

63  
Lej



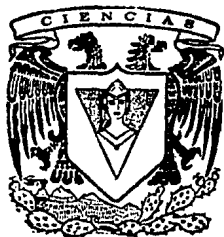
# UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

"ANALISIS ESTADISTICO DE UN ESTUDIO  
EPIDEMIOLOGICO DE LA ENFERMEDAD DE  
CHAGAS EN EL ESTADO DE MORELOS"

T E S I S  
Que para obtener el Título de  
A C T U A R I O  
p r e s e n t a

LETICIA VIGUERAS FLORES



MEXICO, D. F.

1991

TESIS CON  
FALLA DE ORIGEN



## **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 está 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.

## INDICE

INTRODUCCION .....	i
CAPITULO I .....	1
EPIDEMIOLOGIA .....	2
INTRODUCCION A LA EPIDEMIOLOGIA .....	2
LA INVESTIGACION EPIDEMIOLOGICA .....	4
TIPOS DE ESTUDIO .....	6
MODELOS EPIDEMIOLOGICOS .....	15
MEDIDAS MAS USADAS EN EPIDEMIOLOGIA .....	17
CAPITULO II .....	19
PROTOCOLO DE INVESTIGACION .....	20
JUSTIFICACION .....	23
PLANTEAMIENTO DEL PROBLEMA .....	24
OBJETIVOS .....	25
ESTUDIO PILOTO .....	30
CAPITULO III .....	33
TECNICAS ESTADISTICAS PARA ESTUDIOS EPIDEMIOLOGICOS .....	34
TABLAS DE CONTINGENCIA .....	34
INTRODUCCION .....	34
TABLAS DE DOS DIMENSIONES .....	34
TABLAS DE TRES DIMENSIONES .....	40
TABLAS DE CUATRO O MAS CRITERIOS: .....	46

MODELOS LOGLINEALES .....	47
INTRODUCCION .....	47
MODELOS PARA TABLAS DE DOS DIMENSIONES ..	48
MODELOS PARA TABLAS DE TRES DIMENSIONES ..	50
MODELOS PARA CUATRO O MAS DIMENSIONES: ..	54
ESTADÍSTICA DE BONDAD DE AJUSTE .....	55
SELECCIÓN DE UN MODELO: .....	56
PROCEDIMIENTO DE SELECCIÓN STEPWISE .....	58
ESTIMACIÓN DE PARÁMETROS .....	61
CEROS EN LA TABLA .....	62
CAPITULO IV .....	63
ANALISIS DE DATOS .....	64
INTRODUCCION .....	64
DESCRIPCIÓN DEL ESTUDIO .....	64
INTERPRETACION DE TABLAS DE CONTINGENCIA ..	70
INTERPRETACION DE MODELOS LOGLINEALES ..	77
CONCLUSIONES: .....	83
ANEXOS .....	I
PROGRAMA 1 .....	VII
PROGRAMA 2 .....	X
BIBLIOGRAFIA .....	XXI

## INTRODUCCION

El objetivo de este trabajo es presentar los resultados del análisis numérico de un estudio epidemiológico realizado en el estado de Morelos.

Así mismo presentar que por medio de técnicas estadísticas aplicadas a este tipo de estudios, y el uso de herramientas de cómputo, facilitan a los investigadores la tarea de efectuar operaciones, seguir algoritmos manualmente y hacer cálculos tediosos para llegar a un buen análisis.

El estudio que se presenta en este trabajo fue realizado por un grupo de investigadores del depto. de Epidemiología de la Facultad de Medicina de UNAM, coordinados por la Dra. Laura Moreno jefe del departamento, quien nos proporcionó los datos para la realización de este trabajo.

El trabajo está dividido en cuatro capítulos: introducción en el campo de la Epidemiología, protocolo de investigación, técnicas estadísticas para el análisis de los datos y los resultados obtenidos.

El Capítulo I se da una introducción a la epidemiología y un panorama general sobre los tipos de estudios que se utilizan en la investigación epidemiológica.

El Capítulo II presenta el protocolo de investigación del estudio epidemiológico de la enfermedad de Chagas en el edo. de Morelos, el cual fue realizado por la Dra. Guadalupe García del depto. de Epidemiología, y donde se presenta la justificación, objetivos y metodología del trabajo. Así mismo se describe un estudio piloto que dio la pauta para realizar la fase de campo, y así terminar el trabajo sin contratiempos.

El Capítulo III describe las técnicas estadísticas utilizadas para el análisis de los datos recopilados en la encuesta Tripanosomiasis Americana en el edo. de Morelos.

En la primera parte de este capítulo se da una introducción a las tablas de contingencia, y se presenta desde tablas de dos dimensiones hasta multidimensionales. En la segunda parte se da una introducción a los modelos loglineales para tablas de dos dimensiones hasta modelos para tabla multidimensionales.

**El capítulo IV tiene como objetivo presentar el análisis de los datos y la interpretación de los programas realizados en el paquete estadístico Biomedical Packet.**

**Al final del documento se presentan las conclusiones y un anexo donde aparece la encuesta aplicada y los programas utilizados.**

# CAPITULO I

## EPIDEMIOLOGIA

### INTRODUCCION A LA EPIDEMIOLOGIA

La epidemiología estudia la distribución de enfermedades o de condiciones fisiológicas en poblaciones humanas y los factores que determinan tal distribución. En términos más generales, puede decirse que la epidemiología estudia la salud y sus enfermedades en poblaciones humanas, para responder a cuatro objetivos generales:

**Describir el estado de salud de las poblaciones**

**Explicar la etiología de las enfermedades**

**Predecir la frecuencia de las enfermedades y el estado de salud de las poblaciones**

**Controlar la distribución de las enfermedades**

La epidemiología es una disciplina que ha desarrollado métodos relativamente especializados para investigar la etiología de las enfermedades y a su vez utiliza, de acuerdo con las necesidades del momento, conocimientos específicos y habilidades de otras ciencias. Las ciencias que pueden intervenir en un estudio epidemiológico son: Bioquímica, Física, Microbiología, Anatomía, Patología, Demografía, Antropología, Sociología y Estadística:

**Bioquímica:** Proporciona pruebas útiles en la demostración de desordenes metabólicos

**Física:** Proporciona varios instrumentos de diagnóstico.

**Microbiología:** proveen de métodos para aislar y reconocer a los agentes virales, bacterianos, fúngicos y parasitarios, y demostrar el desarrollo de anticuerpos específicos estimulados por algún agente infeccioso.

**Anatomía Patológica:** El diagnóstico basado en el reconocimiento de los cambios gruesos y microscópicos característicos de la enfermedad, por ejemplo el cáncer.

**Demografía:** Ciencia que provee de información sobre la distribución de la población con respecto a sus características necesarias para



describir la frecuencia relativa de la enfermedad en segmentos de la población definidos por esos atributos.

**Antropología-Sociología:** Contribuyen con su acervo de información existente y con sus métodos especiales de encuesta, para adquirir información adicional y factores sociales que puedan influir sobre la salud.

**Estadística:** El epidemiólogo confronta problemas en el diseño de estudio, recolección, tabulación, análisis e interpretación de datos, y la consulta en estadística es esencial como guía en la resolución de los problemas antes mencionados

El propósito más importante de la epidemiología es el descubrimiento de los factores esenciales o contribuyentes a la aparición de cualquier enfermedad particular.

En epidemiología una comprensión de ocurrencia natural de la enfermedad puede dar lugar al desarrollo de métodos efectivos para la prevención de la enfermedad, que interceptan la acción de los factores causales.

Los usos que refuerzan la interrelación entre la clínica y la epidemiología fueron clasificados por Morris en:

**Describir la Enfermedad en la Comunidad:**

La epidemiología describe la enfermedad en una comunidad y la sitúa dentro de los parámetros en que se mueve, traza un perfil de los problemas de salud y sugiere prioridades según sea la naturaleza y la magnitud de los datos.

**Como Instrumento de Predicción:**

El conocer a fondo un problema nos permitirá establecer hipótesis para el futuro, y al no aparecer ningún acontecimiento inesperado se podrán hacer proyecciones futuras, siempre y cuando las cosas se muevan sobre un marco conocido.

**Identificación de Grupos más vulnerables:**

De acuerdo al análisis de las probabilidades individuales de enfermar y morir, se conocen a los grupos expuestos a un mayor riesgo, lo cual podría permitir aplicar programas de bajo costo y alto rendimiento. Por lo tanto se pueden situar de acuerdo a la enfermedad a los grupos humanos más expuestos al riesgo.

**Evaluar:**

Evaluar implica medir utilizando los indicadores adecuados para asignar el valor objetivo que tienen las múltiples acciones de salud, desde el uso de una vacuna, hasta el rendimiento de estructuras complejas de salud.

**Completar el Cuadro Clínico de la Enfermedad:**

Un epidemiólogo por medio de la observación puede percibir las variaciones de una enfermedad en su escenario natural a diferencia de aquello que puede percibir en un consultorio u hospital. Y si conjuntamos visión y conocimiento se pueden identificar estadios preclínicos.

**Identificar Síndromes Nuevos:**

Debido a la constante observación de los acontecimientos del ambiente natural se han logrado grandes aportaciones a la medicina.

**Identificar Factores Etiológicos (o causales):**

Mediante el análisis de asociación de variables y uso de una adecuada metodología epidemiológica, se han identificado factores etiológicos.

**LA INVESTIGACION EPIDEMIOLOGICA**

La investigación epidemiológica, además de describir la ocurrencia de la enfermedad y la características del agente primario, ha tratado de aumentar la precisión en los diagnósticos, para entender la patogenia, para determinar los mecanismos de transmisión y reservorio, para descubrir y evaluar los factores contribuyentes, y para desarrollar y evaluar medidas de prevención y control.

El objetivo de la investigación epidemiológica al nivel de comprensión es hacer generalizaciones acerca de la historia natural de una enfermedad, la cual puede dividirse en cuatro etapas (Mausner J. y Bahn A. 1977)

- 1) Etapa de susceptibilidad: En esta etapa la enfermedad no se ha desarrollado pero se encuentran presentes los factores que favorecen su ocurrencia.

- 2) Etapa de la Enfermedad Preclínica: En esta etapa no se ha manifestado la enfermedad, pero ya se ha iniciado el proceso patológico.
- 3) Etapa de la enfermedad clínica: Es cuando los cambios anatómico-funcionales son suficientes para que haya signos y síntomas reconocibles de enfermedad.
- 4) Etapa de Resultado de Enfermedad: En esta etapa se manifiesta el resultado de la enfermedad, ya sea recobrase, remitir, cambiar, incapacidad o muerte.

El objetivo de la investigación epidemiológica a nivel de intervención, es desarrollar prácticas, programas y políticas en materia de salud, con la finalidad de prevenir y controlar las enfermedades y promover buena salud.

Una de las metas de mayor importancia en la investigación epidemiológica es establecer el potencial de causalidad de agentes que todavía no se han encontrado asociados a la enfermedad.

Como ya se mencionó las investigaciones epidemiológicas se ocupan de entender la ocurrencia de la enfermedad y la formulación de métodos para su solución, y estas se han clasificado en:

#### **Laboratorio o Campo**

Son por naturaleza experimentales y pueden tomar diversas formas. Un ejemplo es el trabajo de Enders al demostrar que los polivirus se multiplican en histocultivos produciendo efectos destructores visibles. Este investigador aportó una técnica para detectar los polivirus en el hombre y en sus fuentes ambientales sospechosas.

#### **Experimentales**

Las investigaciones experimentales manipulan a las poblaciones con las siguientes características: El factor bajo estudio es propuesto por el investigador, se asignan los grupos experimentales a los grupos de tratamiento, y los grupos testigos mediante un método formal de distribución al azar.

#### **Observacionales:**

Son los más usados, se enfoca sobre eventos principalmente "posibles", exposiciones y casos de enfermedad que han ocurrido ó están

ocurriendo en la población bajo estudio, y la agrupación de datos se lleva a cabo conforme a las características que presente la población.

La observación implica que el investigador no manipula la población bajo estudio, y el análisis implica la búsqueda de una asociación entre la ocurrencia de la enfermedad y las posibles influencias causales.

En las investigaciones epidemiológicas, un primer paso en el estudio de una enfermedad es el describir con exactitud su ocurrencia en la población.

Con respecto a la información básica que se recolecta de cada paciente, se refiere al tiempo, lugar y atributos personales, y el análisis cuidadoso de esta información, revelará los patrones generales de ocurrencia de la enfermedad y permitirá reconocer los grupos que tienen un mayor riesgo de adquirir la enfermedad.

Las pistas sobre los factores causales de una enfermedad pueden deducirse por analogías con enfermedades conocidas que poseen patrones de ocurrencias similares ó al grupo de miembros con alto riesgo.

Un principio básico es que una hipótesis aceptable debe ser compatible con los hechos conocidos sobre la ocurrencia de la enfermedad.

La investigación puede centralizarse en métodos de diagnóstico, o en posibles mecanismos de transmisión, y esta tomará la forma de estudio de campo centrado alrededor de la ocurrencia de la enfermedad.

## TIPOS DE ESTUDIO

El proceso salud-enfermedad ha planteado la necesidad de utilizar diversos enfoques con base en el objetivo de la investigación, el avance de ciencias médicas y consideraciones de tipo ético.

El objeto central de la investigación clínico-epidemiológico es el estudio de una población, a diferencia del estudio de casos que maneja únicamente numeradores.

Para clasificar una investigación se considerarán las siguientes características:

- 1) El período en que se capta la información
- 2) La evolución del fenómeno estudiado
- 3) La comparación de poblaciones
- 4) La inferencia del investigador en el estudio.

#### **Información**

**Retrospectivo:** Un estudio es retrospectivo, si la información se capta antes de que se planea la investigación y con fines ajenos a está.

**Retrospectivo Parcial:** Se presenta cuando un estudio, cuenta con una parte de la información, y el resto se obtendrá conforme lo planteo la investigación.

**Prospectivo:** Un estudio es prospectivo, cuando toda la información se obtendrá conforme a los criterios de la investigación y para los fines específicos de esta.

#### **Evolución del fenómeno**

**Longitudinal:** Un estudio es longitudinal, cuando se miden las variables involucradas, y se hace un seguimiento para observar la evolución de los sujetos en el tiempo, llevando a la comparación de los valores de las variables en cada sujeto y en diferentes momentos.

**Transversal:** Un estudio es considerado transversal, si las variables se miden una sola vez, y los grupos de unidades se miden en un momento determinado.

#### **Comparación de Poblaciones**

**Descriptivo:** Un estudio es descriptivo, cuando existe una población en la que no existen hipótesis centrales y que se pretende describir en función de un grupo de variables.

**Comparativo:** Un estudio es Comparativo, si existen dos o más poblaciones y se pretende comparar algunas variables para contrastar una o varias hipótesis centrales.

Los estudios comparativos, en cuanto abordar el fenómeno se divide en:

- 1) De efecto a causa: Se investigan dos o más grupos de unidades de estudio que se diferencian en varias modalidades de un factor causal y se estudia el desarrollo de éstas para evaluar, conocer y analizar el efecto y la frecuencia de la aparición de aquél dentro de cada grupo.
- 2) De causa a efecto: Se parte de dos o más grupos de unidades de estudio que presentan cierto fenómeno considerado como efecto en varias modalidades y se retrocede al pasado para determinar o conocer el factor causal, y la proporción en que éste presentó en los diferentes grupos

#### Inferencia del Investigador

**Observacional:** Si el estudio es de tipo observacional, el investigador no debe intervenir para modificar los factores que intervienen en la investigación, sólo puede describir o medir el fenómeno en estudio.

**Experimental:** En un estudio de tipo observacional, el investigador puede modificar las variables que intervienen en el proceso. Es posible en este tipo de estudio asignar al azar las unidades a la diversas variantes del factor causal.

La combinación lógica de estas características nos dan como resultado la estructura específica de cada uno de los estudios.

Las relaciones de cuasalidad en el comportamiento, aprendizaje, estado de salud, y demás aspectos de los seres humanos es laborioso, y a menudo se requiere de investigaciones que primero sugieran una hipótesis y después sea apoyada dentro de contextos diferentes.

Ningún estudio puede corroborar una hipótesis únicamente apoya la veracidad de ésta. Los diferentes tipos de estudios sugieren o proporcionan grados variables de apoyo a una hipótesis; de aquí que se tengan numerosos estudios sobre una misma relación de causalidad. Sin embargo no es necesario que se elaboren todos los tipos de estudios ni que éstos incluyan orden determinado. A continuación se presentan los diferentes tipos de estudios.

## ENCUESTA DESCRIPTIVA

### Características del estudio:

OBSERVACIONAL/PROSPECTIVO/RETROSPECTIVO/TRANSVERSAL/DESCRIPTIVA

La encuesta descriptiva es un estudio que considera una sola población, en la cual deben ser definidas las características generales y la ubicación en el tiempo y en el espacio; se pretende describir la situación de la población en un momento determinado, de acuerdo con algunas variables.

No existe una hipótesis central aunque pueden plantearse hipótesis que posteriormente puedan ser contrastadas.

### Características Generales

En el caso de que la población esté constituida por una cantidad considerable de unidades, que no puedan estudiarse con los recursos humanos, financieros y físicos disponibles, se requiere de utilizar métodos de muestreo, que permitan estudiar a la población con un número más reducido de unidades pero sin perder representatividad de la población objetivo.

Para hacer la selección de variables tenemos que considerar cual o cuales son los objetivos de la investigación y definir adecuadamente las que se utilizarán, para evitar sesgos de información que alteren los resultados.

En el caso de estudios prospectivos se elaboran hojas de captación de datos para registrar la principales variables, en general cuando el número de variables es grande se requiere precodificar, para facilitar el proceso de información.

En los estudios retrospectivos o descriptivos la información es captada de fuentes primarias: Estadísticas demográficas, de mortalidad, de morbilidad y hospitalarias; documentos personales, y expedientes clínicos.

Esta información debe ser tratada con especial atención, ya que puede presentar características que no coincidan con los conceptos que el investigador esta manejando en su trabajo, ya que pudo ser captada por diferentes causas y por medios diferentes, por lo tanto una vez obtenida la información debe ser clasificada y ordenada, posteriormente agrupar los resultados de medición, en función del objetivo de la investigación.

## ENCUESTA COMPARATIVA

### OBSERVACIONAL/PROSPECTIVO/RETROSPECTIVO,TRANSVERSAL/COMPARATIVA

En este tipo de estudio se consideran dos o más poblaciones se comparan algunas de las variables en ocasión única, con el objeto de contrastar una o varias hipótesis.

#### Características generales

En los estudios comparativos es de gran importancia definir a cada una de las poblaciones, ya que los factores que se consideran causa o efecto dentro de la relación de causalidad, pueden aparecer sólo en una de las poblaciones; otra manera de definir a la población, es la presencia del factor efecto en varias modalidades.

La población objetivo debe ser definida espacio-temporalmente, y en el caso de no poder ser ubicada alguna de las poblaciones, se deberá tomar una muestra representativa, es decir, el tener que asegurarse de que las conclusiones obtenidas en los grupos de estudio se puedan extrapolar a la población objetivo..

Si la población objetivo es muy grande deberán emplearse métodos de muestreo, cuidando que la muestra resulte representativa de la población objetivo.

Debido a que este tipo de estudio tiene dos ó más poblaciones, se pueden evaluar las diferencias entre las poblaciones, siempre y cuando estas sean comparables,es decir exista homogeneidad en las muestras respecto a los posibles factores de confusión, en caso de no serlo, se podría aplicar igualación de atributos o formación de bloques.

La información que se capta para este tipo de estudio se obtiene de fuentes secundarias:

Para estudios retrospectivos: expedientes, radiografías, grabaciones, censos.

Para estudios prospectivos: Estadísticas vitales, entrevista abierta o estructurada, cuestionarios, entre otros.

Esta información deberá ser verificada para garantizar su precisión y validez, una vez que fué validada. es necesario contrastar la hipótesis central por medio de procesos estadísticos, los cuales se encargaran de seleccionar con base en los objetivos las variables involucradas.



## REVISION DE CASOS

Características del estudio:

OBSERVACIONAL/RETROSPECTIVO/LONGITUDINAL/DESCRIPTIVO

En este tipo de estudio se pretende conocer la evolución del fenómeno estudiado en el pasado, en relación con ciertas variables que se miden en diversas ocasiones a las unidades de una población.

Características Generales

En el estudio de revisión de casos se trabaja con una sola población, a la cual no es posible ubicar en el espacio y en tiempo, dado que el origen del registro de las unidades que constituyen a la población proviene de documentos como: expedientes, placas, registros, resultados de exámenes, entre otros.

Debido a que estudiar a el total de la población en una investigación resulta muy complicado, es necesario utilizar técnicas de muestreo, sin olvidarse de no perder representatividad, es decir, si las conclusiones obtenidas en un grupo que se está estudiando se puede extrapolar a la población objetivo; en caso de tener un gran número de expedientes es necesario tomar una submuestra, que sea representativa, y que el número de unidades sea suficiente para que las inferencias o extrapolaciones sean validas.

En la información recabada por medio de estos documentos se tiene el problema de la credibilidad, dado que se reunió por diferentes medios y personas ajenas a la investigación, por lo tanto es necesario verificar la validez de los datos.

En general se recomienda diseñar una hoja de registro de datos elaborada en forma especial para la investigación.

## CASOS Y CONTROLES

Características del estudio:

OBSERVACIONAL/RETROSPECTIVO/LONGITUDINAL/COMPARATIVO\* EFECTO/CAUSA

El propósito de este estudio es conocer si el o los grupo de casos tiene mayor frecuencia de presentación del posible factor etiológico del padecimiento que los grupos de control. Para este tipo de estudio se forman uno ó más grupos de sujetos que presenten un determinado resultado-enfermedad y uno o más grupos que no presenten dicho efecto.

Se recaba información acerca de la exposición basada en uno o más factores considerados como riesgos, determinándose la proporción o grado de exposición a este factor entre los grupos casos y controles.

### **Características Generales**

En este tipo de estudios se tienen dos poblaciones objetivo, una es la población de casos y otra la de controles, y en relación con estas poblaciones deberán definirse las características generales de ubicación en el espacio y en el tiempo.

Para que una unidad (individuo, familia) sea parte de la población de casos y/o controles, debe definirse operacionalmente el factor que se considera como efecto y especificar los métodos de diagnóstico y la forma en que se medirá dicho factor. Cuando los estudios se realizan en hospitales la selección de controles deberá hacerse a partir de individuos cuyo diagnóstico no este asociado al factor de riesgo que se estudia.

Los estudios en los que la fuente de información está basada en los documentos es difícil ubicarla en el espacio y el tiempo, ya que no se tienen criterios definidos sobre la procedencia geográfica o temporal de los sujetos que la integrarán.

Es importante mencionar que la población objetivo se determina en base al origen de los documentos, de tal manera que el documento es tan importante como la población misma.

En el caso de comparar las muestras y contrastar una hipótesis, deberán existir las diferencias entre poblaciones y los factores de causa.

En general es muy difícil estudiar todas las unidades de la población objetivo de casos y controles, por lo que se requiere de muestreo o tomar submuestras sin perder representatividad.

Para este tipo de estudios la captación de información debe planearse cuidadosamente ya que está se obtiene a partir de muestras de suero, grabaciones o expedientes, por los que se debe tener especial atención en el origen de está; en el caso de que se desee información adicional se puede obtener directamente en las unidades de estudio.

Una vez que se obtuvo la información, será necesario ordenarla y clasificarla y agrupar los resultados que se obtengan en función de los objetivos de la investigación. Posteriormente se podrán determinar la asociación de variables y comparar parámetros estadísticos entre las poblaciones de interés

**PERSPECTIVA HISTORICA****Características del estudio:****OBSERVACIONAL/RETROSPECTIVO/LONGITUDINAL/COMPARATIVO\* CAUSA/EFEECTO**

El propósito del estudio es conocer que grupo tuvo mayor incidencia en cuanto al efecto, es decir si el grupo expuesto al factor considerado como riesgo o el no expuesto. Se forman muestras con sujetos que pertenecían a poblaciones que estuvieran expuestas o no a uno o más factores, considerados como el riesgo, tales como tratamientos, agentes o atributos.

Es de importancia considerar que los sujetos que formen a la población, deberán cumplir con las características que los ubiquen en alguno de los dos grupos: expuestos o no expuestos, considerando principalmente que el factor causal potencial no debe estar presente en la población no expuesta, y definir otros aspectos tales como: la entidad nosológica, criterios de morbilidad, edad, sexo, raza, lugar de nacimiento, entre otros.

En particular deberá definirse operacionalmente el factor considerado como causa, y la forma en que será medido.

La población objetivo está dada por los individuos que sus documentos estén disponibles, y se consideren como una muestra de la población.

Generalmente este tipo de estudios no tiene ubicación espaciotemporal, ya que la información se encuentra en expedientes o archivos, aunque en algunos casos existen archivos especiales donde es posible localizar a la población que estuvo expuesta a un factor de riesgo, y en estos casos podremos ubicarla en el espacio y en el tiempo.

Respecto a la información, se encuentra generalmente en: expedientes clínicos, resultados de exámenes, grabaciones, fotografías, y para captarla es necesario que la forma de vaciamiento de datos que se diseña especialmente para cada proyecto, sea homogéneo y estandarizado.

En ocasiones se requiere de mayor información obtenida prospectivamente, y se necesita localizar al sujeto, y tener cuidado en los posibles sesgos, dado que es posible que el grupo de expuestos al factor de riesgo esté mucho más sensible sobre la posible ocurrencia del efecto, que el grupo de no expuestos.

Por último una vez obtenidos los resultados, se podrá determinar la asociación entre variables, hacer comparaciones entre poblaciones de interés, aquellas que presenten el factor causal y las que no lo presenten.

## ESTUDIO DE UNA O VARIAS COHORTES

Características del estudio:

OBSERVACIONAL/PROSPECTIVO/LONGITUDINAL/DESCRPTIVO/COMPARATIVO

En el estudio de una cohorte se tiene una población, la cual presenta un proceso de seguimiento para conocer su evolución. Aquí se presenta un grupo de individuos que tuvieron alguna experiencia en común como enfermedad, tratamiento, o que compartan alguna característica específica

En un estudio de dos o más cohortes, se consideran dos o más poblaciones, que presentan diferentes modalidades del factor causal, y que se someten a un proceso de seguimiento con el fin de conocer su evolución.

### Características Generales

La población objetivo es aquella que está constituida por personas que pudieron haber sufrido alguna experiencia de exposición a riesgos inusitados cuyos efectos se desea evaluar, y sobre la cual se inferirán los hallazgos de la cohorte, en caso de ser hallazgos negativos no excluyen la posibilidad de que haya ocurrido un exceso de enfermedad antes de la definición de la cohorte, por lo tanto deberá definirse el período durante el cual se seguirá a los individuos de esa población, y las características que determinen al sujeto parte de la cohorte o cohortes.

Se presentan tres casos cuando queremos ubicar a la población espacio-temporalmente:

- 1) La población puede ser ubicada en el tiempo y en el espacio.
- 2) La población no puede ser ubicada, cuando una de las poblaciones está ubicada espacio-temporalmente y la otra no.
- 3) En el caso de que los grupos que forman a la cohorte sean heterogéneos, se toma una submuestra de cada estrato, verificando la representatividad.

En estudios de varias cohortes, se requiere de información adicional que clasifique a la población en expuesto y no expuestos al factor que se investiga, ya que pueden presentarse poblaciones que no estuvieron expuestas a un factor en especial.

## MODELOS EPIDEMIOLOGICOS

El problema central sobre la causalidad, es cómo determinar o bajo que criterio decir que un factor es causa de otro, para ello se requiere de un buen estimador de la asociación entre el supuesto factor causal y el factor respuesta.

Para decidir si es estadísticamente significativo y con ello determinar si existe una relación causal para ello requerimos de:

- Fuerza de asociación
- Claridad en la secuencia de eventos en el tiempo
- Existencia de una Dosis-Respuesta
- Consistencia en los resultados
- Factibilidad biológica de la hipótesis
- Coherencia en los resultados

Sin embargo ninguno de los criterios anteriores son determinantes y se consideran sólo como auxiliares en la evaluación del factor de riesgo.

Se mencionarán algunos de los modelos que permiten describir la relación de causalidad:

### Modelo de causalidad determinística unifactorial:

Bajo este modelo se dice que X es causa de Y, si en un sistema donde todos los factores están inicialmente fijos, cualquier manipulación o cambio sólo en X induce un cambio subsecuente sólo en Y. Esta definición requiere que se cumplan dos condiciones:

- La especificidad de la causa.
- La especificidad del efecto.

**Modelo de causalidad determinístico multifactorial**

Este es un modelo de determinismo modificado para permitir múltiples causas de una enfermedad; se definen conglomerados de factores causales, los cuales son tratados como causas suficientes y cada conglomerado suficiente tiene un efecto en la enfermedad, que es independiente de los efectos de los factores que están en otro conglomerado.

**Modelo Epidemiológico:**

Actualmente en epidemiología no se utiliza un modelo causal determinístico, sino un probabilístico, es decir, se utiliza teoría de probabilidad y técnicas estadísticas para probar y estimar a partir de observaciones, relaciones que se han hipotetizado como causales.

Esto no niega que no se utilice un modelo determinístico, sino que se utiliza la teoría de probabilidad porque se tiene un conocimiento limitado del proceso causal y una capacidad limitada de observarlo.

Debido a que se tiene un conocimiento limitado del proceso causal, no se utiliza el término factor causal, sino factor de riesgo, para indicar que se trata de una variable que se cree está relacionada con la probabilidad de que un individuo desarrolle la enfermedad.

El factor de riesgo es el agente cuyo efecto en el resultado de interés que está siendo estudiado. El factor tratamiento es aquel cuyo efecto es el resultado de interés que está siendo estudiado, cuando dicho agente se aplicó específicamente para evaluar el factor resultado bajo consideración.

Para encontrar un factor de riesgo es necesario que este asociado estadísticamente con el desarrollo de la enfermedad (factor respuesta); que la presencia del factor preceda a la ocurrencia de la enfermedad y finalmente que el efecto del factor (asociación) que se observa no sea completamente explicado por cualquier fuente de error, incluyendo el error por muestreo, la presencia de otros factores de riesgo o en general cualquier problema en el diseño o análisis de la investigación.

## MEDIDAS MAS USADAS EN EPIDEMIOLOGIA

### Medidas de Morbilidad:

Esta clasificada en dos tipos: de Incidencia y Prevalencia.

**Incidencia:** son medidas de frecuencia basadas en casos nuevos que ocurren en un período determinado, las cuales reflejan cambios en el estado de la enfermedad y son usados para identificar el factor de riesgo; a su vez están clasificadas en dos grupos: de tasa (es el potencial instantaneo para el cambio en el estado de enfermedad por unidad de tiempo) y las de riesgo(probabilidad de que un individuo no enfermo desarrolle una enfermedad en un período de tiempo determinado, pero condicionado a que no muera por otra enfermedad.

**Prevalencia:** son medidas de frecuencia basadas en los casos existentes en una población en un periodo de tiempo determinado. Los casos de prevalencia representan sobrevivientes de una enfermedad, y no son adecuados para identificar un factor de riesgo, son utilizadas con frecuencia para administrar y planear servicios medicos.

### Medidas de Mortalidad

Son medidas de frecuencia en las que el resultado de interés es la muerte; por la facilidad de obtención y confiabilidad son usados para probar hipótesis etiológicas, así como para evaluar programas preventivos y terapéuticos.

### Medidas del efecto del Factor de Riesgo

Las medidas del factor de riesgo son aquellas que sirven para evaluar el efecto que el factor de riesgo tiene en el evento respuesta, con la finalidad de investigar la asociación.

Para elegir una medida requerimos:

Escala de medición del factor de riesgo.

Modelo para expresar como el factor de riesgo afecta al factor respuesta.

Objetivo del estudio

De acuerdo a la escala en que se mide el factor de riesgo y respuesta tenemos:

**CATEGORICO:**

NOMINAL

ORDINAL

**CUANTITATIVO:**

CONTINUA

DISCRETA



## **CAPITULO II**

## PROTOCOLO DE INVESTIGACION

La enfermedad de Chagas o Tripanosomiasis Americana, es actualmente un problema de salud pública en muchos países de Centro y Sudamérica y es una de las seis enfermedades tropicales más importantes del mundo. Esta enfermedad fue descrita por el médico brasileño Carlos Chagas, quién descubrió su agente etiológico, patogénesis y modo de transmisión en 1909 en Minas Gerais, Brasil.

Se han realizado estudios donde se estima que 90 millones de personas viven en áreas endémicas, expuestas al riesgo de enfermar. Entre 16 y 18 millones de personas tienen actualmente el padecimiento y alrededor de 200 mil personas adquieren la enfermedad cada año.

Se estima que el 62% de los individuos infectados nunca desarrollan sintomatología, del resto, el 27% desarrollan cardiopatía, y el 3% patología digestiva, mientras el 8% desarrollan ambas.

Esta parasitosis fue notificada en nuestro país por primera vez en el año de 1940, cuando Mazzotti descubrió los dos primeros casos de enfermedad de Chagas en humano, en una comunidad del estado de Oaxaca.

A partir de entonces se iniciaron una serie de estudios encaminados a conocer la distribución geográfica, reservorios, características clínicas, de este padecimiento en el país.

Se han reportado más de 100 especies en el mundo, en México son 40 las reportadas, algunas con hábitos intradomiciliarios muy acentuados como el caso de triatoma barberi y triatoma dimidiata, consideradas como los transmisores más importantes para el país debido a sus hábitos domiciliarios.

En nuestro país estos insectos hematófagos son conocidos con diferentes nombres vulgares como: chinche besucona, hocicona, del monte, de campestela, ahorcadora, voladora, y en algunos países de Sudamérica se le conoce como: vinchuca, pitos, barberios o chipos.

Como se mencionó, las triatomas son hematófagos estrictos y bastante voraces, los buenos transmisores defecan, mientras se alimentan o inmediatamente después. En sus heces se encuentran los Tripomastigotes Metacíclicos, fase infectante para el hombre, los cuales penetran por la piel sana, microheridas o mucosas, favoreciendo su entrada el rascado que se produce por las sustancias pruriginosas de la saliva del transmisor.

Se inicia así la fase aguda o aparente de la enfermedad, en la que se presenta el signo de Romaña y/o el chagoma de inoculación, edema facial y fiebre. Después de la etapa aguda, el paciente suele "curar" espontáneamente y pasar asintomático durante algún período, que a veces es de varios años.

La fase crónica está caracterizada por bloqueos de conducción eléctrica cardíaca, que puede ocasionar desde insuficiencia cardíaca progresiva, hasta la muerte súbita. En aparato urinario puede provocar megauréteres y megavejiga. Una vez que se presenta el daño, esté es irreversible e incurable.

El riesgo de infectarse está dado por:

- Infestación (vector presente)
- Colonización (presente en el domicilio)
- Infección (que esté infectado)

Los reservorios vertebrados no humanos de *Triatima cruzi*, que se han encontrado en México son: tlacuache, rata noruega, armadillo, perro, ratón, ardilla y ratones de campo.

La enfermedad está limitada al continente americano y se encuentra distribuida desde México, hasta Argentina y Chile.

Es importante tomar en cuenta que existe una activa y constante migración hacia las áreas urbanas, de personas infectadas por el parásito, las cuales proceden de sectores en donde la enfermedad tiene un carácter endémico, por lo que está parasitosis, puede estar extendida por todo el territorio nacional.

Lo anterior podría presentar un problema para la Salud Pública en los Centros Urbanos de México, tal como ocurre en algunos países de Sudamérica.

En México, en el edo. de Oaxaca, se realizó un estudio en 60 comunidades, en el que se colectaron 4,023 muestras sanguíneas, de las cuales resultaron Seropositivas el 16.3%(8). Esto varió de comunidad en

comunidad, encontrándose desde Seronegatividad en seis localidades, hasta un 50% de Seropositividad en la comunidad de Cerro del Aire. En esta última se observó que el 76% de los individuos mayores de 20 años, produjeron una reacción seropositiva, lo cuál representa uno de los porcentajes más altos reportados en México<sup>1</sup>

Investigaciones sobre vectores en Oaxaca, muestran que existen en el Estado, varios tipos de Triatomas, todos ellos con infección natural por *Triatoma cruzi* y por lo tanto capaces de transmitir la enfermedad al hombre.

Jalisco es uno de los estados en que se han realizado mayor número de estudios. Desde 1967 Cuarte y Cols., describieron los primeros casos de tripanosomiasis, y hasta 1987 se habían encontrado tres pacientes asintomáticos con seropositividad para *Triatoma cruzi* o en fase crónica

En Guerrero se encontró, en un estudio realizado en 216 individuos, que el 8.5% tenían reacción de Fijación del complemento positivo. El transmisor principal fué el armadillo.

En Chiapas, la mayoría de los estudios se han realizado al norte del edo. En la comunidad de Agua Azul Chiquito el 28% de 80 personas, resultaron seropositivas, en la comunidad de León Brindis el 14% de 73 personas lo fueron. También se observó en el mismo estudio, que de 150 niños menores de 10 años, de 5 comunidades, el 4% resultó seropositivo, indicando que la transmisión continúa.

En Yucatán se encontró que de 1545 individuos, el 11.2% resultó seropositivo, también se encontró un índice de infección natural en las triatomas del 16%, siendo el 63.9% capturadas intradomiciliariamente<sup>2</sup>.

En el estado de Morelos, se llevó a cabo una encuesta seroepidemiológica en 600 individuos de los cuales el 28.6% resultó positivo<sup>3</sup>.

- 
- 1 Tay J., Salazar op. cit pag 325-332.
  - 2 Quintal R.E., Zavala op. cit. p.255-258
  - 3 Sanchez B.B. [1988]

En el estado de Tabasco en una muestra de 561 personas, el 13% resultó seropositivo.

En algunos estados sólo se tiene información sobre los transmisores involucrados en la enfermedad: Distrito Federal, Zacatecas y Michoacan.

En la actualidad, son pocos los países que tienen bien establecido un sistema de control, como el que se tiene en algunas zonas de Sudamérica. Argentina por ejemplo, posee un programa que abarca el 50% de la zona endémica, protegiendo al 25% de las personas expuestas al riesgo de enfermar. Brasil, cuyo programa incluye al 84.2% del área endémica, y protege así al 86.9% de las personas expuestas al riesgo de enfermar. Venezuela, cuyo programa de control cubre al 43.5% del área endémica y protege al 79.5% de la población expuesta. Uruguay cubre aproximadamente al 60.8% de la zona endémica y protege así al 48.2% de la población expuesta. Por último Chile cubre al 16.5% de su área endémica y protege al 1.5% de la población expuestas.

## JUSTIFICACION

En nuestro país se han llevado a cabo esfuerzos por desarrollar estudios sobre prevalencia de la enfermedad de Chagas (referentes a su sero-epidemiología, vectores, reservorios, manifestaciones clínicas), sin embargo, se observa que la información que se obtiene de ellos, no es suficiente para caracterizar un comportamiento de esta enfermedad en México (en lo que se refiere a distribución, prevalencia y dinámica de transmisión) que nos marque la pauta para desarrollar estrategias para su control y vigilancia.

Es importante mencionar que estos estudios no se han realizado bajo criterios uniformes, que garanticen la comparabilidad de los resultados en las diferentes zonas geográficas en las que se han llevado a cabo. El grupo de trabajo sobre Normas para Investigación Multidisciplinarias acerca de la epidemiología de la enfermedad de Chagas, se reunió en Brasilia en julio de 1979, para establecer las normas y criterios para la elaboración de estudios de prevalencia y longitudinales, con el objeto de

que los resultados de los estudios que se realicen sobre esta enfermedad, sean comparables, según las diferentes regiones y así los conocimientos adquiridos sirvan para establecer las normas necesarias a fin de elaborar mejores estrategias de lucha, en las zonas estudiadas.

En nuestro país no se conoce la prevalencia de la enfermedad, sin embargo los resultados que se han obtenido en los lugares de la república en donde se ha buscado la enfermedad por encuestas seroepidemiológicas han mostrado altos índices de infección.

Uno de los estados que cuenta con las condiciones socioeconómicas, climatológicas y ecológicas para permitir la transmisión activa de esta enfermedad es Morelos, resultando ser el lugar idóneo para realizar este estudio, bajo criterios sugeridos por el grupo de trabajo sobre Normas para Investigaciones Multidisciplinarias acerca de la epidemiología de la enfermedad de Chagas.

### PLANTEAMIENTO DEL PROBLEMA

Los factores de riesgo para tripanosomiasis americana están bien identificados, pero se desconoce la forma en que interactúan, lo que repercute en la frecuencia del padecimiento en las diversas comunidades. Por lo tanto es necesario, identificar y jerarquizar estos factores para conocer la prevalencia de la enfermedad en sus diferentes etapas en dichas comunidades. Lo anterior servirá para desarrollar estudios de casos y controles o de cohorte, para analizar la interacción de los factores en la comunidad.

---

5 Salazar, Bucio op. cit p.11-51

## OBJETIVOS

Este estudio tiene como objetivos :

- 1) Establecer la prevalencia de tripanosomiasis americana en la Jurisdicción dos del Estado de Morelos, con base en aspectos serológicos.
- 2) Clasificar las poblaciones, según la seropositividad encontrada.
- 3) Identificar los factores de riesgo existentes en cada categoría y evaluar su iteracción

## METODOLOGIA

El estudio es de tipo transversal, se realizó un tamizaje en comunidades del estado de Morelos, con el fin de determinar la prevalencia de la enfermedad de Chagas, se tomaron comunidades con de seropositividad para identificar los factores de riesgo y determinar los índices entomológicos en ellas.

El estado de Morelos se encuentra dividido en tres Jurisdicciones:

Cuernavaca  
Jojutla  
Cuautla

Se toma una muestra de 24 comunidades de la jurisdicción dos, elegidas por conveniencia, accesibilidad, así como por las características climatológicas y ecológicas que presenta dicha región, que facilitan las condiciones para que se lleve a cabo la transmisión activa de la enfermedad.

De cada comunidad, se obtuvo el total poblacional y se sumaron dichas cantidades para obtener un total, del cuál se calculó el tamaño de muestra mediante la fórmula:

$$n = \frac{Npq}{(N-1)D + pq}$$

n = Tamaño de muestra

N = Total (140.409)

p = prevalencia de tripanosomiasis americana estimada en otros estudios (15%)

$$q = (1-p)$$

D = Intervalo de confianza al 95%

$$D = \frac{B}{4}$$

B = Magnitud del limite de error: 2%

$$P(p_1 = p = p_2) = 0.95$$

$$P(.13 = .15 = .17) = 0.95$$

$$B = \frac{p_2 - p_1}{2} = .02$$

$$D = \frac{.02}{4} = 0.0001$$

$$n = 1264$$

Se aumenta un 20% a la muestra, para conservar el tamaño de la muestra aún después de la aplicación de los criterios de selección y la eliminación de cuestionarios ilegibles o incompletos, lo cuál dá un total de 1517 individuos.

Al total se dividirá entre cinco (que es el número promedio de personas que se espera encontrar en cada vivienda), para obtener así el número de viviendas que se muestrearán (312 en total).

Para distribuir este tamaño muestral en cada una de las comunidades, se aplicará el procedimiento de asignación proporcional de la muestra con la siguiente formula:

$$\frac{N_h}{N} \cdot (n)$$

$N_h$  = Total de la población en cada una de las comunidades

$N$  = Total de individuos (140,409)

$n$  = tamaño de muestra de viviendas (312)



De esta manera podremos estimar la prevalencia de la enfermedad de Chagas en la Jurisdicción dos (Jojutla) estudiadas, con un límite de error no superior al 2% y un intervalo de confianza del 95%.

COMUNIDADES	TOTAL POBLACIONAL Nh	MUESTRA DE VIVIENDAS Nh
ACAMILPA	1490	5
AHUEHUETZINGO	2040	5
EL HIGUERON	3440	7
HUAJINTLAN	2369	5
HUAUTLA	1659	5
JOJUTLA	28383	56
LA TIGRA	263	5
LORENZO VAZQUEZ	880	5
PEDRO AMARO	3545	7
PUENTE DE IXTLA	23737	47
QUILAMULA	634	5
SAN JOSE DE PALA	398	5
SAN JOSE VISTA H.	2509	6
SANTA ROSA TREINTA	4450	8
TEHUIXTLA	5274	11
TEQUESQUITENGO	3144	6
TICUMAN	3390	6
TILZAPOTLA	4729	9
TLALTIZAPAN	7204	14
TLALQUILTENANGO	21223	42
TLATENCHI	1500	5
XOCHIPALA	164	5
XOXOCOTLA	19220	38
<b>TOTALES</b>	<b>141909</b>	

Una vez conocido el número de viviendas con las que contribuirá cada comunidad, se elegirá al azar una vivienda en cada manzana. Se incluirán en el estudio a las personas que habiten en esa vivienda y cumplan los siguientes criterios de selección:

**CRITERIOS DE INCLUSION:**

- Edad mínima 3 años cumplidos
- Haber habitado en la comunidad por lo menos un año

## CRITERIOS DE EXCLUSION

- Personas de 51 cumplidos o mayores
- Personas que hayan sido transfundidas

A cada una de las personas que vivan en ese domicilio se le aplicó un cuestionario en el que se obtuvo información de las siguientes variables:

**EDAD:** Años de vida cumplidos.

**SEXO:** Género al que pertenece el individuo. Se califica al individuo según su sexo: Femenino ó Masculino, la escala de medición es cualitativa y los niveles son: masculino y femenino.

**OCUPACION:** Empleo u oficio en el que trabaja un individuo. Es medida según la rama de actividad, la escala de medición es cualitativa y los niveles son: agricultura, artesanía, hogar, estudiante, profesionista, técnico, comerciante, obrero, otro.

**ESCOLARIDAD:** Años de estudio de un individuo. Es medida por los ciclos escolares terminados, la escala de medición es cualitativa y sus niveles son: analfabeta, primaria incompleta, primaria completa, secundaria, técnico, preparatoria, profesionista.

**LUGAR DE NACIMIENTO:** Nombre de la comunidad y el estado de la República, en donde nació el individuo entrevistado. Es medida en escala cualitativa y su nivel será el nombre específico de dicha comunidad y estado.

**LUGAR DE RESIDENCIA ANTERIOR:** Comunidad en la que habitó la persona antes de hacerlo en la actual, se determinó el nombre específico de la comunidad en la que residió, así como el tiempo en durante en cual vivió en dicha comunidad. Es medida en la escala cualitativa, siendo su nivel el nombre específico de la comunidad y los años y meses en que radicó en ella.

Debido a que la enfermedad se puede adquirir en la comunidad actual o en la anterior, se hicieron preguntas que permitan relacionar los datos obtenidos con los resultados de registro del laboratorio, para establecer en donde fué adquirida la enfermedad.

**TIEMPO DE RESIDENCIA EN LA COMUNIDAD:** Tiempo durante el que un individuo ha vivido dentro de la comunidad. Es medida en años, es medida en escala cuantitativa discreta y su nivel es el número de años que la persona ha vivido en la comunidad.

**PRESENCIA DE ANIMALES DOMESTICOS:** Se refiere a la existencia de animales domésticos dentro de la vivienda. Su escala de medición es cualitativa y los niveles son: perros, gatos, ardillas, otros.

**MATERIAL DE CONSTRUCCION DE TECHOS:** Se refiere al material principal con que está construida la cubierta de la vivienda. Se indicará el material predominante en la construcción de dicha cubierta, y es medida en escala cualitativa y los niveles son: madera, carrizo-bambú, lámina de asbesto, lámina de aluminio, lámina de fibra de vidrio, palma o zacate, teja, cemento o concreto.

**MATERIAL DE CONSTRUCCION DEL MUROS:** Se refiere al principal material con que se construyen los muros o paredes de la vivienda. Se indica el material predominante en la construcción de los muros, es medida en la escala cualitativa nominal y sus niveles son: adobe, piedra, carrizo-bambú, ladrillo, barro, cemento, madera.

**MATERIAL DE CONSTRUCCION DE PISOS:** Se refiere al principal material que constituye el piso de la vivienda. Se indica el material predominante en la construcción del piso, es medida en la escala cualitativa, y sus niveles son: tierra, cemento, madera, mosaico, piedra.

**FIURAS:** Existencia de hendiduras en los muros, techos o piso de la vivienda. Se considera fisura, cuando se encuentren aberturas (regulares o irregulares), sobre los muros, techos o piso sin importar su longitud o profundidad, su escala de medición es cualitativa y sus niveles son: sí o no. Las triatomas tienen preferencia por sitios oscuros, silenciosos y tibios para su desarrollo, características que les brindan las fisuras, motivos por el cual es importante esta información.

**ILUMINACION:** Se refiere a la cantidad de luz natural que reciben las habitaciones durante el día. Es medida por la cantidad de luz natural que reciban las habitaciones durante el día, su escala de medición es cualitativa ordinal y sus niveles son: Buena (cuando la luz que penetra en las habitaciones, permita visualizar lo que hay en la habitación durante el día), Regular (cuando la luz natural que recibe la habitación durante el día sólo penetra cuando es muy intensa), Mala (cuando la habitación carece de penetraciones luminosas durante el día).

**VENTILACION:** Se refiere a la forma en que penetra y circula el aire dentro de la vivienda. Es medida por la calidad de la ventilación dentro de los cuartos de la vivienda, así como el sitio donde duermen los animales, su escala de medición es cualitativa ordinal, y sus niveles son: Buena (cuando la vivienda cuenta con una ventana lo suficientemente amplia y se encuentran abiertas por lo menos 3 hrs. al día), Regular (cuando la habitación cuenta con una ventana pequeña, o que esté

orientada hacia un lugar carente de corrientes de aire), Mala (cuando la habitación no cuenta con ventanas o éstas se encuentran clausuradas). Debido a que las triatomas tiene preferencia por sitios mal ventilados y poco iluminados esta información es valiosa.

Se tomó a cada persona una muestra de sangre obtenida por punción digital, para realización de la prueba serológica. Una vez que se tienen las muestras se procesarán en el Depto. de Ecología Humana de la Facultad de Medicina.

### ESTUDIO PILOTO

Previo a la realización de la fase de campo se llevó a cabo un estudio piloto, el cual tenía los siguientes objetivos:

-Probar el instrumento de recolección de información, es decir detectar fallas en la secuencia de las preguntas, así como la detección de preguntas mal formuladas incoherentes o agresivas.

-Valorar la confiabilidad de las variables: material de construcción de techos, pisos y muros, ventilación e iluminación. dado que estas variables serían clasificadas por observación directa de cada uno de los médicos pasantes en servicio social, quienes fueron entrevistadores, y se deseaba valorar la concordancia de los observadores teniendo como base la operacionalización de cada una de las variables y los adiestramientos que se les brindó.

Para alcanzar dichos objetivos se eligieron cuatro comunidades: Jojutla, Tlaltizapan, Quilamula, Ahuehuetzingo, de las veinticuatro totales.

A cada persona se le aplicó el cuestionario y se le tomó una muestra sanguínea para realización de la prueba serológica. Las variables que se estudiaron son las que se mencionaron con anterioridad.

Además de la proporción de seropositividad obtenida, se observó que el cuestionario no presentó ningún problema en su aplicación, en su secuencia o en algún otro punto.

COMUNIDADES	FAMILIAS	INDIVIDUOS	SEROPOS
JOJUTLA	5	27	02
TLALTIZAPAN	4	18	10
QUILAMULA	3	12	01
AHUEHUTZINGO	1	08	05
TOTAL	13	65	18

Los resultados obtenidos en esta prueba fuerón los siguientes:

COMUNIDADES	INDIVIDUOS	SERO +
JOJUTLA	27	02
TLALTIZAPAN	18	10
QUILAMULA	12	01
AHUEHUTZINGO	08	05
TOTAL	65	18

Se visitarón treinta viviendas de Jojutla, Morelos solicitando que calificaran cada una de las variables como positivo o negativo en forma individual.

Cada variable se dicotomizó en positivo y negativo, según se considerara como factor de riesgo o no. La clasificación se hizo a priori.

Los resultados encontrados, se analizarón con la prueba de Kappa:

$$K = \frac{Pa - Po}{1 - Po}$$

Pa = Concordancia esperada

Po = Concordancia observada

## PROTOCOLO DE INVESTIGACION

Se calculó el promedio del índice de kappa, y se encontró una concordancia de:

VARIABLE	PROMEDIO
TECHO	70%
MURO	100%
PISO	100%
VENTILACION	78%
ILUMINACION	82%

Con lo anterior podemos decir según parámetros bien establecidos (FLEISS), que la concordancia de las variables probadas, es adecuada y su medición es reproducible

Los resultados obtenidos en el estudio piloto, dieron bases para iniciar la fase de campo la cuál se llevó a cabo en 24 comunidades de la jurisdicción dos de Morelos.

## **CAPITULO III**

## TECNICAS ESTADISTICAS PARA ESTUDIOS EPIDEMIOLOGICOS

### TABLAS DE CONTINGENCIA

#### INTRODUCCION

En el análisis estadístico existen diferentes métodos que son aplicados al campo de la biología, ciencias sociales, medicina, entre otros, para el análisis de datos, con un interés específico.

Quando se presenta un estudio en donde el interés es hacer el análisis de una muestra clasificada con respecto a varios criterios cualitativos, podemos utilizar las Tablas de Contingencias.

Las Tablas de Contingencia nos permiten detectar relaciones existentes entre variables, y son apropiadas cuando estas son medidas en escala nominal u ordinal.

Las Tablas de Contingencia pueden ser de dos o más dimensiones, y el número de dimensiones corresponde al número de variables bajo estudio.

Se describirán algunos de los métodos estadísticos que permiten diseñar e interpretar las Tablas de Contingencia de dos o más criterios.

#### TABLAS DE DOS DIMENSIONES

Quando se presenta el caso de dos variables tenemos arreglos rectangulares con columnas correspondientes a las categorías de una variable y renglones para las categorías de la segunda variable y la posición de las celdas especifica las características de los individuos.

La construcción de una tabla de contingencia se hace con el fin de evidenciar las relaciones entre variables involucradas, es decir, la asociación que existe cuando se involucran dos, tres o más variables categóricas.



Las categorías de las variables que se encuentren en estudio, deberán ser exhaustivas (la clasificación que fué elegida agrupa en sus categorías a todos los miembros de la población), y mutuamente exclusivas ( un miembro de la población pertenece sólo a una celda)

Cada variable tendrá sus categorías correspondientes y en caso de tener sólo dos será denominada variable dicotómica.

Una tabla de dos dimensiones o dos entradas con variables binarias se representa de la siguiente forma:

	Variable 2		Totales
Variable 1	x <sub>11</sub>	x <sub>12</sub>	x <sub>1.</sub>
	x <sub>21</sub>	x <sub>22</sub>	x <sub>2.</sub>
Totales	x <sub>.1</sub>	x <sub>.2</sub>	x <sub>..</sub>

En esta tabla tenemos una muestra con N observaciones, clasificada con respecto a dos variables. Cada una con dos categorías, considerado este el caso más simple.

Si seleccionamos a un miembro de la población aleatoriamente podría caer en alguna de las categorías, por ejemplo la i-ésima, con probabilidad P<sub>i</sub> donde esta probabilidad es el vector de celdas de probabilidad:

$$\sum_i P_i = 1$$

Este arreglo no refleja ninguna característica acerca de los individuos que caigan dentro de una celda en particular. La P<sub>i</sub> refleja la frecuencia relativa de cada categoría

Si consideramos una muestra aleatoria y cada una de las probabilidades de las celdas suman uno.

$$\sum_i \sum_j P_{ij} = 1$$

Para examinar las probabilidades marginales tenemos que:

$$P_{i.} = \sum_k P_{ik}$$

$$P_{.j} = \sum_k P_{kj}$$

Ahora los totales tienen la probabilidad de que un individuo caiga en la categoría  $i$  de la variable  $1$  y la categoría  $j$  de la variable  $2$ , estas son probabilidades no condicionales de pertenecer a cualquiera de las dos variables.

Cada conjunto de probabilidades debe sumar uno, y a cada una de las intersecciones de categorías de variables le llamaremos celdas o casilla, y estarán constituidas por los elementos de la muestra que simultáneamente presenten el atributo de la variable renglón, y el atributo de la variable columna.

En cada casilla tenemos el total de elementos de la muestra clasificada de la categoría  $i$  de la variable  $1$  y de la categoría  $j$  de la variable  $2$ , ese total es llamado frecuencia observada y se denota como  $x_{ij}$ .

A la suma de las frecuencias en el renglón de la  $i$ -ésima categoría de la variable  $1$  se denota como  $x_{i.}$  y a la suma de las frecuencias en la columna de la  $j$ -ésima categoría de la variable  $2$  se denota como  $x_{.j}$  y a la suma de las frecuencias del total de la tabla como  $x_{..}$ .

Si generalizamos, considérese la variable  $1$  con  $t$  categorías y la variable  $2$  con  $s$  categorías por lo que se tiene una tabla:

Variable 1	Variable 2					Totales	
	$x_{11}$	$x_{12}$	..	$x_{1j}$	..	$x_{1s}$	$x_{1.}$
	$x_{21}$	$x_{22}$	..	$x_{2j}$	..	$x_{2s}$	$x_{2.}$
	.						
	$x_{i1}$	$x_{i2}$	..	$x_{ij}$	..	$x_{is}$	$x_{i.}$
	.						
	$x_{t1}$	$x_{t2}$	..	$x_{tj}$	..	$x_{ts}$	$x_{t.}$
	.						
Totales	$x_{.1}$	$x_{.2}$	..	$x_{.j}$	..	$x_{.s}$	$x_{..} = N$

Tenemos una tabla de contingencia de  $t \times s$ , donde los renglones tienen  $t$  categorías y las columnas tienen  $s$  categorías.

Ahora los totales tienen la probabilidad de que un individuo caiga en la categoría  $i$  de la variable  $1$  y la categoría  $j$  de la variable  $2$ , estas son probabilidades no condicionales de pertenecer a cualquiera de las dos variables. Cada conjunto de probabilidades debe sumar uno.

Los totales marginales de cada una de las categorías se representarán:

$$x_{i.} = \sum_{j=1}^J x_{ij} \quad 1.1$$

$$x_{.j} = \sum_{i=1}^I x_{ij} \quad 1.2$$

y la suma de las frecuencias de la tabla:

$$x_{..} = \sum_{i=1}^I \sum_{j=1}^J x_{ij} = N \quad 1.3$$

Para obtener los valores esperados de las observaciones ( $x_{ij}$ ) tenemos que:

$$m_{ij} = NP_{ij} \quad 1.4$$

Donde  $P_{ij}$  es la probabilidad de que una observación pertenezca a la celda ( $i,j$ ).

Si se supone independencia entre variables se cumple que 1:

$$P_{ij} = P_i \cdot P_j \text{ para todo } i,j \quad 1.5$$

Sea  $P_i$  la probabilidad de que un elemento de la población pertenezca a la  $i$ -ésima categoría de la variable renglón, y  $P_j$  la probabilidad de que un elemento de la población pertenezca a la  $j$ -ésima categoría de la variable columna.

---

1 Mood y Graybill pag. 51.

Y se tendr a bajo la hip tesis de independencia:

$$m_{ij} = N P_i \cdot P_j \quad 1.6$$

Debido a que se basa sobre probabilidades de la poblaci n estimamos con el estimador de m xima verosimilitud <sup>2</sup>, a partir de los totales marginales de las frecuencias observadas:

$$\hat{P}_i = x_i / N \quad 1.7$$

$$\hat{P}_j = x_j / N \quad 1.8$$

Una vez que se obtuvieron estos valores, podemos estimar la frecuencia esperada:

$$m_{ij} = N (\hat{P}_i) (\hat{P}_j) \quad 1.9$$

entonces obtenemos los valores estimados para dos variables bajo la hip tesis de independencia:

$$\hat{m}_{ij} = x_i x_j / N \quad 1.10b$$

Y se debe cumplir que:

$$x_{i.} = m_{i.}$$

$$x_{.j} = m_{.j}$$

Donde:

$$\sum_{j=1}^J m_{ij} = m_{i.} = (x_{i.}) (x_{.j.}) = x_{i.} \quad 1.11.2$$

$$\sum_{i=1}^I m_{ij} = m_{.j} = (x_{i.}) (x_{.j.}) = x_{.j} \quad 1.11.1$$

Esto nos dice que los totales esperados coinciden con los totales observados.

<sup>2</sup> Mood y Graybill pag 206.

EJEMPLO:

Tomamos de la encuesta realizada en el Estado de Morelos las variables de Serología, denotada por (Serolo), y Municipio, denotada por (MUNICI), y consideramos a P1 como la probabilidad de que un miembro en el grupo sea Seropositivo y clasificado en los Municipios de AMA, TIZ, TEN JOJ Y TXT

Esta clasificación es tomada en base a un estudio para determinar cuál es la relación que existe entre las variables.

MUNICI	SEROLO		TOTAL
	POS	NEG	
AMA	9	43	52
TIZ	17	157	174
TEN	31	158	189
JOJ	57	209	266
IXT	55	409	464
TOTAL	169	976	1145

Obtenemos los valores observados para el municipio de JOJ es:

$$x_{14} = 57 \text{ y } x_{24} = 209$$

Para Obtener los valores esperados para el municipio de JOJ tenemos:

$$m_{14} = (266)(169)/1145$$

Una vez que se obtienen los valores estimados podemos recurrir a una prueba estadística que nos permita probar la hipótesis de independencia.

Uno de nuestros principales propósitos es determinar si las variables son independientes o no y para ello recurriremos a una prueba estadística:  $\chi^2$  de Pearson.

Está prueba básicamente compara las frecuencias observadas con las frecuencias que se esperan, bajo la hipótesis de interés.

$$\chi^2 = \sum_i \sum_j (x_{ij} - m_{ij})^2 / m_{ij} \quad 1.12$$

## TECNICAS ESTADISTICAS PARA ESTUDIOS EPIDEMIOLOGICOS

Si las frecuencias fueran muy parecidas o iguales se dice que las variables son independientes. Para determinar las frecuencias son muy parecidas o iguales se toma el limite que establece el valor que proporciona la distribución teórica de la  $\chi^2$ , y esto se debe a que las diferencias encontradas pueden deberse a factores aleatorios al momento de obtener la muestra.

Para tomar la decisión de rechazar o no la hipótesis de independencia depende de la probabilidad de obtener por azar un valor dado por  $\chi^2$ , y mientras más estrictos más probable será no rechazar la hipótesis, ya que se pide para considerar independencia los valores observados y esperados sean muy próximos.

Es importante mencionar que la estadística de prueba que resulta de comparar los valores esperados contra los observados, se distribuye aproximadamente como una normal estandarizada con media cero y varianza uno.

Se fija un nivel de significancia (Probabilidad de rechazar la hipótesis nula cuando en realidad es cierta) lo cual determina una región crítica en función del parámetro "grados de libertad".

Al parámetro que incluye el número de casillas en una tabla se le conoce como grados de libertad, este nombre se deriva del número de casillas que se puede modificar en el valor de su frecuencia sin alterar los totales marginales.

La forma de calcular los grados de libertad en una tabla de dos dimensiones es:  $(r-1)(c-1)$  donde r son los renglones y c las columnas.

### TABLAS DE TRES DIMENSIONES

Sólo hemos tratado el análisis para tablas de dos dimensiones, sin embargo es muy común que se presente el caso en donde intervienen tres o más variables.

Podemos observar que conforme el número de criterios que describe a cada individuo aumenta, la interpretación del análisis resulta más complicada.

Como ejemplo tenemos una tabla de tres criterios donde involucramos a la población encuestada en nuestro estudio, donde se tiene las variables Serología (Sero), Municipio (Munici) y Material de

Muro (Mamu). Lo que interesa es determinar cuales son las relaciones que existen entre estas variables.

MUNICI	MAMU	SEROLOGIA		TOTAL
		POS	NEG	
AMA	EXP	8	7	8
	NEXP	1	36	44
	TOTAL	9	43	52
TIZ	EXP	4	43	47
	NEXP	6	106	112
	TOTAL	10	149	159
TEN	EXP	21	100	121
	NEXP	10	57	67
	TOTAL	31	157	188
JOJ	EXP	7	30	37
	NEXP	50	177,227	
	TOTAL	57	207	264
IXT	EXP	16	87	103
	NEXP	35	274	309
	TOTAL	51,361	412	

Podremos revisar que ocurre con Serología y Material de muro, Serología y Municipio, Material de muro y Municipio y tratar de encontrar asociación entre ellas.

Cuando se presenta el caso de una tabla de tres dimensiones se puede trabajar de dos formas:

- a) Tratando la tabla como si fuera de dos dimensiones y analizar la relación entre variables dos a dos, es decir hacer las pruebas de independencia para encontrar asociación entre cada par de variables.
- b) Otra manera es cuando las tres variables se analizan simultáneamente, de aquí que se desprenden nuevos planteamientos, y no sólo probar independencia entre variables, sino también:
  - 1) Probar si Ocupación y Municipio son independientes de la Serología.

2) Probar si existe asociación entre la Serología y alguna de las otras dos variables

3) Probar que las tres variables sean mutuamente independientes.

Las Hipótesis que podemos probar en el caso de presentarse tablas tridimensionales son las siguientes :

**Hipótesis de Independencia Mutua:**

Una hipótesis de independencia mutua es equivalente a plantear la prueba de independencia en una tabla de dos dimensiones; siguiendo la notación utilizada para estas tablas sólo agregaremos un subíndice más que nos indique que aumentamos una nueva variable.

En términos de probabilidades, tenemos que  $P_{ijk}$  es la probabilidad estimada de que una observación tenga la categoría  $i$  de la variable  $1$  renglón, la categoría  $j$  de la variable  $2$  columna y la categoría  $k$  de la variable  $3$  control.

Se pretende probar independencia:

$$P_{ijk} = P_{i.} P_{.j} P_{..k} \quad 2.1$$

donde  $P_{i.}$ ,  $P_{.j}$  y  $P_{..k}$  son las probabilidades marginales de la variable renglón, de la variable columna, y de la variable control respectivamente.

Para encontrar los estimadores de los valores esperados tenemos:

$$m_{ijk} = N \hat{P}_{i.} \hat{P}_{.j} \hat{P}_{..k} \quad 2.2$$

Podemos calcular los estimadores a partir de los totales marginales de cada variable:

$$\hat{P}_{i.} = x_{i.} / N \quad 2.3$$

$$\hat{P}_{.j} = x_{.j} / N \quad 2.4$$

$$\hat{P}_{..k} = x_{..k} / N \quad 2.5$$



y los totales marginales son:

$$x_{i..} = \sum_{j=1}^J \sum_{k=1}^K x_{ijk} \quad 2.6$$

$$x_{.j.} = \sum_{i=1}^I \sum_{k=1}^K x_{ijk} \quad 2.7$$

$$x_{.k.} = \sum_{i=1}^I x_{i.k.} \quad 2.8$$

Ahora si sustituimos en 2.2 los valores esperados calculados por máxima verosimilitud 3 son:

$$m_{ijk} = (N)(x_{i..}/N)(x_{.j.}/N)(x_{.k.}/N) \quad 2.10$$

$$\hat{m}_{ijk} = x_{i..}x_{.j.}x_{.k.}/N^2 \quad 2.11$$

En general el procedimiento consiste en calcular los valores estimados de las frecuencias esperadas en caso de que la hipótesis de interés sea cierta, después comparar estos valores con las frecuencias observadas usando  $X^2$ .

Con estos valores calculamos la estadística  $X^2$  y los grados de libertad, para un caso de independencia mutua se calculan 4:

$$g.l. = r-1 - c-1 + 2$$

#### Hipótesis de Independencia parcial:

En caso que la hipótesis de independencia sea rechazada, una de las razones por las que podría haber sido es que dos de las variables esten asociadas y la tercera sea independiente de ambas.

3 Mood y Graybill, pag.213.

4 Everitt, op. cit.

Tenemos tres posibles hipótesis que pueden ser planteadas de la siguiente manera:

$$H_0(1) = P_{ijk} = P_{.k}P_{ij} \quad 2.17$$

Las variables 1 y 2 están asociadas y la variable 3 es independiente de ambas.

$$H_0(2) = P_{ijk} = P_{i.}P_{.jk} \quad 2.18$$

Las variables 2 y 3 están asociadas y la variable 1 es independiente de ambas

$$H_0(3) = P_{ijk} = P_{.j}P_{i.k} \quad 2.19$$

Las variables 1 y 3 están asociadas y la variable 2 es independiente de ambas.

Si la hipótesis  $H_0(1)$  es cierta se estiman los valores esperados y con  $m_{ijk}$  se calcula la  $X^2$  con grados de libertad:  $(r-1)(c-1) + 1$ .

#### Hipótesis de Independencia Condicional:

En caso de rechazar la hipótesis de independencia mutua tenemos otra alternativa: que exista independencia entre dos variables dentro de cada nivel de la tercera variable, pero que cada una de ellas este asociada con la tercera.

Las posibles hipótesis nulas son:

$$H_0(1) = P_{ijk} = P_{i.k}P_{.jk}/P_{.k} \quad 2.20$$

Las variables 1 y 2 son independientes en cada nivel k de la tercera variable.

$$H_0(2) = P_{ijk} = P_{ij.}P_{.jk}/P_{.j} \quad 2.21$$

Las variables 1 y 3 son independientes en cada nivel j de la segunda variable.

$$H_0(3) = P_{ijk} = P_{ij.}P_{i.k}/P_{i.} \quad 2.22$$

Las variables 2 y 3 son independientes en cada nivel i de la primera variable.

Si estimamos la frecuencia por su estimador de máxima verosimilitud para el caso de  $H_0(1)$  se tendría:

$$m_{ijk} = (x_{i.k})(x_{.jk})/x_{.k} \quad 2.23$$

$$g.l. = (r-1)-(r-1)-(c-1)-(l-1)$$

Las restricciones sobre los totales marginales para la hipótesis  $H_0(1)$  son:

$$x_{j.} = m_{i.x} \cdot j = m_{j.x} \cdot k = m_{.k}$$

$$x_{i.k} = m_{i.k} \cdot x_{.jk} = m_{.jk}$$

Si estimamos la frecuencia por su estimador de máxima verosimilitud para el caso de  $H_0(2)$  se tendría:

$$m_{ijk} = (x_{ij.})(x_{.jk})/x_{.j} \quad 2.24$$

$$g.l. = (r-1)-(r-1)-(c-1)-(l-1)$$

Las restricciones sobre los totales marginales para la hipótesis  $H_0(2)$  son:

$$x_{j.} = m_{i.x} \cdot j = m_{j.x} \cdot k = m_{.k}$$

$$x_{ij.} = m_{ij.x} \cdot jk = m_{.jk}$$

Si estimamos la frecuencia por su estimador de máxima verosimilitud para el caso de  $H_0(3)$  se tendría:

$$m_{ijk} = (x_{ij.})(x_{i.k})/x_{i.} \quad 2.25$$

$$g.l. = (r-1)-(r-1)-(c-1)-(l-1)$$

Las restricciones sobre los totales marginales para la hipótesis  $H_0(3)$  son:

$$x_{.j} = m_{i..}x_{.j} = m_{.j}x_{.k} = m_{.k}$$

$$x_{ij} = m_{ij}x_{i.k} = m_{i.k}$$

### TABLAS DE CUATRO O MAS CRITERIOS:

Cuando se desea hacer una generalización para  $n$  variables, la cantidad de hipótesis que se pueden considerar aumenta considerablemente al tener un número mayor de variables.

Como en el caso de tres dimensiones donde se podían analizar varias tablas de dos criterios, hasta los 70s los investigadores manejaban las tablas de dos dimensiones utilizando los totales marginales de las tabla multidimensionales, pero este tratamiento en ocasiones da una visión no muy buena de las relaciones entre variables, y presenta los siguientes problemas::

- 1) Se confunde la relación marginal entre un par de variables, con la relación entre ellas cuando otras están presentes.
- 2) No permite el análisis simultáneo de la relación entre varios pares de variables.
- 3) Ignora la posibilidad de asociación entre tres o más variables.

En la actualidad para el análisis de tablas de contingencia multidimensionales se utilizan los modelos loglineales, los cuales son usados para describir las relaciones estructurales entre las variables categóricas.

## MODELOS LOGLINEALES

### INTRODUCCION

Un modelo en términos generales es utilizado para acumular y describir los diferentes aspectos de la realidad que ya se conocen, y sirven como instrumento para explicar el pasado y el presente y prevenir el futuro, es decir los modelos nos sirven para describir y explicar la realidad.

Dentro de los modelos encontramos los matemáticos en los cuales las propiedades de las cosas son expresadas simbólicamente y a su vez los símbolos que se emplean representan relaciones entre los objetos.

En este trabajo utilizamos un tipo muy específico de modelos matemáticos los cuales llamamos modelos loglineales y al utilizarlos trataremos de explicar o ajustar los datos en base a las frecuencias esperadas bajo un modelo que se espera se aproxime a las frecuencias observadas.

Se expondrán modelos para tablas de dos dimensiones, y posteriormente extendernos a modelos donde intervengan tres o más dimensiones

#### Modelos Loglineales:

Para el análisis de tablas multidimensionales los modelos loglineales son usados con frecuencia, estos son modelos lineales en los logaritmos de las frecuencias esperadas en las celdas. Este tipo de modelos tienen una cierta analogía con los modelos de análisis de varianza.

En general un modelo Loglineal no distingue entre las variables dependientes e independientes, es decir todas las variables son tratadas como respuesta, tomando la asociación entre variables. El criterio que se analiza son las frecuencias esperadas como una función de todas las variables en un modelo.

Los modelos Loglineales están definidos para miembros de la familia exponencial, y cabe mencionar que generalmente los datos de las frecuencias de las tablas de contingencia de cualquier dimensión proviene de alguno de los siguientes modelos de muestreo:

**Poisson:** Se observa un conjunto de procesos Poisson una para cada celda sobre un período de tiempo sin un conocimiento a priori respecto al número de observaciones, es decir, no se fija el tamaño de muestra.

**Multinomial:** Se toma una muestra de tamaño  $N$  y se clasifica cada miembro de la muestra de acuerdo a sus valores según las variables que se tienen.

**Producto-Multinomial:** Generalmente en los estudios de las observaciones solo una muestra se examina, pero en ocasiones es usual tener varios grupos. Cuando miembros de diferentes grupos no están apareados, las frecuencias de las celdas representan individuos y los totales de grupos forman los totales marginales

### Modelos para tablas de dos dimensiones

La construcción de un modelo se basa en que si dos variables son independientes se debe cumplir que:

$$P_{ij} = P_i P_j \text{ para todos los valores de } i, j$$

Esto es la probabilidad de que una observación caiga en la  $ij$ -ésima casilla, es decir tenemos el producto de las probabilidades marginales.

Para expresar esta probabilidad como la suma de marginales y como función lineal, tomamos el logaritmo:

$$\ln P_{ij} = \ln P_i + \ln P_j \quad 3.1$$

Si esto lo expresamos en términos de frecuencias tenemos que:

$$\ln m_{ij} = \ln m_i + \ln m_j - \ln N \quad 3.2$$

Si sumamos sobre  $i$  y sobre  $j$  tenemos que se representa a cada parámetro con "u", entonces tenemos que:

$$\ln m_{ij} = u + u_1(i) + u_2(j) \text{ para } i = 1..I, j = 1..J \quad 3.4$$

donde:

$$u = \sqrt{1/I} \sum_{i=1}^I \sum_{j=1}^J \ln m_{ij} \quad 3.5.1$$

$$u_1(i) = \sqrt{1/J} \sum_{j=1}^J \ln m_{ij} - u \quad 3.5.2$$

$$u_2(j) = \sqrt{1/I} \sum_{i=1}^I \ln m_{ij} - u \quad 3.5.3$$

$u$  es la media de los logaritmos de los valores esperados, y representa el efecto medio global

$u_1(i)$  representa el efecto principal del renglón de la  $i$ -ésima categoría menos el efecto medio global.

$u_2(j)$  representa el efecto principal de la columna de la  $j$ -ésima categoría menos el efecto medio global.

Esto tiene una similitud con la notación de un análisis de varianza, sólo que aquí los términos describen las relaciones estructurales entre las variables categóricas.

En este modelo intervienen solo el efecto medio global y el efecto principal de cada variable.

Ahora tenemos que estimar los parámetros del modelo para someter a prueba si este se ajusta a los datos.

El modelo 3.4 está considerado de independencia completa para tablas bidimensionales.

En caso de no existir independencia estadística se incluirá un nuevo término que nos de el efecto de asociación entre las variables.

Si se presenta este caso el modelo sería:

$$\log m_{ij} = u + u_1(i) + u_2(j) + u_{12}(ij) \quad 3.6$$

donde  $u_{12}(ij)$  es el efecto de interacción de las categorías  $(ij)$  de las variables  $(1,2)$ .

Ahora se plantea la hipótesis de independencia, y se espera que el término  $u_{12}(ij) = 0$  para toda  $i, j$ , y con esto el modelo 3.4 será el que mejor se ajusta a los datos.

### Modelos para Tablas de tres dimensiones

Los modelos para tablas de contingencia son muy similares a un análisis de varianza por ello se han tomado términos como asociación e interacción entre las variables de tipo cuantitativo que forman una tabla.

Para el caso de una tabla de tres dimensiones tenemos  $x_{ijk}$  como la frecuencia observada en las categorías  $i$  de la variable 1, en las categorías  $j$  de la variable 2, y en las categorías  $k$  de la variable 3, y el valor esperado bajo algún modelo es  $m_{ijk}$

El estimador del valor esperado sería:

$$\hat{m}_{ijk} = (x_{i..}/n)(x_{.j.}/n)(x_{...k}/n)n$$

Ahora si tomamos el logaritmo de la estimación de los valores esperados, y tomamos la notación de un análisis de varianza tenemos:

$$\log m_{ijk} = u + u_1(i) + u_2(j) + u_3(k) \quad 3.8$$

donde:

$$u = \frac{1}{IJK} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K 3.8.1$$

$$u_1(i) = \frac{1}{JK} \sum_{j=1}^J \sum_{k=1}^K \ln m_{ijk} - u \quad 3.8.2$$

$$u_2(j) = \frac{1}{IK} \sum_{i=1}^I \sum_{k=1}^K \ln m_{ijk} - u \quad 3.8.3$$

$$u_3(k) = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \ln m_{ijk} - u \quad 3.8.4$$

El modelo antes mencionado se supone cuando existe completa independencia entre las variables.



**Modelo Saturado**

Un modelo saturado es aquel que contiene un parámetro para cada celda, es decir el número de parámetros es igual al número de celdas.

El modelo saturado para una tabla de tres dimensiones está dado por:

$$\log m_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) + u_{23}(jk) + u_{123}(ijk)$$

En este modelo aparece el efecto principal de cada una de las variables y los parámetros que miden el efecto de interacción de primer orden y además el efecto de interacción de segundo orden entre las tres variables.

**Efecto de interacción primer orden:**

Representa la desviación del promedio de la media debido a la última variable que se agrega. Todas las posibles combinaciones de dos variables nos dan los términos de segundo orden.

**Efecto de interacción segundo orden:**

Los términos de segundo orden relacionan las variables uno y dos, es el promedio de los efectos de segundo orden en cada tabla:

**Efecto de interacción tercer orden:**

La definición de términos de tercer orden varía de acuerdo al número de dimensiones

Cuando se desean plantear hipótesis a través de los modelos loglineales se deben eliminar del modelo saturado aquellos términos que se consideran que no deben aparecer, por ejemplo la hipótesis a plantear indica que no existe asociación entre las tres variables simultáneamente, entonces se debe cumplir que:  $u_{123} = 0$  y el modelo quedaría:

$$\log m_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) + u_{23}(jk)$$

En caso de no existir independencia entre las tres variables se pueden plantear las siguientes hipótesis para los modelos loglineales tridimensionales:

**Modelo de Independencia Mutua:**

Este modelo especifica que no existe asociación entre ninguna de las variables.

La hipótesis que se cumple es:

$$u_{12} = u_{13} = u_{23} = u_{123} = 0$$

Y el modelo que resulta es:

$$\ln m_{ijk} = u + u_1(i) + u_2(j) + u_3(k) \quad 3.10$$

Este modelo sólo incluye el efecto medio global y el efecto principal de cada una de las variables, es decir solo refleja las diferencias entre los totales marginales de de cada variable

**Modelo de Independencia Condicional:**

El tipo de interacción que se puede presentar entre las tres variables es que la interacción entre las variables uno y dos sea nula en cada nivel de la tercera variable.

Si se cumple la hipótesis en que  $u_{12} = 0$ , y dado que trabajamos con modelos jerárquicos el termino  $u_{123} = 0$ , entonces el modelo quedaría:

$$\log m_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{13}(ik) + u_{23}(jk) \quad 3.13$$

donde  $u_{12} = u_{123} = 0$

Este modelo se puede expresar como la combinación lineal del efecto medio global y el efecto principal de cada una de las variables y los parámetros  $u_{13}$  y  $u_{23}$  de primer orden

Los modelos que podríamos tener cuando existe independencia condicional entre dos variables dada otra son:

$$\log m_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ik)$$

$$\log m_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{23}(jk)$$

$$\log m_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{13}(ik)$$

**Modelo de Independencia parcial:**

Uno de los modelos que nos interesaría proponer sería aquel que incluye solamente un efecto de interacción de primer orden entre dos variables específicas.

Una de las hipótesis que se podría plantear sería:

$$P_{ijk} = P_i \cdot P_{ijk}$$

Está hipótesis implícita nos dice que entre el conjunto de interacciones solo están presentes las variables dos y tres.

El modelo sería:

$$\log m_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{23}(jk) \quad 3.13$$

Este modelo es el que incluye sólo un efecto de primer orden, es decir el menor número posible de interacciones.

Si el modelo ajusta bien a los datos, se puede concluir que las interacciones que fuerón excluidas son redundantes, por lo tanto se tiene que reducir el tamaño de la tabla de acuerdo a las interacciones que se excluyan.

#### Modelos Jerárquicos:

Estos modelos son aquellos en los que siempre que se incluye un término todos los términos de orden menor también deben estar incluidos. Ejemplo:

En el caso de incluir el término  $u_{123}(ijk)$ , en un modelo donde solo tenemos deben estar incluidos los términos  $u_{12}(ij)$ ,  $u_{13}(ik)$  y  $u_{23}(jk)$

y el modelo sería:

$$\log m_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{23}(jk) + u_{13}(ik) + u_{123}(ijk) \quad 3.14$$

La familia de los modelos jerárquicos está definida como una familia en la cual si alguno de  $u$  términos es cero, todos los términos de orden mayor también deben de ser igual a cero.

Ya que si se consideran modelos no jerárquicos nos enfrentamos al problema de no hacer una interpretación adecuada.

En tablas largas con estructura no jerárquica una posible estrategia es particionar y ver la sección de los datos. y si  $u_{123}$  no es igual a cero pero  $u_{12} = 0$ , se puede particionar de acuerdo a la categoría de la variable 3, entonces se tienen algunas tablas con interacciones entre la variable 1 y 2

**Modelos para cuatro o mas dimensiones:**

Cuando tenemos el caso de cuatro dimensiones, tenemos 113 modelos jerárquicos, donde se incluye a todos los términos  $u$  de los efectos principales. El número de modelos crece rápidamente conforme aumenta el número de dimensiones.

El modelo más simple es el de completa independencia para cuatro variables:

$$\log m_{ijkl} = u + u_1 + u_2 + u_3 + u_4 \quad 3.16$$

En ocasiones es difícil encontrar los valores esperados en una tabla de cuatro dimensiones para ello se sugiere utilizar los métodos: Interactivo de Ajuste Proporcional o el algoritmo de Newton-Raphson.

Algunos autores como Bishop[1971], Goodman[1971b] Bishop, Fienberg y Holland [1975] tienen reglas para decidir si se puede o no encontrar directamente los valores estimados de un modelo en particular.

Cuando se presenta el caso de modelos de independencia condicional o completa independencia es posible encontrar los estimadores directamente, a continuación se darán algunos ejemplos donde se encuentran directamente 6:

Modelo:

$$\log m_{ijkl} = u + u_1 + u_2 + u_3 + u_4 \quad 3.16$$

---

6 Bishop, op. cit. pag.38

valores esperados:

$$m_{ijkl} = x_{i...} x_{.j..} x_{...k.} x_{...l}/N^3 \quad 3.16.1$$

Modelo:

$$\log m_{ijkl} = u + u_1 + u_2 + u_3 + u_4 + u_{12} + u_{34} \quad 3.17$$

valores esperados:

$$m_{ijkl} = x_{ijk.} x_{...l}/N^3 \quad 3.17.1$$

Cuando se tiene un gran número de posibles modelos loglineales jerárquicos que pueden ser ajustados en una tabla multidimensional, se pueden utilizar varias técnicas para hacer una buena selección las cuales se describen más adelante.

### Estadística de Bondad de Ajuste

Una vez que se obtuvieron los valores esperados se procede a checar cual es el modelo que mejor se ajusta a los datos y para ello se propone utilizar las estadísticas:

$$\chi^2 = \sum (\text{observados} - \text{esperados})^2 / \text{esperados} \dots 1$$

$$G^2 = 2 \sum (\text{observados}) \log(\text{observados} / \text{esperados}) \dots 2$$

La estadística  $\chi^2$  es la Ji cuadrada de Pearson. En particular para una tabla de  $2 \times 2$  se tiene que  $\chi^2 = z^2$ , donde  $z$  es la devianza usada para comparar dos proporciones binomiales. La  $\chi^2$  esta distribuida asintóticamente como una  $\chi^2$  bajo la hipótesis nula.

La  $G^2$  es la estadística de bondad de ajuste (devianza), se distribuye asintóticamente como una  $\chi^2$  con grados de libertad apropiados para los valores estimados.

La estadística  $G^2$  es menos familiar que la  $\chi^2$ . Una de las desventajas que tiene la  $G^2$  es que para calcularla se requiere de las tablas logarítmicas, sin embargo posee propiedades importantes que la  $\chi^2$  no tiene: Es una estadística que minimiza los valores esperados y puede partirse convenientemente en dos formas condicional y estructuralmente.

Dichas estadísticas son utilizadas para probar que el modelo que se tiene es el que mejor se ajusta a los datos.

Ambas estadísticas son asintóticamente equivalentes, esto es ellas son equivalentes en muestras muy grandes si la hipótesis nula es verdadera.

#### Selección de un modelo:

Para elegir el modelo más adecuado para un conjunto de datos, se requiere de métodos establecidos basados en criterios como la de bondad de ajuste que permitan comparar varias posibilidades de ajuste.

Se tiene una técnica general que sirve para comparar los valores esperados de dos diferentes modelos loglineales. Se tiene el modelo<sub>1</sub> y el modelo<sub>2</sub> sus estimaciones serían:

$$m_{ijk}(1) = (\text{esperados}_1)$$

$$m_{ijk}(2) = (\text{esperados}_2)$$

Donde el modelo<sub>2</sub> es un caso especial del modelo<sub>1</sub> (el modelo de completa independencia en una tabla de tres dimensiones es un caso especial del modelo de independencia de la variable tres con las variables uno y dos juntas) por lo tanto :

$$G^2 = 2 \sum (\text{observados}) \log \left( \frac{\text{observados}}{\text{esperados}_1} \right) \quad 4.1$$

Esta prueba puede ser usada si la diferencia entre los valores esperados de los dos modelos simplemente se debe a una variación aleatoria, dado que los valores esperados satisfacen al modelo<sub>1</sub>

Esta prueba estadística condicional tiene una distribución asintótica de un  $X^2$  (bajo la hipótesis nula) con grados de libertad iguales a la diferencia de los grados de libertad de los dos modelos  $\gamma$

Este método se puede utilizar para escoger el mejor modelo, el cuál consiste en particionar la estadística de bondad de ajuste para Modelos Jerárquicos en varias partes aditivas. Cuando se hace uso de las particiones los modelos también deben ser jerárquicos anidados es decir cada modelo debe estar incluido en el siguiente.

Para ejemplificar en una tabla de tres dimensiones tenemos los modelos:

$$1) u + u_1 + u_2 + u_3$$

$$2) u + u_1 + u_2 + u_3 + u_{13}$$

$$3) u + u_1 + u_2 + u_3 + u_{13} + u_{23}$$

$$4) u + u_1 + u_2 + u_3 + u_{12} + u_{13} + u_{23}$$

y la estadística de bondad de ajuste satisface:

$$G^2_{(1)} = G^2_{(2)} = G^2_{(3)} = G^2_{(4)} =$$

Este resultado es importante ya que es una de las razones por la que se utiliza está estadística y no la  $X^2$ , debido a que este resultado no es cierto para cualquier conjunto anidado de modelos.

Además tenemos que  $G^2_{(1)} - G^2_{(2)}$ ,  $G^2_{(2)} - G^2_{(3)}$ , y  $G^2_{(3)} - G^2_{(4)}$  son estadísticas de la forma 4.1 y pueden ser usadas par probar diferencias entre los modelos 1,2,3 y 4.

En general el método de selección consiste en tomar el modelo mas simple y comparar con el modelo que le sigue y la diferencia de la estadística  $G^2$  con grados de libertad igual a la diferencia de los grados de libertad de los dos modelos. En caso de que no exista significancia en la prueba el modelo menos simple se desecha y se procede a repetir el procedimiento. Si la prueba es significativa se toma el modelo menos simple que corresponda a la diferencia y ese se toma como el mejor modelo

## Procedimiento de selección Stepwise

Describiremos el procedimiento de selección Stepwise para modelos sugeridos por Goodman (1971). Se describirá para cuatro variables, pero puede ser extendida para cinco o más variables.

Primeramente se toma un nivel de significancia, por ejemplo de 0.05 y se proponen modelos y se hace la prueba de bondad de ajuste ( $G^2$ ):

$$1) u_{12} = u_{13} = u_{14} = u_{23} = u_{24} = u_{34} = 0$$

$$2) u_{123} = u_{124} = u_{134} = u_{234} = 0$$

$$3) u_{1234} = 0$$

Podemos notar que el modelo<sub>1</sub> implica el modelo<sub>2</sub>, y el modelo<sub>2</sub> implica el modelo<sub>3</sub>. Si el modelo<sub>3</sub> no se ajusta cambiamos el modelo que esta presente por el modelo que tiene todos los términos  $u$  presentes. Esto es posible cuando el modelo 1 o 2 se ajusta mientras que el 3 no (estas situaciones ocurren raramente).

Pero si se presenta el caso en que el modelo<sub>3</sub> se ajusta y el modelo<sub>2</sub> no se ajusta entonces se hace lo siguiente:

a) Selección hacia adelante.

Del modelo 2 agregar una interacción de tercer orden .

b) Eliminación hacia atrás:

Del modelo 3 se borran las interacciones de tercer orden.

Suponemos que el modelo 2 y 3 se ajusta y el 1 no se ajusta.

c) Selección hacia adelante.

Tomar el modelo 1 y agregar una interacción de segundo orden.

d) Eliminación hacia atrás

Se toma el modelo 2 y se borran las iteracciones de segundo orden



Suponemos que el modelo 2 y 3 se ajusta y modelo 1 no se ajusta, los pasos involucrados hacia adelante son:

C1) Agregar aquel término del segundo orden que tenga mayor significancia, utilizando la estadística de bondad de ajuste, la cual tiene un nivel de significancia que no excede al 0.05.

C2) Agregar el siguiente término de mayor significancia de los términos de interacción de segundo orden, utilizando la prueba que se menciona en el paso anterior.

C3)(opcional) Borrar alguno de los términos de las iteración de segundo orden que no de una contribución significativa al modelo (usando la prueba estadística de bondad de ajuste).

C4) Repetir los pasos C2 y C3 hasta que los términos de la interacción de segundo orden, puedan ser agregados o dejados.

Los pasos para la eliminación hacia atrás son muy similares pero van en dirección contraria.

D1) Eliminar el término de menor significancia de los términos de segundo orden, que tiene un valor de significancia que no excede de 0.05.

D2) Eliminar el siguiente de los términos de menor significancia, usando en la prueba estadística del mejor ajuste.

D3) Agregar atrás los términos de la interacción de segundo orden que ahora son significantes para el ajuste del modelo.

D4) Repetir los pasos D2 y D3 hasta que los términos del segundo orden sean agregados o borrados.

Para ejemplificar el procedimiento de selección usando datos de Ries-Smith y tomando un nivel de significancia de 0.05 tenemos:

Un modelo que no se ajusta es:

1) 1,2,3,4

los modelos que se ajustan son:

2) 12,13,14,23,24,34

3) 123,124,134,234

\*tabla de estadística  $G^2$  y g.l. Bishop.

Se desea encontrar un modelo entre 1 y 2; empezamos con 1 aumentamos  $u_{24}$ , y la diferencia entre la  $G^2$  de 1 y de 1'

A)1,3,24

es  $42.9-22.4 = 20.5$  con un g.l., esto es muy significativo se puede agregar el término. Ahora agregamos el término  $u_{34}$ , y la diferencia entre la  $G^2$  de A y B es:

B)1,24,34

es  $22.4-18.0 = 4.4$  con un g.l.. En este punto no podemos borrar algún término. Ahora aumentamos el término  $u_{13}$ , y la diferencia entre la  $G^2$  de B y C es:

C)13,24,34

es  $18-11.9 = 6.1$  con 2 g.l.. Ahora no podemos en ninguno de los dos borrar o agregar términos de segundo orden, entonces el procedimiento hacia adelante nos da C como el mejor modelo.

En la eliminación hacia atrás empezamos con el modelo 2 y borramos el término  $u_{14}$ , la diferencia entre la  $G^2$  de 2 y D es:

D)12,13,23,24,34

es  $10.1-9.9 = .2$  con 2 g.l.. Ahora quitamos  $u_{12}$ , la diferencia entre la  $G^2$  de D y E es:

E)13,23,24,34

es  $11.2-10.1 = 1.1$  con 2 g.l.. En esta etapa no podemos agregar hacia adelante algún término. Procedemos a borrar  $u_{23}$  y la diferencia entre la  $G^2$  de E y C es  $11.9-11.2 = .7$  con 1 g.l.. Ahora no podemos agregar o borrar ningún término más. Entonces tomamos el mejor modelo a C.

Solamente con los métodos utilizados en un análisis de regresión es posible hacer una selección hacia adelante o una eliminación hacia atrás,

para elegir de los diferentes modelos el mejor. Como en un procedimiento de regresión de stepwise los valores de significancia deben ser interpretados con precaución porque las pruebas estadísticas son altamente dependientes 8.

Goodman [1971] sugiere abreviar los procedimientos stepwise utilizando métodos similares al descrito con anterioridad. Wermuth [1976] propone un procedimiento de selección hacia atrás no iterativo basado inicialmente en los modelos donde interviene la estimación directa. Benedetti y Brown [1976] comparan estrategias de selección de modelos incluidos en los métodos de selección de Goodman..

Cabe mencionar que algunos de los métodos stepwise requiere de computación, y ellos no deben ser pensados como inventos automáticos para decidir sobre el modelo loglineal apropiado.

### Estimación de parámetros

Si suponemos que se ha encontrado el mejor modelo, es decir aquel que se ajusta a las observaciones, los parámetros estimados ( $\hat{u}$ ) nos sirven para medir el efecto que cada variable tiene sobre la frecuencia observada en cada celda. Esto mide la magnitud del impacto de cada variable y de los efectos de interacción.

Los estimadores de los parámetros en el modelo ajustado se obtienen como funciones de los logaritmos de los valores esperados estimados  $m_{ijk}$ .

Dentro de las características de los estimadores se tiene que:

- 1) Los estimadores de los efectos principales sólo reflejan el impacto de los totales marginales y la magnitud más alta es el total marginal más grande.
- 2) La suma de estimadores para cada variable es cero. El signo indica que esta arriba o abajo del promedio de los logaritmos para la variable.

---

8 Bishop, op. cit pag 68

Al tener la propiedad de que la suma sobre cualquier variable es cero no es necesario hacer el cálculo de todos los parámetros, es decir se obtiene la estimación de un parámetro y por diferencias se obtienen los demás.

Para cualquier término que sea igual a cero, todos los términos de orden superior deben ser iguales a cero.

### CEROS EN LA TABLA

Cuando en una tabla tenemos casillas que contengan ceros, existen dos causas principales:

- 1) La combinación entre las categorías de cada celdas no tenga posibilidad de ocurrir empíricamente. En este caso se llaman ceros estructurales
- 2) Que el tamaño de la muestra no sea suficiente para el número de celdas de la tabla. Se llaman cero muestrales

Algunos autores como Everitt y Fienberg recomiendan métodos para eliminar los ceros en tablas. El primero recomienda aumentar el tamaño de muestra, si esto no da resultado, aumentar con una constante todas las frecuencias de la tabla. Fienber en 1969 desarrollo un método para calcular el valore de la constante en cada casilla.

Cuando tenemos ceros estructurales en una tabla ya no es necesario calcular los parámetros de esas celdas, ya que se sabe que serán cero. Los grados de libertad si tendrán que ser modificados, restandole las casillas que tienen ceros.

## **CAPITULO IV**

## ANALISIS DE DATOS

### INTRODUCCION

Como se ha visto hasta el momento los cálculos que se tienen que realizar para encontrar los resultados estadísticos en relación a las técnicas antes mencionadas es muy complicado y exahustivo, por ello es necesario utilizar como herramienta a la computadora para el desarrollo de este tipo de procedimientos, ya que sin ellas el trabajo requeriría de mucho tiempo.

Una gran ventaja que en la actualidad se cuenta con diversos programas que realizan esta tarea.

En especial para el desarrollo de este trabajo utilizaremos un paquete estadístico llamado BMDP(Biomedical Package), el cuál está formado por un conjunto de programas diseñados para el análisis de datos, en especial para el análisis de tablas de contingencia y modelos loglineales.

Cada programa se identifica por medio de un código, el que se utilizó para Tablas de Contingencia y Modelos Loglineales es 4F. y cada programa que se diseño consta de enunciados que inician con una diagonal seguido de una o varias palabras reservadas, llamado Lenguaje de Control.

El uso de este paquete se hizo en una Minicomputador VAX, y el proceso de los datos en una microcomputadora IBM compatible con el paquete DBASE.

### Descripción del Estudio

El estudio que se presenta en este trabajo se realizó en el Edo. de Morelos con el fin de determinar la prevalencia de la enfermedad de Chagas en 24 comunidades de la Jurisdicción dos de este Edo., e identificar los factores de riesgo existentes en cada categoría y evaluar su interacción.

De cada una de las comunidades se obtuvo una muestra, y se aplicó la Encuesta "Tripanosomiasis Americana en el Edo. de Morelos" que se encuentra en el anexo A. Dicha encuesta fué aplicada en 312 viviendas y a un total de 1786 individuos.

La muestra fué clasificada en 41 variables, éstas variables fueron seleccionadas con el fin de establecer una pauta para desarrollar estrategias para el control y vigilancia epidemiológica de la enfermedad.

Se describirá cada una de las variables con sus categorías correspondientes, y entre parentesis el nombre abreviado que se le dio a cada una:

PARENTESCO(PAREN): Presenta siete categorías (T01):

JEFE DE FAMILIA	HIJO	YERNO	NIETO
ESPOSA	PADRE	OTRO	

*Se observó que en la categoría de Hijo es la más predominante ya que se tienen 933 encuestados de esta categoría del total de la muestra.*

EDAD((EDAD): Se clasifica en intervalos de

01 a 04 años	05 a 09	10 a 14
15 a 24	25 a 34	35 a 44
45 a 54	55 a 65	65 o mas

*En esta variable encontramos que el intervalo de edad que contiene más población encuestada fué la que está entre 10 y 20 años de edad con 947 del total.*

SEXO(SEXO):

FEMENINO	MASCULINO
----------	-----------

*Se encontró que en la población predomina el sexo Masculino con 929 del total de la muestra.*

LUGARA DE NACIMIENTO(LUGARNAC): Se considerarán los 16 principales estados de la República.

MORELOS	VERACRUZ	GUERRERO
D. F.	HIDALGO	ESTADO DE MEXICO
PUEBLA	GUANAJUATO	MICHOACAN
CHIHUAHUA	ZACATECAS	JALISCO
OAXACA	NAYARIT	BAJA CALIFORNIA
CAMPECHE		

*Dado que la encuesta fué realizada en el estado de Morelos y está se encuentra muy cerca del Edo. de México el 30% de la población nació en este lugar.*

ESCOLARIDAD(ESCOLARI): Se consideró el grado de estudio siguiente:

ANALFABETA	PRIMARIA IN-COMP.	PRIMARIA COM-PLETA
SECUNDARIA	PREPARATORIA	TECNICO
NORMAL	PROFESIONISTA	OTRO

*De los municipios en que fué aplicada la encuesta se encontró que el grado de escolaridad predominante es la primaria incompleta con 574.*

OCUPACION(OCUPA): Las principales actividades fuerón:

AGRICULTURA	GANADERIA	ARTESANIAS
COMERCIO	HOGAR	ESTUDIANTES
PROFESOR	TECNICO	OBRERO
OTRO		

LUGAR DE RESIDENCIA: Se considerarán los 16 estados de la República clasificados en la variable Lugar de Nacimiento.

*En esta variable el estado que predominó fué el de Morelos.*

Las siguientes variables son de respuesta dicotómica:

RESIDENCIA(RESIDE). *Se encontró que 1396 de 1756 nunca cambiaron de residencia.*

OTRA RESIDENCIA(OTRARES)

CHINCHES EN LA RESIDENCIA(CHINRES). *Se encontraron 292 viviendas sin la presencia de chinches*



**CUAL RESIDENCIA(CUALRES):** Se considerarán los mismos 16 estados de la República que en la variable Lugar de Nacimiento.

*El estado que predominó fué el Edo. de México*

**CLIMA DE LA RESIDENCIA(CLIMAR2):**

FRIO	TEMPLADO	CALUROSO
------	----------	----------

Se compactó como Adecuado y No adecuado.

*El clima predominante es el templado.*

Estas variables de respuesta dicotómica(Si o No)

**VIVIR EN OTRO PAIS(OTROPA)** *Se encontró que la mayoría de los encuestados 1610 nunca viajarán a otro país.*

**PRESENCIA DE CHINCHES EN ESE PAIS(CHINPA).** *De las 12 personas que vivieron en otro país ninguna vio una chinche.*

**CUAL PAIS(CUALPA):** *El único país que fué visitado es EU.*

CANADA	ESTADOS UNIDOS	SUDAMERICA
EUROPA		

**PRUEBA SEROLOGICA(SEROLO):**

POSITIVA	NEGATIVA
----------	----------

**LOCALIDAD(LOCALIDA):**

ACA	AMU	AJU	HIG
HUA	AUT	JOJ	TIG
LOR	AMA	IXT	QUI
PAL	VIS	ROS	TEH
TEQ	TIC	TIL	TIZ
TEN	XOC	XOX	CHI

## MUNICIPIO(MUNICI):

AMA IXT	TIZ	TEN	JOJ
------------	-----	-----	-----

*El municipio de mayor población es Jojutala.*

**ESTADO(ESTADO):** Se considerarán los 16 estados de la República clasificados en la variable lugar de nacimiento.

## JURISDICCION(JURISDI):

JOJUTLA
---------

Variables de respuesta dicotómica:

**PRESENCIA DE GATO(GATO22)** *En 227 viviendas de 350 se encontraron gatos.*

**PERRO(PERRO22)** *En 226 viviendas tenían perros.*

**ARDILLA(ARDILLA):** *Solo en 10 viviendas se encontraron ardillas.*

## DORMITORIO DE LOS ANIMALES(DORAN2):

CASA	ESTABLO	PATIO	OTRO
------	---------	-------	------

Se compactó como Adentro de la casa y Fuera de la casa.

*La mayoría de las viviendas tenían a los animales en el patio.*

**MATERIAL DEL MURO(MAMU2):** Está variable se compacto en dos categorías

1) LADRILLO, CEMENTO

2) ADOBE, BARRO, PIEDRA

*El material que predominó en la construcción de muros es el ladrillo con 135 viviendas.*

**MATERIAL DE PISO(MAPI26):** Está variable se compactó en las siguientes categorías

1)CEMENTO,MOSAICO,PIEDRA

2)TIERRA,MADERA

*El material que predomina para el piso es el de cemento con 222 viviendas*

**MATERIAL DE TECHO(MATE23):** Está variable se compactó en dos categorías

1)LAMINA DE: ASBESTO,ALUMINIO,FIBRA DE VIDRIO CEMENTO/CONCRETO

2)MADERA, CARRIZO, PALMA, TEJA

*El material de techo que se encontró con más frecuencia en 136 viviendas fué de lamina de asbesto.*

Variables de respuesta dicotomica:

**LA PRESENCIA DE FISURAS EN TECHO(FISTE27)***En los techos de 234 viviendas no presentaban fisura.*

**FISURAS EN PISO(FISPI27)** *La presencia de fisuras en el piso fue detectada en 80 viviendas*

**FISURAS EN MURO(FISMU27).** *La presencia de fisuras en el muro fué detectada en 162 viviendas.*

**PRESENCIA DE CHINCHES EN LAS FISURAS(SALCHI28).** *Solo 37 viviendas, han visto la presencia de chinches en fisuras*

**VENTILACION DEL DORMITORIO(VENDOR2).** *En 166 viviendas hay buena ventilación.*

**DEL ALMACEN(VALMA)** *En 51 viviendas la ventilación es buena.*

**DONDE VIVEN ANIMALES(VENAN2)**

1)BUENA

2)REGULAR

3)MALA

*En 166 viviendas la ventilación donde los animales se encuentra es buena*

**ILUMINACION DEL DORMITORIO(ILUM2)**

**DEL ALMACEN(ILUMA2)**

**DONDE HABITAN ANIMALES(ILUMN2)**

1)BUENA

2)REGULAR

3)MALA

**INTERPRETACION DE TABLAS DE CONTINGENCIA**

En la siguiente sección tenemos el desarrollo de programas específicos para el análisis de datos en tablas de contingencia.

**Programa uno**

Este programa se diseño para el análisis de tablas de contingencia de dos entradas.

Para este programa se definió un archivo de datos que contiene toda la información recopilada en la Encuesta "Tripanosomiasis Americana en el estado de Morelos".

Una a vez que las variables fuerón definidas con sus categorías correspondientes se elaborarán tablas de contingencia, y apartir de los valores de la tabla se calculan los valores de frecuencias que se desean encontrar, bajo el supuesto de no asociación .

Se elaborarán tablas de contingencia, donde se relacionarán: la variable de interés (Serología) con todas las demás variables, y se utilizó la estadística  $\chi^2$  para decidir cuáles son las variables que están asociadas, los resultados obtenidos son los siguientes:

## Serología y Parentesco :

PARENTESCO	SERO +	$\chi^2$	N
JEFE DE FAMILIA	12.4%	.0517	1786
ESPOSA	17.3%		
HIJO	13.3%		
YERNO	12.0%		
NIETO	27.5%		
PADRE	11.1%		
OTRO	25.6%		

## Serología y edad :(T02)

EDAD	SERO +	$\chi^2$	N
1	18.5%	.3669	1786
1-4	14.6%		
5-9	17.0%		
10-14	16.8%		
15-24	13.6%		
25-34	9.3%		
35-44	10.5%		
45-54	19.4%		
55-64	21.3%		
A	15.2%		

## Serología y SEXO :(T03)

SEXO	SERO +	$\chi^2$	N
FEM	16.7%	.0464	1786
MASC	12.5%		

## Serología y Lugar de Nacimiento :(T04)

LUGARNAC	SERO +	$X^2$	N
MOR	14.3%	.6977	1786
GUE	16.9%		
EDO	30%		
PUE	20%		
JAL	25%		

## Serología y Escolaridad :(T05)

ESCOLARI	SERO +	$X^2$	N
ANALF	13.9%	.2665	1786
PRIM. INCOM.	18.0%		
PRIM. COM.	12.7%		
SECUNDARIA	12.6%		
PREPA	17.2%		
TECNICO	8.6%		
NORMAL	22.2%		
PROF	4.3%		

## Serología y Residencia en el mismo lugar :(T07)

RESIDE	SERO +	$X^2$	N
SI	20%	.9460	1786
NO	19.4%		

## Serología y Otra residencia :(T08)

OTRARES	SERO +	$X^2$	N
SI	14.8%	.8656	1786
NO	14.4%		

## Serología y Cual residencia :(T09)

CUALRES	SERO +	$X^2$	N
MOR	38.1%	.7556	1786
GUE	28.6%		
DF	17.8%		
EDO	33.3%		
PUEB	16.7%		

## Serología y Clima de la residencia indica:(T10)

	CLIMAR	SERO +	X <sup>2</sup>	N
ADEC		14.0%	.7144	1786
NADEC		16.1%		

## Serología y Chinchas en la residencia :(T11)

	CHINRES	SERO +	X <sup>2</sup>	N
SI		17.1%	.5707	1786
NO		14.7%		

## Serología y Otro país :(T12)

	OTROPA	SERO +	X <sup>2</sup>	N
SI		12.5%	.9169	1786
NO		13.8%		

## Serología y cual es el país donde vivio :(T13)

	CUALPA	SERO +	X <sup>2</sup>	N
EU		12.5%	.7077	1786

## Serología y clima del país :(T14)

	CLIMAPA	SERO +	X <sup>2</sup>	N
TEMP		16.7%	.6376	1786
CALU		14.7%		

## Serología y Chinchas en el país en que vivio:(T15)

	CHINPA	SERO +	X <sup>2</sup>	N
NO		10%	.7646	1786

## Serología y Municipio nos :(T16)

	MUNICI	SERO +	X <sup>2</sup>	N
AMA		17.3%	.0000	1786
TIZ		9.8%		
TEN		16.4%		
JOJ		21.4%		
IXT		11.9%		

## Serología y Localidad :(T17)

LOCALIDA	SERO +	X <sup>2</sup>	N
AMU	9.0%	.0023	1786
IXT	13.2%		
TEH	20.4%		
XOX	8.4%		

## Serología y Estado :(T18)

ESTADO	SERO +	X <sup>2</sup>	N
MOR	14.8%	.5559	1786

## Serología y Jurisdicción :(T19)

JURISDI	SERO +	X <sup>2</sup>	N
JOJU	14.8%	.000	1786

## Serología y presencia de Perro :(T20)

PERRO	SERO +	X <sup>2</sup>	N
SI	14.6%	.7177	1786
NO	15.5%		

## Serología y presencia de gato indica :(T21)

GATO	SERO +	X <sup>2</sup>	N
SI	12.8%	.1704	1786
NO	15.7%		

## Serología y ardilla :(T22)

ARDILLA	SERO +	X <sup>2</sup>	N
SI	14.3%	.9737	1786
NO	14.5%		

## Serología y dormitorio de los animales :(T23)

DORANI	SERO +	X <sup>2</sup>	N
DENTRO CASA	10.8%	.3312	1786
FUERA CASA	14.8%		



## Serología y material de techo :(T24)

MATE	SERO +	X <sup>2</sup>	N
EXPUESTO	.2074	1786	
NO EXP.	14.3%		

## Serología y material de muro indica:(T25)

MAMU	SERO +	X <sup>2</sup>	N
EXPUESTO	21%	.0096	1786
NO EXP.	17%		

## Serología y material de piso :(T26)

MAPI	SERO +	X <sup>2</sup>	N
EXPUESTO	24%	.3937	1786
NO EXP	18%		

## Serología y Fisura de muro :(T27)

FISMU	SERO +	X <sup>2</sup>	NTT2
SI	16.0%	.2162	1786
NO	13.4%		

## Serología y Fisura de Piso :(T28)

FISPI	SERO +	X <sup>2</sup>	N
SI	16.1%	.5932	1786
NO	14.6%		

## Serología y Fisura de techo :(T29)

FISTE	SERO +	X <sup>2</sup>	N
SI	14.9%	.8984	1786
NO	15.2%		

## Serología y presencia de chinches :(T30)

SALCHI	SERO +	X <sup>2</sup>	N
SI	14.4%	.9659	1786
NO	14.6%		

Serología y ventilación del dormitorio :(T31)

VENDOR	SERO +	X <sup>2</sup>	N
ADEC	11.4%	.0036	1786
NO ADEC	17.5%		

Serología y ventilación del almacén :(T32)

VENALMA	SERO +	X <sup>2</sup>	N
ADEC	16.4%	.9392	1786
NOADEC	23.2%		

Serología y ventilación donde duermen los animales :(T33)

VENAN	SERO +	X <sup>2</sup>	N
ADEC	17.7%	.8262	1786
NADEC	21%		

Serología e Iluminación del dormitorio :(T34)

ILUM	SERO +	X <sup>2</sup>	N
NOADE	15.7%	.0755	1786
ADEC	13.5%		

Serología e iluminación del almacén :(T35)

ILUMA	SERO +	X <sup>2</sup>	N
ADEC	17.4%	.8654	1786
NOADE	18%		

Serología e iluminación donde duermen los animales :(T36)

ILUMN	SERO +	X <sup>2</sup>	N
ADEC	15.4%	.7373	1786
NADEC	14.4%		

El interés de hacer estas tablas es probar la hipótesis de independencia entre las dos variables que aparecen en cada una de ellas.

b) Y las variables que no son independientes y se rechaza la hipótesis de independencia, son las que incluiremos en el modelo.

Basándonos en la prueba estadística  $X^2$ , las variables que estuvieron significativamente asociadas con serología son: PARENTESCO SEXO LOCALIDAD MUNICIPIO ILUMINACION DEL DORMITORIO MATERIAL DE MURO Y VENTILACION DEL DORMITORIO

### INTERPRETACION DE MODELOS LOGLINEALES

Es importante recordar que las pruebas que se utilizan en los programas, tienen un valor de significancia de .05 y si el modelo que ajustamos tiene una probabilidad menor que este valor o igual diremos que el modelo es bueno, por lo tanto al interpretar la  $X^2$  y la  $G^2$  nos basaremos en este valor.

En cuanto a las pruebas Marginal y Parcial tenemos que:

La prueba de Asociación parcial: Diferencia entre ajustar un modelo que contiene todos los marginales del mismo orden y el modelo que contenga todos los marginales del mismo orden excepto con el que empiece la prueba.

Prueba de Asociación Marginal: Esta prueba consiste en incluir todos los efectos que contiene el modelo, excepto con el efecto que empieza la prueba.

Para estas pruebas también se toma el valor de significancia .05.

#### Programa uno

La relación que existe entre los factores de riesgo con la variable de interés serología se ha visto en las tablas; ahora encontraremos las relaciones que existen entre estas variables y para ello utilizaremos los modelos loglineales.

De las variables que fueron más significativas, tenemos a la variable sexo, sin embargo está no será considerada posteriormente ya que se construyeron los modelos:

1) SV,SV,SI,VU,VI,UI

con una  $X^2 = .0765$

2) SVU,SVI,VUI

con una  $X^2 = .9735$

Como se puede observar los modelos no son significativos por lo tanto esta variable queda fuera; en cuanto a la variable localidad no será considerada ya que dentro de cada municipio se considera a sus localidades correspondientes, por lo tanto se trabajará con la variable Municipio; y en relación a la variable parentesco se consideró difícil de interpretar, dado que resulta un poco dudoso el origen de la pregunta.

Por lo tanto sólo se considerarán las variables de: municipio, ventilación del dormitorio, material de muro e iluminación del dormitorio.

Una vez que ya se han definido las variables que serán utilizadas en el modelo, se diseñó este programa que se encuentra en el anexo .

En este programa se hace una prueba de hipótesis sobre los efectos de interacción con el propósito de decidir cuál será la interacción de mayor orden que se tendrá en el modelo, y en cada renglón se incluye el término y la prueba  $X^2$  y su nivel de significancia.

Por ejemplo si tomamos el segundo renglón tenemos un modelo con interacciones de segundo orden, más todas las interacciones de orden mayor y la probabilidad de las prueba para esas interacciones se considera de  $p = 0$ , y esto implica que se necesita al menos una de esas interacciones, pero no se sabe cuál.

Como se puede observar lo que nos indica esta salida es que para ajustar el modelo incluiremos los efectos de primero, segundo y algunos de tercer orden.

En la siguiente tabla tenemos que los modelos de orden menor se están anidados uno dentro de los de mayor orden, y lo que se obtiene aquí es la diferencia de los modelos considerados en la tabla anterior, permitiéndonos corroborar que podemos incluir hasta efecto de tercer

orden, y esto lo vemos por medio de la pruebas estadísticas: ji-cuadrada y Bondad de ajuste.

Si tomamos el renglón tres tenemos que si hacemos la diferencia entre el modelo de orden dos y de orden tres equivale a probar que las interacciones de tercer orden son iguales a cero, y esto produce que las interacciones de tercer orden igual cero.

En la tercera y última tabla de este programa tenemos la prueba de Asociación Parcial y Asociación Marginal, estas dos pruebas son usadas simultáneamente para determinar:

- a)cuales interacciones son necesarias, se dice que una interacción es necesaria cuando el valor de significancia se toma al .05, y las dos pruebas se encuentran muy cerca de este nivel .
- b)cuales son cuestionables (si alguna de las dos pruebas es significativa), se dice que una interacción es cuestionable cuando alguna de las pruebas esta por arriba del nivel de significancia y la otra es significativa respetando el nivel de .05
- c)cuales interacciones no se requieren para el modelo final, esto es cuando ninguna de ls dos pruebas es significativa.

Los efectos están señalados con letras, lo que significa que:

Variable<sub>1</sub> = U (Municipio)

Variable<sub>2</sub> = M (Material de muro)

Variable<sub>3</sub> = I (Iluminación)

Variable<sub>4</sub> = V (Ventilación)

Como podemos observar en está última tabla en el renglón donde se encuentra la pareja que se espera sean independientes es municipio y ventilación ( MV con  $p = .3182$ ) que bajo los criterios antes mencionados está considerada como cuestionable y bien podría entrar en el modelo final., y cuando intervienen tres variables tenemos que algunas interacciones son cuestionables y la que podría entrar en el modelo seria: UV con  $p = .00$ .

ESTA TESIS NO DEBE SALIR DE LA BIBLIOTECA

Con esta información se está en condiciones de proponer el modelo que mejor explica a las frecuencias observadas.

El modelo que se proponen es:

$$\log m_{ijkl} = u + u_1 + u_2 + u_3 + u_4 + u_{12} + u_{13} + u_{14} + u_{23} + u_{24} + u_{34}$$

y en el programa se pide un procedimiento de stepwise y produce los siguientes modelos:

$$1) \log m_{ijkl} = u + u_1 + u_2 + u_3 + u_4 + u_{12} + u_{13} + u_{14} + u_{23} + u_{34}$$

$$2) \log m_{ijkl} = u + u_1 + u_2 + u_3 + u_4 + u_{12} + u_{13} + u_{23} + u_{34}$$

Debido a que todos son significativos, tratamos de seleccionar el que menos número de parámetros incluya, o el que más se parezca a la tabla, además de basarse en lo que el investigador proponga en base a otros estudios ya realizados.

Para hacer la mejor elección hacemos otra prueba que consiste en tomar la diferencia de  $G^2$  entre el modelo uno y el modelo dos, de ahí que seleccionamos el modelo que mejor se ajusta:

$$\log m_{ijkl} = u + u_1 + u_2 + u_3 + u_4 + u_{12} + u_{13} + u_{23} + u_{34} \text{ con } p = .00$$

Este modelo incluye solo interacciones de segundo orden.

#### Programa dos.

Una vez que tenemos el modelo final se calculan los parámetros de los términos de interacción de ese modelo.

Primero aparece el modelo propuesto con su valor de significancia de cada prueba estadística, indicando que tan bueno es el modelo, observando que el modelo es bueno con una  $p = .00$

En las primeras tablas están los valores esperados de cada variable, en seguida se muestran los residuales estandarizados

donde el cálculo es la diferencia del valor esperado y el valor observado entre la raíz cuadrada del valor esperado; y la suma de estos términos reproduce la  $X^2$ .

Estos nos permiten ver si existe un patrón que nos indique si alguna interacción falta, para que nuestro modelo sea bueno.

Estos nos permiten ver si existe un patrón que nos indique si alguna interacción falta, para que nuestro modelo sea bueno.

Como podemos observar en esta tabla al parecer no falta ninguna interacción, en general no se ven problemas con las interacciones que tenemos presentes.

En las siguientes tablas tenemos los parámetros U's y la estimación de las U's y podemos llegar a lo siguiente:

- a) Podemos observar que el material de muro no expuesto predomina sobre el material de muro expuesto.
- b) El municipio de mayor población es el de Ixtla sobre el resto de los municipios.
- c) Predomina la iluminación regular, y hay muy poca iluminación en buenas condiciones.
- d) La ventilación del dormitorio no adecuada predomina sobre la adecuada.

En relación a las asociaciones entre variables tenemos que:

1) Iluminación del dormitorio y material de muro

Cuando la iluminación es mala, el material del muro está muy expuesto; si la iluminación es buena el material de muro no está expuesto.

2) Ventilación del dormitorio e iluminación del dormitorio:

La iluminación del dormitorio cuando es buena las condiciones de ventilación también son buenas, y cuando la ventilación es no adecuada la iluminación es regular.

3) Iluminación del dormitorio y municipio:

En el municipio de Puente de Ixtla es donde se tiene la mejor iluminación en comparación con los demás municipios y el municipio que tiene en malas condiciones la iluminación es P. amaro.

4) Municipio y material de muro:

En el municipio de Tequequistingo el material de muro es más expuesto que en cualquier otro municipio.

**Por lo tanto podemos concluir que las cuatro variables están relacionadas entre sí.**



## CONCLUSIONES:

A lo largo de este trabajo se dio un panorama general de lo que es la investigación epidemiológica, y como en los últimos años y merced al vertiginoso desarrollo de las computadoras, la epidemiología ha incorporado una verdadera avalancha de procedimientos matemáticos. Cabe destacar la aplicación de los modelos matemáticos para el estudio del comportamiento de las enfermedades infecciosas o las leyes que gobiernan la propagación de las epidemias.

En la investigación epidemiológica interesan las enfermedades en cuanto atacan a los grupos humanos, razón por la cual su enfoque difiere del utilizado por la medicina clínica, en cuyo caso su universo es el individuo enfermo.

En la actualidad, el método epidemiológico es utilizado para estudiar una gran diversidad de problemas, y los cuales podrían resumirse de la siguiente manera:

1. Identificación de los determinantes o causas de las enfermedades con la finalidad de facilitar su prevención o control
2. Distribución de características biológicas en la población.
3. Identificación de poblaciones humanas con mayor riesgo para el desarrollo de una enfermedad determinada.
4. Desarrollo y evaluación de métodos para la prevención y control de las enfermedades.
5. Generación de datos esenciales para la planeación, evaluación y administración de los servicios de atención a la salud.

Por otro lado, aunque la planeación y evaluación de los servicios de atención a la salud no constituyen estrictamente campos de la epidemiología, recientemente estas actividades comenzaron a basarse en el conocimiento generado por la aplicación del método epidemiológico a los problemas de gestión de dichos servicios.

Después de dar una introducción a los conceptos básicos de epidemiología se presenta el protocolo de Investigación y se desglosa como fue planeado el trabajo y bajo que condiciones se trabajó.

Se presentaron las técnicas estadísticas que pueden ser utilizadas para tratar este tipo de estudios, demostrando que importancia tiene el uso de computadoras y programas para el análisis de los datos y en especial para tablas de contingencia y modelos loglineales .

Las tablas obtenidas nos dieron la relación que existe entre la variables y la variable de interés, permitiendo darse cuenta de las condiciones sociales y económicas en que vive la jurisdicción dos del estado de Morelos..

Al proponer el modelo final observamos que resultó ser un modelo que se ajusta a los datos, observando que las variables de Iluminación del dormitorio, Ventilación del dormitorio, Municipio y Material de muro están relacionadas entre sí.

El estudio que se planteo en este trabajo tiene como objetivo principal prevenir y controlar la enfermedad de Chagas en el Edo. de Morelos, con el modelo que se planteo se espera colaborar para la realización de estudios posteriores y tratar de prevenir esta enfermedad.

## **ANEXOS**

**UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO**

**FACULTAD DE MEDICINA**

**DEPTO. EPIDEMIOLOGIA**

**TRIPANOSOMIASIS AMERICANA EN EL EDO. DE MORELOS**

**CLAVE DEL ENTREVISTADOR:**

**FOLIO:**

**FECHA:**

**LOCALIDAD**

**MUNICIPIO**

**ESTADO**

**JURISDICCION SANITARIA**

**SECCION A: ESTA PARTE DEL CUESTIONARIO SE APLICARA A CADA PERSONA QUE SE INCLUYA EN EL ESTUDIO.**

**HEMOS SABIDO DE CASOS DE PERSONAS QUE SE ENFERMAN POR LA PICADURA DE CHINCHES EN ESTA ZONA. QUISIERAMOS QUE UD. NOS HICIERA FAVOR DE RESPONDER ALGUNAS PREGUNTAS QUE NOS AYUDARIAN A SABER SI UD. Y SU FAMILIA ESTAN EXPUESTOS A ESTE RIESGO. POR SUPUESTO QUE LO QUE UD. NOS DIGA, ES CONFIDENCIAL Y SOLO LO USAREMOS PRA PREVENIR LA PRESENTACION DE LA ENFERMEDAD EN SU COMUNIDAD.**

**1)DOMICILIO**

**2)NOMBRE DEL JEFE DE FAMILIA**

**3)COMPOSICION FAMILIAR:**

**1)JEFE DR FAMILIA 5)NIETO(A)**

**2)ESPOSA(O) 6)PADRE(MADRE)**

**3)HIJO(A) 7)OTRO(ESPECIFIQUE)**

4)YERNO(NUERA)

NOMBRE DEL MIEMBRO DE LA FAMILIA

4)PARENTESCO CON EL J. DE F.

5)EDAD

6)SEXO

7)LUGAR DE NACIMIENTO

8)ESCOLARIDAD

1.ANALFABETA5.PREPARATORIA

2.PRIMARIA INCOMPLETA 6.TECNICO

3.SECUNDARIA COMPLETA 7.NORMALISTA

4.SECUNDARIA COMPLETA 8.PROFESIONISTA

MIEMBRO DE LA FAMILIA

9)FECHA DE NACIMIENTO

10)OCUPACION

1.AGRICULTURA 6.ESTUDIANTE

2.GANADERIA 7.PROFESIONISTA

3.ARTESANO 8.TECNICO

4.COMERCIANTE 9.OBRERO

5.HOGAR 10.OTRO (ESPECIFIQUE)

11)CUANTO TIEMPO DE VIVIR EN LA COMUNIDAD?

MIEMBRO DE LA FAMILIAAÑOS

12)HA VIVIDO EN OTRA CUMUNIDAD DEL PAIS

SI NO

13) CUAL (ES)?

LOCALIDAD ESTADO

14) CUANTO TIEMPO VIVIO EN CADA UNA DE ELLAS?

AÑOS

15) EL CLIMA DE ESE(ESOS) LUGAR(ES) LA MAYOR PARTE DEL AÑO ERA?

FRIO TEMPLADO CALUROSOS

16) HABIA CHINCHES EN AQUELLAS(S) COMUNIDADES(ES)?

SI NO

17) HA VIVIDO EN OTRO PAIS?

SI NO

18) CUAL(ES) MIEMBRO DE LA FAMILIA?

LOCALIDAD PAIS

19) CUANTO TIEMPO VIVIO EN CADA UNA DE ELLAS?

AÑOS

20) EL CLIMA DE ESE(ESOS) LUGAR(ES) LA MAYOR PARTE DEL AÑO ERA?

FRIO TEMPLADO CALUROSO

21) HABIA CHINCHES EN AQUELLA COMUNIDAD?

SI NO

SECCION B.LA INFORMACION QUE A CONTINUACION SE SOLICITA, SE RECABARA MEDIANTE OBSERVACION DIRECTA DEL ENTREVISTADOR. PARA ELLO SOLICITARA PERMISO, PARA ENTRAR A LA VIVIENDA.

22)HAY ANIMALES DOMESTICOS?

PERRO(S) SI NO

GATO(S) SI NO

ARDILLA SI NO

23)EN DONDE DUERMEN LOS ANIMALES DOMESTICOS?

DENTRO DE LA CASA

ESTABLO

EN EL PATIO

OTRO (ESPECIFIQUE)

24)EL PRINCIPAL MATERIAL DE CONSTRUCCION DEL TECHO DE ESTA CASA ES:

MADERA LAMINA DE FIBRA DE VIDRIO

CARRIZO/BAMBU PALMA O ZACATE

LAMINA DE ASBESTO TEJA

LAMINA DE ALUMINIO CEMENTO/CONCRETO

25)EL PRINCIPAL MATERIAL DE LOS MUROS DE ESTA CASA ES:

ADOBE CEMENTO

PIEDRA MADERA

LADRILLO CARRIZO/BAMBU

BARRO

26)EL PRINCIPAL MATERIAL DEL PISO DE ESTA CASA ES:

TIERRA MOSAICO

CEMENTO PIEDRA

MADERA

27)SE OBSERVAN FISURAS O GRIETAS EN LA VIVIENDA:

MURO SI NO

TECHO SI NO

PISO SI NO

28)HA VISTO SALIR CHINCHES DE LAS FISURAS?

SI NO

29)LA VENTILACION DE LA VIVIENDA ES:

DORMITORIA BUENA REGULAR MALA

ALMACEN BUENA REGULAR MALA

LUGAR DONDE

DUERMEN SUS ANIMALES BUENA REGULAR MALA

30)LA ILUMINACION DE LA VIVIENDA ES:

DORMITORIO BUENA REGULAR MALA

ALMACEN BUENA REGULAR MALA

LUGAR DONDE

DUERMEN SUS ANIMALES BUENA REGULAR MALA

SECCION C. REGISTRO DE LABORATORIO.

ESTA PARTE DE LA ENCUESTA SE APLICARA A CADA UNA DE LA PERSONAS INCLUIDAS EN EL ESTUDIO:

LABORATORIO EN LA PERSONA: #PERSONA



**SEROLOGIA:**

**TITULOS DE LA PRUEBA:**

**POSITIVA NEGATIVA**

# PROGRAMA 1

/TABLE INDEX = MUNICI,MANU2,ILUMDORM,VENDOR2.

SYMBOLS = 'U','M','T','V'.

/FIT ALL

ASSOCIATION IS 5.

ITERATION = 30.

CONV = .01,.00001.

/PRINT NO OBS.

NO EXCL.

\*\*\*\*\*  
 \* TABLE PARAGRAPH 1 \*  
 \*\*\*\*\*

TOTAL OF THE OBSERVED FREQUENCY TABLE IS 1751

\*\*\*\*\* THE RESULTS OF FITTING ALL K-FACTOR MARGINALS.

SIMULTANEOUS TEST THAT ALL K+1 AND HIGHER FACTOR INTERACTIONS ARE ZERO.

K-FACTOR	D.F.	LR CHISQ	PROB.	PEARSON CHISQ	PROB.	ITERATION
0-MEAN	59	3379.54	0.00000	4811.21	0.00000	
1	51	1740.16	0.00000	1900.57	0.00000	2
2	27	130.56	0.00000	133.05	0.00000	18
3	5	58.00	0.00000	46.22	0.00000	11
4	0	0.	1.	0.	1.	

\*\*\*\*\*SIMULTANEOUS TEST THAT ALL K-FACTOR INTERACTIONS ARE SIMULTANEOUSLY ZERO.

THE CHI-SQUARES ARE DIFFERENCES IN THE ABOVE TABLE.

K-FACTOR	D.F.	LR CHISO	PROB.	PEARSON CHISO	PROB.
1	8	1639.38	0.00000	2910.63	0.00000
2	24	1609.60	0.00000	1767.52	0.00000
3	22	72.56	0.00000	86.83	0.00000
4	5	58.00	0.00000	46.22	0.00000

\*\*\*\*\* ASSOCIATION OPTION SELECTED FOR ALL TERMS OF ORDER LESS THAN OR EQUAL TO 4

EFFECT	PARTIAL ASSOCIATION		MARGINAL ASSOCIATION					
	D.F.	CHISQUARE	PROB	ITER	D.F.	CHISQUARE	PROB	ITER
U.	4	679.92	0.0000					
M.	1	325.49	0.0000					
I.	2	620.58	0.0000					
V.	1	13.39	0.0003					
UM.	4	167.05	0.0000	15	4	154.42	0.0000	2
UI.	11	119.60	0.0000	11	11	111.37	0.0000	2
UV.	4	59.41	0.0000	11	4	62.67	0.0000	2
MI.	2	106.38	0.0000	15	2	196.44	0.0000	2
MV.	1	1.00	0.3182	15	1	102.55	0.0000	2
IV.	2	964.37	0.0000	7	2	1070.32	0.0000	2
UMI.	6	12.91	0.0445	8	8	27.92	0.0005	7
UMV.	3	1.37	0.7114	5	6	10.28	0.1134	5
UIV.	8	35.18	0.0000	15	10	30.88	0.0006	17
MIV.	2	7.64	0.0219	15	2	7.97	0.0185	11
UMIV.	5	58.00	0.0000					

NUMBER OF INTEGER WORDS USED IN PRECEDING PROBLEM 4188

**CPU TIME USED 99.920 SECONDS**

**BMDP4F-TWO-WAY AND MULTWAY FREQUENCY TABLES - MEASURES OF ASSOCIATION**

**Version: 1990 (VAX/VMS) DATE: 14-AUG-91 AT 15:12:09**

**END OF INSTRUCTIONS**

**PROGRAM TERMINATED**

## PROGRAMA 2

/TABLE INDEX = MUNICI,MAMU2,ILUMDORM,VENDOR2

SYMBOLS = 'U','M','T','V'.

/FIT MODEL = MU,MLVI,UI.

ASSOCIATION IS 5.

ITERATION = 30.

CONV = .01,.00001.

/PRINT NO OBS.

NO EXCL.

EXPECTED.STANDARIZED.LAMBDA.

/END

\*\*\*\*\*  
\* TABLE PARAGRAPH 1 \*  
\*\*\*\*\*

TOTAL OF THE OBSERVED FREQUENCY TABLE IS 1751

\*\*\*\*\* THE RESULTS OF FITTING ALL K-FACTOR MARGINALS.

SIMULTANEOUS TEST THAT ALL K+1 AND HIGHER FACTOR INTERACTIONS ARE  
ZERO.

K-FACTOR	D.F.	LR CHISQ	PROB.	PEARSON CHISQ	PROB.	ITERATION
0-MEAN	59	3379.54	0.00000	4811.21	0.00000	
1	51	1740.16	0.00000	1900.57	0.00000	2
2	27	130.56	0.00000	133.05	0.00000	18
3	5	58.00	0.00000	46.22	0.00000	11
4	0	0. 1.	0. 1.			

\*\*\*\*\*SIMULTANEOUS TEST THAT ALL K-FACTOR INTERACTIONS ARE SIMULTANOUSLY ZERO.

THE CHI-SQUARES ARE DIFFERENCES IN THE ABOVE TABLE.

K-FACTOR	D.F.	LR CHISQ	PROB.	PEARSON CHISQ	PROB.
1	8	1639.38	0.00000	2910.63	0.00000
2	24	1609.60	0.00000	1767.52	0.00000
3	22	72.56	0.00000	86.83	0.00000
4	5	58.00	0.00000	46.22	0.00000

\*\*\*\*\* ASSOCIATION OPTION SELECTED FOR ALL TERMS OF ORDER LESS THAN OR EQUAL TO 4

EFFECT	PARTIAL ASSOCIATION				MARGINAL ASSOCIATION			
	D.F.	CHISQUARE	PROB	ITER	D.F.	CHISQUARE	PROB	ITER
U.	4	679.92	0.0000					
M.	1	325.49	0.0000					
L.	2	620.58	0.0000					
V.	1	13.39	0.0003					
UM.	4	167.05	0.0000	15	4	154.42	0.0000	2

UI.	11	119.60	0.0000	11	11	111.37	0.0000	2
UV.	4	59.41	0.0000	11	4	62.67	0.0000	2
MI.	2	106.38	0.0000	15	2	196.44	0.0000	2
MV.	1	1.00	0.3182	15	1	102.55	0.0000	2
IV.	2	964.37	0.0000	7	2	1070.32	0.0000	2

UMI.	6	12.91	0.0445	8	8	27.92	0.0005	7
UMV.	3	1.37	0.7114	5	6	10.28	0.1134	5
UIV.	8	35.18	0.0000	15	10	30.88	0.0006	17
MIV.	2	7.64	0.0219	15	2	7.97	0.0185	11

UMIV. 5 58.00 0.0000

\*\*\*\*\*  
\* MODEL 1 \*

MODEL	LIKELIHOOD-RATIO		PEARSON		ITER.
	D.F.	CHI-SQUARE	PROB	CHI-SQUARE	
UM,MI,IV,UI	32	192.98	0.0000	181.10	0.0000 7

\*\*\*\*\* EXPECTED VALUES USING ABOVE MODEL

VENDOR2 ILUMDORM MAMU2

MUNICI

AMA TIZ TEN JOJ IXI @PROGRAM = TOTAL

ADEC BUENA EXP	0.2	9.8	31.5	8.1	49.0	98.6
NEXP	3.3	63.5	90.6	183.8	229.3	@PROGRAM = 570.4
----- -----						
TOTAL	3.5	73.3	122.1	191.9	278.2	669.0

REGULAR EXP	1.4	4.6	15.1	5.3	16.0	42.4
NEXP	6.0	8.1	11.9	33.1	20.5	79.6
----- -----						
TOTAL	7.4	12.7	26.9	38.4	36.5	122.0

MALA EXP	0.0	1.2	2.3	1.0	1.0	5.5
NEXP	0.0	0.5	0.4	1.4	0.3	2.5
----- -----						
TOTAL	0.0	1.6	2.7	2.4	1.3	8.0

---

NADEC BUENA EXP	0.0	1.4	4.6	1.2	7.2	14.4
NEXP	0.5	9.3	13.3	26.9	33.6	83.6
----- -----						
TOTAL	0.5	10.7	17.9	28.1	40.8	98.0

REGULAR EXP	8.3	26.7	88.3	31.3	93.9	248.6
NEXP	35.2	47.6	69.7	194.3	120.5	467.4
----- -----						
TOTAL	43.6	74.3	158.1	225.6	214.5	716.0

MALA EXP	0.0	20.3	40.2	17.1	16.9	94.5
NEXP	0.0	8.0	7.1	23.6	4.8	43.5
----- -----						



TOTAL 0.0 28.4 47.3 40.6 21.7 | 138.0

\*\*\*\* STANDARDIZED DEViates = (OBS - EXP)/SQRT(EXP) FOR ABOVE MODEL

VENDOR2 ILUMDORM MAMU2 MUNICI

-----  
AMA TIZ TEN JOJ IXT  
-----

ADEC BUENA EXP -0.5 -1.8 1.5 -0.4 -1.9  
NEXP 0.4 1.1 -0.3 0.8 -0.3

REGULAR EXP -1.2 4.9 -0.8 1.6 -2.3  
NEXP -2.5 4.2 -3.4 1.0 -0.1

MALA EXP 0.0 2.6 -0.2 -1.0 -1.0  
NEXP 0.0 -0.7 -0.6 -1.2 3.3

-----  
NADEC BUENA EXP -0.2 2.1 3.4 -1.1 1.1  
NEXP -0.7 -1.7 -3.6 -1.5 2.7

REGULAR EXP 0.6 -1.9 -1.6 0.3 2.2  
NEXP 1.0 -1.8 3.6 -0.8 -1.0

MALA EXP	0.0	-0.1	0.4	-0.5	-0.2
NEXP	0.0	-0.7	-0.8	0.9	0.1

ASYMPTOTIC STANDARD ERRORS OF THE PARAMETER ESTIMATES ARE  
COMPUTED

BY INVERTING THE INFORMATION MATRIX.

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

THETA(MEAN) 2.1840

\*\*\*\*\* ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL  
ABOVE

MUNICI

----

AMA	TIZ	TEN	JOJ	INT
-----	-----	-----	-----	-----

-----

-1.253	-0.161	0.448	0.396	0.570
--------	--------	-------	-------	-------

\*\*\*\*\* RATIO OF THE LOG-LINEAR PARAMETER ESTIMATE TO ITS STANDARD  
ERROR

MUNICI

----

AMA	TIZ	TEN	JOJ	INT
-----	-----	-----	-----	-----

-----

-4.082	-1.500	4.517	4.961	5.245
--------	--------	-------	-------	-------

\*\*\*\*\* ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL  
 ABOVE

MAMU2

-----  
 EXP NEXP

-----  
 -0.351 0.351

\*\*\*\*\* RATIO OF THE LOG-LINEAR PARAMETER ESTIMATE TO ITS STANDARD  
 ERROR

MAMU2

-----  
 EXP NEXP

-----  
 -6.830 6.830

\*\*\*\*\* ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL  
 ABOVE

ROGRAMA MAMU2

MUNICI

-----  
 AMA TIZ TEN JOJ IXT

-----  
 EXP -0.335 0.097 0.504 -0.527 0.261

NEXP 0.335 -0.097 -0.504 0.527 -0.261

\*\*\*\*\* RATIO OF THE LOG-LINEAR PARAMETER ESTIMATE TO ITS STANDARD  
 ERROR

MAMU2

MUNICI

	AMA	TIZ	TEN	JOJ	IXT
EXP	-2.325	1.240	8.103	-7.456	4.477
NEXP	2.325	-1.240	-8.103	7.456	-4.477

\*\*\*\*\* ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

ILUMDORM		
BUENA	REGULAR	MALA
-0.079	0.892	-0.813

\*\*\*\*\* RATIO OF THE LOG-LINEAR PARAMETER ESTIMATE TO ITS STANDARD ERROR

ILUMDORM		
BUENA	REGULAR	MALA
-0.548	9.356	-3.981

\*\*\*\*\* ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

ILUMDORM	MUNICI				
AMA	TIZ	TEN	JOJ	IXT	
BUENA	-1.994	0.311	0.464	0.192	1.027

REGULAR 0.134 -0.225 -0.045 0.000 0.137

MALA 1.861 -0.086 -0.419 -0.192 -1.164

\*\*\*\*\* RATIO OF THE LOG-LINEAR PARAMETER ESTIMATE TO ITS STANDARD  
ERROR

ILUMDORM	MUNICI				
---	---				
	AMA	TIZ	TEN	JOJ	IXT
BUENA	-4.029	2.023	3.175	1.412	6.779
REGULAR	0.490	-1.802	-0.421	0.000	1.214
MALA	2.737	-0.382	-1.991	-1.412	-5.035

\*\*\*\*\* ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL  
ABOVE

ILUMDORM	MAMU2	
----	----	
	EXP	NEXP
BUENA	-0.682	0.682
REGULAR	-0.035	0.035
MALA	0.717	-0.717

\*\*\*\*\* RATIO OF THE LOG-LINEAR PARAMETER ESTIMATE TO ITS STANDARD  
ERROR

ILUMDORM	MAMU2	
----	----	
	EXP	NEXP
BUENA	-13.420	13.420
REGULAR	-0.773	0.773

MALA 10.257 -10.257

\*\*\*\*\* ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

VENDOR2

-----

ADEC NADEC

-----  
-0.449 0.449

\*\*\*\*\* RATIO OF THE LOG-LINEAR PARAMETER ESTIMATE TO ITS STANDARD ERROR

VENDOR2

-----

ADEC NADEC

-----  
-6.882 6.882

\*\*\*\*\* ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

VENDOR2 ILUMDORM

-----

BUENA REGULAR MALA

-----  
ADEC 1.410 -0.435 -0.974

NADEC -1.410 0.435 0.974

\*\*\*\*\* RATIO OF THE LOG-LINEAR PARAMETER ESTIMATE TO ITS STANDARD ERROR

VENDOR: ILUMDORM

-----  
BUENA REGULAR MALA  
-----

ADEC 19477 6.118 7.882

NADEC -19477 6.118 7.882

## BIBLIOGRAFIA

Everitt, B.S. *The Analysis of Contingency Tables* London Chpman and Hall, New York, John Wiley & Sons 1977.

Agresti, Alan *Analysis of Ordinal Categorical Data*. John Wiley & Sons. Universidad de Florida. 1984.

BMDP Biomedical Computer Program Series, Berkeley University of California.

Fienberg, S. *The analysis of Cross- Classified Categorical Data*. 1977.

Bishop, Fienberg y Holland. *Discrete Multivariate Analysis: Theory and Practice*. London 1975.

Mood y Graybill. *Introducción a la teoría de la estadística*. Ed. Aguilar. 1972.

Rodriguez Chavez Jose Javier. Tesis: Análisis de tablas de contingencia multidimensionales mediante el uso de modelos loglineales. 1989.

Enriquez Bolaños Jaqueline. Tesis: El uso de paquetes estadísticos para modelos loglineales. 1989.

Olguin Uribe Jorge Manuel. Tesis: Residuales en modelos para datos categóricos. 1986.

Ignacio Méndez R., Namihira, Moreno, Sosa. *El protocolo de Investigación*. Ed. Trillas. 1988.

Sánchez B.B. Tesis: Miocardiopatía Crónica e Infección por T. Cruzi en una localidad de Morelos y Tabasco. 1988.

Reseña. Situación de la enfermedad de Chagas en las Americas. BOL. OF. SANIT. PANAM. 1984.

Mazzotti, I. Dos casos de enfermedad de Chagas en el estado de Oaxaca. *Gaceta Medica de México*. 1940.



Tay j. Salazar, Ontiveros, Haro, Jiménez. Epidemiología de la enfermedad de Chagas en una población de Oaxaca. 1987.

Quintal R., Zavala, Rodríguez. La enfermedad de Chagas en el Edo. de Yucatán.

Salazar S.P.M., Bucio, Haro, Tay. Reservorios y transmisores de T. cruzi en el estado de Oaxaca. 1979.

Schenone, H., Villarroel, Rojas, Alfaro. Factores biológicos y ecológicos en la Epidemiología de la enfermedad de Chagas en Chile. 1983.