dispone de una distribución de características y estados sobre una muestra de  $N$  individuos formando una tabla de contingencia.

		Características				
		A <sub>1</sub>	A <sub>2</sub>	. . .	A <sub>n</sub>	
F=	Poblaciones					
	H <sub>1</sub>	f <sub>11</sub>	f <sub>12</sub>	. . .	f <sub>1j</sub>	f <sub>1.</sub>
	H <sub>2</sub>	f <sub>21</sub>	f <sub>22</sub>	. . .	f <sub>2j</sub>	f <sub>2.</sub>
	.	.	.	.	.	.
	H <sub>k</sub>	f <sub>k1</sub>	f <sub>k2</sub>	. . .	f <sub>kj</sub>	f <sub>k.</sub>
		f <sub>.1</sub>	f <sub>.2</sub>	. . .	f <sub>.j</sub>	N

Donde  $f_{ij}$  es la frecuencia de  $A_j$  en la población  $H_i$ , las estimaciones máximo verosímiles son

$$p_{ij} = \frac{f_{ij}}{f_{k.}}$$

$$q_i = \frac{f_{k.}}{N}$$

La expresión de verosimilitud  $L_i(x_1, \dots, x_n)$  en el caso de que las variables sean dependientes ha sido abordada bajo diferentes perspectivas, entre algunas hay:

- 1) Tratar la información binaria  $(x_1, \dots, x_n)$  como variables cuantitativas y aplicar el discriminador lineal o el discriminador cuadrático.
- 2) Desarrollar una serie  $L_i(x_1, \dots, x_n)$ . Existen diferentes métodos de expansión ortonormal, expansión de Rademacher-Walsh y otros.

### 2.3 Análisis Canónico.

El análisis canónico o también llamado de correlación canónica (ACC) fue desarrollado por Hotelling en 1936 y puede ser visto como una extensión del análisis de regresión múltiple. Su objetivo es correlacionar simultáneamente varias variables, ya sean dependientes o covariables. El principio subyacente es desarrollar combinaciones lineales de cada conjunto de variables (ambas dependientes o independientes) de una manera que maximice la correlación entre los dos conjuntos.

**Análisis de tablas de contingencia.**

Considérese dos variables cualitativas H y A con n y p niveles respectivamente, y una muestra de K observaciones dando lugar a una tabla de contingencia  $F = (f_{ij})$ , en el mismo sentido que en la sección anterior. Sean entonces las variables dicotómicas

$$X_i(w) = \begin{cases} 1 & \text{si } w \text{ pertenece a la } i\text{-ésima columna} \\ & i = 1, \dots, n \\ 0 & \text{en caso contrario} \end{cases}$$

$$Y_j(w) = \begin{cases} 1 & \text{si } w \text{ pertenece a la } j\text{-ésima columna} \\ & j = 1, \dots, p \\ 0 & \text{en caso contrario} \end{cases}$$

las observaciones sobre un individuo  $w$  es una sucesión de  $n+p$  ceros y unos y la tabla de frecuencias  $F = (f_{ij})$  condensa la información contenida en una matriz de datos de orden  $k(n+p)$ , formada por ceros y unos. Una forma de estudiar la dependencia estocástica entre los dos conjuntos de variables consiste en aplicar un análisis de correlación canónica relacionando combinaciones lineales de las variables X con combinaciones lineales de las variables Y.

La super matriz de covarianzas es,

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \frac{1}{N} \begin{bmatrix} D_p - k^{-1}ff' & F' - Fs' \\ F - Fs & D_n - k^{-1}gg' \end{bmatrix}$$

siendo  $f = (f_1, \dots, f_p)'$ ,  $g = (g_1, \dots, g_k)'$ ,  $Fs = fg'$ ,  $D_p = \text{diag}(f)$ ,  $D_n = \text{diag}(g)$ . Sin embargo, las matrices  $S_{11}$  y  $S_{22}$  son singulares. Utilizando g-inversas, los vectores canónicos resultan de

$$(S_{11}^{-1}S_{12}S_{22}^{-1}S_{21} - \lambda I_p)v_i = 0.$$

$$(S_{11}^{-1} \quad S_{12}^{-1} \quad S_{22}^{-1} \quad S)v_i = 0.$$

Indicando  $J_p = I_p^{-1}/p$   $EE'$ , esta ecuación equivale a

$$(J_p D_p^{-1} F' D_n^{-1} F^{-1} I_p)u_i = 0.$$

La relación entre los vectores canónicos es  $v_i = S_{22}^{-1}S_{21}u_i$ ,  $u_i = J_n D_n^{-1} F u_i$ . Introduciendo el cambio de escala  $v'_i = J_n v_i$ ,  $u'_i = J_p u_i$ , la relación es

$$v'_i = D_n^{-1} F u'_i$$
$$u'_i = D_p^{-1} F' v'_i.$$

Las coordenadas de  $v'_i$  son el promedio ponderado de las coordenadas de  $u'_i$ .

La asociación entre las variables cualitativas se descompone en  $\min(n,p) - 1$  pares de variables canónicas, existiendo independencia completa si la primera correlación canónica no es significativa. La descomposición canónica de una tabla de contingencia mantiene bastantes analogías con el análisis de correspondencias y la descomposición aditiva de la ji-cuadrado.

### 3. ANÁLISIS DE CORRESPONDENCIAS.

#### Introducción.

El análisis de correspondencias (AC) tiene sus orígenes en el trabajo realizado por Jean Paul Benzècri en 1965. Se desarrolla como técnica estadística aplicada de naturaleza euclídea a la hora de analizar las tablas de contingencia, cuyo objeto es obtener una representación geométrica simultánea de los datos que ayude a visualizar y localizar las relaciones establecidas en grandes matrices de datos. Este nuevo enfoque del análisis de las tablas de contingencia, fuera del marco de la estadística clásica ha derivado posteriormente a nuevas formulaciones. De este modo, es plausible utilizarlo en tablas de datos cualitativos, y en general a cualquier estructura (poblaciones por caracteres o variables 1 por variable 2) que contengan valores enteros positivos.

El AC es un tipo de análisis de componentes principales sobre una distancia especial llamada ji-cuadrado, cuya principal propiedad es la posibilidad de una representación geométrica de dos conjuntos de datos que comprenden los renglones y columnas de una matriz de datos. Además, frecuentemente es expuesto en una estructura probabilística. Los números  $k_{ij}$  de la tabla de datos están presentados como los resultados del conteo efectuado sobre una muestra obtenida de un muestreo aleatorio independiente. Pero, los datos reales rara vez se ajustan al método.

Este capítulo describe los conceptos básicos del análisis de correspondencias.

#### 3.1 Fundamentos.

Supóngase que los datos de una tabla de frecuencias corresponden a dos criterios de clasificación, los que se determinan como "caracteres" (síntomas) y "poblaciones" (enfermedades), los cuales se disponen en una tabla de contingencia:

		J caracteres				
		I	B <sub>1</sub>	B <sub>2</sub>	... B <sub>p</sub>	
poblaciones	A <sub>1</sub>	k <sub>11</sub>	k <sub>12</sub>	...	k <sub>1p</sub>	k <sub>1.</sub>
	A <sub>2</sub>	k <sub>21</sub>	k <sub>22</sub>	...	k <sub>2p</sub>	k <sub>2.</sub>
	.	.	.	.	.	.
	.	.	.	.	.	.
	A <sub>n</sub>	k <sub>n1</sub>	k <sub>n2</sub>	...	k <sub>np</sub>	k <sub>n.</sub>
		k <sub>.1</sub>	k <sub>.2</sub>	...	k <sub>.p</sub>	k

Fig 1. Modelo de una tabla de contingencia, donde hay p formas para J y n formas para I.

$$k_{i.} = \sum_j^p k_{ij} \quad (3.1)$$

$$k_{.j} = \sum_i^n k_{ij} \quad (3.2)$$

donde  $k_{ij}$  es la frecuencia de presencia de la población  $A_i$  y el carácter  $B_j$ ,  $k_{i.}$  es la frecuencia de  $A_i$  o renglón marginal,  $k_{.j}$  es la frecuencia de  $B_j$  o columna marginal,  $k$  es el número total de individuos en la muestra.

$$k = \sum k_{ij} \quad (3.3)$$

La distribución de frecuencias o densidades marginales de los caracteres en la población  $A_i$  viene dada por el vector de coordenadas

$$A_i : \left( \frac{k_{i1}}{k_{i.}}, \dots, \frac{k_{ip}}{k_{i.}} \right) \quad (3.4)$$

que se puede entender como la densidad de probabilidad discreta de  $A_i$ , si se interpreta  $k_{ip}/k_{i.}$  como la "probabilidad del suceso  $B_j$  condicionado a la población  $A_i$ ".

Para aplicar el AC no se emplea directamente la tabla de frecuencias, sino que se transforma en una matriz de probabilidades, de acuerdo a la siguiente expresión:

$$f_{ij} = \frac{k_{ij}}{k} \quad \sum_{i=1}^n \sum_{j=1}^p f_{ij} = 1 \quad (3.5)$$

que se puede interpretar como la frecuencia relativa de  $i$  y  $j$ . De forma semejante se obtiene la masa de los renglones y columnas (perfiles marginales), las cuales miden la importancia relativa de sus elementos:

$$f_{i.} = \sum_{j=1}^p f_{ij} = \frac{k_{i.}}{k} \quad \sum_{j=1}^p f_{ij} = 1 \quad (3.6)$$

$$f_{.j} = \sum_{i=1}^n f_{ij} = \frac{k_{.j}}{k} \quad \sum_{i=1}^n f_{ij} = 1 \quad (3.7)$$

Por último, se verifica que las frecuencias condicionales o perfiles (profiles), porcentaje relativo a la suma de los renglones o columnas, construyen la configuración de puntos:

$$\begin{aligned} f_{ij} &= \frac{f_{ij}}{f_{.j}} = \frac{k_{ij}}{k_{.j}} \\ f_{ij} &= \frac{f_{ij}}{f_{i.}} = \frac{k_{ij}}{k_{i.}} \end{aligned} \quad (3.8)$$

**Importancia de los perfiles.**

Cualquier elemento  $i$  de  $I$  esta representado geoméricamente por un vector en  $R^p$ . El  $j$ -ésimo componente de este vector tiene los valores  $k_{ij}/k_{i.}$ , los cuales son los valores del perfil de  $i$  sobre  $J$ . A la inversa, cualquier elemento  $j$  de  $J$  esta representado geoméricamente por un vector en  $R^n$ . El  $i$ -ésimo componente de este vector tiene los valores  $k_{ij}/k_{.j}$ , los valores del perfil de  $j$  sobre  $I$ . El modelo básico de análisis de correspondencias involucra cálculos basados sobre los perfiles de  $I$  y  $J$ . De este modo, los elementos  $i$  e  $i'$  de  $I$  son considerados semejantes si y sólo si sus perfiles asociados  $k_{ij}/k_{i.}$  y  $k_{i'j}/k_{i'.}$  son semejantes, es decir; si la  $j$ -ésima columna y la  $j'$ -ésima columna de la tabla de contingencia son proporcionales. De igual manera se realiza para  $i$ .

**Definición.** La nube de puntos  $N(J)$  es el conjunto de elementos de  $j \in J$ , cuyas coordenadas son los componentes del perfil  $f_j' = \{f_{ij}, j \in J\}$  y cuya masa es  $f_{.j}$ .  $N(J)$  es un subconjunto de  $P(I)$ , es decir; si se aplican las leyes de probabilidad sobre  $I$ , este mismo; es un subconjunto del espacio  $R^p$ .

De este modo,

$$N(I) = \left\{ \frac{k_{ij}}{k_{i.}}, f_i \right\} \subset P(J) \subset R^p,$$

$$N(J) = \left\{ \frac{k_{ij}}{k_{.j}}, f_j \right\} \subset P(I) \subset R^n.$$

**Espacio de los perfiles**

El conjunto  $I$  esta representado en el espacio de perfiles sobre  $J$ . Contrariamente, el conjunto  $J$  esta representado en el espacio de perfiles sobre  $I$ :

$$\left\{ \frac{k_{ij}}{k_{i.}}; i \in I \right\} = \text{conjunto de perfiles de } i \in I.$$

$$\left\{ \frac{k_{ij}}{k_{.j}}; j \in J \right\} = \text{conjunto de perfiles de } j \in J.$$

Considere las leyes de probabilidad sobre el conjunto  $J$ , denotado por  $P(J)$ . Un elemento del perfil sobre  $J$ , de esta manera es

$$\pi_j = \{\pi_1, \dots, \pi_i, \dots, \pi_n\}$$

tal que  $\pi_1 + \pi_2 + \dots + \pi_n = 1$  con  $\pi_i \geq 0, \forall i \in I$ . Cuando la dimensión de  $J$  es igual a 2, entonces,  $\pi_j = (\pi_1, \pi_2)$  con  $\pi_1 + \pi_2 = 1$  y  $\pi_1, \pi_2 \geq 0$ , se tiene un segmento de recta como se muestra en la figura 1. Si la dimensión de  $J$  es igual a 3, entonces,  $\pi_j = (\pi_1, \pi_2, \pi_3)$  con  $\pi_1 + \pi_2 + \pi_3 = 1$  y  $\pi_1, \pi_2, \pi_3 \geq 0$ , cuya representación geométrica corresponde a un triángulo equilátero. Fig. 2. En el caso general, las leyes de probabilidad se aplican al espacio de perfiles sobre  $J$ , tal que

$$\begin{matrix} k_{ij} \\ k_{i\cdot} \end{matrix} : \left( k_{i1}, \dots, k_{ip} \right) \text{ con } \sum_{j \in J} \frac{k_{ij}}{k_{i\cdot}} = 1 \text{ y } \frac{k_{ij}}{k_{i\cdot}} \geq 0, \forall i \in I.$$

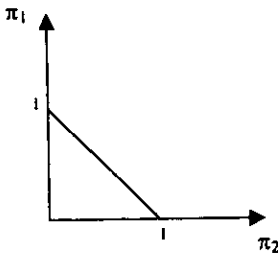


Figura 1

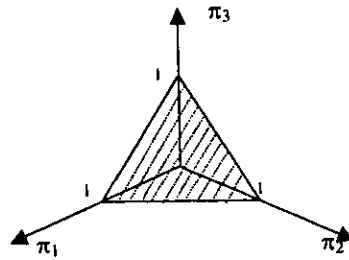


Figura 2

La nube de puntos  $N(J)$  es el conjunto de elementos de  $j \in J$ , cuyas coordenadas son los componentes del perfil  $\frac{f_{ij}}{f_{i\cdot}} = \{f_{ij}, i \in I\}$  y cuya masa es  $f_{j\cdot}$ .

### 3.2 Distancia ji-cuadrado.

Uno de los objetivos del AC es obtener una representación geométrica de las poblaciones  $A_1, \dots, A_n$  en relación a la distribución de frecuencias relativas de los caracteres. Sin embargo, la distancia utilizada es la ji-cuadrado, que es diferente a la distancia basada en la métrica experimental.

La distancia ji-cuadrado entre las poblaciones  $A_i, A_{i'}$  en relación a los caracteres  $B_1, \dots, B_p$  es

$$d^2(i, i') = \sum \frac{1}{f_{\cdot j}} \left( \frac{f_{ij}}{f_{i\cdot}} - \frac{f_{i'j}}{f_{i'\cdot}} \right)^2 = \sum_{j=1}^p \left( \frac{f_{ij}}{f_{i\cdot} f_{\cdot j}} - \frac{f_{i'j}}{f_{i'\cdot} f_{\cdot j}} \right)^2 \quad (3.9)$$

y se aplica de forma similar a  $j$  y  $j'$ .



A este tipo de distancia se le denomina también “ponderada”, por cumplir la propiedad de invarianza o equivalencia distribucional.

De hecho, esta distancia solo se diferencia de la distancia Euclídea usual, en que cada cuadrado es ponderado por la inversa de la frecuencia correspondiente a cada término, en otras palabras, es la distancia  $\chi^2$  entre la distribución de frecuencias condicionales con respecto a la distribución de frecuencias  $\frac{k_{.j}}{k}$ .

Esencialmente, la razón para elegir la distancia ji-cuadrado es que esta verifica la propiedad distribucional. Esta propiedad es importante, porque garantiza invarianza de los resultados sin tener en cuenta de como las variables fueron originalmente codificadas. En otras palabras, esto implica que si se unen dos puntos fila o columna con idénticos perfiles, la distancia entre dos puntos queda invariante.

La originalidad de este tipo de análisis, radica en la utilización de la distancia ji-cuadrado, que si bien no es equivalente al coeficiente de contingencia t (krzanowski, et al 1994)

### Bondad de ajuste

Para construir la configuración de puntos en  $R^p$  y  $R^n$ , elegir los perfiles como coordenadas dando ponderados a cada población y cada carácter igualmente.

Para calcular la bondad de ajuste, es natural dar a cada punto un peso que es proporcional a su frecuencia, así no hay una sobre representación de categorías con totales pequeños, y consecuentemente reflejar la distribución de la población real. Sus pesos son empleados para calcular las coordenadas del perfil medio (centro de gravedad) así como el criterio de bondad de ajuste. La cantidad a ser maximizada es una suma de cuadrados ponderados de esos pesos. El punto  $i$  de  $R^p$  de este modo tiene un peso de  $f_{i.}$ , mientras que el peso del punto  $j$  de  $R^n$  es  $f_{.j}$

De acuerdo con esta distancia, las poblaciones  $A_1, \dots, A_n$  están representadas por una configuración de  $n$  puntos en un espacio euclídeo  $R^p$  de coordenadas

$$P_i = \left( \frac{f_{i1}}{f_{i.}\sqrt{f_{.1}}}, \dots, \frac{f_{ip}}{f_{i.}\sqrt{f_{.p}}} \right) \quad (3.10)$$

separados por la distancia euclídea ordinaria. Se comprueba en seguida que tal configuración esta contenida en el hipérplano de ecuación

$$\sum x_j \sqrt{f_{.j}} = 1. \quad (3.11)$$

Dentro de este contexto, se define  $K$  como una matriz  $(n,p)$  de frecuencias de los datos originales, y por  $F$  la matriz de frecuencias relativas:

$$F = \frac{1}{k} K \quad (3.12)$$

También se definen las matrices diagonales,  $D_p$  de  $(p,p)$ , cuyos elementos son  $f_{.j}$ ; y  $D_n$  de  $(n,n)$  cuyos elementos son  $f_{i.}$ .

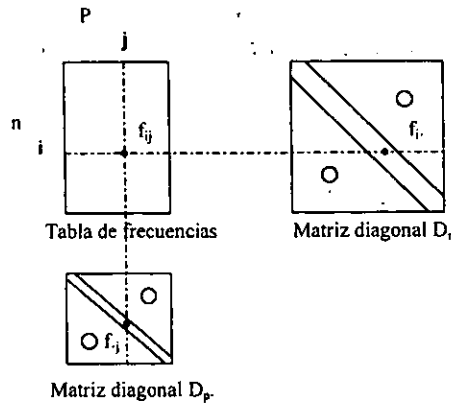


Figura 3. Matriz  $F$ ,  $D_n$  y  $D_p$ .

### 3.3 Representación de las poblaciones.

La existencia de una completa simetría entre los índices  $i$  y  $j$ , permite hacer el análisis sobre el espacio  $R^p$ , la demostración sobre el otro espacio es deducida por permutación de los índices  $i$  y  $j$ , esto es, transformando  $F$  y sustituyendo las matrices  $D_p$  y  $D_n$ .

Se desea representar las distancias gráficamente entre perfiles (profiles). Si los perfiles son coordenadas euclídeas se debe buscar un nuevo sistema de coordenadas euclídeas para el cual, el análisis general pueda ser aplicado.

Si se transforma  $f_{ij}/f_{i.}$  en  $f_{ij}/\sqrt{f_{i.}f_{.j}}$  se obtiene la distancia ji-cuadrado entre  $i$  e  $i'$  representada en la ecuación (4.9). Por lo tanto,  $f_{ij}/\sqrt{f_{i.}f_{.j}}$ , es un conjunto de coordenadas euclídeas. Una nueva nube de puntos  $i$ , denotada por  $N'(I)$ , es definida como

$$N'(I) = \{i = 1, \dots, n, f_{ij}(f_{.j})^{-1/2}, \text{ masa} = f_{.j}\} \subset P(J) \subset R^p.$$

Los centros de gravedad de  $N(I)$  y  $N'(I)$ , denotados por  $G$  y  $G'$ , respectivamente son:

$$\begin{aligned} G \text{ tiene } g_1, \dots, g_j, \dots, g_p \text{ como coordenadas en } R^p; \\ G' \text{ tiene } g'_1, \dots, g'_j, \dots, g'_p \text{ como coordenadas en } R^p; \end{aligned}$$

sabiendo que

$$G \text{ tiene } f_{i1}, \dots, f_{ij}, \dots, f_{ip} \text{ como valores, donde } f_{ij} \text{ es la distancia marginal con el perfil } \frac{f_{ij}}{f_{.j}}.$$

Para  $G'$ , lo siguiente se mantiene

$$\begin{aligned} g'_j &= \sum f_{ij} \left( \frac{f_{ij}}{f_{.j}} \right) f_{.j}^{-1/2} \\ &= f_{ij} f_{.j}^{-1/2} = \sqrt{f_{ij}}. \end{aligned}$$

De este modo, el vector de medias se indica por

$$G' = (\sqrt{f_{.1}}, \dots, \sqrt{f_{.j}}, \dots, \sqrt{f_{.p}})$$

Las coordenadas de los puntos  $i$  de  $N(I)$ , son calculados con respecto a  $G'$ . La masa,

$$\overline{GM}_i = \frac{f_{ij}}{f_{i\cdot} \sqrt{f_{.j}}} - \sqrt{f_{.j}}, \text{ asociada con el punto } i \text{ es } f_{i\cdot}.$$

Ahora, el análisis general puede ser aplicado a la matriz  $X$ , donde  $X_{ij}$  esta definida como:

$$X_{ij} = \frac{f_{ij}}{f_{i\cdot} \sqrt{f_{.j}}} - \sqrt{f_{.j}}$$

La matriz  $X'D_p X$ , denotada por  $C_p$  es la matriz de covarianzas entre los caracteres y es diagonalizada como:

$$C_p = \sum_i f_{i\cdot} \left( \frac{f_{ij}}{f_{i\cdot} \sqrt{f_{.j}}} - \sqrt{f_{.j}} \right) \left( \frac{f_{i\ell}}{f_{i\cdot} \sqrt{f_{.j}}} - \sqrt{f_{.j}} \right)$$

dando los eigenvalores  $\lambda_\alpha$  y los eigenvectores  $u_\alpha$ . Note que  $u_0 = (\sqrt{f_{.1}}, \dots, \sqrt{f_{.p}})$  es el vector

propio de  $C_p$  asociado con el valor propio  $\lambda_{\alpha}=0$ . En efecto, se comprueba que

$$\begin{aligned} C_p u_0 &= 0. \\ C_{1j} f_{\cdot 1}, \dots, C_{pj} f_{\cdot p} &= 0. \end{aligned}$$

Ahora,  $u_0 = (f_{\cdot 1}, \dots, f_{\cdot p})$  es ortogonal a cualquier vector propio  $u'$  de  $C_p$ :

$$u_0 u' = \sum_j u_0 u'_j = \sum_j u'_j f_j = 0.$$

También  $u_0$  es vector propio de  $X'D_p X$  considerando sólo los vectores propios de valor propio distinto de 1. Como el valor propio 1 corresponde al vector propio 0 de  $C_p$ , es demostrable ver que los demás valores propios de  $X'D_p X$  son menores que 1.

$$1 > \lambda_2 \geq \dots \geq \lambda_p.$$

Para determinar el porcentaje de variabilidad explicada por una representación en dimensión  $d$ , se divide la variabilidad explicada por los ejes entre la variabilidad total (VT):

$$\begin{aligned} P_d &= 100 \left( \frac{\mu_2 + \dots + \mu_d}{\mu_2 + \dots + \mu_p} \right) \\ VT = \text{traza} C_p &= \sum \frac{f_{ij}^2}{f_{\cdot j} f_{\cdot j}} - f_{\cdot j} = \left( \sum \frac{f_{ij}^2}{f_{\cdot j} f_{\cdot j}} \right) - 1 \end{aligned}$$

siendo  $D_p = \text{diag}(f_{\cdot 1}, \dots, f_{\cdot p})$ , y el porcentaje de varianza explicada por el segundo, tercer, . . . , ejes es

$$P_d = 100 \left[ \frac{\lambda_2 + \dots + \lambda_d}{\text{traza}(X'D_p X)^{-1}} \right] = 100 \left[ \frac{\lambda_2 + \dots + \lambda_d}{\lambda_2 + \dots + \lambda_p} \right]$$

La representación en dimensión  $d$  proporciona una representación de las poblaciones separadas por la distancia ji-cuadrado, salvo la pérdida de información producida al reducir la dimensión.

### 3.4 Representación simultánea.

Considérese nuevamente la tabla de contingencia  $K$  de  $n$  poblaciones y  $p$  caracteres. Se desea representar, sobre los mismos ejes, el conjunto total de poblaciones y caracteres así como aproximar la siguiente situación ideal:

- a) cada punto del carácter  $j$  es un baricentro de los puntos de las poblaciones, cada población  $i$  tiene el ponderado que parte de la población  $i$  está en el perfil del carácter  $j$ , en otras palabras, el ponderado es:

$$p_i = \frac{f_{ij}}{f_{\cdot j}} \quad \sum_i p_i = 1 \quad (3.13)$$

- b) Cada punto de la población  $i$  es un baricentro de los puntos de los caracteres, cada uno de los caracteres tiene el ponderado que parte del carácter  $j$  y está en el perfil de la población  $i$ , en otras palabras, el ponderado es:

$$p_j = \frac{f_{ij}}{f_{i\cdot}} \quad (3.14)$$

Esta situación, frecuentemente es imposible, porque implica que cada conjunto esta contenido totalmente en el otro. Es decir, existe una solución trivial para la cual todos los puntos de ambos conjuntos están unidos en el punto 1 de la abscisa, por lo que, ambas representaciones están estrechamente relacionadas entre si.

El problema de representar las  $i$  poblaciones y los  $j$  caracteres, se resuelve mediante la diagonalización de  $D_n$  y  $D_p$ , respectivamente. Ahora, si se define la matriz  $Y$  de orden  $(n,p)$  de la siguiente forma:

$$Y = \frac{f_{ij}}{\sqrt{f_{i\cdot} \cdot f_{\cdot j}}}$$

con los siguientes resultados:

$$1. X'D_n X = Y'Y = T^* \quad XD_p X' = YY' = S^*$$

2.  $Y'Y$  tiene los mismos valores propios diferentes de cero que  $YY'$ . Considere  $v_\alpha$  los eigenvectores de  $S^*$  y  $u_\alpha$  los vectores propios de  $T^*$ . Entonces

$$v_\alpha = \lambda^{-1/2} Y u_\alpha.$$

$$u_\alpha = \lambda^{-1/2} Y' v_\alpha.$$

Desarrollando esas relaciones se obtiene

$$v_{\alpha i} = \lambda_{\alpha}^{-1/2} \sum_j (f_{ij} f_i^{-1/2} f_{.j}^{-1/2}) u_{\alpha j} \quad 3.15$$

$$u_{\alpha i} = \lambda_{\alpha}^{-1/2} \sum_j (f_{ij} f_i^{-1/2} f_{.j}^{-1/2}) v_{\alpha j} \quad 3.16$$

Si  $\varphi_j$  designa la abscisa de los caracteres  $j$  sobre los ejes ( $\varphi_j$  es el  $j$ -ésimo componente de un vector  $B$ ), y si  $\psi_i$  designa la abscisa de las poblaciones  $i$  sobre el mismo eje, esas condiciones pueden ser escritas de la siguiente manera:

$$\begin{aligned} A &= D_n^{-1} F B & y & & a_i &= \sum_j \frac{f_{ij}}{f_{.i}} b_j \\ B &= D_p^{-1} F' A & y & & b_j &= \sum_i \frac{f_{ij}}{f_{.j}} a_i \end{aligned}$$

Las ecuaciones 3.15 y 3.16 son conocidas como las *relaciones de transición*. Note que  $\lambda_{\alpha}^{-1/2}$  es necesariamente menor o igual que 1; de otro modo, estas implicarían que cada uno de los dos conjuntos cubren un intervalo del eje que está estrictamente contenido en el intervalo cubierto por el otro conjunto.

De esta manera, si se ignora el coeficiente  $\lambda_{\alpha}^{-1/2}$ , el punto representativo de una configuración de puntos está sobre un eje el baricentro de los puntos representativos de la configuración correspondiente a el otro espacio.

Debido a las relaciones de transición, es fácil deducir de un conjunto (I ó J) las relaciones y las coordenadas del otro conjunto. De esta forma, el conjunto que tiene la dimensión más baja es diagonalizada, y entonces las coordenadas del otro conjunto son calculadas. Desde un punto de vista matemático, este es más importante porque un conjunto (frecuentemente I) tiene generalmente una dimensión más grande que el otro conjunto (frecuentemente J).

$$Y = y_{ij} = \frac{f_{ij}}{f_{.i} \cdot f_{.j}} = 1 + \sum_{\alpha=2}^p \lambda_{\alpha}^{-1/2} \varphi_{\alpha i} \phi_{\alpha j} \quad 3.17$$

Esta fórmula muestra como el análisis de correspondencias provee una forma de estudiar la dependencia y la independencia de los conjuntos I y J. La fórmula (3.17), algunas veces es llamada la identidad de Fisher. Esta ecuación toma en cuenta el valor propio uno y su eigenvector asociado.

### 3.5 Interpretación del análisis.

Interpretar un AC, significa dar un sentido a cada eje y a las proximidades entre puntos cuando son proyectadas en los ejes factoriales. No es suficiente calcular las coordenadas factoriales ni una representación gráfica.

Los resultados del AC son presentados en gráficas que representan la configuración de puntos en la proyección de puntos formados por los primeros ejes principales tomados dos a la vez. Como en el análisis de componentes principales es válido interpretar distancias entre los elementos de un conjunto de puntos. También es válido interpretar a posición relativa de un punto de un conjunto con respecto a todos los puntos del otro conjunto, excepto en casos especiales, es extremadamente erróneo interpretar la proximidad de dos puntos que corresponden a diferentes conjuntos de puntos.

El centro de gravedad G, localizado en el origen de los ejes corresponde al promedio de los perfiles de ambos conjuntos de puntos. Además se puede interpretar el ángulo entre puntuaciones fila y puntuaciones columna, tomando el centro de gravedad como el punto de partida, aplicando los siguientes criterios:

- a) Si el ángulo entre dos puntos es agudo (menor de  $90^\circ$ ) las dos características están correlacionadas.
- b) Contrariamente, si el ángulo es obtuso, las puntuaciones están correlacionadas negativamente (aproximadamente  $180^\circ$ ).
- c) Finalmente, si hay un ángulo recto, los dos puntos no interactúan, es decir, no hay asociación.

Recordando que los eigenvalores representan la dispersión de cada eje factorial.

Para interpretar los ejes, es útil calcular dos series de coeficientes para cada eje, esos coeficientes se aplican igualmente para los renglones y columnas de la matriz de datos; se calculan de la siguiente forma:

1. Contribuciones Absolutas. Indican la proporción de varianza explicada por cada variable en relación a cada eje principal: ésta proporción es calculada con respecto a el conjunto total de variables.

La norma de la matriz  $D_p^{-1}$  de los ejes  $u_\alpha$  es igual a 1:  $u_\alpha' D_p^{-1} v_\alpha = 1$ . Si  $u_\alpha = D_p \varphi_\alpha$ , de aquí, sigue que la norma  $D_p$  del factor  $\varphi_\alpha$  es

$$\sum_j f_{.j} \varphi_{\alpha j}^2 = 1 \quad (3.18)$$

La proyección del punto  $j$  de  $R^n$  sobre el eje principal  $\alpha$  es igual a

$$\sum_i \frac{f_{ij}}{f_{\cdot j}} \varphi_{\alpha i} = \sqrt{\lambda_\alpha} \varphi_{\alpha j} = \phi_{\alpha j} \quad (3.19)$$

La varianza del conjunto de puntos proyectados sobre el eje  $\alpha$  con respecto a  $G$  es igual a

$$\sum_j f_{\cdot j} \phi_{\alpha j}^2 = \sum_j f_{\cdot j} (\varphi_{\alpha j} \sqrt{\lambda_\alpha})^2 = \lambda_\alpha \quad (3.20)$$

El cociente  $\frac{\lambda_\alpha \varphi_{\alpha j}^2 f_{\cdot j}}{\lambda_\alpha} = f_{\cdot j} \varphi_{\alpha j}^2 = ca_\alpha(j)$ , representa la contribución absoluta del elemento  $j$  a el eje principal  $\alpha$ . Note que

$$\sum ca_\alpha(j) = 1 \quad (3.21)$$

La contribución absoluta de un elemento  $i$  de  $R^p$  a el eje principal  $\alpha$  es definida en el mismo sentido:

$$ca_\alpha(i) = f_{i \cdot} \varphi_{\alpha i}^2 \quad \text{con} \quad \sum ca_\alpha(i) = 1 \quad (3.22)$$

**2. Correlación Cuadrada.** Indica la parte de la varianza de una variable explicada por un eje principal. En  $R^n$ , el cuadrado de la distancia de la variable  $j$  al centro de gravedad es igual a:

$$d^2(j, G) = \sum_i \frac{1}{f_{i \cdot}} \left( \frac{f_{ij}}{f_{\cdot j}} - f_{i \cdot} \right)^2 \quad (3.23)$$

El cuadrado de la proyección de la variable  $j$  sobre el eje  $\alpha$  es igual a

$$d_\alpha^2(j, G) = \left( \sqrt{\lambda_\alpha} \varphi_{\alpha j} \right)^2 = \lambda_\alpha \varphi_{\alpha j}^2 \quad (3.24)$$

Note que

$$\sum d_\alpha^2(j, G) = d^2(j, G) \quad (3.25)$$

De la figura 4 se observa que la cantidad

$$\cos^2 \omega = \frac{d_\alpha^2(j, G)}{d^2(j, G)} = cr_\alpha(j) \quad (3.26)$$

representa la parte del eje  $\alpha$  en la posición del punto  $j$ . Esta es la correlación cuadrada de  $\alpha$  con elemento  $j$ .



$$\sum_{\alpha} cr_{\alpha}(j) = 1 \quad (3.27)$$

Además los elementos  $p$ , cuyas proyecciones sobre los  $\alpha$ -ésimos ejes son iguales a

$$\{ \sqrt{\lambda_{\alpha}} \varphi_{\alpha j} \}, j = \overline{1, p} \}$$

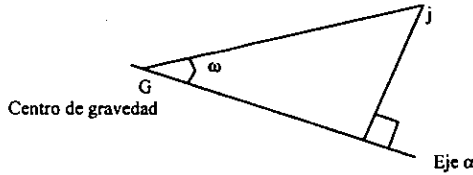


Fig. 4 Representación geométrica del centro de gravedad y una puntuación  $j$ .

pueden ser transformados a los  $n$  elementos del otro conjunto, cuyas proyecciones sobre el  $\alpha$ -ésimo eje son iguales a

$$\{ \sqrt{\lambda_{\alpha}} \varphi_{\alpha i} \}, i = \overline{1, n} \} \quad (3.28)$$

También, las coordenadas de la variable  $j$  sobre los nuevos ejes son proporcionales a  $u_{\alpha j} \sqrt{\lambda_{\alpha}}$ , y se tiene

$$\sum_j (u_{\alpha j} \sqrt{\lambda_{\alpha}})^2 = \lambda_{\alpha} \quad (3.29)$$

donde  $u_{\alpha}$  es unitaria. El número  $100 u_{\alpha j}^2$  expresa el porcentaje de los ejes  $\alpha$  explicados por la variable  $j$  (contribución absoluta). Por el otro lado, la distancia de una variable  $j$  a el origen en  $R^n$  es igual a 1. De este modo, se tiene, en bases de la ortonormalidad de los ejes principales,

$$\sum_{\alpha} \lambda_{\alpha} u_{\alpha j}^2 = 1 \quad (3.30)$$

el término  $\lambda_{\alpha} u_{\alpha j}^2$  expresa la parte de la varianza total de la variable  $j$  que esta asignada a el eje  $\alpha$  (correlación cuadrada).

Recuerde que los eigenvalores representan la dispersión de cada eje. Las coordenadas  $\{ \varphi_{\alpha j}; \alpha = \overline{2, p}; j = \overline{1, p} \}$  están dadas por la siguiente fórmula

$$\varphi_{\alpha j} = \sum \left( \frac{f_{jy}}{f_{\cdot j} \sqrt{f_{\cdot i}}} \right) v_{\alpha i} \quad (3.31)$$

donde  $v_{\alpha i}$  es el  $i$ -ésimo componente del eigenvector  $v_{\alpha}$  de la matriz  $S^* = YY'$ .

Gráficamente, el ángulo más pequeño entre una puntuación y un eje factorial, con el perfil promedio como origen, estará cercano a la unidad y será la correlación de ese eje. Si el ángulo es aproximadamente a  $90^\circ$ , la correlación del eje será cero.

### Puntos explicativos.

Los puntos explicativos de un eje  $\alpha$  son los puntos  $j$  de  $N(J)$  cuyas contribuciones son las más altas que aquellas de los otros puntos de  $N(J)$ . Cada uno de los ejes depende sobre los puntos  $j$  que contribuyen a la mayor varianza total del eje, acorde a lo siguiente:

$$\sum_j f_j \phi_{\alpha j}^2 = \lambda_\alpha \quad (3.32)$$

De esa forma, los puntos explicados de un eje  $\alpha$  son seleccionados cuando su contribución, denotadas por  $ctr_\alpha(j)$ , es más alta que el promedio de las contribuciones totales. El conjunto de puntos explicativos es dividido en dos partes; la parte izquierda contiene los puntos explicativos cuyas coordenadas son negativas, y la parte derecha contiene los puntos cuyas coordenadas son positivas.

### Puntos explicados.

Los puntos explicados por un eje  $\alpha$  son aquellos puntos  $j$  de  $N(J)$  cuyas contribuciones a la excentricidad son mas altos que aquellos de los otros puntos de  $N(J)$ . Un punto  $j$  puede ser explicado sin ser un punto explicativo (del mismo eje), por los valores de  $f_j$  en la fórmula anterior.

Las contribuciones a la excentricidad son semejantes a el coeficiente de correlación cuadrado. Como con los puntos explicativos, los puntos explicados pueden ser seleccionados acorde el valor más alto de  $cor_\alpha(j)$ . Entonces, el conjunto de puntos explicados es dividido en dos partes, la parte izquierda contiene los puntos cuyas coordenadas son negativas y la parte derecha las coordenadas positivas.

Las contribuciones relativas de cada eje  $\alpha$  están dados en porcentaje de varianza de los puntos  $i$  de  $N(I)$ , acorde a la siguiente fórmula.

$$\begin{aligned} ctr_\alpha(i) &= \frac{f_i \phi_{\alpha i}^2}{\lambda_\alpha} & \alpha &= \overline{2, p} \\ ctr_\alpha(j) &= \frac{f_j \phi_{\alpha j}^2}{\lambda_\alpha} & i &= \overline{1, n} \quad j = \overline{1, p} \end{aligned} \quad (3.33)$$

Sin embargo, los valores  $\lambda_\alpha$  pueden tener una variación alta, haciendo difícil la comparación de los valores de  $ctr_\alpha(i)$  para valores diferentes de  $\alpha$ . En este caso, es predecible calcular las

contribuciones absolutas como sigue:

$$\begin{aligned} ca_{\alpha}(i) &= f_i \cdot \phi_{\alpha}^2 \\ ca_{\alpha}(j) &= f_j \cdot \phi_{\alpha}^2 \end{aligned} \quad (3.34)$$

de las siguientes formulas.

$$\begin{aligned} \sum_i \sum_{\alpha=2}^p f_i \phi_{\alpha}^2 &= M^2(N(I)) \\ \sum_j \sum_{\alpha=2}^p f_j \phi_{\alpha}^2 &= M^2(N(J)) \end{aligned} \quad (3.35)$$

### Calidad de explicación.

Para finalizar la interpretación básica de análisis de correspondencias desde un punto de vista numérico, la calidad de explicación de los ejes necesita ser estudiada, también como la calidad de explicación de los puntos  $i$  de  $N(I)$  y  $j$  de  $N(J)$  por los ejes.

### Explicación global de calidad.

Esta dada por la suma de los porcentajes de varianza de los ejes considerados como interpretables y se denota por  $qge_s$ . Sea  $s \leq p$  el número de esos ejes. Entonces,

$$qge_s = \sum_{\alpha=2}^s \tau_{\alpha} = \frac{\sum_{\alpha=2}^s \lambda_{\alpha}}{\sum_{\alpha=2}^p \lambda_{\alpha}}$$

### Explicación local de calidad

Para cada punto  $i$  de  $N(I)$  y  $j$  de  $N(J)$ , la explicación de calidad, denotada por  $qlt_s(i)$  y  $qlt_s(j)$  puede ser calculada en un espacio dado, con rango  $s$ :

$$\begin{aligned} qlt_s(i) &= \sum_{\alpha=2}^s cr_{\alpha}(i) \\ qlt_s(j) &= \sum_{\alpha=2}^s cr_{\alpha}(j) \end{aligned}$$

### 3.6 Prueba de independencia total.

Para poder dar significancia estadística a la tabla de correspondencia, se necesita comparar la distribución observada de elementos en cada casilla de la tabla. Lo más simple es ver cuál sería el contenido de la celda si los elementos están distribuidos semejantemente entre sí. De lo anterior se plantea si existe independencia estocástica entre caracteres y poblaciones, entonces:

$$p_{ij} = P(A_i \cap B_j) = P(A_i)P(B_j) = p_i \cdot p_{.j}$$

y  $\chi^2 = k(\lambda_2 + \dots + \lambda_p)$  sigue la distribución ji-cuadrado con  $(n-1)(p-1)$  grados de libertad. Posteriormente se acepta la hipótesis

$$H_0: \lambda_2 = \dots = \lambda_p = 0$$

si esta  $\chi^2$  es no significativa. En tal caso no hay lugar para la representación por AC, pues todas las poblaciones se distribuyen igualmente a lo largo de los caracteres, la distancia ji-cuadrado entre cada par de poblaciones es nula y la representación queda reducida a un sólo punto. Un AC en el que se observe una nube de puntos carece de significancia estadística si  $\chi^2$  no es significativa. Los puntos observados sólo son resultado de las fluctuaciones del muestreo.

### 3.7 Descomposición aditiva de la ji-cuadrado.

Para estudiar la dimensión significativa  $d$ , se puede utilizar el método que consiste en descomponer la ji-cuadrado total en suma de ji-cuadrados correspondientes a los valores propios. La descomposición es la siguiente:

	Valor propio	ji-cuadrado	g.l.
	$\lambda_2$	$\chi^2_2 = K \lambda_2$	$n + p - 3$
	$\lambda_3$	$\chi^2_3 = k \lambda_3$	$n + p - 5$
	$\lambda_p$	$\chi^2_p = k \lambda_p$	$n + p - (2p - 1) = n - p + 1$
Total	traza $(Y'Y) - 1$	$\chi^2$	$(n-1)(p-1)$

El eje factorial correspondiente a  $\lambda_i$  no es significativo si  $\chi^2_i = N \lambda_i$  con  $(n + p - 2i + 1)$  grados de libertad, no es significativa. En general, si el valor ji-cuadrado residual

$$\chi^2_{(d)} = N (\lambda_2 + \dots + \lambda_d) \quad \text{con } (n+p-3 + \dots + 2m+1) \text{ g.l.}$$

es significativo, mientras que la ji-cuadrado residual

$$\chi^2_{(p-d)} = \chi^2_p - \chi^2_{(d)} = N (\lambda_{d+1} + \dots + \lambda_p) \text{ con } (n+p - 2m - 1 + \dots + n-p+1) \text{ g.l.}$$

no es significativa, es suficiente una representación en dimensión  $d$ . Este criterio se puede interpretar como una descomposición aditiva de la tabla de frecuencias  $F = (f_{ij})$

$$F = F_d + F_{(p-d)}$$

en una tabla  $F_d$ , cuya variabilidad está explicada por  $\lambda_2 + \dots + \lambda_d$ , y una tabla  $F_{(p-d)}$  explicada por  $(\lambda_{d+1} + \dots + \lambda_p)$ . Si  $\chi^2_{(p-d)}$  no es significativa, existe independencia completa en la tabla residual  $F_{(p-m)}$ , y se puede aceptar la hipótesis

$$H_0: \lambda_{d+1} = \dots = \lambda_p$$

### Extensiones y limitaciones.

El análisis de correspondencias es frecuentemente aplicado al análisis de tablas cruzadas con dos variables categóricas, que no es una tabla de contingencia verdadera. Si se respeta el criterio de homogeneidad (todas las celdas son mediciones de las mismas unidades) y exhaustivamente cada caso es clasificado en una y sólo una categoría de cada una de las dos variables.

El análisis de correspondencias es básicamente un doble análisis de componentes principales, uno realizado sobre las filas y el segundo sobre las columnas de la matriz original. Es una herramienta orientada como ayuda en la generación de teorías, facilitando la elección previa de las características o modalidades más significativas, frente a otros criterios de tipo teórico. El análisis de correspondencias múltiple es empleado para el análisis de cuestionarios

### Ejemplos.

Aplicaciones en la investigación empleando análisis de correspondencias.

**Paleontología.** Un estudio de tipología por paleontólogos sobre una muestra de 349 esqueletos de equinos (caballos, cebras, etc....) para tratar de clasificarlos<sup>35</sup>.

**Docencia.** Las funciones del profesor universitario analizadas por sus protagonistas. Un estudio atendiendo al grupo de titulación y los años de experiencia, en la Universidad de Sevilla<sup>32</sup>.

**Economía.** Impacto de la crisis asiática en América Latina<sup>37</sup>.

**Genética.** Identificación de los genes empleando codonw<sup>38</sup>.

**Software.**

Existen actualmente una gran gama de software dedicados para análisis estadísticos, el SPSS es uno de ellos, pero tiene una desventaja. Los gráficos son de una calidad muy baja, además son muy pequeñas y no se pueden manipular libremente, por lo que se recomienda emplear otro paquete para la elaboración de las gráficas, como el MATLAB, MATHEMÁTICA; S-Plus y otros.

**Notas históricas.**

El análisis de correspondencias fue desarrollado simultáneamente a principios de los 50's por Hayashi en Japón y por Benzecri en Francia. El AC es empleado ampliamente en Japón, pero este es conocido como Hayashi III<sup>36</sup>.

El AC es un tipo de análisis de componentes principales sobre una distancia especial llamada ji-cuadrado. cuya principal propiedad es la posibilidad de una representación simultánea de las variables y las poblaciones. El método fue introducido por J. P. Benzecri en 1965; destacan también los resultados de la tesis doctoral de B. Escofier - Cordier (1965), que conjeturó algunos teoremas que serían demostrados más tarde. El análisis de tablas de contingencia mediante el estudio de los valores y vectores propios de  $Y'Y$ , a fin de establecer relaciones de dependencia entre las columnas (o filas) de la tabla, había sido propuesto por E. J. Williams (1952). La representación simultánea había sido estudiada primeramente por Coombs (1950) para una solo dimensión y extendida al caso multidimensional por Bennett y Hays (1960), estableciendo relaciones ordinales no métricas a lo largo de las dimensiones. Hill (1973) ha pretendido dar una versión propia del AC con el nombre de "Reciprocal Averaging".

La dualidad entre caracteres y poblaciones y la representación simultánea había sido advertida por Kendall (1957) y estudiada por Gower (1966). Lefebvre (1976), Chady, Glemarec y Laurec (1976) y otros autores. La representación simultánea es posible, en general, cuando la matriz de covarianzas entre las variables es de la forma  $YY'$ . La relación entre los vectores propios

$$Yv_i = w_i \quad Y'w_i = v_i$$

es la clave de esta representación. Sin embargo, para variables cuantitativas, la obtención de  $Y$  por transformación de los datos originales puede formar la escala de media de los ejes factoriales. El AC es la técnica en donde mejor queda esta dualidad, por la perfecta simetría que existe entre poblaciones y caracteres.

Sin embargo, el AC queda limitado al análisis de tablas de contingencia y de datos de presencia - ausencia.

#### 4. RESULTADOS.

Se reviso un total de 5017 registros de expedientes, de los cuales sólo 3505 (70%) fueron analizados, los restantes son valores incompletos o perdidos (missing). En la tabla 1 se puede apreciar como los patrones de los perfiles fila (neoplasias malignas) son completamente distintos entre si y ocurre lo mismo con los patrones columna (escolaridad de la madre). Lo que indica que las modalidades estudiadas tanto en el sentido de las filas como de las columnas están bastante bien diferenciadas. Nótese que las leucemias son las que presentan mayor frecuencia (en toda la población) en el nivel de escolaridad primaria o menor de las madres 1709, 49% (primaria 16.5%, secundaria 10.6%, bachillerato 6.4% y licenciatura y más 1.7%), seguidas de los linfomas 636, 18.1% ( $\leq$ primaria 9.8%,  $\leq$ secundaria 4.8%, bachillerato 2.7% y  $\geq$ licenciatura 0.9%) predominando más en las madres cuya escolaridad es baja, y ocupando el tercer sitio de importancia se encuentran los tumores del sistema nervioso central (SNC) con 347 (9.9%), predominando de igual forma en las madres de escolaridad baja.

**Tabla 1.** Conjunto de datos registrados durante el periodo 1980-1992 de cáncer en niños menores de 15 años según escolaridad de la madre. Fuente Unidad de Investigación en Epidemiología Clínica del CMN SXXI. IMSS. 1998.

Tipo de cáncer	Escolaridad materna				Profesionistas	Marginal total
	<Primaria incompleta	Primaria completa Secundaria incompleta	Secundaria completa Técnico y bachillerato			
Leucemia	580	372	226	61	1239	
Linfomas	343	167	94	32	636	
SNC	175	99	59	14	347	
SNS	36	31	20	8	95	
Retinoblastoma	153	77	63	14	307	
T. Wilms	92	62	36	8	198	
T. Hepáticos	19	19	12	3	53	
T. Óseos	107	47	36	17	207	
Tej. Blandos	81	32	33	12	158	
C. Germinales	80	42	32	12	166	
Carcinomas	22	16	5	6	49	
No clasificados	21	14	15	0	50	
<b>Marginal total</b>	<b>1709</b>	<b>978</b>	<b>631</b>	<b>187</b>	<b>3505</b>	

La tabla 2, expresa cada una de las dimensiones de las soluciones alcanzadas (no triviales), con la cantidad de la nube de inercia total explicada. De esta forma, si se pone en relación la inercia de cada dimensión con la inercia total, se obtendrá la proporción explicada de la información original (en forma bruta y acumulada, columnas 6 y 7). De este modo, se puede decir que la información de la tabla de contingencia se explicaría por la primer dimensión el 45.5%, mientras en la segunda, la explicación sería del 30.5% y en la tercera del 24.0%. Como con las dos primeras dimensiones

se tendría explicada el 76.0% de la información inicial, se hace plausible reducir la dimensionalidad a dos únicos componentes. Los datos de la siguiente tabla indican que existe una relación estadísticamente significativa entre las dos variables  $p < 0.017$  con 33 grados de libertad, es decir; se verifica la relación entre las dimensiones estudiadas. Por tanto, se encuentra una situación aceptable de análisis de correspondencias.

Tabla 2. Resumen de la inercia explicada por las diferentes dimensiones alcanzadas

Dimensión	Valor Singular	Inercia	Ji cuadrado	Sig.	Proporción de Inercia		Confianza del valor singular	
					Explicado por	Acumulado	Desviación estándar	Correlación
1	.083	.007			.455	.455	.017	0.008
2	.068	.005			.305	.760	.019	
3	.060	.004			.240	1.000		
<b>Total</b>		.015	52.546	.017(a)	1.000	1.000		

a 33 grados de libertad.

De acuerdo a la fórmula  $\chi^2 = Nk$ , se esta en condiciones de establecer la nube de inercia del modelo, que se puede interpretar como la cantidad de la información recogida en la tabla de contingencias, así la nube de inercia es 0.015

Aplicando el método de descomposición aditiva de la ji-cuadrado para las distintas dimensiones se obtiene:

Dimensión	$\chi^2$	g.l.
1	$3505 \cdot 0.007 = 24.54$ ,	$12+4-2(1)-1 = 13$
2	$3505 \cdot 0.005 = 17.53$ ,	$12+4-2(2)-1 = 11$
3	$3505 \cdot 0.004 = 14.02$	$12+4-2(3)-1 = 9$

Descomposición aditiva de la  $\chi^2$ , mediante la fórmula  $k \sum \lambda_{\alpha}$

El autovalor (valor singular, eigenvalor) expresado en la tabla anterior, hace referencia a la cantidad de inercia explicada por la dimensión.

Así, se puede comprobar como el autovalor de la primera dimensión (0.083) es el resultado de operar  $\sqrt{0.007} = 0.83$ . El valor 0.083 (renglón 1 de la dimensión con columna valor singular) se puede expresar en términos de porcentaje, así el 8.3%, es la correlación entre las puntuaciones fila y columna para la primera dimensión, el 6.8% la correlación de las puntuaciones para la segunda dimensión y 6.0% la correlación en la tercera dimensión, únicamente se observan correlaciones positivas muy bajas.



Como se ha visto, dado que las modalidades o categorías, tienen pesos distintos, lo cual indica que pueden tener mayor o menor influencia entre las relaciones de las variables neoplasias malignas y escolaridad materna estudiadas. Para medir la influencia que tiene cada una de las categorías en la relación final, se utilizan los índices de contribución.

**Tabla 3. Resumen de la contribución de las puntuaciones fila, neoplasias malignas (a)**

Tipo de cáncer	Contribución								
	Puntuación en la dimensión			Inercia	Del punto a la inercia de la dimensión		De la dimensión a la inercia del punto		Total
	Masa	1	2		1	2	1	2	
Leucemia	.353	.164	-.034	.001	.114	.006	.807	.029	.836
Linfomas	.181	-.285	.240	.002	.178	.154	.541	.314	.854
SNC	.099	.042	.224	.000	.002	.073	.032	.762	.794
SNS	.027	.318	-.831	.001	.033	.277	.151	.845	.997
Retinoblastoma	.088	.027	.162	.001	.001	.034	.008	.231	.238
Wilms	.056	.282	.057	.000	.055	.003	.784	.026	.810
Hepático	.015	.752	-.574	.001	.103	.074	.677	.323	1.000
Óseos	.059	-.498	-.224	.000	.017	.042	.753	.125	.878
Tejidos blandos	.045	-.396	-.139	.002	.085	.013	.355	.036	.391
Germinales	.047	-.174	-.244	.000	.017	.042	.246	.399	.645
Carcinomas	.014	-.553	-1.073	.002	.052	.238	.188	.579	.767
No clasificados	.014	1.027	.445	.002	.182	.042	.610	.094	.704
<b>Total</b>	<b>1.000</b>			<b>.015</b>	<b>1.000</b>	<b>1.000</b>			

a Normalización simétrica.

Se distingue en primer lugar la contribución relativa de un factor a la inercia de los elementos, que no es nada más que la cantidad relativa de inercia de un elemento (en forma de proporción) que es captada por cada una de las dimensiones alcanzadas. Observando la tabla 3, se puede verificar que las puntuaciones fila, leucemias, linfomas, Tumores del SNS, Tumor de Wilms, Tumores Hepáticos, Tumores Óseos y no clasificados (columna 6, tabla 3); se explican mejor por la primera dimensión, mientras que los Tumores del SNC, Retinoblastomas, Tejidos blandos, Células Germinales y Carcinomas tienen mayor explicación por la segunda dimensión (columna 7).

Para los estratos de escolaridad de la madre, tabla 4, la estructura es semejante a la vista en las categorías de neoplasias malignas, se puede apreciar que las categorías (secundaria incompleta, completa, bachillerato) se distribuyen uniformemente y se explican casi de la misma forma por las dos dimensiones. Mientras que las madres con nivel primaria terminada o menor y profesionistas o más, tienen mayor explicación por la segunda dimensión.

Las contribuciones relativas de los puntos a la dimensión, es la parte de la varianza explicada (en forma de proporción) de cada elemento fila o columna en cada una de las dimensiones. Observando los datos de la tabla 3, se puede determinar que las modalidades fila contribuyen a la formación de la primera dimensión mientras en la tabla 4, se puede observar que las modalidades

1, 2 y 3 forman la primera dimensión, para la segunda dimensión las modalidades 2 y 3, secundaria y preparatoria, la forman básicamente.

Tabla 4. Resumen de la contribución de las puntuaciones a la variable escolaridad de la madre (a)

Escolaridad materna	Puntuación en la dimensión			Contribución					
	1	2	Inercia	Del punto a la inercia de la dimensión		De la dimensión a la inercia del punto			
				1	2	1	2	Total	
<Primaria incompleta	.499	-.209	.187	.003	.258	.253	.603	.396	.999
Primaria completa									
Secundaria incompleta	.279	.288	-.113	.004	.205	.009	.539	.067	.607
Secundaria completa Técnico y bachillerato	.180	.307	-.057	.004	.205	.009	.386	.011	.397
Profesionistas	.053	-.630	-.932	.005	.256	.686	.357	.640	.996
<b>Total</b>	<b>1.000</b>			<b>.015</b>	<b>1.000</b>	<b>1.000</b>			

a Normalización simétrica.

La representación gráfica (Figura 4.1) dentro del análisis de correspondencias ayuda a determinar intuitivamente entre que modalidades se produce la relación. La gráfica muestra una forma simplificada, involucrando las 12 modalidades (categorías) del cáncer y los 4 estratos de escolaridad materna. Los números mostrados sobre los ejes representa la porción de varianza explicada por los ejes, además se pueden interpretar como porcentajes, de este modo, los 2 primeros ejes explican un 76.0% de la varianza total. Este porcentaje da una idea conservativa de la porción de información explicada por los ejes principales. Es conservativo porque hay solo una forma parcial de medir la información.

El centro de gravedad G, localizado en el origen de los ejes, corresponde a el perfil promedio de ambos conjuntos de puntos. De este modo, el perfil o punto, Tumor de Wilms, el cual se encuentra cerca del centro, tiene una distribución no diferenciada entre las neoplasias.

De estas dos variables a simple vista se puede determinar asociación entre las neoplasias malignas y la escolaridad de la madre. Considerando las posiciones relativas a uno de los puntos que pertenecen a la misma nube de puntos. Se observa como las modalidades de neoplasias malignas y el grado de escolaridad presentan distribuciones similares, la elipse encierra tal semejanza. Analizando las distribuciones con respecto a las categorías del cáncer y escolaridad materna se puede notar que la distribución de los linfomas se encuentra asociada a el nivel de escolaridad primaria o menos y a profesionistas, con un nivel de significancia de  $p < 0.017$ , la significancia en análisis de correspondencias se interpreta como "analizar a profundidad".

En relación a la escolaridad del padre, tabla 7, las frecuencias más importantes son las leucemias, 14.2%, las dos primeras dimensiones explican el 89.4% de los datos con una significancia de  $p < 0.011$  con 33 grados de libertad. La representación gráfica de esta asociación se puede observar en la figura 2 (Anexo A), es clara la relación entre ambas variables y se confirma dado que es mayor su significancia comparado con la gráfica de escolaridad materna.

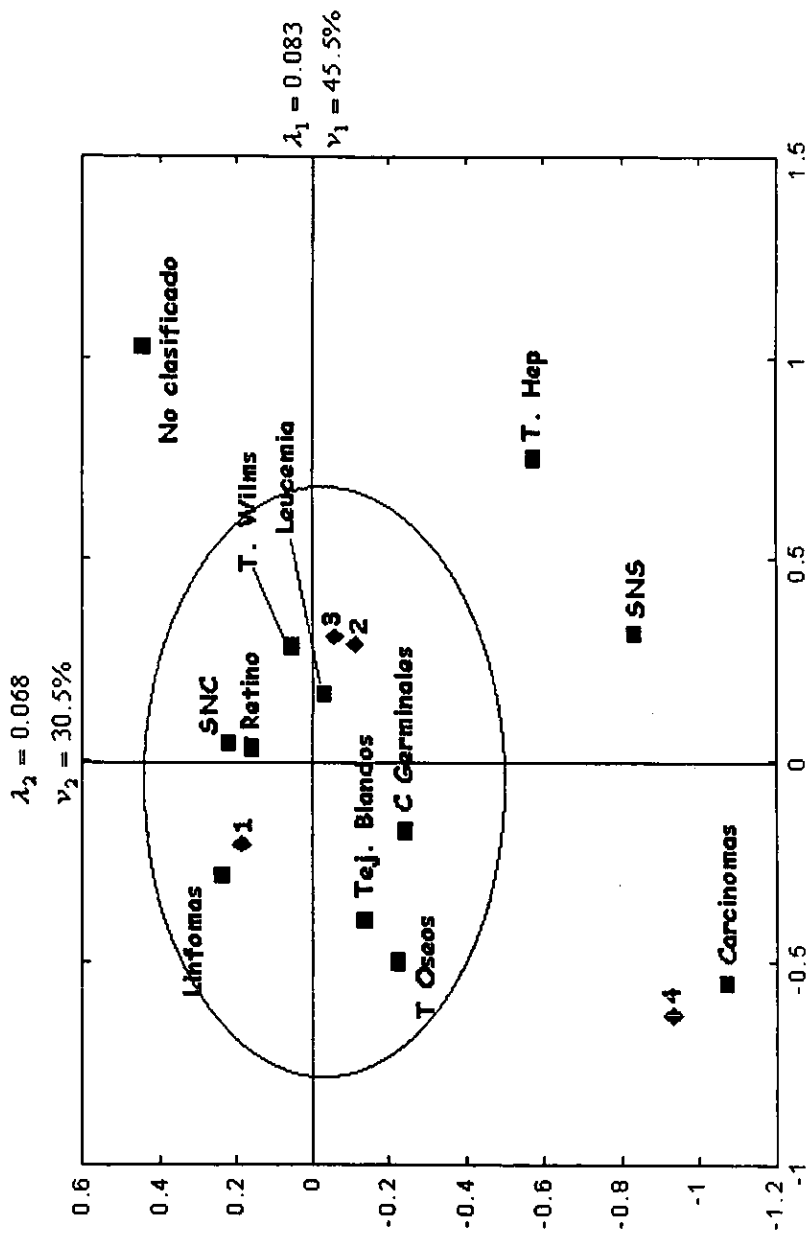


Figura 4.1. Análisis de neoplasias malignas en niños menores de 15 años, representado por un cuadro. Representación del espacio factorial (1,2), donde 1: Primaria incompleta o menos, 2: Primaria completa y secundaria incompleta, 3: Secundaria completa, técnico y bachillerato, y 4: Profesionistas y más, representado por un rombo.

**Tabla 7. Conjunto de datos registrados durante el periodo 1980-1992 de cáncer en niños menores de 15 años según escolaridad de la madre. Fuente Unidad de Investigación en Epidemiología Clínica del CMN SXXI. IMSS. 1998.**

Tipo de cáncer	Escolaridad paterna				Marginal Total
	<Primaria incompleta	Primaria completa Secundaria incompleta	Secundaria completa Técnico y bachillerato	Profesionistas	
Leucemia	496	390	242	111	1239
Linfomas	307	166	107	56	636
SNC	142	110	60	35	347
SNS	25	30	28	12	95
Retinoblastoma	147	78	62	20	307
T. Wilms	78	58	38	24	198
T. Hepático	22	15	9	7	53
T. Óseos	102	51	27	27	207
Tej. blandos	65	43	33	17	158
C. Germinales	74	45	33	14	166
Carcinomas	17	17	9	6	49
No clasificados	16	16	14	4	50
<b>Marginal Total</b>	<b>1491</b>	<b>1019</b>	<b>662</b>	<b>333</b>	<b>3505</b>

Con respecto a la ocupación materna y paterna, las leucemias se presentan en el estrato No clasificado y Personal Administrativo, respectivamente. Se aprecia en la tabla 13 que la ocupación de las madres no clasificada, se conglomeran los diferentes tipos de neoplasias. Con respecto al padre, los estratos Personal Administrativo (Per. Admón.) y Directivos y Funcionarios Públicos (Dir. Y Fun pub) prevalecen más las neoplasias. La significancia de las variables ocupación de la madre y del padre es  $p < 0.378$ ,  $p < 0.004$ , respectivamente. Gráficamente, existe una clara asociación entre la ocupación del padre, más no así, la ocupación de la madre, ya que esta forma un punto en la nube de inercia (figura 4 y 3, respectivamente).

Según el sexo, el cáncer se presenta más en niños que en niñas (tabla 25, Anexo A). Para los grupos de edad, las neoplasias más frecuentes se presentan en los niños mayores de un año, con las leucemias como la neoplasia más predominante (tabla 31, Anexo A).

Realizando un análisis sobre las variables, Institución de salud y escolaridad de la madre y padre, se obtienen aspectos socio-económicos de interés, se nota una clara diferencia entre el Instituto Nacional de Pediatría (INP) y el Hospital Infantil de México, donde se agrupan los niños cuyos padres poseen una escolaridad mínima, el Hospital 20 de Noviembre atiende a los padres con nivel de estudios profesionales y superiores. El CMN Siglo XXI y CMN "La Raza" agrupan los padres cuyo nivel de escolaridad es primaria completa y secundaria incompleta. En relación al lugar de residencia, la incidencia mayor se presenta en el D.F. y en el Estado de México, y los estados Baja California, Quintana Roo y Yucatán presentan la menor incidencia de esta enfermedad, cáncer.

Tabla 13. Conjunto de datos obtenidos de niños con cáncer asociado a la ocupación materna, durante el periodo 1980-1992

Tipo de neoplasias	Ocupación de la madre								Marginal Total
	Profesionistas	Com. y Vend.	Trab. de servicios	Trab. agrícolas	Dir. y Fun.	Per. Admón.	Hogar	No clasificado	
Leucemia	39	13	22	0	1	57	24	1083	1239
Linfomas	28	7	7	1	1	25	12	555	636
SNC	8	7	8	0	0	12	8	304	347
SNS	5	6	2	1	0	4	1	76	95
Retinoblastoma	15	4	3	0	0	10	3	272	307
T. Wilms	9	4	2	0	0	15	1	167	198
T. Hepático	3	0	0	0	0	3	1	46	53
T. Óseos	13	2	3	1	1	12	1	174	207
Tej. blandos	9	2	3	0	0	5	2	137	158
C. Germinales	5	3	3	0	0	9	2	144	166
Carcinomas	5	0	1	0	0	1	1	41	49
No clasificados	1	3	0	0	0	4	2	40	50
<b>Marginal Total</b>	<b>140</b>	<b>51</b>	<b>54</b>	<b>3</b>	<b>3</b>	<b>157</b>	<b>58</b>	<b>3039</b>	<b>3505</b>

Tabla 19. Frecuencias absolutas del conjunto de datos, neoplasias malignas y ocupación paterna.

Tipo de cáncer	Ocupación paterna							Marginal Total
	Profesionistas	Trab. de servicios	Trab. agrícolas	Dir. y Fun pub	Per. Admón.	Hogar	No clasificado	
Leucemia	101	113	38	317	616	49	8	1239
Linfomas	53	42	24	212	267	35	3	636
SNC	36	22	11	115	143	16	4	347
SNS	10	4	3	24	48	6	0	95
Retinoblastoma	19	22	5	128	116	13	4	307
T. Wilms	18	14	7	68	84	5	2	198
T. Hepático	4	8	1	17	22	1	0	53
T. Óseos	24	17	3	72	78	9	4	207
Tej. blandos	21	13	1	54	63	3	3	158
C. Germinales	11	14	7	41	88	4	1	166
Carcinomas	4	2	0	14	27	1	1	49
No clasificados	4	3	1	15	27	0	0	50
<b>Marginal Total</b>	<b>305</b>	<b>274</b>	<b>101</b>	<b>1077</b>	<b>1579</b>	<b>139</b>	<b>30</b>	<b>3505</b>

## 5 DISCUSIÓN.

En este trabajo se ha presentado un análisis sobre la relación entre neoplasias malignas y la escolaridad y ocupación de los padres, cuyos hijos padecen ésta mortal enfermedad, residentes en la República Mexicana. El objetivo principal fue determinar dicha relación, un segundo objetivo fue introducir una técnica estadística multivariada para la exploración de asociaciones complejas en la epidemiología clínica.

Se ha señalado que las neoplasias malignas en la población mexicana son ya un problema importante de salud pública. Sin embargo, no existe suficiente información para tener un panorama claro e integral de estas enfermedades en la población<sup>25</sup>.

Un estudio realizado en la Ciudad de México, mostró que en 1991 existía un incremento en la incidencia de cáncer infantil, una tasa de  $70 \times 10^6$ , aunque posiblemente esta tasa fue subestimada. Además, aquellos pacientes pediátricos cuya residencia se encontraba en la región sureste, presentaron mayor incidencia del cáncer<sup>26</sup>.

A nivel mundial se ha tratado de mejorar la calidad de vida de los pacientes con cáncer, empleando tratamientos menos dolorosos en el niño, a pesar de esto la tendencia de cáncer sigue en aumento. La sobrevida de pacientes menores de 15 años ha sido influenciado positivamente por estudios de cooperación multidisciplinaria, empleando cirugía, quimioterapia y radioterapia, pero uno de 3 niños con cáncer sucumbe a su condición. La identificación de características biológicas y genéticas como factores de riesgo para varios tumores ha conducido a cambiar en tratamientos más sofisticados<sup>27</sup>.

Este análisis exploratorio descriptivo, ha mostrado que la ocupación del padre es más significativa que la ocupación de la madre,  $p < 0.004$  y  $p < 0.378$ , respectivamente. En cuanto a la escolaridad del padre y de la madre presentan un valor  $p$  de  $< 0.011$   $p < 0.017$ , respectivamente. En estudios epidemiológicos se menciona que estas variables poseen una razón de momios (OR) mayor que 3.0. En esta técnica no se calcula el OR, ya que lo que se pretende es esclarecer las interacciones entre variables, sin embargo, se puede decir teóricamente, que la ocupación del padre deberá ser de interés para estudios posteriores. Por otro lado, el lugar de residencia de los niños tienen mayor asociación con las neoplasias malignas, posiblemente el tipo de composición química del aire sea un factor de vital importancia para el desarrollo de un cáncer. Se dice que las características de ambos padres influyen en el desarrollo de una neoplasia maligna, sin embargo; no se especifica que características o factores son necesarios o suficientes para tal desarrollo. En la epidemiología clínica, la escolaridad de ambos padres es importante en cuanto al diagnóstico temprano del cáncer, y así poder brindar un tratamiento que puede combatir el cáncer a tiempo.

A pesar de existir consistencia, en los resultados, con algunas características epidemiológicas de las neoplasias malignas en niños en estudios realizados en México y en otros países, recuérdese que en análisis de correspondencias la significancia, no es estadística, sino quiere decir "analizar a profundidad", con esto en mente, queda abierta la exploración de diversos factores biológicos, químicos y ambientales que determinen relaciones más profundas y que muestren una realidad más clara.

## 5.1 DIRECCIONES FUTURAS

Las matemáticas, como se ha visto a través de la historia, ha sido un actor principal en el desarrollo y generación de nuevas teorías y conocimiento. De este modo, el nuevo milenio requiere la aplicación de las matemáticas para explicar fenómenos considerados complejos, y con la ayuda de la computación, describir y explicar dichos fenómenos.

En diferentes partes del mundo, mayormente en los países desarrollados, existen departamentos de matemáticas, estadística y probabilidad, investigación de operaciones y otros campos que atañen a las matemáticas. En esos países se aplican métodos y teorías que se enseñan en la nuestro país. Sólo que no se aplican por ser complejos, al momento de plantear un problema o modelarlo.

Las matemáticas aplicadas tienen una tarea ardua y compleja por llevar a cabo, la generación de modelos matemáticos capaces de explicar fenómenos como el cáncer y el SIDA, su evolución, desarrollo, o en términos matemáticos sus estados transitorios y absorbentes. Su contribución será en beneficio del desarrollo de la humanidad, por ejemplo, el descubrimiento de la geometría fractal llevo a comprender los errores de transmisión de la información, a través de los ordenadores, una de las aplicaciones más notorias es la telefonía celular, al transmitir grandes cantidades de información a los satélites mediante un fragmento de esa misma información.

Se pueden enumerar muchos ejemplos donde la interferencia del pensamiento matemático a influido en el desarrollo de la ciencia y por ende de la humanidad, o del otro lado de la moneda, ha sido para su exterminio. Tal vez sea, una visión poco clara, o incluso futurista, pero si se quiere crecer siempre hay que proyectar o estimar las situaciones futuras, en base a hechos pasados y presentes.

Una herramienta clave en el entendimiento de fenómenos es la simulación. Actualmente, en el área médica, es empleada para describir poblaciones más realista de pacientes quienes, al inicio de la simulación, son en diferentes etapas de la enfermedad, nuevos pacientes. Este provee una descripción más comprensiva de la evolución de la enfermedad y del paciente mismo<sup>28</sup>, en otras áreas sirve para modelar el comportamiento de colas, procesos para el control de calidad de productos, y el área donde más se desarrolla es en la industria militar y el área espacial.

La conjunción de las matemáticas y de la informática ha demostrado ser más que efectiva, un buen ejemplo de ello, son los paquetes estadísticos que involucran no sólo conceptos matemáticos sino diferentes métodos numéricos para poder realizar cálculos que manualmente se tardaría meses realizarlos manualmente, otro ejemplo son los software para generar imágenes en tres dimensiones, el desarrollo de esos programas implica el conocimiento de cálculo diferencial y álgebra lineal.

Por último, sólo el tiempo permitirá conocer si el desarrollo de fenómenos (enfermedades), como el cáncer, tienen una explicación parsimoniosa y única, que factores intervienen en su desarrollo, explorar aspectos que se han dejado sin analizar, tal vez son los más importantes o quienes contribuyen en el desarrollo de una enfermedad como factor de riesgo.

O tal vez, esperar a que alguien venga a demostrar que siempre hay una explicación sencilla o compleja para todos los "fenómenos".

## CONCLUSIONES

En el área médica, es el punto central de esta aplicación, se emplea para la realización de una investigación, el método científico, además de la estadística para validar los resultados obtenidos, y poder tener un sustento “científico” para inferir una conclusión acorde al fenómeno o problema en estudio. Sin embargo, la medicina no es considerada cien por ciento ciencia, ya que no tiene mediciones fijas ni los tamaños de muestra necesarios para poder inferir acerca de una población en general. Pero cuentan con validez dentro de su población. Se podría discutir más a profundidad este tema, pero no es el caso. También los médicos emplean diferentes variables que son consideradas confusoras o modificadoras de una respuesta. Por ejemplo, en la epidemiología la edad, el sexo, la escolaridad y ocupación de un individuo es considerado un factor de confusión, ya que cierta característica de un individuo con alguna enfermedad no se presenta de la misma forma en hombre que en mujeres ni en las mismas edades.

El análisis que se llevo a cabo en este manuscrito se centro en variables consideradas confusoras, como son la escolaridad y ocupación de los padres, en especial la ocupación del padre es un factor mayormente determinante para el desarrollo de una neoplasia maligna, esto es, no necesariamente como tal, sino lo que implica la ocupación del padre, como son medio en el que se desarrollo y factores químicos, radiaciones a los que esta expuesto el padre; carencia en su sistema inmunológico y susceptibilidad a virus y bacterias, que pueden modificar el ADN o alterar las células del huésped y puede ser heredado a su hijo. En otras palabras, la ocupación del padre es considerada un factor de riesgo porque en la población estudiada, todos los niños tienen cáncer, la asociación es mayor que en la ocupación materna, y es confusora porque no se sabe con certeza que es, en este caso, que categoría de la ocupación del padre produce el cáncer.

Si se analiza desde un punto de vista más lógico, como se mencionó anteriormente, no es la ocupación lo que interesa, sino lo que esta detrás de ésta. La ocupación del padre únicamente sirve para describir los datos y conocer las frecuencias en esta población, además de inducir una conclusión empírica, muy general.

La escolaridad tanto del padre y de la madre, no tiene gran relación para el desarrollo de una neoplasia, sin embargo; es importante considerarlas porque con estas se puede diagnosticar oportunamente un cáncer, prevenirlo, tratarlo y eliminarlo. Ya que si una madre con suficientes conocimientos acerca de las patologías de esta enfermedad puede percatarse más rápidamente que una madre con años de escolaridad mínimos o escasos. Considere el siguiente ejemplo: un pequeño de 5 años sufre un golpe en una de sus rodillas, tiene un moretón, su madre no termino la primaria, ella cree que el moretón fue ocasionado por el golpe, sólo le da una remedio casero para el dolor, transcurridos 3 meses se da cuenta que el moretón no desaparece, lo lleva con el médico de la clínica que le corresponde a la más cercana. El médico pregunta que le sucedió, que ha tomado, y le receta un medicamento para moretones, transcurren 2 meses después de acudir con el médico familiar, el niño presenta una alteración globular en su rodilla, se toman radiografías de la extremidad del pequeño y el radiólogo junto con el médico se dan cuenta de la probabilidad de un tumor óseo. Se envía al nivel de atención correspondiente, pero el cáncer ya ha avanzado y lo único que se puede hacer, para salvar la vida del niño, es amputar el miembro.

El anterior párrafo, trata de describir un ejemplo de la vida real, donde se destaca la importancia del nivel de escolaridad de los padres. Tanto los padres como los médicos deben conocer las



diferentes patologías de una enfermedad, porque de ello depende la vida de un niño o individuo. En ese sentido, la escolaridad es importante, la anterior situación no hubiera sido tan traumática para el menor, si su madre hubiese tenido más conocimiento acerca de enfermedades tan mortales como el cáncer.

Actualmente, en los estudios clínicos se emplean este tipo de variables para determinar asociaciones o interacciones que puedan vislumbrar el desarrollo de una enfermedad como tal, claro está, que existen diversas enfermedades, pero en la mayoría de estas no se determina un solo factor, sino múltiples factores que intervienen en su desarrollo.

Aún no queda claro, que factores intervienen en el desarrollo de una neoplasia maligna. Lo que sí queda claro, es que deben realizarse estudios más profundos que involucren variables específicas que vislumbren relaciones más verosímiles. En la actualidad, muy pocos son los estudios que involucran factores como medio ambiente, composición del aire, exposición a químicos contaminación u otros elementos que disminuyen la capacidad inmunológica de los individuos y los hace susceptibles a enfermedades como el cáncer, diabetes, y otras relacionadas con la modificación de los genes. Para ser una fuente de cultivo de bacterias y virus, es decir, que compuestos químicos conforman el aire de una determinada zona geográfica, determinar cómo afectan a la población, y como se asocia esta con el desarrollo de una enfermedad.

Por otra parte, el análisis de correspondencias es una herramienta orientada como ayuda en la generación de nuevas teorías, facilitando la elección previa de las características o modalidades más significativas, frente a otros criterios de tipo teórico. Además demostró ser útil en la exploración de asociaciones consideradas espurias, cabe mencionar que es una técnica descriptiva exploratoria que tiene desventajas, pero queda la posibilidad de seguirla desarrollando o aplicar otras tecnologías aunadas a esta para su mayor efectividad.

Como conclusión final, este trabajo ha dado la oportunidad de saber que las matemáticas tienen aplicación y mucha en diversas áreas, pero más en la medicina. La teoría que se enseña en las aulas de aprendizaje es suficiente para comprender el proceso de una técnica, pero siempre hace falta la práctica, que en mi caso y el de muchos compañeros fue escasa, valdría la pena que se retomara el plan de estudios a orientar a los alumnos de nuevo ingreso a la aplicación de fenómenos reales.

**REFERENCIAS****BIBLIOGRÁFICAS**

1. Hair FJ, Anderson ER y otros. *Multivariate Data Analysis with Readings*. Third Edition. MacMillan Publishing Company. New York 1987;5-10.
2. Mendenhall W, Dennis D W, Scheaffer R L. *Estadística Matemática con Aplicaciones*. Segunda Edición. Grupo Editorial Iberoamérica. 1990; 1-3.
3. Jambu M. *Exploratory and Multivariate Data Analysis*. Academic Press Inc. London. 1991;1,169-239.
4. Sahai H, Khurshid A. *Statistics in Epidemiology*. CRC Press. 1996;143-162.
5. Cuadras CM. *Métodos de Análisis Multivariante*. PPU Barcelona. 1991;265-280.
6. Wayne WD. *Biostatistics: A Foundation for Analysis in the Health Sciences*. Fifth Edition. John Wiley & Sons, Inc. 1991;prefacio.
7. Tortora GJ, Anagnostakos P. *Principios de Anatomía y Fisiología*. 5ª Edición. Harla. 1989;90-95.
8. Feinberg E. Stephen. *The Analysis of Cross-Classified Categorical Data*. Second Edition, The MIT Press, 1980.
9. Benzecri J P. *Correspondence Analysis Handbook*. Marcel Dekker Inc. New York. 1992.
10. Krzanowski W J, Marriott FHC. *Multivariate Analysis Part I: Distributions, ordination and inference*. Edward Arnold. 1994;127-144.
11. Rijkevorsel JLAV, Leeuw JD. *Component and Correspondence Analysis: Dimension Reduction by Functional Approximation*. John Wiley & Sons. New York. 1988.
12. Lebart L, Morineau A, Warwick KM. *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for large matrices*. John Wiley & Sons. New York. 1984;30-62.
13. Greenacre M J. *Theory and applications of correspondence analysis*. Academic Press. 1984.

## OTRAS FUENTES

14. Concato J, Feinstein RA, Holford RT. The Risk of Determining Risk with Multivariate Models. *Annals of Internal Medicine*. American College of Physicians. 1993;118:201-210.
15. Bowman A, Curren J, Lyall R. The birds and the bees: interactive graphics and problem solving in the teaching of statistics. *Statistics and Computing*. Chapman & Hall. 1997;7:237-246.
16. Malkin D. Age specific oncogenesis: the genetics of cancer susceptibility. *Environ Health Perspect*. 1995;103:37-39.
17. Goldman R. Children Unique and vulnerable. Environmental Risk facing children and recommendations for response. *Environ Health Perspect*. 1995;103(supl):13-18.
18. Valdespina G, Velázquez M O, López A P. Panorama Epidemiológico del cáncer en México. 1987. *Gac Med Mex*. 1992;128:329-332.
19. Parkin AM, Stiller CA y otros. The international incidence of childhood cancer. *International Journal Cancer*. 1988;42:511-520.
20. Cassab HG, Barroso E y otros. Progresos en el registro del cáncer en el Instituto Mexicano del Seguro Social. *Rev Med IMSS*. 1986;24:125-131.
21. Stockwell EG. A critical examination of relationship between socio-economic status and mortality. *American Journal Public Health*. 1963;53:956-64.
22. Cordier S, Lefeuvre B y otros. Parental occupation, occupational exposure to solvents and polycyclic aromatic hydrocarbons and risk of childhood brain tumor (Italy, France, Spain). *Cancer Causes and Control*. 1995;8(5):no disponibles.
23. Clasificación Internacional Uniforme de Ocupaciones. Berna: Organización Internacional Uniforme de Ocupaciones. 1968:1
24. Lerner IJ, Kennedy BJ. The Prevalence of Questionable Methods of Cancer Treatment in the United States. *CA-A Cancer Journal* 1992;42:181-191.
25. Arturo Fajardo G, Manuel Mejía A, y cols. Epidemiología de las neoplasias malignas en niños residentes del Distrito Federal (1982-1991). *Boletín Med Hosp. Infant Mex*. Septiembre 1995;9(52):507-516.
26. Fajardo G A, Navarrete M A, and cols. Incidence of malignant neoplasms in children attended at social security hospitals in Mexico City. *Med Pediatr Oncol*. 1997; 29:208-212
27. Grosfeld J L. Risk based management: current concepts of treating malignant solid tumors of childhood. *J Am Coll Surg*, 1999 Oct; 189(4):407-25.

28. Ruth Davies, Chris Canning. Discrete Event Simulation to Evaluate Screening for Diabetic Eye Disease. *Simulation*. April 1996;66:4,209-216.
29. Enciclopedia Encarta 2000  
 "Álgebra", Enciclopedia Encarta 2000, Microsoft Corporation 1993-1999.
30. Historia de las matemáticas  
<http://www.alared.com/oposiciones/matematicas/>
31. Herrero JF, Cuesta M. Introducción al Análisis de Correspondencias. Universidad de Oviedo. 1998.  
[http://www.uniovi.es/user\\_html/herrero/Dpto\\_Psicologia/metodos/hardcopy/98\\_1.html](http://www.uniovi.es/user_html/herrero/Dpto_Psicologia/metodos/hardcopy/98_1.html)
32. Cristina M R. Las Funciones del Profesor Universitario analizadas por sus protagonistas. *Revista Electrónica de Investigación y EE*. 1996;1(2);1134-4032.  
<http://www.uca.es/RELIEVE/v2n1.htm>
33. Tutorial SPSS  
<http://www.vulture.geog.ubc.ca/tutorials/spss/Index.html>.
34. NATIONAL CANCER INFORMATION (NCI)  
<http://www.nci.gov.uk/cancernet/Index.html>  
[http://cancernet.nci.nih.gov/clinpdq/facts\\_span2/#1](http://cancernet.nci.nih.gov/clinpdq/facts_span2/#1)
35. Francois Micheloud.  
<http://www.micheloud.com/FXM/COR/S-marg.htm>
36. Kochevar Research (Correspondence Analysis)  
<http://www.kochevarresearch.com/Correspondence%20analysis.htm>
37. Impacto de la crisis asiática en América Latina  
<http://www.lanic.utexas.edu/project/sela/docs/spdredi21-98a.htm>
38. Correspondence Analysis of Codon Usage  
<http://www.molbiol.ox.ac.uk/cu/culong.html>

# ANEXOS

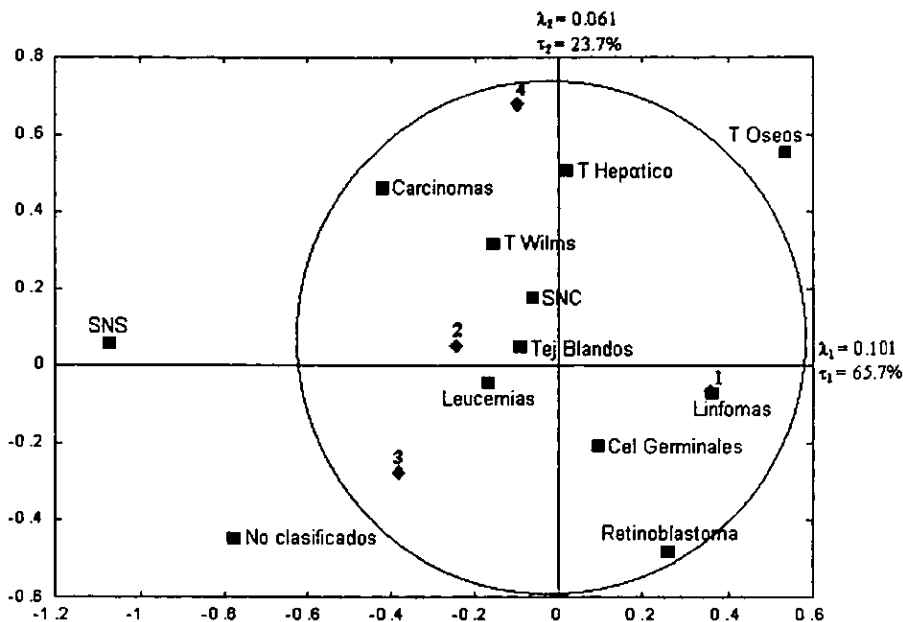


Figura 2. Análisis de datos retrospectivos, neoplasias malignas (representado por un cuadro) en niños representada en el espacio factorial (1,2), donde 1: Primaria incompleta o menos, 2: Primaria completa y secundaria incompleta, 3: Secundaria completa, técnico y bachillerato, y 4: Profesionistas y más, representado por un rombo.

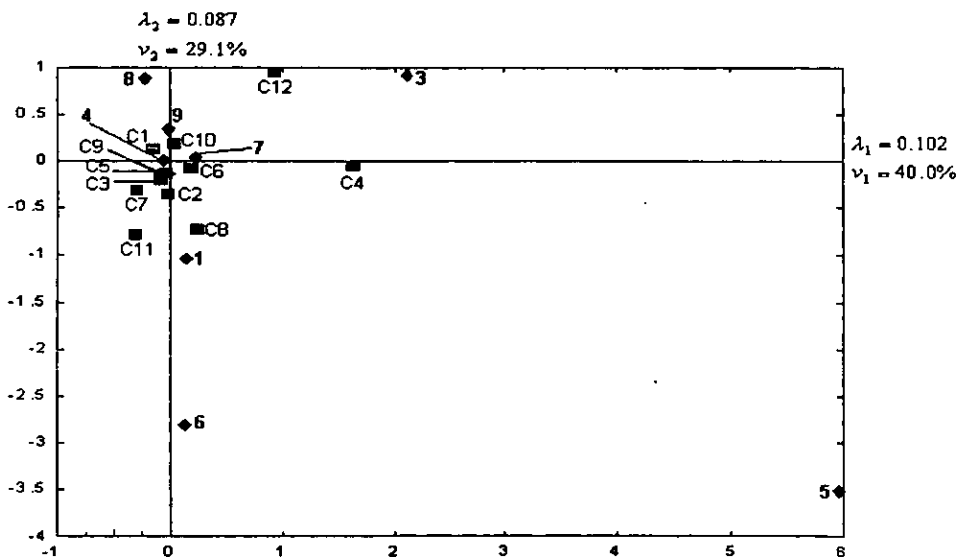


Figura 3. Gráfica de las 12 modalidades del cáncer según clasificación de Birch y Marsden, cuadro azul, asociada a la ocupación materna, representación sobre los 2 principales ejes, donde 1: Profesional o técnico, 3: Comerciante o vendedor, 4: Trabajador de los servicios, 5: Trabajadores agrícolas, 6: Directores y funcionarios, 7: Personal administrativo, 8: Hogar sin recibir salario, y 9: No clasificado; rombo rojo.

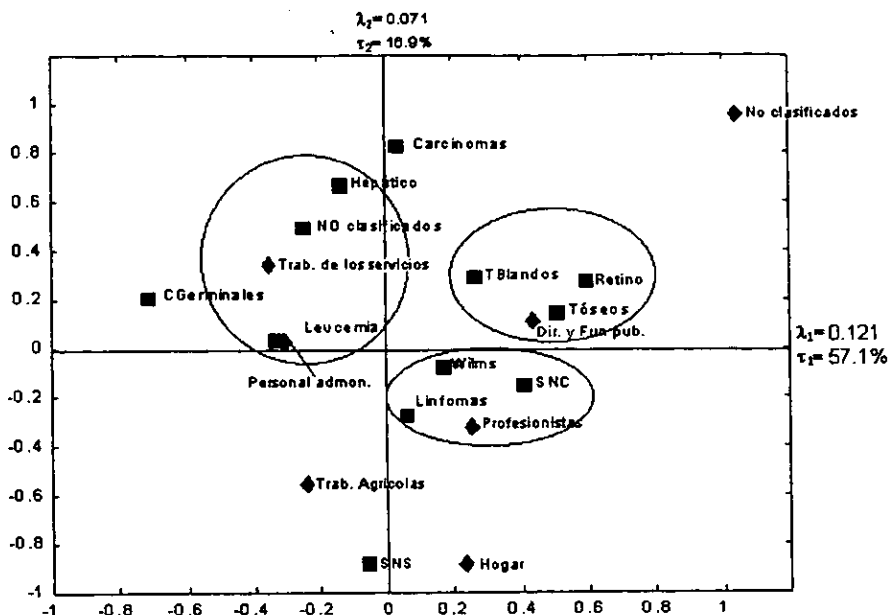


Figura 4. Gráfica de las 12 modalidades del cáncer según clasificación de Birch y Marsden, cuadro rojo, asociada a la ocupación paterna, representación sobre los 2 principales ejes; donde 1: Profesional o técnico, 4: Trabajador de los servicios, 5: Trabajadores agrícolas, 6: Directores y funcionarios, 7: Personal administrativo, 8: Hogar sin recibir salario, y 9: No clasificado; rombo azul.

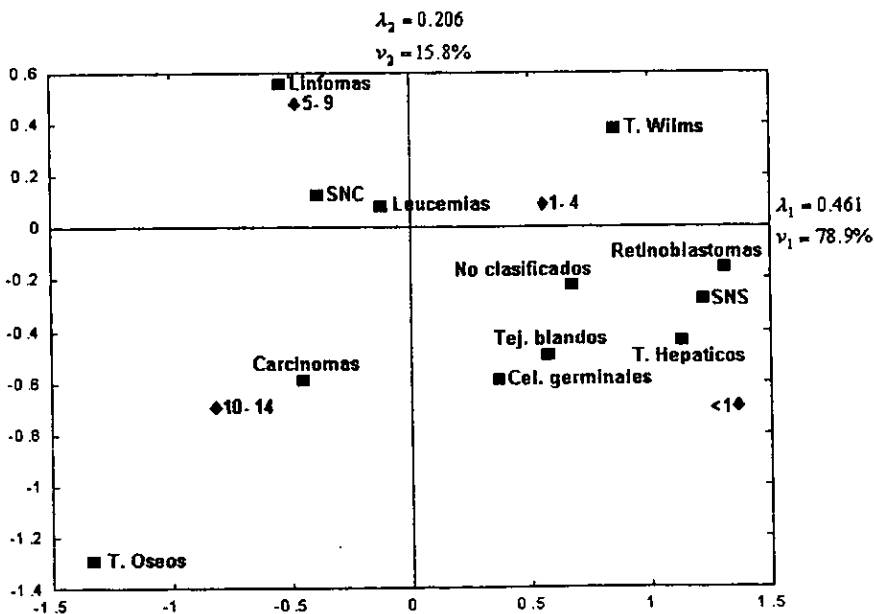


Figura 5. Representación de las puntuaciones fila y columna, tablas 33 y 34, sobre los 2 primeros ejes. Neoplasias malignas asociadas a la edad pediátrica.

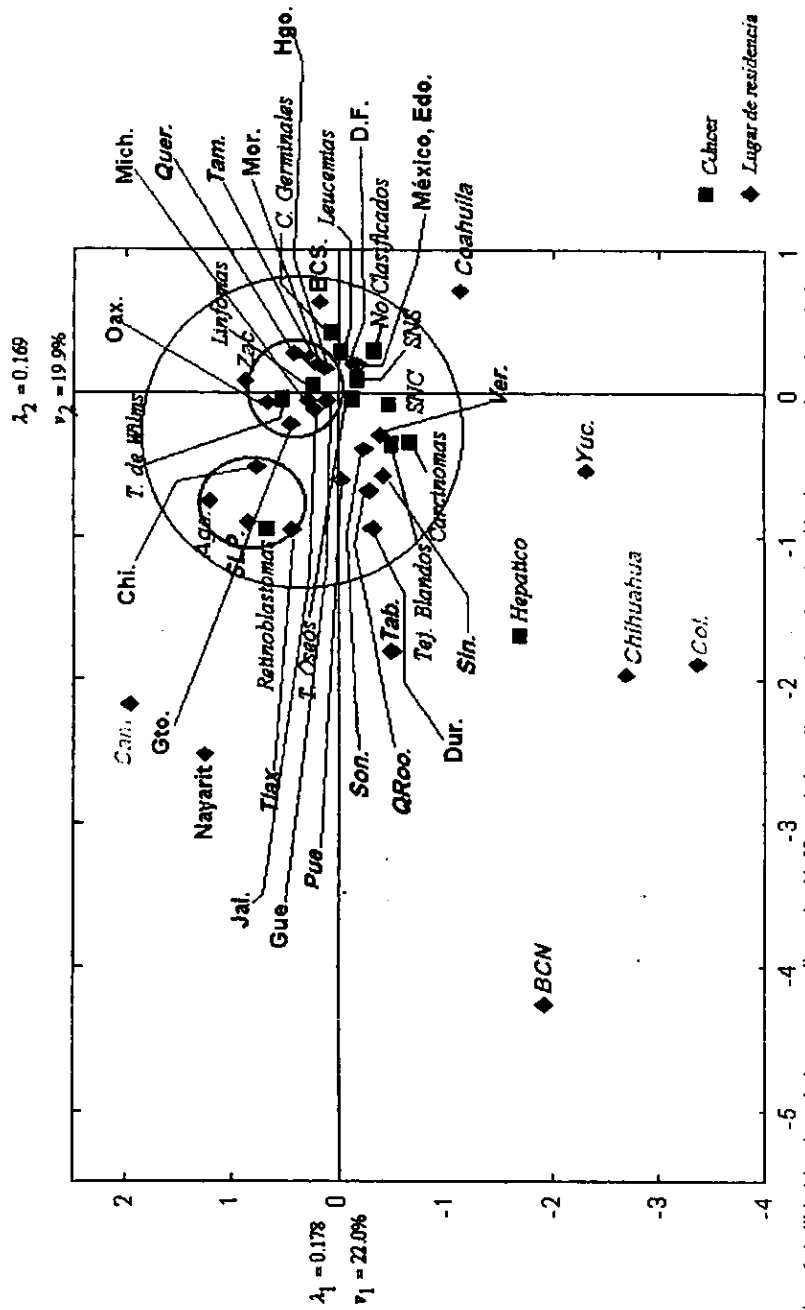


Figura 6. Análisis del conjunto de datos correspondiente a la tabla 37, neoplasias malignas en niños y lugar de residencia representada en el espacio factorial (1,2).



Tabla 1. Conjunto de datos registrados durante el periodo 1980-1992 de cáncer en niños menores de 15 años según escolaridad de la madre. Fuente Unidad de Investigación en Epidemiología Clínica del CMN SXXI. IMSS. 1998.

Tipo de cáncer	Escolaridad materna				Marginal total
	<Primaria incompleta	Primaria completa Secundaria incompleta	Secundaria completa Técnico y bachillerato	Profesionistas	
Leucemia	580	372	226	61	1239
Linfomas	343	167	94	32	636
SNC	175	99	59	14	347
SNS	36	31	20	8	95
Retinoblastoma	153	77	63	14	307
T. Wilms	92	62	36	8	198
T. Hepáticos	19	19	12	3	53
T. Óseos	107	47	36	17	207
Tej. Blandos	81	32	33	12	158
C. Germinales	80	42	32	12	166
Carcinomas	22	16	5	6	49
No clasificados	21	14	15	0	50
Marginal total	1709	978	631	187	3505

Tabla 2. Resumen del análisis sobre las variables en estudio.

Dimensión	Valor Singular	Inercia	Ji cuadrado	Sig.	Proporción de Inercia		Confianza del valor singular	
					Explicado por	Acumulativo	Desviación estándar	Correlación
1	.083	.007			.455	.455	.017	.008
2	.068	.005			.305	.760	.019	
3	.060	.004			.240	1.000		
Total		.015	52.546	.017 <sup>a</sup>	1.000	1.000		

a 33 grados de libertad.

Tabla 3. Análisis sobre las puntuaciones renglón, tipo de cáncer (a).

Tipo de cáncer	Masa	Puntuación en la dimensión		Inercia	Contribución				
		1	2		Del punto a la inercia de la dimensión		De la dimensión a la inercia del punto		Total
					1	2	1	2	
Leucemia	.353	.164	-.034	.001	.114	.006	.807	.029	.836
Linfomas	.181	-.285	.240	.002	.178	.154	.541	.314	.854
SNC	.099	.042	.224	.000	.002	.073	.032	.762	.794
SNS	.027	.318	-.831	.001	.033	.277	.151	.845	.997
Retinoblastoma	.088	.027	.162	.001	.001	.034	.008	.231	.238
T. Wilms	.056	.282	.057	.000	.055	.003	.784	.026	.810
T. Hepático	.015	.752	-.574	.001	.103	.074	.677	.323	1.000
T. Óseos	.059	-.498	-.224	.000	.017	.042	.753	.125	.878
Tej. blandos	.045	-.396	-.139	.002	.085	.013	.355	.036	.391
C Germinales	.047	-.174	-.244	.000	.017	.042	.246	.399	.645
Carcinomas	.014	-.553	-1.073	.002	.052	.238	.188	.579	.767
No clasificados	.014	1.027	.445	.002	.182	.042	.610	.094	.704
<b>Total</b>	<b>1.000</b>			<b>.015</b>	<b>1.000</b>	<b>1.000</b>			

a. Normalización simétrica.

Tabla 4. Análisis sobre las puntuaciones columna, escolaridad de la madre(a).

Escolaridad materna	Masa	Puntuación en la dimensión		Inercia	Contribución				
		1	2		Del punto a la inercia de la dimensión		De la dimensión a la inercia del punto		Total
					1	2	1	2	
<Primaria incompleta	.499	-.209	.187	.003	.258	.253	.603	.396	.999
Primaria completa Secundaria incomp.	.279	.288	-.113	.004	.205	.009	.539	.067	.607
Secundaria completa Técnico y Bach.	.180	.307	-.057	.004	.205	.009	.386	.011	.397
Profesionistas	.053	-.630	-.932	.005	.256	.686	.357	.640	.996
<b>Total</b>	<b>1.000</b>			<b>.015</b>	<b>1.000</b>	<b>1.000</b>			

a. Normalización simétrica.

Tabla 5. Confianza de las puntuaciones renglón

Tipo de cáncer	Desviación estándar en la dimensión		Correlación
	1	2	
Leucemia	.065	.182	.293
Linfomas	.248	.380	.559
SNC	.219	.210	-.036
SNS	.755	.297	.867
Retinoblastoma	.243	.534	-.059
T. Wilms	.121	.332	-.231
T. Hepático	.525	.570	.978
T. Óseos	.260	.517	-.495
Tej. blandos	.410	1.047	-.050
C. Germinales	.274	.424	-.164
Carcinomas	1.061	1.327	-.331
No clasificados	.661	1.665	-.305

Tabla 6. Confianza de las puntuaciones columna

Escolaridad materna	Desviación estándar en la dimensión		Correlación
	1	2	
<Primaria incompleta	.170	.160	.972
Primaria completa	.208	.513	.151
Secundaria incompleta	.293	.811	.069
Secundaria completa	.840	.470	-.922
Técnico y bachillerato			
Profesionistas			

Tabla 7. Conjunto de datos registrados durante el periodo 1980-1992 de cáncer en niños menores de 15 años según escolaridad de la padre. Fuente Unidad de Investigación en Epidemiología Clínica del CMN SXXI. IMSS. 1998.

Tipo de cáncer	Escolaridad paterna				Marginal Total
	<Primaria incompleta	Primaria completa Secundaria incompleta	Secundaria completa Técnico y bachillerato	Profesionistas	
Leucemia	496	390	242	111	1239
Linfomas	307	166	107	56	636
SNC	142	110	60	35	347
SNS	25	30	28	12	95
Retinoblastoma	147	78	62	20	307
T. Wilms	78	58	38	24	198
T. Hepático	22	15	9	7	53
T. Óseos	102	51	27	27	207
Tej. blandos	65	43	33	17	158
C. Germinales	74	45	33	14	166
Carcinomas	17	17	9	6	49
No clasificados	16	16	14	4	50
Marginal Total	1491	1019	662	333	3505

Tabla 8. Resumen de valores obtenidos por el análisis de ambas variables, neoplasias malignas y escolaridad paterna.

Dimensión	Valor Singular	Inercia	JI cuadrado	Sig.	Proporción de inercia		Confianza del valor singular	
					Explicado por	Acumulado	Desviación estándar	Correlación 2
1	.101	.010			.657	.657	.016	-.028
2	.061	.004			.237	.894	.017	
3	.041	.002			.106	1.000		
<b>Total</b>		.015	54.336	.011 <sup>a</sup>	1.000	1.000		

a. 33 grados de libertad.

Tabla 9. Análisis sobre las puntuaciones renglón (a).

Tipo de cáncer	Masa	Puntuación en la dimensión		Inercia	Contribución				
		1	2		Del punto a la inercia de la dimensión		De la dimensión a la inercia del punto		Total
					1	2	1	2	
Leucemia	.353	-.167	-.043	.001	.097	.011	.701	.028	.729
Linfomas	.181	.361	-.070	.002	.234	.015	.974	.022	.996
SNC	.099	-.067	.178	.000	.004	.052	.095	.410	.505
SNS	.027	-1.071	.060	.004	.308	.002	.865	.002	.867
Retinoblastoma	.088	.259	-.482	.002	.058	.336	.314	.651	.964
T. Wilms	.056	-.155	.314	.001	.013	.092	.252	.622	.873
T. Hepático	.015	.017	.506	.000	.000	.064	.002	.920	.921
T. Óseos	.059	.534	.553	.003	.167	.298	.590	.380	.970
Tej. blandos	.045	-.093	.048	.000	.004	.002	.161	.026	.187
C. Germinales	.047	.093	-.203	.000	.025	.032	.218	.622	.840
Carcinomas	.014	-.423	.462	.000	.025	.049	.541	.388	.929
No clasificados	.014	-.777	-.447	.001	.085	.047	.808	.161	.968
<b>Total</b>	1.000			.016	1.000	1.000			

a. Normalización simétrica

Tabla 10. Análisis sobre las puntuaciones columna(a).

Escolaridad paterna	Masa	Puntuación en la dimensión		Inercia	Contribución				
		1	2		Del punto a la inercia de la dimensión		De la dimensión a la inercia del punto		Total
					1	2	1	2	
<Primaria incompleta	.425	.359	-.064	.006	.544	.029	.981	.019	1.000
Primaria completa	.291	-.245	.053	.003	.172	.013	.659	.018	.677
Secundaria incompleta									
Secundaria completa	.189	-.383	-.277	.004	.274	.239	.671	.211	.882
Técnico y bachillerato	.095	-.100	.677	.003	.009	.719	.032	.872	.904
Profesionistas									
Total	1.000			.016	1.000	1.000			

a Normalización simétrica

Tabla 11. Confianza de las puntuaciones renglón.

Tipo de cáncer	Desviación estándar en la dimensión		Correlación
	1	2	1-2
Leucemia	.061	.137	-.054
Linfomas	.043	.092	.607
SNC	.111	.190	.114
SNS	.239	.567	.077
Retinoblastoma	.199	.140	.461
T. Wilms	.136	.149	.201
T. Hepático	.206	.163	-.053
T. Óseos	.230	.217	-.572
Tej. blandos	.113	.260	.049
C. Germinales	.093	.105	.241
Carcinomas	.207	.225	.440
No clasificados	.206	.262	-.552

Tabla 12. Confianza de las puntuaciones columna.

Escolaridad paterna	Desviación estándar en la dimensión		Correlación
	1	2	1-2
<Primaria incompleta	.039	.088	.638
Primaria completa	.099	.216	.080
Secundaria incompleta			
Secundaria completa	.148	.213	.080
Técnico y bachillerato			
Profesionistas	.289	.236	.101

Tabla 17. Confianza de las modalidades: cáncer

Tipo de cáncer	Desviación estándar en la dimensión		Correlación
	1	2	1-2
Leucemia	.128	.155	.868
Linfomas	.157	.120	-.194
SNC	.421	.153	.116
SNS	.5795	1.493	.213
Retinoblastoma	.202	.226	-.154
T. Wilms	.324	.415	.205
T. Hepático	.380	.482	-.443
T. Óseos	.831	.394	.402
Tej. Blandos	.216	.264	-.259
C. Germinales	.205	.161	-.103
Carcinomas	.748	.768	-.435
No clasificados	1.488	1.050	-.572

Tabla 18. Confianza de las modalidades ocupación materna.

Ocupación materna	Desviación estándar en la dimensión		Correlación
	1	2	1-2
Profesionista	1.042	.411	.345
Com. Y Vend.	1.439	1.945	-.914
Trab. de serv.	.641	.661	-.036
Trab. agrícolas	6.994	6.000	.711
Dir. y fun. pub.	3.005	3.645	-.326
Per. Admón..	.468	.609	-.031
Hogar	.986	.646	.447
No clasificados	.026	.057	.278

Tabla 19. Frecuencias absolutas del conjunto de datos, neoplasias malignas y ocupación paterna.

Tipo de cáncer	Ocupación paterna							Marginal Total
	Profesionistas	Trab. de servicios	Trab. agrícolas	Dir. y Fun pub	Per. Admón.	Hogar	No clasificado	
Leucemia	101	113	38	317	616	49	8	1239
Linfomas	53	42	24	212	267	35	3	636
SNC	36	22	11	115	143	16	4	347
SNS	10	4	3	24	48	6	0	95
Retinoblastoma	19	22	5	128	116	13	4	307
T. Wilms	18	14	7	68	84	5	2	198
T. Hepático	4	8	1	17	22	1	0	53
T. Óseos	24	17	3	72	78	9	4	207
Tej. blandos	21	13	1	54	63	3	3	158
C. Germinales	11	14	7	41	88	4	1	166
Carcinomas	4	2	0	14	27	1	1	49
No clasificados	4	3	1	15	27	0	0	50
<b>Marginal Total</b>	<b>305</b>	<b>274</b>	<b>101</b>	<b>1077</b>	<b>1579</b>	<b>139</b>	<b>30</b>	<b>3505</b>

Tabla 20. Análisis de la inercia explicada por las diferentes dimensiones.

Dimensión	Valor Singular	Inercia	Ji cuadrado	Sig.	Proporción de inercia		Confianza del Valor Singular	
					Explicada por	Acumulada	Desviación estándar	Correlación
								2
1	.121	.017			.571	.571	.015	.008
2	.071	.005			.169	.741	.015	
3	.066	.004			.145	.886		
4	.043	.002			.063	.949		
5	.033	.001			.036	.985		
6	.021	.000			.015	1.000		
<b>Total</b>		.030	127.034	.004(a)	1.000	1.000		

a 88 grados de libertad.

Tabla 21. Análisis de las puntuaciones fila para la formación de las dimensiones(a).

Tipo de cáncer	Masa	Puntuación en la dimensión			Inercia	Contribución			
		1	2	Inercia		Del punto a la inercia de la dimensión		De la dimensión a la inercia del punto	
						1	2	1	2
Leucemia	.350	-.332	.030	.005	.294	.005	.956	.004	.961
Linfomas	.189	.063	-.276	.002	.006	.202	.064	.667	.731
SNC	.097	.413	-.154	.003	.126	.032	.677	.051	.728
SNS	.026	-.062	-.887	.002	.001	.290	.007	.735	.741
Retinoblastoma	.085	.594	.273	.006	.229	.089	.629	.073	.701
T. Wilms	.056	.175	-.076	.001	.013	.004	.433	.044	.477
T. Hepático	.014	-.137	.663	.001	.002	.087	.031	.395	.427
T. Óseos	.061	.511	.144	.002	.121	.018	.858	.037	.895
Tej. blandos	.046	.264	.290	.002	.025	.054	.274	.180	.454
C. Germinales	.045	-.715	.206	.003	.175	.027	.866	.039	.905
Carcinomas	.014	.037	.827	.002	.000	.136	.001	.381	.382
No clasificados	.017	-.249	.488	.001	.008	.056	.157	.328	.485
<b>Total</b>	1.000			.030	1.000	1.000			

a Normalización simétrica

Tabla 22. Análisis sobre las puntuaciones columna para la formación de las dimensiones(a).

Ocupación del padre	Masa	Puntuación en la dimensión			Contribución				
		1	2	Inercia	Del punto a la inercia de la dimensión		Del punto a la inercia de la dimensión		Total
					1	2	1	2	
Profesionistas	.088	.255	-.322	.003	.044	.129	.217	.189	.406
Trab. de los servicios	.077	-.353	.340	.003	.074	.125	.413	.208	.621
Trab. agrícolas	.030	-.243	-.557	.002	.014	.132	.128	.364	.492
Dir. y Fun. públicos	.302	.437	.109	.009	.441	.050	.885	.030	.915
Pers. Admon.	.453	-.308	.026	.006	.328	.004	.932	.004	.935
Hogar	.039	.236	-.885	.003	.017	.433	.091	.699	.791
No clasificados	.010	1.045	.954	.004	.083	.127	.362	.165	.527
<b>Total</b>	<b>1.000</b>			<b>.030</b>	<b>1.000</b>	<b>1.000</b>			

a Normalización simétrica.

Tabla 23. Bondad de ajuste de las categorías del cáncer

Tipo de cáncer	Desviación estándar en la dimensión		Correlación 1-2
	1	2	
Leucemia	.041	.077	.110
Linfomas	.093	.318	.087
SNC	.114	.772	.083
SNS	.295	.405	-.083
Retinoblastoma	.172	1.185	.015
T. Wilms	.123	.200	.051
T. Hepático	.312	1.197	.068
T. Óseos	.082	.483	-.029
Tej. blandos	.168	1.028	.042
C. Germinales	.131	.222	.226
Carcinomas	.396	1.321	.135
No clasificados	.245	.742	.062

Tabla 24. Bondad de ajuste de las categorías ocupación paterna

Ocupación del padre	Desviación estándar en la dimensión		Correlación 1-2
	1	2	
Profesionistas	.182	1.169	.008
Trab. de los servicios	.175	.386	.049
Trab. agrícolas	.276	.424	-.167
Dir. y Fun. públicos	.064	.406	-.005
Pers. Admon.	.040	.100	.077
Hogar	.278	.713	.067
No clasificados	.571	3.378	.078



**Tabla 25** Conjunto de datos de neoplasias malignas registrados durante el periodo 1980 - 1991 según el sexo del niño.

Tipo de cáncer	Sexo		Marginal Total
	niña	niño	
Leucemia	550	689	1239
Linfomas	189	447	636
SNC	159	188	347
SNS	39	56	95
Retinoblastoma	142	165	307
T. Wilms	100	98	198
T. Hepático	20	33	53
T. Óseos	91	116	207
Tej. blandos	63	95	158
C. Germinales	74	92	166
Carcinomas	21	28	49
No clasificados	26	24	50
<b>Marginal Total</b>	<b>1474</b>	<b>2031</b>	<b>3505</b>

**Tabla 26.** Resumen del análisis de inercia explicada por las diferentes dimensiones.

Dimensión	Valor Singular	Inercia	Ji cuadrado	Signif.	Proporción de Inercia		Confianza del valor Singular
					Explicada por	Acumulativa	Desviación estándar
I	.126	.016			1.000	1.000	.016
Total		.016	56.085	.000(a)	1.000	1.000	

a 11 grados de libertad.

Tabla 27. Resumen de los puntos renglón, tipo de cáncer(a).

Tipo de cáncer	Puntuación en la dimensión			Contribución		Total
	Masa	1	Inercia	Del punto a la inercia	De la dimensión a la inercia	
				De la dimensión	del punto	
			1	1		
Leucemia	.353	-.133	.001	.049	1.000	1.000
Linfomas	.181	.703	.011	.708	1.000	1.000
SNC	.099	-.215	.001	.036	1.000	1.000
SNS	.027	.057	.000	.001	1.000	1.000
Retinoblastomas	.088	-.239	.001	.040	1.000	1.000
T. Wilms	.056	-.481	.002	.103	1.000	1.000
T. Hepáticos	.015	.246	.000	.007	1.000	1.000
T. Óseos	.059	-.109	.000	.006	1.000	1.000
Tej. Blandos	.045	.124	.000	.005	1.000	1.000
C. Germinales	.047	-.144	.000	.008	1.000	1.000
Carcinomas	.014	-.046	.000	.000	1.000	1.000
No clasificados	.014	-.566	.001	.036	1.000	1.000
<b>Total</b>	<b>1.000</b>		<b>.016</b>	<b>1.000</b>		

a Normalización simétrica.

Tabla 28 Resumen de los puntos columna, sexo(a).

Sexo	Puntuaciones en la dimensión			Contribución		Total
	Masa	1	Inercia	Del punto a la inercia	De la dimensión a la inercia	
				de la dimensión	del punto	
			1	1		
Niña	.421	-.417	.009	.579	1.000	1.000
Niño	.579	.303	.007	.421	1.000	1.000
<b>Total</b>	<b>1.000</b>		<b>.016</b>	<b>1.000</b>		

a Normalización simétrica

Tabla 29. Confianza de los puntos renglón

Tipo de cáncer	Desviación estándar en la dimensión
	1
Leucemia	.121
Linfomas	.046
SNC	.015
SNS	.007
Retinoblastomas	.016
T. Wilms	.031
T. Hepático	.017
T. Óseos	.009
Tej. blandos	.010
C. Germinales	.011
Carcinomas	.006
No clasificados	.036

Tabla 30. Confianza de los puntos columna

Sexo	Desviación estándar de la dimensión
	1
Niña	.027
Niño	.020

Tabla 31. Conjunto de datos organizados según neoplasias malignas y edad pediátrica, durante el periodo 1980-1992.

Tipo de cáncer	Edad pediátrica (años)				Marginal Total
	<1	1-4	5-9	10-14	
Leucemia	43	517	408	271	1239
Linfomas	11	163	330	132	636
SNC	7	116	136	88	347
SNS	23	55	15	2	95
Retinoblastoma	57	220	27	3	307
Tumor de Wilms	18	136	40	4	198
Hepático	13	29	8	3	53
T Óseos	1	13	67	126	207
Tejidos blandos	25	78	28	27	158
C. Germinales	22	77	30	37	166
Carcinomas	2	15	14	18	49
No clasificados	10	22	13	5	50
<b>Marginal Total</b>	<b>232</b>	<b>1441</b>	<b>1116</b>	<b>716</b>	<b>3505</b>

Tabla 32. Resumen de estadísticos sobre la inercia de cada dimensión.

Dimensión	Valor Singular	Inercia	Ji cuadrado	Signif.	Proporción de Inercia		Confianza del Valor Singular	
					Explicado por	Acumulada	Desviación estándar	Correlación 2
1	.461	.212			.789	.789	.013	.165
2	.206	.042			.158	.946	.019	
3	.120	.014			.054	1.000		
Total		.269	942.841	.000(a)	1.000	1.000		

A 33 grados de libertad.

Tabla 33. Resumen de los puntos renglón, neoplasia maligna (a).

Tipo de cáncer	Puntuaciones en la dimensión				Contribución				Total
	Masa	1		Inercia	Del punto a la inercia de la dimensión		De la inercia a la dimensión del punto		
		2	1		2	1	2		
Leucemia	.353	-.123	.085	.006	.012	.013	.422	.091	.513
Linfomas	.181	-.547	.554	.040	.118	.271	.624	.286	.910
SNC	.099	-.393	.127	.007	.033	.008	.947	.044	.992
SNS	.027	1.216	-.278	.021	.087	.010	.872	.020	.893
Retinoblastoma	.088	1.309	-.154	.070	.326	.010	.993	.006	.999
T. Wilms	.056	.855	.382	.023	.090	.040	.836	.075	.911
T. Hepáticos	.015	1.131	-.438	.011	.042	.014	.814	.055	.868
T. Óscos	.059	-1.327	-1.289	.068	.226	.477	.703	.297	1.000
Tej. blandos	.045	.579	-.491	.010	.033	.053	.725	.233	.959
C. Germinales	.047	.371	-.584	.006	.014	.078	.470	.522	.992
Carcinomas	.014	-.458	-.584	.002	.006	.023	.567	.412	.979
No clasificados	.014	.677	-.223	.005	.014	.003	.628	.031	.658
Total	1.000			.269	1.000	1.000			

a Normalización simétrica.

Tabla 34. Resumen de los puntos columna, edad pediátrica(a).

Grupo de edad pediátrica (en años)	Puntuaciones en la dimensión				Contribución				
	Masa			Inercia	Del punto a la inercia de la dimensión		De la dimensión a la inercia del punto		
		1	2		1	2	1	2	Total
<1	.066	1.364	-.694	.071	.267	.155	.802	.093	.895
1-4	.411	.558	.086	.064	.278	.015	.923	.010	.933
5-9	.318	-.479	.478	.051	.159	.353	.659	.293	.952
10-14	.204	-.817	-.693	.083	.296	.477	.754	.243	.996
<b>Activo Total</b>	<b>1.000</b>			<b>.269</b>	<b>1.000</b>	<b>1.000</b>			

A Normalización simétrica

Tabla 35. Confianza de las puntuaciones fila

	Desviación estándar en la dimensión		Correlación 1-2
	1	2	
Leucemia	.027	.049	-.280
Linfomas	.056	.089	.092
SNC	.020	.019	.221
SNS	.079	.155	-.231
Retinoblastoma	.029	.044	.355
T. Wilms	.058	.107	-.373
T. Hepático	.088	.164	-.230
T. Óseos	.088	.063	-.375
Tel. Blandos	.047	.054	-.014
C. Germinales	.046	.033	.170
Carcinomas	.042	.045	-.183
No clasificados	.086	.181	-.285

Tabla 36. Confianza de las puntuaciones  
columna

Edad pediátrica	Desviación estándar en la dimensión		Correlación 1-2
	1	2	
<1	.109	.166	-.214
1-4	.036	.058	-.415
5-9	.041	.060	.295
10-14	.048	.043	-.222

Tabla 37. Conjunto de datos organizados según el lugar de residencia, durante el periodo 1980-1992

Lugar de Residencia	Tipo de neoplasia													Total Marginal
	Leucemia	Linfomas	SNC	SNS	Retinoblastomas	Wilms	T. Hepáticos	T. Óscos	Tej. Blandos	C. Germinales	Carcinomas	No clasificados		
Aguascalientes	5	2	0	0	3	3	0	1	2	0	0	0	16	
Baja California	0	0	0	0	1	0	1	1	0	0	0	0	3	
B.C.S.	0	1	0	0	0	0	0	0	0	0	0	0	1	
Campeche	1	2	0	0	1	1	0	0	0	0	1	0	6	
Coahuila	2	0	0	0	0	0	0	1	0	0	0	0	3	
Colima	1	0	2	0	0	0	1	0	1	0	0	0	5	
Chiapas	20	12	6	3	13	7	1	6	0	7	2	0	77	
Chihuahua	4	0	0	0	1	0	1	1	1	1	2	0	11	
D.F.	347	128	78	18	55	43	12	54	31	45	20	9	840	
Durango	7	1	4	1	6	1	1	2	3	0	0	0	26	
Guanajuato	44	37	18	4	21	18	1	9	7	6	0	2	167	
Guerrero	86	62	22	3	21	9	4	13	12	6	1	1	240	
Hidalgo	69	26	19	5	17	11	0	14	11	14	2	4	192	
Jalisco	4	3	1	0	1	1	0	4	1	1	0	0	16	
México, Edo.	320	172	95	25	55	45	15	43	33	42	9	18	872	
Michoacán	61	36	18	9	18	11	2	11	9	8	0	2	185	
Morelos	68	21	9	7	15	12	0	14	8	6	2	5	166	
Nayarit	2	3	1	0	3	1	0	0	0	0	0	0	10	
Oaxaca	39	24	7	2	13	6	2	5	1	9	0	1	109	
Puebla	31	20	17	3	16	5	3	5	9	6	5	1	121	
Querétaro	27	32	5	2	7	3	1	4	2	6	2	0	89	
Quintana Roo	0	0	2	0	0	0	0	0	0	0	0	0	2	
Sn Luis Potosí	3	4	0	0	5	1	0	1	3	2	0	0	19	
Sinaloa	7	9	7	5	5	4	2	2	3	1	0	2	47	
Sonora	1	2	2	0	1	0	0	0	0	0	0	0	6	
Tabasco	3	1	2	1	4	1	2	0	2	0	0	0	16	
Tamaulipas	8	2	0	1	1	1	0	1	2	1	0	0	17	
Tlaxcala	10	8	4	0	6	4	0	4	1	0	2	1	40	
Veracruz	60	21	24	6	14	6	4	11	15	5	3	4	173	
Yucatán	1	0	2	0	0	0	0	0	1	0	0	0	4	
Zacatecas	8	7	2	0	4	4	0	1	0	0	0	0	26	
Total Marginal	1239	636	347	95	307	198	53	207	158	166	49	50	3505	

Tabla 38. Análisis de la inercia explicada por las diferentes dimensiones

Dimensión	Valor singular	Inercia	Ji cuadrado	Signif.	Proporción de Inercia		Confianza del valor singular	
					Explicado por	Acumulativa	Desviación estándar	Correlación 2
1	.178	.032			.220	.220	.023	.004
2	.169	.029			.199	.420	.017	
3	.147	.022			.151	.571		
4	.126	.016			.111	.682		
5	.111	.012			.087	.769		
6	.097	.009			.065	.834		
7	.086	.007			.051	.886		
8	.082	.007			.047	.933		
9	.071	.005			.035	.968		
10	.051	.003			.018	.986		
11	.044	.002			.014	1.000		
Total		.143	501.446	.000(a)	1.000	1.000		

a 341 grados de libertad.

Tabla 39. Resumen de las puntuaciones fila, tipo de cáncer(a).

Tipo de cáncer	Masa	Puntuación en la Dimensión		Inercia	Contribución				
		1	2		Del punto a la inercia de la dimensión		De la dimensión a la inercia del punto		Total
					1	2	1	2	
Leucemia	.353	.181	-.331	.011	.065	.229	.192	.610	.802
Linfomas	.181	.403	.416	.016	.16	.186	.321	.326	.647
SNC	.099	-.508	-.003	.015	.144	.000	.296	.000	.296
SNS	.027	-.313	.114	.007	.015	.002	.064	.008	.073
Retinoblastoma	.088	-.329	.744	.016	.053	.287	.105	.513	.618
Wilms	.056	.179	.466	.009	.010	.073	.037	.240	.277
Hepático	.015	-1.893	.018	.018	.305	.000	.538	.000	.538
T. Óseos	.059	0.34	-.252	.008	.000	.022	.002	.080	.081
Tej. Blandos	.045	-.858	-.022	.013	.187	.000	.461	.000	.462
C. Germinales	.047	.317	-.252	.007	.027	.016	.118	.071	.188
Carcinomas	.014	-.585	-1.433	.018	.027	.170	.048	.272	.319
No clasificados	.014	-.086	-.372	.005	.001	.012	.004	.065	.068
Total	1.000			.143	1.000	1.000			

a Normalización simétrica.

Tabla 40. Resumen de las puntuaciones columna, lugar de residencia(a).

Lugar de residencia	Puntuación en la dimensión				Contribución				Total
	Masa			Inercia	Del punto a la inercia de la dimensión		De la dimensión a la inercia del punto		
		1	2		1	2	1	2	
Aguascalientes	.005	-.147	.930	.004	.001	.023	.005	.183	.188
Baja California	.001	-4.106	1.007	.008	.081	.005	.0315	.018	.333
B.C.S.	.000	2.270	2.467	.001	.008	.010	.203	.228	.431
Campeche	.002	.237	.275	.004	.001	.001	.004	.005	.009
Coahuila	.001	.743	-1.805	.002	.003	.017	.046	.257	.303
Colima	.001	-4.038	-.405	.006	.131	.001	.679	.006	.686
Chiapas	.022	.059	.425	.006	.000	.023	.002	.121	.124
Chihuahua	.003	-1.625	-2.129	.009	.047	.084	.163	.266	.429
D.F.	.240	.088	-.396	.008	.010	.223	.041	.804	.845
Durango	.007	-1.487	.588	.005	.092	.015	.619	.092	.711
Guanajuato	.048	.099	.731	.006	.003	.151	.015	.755	.770
Guerrero	.068	.163	.263	.005	.010	.028	.065	.160	.225
Hidalgo	.055	.062	-.166	.003	.001	.009	.013	.090	.103
Jalisco	.005	.306	-.055	.003	.002	.000	.022	.001	.023
México D.F.	.249	.080	-.077	.004	.009	.009	.067	.058	.124
Michoacán	.053	.027	.276	.002	.000	.024	.003	.341	.344
Morelos	.047	.173	-.211	.005	.008	.013	.053	.075	.128
Nayarit	.003	.144	1.945	.003	.000	.064	.004	.698	.702
Oaxaca	.031	.395	.320	.003	.027	.019	.254	.158	.412
Puebla	.035	-.680	.105	.005	.090	.002	.515	.012	.527
Querétaro	.025	.714	.577	.007	.073	.050	.343	.213	.556
Quintana Roo	.001	-2.861	-.018	.005	.026	.000	.160	.000	.160
Sn Luis Potosí	.005	-.360	1.259	.005	.004	.051	.023	.271	.294
Sinaloa	.013	-.874	.761	.007	.058	.046	.251	.181	.431
Sonora	.002	-.336	1.224	.002	.001	.015	.018	.224	.242
Tabasco	.005	-2.470	1.098	.007	.157	.033	.684	.128	.812
Tamaulipas	.005	.141	-.362	.002	.001	.004	.010	.060	.069
Tlaxcala	.011	-.032	.307	.004	.000	.006	.001	.047	.047
Veracruz	.049	-.614	-.252	.005	.105	.019	.646	.104	.749
Yucatán	.001	-2.383	-.532	.004	.036	.002	.326	.015	.342
Zacatecas	.007	.583	1.105	.003	.014	.054	.137	.467	.604
<b>Total</b>	<b>1.000</b>			<b>.143</b>	<b>1.000</b>	<b>1.000</b>			

a Normalización simétrica.



Tabla 41. Confiianza de los puntos renglón

Tipo de cáncer	Desviación estándar en la dimensión		Correlación
	1	2	
Leucemia	.445	.242	.952
Linfomas	.574	.546	-.904
SNC	.387	.858	-.043
SNS	.453	.634	.175
Retinoblastoma	1.078	.542	.591
T. Wilms	.659	.461	-.521
T. Hepático	1.253	2.490	.030
T. Óseos	.431	.327	.195
Tej. blandos	.424	1.189	-.106
C. Germinales	.388	.521	.762
Carcinomas	2.294	1.553	-.148
No clasificados	.646	.536	-.286

Tabla 42. Confiianza de los puntos columna

Lugar de residencia	Desviación estándar en la dimensión		Correlación
	1	2	
Aguascalientes	1.443	.926	.057
Baja California	4.324	5.406	.327
B.C.S.	3.503	3.330	-.799
Campeche	1.608	2.563	.252
Coahuila	2.561	1.406	.661
Colima	1.693	5.707	-.294
Chiapas	.747	.677	-.183
Chihuahua	3.475	2.732	-.462
D.F.	.552	.157	.808
Durango	.966	1.948	.873
Guanajuato	.999	.194	-.588
Guerrero	.415	.301	-.582
Hidalgo	.317	.213	.226
Jalisco	.668	.794	.040
México D.F.	.174	.196	.203
Michoacán	.430	.189	-.010
Morelos	.392	.366	.449
Nayarit	2.715	.674	-.430
Oaxaca	.543	.590	-.817
Puebla	.301	.900	.365
Querétaro	.842	.991	-.877
Quintana Roo	3.105	5.326	-.049
Sn Luis Potosí	1.888	.928	.250
Sinaloa	1.167	1.319	.828
Sonora	1.881	1.089	.453
Tabasco	1.750	3.169	.900
Tamaulipas	.709	.517	.192
Tlaxcala	.653	.767	.067
Veracruz	.394	.859	-.838
Yucatán	2.106	3.911	-.354
Zacatecas	1.524	.895	-.852

Tabla 43. Conjunto de datos registrados durante el periodo 1980-1992 de cáncer en niños menores de 15 años según escolaridad de la madre. Fuente Unidad de Investigación en Epidemiología Clínica del CMN SXXI. IMSS. 1998.

Escolaridad materna					
Institución de salud	<Primaria incompleta	Primaria completa Secundaria incompleta	Secundaria completa Técnico y bachillerato	Profesionistas	Total
CMNyCMR	225	250	163	37	675
20deNov	17	26	38	52	133
HGSS	62	14	5	1	82
INFMEEX	443	234	107	27	811
INP	962	454	318	70	1804
<b>Total</b>	<b>1709</b>	<b>978</b>	<b>631</b>	<b>187</b>	<b>3505</b>

Tabla 44. Resumen del análisis sobre las variables en estudio.

Dimensión	Valor Singular	Inercia	Ji cuadrado	Sig.	Proporción de inercia		Confianza del valor singular	
					Explicada por	Acumulada	Desviación Estándar	Correlación 2
1	.327	.107			.805	.805	.029	.304
2	.154	.024			.178	.983	.017	
3	.048	.002			.017	1.000		
<b>Total</b>		.133	465.366	.000a	1.000	1.000		

a 21 grados de libertad.

Tabla 45. Resumen de los puntos renglón(a)

Institución de salud	Contribución								Total
	Masa	Puntuación en la dimensión		Inercia	Del punto a la inercia de la dimensión		De la dimensión a la inercia del punto		
		1	2		1	2	1	2	
CMNyCMR	.193	-.227	.734	.019	.030	.674	.168	.831	.999
20deNov	.038	-2.726	-.623	.094	.863	.096	.976	.024	1.000
HGSS	.023	.665	-1.002	.007	.032	.153	.481	.514	.996
INFMEEX	.231	.242	-.134	.006	.041	.027	.687	.099	.786
INP	.515	.147	-.123	.006	.034	.051	.642	.212	.853
<b>Total</b>	1.000			.133	1.000	1.000			

a Normalización simétrica.

Tabla 46. Resumen de las puntuaciones columna (a).

Escolaridad materna	Contribución								Total
	Masa	Puntuación en la dimensión		Inercia	Del punto a la inercia de la dimensión		De la dimensión a la inercia del punto		
		1	2		1	2	1	2	
<Primaria incompleta	.488	.344	-.324	.027	.177	.333	.706	.294	1.000
Primaria completa, Secundaria incompleta	.279	.016	.439	.009	.000	.350	.002	.906	.909
Secundaria completa, técnico y bachillerato	.180	-.313	.386	.011	.054	.175	.514	.368	.881
Profesionistas	.053	-2.171	-.641	.086	.769	.143	.960	.039	.999
<b>Total</b>	1.000			.133	1.000	1.000			

a Normalización simétrica.

Tabla 47. Confianza de los puntos renglón

Institución de salud	Desviación en la dimensión		Correlación
	1	2	
CMNyCMR	.101	.047	.421
20deNov	.182	.164	-.765
HGSS	.131	.085	.535
INFMEEX	.046	.082	-.024
INP	.026	.043	-.031

Tabla 48. Confianza de los puntos columna

Escolaridad materna	Desviación en la dimensión		Correlación
	1	2	
<Primaria incompleta	.037	.031	.642
Primaria completa Secundaria incompleta	.061	.065	-.090
Secundaria completa Técnico y bachillerato	.100	.106	.317
Profesionistas	.182	.139	-.796

Tabla 49. Datos de las principales instituciones de salud en el D.F. que atienden a la población mexicana.

## Escolaridad paterna

Institución de salud	<Primaria incompleta	Primaria completa Secundaria incompleta	Secundaria completa Técnico y bachillerato	Profesionistas	Marginal Total
CMNyCMR	149	254	209	63	675
20deNov	6	25	24	78	133
HGSS	51	23	7	1	82
INFMEEX	391	248	123	49	811
INP	894	469	299	142	1804
<b>Marginal Total</b>	<b>1491</b>	<b>1019</b>	<b>662</b>	<b>333</b>	<b>3505</b>

ESTA TESIS NO SALE  
DE LA BIBLIOTECA.

Tabla 50. Resumen de la inercia de las diferentes dimensiones alcanzadas.

Dimensión	Valor Singular	Inercia	Ji cuadrado	Sig.	Proporción de la inercia		Confianza del Valor Singular	
					Explicado por	Acumulativa	Desviación estándar	Correlación 2
1	.355	.126			.734	.734	.024	.290
2	.211	.044			.259	.992	.017	
3	.037	.001			.008	1.000		
<b>Total</b>		.172	601.678	.000a	1.000	1.000		

a 21 grados de libertad.

Tabla 51. Resumen de los puntos renglón(a)

Institución de salud	Masa	Puntuación en la dimensión		Inercia	Contribución				Total
		1	2		Del punto a la inercia de la dimensión		De la dimensión a la inercia del punto		
					1	2	1	2	
CMNyCMR	.193	-.287	.903	.039	.045	.745	.145	.855	1.000
20deNov	.038	-2.805	-.773	.111	.841	.108	.957	.043	1.000
HGSS	.023	.680	-.488	.005	.030	.027	.745	.228	.974
INFMEX	.231	.253	-.106	.007	.042	.012	.802	.084	.886
INP	.515	.170	-.211	.011	.042	.109	.499	.458	.957
<b>Total</b>	1.000			.172	1.000	1.000			

a Normalización simétrica.

Tabla 52. Resumen de los puntos columna(a)

Institución de salud	Masa	Puntuación en la dimensión		Inercia	Contribución				Total
		1	2		Del punto a la inercia de la dimensión		De la dimensión a la inercia del punto		
					1	2	1	2	
<Primaria incompleta	.425	.426	-.398	.042	.218	.320	.658	.341	.999
Primaria completa	.291	.041	.342	-.008	.001	.162	.022	.888	.909
Secundaria incompleta	.189	-.173	.650	.019	.016	.378	.104	.867	.971
Secundaria completa	.095	-1.690	-.557	.103	.765	.140	.940	.060	1.000
Profesionistas									
<b>Total</b>	<b>1.000</b>			<b>.172</b>	<b>1.000</b>	<b>1.000</b>			

a Normalización simétrica.

Tabla 53. Confianza de los puntos renglón

Institución de salud	Desviación estándar En la dimensión		Correlación
	1	2	
CMNyCMR	.125	.048	.582
20deNov	.168	.209	-.822
HGSS	.086	.099	.306
INFMEEX	.042	.062	-.039
INP	.032	.034	.160

Tabla 54. Confianza de los puntos columna

Escolaridad del padre	Desviación estándar En la dimensión		Correlación
	1	2	
<Primaria incompleta	.044	.041	.785
Primaria completa	.057	.060	-.112
Secundaria incompleta			
Secundaria completa	.101	.074	.185
Técnico y bachillerato			
Profesionistas	.122	.131	-.888

**CLASIFICACIÓN DE BIRCH Y MARSDEN PARA CÁNCER EN NIÑOS.**

## GRUPO DIAGNÓSTICO

<b>I. Leucemias</b>	<ul style="list-style-type: none"> <li>a) Leucemias agudas linfocíticas</li> <li>b) Leucemias agudas no linfocíticas</li> <li>c) Leucemia mieloide crónica</li> <li>d) Otras leucemias</li> <li>e) Leucemia inespecífica</li> </ul>
<b>II. Linfomas y otras neoplasias reticuloendoteliales</b>	<ul style="list-style-type: none"> <li>a) Linfoma Hodgkin</li> <li>b) Linfoma no Hodgkin</li> <li>c) Linfoma de Burkitt</li> <li>d) Otras neoplasias reticuloendoteliales</li> </ul>
<b>III. Neoplasias intracraneales e intraespinales</b>	<ul style="list-style-type: none"> <li>a) Ependimoma</li> <li>b) Astrocitoma</li> <li>c) Tumor neuroectodermico primitivo</li> <li>d) Otros gliomas</li> <li>e) Miscelánea de neoplasias intracraneales e intraespinales</li> <li>f) Neoplasia intracraneal e intraespinal inespecífica</li> </ul>
<b>IV. Tumores de Sistema Nervioso Simpático</b>	<ul style="list-style-type: none"> <li>a) Neuroblastoma y ganglioneuroblastoma</li> <li>b) Otros.</li> </ul>
<b>V. Retinoblastoma</b>	
<b>VI. Tumores Renales</b>	<ul style="list-style-type: none"> <li>a) Tumor de Wilms, Rabdoide y sarcoma de células claras</li> <li>b) Carcinoma renal</li> <li>c) Tumor renal maligno inespecífico</li> </ul>
<b>VII. Tumores Hepáticos</b>	<ul style="list-style-type: none"> <li>a) Hepatoblasma</li> <li>b) Carcinoma hepático</li> <li>c) Tumores hepáticos malignos inespecíficos</li> </ul>
<b>VIII. Tumores Malignos de Hueso</b>	<ul style="list-style-type: none"> <li>a) Osteosarcoma</li> <li>b) Condrosarcoma</li> <li>c) Sarcoma de Edwing's</li> <li>d) Otros específicos</li> <li>e) Tumores malignos inespecíficos de hueso</li> </ul>
<b>IX. Sarcoma de Tejidos Blandos</b>	<ul style="list-style-type: none"> <li>a) Rabdomiocarcinoma</li> <li>b) Fibrosarcoma, neurofibrosarcoma y otras neoplasias neurofibromatosas</li> <li>c) Sarcoma de Kaposi</li> <li>d) Otros sarcomas específicos de tejido blando</li> <li>e) Sarcoma inespecífico de tejidos blandos</li> </ul>
<b>X. Tumores de células germinales</b>	<ul style="list-style-type: none"> <li>a) Tumores de células germinales intracraneal e intraespinal</li> <li>b) Otros tumores de células germinales inespecíficas no gonadales</li> <li>c) Tumores de células gonadales</li> <li>d) Carcinoma gonadal</li> <li>e) Otros tumores específicos gonadales</li> </ul>
<b>XI. Carcinoma y otras neoplasias epiteliales</b>	<ul style="list-style-type: none"> <li>a) Carcinoma adrenocortical</li> <li>b) Carcinoma tiroideo</li> <li>c) Carcinoma nasofaríngeo</li> <li>d) Melanoma maligno</li> <li>e) Otros carcinomas no específicos</li> </ul>
<b>XII. Otras Neoplasias Malignas</b>	<ul style="list-style-type: none"> <li>a) Otros tumores malignos específicos</li> <li>b) Otros tumores malignos no específicos</li> </ul>

Birch J M, Marsden H B. A classification scheme for childhood cancer. International Journal Cancer. 1987;40:620-624.

**PROYECTO:**

**INSTRUMENTO DE RECOLECCION DE DATOS**

NUM. DE AFILIACION ..... FOLIO .....

NOMBRE DEL PACIENTE .....

SEXO \_\_\_\_\_ 1) MASCULINO 2) FEMENINO

EDAD AL DIAGNOSTICO \_\_\_\_\_ (meses)

AÑO DEL DIAGNOSTICO \_\_\_\_\_ EDAD ACTUAL \_\_\_\_\_ (meses)

LUGAR DE RESIDENCIA .....

DOMICILIO .....

DELEGACIÓN .....

DELEGACION I.M.S.S. ....

**ANTECEDENTES HEREDOFAMILIARES:**

FAMILIARES CON CANCER \_\_\_\_\_ 1) SI 2) NO 3) NO SABE

PARENTESCO .....

TIPO DE CANCER EN EL FAMILIAR .....

**ANTECEDENTES PERSONALES NO PATOLOGICOS:**

LUGAR DE NACIMIENTO .....

FECHA DE NACIMIENTO ..... MES \_\_\_\_\_ AÑO \_\_\_\_\_

MES DE CONCEPCIÓN .....

ORDEN AL NACIMIENTO .....

EMBARAZO DE TERMINO \_\_\_\_\_ 1) SM 2) NO

PESO AL NACIMIENTO \_\_\_\_\_ (gramos)

OCUPACION DEL PADRE .....

OCUPACION DE LA MADRE .....

ESCOLARIDAD DEL PADRE (años cursados) .....

ESCOLARIDAD DE LA MADRE (años cursados) .....

SALARIO DEL PADRE (nuevos pesos) .....

SALARIO DE LA MADRE (nuevos pesos) .....

SALARIO CONJUNTO (nuevos pesos) .....

NIVEL SOCIECONOMICO \_\_\_\_\_ 1) ALTO 2) MEDIO 0) BAJO

Nº. DE CUARTOS EN LA CASA .....

Nº. DE PERSONAS QUE VIVEN GN LA CASA .....

AGUA INTRADOMICILIARIA \_\_\_\_\_ 1) SI 0) No

ELIMINACION DE EXCRETAS \_\_\_\_\_ 1) FECALISMO 2)LETRINA 3) TIPO INGLES

EDAD DEL PADRE AL NACIMIENTO DEL HIJO \_\_\_\_\_ (años)

EDAD DE LA MADRE AL NACIMIENTO DEL HIJO \_\_\_\_\_ (años)

**ANTECEDENTES PERSONALES PATOLÓGICOS:**

ENFERMEDADES EXANTEMATICAS .....

00) NINGUNA 01) SARAMPION 02) VARICELA 04) RUBEOLA,

08) HEPATITIS 16) ESCARLATINA 32) OTRAS 99) NHD

INTERNAMIENTOS HOSPITALARIOS DURANTE EL PRIMER AÑO DE VIDA .....

1) SI 2) NO 3) NO SABE

MALFORMACIONES \_\_\_\_\_ 1) SI 2) NO 3) NO SABE



TIPO .....

ANTECEDENTES RELACIONADOS CON EL PADECIMIENTO:

FECHA DE INICIO DE SIGNOS Y SÍNTOMAS ..... MES \_\_\_\_ AÑO \_\_\_\_

FECHA DE DIAGNOSTICO DE CANCER ..... MES \_\_\_\_ AÑO \_\_\_\_

TIEMPO TRANSCURRIDO ENTRE EL INICIO DE LOS SÍNTOMAS Y LA FECHA DE DIAGNOSTICO  
 \_\_\_\_ (meses)

TIPO DE CANCER .....

LOCALIZACIÓN .....

01) CABEZA    02) CUELLO                    04) TORX                    08) ABDOMEN

16) PELVIS    32) GENITALES                    64) EXTREMIDADES SUP.

128) EXTREMIDADES INF.                    156) OTRAS

SITIO \_\_\_\_ 1) UNILATERAL                    2) BILATERAL                    3) GENERALIZADA

ESTUDIO HISTOPATOLOGICO \_\_\_\_ 1) SI 0) No.

ESTADIO .....

VARIEDAD HISTOLÓGICA .....

CLASIFICACION INTERNACIONAL DE LAS ENFERMEDADES .....

CLASIFICACION DETALLADA .....

VIVE EL PACIENTE \_\_\_\_ 1) SI 0) NO.

TIENE OTRO CANCER \_\_\_\_ 1) SI 0) NO

DIAGNOSTICO DE SEGUNDO CANCER .....

ESTUDIO HISTOPATOLOGICO II \_\_\_\_ 1) SI 0) NO

**MANUAL DE CODIFICACIÓN.****SECUENCIA DEL CÁNCER EN NIÑOS RESIDENTES EN LA REÚBLICA MEXICANA.**

FOLIO. No necesita codificarse

NOMBRE DEL PACIENTE. No necesita codificarse.

NÚMERO DE AFILIACIÓN. No necesita codificarse.

EDAD (MESES). Ocupa 3 campos

SEXO. Ocupa un campo: 1) Masculino; 0) Femenino.

DOMICILIO. No necesita codificarse

LUGAR DE RESIDENCIA. Ocupa 2 campos. \_\_\_\_\_

- |                         |                     |
|-------------------------|---------------------|
| 01. Aguascalientes      | 17. Morelos         |
| 02. Baja California     | 18. Nayarit         |
| 03. Baja California Sur | 19. Nuevo León      |
| 04. Campeche            | 20. Oaxaca          |
| 05. Coahuila            | 21. Puebla          |
| 06. Colima              | 22. Querétaro       |
| 07. Chiapas             | 23. Quintana Roo    |
| 08. Chihuahua           | 24. San Luis Potosí |
| 09. Distrito Federal    | 25. Sinaloa         |
| 10. Durango             | 26. Sonora          |
| 11. Guanajuato          | 27. Tabasco         |
| 12. Guerrero            | 28. Tamaulipas      |
| 13. Hidalgo             | 29. Tlaxcala        |
| 14. Jalisco             | 30. Veracruz        |
| 15. México              | 31. Yucatán         |
| 16. Michoacán           | 32. Zacatecas.      |

INSTITUCIÓN DE SALUD. Ocupa 2 campos: \_\_\_\_\_

01. Hospital de Pediatría Centro Médico Nacional Siglo XXI (CMN)
02. Hospital de Pediatría de Centro Médico La Raza (CMR).
03. Hospital Clínica N°25.
04. Hospital 20 de Noviembre ISSSTE
05. Hospital López Mateos ISSSTE
06. Hospital General de la Secretaría de Salud.
07. Hospital Infantil de México.
08. Instituto Nacional de Pediatría.

LUGAR DE RESIDENCIA. Ocupa 2 campos: \_\_\_\_\_

- |                           |                |
|---------------------------|----------------|
| 01. Aguascalientes        | 17. Morelos    |
| 02. Baja California Norte | 18. Nayarit    |
| 03. Baja California Sur   | 19. Nuevo León |
| 04. Campeche              | 20. Oaxaca     |
| 05. Coahuila              | 21. Puebla     |
| 06. Colima                | 22. Querétaro  |

- |                      |                     |
|----------------------|---------------------|
| 07. Chiapas          | 23. Quintana Roo    |
| 08. Chihuahua        | 24. San Luis Potosí |
| 09. Distrito Federal | 25. Sinaloa         |
| 10. Durango          | 26. Sonora          |
| 11. Guanajuato       | 27. Tabasco         |
| 12. Guerrero         | 28. Tamaulipas      |
| 13. Hidalgo          | 29. Tlaxcala        |
| 14. Jalisco          | 30. Veracruz        |
| 15. México           | 31. Yucatán         |
| 16. Michoacán        | 32. Zacatecas.      |

FECHA DE NACIMIENTO. Ocupa 2 campos (mes): \_\_\_\_\_

- |            |               |
|------------|---------------|
| 01 Enero   | 07 Julio      |
| 02 Febrero | 08 Agosto     |
| 03 Marzo   | 09 Septiembre |
| 04 Abril   | 10 Octubre    |
| 05 Mayo    | 11 Noviembre  |
| 06 Junio   | 12 Diciembre. |

AÑO DE DIAGNÓSTICO. Ocupa 2 campos: \_\_\_\_\_

- |    |    |    |
|----|----|----|
| 81 | 86 | 91 |
| 82 | 87 |    |
| 83 | 88 |    |
| 84 | 89 |    |
| 85 | 90 |    |

OCUPACIÓN DEL PADRE. Ocupa 2 campos:

ESCOLARIDAD DEL PADRE. Ocupa 2 campos: \_\_\_\_\_

- |                            |                          |
|----------------------------|--------------------------|
| 01. Analfabeta             | 02. Sabe leer y escribir |
| 03. Primaria incompleta    | 04. Primaria completa    |
| 05. Secundaria incompleta  | 06. Secundaria completa  |
| 07. Bachillerato           | 08. Técnico              |
| 09. Otros (profesionistas) |                          |

ESCOLARIDAD DE LA MADRE. Ocupa 2 campos: \_\_\_\_\_

- |                            |                          |
|----------------------------|--------------------------|
| 01. Analfabeta             | 02. Sabe leer y escribir |
| 03. Primaria incompleta    | 04. Primaria completa    |
| 05. Secundaria incompleta  | 06. Secundaria completa  |
| 07. Bachillerato           | 08. Técnico              |
| 09. Otros (profesionistas) |                          |

NIVEL SOCIOECONÓMICO. Ocupa 2 campos: \_\_\_\_\_

- |         |          |         |
|---------|----------|---------|
| 2) Alto | 1) Medio | 0) Bajo |
|---------|----------|---------|

SALARIO DE LA MADRE. No se codifica

SALARIO DEL PADRE. No se codifica.

EDAD DEL PADRE AL NACIMIENTO DEL HIJO. Ocupa 2 campos: \_\_\_\_\_

EDAD DE LA MADRE AL NACIMIENTO DEL HIJO. Ocupa 2 campos: \_\_\_\_\_

TIEMPO TRANSCURRIDO ENTRE EL INICIO DE SÍNTOMAS Y LA FECHA DEL DIAGNÓSTICO EN EL PACIENTE (Tiempo transcurrido, tiempo en meses) Ocupa 2 campos (cuando sea de un mes se codificará como 90)

TIPO DE CÁNCER. Ocupa 2 campos: \_\_\_\_\_

- |                                                                           |                                                          |
|---------------------------------------------------------------------------|----------------------------------------------------------|
| 01. Leucemia linfoblástica aguda                                          | 48. Ganglioneuroma                                       |
| 02. Meduloblastoma                                                        | 49. Hemolinfagioma                                       |
| 03. Neuroblastoma                                                         | 50. Hemangioendotelioma hepático                         |
| 04. Osteosarcoma de miembros inferiores                                   | 51. Hemangioma                                           |
| 05. Linfoma no Hodgkin                                                    | 52. Hepatocarcinoma                                      |
| 06. Carcinoma embrionario de senos endodérmicos (sacro)                   | 53. Tumor mediastinal                                    |
| 07. Sarcoma de Edwin                                                      | 54. Tumor laríngeo                                       |
| 08. Enfermedad de Hodgkin                                                 | 55. Germinoma                                            |
| 09. Tumor rabdoide fosa posterior/rabdomiosarcoma                         | 56. Teratoma maduro                                      |
| 10. Carcinoma embrionario glándula pineal                                 | 57. Cáncer Renal de células claras                       |
| 11. Teratoma inmaduro con tumor de senos endodérmicos (testículo derecho) | 58. Tumor saco vitelino                                  |
| 12. Hepatoblastoma fet                                                    | 59. Mixoma                                               |
| 13. Glioblastoma multiforma                                               | 60. Papiloma plexos coroides                             |
| 14. Leucemia mieloide aguda                                               | 61. Tumor de células de leydin                           |
| 15. Retinoblastoma                                                        | 62. Linfoma histiocítico                                 |
| 16. Carcinoma de glándula suprarrenal                                     | 63. Cáncer de parótida izquierda                         |
| 17. Ependimoma cuarto ventrículo.                                         | 64. Carcinoma indiferenciado                             |
| 18. Tumor de Wilms                                                        | 65. Sarcoma meníngeo                                     |
| 19. Disgerminoma ovario                                                   | 66. Cáncer Verrugoso de pené                             |
| 20. Tumor neuroectodérmico primitivo                                      | 67. Timoma                                               |
| 21. Leucemia no linfoblástica aguda                                       | 68. Gonadoblastoma                                       |
| 22. Fibromatosis maligna                                                  | 69. Cáncer de testículo                                  |
| 23. Carcinoma papilar de tiroides                                         | 70. Cistosarcoma de glándula mamaria                     |
| 24. Astrocitoma fibrilar                                                  | 71. Ameloblastoma                                        |
| 25. Histiocitosis                                                         | 72. Teratoma presacro                                    |
| 26. Sarcoma indiferenciado de hígado                                      | 73. Pinealoblastoma                                      |
| 27. Tumor de ganglios basales                                             | 74. Tumor de senos endodérmicos                          |
| 28. Sinovioma                                                             | 75. Blastoma pulmonar                                    |
| 29. Adenocarcinoma de glándulas salivales                                 | 76. Cáncer nasofaríngeo                                  |
| 30. Tumor de ovario de células glanulomasas                               | 77. Angiofibroma nasofaríngeo                            |
| 31. Nefroblastoma                                                         | 78. Sarcoma indiferenciado de tejidos o partes blandas   |
| 32. Fibrosarcoma                                                          | 79. Adenocarcinoma de riñón                              |
| 33. Linfoma linfocítico maligno                                           | 80. Shwaroma maligno                                     |
| 34. Leiomiomasarcoma                                                      | 81. Tumor de tritón                                      |
| 35. Cáncer de colon                                                       | 82. Pancreatoblastoma                                    |
| 36. Linfangioma                                                           | 83. Melanome maligno                                     |
| 37. Craneofaringeoma                                                      | 84. Tumor benigno de otros órganos genitales de la mujer |
| 38. Neoplasia germinal maligna cerebral                                   | 85. Adenocarcinoma submaxilar                            |
| 39. Ectomesenquioma                                                       | 86. Adenocarcinoma de plexos coroides                    |
| 40. Osteocondroma                                                         | 87. Sarcoma pleomorfo de ojo                             |
| 41. Linfoma de células grandes                                            | 88. Teratoma retroperitoneal                             |
| 42. Cáncer de recto y ano                                                 | 89. Germinoma intraorbitario                             |
| 43. Adenoma hipofisiario                                                  | 90. Carcinoma sacrococcigeo                              |
| 44. Carcinoma epidermoide                                                 | 91. Granuloma de pared abdominal                         |
| 45. Epitelioma                                                            | 92. Fibrohistiocitoma                                    |
| 46. Disgerminoma intracraneano                                            | 93. Carcinoma ambrionario de retroperitoneo              |
| 47. Neuroepitelioma                                                       | 94. Meningioma intradural cervical                       |
|                                                                           | 95. Carcinoma embrionario de ovario                      |