

2940
UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS



CONCEPTUALIZACION Y ANALISIS DE UN
SISTEMA DE INFORMACION (ESTADISTICO)

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

JOSE DE JESUS VALLADOLID DIAZ



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis 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.

I N D I C E

	Página
INTRODUCCION	1
CAPITULO I.- CONCEPTOS BASICOS EN LA TEORIA DE SISTEMAS Y SU ENFOQUE A SISTEMAS DE INFORMACION.	4
1.1.Historia y desarrollo de la teoría general de sistemas.	5
1.1.1.- Características de un sistema	10
1.1.2.- Estructura del sistema	13
1.1.2.1.- Características de las estructuras alternativas.	14
1.1.2.2.- Criterios para seleccionar una estructura.	16
1.1.2.3.- Diseño de la estructura del sistema.	18
1.2.Tecnología de los Sistemas de Información	19
1.2.1.- Función de un sistema de información.	19
1.2.2.- Funciones que se cumplen en el sistema de información.	21
1.2.3.- Sistemas integrados de información.	27
CAPITULO II.- DISEÑO Y JUSTIFICACION DE UN SISTEMA DE INFORMACION A TRAVES DEL ANALISIS DE UN CASO PRACTICO.	29
2.1.Ejemplo de un sistema de información estática.	30
2.2.VARIABLES que integran el sistema.	33
2.3.Fuentes de información del sistema	37
2.4.Diseño de los registros de captura de la información.	38

CAPITULO III.- ANALISIS ESTADISTICO DE UN SISTEMA DE INFORMACION.	57
3.1.Tablas de contingencia.	58
3.1.1.- "Tablas de entrada"	59
3.1.2.- Tablas de doble entrada.	64
3.1.2.1.- Para el caso de muestras pequeñas.	70
3.1.3.- Tablas de K entradas.	73
3.2.Ejemplos de tablas de contingencia: de doble y triple entrada.	82
3.3.Casos especiales de tablas de contingencia y proposición de la aplicación de la teoría.	92
3.3.1.- Tablas de contingencia con clasificación incompleta de datos, en el caso de doble entrada.	93
3.3.2.- Caso especial donde algunas celdas en la clasificación son a priori cero.	98
3.3.3.- Método de particiones de Ji-cuadrada para tablas de contingencia de K-entradas.	100
3.3.4.- Precisión de una aproximación a la potencia de las pruebas de bondad de ajuste - Ji-cuadrada con pequeñas pero iguales - frecuencias esperadas.	101
3.3.5.- Sobre el análisis de tablas de contingencia multidimensional.	102
 CAPITULO IV.- CONCLUSIONES	 103
 APENDICE.- TEORIA DE REDES (CONCEPTOS FUNDAMENTALES)	 108
 BIBLIOGRAFIA	 112

INTRODUCCION

El desarrollo de todo País va creando como consecuencia una serie de necesidades en sus diferentes sectores, lo que requiere de la intervención de técnicos y profesionistas especializados en áreas interdisciplinarias.

Uno de los sectores que requiere la intervención directa del actuario es el asegurador, el que actualmente por los problemas económicos y tecnológicos que vive nuestro País, ha manifestado la urgencia de actualización en esta actividad, por lo que es de mi interés contribuir en el análisis de un problema que en la práctica he observado, ya que actualmente en esta área es donde me desarrollo profesionalmente. Entre otras, una de mis observaciones corresponde a la carencia de sistemas de información, lo que consecuentemente provoca que los tomadores de decisiones programen actividades y determinen medidas de manera improvisada, empírica y en muchas ocasiones sin tener la menor idea de las consecuencias que estas decisiones puedan dar lugar.

El presente trabajo consiste en el desarrollo básico de lo que es un sistema de información y el tratamiento estadístico fundamental en el diseño del mismo, con la idea que éste pueda brindar a un tomador de decisiones el apoyo necesario, para contar con información confiable y oportuna, en la actualización técnica y administración de los diferentes productos o servicios que ofrece el sector asegurador en nuestro País.

La estructura del trabajo consiste en cuatro capítulos y un apéndice; el primer capítulo se denomina "Conceptos Básicos en la Teoría de Sistemas y su Enfoque a Sistemas de Información", en él se tratan los aspectos básicos de la Teoría General de Sistemas, un breve comentario de su origen, se expone la tecnología de los sistemas de información, que --

consiste en el análisis de los canales de comunicación de sistemas formales e informales, y las funciones que se cumplen en el sistema de información, tales como: recolección de datos, preparación de índices, comprensión y exposición o salida de datos entre otras.

El segundo capítulo se denomina "Diseño y Justificación de un Sistema de Información a través del Análisis de un caso Práctico", aquí se presenta el ejemplo de un sistema de información actualmente en funcionamiento, empezando por su estructuración, que para dárle validez se justifica identificando cada una de las partes del sistema, sus variables, fuentes de información, diseño de registros de captura, planteamiento de los objetivos general y particular, finalizando con un comentario sobre el método de evaluación del sistema.

El tercer capítulo se titula "Análisis Estadístico de un Sistema de Información", en éste se presentan los aspectos básicos para efectuar el análisis de independencia de variables utilizando tablas de contingencia, dando el desarrollo de "Tablas de una Entrada", que propiamente corresponde a bondad de ajuste, tablas de doble entrada y generalizando con el caso de tablas de contingencia de K entradas, presentando algunos ejemplos ilustrativos con información del propio sistema, para finalizar con casos especiales de tablas de contingencia.

El cuarto capítulo se titula "Conclusiones y Recomendaciones", aquí se presentan comentarios básicos sobre todo lo expuesto en los tres anteriores capítulos, así como las recomendaciones para su aplicación en la práctica.

Por último concluimos este trabajo con la inclusión de un apéndice denominado "Teoría de Redes (conceptos fundamentales)".

Para concluir, es mi deseo aclarar que ante la magnitud del enfoque sistemático, no me fue posible detallar, como me hubiera satisfecho, todo

lo relativo a Teoría de Sistemas, en esas condiciones me limité a exponer los conceptos básicos sobre el tema, pero con la idea de proporcionar un panorama completo.

Atentamente,

Jesús Valladolid Díaz.

CAPITULO I

**CONCEPTOS BASICOS EN LA TEORIA
DE SISTEMAS Y SU ENFOQUE
A SISTEMAS DE INFORMACION.**

1.1.- HISTORIA Y DESARROLLO DE LA TEORIA GENERAL DE SISTEMAS.

Creación de la Teoría General de Sistemas.- La noción de la Teoría General de Sistemas se debe a Von Bertalanffy, quien la formuló oralmente por primera vez en los años treinta y en varias publicaciones después de la segunda guerra mundial.

Von Bertalanffy esbozó la Teoría Dinámica de Sistemas y describió matemáticamente varias propiedades sistemáticas (totalidad, suma, crecimiento, competencia, alometría, mecanización, centralización, finalidad, equifinidad, etc.) derivadas de la descripción del sistema mediante ecuaciones diferenciales simultáneas.

A él le interesaba elaborar especialmente la Teoría de "Sistemas - Abiertos", es decir, sistemas que intercambian materiales con el medio ambiente, como es el caso de todo sistema "vivo", los que guardan múltiples relaciones con aspectos biológicos, bioquímica, fisiología, etc. en tre otras ramas de la ciencia.

Resulta erróneo afirmar que la Teoría de Sistemas surgió del esfuerzo militar emprendido en la segunda guerra mundial, pues de hecho tiene raíces muy ajenas al diseño de maquinaria bélica y a los desarrollos tecnológicos que éste ha producido; la cibernética y enfoques a fines fueron desarrollados en forma independiente mostrando muchas analogías con la Teoría General de Sistemas.

El término Teoría General de Sistemas fue propuesto por Bertalanffy con una acepción deliberada de universalismo, ya que anotado lo que nos rodea existen los sistemas en abundancia. Define sistema como un conjunto de elementos que se relacionan entre ellos y con el medio.

A continuación se mencionan varias definiciones de "Sistemas".

- 1a. "Conjunto de reglas o principios enlazados entre sí" [1].
- 2a. "Conjunto de cosas que ordenadamente relacionadas entre sí contribuyen a determinado objetivo" [2].
- 3a. "Simplemente, en un sentido común organizado y en una definición mas formal, tenemos que función de sistemas y procedimientos podría ser el análisis de los planes de acción colectivos, procedimientos, formas y equipo, con el fin de simplificar y estandarizar las operaciones de la oficina" [3].
- 4a. "Un sistema es un todo organizado y complejo, implica un complejo interconectado de componentes o partes fundamentales relacionadas, que forman un todo unitario" [4].
- 5a. "Un sistema es un conjunto de atributos y la historia de los cambios que ocurren en ese conjunto" [3].
- 6a. "Un sistema es un conjunto de partes interdependientes o interactuantes, cuyas relaciones entre sí o entre sus atributos, determinan un todo unitario que realiza determinado efecto, función u objetivo"[3].
- 7a. Miguel Duhalt Krauss destaca las siguientes notas esenciales del concepto de sistemas:
 - a) "Un conjunto de combinación de cosas o partes.
 - b) Integrados o Interdependientes.
 - c) Cuyas relaciones entre sí y con sus atributos las hacen tomar un todo unitario u organizado.
 - d) que cumplen determinado propósito o realizan determinada función,
 - e) y que puede mantener cierto grado de estabilidad aunque la materia y la energía que lo compongan están sujetas a cambios constantes", [3].
- 8a. De Ackoff define el sistema como un complejo de entidades relacionadas entre sí, esto incluye tanto los sistemas conceptuales como los concretos [4]

Para situar sistema en el sentido que nos interesa, diremos que es un conjunto de elementos coordinados y relacionados entre sí pero independientes, elementos que serán información estadística, con un fin determinado explícitamente como ser instrumento de apoyo para un tomador de decisiones, planeación, administración, etc.

Los enfoques sistemáticos incluyen la Teoría General de Sistemas - (en su acepción más restringida) como son, la cibernética, la teoría de - autómatas, la informática, la de control, la de conjuntos, la de gráficas y redes, la de juegos y decisiones, la computarización y simulación, etc., por lo que generalmente se acepta que un sistema es un modelo de índole - general.

En el universo que nos rodea existen sistemas en abundancia de tal - manera que determinan si la multitud de sistemas tiene realidad objetiva - o si solo existen en la mente del observador, se convierte en un problema de índole filosófica.

Los sistemas podemos dividirlos en dos grupos: Reales o Concretos y Abstractos.

Sistema Real o Concreto.- Es aquel en que al menos dos de sus elementos son objetos. El establecimiento de la existencia y propiedades - de sus elementos y la naturaleza de las relaciones entre ellos requiere de la investigación con un componente empírico. Estos sistemas son temas de estudio de las ciencias no-formales. Son las entidades percibidas o deducidas de la observación, cuya existencia es independiente del observador, por ejemplo: un sistema de defensa antiaérea, una computadora, una red telefónica, sistema nervioso central en un mamífero y el universo físico que sustituye un supersistema gigantesco, entre otros, que como se ven están constituidos por el hombre o por la naturaleza con un fin preciso.

Sistema Abstracto .- Es aquel en que todos sus elementos son conceptos y creados por definición y las relaciones entre ellos son creadas por supuestos (axiomas y postulados). Estos sistemas son los temas de estudio de las llamadas Ciencias Formales. Por tanto son sistemas conceptuales que tienen correspondencia con la realidad, sin embargo su distribución no resulta tan clara como parece a primera vista, por ejemplo: la lógica y las matemáticas.

Surgen del esfuerzo humano por abarcar todo lo complejo de la siguiente manera: está demostrado que es habitual subdividir los problemas complejos en sub-problemas menos complejos, los que eventualmente por repetición del proceso es factible llegar a solucionar subproblemas. De ese modo, la solución compuesta de una jerarquía de subproblema representa una solución satisfactoria del problema global, por ejemplo la química, que exhibe una estructura jerárquica de subramas.

El estado de un sistema, en un momento de tiempo.- Es el conjunto de propiedades relevantes que el sistema tiene en ese tiempo. Cualquier sistema tiene un número ilimitado de propiedades. Sólo algunas de esas son importantes para cualquier investigación en particular. Los valores de las propiedades importantes constituye el estado del sistema en algunos casos pueden ser interesantes dos posibles estados, por ejemplo.- el cerrado y el abierto ó despertar y dormir.

El medio ambiente de un sistema.- El conjunto de elementos y sus propiedades importantes los cuales no son parte del sistema, pero un cambio en cualquiera de ellos puede producir un cambio en el estado del sistema. Elementos externos que afectan a propiedades irrelevantes de un sistema no son parte de ese medio ambiente.

El medio ambiente del estado de un sistema en un momento de tiempo.- Es el conjunto de propiedades relevantes en ese tiempo. El estado de un elemento o subconjunto de elementos de un sistema o el medio ambiente puede

ser definido de manera semejante.

Los elementos que forman el medio ambiente de un sistema y el medio ambiente mismo pueden ser conceptualizados como sistema cuando llegan los focos de atención. Todos los sistemas pueden ser conceptualizados como parte de otro y gran sistema.

Sistema Cerrado.- Es aquel que no tiene medio ambiente. No tiene interacción con cualquier elemento no contenido en él, es completamente contenido el mismo.

Sistema abierto.- Es aquel que hace su medio ambiente.

Sistema (ambiental) Evento.- Es un cambio en una o más propiedades de la estructura de un sistema en un período de tiempo de duración determinada; eso es un cambio en el estado estructural del sistema (o ambiente).

Sistema (de un estado) estático.- Es aquel en el que no ocurren eventos.

Sistema Dinámico (multi-estático).- Es aquel en el cual ocurren eventos y cambian sus estados en el tiempo.

Sistema Homeostático.- Es un sistema estático del cual sus elementos y medio ambiente son dinámicos.

Reacción de un Sistema.- Es un sistema evento para el cual otro evento que ocurre al mismo evento o medio ambiente es suficiente; es deterministicamente causado por otro evento.

Respuesta de un Sistema.- Es un sistema evento para el cual otro evento que ocurre al mismo sistema o al medio ambiente es necesario pero

no suficiente; esto es un sistema evento producido por otro sistema evento.

Acto de un sistema.- Sistema evento para el cual no hay cambios en el medio ambiente del sistema es en cada uno necesario o suficiente. Son por tanto, eventos determinados por ellos mismos, cambios autónomos.

Comportamiento de un Sistema.- Sistema evento el cual es necesario o suficiente para otro evento en ese sistema o su medio ambiente. Es un -- cambio en el sistema el cual inicia otros eventos. Las reacciones, respuestas y acciones pueden construir por si mismas el comportamiento.

1.1.1. CARACTERISTICAS DE UN SISTEMA.

La característica esencial de un sistema consiste en estar compuesto por partes que ejercen interacción, cada una de las cuales reviste un interés propio. Sin estas interacciones el estudio de los sistemas sería relativamente poco interesante, pues son ellas las que enriquecen el comportamiento de un sistema y hacen de su análisis una tarea muy compleja.

Antes de continuar es importante mencionar una serie de consideraciones sobre el significado de un sistema, mismas que comentaré brevemente.

Aspectos básicos sobre el significado de un Sistema:

- 1.- Los objetivos del sistema considerado como un todo y más específicamente las medidas de actuación del sistema completo.
- 2.- El medio ambiente del sistema: las restricciones fijas.
- 3.- Los recursos del sistema.
- 4.- Los componentes del sistema, sus actividades, metas y medidas de -- actuación.
- 5.- La administración del sistema.

Al determinar los objetivos de un sistema se corren ciertos riesgos ya que la tarea de un diseñador consiste en probar si no se descuida el - objetivo mas importante y se cumplen con otros de menor importancia, por lo que resulta relativo definir los objetivos de un sistema, y aquí se presenta un error muy común ya que se recalca lo que es obvio.

El medio ambiente del sistema es el contexto donde se encuentra, es decir, lo que está "fuera" del sistema, lo cual implica que el sistema no puede hacer nada al respecto a sus características o su comportamiento. - Integra las cosas y las personas que son "constantes" o dadas desde el punto de vista de sistemas. Ejemplo: restricciones presupuestadas que pueden - estar en el medio ambiente.

No es tan solo algo que está fuera de control del sistema, sino que es algo que determina como opera el sistema. Uno de los aspectos más importantes en el ambiente del sistema es el "Programa de Requerimientos".

Los recursos del sistema, se encuentran dentro del sistema; son los medios que utiliza el sistema para hacer sus trabajos. Son las cosas que el sistema puede cambiar y utilizar para su propio provecho.

Un aspecto importante en la determinación de los recursos, es poner atención especial a los avances tecnológicos que pueden ayudar y aumentar considerablemente el empleo eficaz de los recursos (aquí se debe poner atención a los recursos existentes y como en el futuro se pueden aumentar).

Los recursos son el depósito general fuera del cual los actos específicos del sistema pueden moldearse. Los actos específicos son ejecutados por los componentes o partes del sistema.

Los componentes de un sistema están integrados por subpartes y están ligados mediante interfases.

Las interfases aumentan la forma de entradas y salidas tales como entidad física o material o energía o información.

Ciertos componentes del sistema transforman las entradas y salidas y se dice que está perfectamente bien definido cuando se conoce el proceso de transformación. En muchos casos la transformación se refiere a entradas y salidas de orden físico y la información solo actúa como medio de coordinación, por otra parte, un sistema puede referirse únicamente a la información, como ocurre en el caso de un sistema de información estadística y que es el objeto del presente trabajo.

La descripción completa del comportamiento de un sistema exige el conocimiento del comportamiento de cada uno de sus componentes, así como las interrelaciones entre sus componentes.

Cuando hablamos de la magnitud de un sistema, resulta relativo, ya que los límites de un sistema son necesariamente arbitrarios, si consideramos que cualquier rama de la jerarquía de un sistema puede ser considerada como un sistema en sí mismo, por tanto los puntos terminales inferiores en la jerarquía de un sistema son asimismo, esencialmente arbitrarios, de ahí que no exista una preocupación por detallar minuciosamente los niveles jerárquicos de un sistema.

Por último la administración de un sistema tiene que referirse a la generación de los planes para el sistema, o sea la consideración de todos los aspectos que se han mencionado, las metas generales, el medio ambiente, la utilización de recursos y los componentes. La administración establece las metas de los componentes, asigna los recursos y controla la actuación del sistema. Para el administrador científico, el enfoque de sistemas implica la construcción de un "Sistema de Información" que habrá de registrar la información relevante para la toma de decisiones y específicamente habrá de señalar la mejor información sobre el uso de los re-

cursos incluyendo las oportunidades despreciadas.

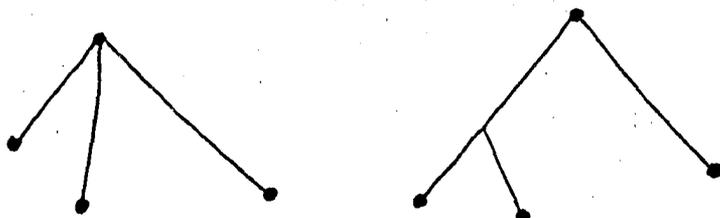
1.1.2. ESTRUCTURA DEL SISTEMA.

Cuando se busca una estructura para el sistema, se debe relacionar una que combine la interrelación y las actividades complementarias de manera de reducir la interdependencia entre los componentes y alcanzar eficiencia en los objetivos globales del sistema.

La estructura de un sistema es la resultante de la subdivisión o factorización de metas globales en una escala jerárquica de subsistemas menos complejos. Para ello se realiza un análisis de medios y fines que relaciona los resultados deseados (metas globales) con los medios para lograrlos.

La estructura del sistema describe la manera en que se combina jerárquicamente el conjunto de tareas elementales a fin de construir el sistema total, debiendo entender por tarea elemental como el subsistema de nivel más bajo, a final de cada una de las diferentes cadenas (actividad-eventos) 1), la que puede representar un subsistema relativamente amplio, pero sin tener una estructura definida en términos de actividades de nivel inferior.

Cada estructura representa un sistema diferente, en esta exposición nos limitaremos a la denominada "árbol" 2), que se define como aquella que no tiene ciclos (es decir, cada subsistema es componente de un solo subsistema superior) y todas las partes están conectadas. Las "hojas", al final de cada rama del árbol, representan las tareas elementales del sistema.



Dos estructuras con tres tareas cada una.

- 1) En algunos libros aparece como medios-fines.
- 2) Para un desarrollo más amplio, consultar apéndice.
Teoría de Redes.

1.1.2.1. CARACTERISTICAS DE LAS ESTRUCTURAS ALTERNATIVAS.

Examinaremos algunas de las características mas importantes de las estructuras alternativas como son: a) Grado de Fragmentación del Sistema, b) Complejidad de las tareas de los subsistemas, y c) Comunicaciones dentro del sistema.

a) Grado de Fragmentación del Sistema.- La fragmentación jerárquica del sistema produce, invariablemente interacciones entre sus partes, e impone asimismo, una barrera a la coordinación a través de los límites que separan las diversas partes del sistema. De este modo simplifica el sistema al crear subsistemas relativamente aislados, aunque se corre el riesgo de introducir barreras en la comunicación entre actividades interactuantes.

b) Complejidad de las tareas de los subsistemas.- Veamos cuales son los efectos de la estructura sobre la complejidad de la tarea de un subsistema. El número de relaciones potenciales entre tareas contiguas de nivel inferior nos proporciona un índice aproximado de la relativa complejidad de una tarea de nivel superior. El número total de subconjuntos dentro de un conjunto con "n" objetos es 2^n (incluyendo el conjunto vacío). De ahí surge que cuando no hay fragmentación, un sistema con "n" tareas elementales comprende $2^n - 1$ relaciones.

Al producirse fragmentación, el subsistema solo debe ocuparse detalladamente de aquellas tareas que le atañen en forma directa. Si se utiliza un alcance de SCN* el número de relaciones potenciales se reduce de $2^n - 1$ a $2^s - 1$, así el problema se reduce y deja de ser evidentemente irresoluble para hacerse manejable. Si existen fuertes interacciones, el alcance se debe mantener dentro de los límites estrechos a fin de que las tareas se mantengan, asimismo, dentro de las factibilidades de realización. Si las tareas implican interacciones relativamente débiles, el alcance puede ser aumentado, sin afectar la complejidad de aquellas.

Por lo tanto, es arriesgado generalizar más acerca del alcance ideal. Su determinación implica un compromiso excesivamente complejo entre fragmentación y coordinación.

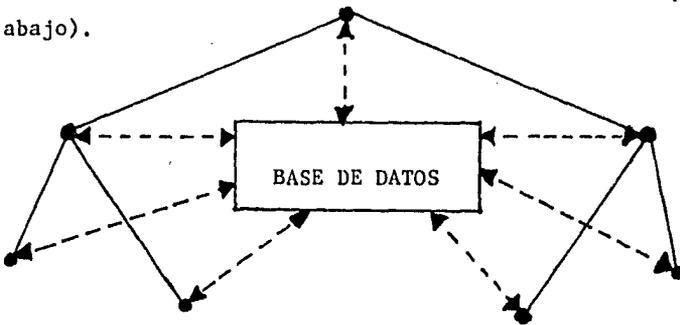
c) Comunicaciones dentro de Sistemas.- El comportamiento "óptimo" de cualquier subsistema dado, depende de las actividades de otros subsistemas, por lo que exige que los canales de comunicación proporcionen información acerca de tales actividades.

Los canales jerárquicos de comunicación al contrario de los directos reducen en gran medida tanto el número de eslabones totales en la red, como el de los que debe mantener cada subsistema por separado.

* Denota S subconjunto de N (Conjunto de Números Naturales).

Debido a la "distancia" que separa las tareas a lo largo de los eslabones jerárquicos, es típico que la red de informaciones de un sistema incluya cierto número de importantes eslabones directos entre actividades estrechamente vinculadas.

Una alternativa mas atrayente consiste en superponer sobre los canales jerárquicos un sistema de información que eslabona cada tarea con un conjunto de información común o base de datos. (Como se aprecia en la figura de abajo).



1.1.2.2. CRITERIOS PARA SELECCIONAR UNA ESTRUCTURA.

Como hemos visto para diseñar una estructura se dispone de un gran número de alternativas, de las cuales se debe elegir la más adecuada, de acuerdo a las necesidades y objetivos del sistema, para lo cual mencionamos enseguida algunos criterios para seleccionar una estructura, entre los que se encuentran los siguientes: a) Requerimientos del Manejo de la Información, b) Interacción entre las Actividades, c) Complementación entre las Actividades, d) Concordancia de los Objetivos Asignados con los de Mayor Nivel.

a) Requerimientos del Manejo de la Información.- Partiendo de la base que no se tienen limitaciones para el manejo de la información, ya que en caso contrario se restringe severamente la elección de una estructura adecuada que se elija debe presentar la mayor fragmentación posible y el -

menor número de canales de comunicación, siempre pensando que las exigencias de computo sean en lo posible lo menos complejas.

El manejo de la información que se requiere en un determinado nodo depende de la naturaleza y el número de las subtareas y del tipo de las - restricciones de alto nivel que se imponen, las que puedan aumentar o disminuir las exigencias en el manejo de la información en el nivel inferior. Por lo que al seleccionar una estructura, esencialmente hay que imponer restricciones tales que se elimine toda consideración de las alternativas que pudieran tener una probabilidad baja de figurar dentro de una escala ceptable de exigencias relacionadas con el manejo de infomación.

b) Interacción entre las Actividades.- Normalmente, las activida--des combinadas en un determinado nodo presentan una interacción estrecha. Existen dos razones principales para esto: tal estructura facilita las - comunicaciones y simplifica la tarea de ocuparse de las interacciones.

Si se agrupan las tareas estrechamente vinculadas por interacciones, se reducirá en gran parte el uso de la vía jerárquica por la que se debe - fluír la información ó por otra parte las tareas cuya interacción es débil, pueden hacerse bastante distantes, ya que es muy poca comunicación que existe entre esas actividades.

c) Complementación entre las actividades.- Dos o más tareas son -- complementarias cuando el resultado está mas cerca del objetivo al efectuar las juntas que separadas.

El grado en que una determinada estructura permita la explotación de esas complementarias constituye un criterio importante para juzgar si es - adecuada o no; luego entonces la combinación de actividades complementarias crea interacciones adicionales o dicho de otra forma las reconoce en forma explícita.

d) Concordancia de los Objetivos Originados con los de Mayor Nivel.- Un sistema fragmentado en una jerarquía de subsistemas implica que cada -- una de las tareas fragmentadas resultantes es parcialmente independiente - de todas las otras tareas.

Al diseñar subsistemas de nivel superior, se pretende lograr en los niveles inferiores un comportamiento acorde con los objetivos de nivel superior, pa ra lo cual se imponen diversos subobjetivos y restricciones sobre esos sub- sistemas inferiores. Por lo tanto, la eficiencia de determinado conjunto - de tareas de nivel inferior está determinada, en parte por la capacidad de imponer cuando se hace el diseño del subsistema en alto nivel, restricciones razonables que promuevan el comportamiento adecuado.

1.1.2.3. DISEÑO DE LA ESTRUCTURA DEL SISTEMA.

Al respecto se analizarán brevemente dos aspectos que se consideran los más importantes en el diseño de la estructura del sistema y son:

a) Optimización del Sistema, y b) Creatividad o Heurística para - la determinación de la Estructura.

a) Optimización de la Estructura.- Los procedimientos que se siguen para determinar la fragmentación de todo el sistema están personificados - en la heurística, mas que en reglas bien definidas. Se han desarrollado - algunos procedimientos de optimización, pero su aplicación general es muy reducida.

Por ejemplo Alexander (1964) ha realizado una interesante tentativa de desarrollo de un algoritmo para la síntesis de los sistemas.

El primer paso de un método consiste en especificar un conjunto de necesidades de diseño (correspondientes a las tareas elementales), y luego

se determina la interacción entre cada par de necesidades.

También he preparado un método de distribución de estas necesidades en una jerarquía de subconjuntos, de modo tal que sus interacciones se reduzcan a lo mínimo. Cada uno de los subconjuntos, consiste en necesidades con fuerte interacción, pueden ser atendidos como un problema de diseño - relativamente independiente; sólo que este método está limitado por el hecho de que únicamente permite interacciones entre pares, y se supone que - las que surgen entre tres o mas actividades son insignificantes.

b) Heurísticamente para la Determinación de la Estructura.- La descripción de una buena metodología para el diseño de sistemas es en rigor, sólo una descripción de resolución de problemas. Es evidente que la mente humana capaz de tratar al mismo tiempo sino un número reducido de variables que actúan entre sí.

Cuando se diseña un sistema complejo, implica una solución de una - serie de subproblemas locales, una vez resuelto, éste entra consecuentemente a agregarse a otros subsistemas de un modo similar; de esta manera, el proceso de diseño se convierte en un desmenuzamiento progresivo del problema global.

En consecuencia lo que se busca en una heurística relativamente simple para eliminar aquellas alternativas que sin duda, resultan inadecuadas.

1.2. TECNOLOGIA DE LOS SISTEMAS DE INFORMACION.

1.2.1. FUNCION DE UN SISTEMA DE INFORMACION.

La función básica consiste en proporcionar información a fin de tomar decisiones y proceder a su coordinación, lo que se lleva a cabo a -

través de sus componentes con funciones tales como captación, dosificación, transmisión, almacenaje, recuperación, transformación y presentación de la información.

Un sistema de información está superpuesto a una estructura organizativa subyacente. Sus canales de comunicación incluyen los nexos jerárquicos de la organización, por lo que la red de información representa un sistema definido, cuyo objetivo es proporcionar información a los diferentes nodos de decisión dentro de la organización.

Los canales de comunicación de un sistema pueden ser de dos tipos:

a) formales y b) informales.

a) Formales.- La característica esencial consiste en que su existencia es reconocida explícitamente y definida como parte del sistema total, cuya descripción determina el origen y destino de la información, características como frecuencia de la misma, probabilidades de error, la prioridad, - entre otras; y se manifiesta generalmente de manera escrita.

b) Informales.- La característica esencial consiste en que suple - las omisiones que puede dejar de considerar los canales formales, cumpliendo de esta manera una función extremadamente importante de reforzar la información, sin embargo, constituye un apoyo poco seguro, como es el caso de los rumores, que se constituyen por canales poco fidedignos y muy "ruidosos".

De esta manera podemos decir que un sistema de información, está formado de componentes, tanto humanos como automáticos, involucrados en la toma de decisiones, la coordinación y la monitorización; pero para nuestra exposición sólo nos interesamos en lo que involucre el procesamiento de información.

Por otro lado y en consecuencia tenemos que la salida de un sistema de información puede ser clasificada en dos categorías:

- 1) Información para Toma de Decisiones no Programadas.
- 2) Información para Actividades Programadas, Humanas y Mecanizadas.

Información que puede ser extraída simplemente de algún tipo de archivo, hasta la aplicación de diversas transformaciones tanto simples como sumamente complejas.

1.2.2. FUNCIONES QUE SE CUMPLEN EN EL SISTEMA DE INFORMACION.

Existen diversos tipos de entradas y salidas en los sistemas de información, en las transformaciones cumplidas por los diversos subsistemas y su estructura. Sin embargo podemos señalar una serie de funciones básicas entre ellas, que es importante destacar:

- a) Recolección de datos.
- b) Clasificación de datos y preparación de índices.
- c) Compresión de los datos.
- d) Archivo de datos.
- e) Administración de los datos.
- f) Computación.
- g) Transmisión de los datos.
- h) Exposición (salida) de datos.

Se explicará brevemente en que consiste cada uno de ellos:

a) Recolección de Datos.- Consiste en la captura de la información que atañe a la organización y a su ambiente. Toda la información procesada por el sistema deberá, en un momento u otro, ser recolectada de algún

modo, aquí un aspecto importante el que se refiere al volumen de datos recolectados, por más modesta que sea una organización es capaz de generar - un gran volumen, siendo que cualquier acontecimiento que se registre deberá describirse mediante algún tipo de codificación, a fin de ser incorporado al sistema de información. Según la eficiencia y amplitud del código, cada uno de los acontecimientos puede dar origen a un volumen considerable - de datos.

La recolección de datos consiste de dos etapas, Captación y Registro, para lo cual existen métodos, que van desde los completamente manuales hasta los anatómicos en su totalidad, lo que desde el punto de vista presupues- tal pueden ser caros y susceptibles de un margen relativamente alto error, por lo que es importante reducir el volumen de datos recolectados, pudiendo hacerse simplemente a través de disminuir la vigilancia sobre el ambiente. Teniendo menor número de "Puntos de Muestreo", es factible muchas veces reducir el volumen de los datos recogidos, sin que, descienda el valor de la información; también es importante recolectar la misma información más de - una vez, otra manera de reducir el volumen de los datos recolectados consis- te en predecir los datos mas que en recolectarlos.

b) Clasificación de Datos y Preparación de Índices.- Consiste en aso- ciar un acontecimiento con otros que guardan semejanzas significativas y su objetivo es particularizar los acontecimientos según ciertas dimensiones -- significativas para la toma de decisiones.

La dificultad de clasificar y preparar el índice depende, en gran par- te del tipo de complejidad de datos que se manejan.

Los términos índices que describen los datos narrativos y representan una abstracción: tratan de resumir la esencia del dato, de modo que se -- pueda inferir el contenido por los términos usados. Esto presenta la evi- dente ventaja de reducir sustancialmente la cantidad de datos que ha de --

examinarse para localizar la información deseada. Además estos términos presentan otra ventaja tal vez menos evidente: una terminología y formatos normalizados, pero existe el riesgo de que el sistema pase por alto información importante.

c) Compresión de Datos.- Consiste en reducir el volumen de datos que se transmiten dentro de la organización, sin reducir demasiado su contenido de información. Para lo que existen varios métodos, entre los que se encuentran: Filtrado de Información Insignificante, Compresión de Datos Probabilísticos y Agregado de Datos.

Filtrado de Información Insignificante.- Dado que la gran masa de datos que se introducen en un sistema reviste relativamente poca importancia, en forma ideal tenemos que, el sistema de información debería filtrar datos insignificantes y dejar pasar solamente las "excepciones" que obligan a una modificación de los planes urgentes (Ackoff 1967). La identificación de todas las excepciones requiere una formalización del proceso de planeamiento. Normalmente, esto no se puede hacer, de modo que los sistemas de detención de las excepciones están ideales para eliminar solo las no excepciones relativamente definidas que encuadran dentro de las tolerancias permitidas para un patrón establecido. Al fijar límites para el proceso de eliminación, el diseñador del sistema se enfrenta, evidentemente, con el problema de balancear el error de dejar pasar un dato no significativo frente al de eliminar una excepción verdadera.

Agregado de Datos.- Consiste en agregar los datos de las dimensiones de clasificación cuya conservación no se considera importante o no se desea conservar y reducir de esta manera el volumen de los datos presentados. La utilidad de este proceso se basa en las dimensiones eliminadas no tienen importancia para el propósito perseguido.

Compresión de Datos Probabilísticos.- Partiendo de que muchos de los

datos recolectados por la organización se refieren a variables probabilísticas, las que podemos describir de varias maneras que difieren considerablemente en su grado de comprensión. El mayor grado se obtiene cuando una variable se describe mediante un solo parámetro.

La comprensión de datos de este tipo tiene dos propósitos. En primera, reducir el volumen de los datos comunicados. En segunda, generar una mayor consistencia a través de toda la organización.

d) Archivo de Datos.- El archivo de datos cumple la función de memoria y permite que la organización actúe sobre la base de informaciones relativas a un lapso de tiempo arbitrario. Esto se justifica ya que todo comportamiento actual tiene su origen en el pasado; la falta de memoria paralizaría toda la organización. Aquí se destacan cuatro aspectos importantes que son: Organización de la Base de Datos, Identificación de los Datos Archivados, Medios de Archivo y Jerarquía de Archivos.

Organización de Datos.- Consiste en la suma total de la información archivada, la que debe ser estructurada de manera que muestre las importantes relaciones entre los elementos de los datos, para facilitar la recuperación de la información. Además, cada elemento ha de tener un formato especializado y se le debe asignar un lugar de archivo especificado. (También ese lugar puede ser variable).

Identificación de los Datos Archivados.- Partiendo de que los datos archivados deben ser identificados de alguna manera, tenemos que esta identificación puede ser explícita o implícita.

La identificación en forma explícita consiste en que, los datos se acompañan cada uno con una etiqueta o marca, debidamente codificada; luego la forma implícita, consiste en que la identificación de los datos está en la posición de los mismos, ya sea la posición física absoluta o la relativa

a otros datos. La gran ventaja de la identificación implícita consiste en eliminar la necesidad de archivar la etiqueta junto con los datos sustantivos.

Medios de Archivo.- Cualquier medio que pueda adoptar dos ó más estados estables puede ser empleado para archivar datos. Estos se codifican de tal manera que existe una correspondencia uno a uno entre determinado valor de un dato y una disposición de los estados estables.

Para aquellos datos que se van a procesar mecánicamente existen evidentes ventajas en archivar la información de manera tal que resulte factible leer directamente por medio de algún mecanismo automático.

Los medios magnéticos ofrecen una combinación especial de propiedades que los convierte en la forma de archivo más usada.

Jerarquía de Archivos.- Debido a las amplias diferencias en volumen, velocidad y costo, la base de datos se archiva por lo general en una jerarquía de mecanismos físicos de archivo.

A fin de minimizar el tiempo de acceso promedio, los datos que tienen la mayor probabilidad de recuperación deben ser asignados a los mecanismos que presentan el mayor tiempo de acceso.

e) Administración de los Datos.- El objetivo de esta función consiste en facilitar el acceso a la información contenida en la base de datos, en el que destacan dos aspectos importantes: Requerimientos de la Administración de Datos y en consecuencia Sistemas de Administración de Datos.

Requerimientos de la Administración de Datos.- Este es uno de los aspectos más críticos de todo el sistema de información, por lo que debe cumplir con los siguientes requisitos: eficiencia en el procesamiento de rutina -

consultas específicas, seguridad en sus tres aspectos (protección contra - pérdida, el secreto y la validez), generalización y flexibilidad.

Aquí sólo analizaremos la parte seguridad, ya que en la práctica - se habla mucho de pérdida de archiveros, mencionando brevemente en que con siste cada aspecto:

- 1) La seguridad contra pérdida o destrucción.- Se obtiene mediante la duplicación.
- 2) La seguridad de conservación de información secreta o - confidencial.- Consiste en que ciertas partes de la base de datos están completamente reservadas a una categoría especificada de usuarios.
- 3) La seguridad de validez.- Consiste en que la informa-- ción tiene que ser sometida a una serie de controles en cuanto a formato, alcance, etc. de donde se infiere si un dato es fidedigno o no.

Sistemas de Administración de Datos.- Una función de administración de datos que cumpla con todas las exigencias de procesamiento de rutina, se guridad, generalización y flexibilidad debe ser necesariamente muy compleja, de ahí que se han desarrollado sistemas de administración de datos, los que descuidan algún aspecto de los mencionados, concentrando mayor importancia sobre todo, en la eficiencia del procesamiento de las operaciones de rutina y descuidando sobre todo las exigencias de generalización y flexibilidad; - y esto en gran parte está propiciado por la capacidad limitada de la mayoría de los equipos de procesamiento.

f) Computación.- Su función consiste en que, como incluye todos los procesos dentro del sistema de información, transforma los datos de entrada

en datos de salida; luego dentro de los procesos se pueden manejar datos de toda clase numéricos, alfabéticos y alfanuméricos y el procesamiento puede ser secuencial o al azar, de acuerdo a las necesidades de cada usuario del sistema y del tipo de proceso efectuado.

g) Transmisión de Datos.- Consiste en la comunicación entre puntos del sistema separados geográficamente. La comunicación de la información puede efectuarse mediante el movimiento físico del medio en que está registrado el dato (papel, cinta magnética, etc) por transmisión de una señal eléctrica en la que está codificada la información.

La exhibición de la información comúnmente se presenta en listados -- limitándose al sentido de la vista, pero actualmente hay mecanismos que -- expresan la información en exposición auditiva, a través de teléfono, entre otros.

1.2.3. SISTEMAS INTEGRADOS DE INFORMACION.

Partiendo de que el objetivo común de todo aquel que diseña sistemas de información es de construir un sistema de información integrado o total, analizaremos los dos aspectos de la integración:

- a) El mayor acoplamiento del mismo sistema de información.
- b) Un sistema de información que permite un acoplamiento más estrecho de las diversas partes de la organización.
- c) Acoplamiento mas estrecho del sistema de información.

El sistema de información, como todos los sistemas, puede ser diseñado con diferentes grados de independencia entre sus partes. En la práctica se ha visto que, los subsistemas de información han exigido una independencia considerable, simplemente porque la capacidad de procesamiento --

de la información de que disponían los diseñadores era limitada, que la tarea solo se podría cumplir en forma muy fragmentada. Fragmentación que se manifiesta, en principio, en la recolección de datos, en la organización de los archivos y en el alcance de la computación.

En un sistema de información fragmentado, cada subsistema es responsable de la recolección de sus propios datos y en otro caso el sistema integrado de información está organizado alrededor de una base común de datos.

Trayendo como consecuencia que en el sistema fragmentado existe una duplicidad de datos, lo que en el caso de tener una base de datos común se elimina y en consecuencia dando mayor fluidez a los procesos y sobre todo a un menor costo.

d) Acoplamiento mas estrecho entre las actividades de la organización:

La integración de un sistema constituye una condición necesaria, pero no suficiente para la integración de actividades. Un acoplamiento mas estrecho de estas últimas requiere de planeación que abarque un campo más amplio. Esto significa que dicha planeación debe cumplir mas variables y tener en cuenta mayor cantidad de interacciones entre las subunidades de la organización. De ahí que el sistema de información debe ser capaz de:

1) Proporcionar un acceso fácil a los datos relativos a una amplia gama de actividades, y 2) cumplir el complejo procesamiento que se requiere para lograr planeación comprensiva.

CAPITULO II

DISEÑO Y JUSTIFICACION DE UN SISTEMA
DE INFORMACION A TRAVES DEL ANALISIS
DE UN CASO PRACTICO.

2.1 EJEMPLO DE UN SISTEMA DE INFORMACION ESTADISTICA.

Debido a que en el presente trabajo se pretende exponer un ejemplo ilustrativo pero válido, únicamente se analizará un subsistema del sistema de información de una compañía de Seguros de Daños, ya que si se pretendiera analizar por completo resultaría muy complejo por el volumen - de conceptos y variables que contiene.

Se tratará de hacer un análisis lo más detallado posible de los aspectos que intervienen en un sistema de información, así como de los datos que resultan básicos para una actualización y administración adecuada de los seguros que se manejan en el ramo de daños, en base al comportamiento que describen las variables que intervienen.

También quiero hacer mención, que el tratamiento que se le dá a esta información, es básico, por lo que de manera semejante se pueden analizar los demás subsistemas del sistema de información, a excepción de lo que se refiere a riesgos catastróficos como es el seguro de terremoto y erupción volcánica, que requiere de análisis mucho más profundos y complejos.

El sistema a que se hace mención se refiere al seguro sobre equipo e instalaciones electrónicas, que corresponde al ramo de diversos.

Brevemente y de manera concisa mencionaré los riesgos que pueden cubrirse mediante este seguro, para que de manera introductoria nos formemos una idea de la información que contendrá el sistema y así resulte mas sencillo el análisis de la justificación del mismo.

Este seguro ampara contra daños y pérdidas que sufren los bienes asegurados en forma súbita o imprevista, a consecuencia de:

- a) Incendio, Impacto de Rayo. Explosión, Implosión.

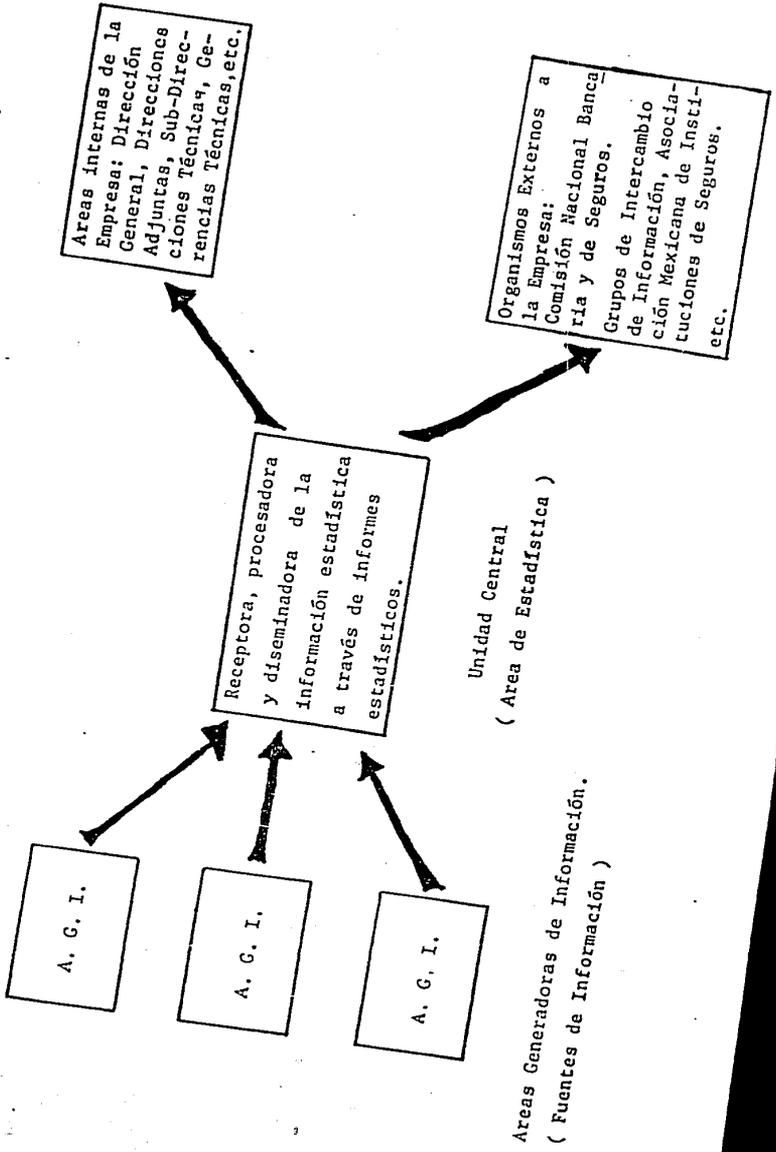
- b) Humo, Hollín, Gases, Líquidos o Polvos Corrosivos.
- c) Corto Circuito, Azogamiento, Arco Voltáico, Perturbaciones por Campos Magnéticos; Aislamiento insuficiente, Sobretensiones causadas por rayo, Tostadura de Aislamiento.
- d) Errores de construcción, Fallas de Montaje, Defectos de Material.
- e) Errores de Manejo, Descuido, Negligencia e Impericia.
- f) Daños mal intencionados y dolo de terceros.
- g) Robo con Violencia.
- h) Granizo, Helada, Tempestad.
- i) Hundimiento del Terreno, Deslizamiento de Tierra, Caída de Rocas, Aludes.
- j) Otros Accidentes no Excluidos en esta Póliza.

Así como los siguientes que pueden ser cubiertos pero de manera adicional:

- k) Terremoto y Erupción Volcánica.
- l) Ciclón y Huracán.
- m) Inundación y Daños por Agua
- n) Huelgas, Alborotos Populares y Conmoción Civil.
- o) Robo sin Violencia y Asalto.
- p) Gastos Adicionales: se amparan los gastos por concepto de flete expreso, trabajos en días festivos y horas extras.
- q) Gastos por Flete Aéreo.
- r) Riesgos de Climatización.
- s) Equipos Móviles y Portátiles fuera de los Predios Asegurados.

La siguiente gráfica presenta el funcionamiento del Sistema de Información a través de un flujograma, en el que se notarán los aspectos fundamentales del mismo, los que a continuación se analizarán.

FLUJOGRAMA DEL SISTEMA DE INFORMACION DE UNA COMPAÑIA DE SEGUROS DE DAÑOS



Areas Usuarías de la Información

Con una breve explicación del flujograma, obtendremos lo siguiente:

El sistema se alimenta de la información que le proporcionan las áreas generadoras, las cuales constituyen sus fuentes directas, el sistema de información se desarrolla en todas sus actividades de recepción, - proceso y diseminación de la información en una Unidad Central que en nuestro caso se le denomina Area de Estadística, la cual tiene como función - final el informar y asesorar a las Areas Internas de la Empresa como a los Organismos Externos de la Empresa, las que constituyen las áreas usuarias de la información y en consecuencia ellos constituyen los tomadores de decisiones tanto a nivel microestadístico (interno de la empresa), como macroestadístico (a nivel de mercado).

2.2. VARIABLES QUE INTEGRAN EL SISTEMA.

Se presenta de manera concentrada cada variable que interviene en el sistema de información, explicando el significado del concepto que integra, aunque esto mismo se expondrá en detalle en la parte referente a "Condiciones Generales, Reglas, Tarifa y Coberturas del Seguro de Equipo e Instalaciones Electrónicas o Electromagnéticas".

El conjunto de variables que constituyen el sistema, son los siguientes:

- 1.- Clasificación de Equipo.
- 2.- Sección.
- 3.- Riesgo.
- 4.- Unidades expuestas.
- 5.- Sumas Aseguradas expuestas

- 6.- Primas emitidas.
- 7.- Primas devengadas.
- 8.- Número de siniestros.
- 9.- Monto de siniestros.
- 10.- Deducibles afectados.
- 11.- Gastos de ajuste.
- 12.- Salvamentos.

CLASIFICACION DE EQUIPO.- Se refiere a la clave correspondiente a cada equipo, de acuerdo al catálogo de claves para tarificación del inciso No. 2 de las reglas y tarifas de este seguro como sigue:

(Se anexa un ejemplar del catálogo en la parte referente a las condiciones generales, reglas y tarifas de este seguro).

- 0100 Instalaciones y equipos de procesamiento electrónico de datos.
- 0200 Instalaciones de telecomunicación.
- 0300 Instalaciones emisoras y receptoras, incluyendo equipos de radio, televisión y cinematografía.
- 0400 Instalaciones de luz.
- 0500 Instalaciones para investigación y análisis de materiales.
- 0600 Instalaciones y equipos eléctricos de medicina.
- 0700 Instalaciones de señales y transmisión.
- 0900 Instalaciones técnicas y equipo para oficinas y técnica gráfica y reproducción.
- 1000 Instalaciones de control y registro.

SECCION.- Se refiere a las secciones de la tabla de cotización (I.- Daños Materiales, II.- Portadores Externos de Datos y. III.- Incremento en el costo de operación.)

RIESGO.- Se refiere a los siguientes riesgos cubiertos:

- DM-a) Incendio, impacto de rayo, explosión, implosión.
- DM-b) Humo, hollín, gases, líquidos o polvos corrosivos.
- DM-c) Corto circuito, azogamiento, arco voltáico, perturbaciones por campos magnéticos; aislamiento insuficiente, sobretensiones -- causadas por rayo, tostadura de aislamientos.
- DM-d) Errores de construcción, fallas de montaje, defectos de material.
- DM-e) Errores de manejo, descuido, negligencia e impericia.
- DM-f) Daños mal intencionados y dolo de terceros.
- DM-g) Robo con violencia.
- DM-h) Granizo, helada, tempestad.
- DM-i) Hundimiento del terreno, deslizamiento de tierra, caída de rocas, aludes.
- DM-j) Otros accidentes no excluidos.
 - a) Terremoto y erupción volcánica.
 - b) Ciclón y Huracán.
 - c) Inundación y daños por agua.
 - d) Huelgas, alborotos populares y conmoción civil.
 - e) Robo sin violencia y asalto.
 - f) Gastos adicionales.
 - g) Gastos por flete aéreo
 - h) Climatización.
 - i) Equipos móviles y portátiles.

UNIDADES EXPUESTAS.- Corresponden al número de unidades aseguradas calculadas en veinticuatroavos y de acuerdo al período transcurrido de la póliza dentro del ejercicio que se reporta.

SUMAS ASEGURADAS EXPUESTAS.- Se refiere a la suma de los montos de sumas aseguradas de las unidades expuestas en el ejercicio que se -

reporta.

PRIMAS EMITIDAS.- Se refieren a los importes, en miles, de primas de las pólizas emitidas con principio de vigencia en el año que se reporta y endosos que las afectan, así como endosos que afectan a las pólizas emitidas en el año anterior.

PRIMAS DEVENGADAS.- Son los importes, en miles, de las primas emitidas calculadas en veinticuatroavos, de acuerdo al período transcurrido dentro del ejercicio que se reporta, más las primas que quedaron por devengarse de lo emitido en el ejercicio anterior.

NUMERO DE SINIESTROS.- Es la cantidad de siniestros ocurridos en el ejercicio que se reporta correspondientes a pólizas con vigencia en el mismo.

MONTO DE SINIESTROS.- Es el importe, en miles, de todos los siniestros ocurridos en el ejercicio que se reporta afectando a pólizas con vigencia en el mismo.

DEDUCIBLES AFECTADOS.- Se refiere al porcentaje de la suma asegurada, con la cual el asegurado participa en el monto de los siniestros.

GASTOS DE AJUSTE.- Son los importes, en miles, de pagos efectuados en el ejercicio que se reporta por la atención a siniestros ocurridos correspondientes a pólizas en vigencia en el mismo.

SALVAMENTOS.- Son los importes, en miles, recuperados por salvamentos en el ejercicio que se reporta de los siniestros ocurridos, correspondientes a pólizas con vigencia en el mismo.

2.3. FUENTES DE INFORMACION DEL SISTEMA.

Para explicar las fuentes de información primero mencionaremos las áreas generadoras de la información y luego el documento fuente el cual sirve como auxiliar del sistema de captura de información - del sistema.

Las áreas fuente o generadoras de la información, las clasificaré en dos: a) Emisión y, b) Siniestros.

a) Areas de emisión:

<u>NOMBRE</u>	<u>DOCUMENTO FUENTE</u>
- Depto. de Emisión del Ramo de Diversos.	- Póliza, endoso "A" y endoso "D"
- Depto. de Tarifas.	- Reglas y Tarifas.
- Depto. de Ingeniería de Riesgos Diversos.	- Expediente de la póliza.

b) Areas de Siniestros:

- Depto. de Siniestros del Ramo de Diversos.	- Aviso de Siniestro, ajustes y orden de pago de siniestro.
- Depto. de Ingeniería de Riesgos Diversos.	- Expediente de la póliza.

2.4. DISEÑO DE LOS REGISTROS DE CAPTURA DE LA INFORMACION.

Sobre este punto hablaremos de dos aspectos: 1ro. la codificación de la información y, 2do. la captura de la información.

Codificación de la información.- Se realizará de acuerdo al diseño que a continuación se describe, en formas debidamente requisitadas.

A continuación se presenta la descripción de los formatos de co
dificación y se incluyen formatos:

<u>NOMBRE DEL DATO</u>	<u>COLUMNAS</u>	<u>C O M E N T A R I O</u>
SUMA ASEGURADA SECC.I	32 - 37	SUMA ASEGURADA CORRESPONDIENTE - A LA SECCION 1, SE ANOTA EN MILES.
SUMA ASEGURADA SECC.II	38 - 43	SUMA ASEGURADA CORRESPONDIENTE - A LA SECCION II, SE ANOTA EN MILES.
SUMA ASEGURADA SECC.III	44 - 49	SUMA ASEGURADA CORRESPONDIENTE A LA SECCION III, SE ANOTA EN MILES.
PRIMA SECCION I	50 - 55	PRIMA CORRESPONDIENTE A LAS COBERTURAS DE LA SECCION I, SE ANOTA EN PESOS.
PRIMA SECCION II	56 61	PRIMA CORRESPONDIENTE A LAS COBERTURAS DE LA SECCION II, SE ANOTA EN PESOS.
PRIMA SECCION III	62 - 67	PRIMA CORRESPONDIENTE A LAS COBERTURAS DE LA SECCION III, SE ANOTA EN PESOS.
CUOTA SECCION I	68 - 71	CUOTA CORRESPONDIENTE POR COBERTURA EN SECCION I.
CUOTA SECCION II	72 - 75	CUOTA CORRESPONDIENTE POR COBERTURA EN LA SECCION II.
CUOTA SECCION III	76 - 79	CUOTA CORRESPONDIENTE POR COBERTURA EN LA SECCION III.
DEDUCIBLE SECCION I	80	DEDUCIBLE ES POR TIPO DE EQUIPO SEGUN CATALOGO ANEXO.
DEDUCIBLE SECCION II	81	DEDUCIBLE POR TIPO DE EQUIPO, - SEGUN CATALOGO ANEXO.
DEDUCIBLE SECCION III	82	DEDUCIBLE POR TIPO DE EQUIPO - SEGUN CATALOGO ANEXO.

DESCRIPCION DEL FORMATO DE EMISION.

<u>NOMBRE DEL DATO</u>	<u>COLUMNAS</u>	<u>COMENTARIO</u>
--	1 - 2	b. SON PARA USO FUTURO, POR LO QUE SE DEJAN EN BLANCO.
NUMERO DE POLIZA	3 - 9	EN LAS COLUMNAS 3 A 8 SE ANOTA EL NUMERO DE POLIZA, ENDOSO "A" O ENDOSO "D" Y EN LA COLUMNA 9, SE ANOTA EL DIGITO VERIFICADOR.
INCISO	10 - 14	SE ANOTA EL NUMERO PROGRESIVO DE IDENTIFICACION DE CADA UNIDAD, PARA EL CASO DE POLIZAS O ENDOSOS QUE CUBRAN VARIAS UNIDADES.
VIGENCIA	15 - 18	EN LAS COLUMNAS 15 Y 16 SE ANOTA EL MES DE VENCIMIENTO DE LA POLIZA O ENDOSO Y EN LAS COLUMNAS 17 Y 18 SE ANOTA EL AÑO DE VENCIMIENTO.
AGENTE	19-23	CLAVE DEL AGENTE A QUE CORRESPONDE ESE DOCUMENTO.
TIPO DE EQUIPO	24 - 27	SE ANOTA LA CLAVE CORRESPONDIENTE A CADA EQUIPO DE QUE SE TRATA, SEGUN CATALOGO ANEXO.
SECTOR	28	SE ANOTA UN 1 SI ES SECTOR OFICIAL Y UN 2 SI ES SECTOR PARTICULAR.
POBLACION	29 - 30	SE ANOTA LA CLAVE DE UBICACION Y LUGAR DE OPERACION DEL EQUIPO SEGUN CATALOGO ANEXO.
MONEDA	31	SE ANOTA 1 SI ES MONEDA NACIONAL Y SE ANOTA UN 2 SI ES DOLARES. EN CASO DE OTRA MONEDA, HAY QUE HACER LA CONVERSION A DOLARES.

<u>NOMBRE DEL DATO</u>	<u>COLUMNAS</u>	<u>C O M E N T A R I O</u>
COBERTURAS POR SECCIONES	83 - 94	COLUMNA 83 SE ANOTA UN 1 SI TIENE LA COBERTURA BASICA DE LA SECCION I, COLUMNA 84 SE ANOTA UN 1 SI TIENE LA COBERTURA BASICA DE LA SECCION II, COLUMNA 85 SE ANOTA UN 1 SI TIENE LA COBERTURA BASICA DE LA SECCION III. DE LA COLUMNA 86 A LA 94, SE ANOTA UN 1 SI TIENE CADA COBERTURA MEDIANTE ENDOSO, SEGUN CATALOGO ANEXO.
SUMA REASEgurADA	95 - 100	SUMA ASEGURADA CEDIDA EN REASEGURO.

NOTA: LAS COLUMNAS QUE QUEDAN EN BLANCO, SON PARA USO FUTURO.

DESCRIPCION DEL FORMATO DE SINIESTROS.

<u>NOMBRE DEL DATO</u>	<u>COLUMNAS</u>	<u>C O M E N T A R I O</u>
--	1 - 2	PARA USO FUTURO, QUEDAN EN BLANCO.
NUMERO DE POLIZA	3 - 9	NUMERO DE POLIZA O ENDOSO SINIESTRADO.
INCISO	10 - 14	NUMERO DE INCISO DE IDENTIFICACION DE LA POLIZA SINIESTRADA.
NUMERO DE SINIESTRO	15 - 20	NUMERO PROGRESIVO QUE SE ASIGNA A CADA SINIESTRO.
FECHA DE OCURRENCIA	21 - 24	COLUMNAS 21 Y 22 MES DE OCURRENCIA Y COLUMNAS 23 Y 24 SE ANOTA EL AÑO DE OCURRENCIA.
TIPO DE EQUIPO	25 - 28	SE ANOTA LA CLAVE DE TIPO DE EQUIPO SEGUN CATALOGO ANEXO.
SECTOR	29	SE ANOTA 1 SI ES SECTOR OFICIAL Y 2 SI ES PARTICULAR.

<u>NOMBRE DEL DATO</u>	<u>COLUMNAS</u>	<u>C O M E N T A R I O</u>
POBLACION	30 - 31	SE ANOTA LA CLAVE DE LA ENTIDAD EN DONDE OCURRIO EL SINIESTRO SEGUN - CATALOGO ANEXO.
MONEDA	32	SE ANOTA 1 SI ES MONEDA NACIONAL Y 2 SI ES DOLARES, EN CASO DE QUE LA MONEDA EXTRANJERA SEA OTRA, HAY - QUE CONVERTIRLA A DOLARES.
COBERTURAS AFECTADAS	34 - 44	COLUMNA 33 COBERTURA BASICA DE LA SECCION I, COLUMNA 34 COBERTURA BA SICA DE LA SECCION II, COLUMNA 35, COBERTURA BASICA DE LA SECCION III COLUMNAS 36 A 44, SE ANOTA UN 1 EN CADA COLUMNA SEGUN CATALOGO ANEXO.
IMPORTE DEL SINIESTRO	45 - 50	MONTO ESTIMADO O AJUSTE DE MAS O MENOS AL SINIESTRO.
CAUSA DEL SINIESTRO	51 - 52	CAUSA DEL SINIESTRO, SEGUN CATALOGO ANEXO.
IMPORTE DEL SALVAMENTO	53 - 58	MONTO DE LA RECUPERACION POR SALVA MENTO DEL SINIESTRO.
GASTOS DE AJUSTE	59 - 63	MONTO DE GASTOS DE AJUSTE POR CON- CEPTO DE SINIESTRO.
SUMA REASEGURADA RECU PERADA.	64 - 69	MONTO DE LA SUMA ASEGURADA RECUPE- RADA POR REASEGURO.

Captura de Información.- Se realizará de medios magnéticos de captura y constituirá nuestro banco de datos, el cual quedará graba do en cintas magnéticas, de acuerdo al registro presentado en el -- punto anterior. Este aspecto será tratado posteriormente en la jus tificación del sistema de información, presentando también la salida del sistema que está constituida por un conjunto de reportes básicos de información, a los cuales en un capítulo posterior se les dará - un tratamiento estadístico.

Expondremos la justificación de la validez del sistema de información que se analiza conforme a las consideraciones básicas siguientes, mismas que conceptualmente mencionamos en la primera parte de esta exposición:

- 1a.- Identificación del objetivo del sistema.
- 2a.- Identificación del medio ambiente del sistema.
- 3a.- Identificación de los recursos del sistema.
- 4a.- Identificación de los componentes del sistema.
- 5a.- Administración del sistema.
- 6a.- Método de Evaluación del sistema.

Para la identificación del objetivo del sistema lo dividiremos en un objetivo general y varios objetivos particulares, de la siguiente manera:

OBJETIVO GENERAL

Proporcionar a un tomador de decisiones un medio de apoyo, para una debida administración del seguro de equipo electrónico, así como una adecuada actualización del producto en base a los avances tecnológicos que determinen la necesidad de modificar las condiciones de aseguramiento.

OBJETIVOS PARTICULARES

- Conocer la estructura de nuestra cartera del seguro de equipo electrónico.
- Determinar la eficiencia de las cuotas a que se tasa el seguro en base la siniestralidad registrada y como consecuencia actualización de las mismas.
- Determinar las capas de reaseguro que se afectan y de esta manera medir la eficiencia de nuestros planes de retención.

- Determinar los intervalos en que se acumula la siniestralidad y medir la eficiencia de los deducibles aplicados.
- Determinar los riesgos mas frecuentemente afectados, para hacer recomendaciones en la administración de los mismos y en consecuencia una mejor selección de riesgos.
- Conocer la explotación del seguro, en base al potencial del mercado.

La medida de actuación del sistema quedará asentada, aunque de manera teórica, posteriormente en el capítulo en que se efectúe el análisis estadístico de la información y a través de ello se puedan obtener resultados favorables.

La identificación del medio ambiente del sistema resulta un poco complicado y además relativo porque hay que plantear dos interrogantes sobre los elementos que actúan en el sistema; primera, ¿Podría yo hacer algo acerca de ello?; segunda, ¿Influyen en mis objetivos?. Si la respuesta a la primera pregunta es NO, y SI a la segunda, entonces sí está en el medio ambiente. Trataré de ser breve al mencionar cada uno de ellos.

- 1.- El presupuesto que se determine para llevar acabo la implantación del sistema, el cual por política de la compañía es fijo -- anualmente.
- 2.- Los riesgos catastróficos que pueden crear desvíos importantes en mi cartera asegurada.
- 3.- La falta de conocimiento de la teoría estadística de las personas que generan la información (policitadores, agentes de seguros, - ajustadores de siniestros, etc..)
- 4.- La existencia de criterios flexibles para la selección de riesgos y en consecuencia una inapropiada clasificación de la información.

- 5.- La existencia de eventos fortuitos que provoquen una desviación en las actividades programadas anteriormente.

Considero que principalmente estos elementos componen el medio ambiente del sistema, ya que la ocurrencia de ellos pueden provocar un cambio importante en los objetivos del sistema e inclusive el que fracase.

Para la identificación de los recursos del sistema, partiremos de lo siguiente: primero estos se encuentran dentro del sistema, y segundo son los medios que utiliza el sistema para efectuar sus funciones. Estos al contrario del medio ambiente, son las cosas que el sistema puede cambiar y utilizar para su propio provecho.

Los principales recursos del sistema son los siguientes:

- 1.- El presupuesto destinado a la implantación y desarrollo del sistema en todos sus aspectos.
- 2.- El personal que interviene en la generación de la información.
- 3.- El equipo electrónico que se usa en la captura y proceso de la información generada.
- 4.- El personal que interviene en la captura, proceso y diseminación de la información.
- 5.- El apoyo que se le da al plan de actividades del sistema por las áreas generadoras y usuarias de los productos de la información.
- 6.- La información fuente generada por las áreas de emisión, reaseguro, siniestros, salvamentos, etc.

Luego en lo referente a la identificación de los componentes del sistema, se dice que la única razón de separar el sistema en componentes es para proporcionar al analista el tipo de información que necesita para poder decir si el sistema está operando adecuadamente y lo -

que se debe hacer a continuación. Podemos mencionarlo como sigue:

El componente básico queda determinado por el conjunto de diseños y programas elaborados para la captura, proceso y diseminación de la información, además del banco de datos que constituye el sistema de información el cual se presenta en una serie de reportes básicos denominados cuadros estadísticos del seguro de equipos e instalaciones -- electrónicas o electromagnéticas y que son los que a continuación se mencionan:

Cuadro 1.- Datos estadísticos de 1982 del seguro de equipos e instalaciones eléctricas o electromagnéticas, por clase de equipo y sección asegurada.

Cuadro 2.- Datos estadísticos de 1982 del seguro de E.E., por clase de equipo y riesgo asegurado.

Cuadro 3.- Datos estadísticos de 1982 del seguro de E.E. por deducible, riesgo afectado y clase de equipo.

Cuadro 4.- Datos estadísticos de 1982 del seguro de E.E. por rangos de sumas aseguradas, riesgo y clase de equipo.

Cuadro 5.- Datos estadísticos del año de 1982 del seguro de E.E. por rangos de siniestralidad, según riesgo y clase de equipo.

(Se anexan formatos de los cinco cuadros)

Por último podríamos decir que los tomadores de decisiones en los diferentes aspectos de la información son los usuarios del sistema - que van a determinar las necesidades de modificación y actualización del mismo, así como del seguro a que se refiere.

En lo que respecta a la administración del sistema, será llevado al cabo mediante el control de la medida de actuación y el cumplimiento de los objetivos antes mencionados, que a través de nuestros usuarios nos determinarán, como ya se dijo en el párrafo anterior, las necesidades de actualización del sistema en base, a los desarrollos tecnológicos que los bienes asegurables vayan presentando y en consecuencia, las nuevas condiciones del mercado para este fin.

Por último, el método de evaluación del sistema será tratado posteriormente, y aquí únicamente mencionaremos que nos servirá como medio de apoyo en la optimización de los objetivos.

DATOS ESTADISTICOS DE 198
 DEL SEGURO DE
 EQUIPO E INSTALACIONES
 ELECTRONICAS O ELECTROMAGNETICAS

SECCION _____
 (MILES DE PESOS)

<u>CLASIFICACION DE EQUIPO</u>	<u>UNIDADES EXPUESTAS</u>	<u>SUMAS ASEGURADAS EXPUESTAS</u>	<u>P R I M A S EMITIDAS DEVENGADAS</u>	<u>NUMERO DE SINIESTROS</u>	<u>MONTO SIN</u>	<u>GASTOS DE AJUSTE</u>	<u>SALVAMENTOS</u>
0100							
0200							
0300							
0400							
0500							
0600							
0700							
0800							
0900							
1000							

T O T A L

* SECCION: DAÑOS MATERIALES, MATERIAL PORTADOR EXTERNO DE DATOS, INCREMENTO EN LE COSTO DE OPERACION.

DATOS ESTADISTICOS DE 198
DEL SEGURO DE

EQUIPO E INSTALACIONES
ELECTRONICAS O ELECTROMAGNETICAS
POR COBERTURAS DE LA SECCION I Y II

RIESGO

(MILES DE PESOS)

CLASIFICACION DE EQUIPO	UNIDADES EXPUESTAS	SUMAS ASEGURADAS EXPUESTAS	P R I M A S EMITIDAS DEVENGADAS	NUMERO DE		MONTO		GASTOS DE	
				SINIESTROS	SIN	CON	SIN	AJUSTE	SALVAMENTOS
0100									
0200									
0300									
0400									
0500									
0600									
0700									
0800									
0900									
1000									

T O T A L

RIESGO: INCENDIO, NEGLIGENCIA, CORTO CIRCUITO, AGUA, RAYO, HURTO; (BASICA) TERREMOTO, CICLON, INUNDACION, HUELGAS, ROBO SIN VIOLENCIA, GASTOS ADICIONALES, FLETE AEREO, CLIMATIZACION Y EQUIPOS MOVILES.

DATOS ESTADISTICOS DE 198
 DEL SEGURO DE
 EQUIPOS E INSTALACIONES
 ELECTRONICAS O ELECTROMAGNETICAS
 POR DEDUCIBLE APLICADO

** RIESGO _____
 CLASE DE EQUIPO _____
 (MILES DE PESOS)

* DEDUCIBLE	UNIDADES EXPUESTAS	SUMAS ASEGURADAS EXPUESTAS	P R I M A S EMITIDAS DEVENGADAS	NUMERO DE SINIESTROS	MONTO SIN	GASTOS DE AJUSTE	SALVAMENTOS
0.10							
0.15							
0.20							
0.25							
0.30							
0.35							
0.40							
0.45							
0.50							
1.00							
2.00							
3.00							
10.00							

T O T A L

* EL DEDUCIBLE ES EN POR CIENTO DE LA SUMA ASEGURADA.

** UN CUADRO RESUMEN PARA CADA CLASE DE EQUIPO POR GRUPO GENERICO (100, 200, 300,1000) Y POR RIESGO: INCENDIO, NEGLIGENCIA
 CORTO CIRCUITO, AGUA, RAYO, HURTO, OTROS. BASICA, TERREMOTO, CICLON, INUNDACION, HUELGAS, ROBO SIN VIOLENCIA, GASTOS ADICIO-
 NALES, FLETE AEREO, CLIMATIZACION Y EQUIPOS MOVILES.

DATOS ESTADISTICOS DE 198
 DEL SEGURO DE
 EQUIPOS E INSTALACIONES
 ELECTRONICAS O ELECTROMAGNETICAS
 POR RANGOS DE SUMAS ASEGURADAS

RIESGO _____
 CLASE DE EQUIPO _____
 (MILES DE PESOS)

R A N G O S UNIDADES SUMAS ASEGURADAS P R I M A S NUMERO DE MONTO GASTOS DE
EXPUESTAS EXPUESTAS EMITIDAS DEVENGADAS SINIESTROS SIN AJUSTE SALVAMENTOS

HASTA 300
 300 - 2000
 2000 - 8000
 8000 - 16000
 16000 - 32000
 32000 +

T O T A L

RIESGOS: MISMOS DEL CUADRO No.2

DATOS ESTADISTICOS DEL AÑO 198__
 DEL SEGURO DE
 EQUIPOS E INSTALACIONES
 ELECTRONICAS O ELECTROMAGNETICAS
 POR RANGOS DE SINIESTRALIDAD

RIESCO _____
 CLASE DE EQUIPO _____
 (MILES DE PESOS)

R A N G O S UNIDADES SUMAS ASEGURADAS P R I M A S NUMERO DE MONTO GASTOS DE
EXPUESTAS EXPUESTAS EMITIDAS DEVENGADAS SINIESTROS SIN AJUSTE SALVAMENTOS

HASTA 300
 300 - 2000
 2000 - 8000
 8000 - 16000
 16000 - 32000
 32000 - +

T O T A L

RIESGOS: MISMOS DEL CUADRO No.2

DATOS ESTADISTICOS DEL AÑO 198
 DEL SEGURO DE
 EQUIPOS E INSTALACIONES
 ELECTRONICAS O ELECTROMAGNETICAS

CUADRO No. 6

* RIESGO

CLASE DE EQUIPO

(MILES DE PESOS)

<u>P E R D I D A</u>	<u>UNIDADES EXPUESTAS</u>	<u>SUMAS ASEGURADAS EXPUESTAS</u>	<u>P R I M A S EMITIDAS DEVENGADAS</u>	<u>NUMERO DE SINIESTROS</u>	<u>MONTO</u>		<u>GASTOS DE AJUSTE</u>	<u>SALVAMENTOS</u>
					<u>SIN</u>	<u>CON</u>		

T O T A L E S

P A R C I A L E S

T O T A L

* MISMOS RIESGOS Y CLASE DE EQUIPO QUE EN CUADRO No. 3

CAPITULO III

**ANALISIS ESTADISTICO DE UN SISTEMA
DE INFORMACION.**

Parte importante en el diseño de un sistema de información es determinar el uso que se le va a dar al mismo, para que sea verdaderamente -- una herramienta de apoyo para un tomador de decisiones; se ha elegido el uso de tablas de contingencia dentro del análisis estadístico de información como punto fundamental, ya que antes de efectuar cualquier tipo de análisis es importante conocer la relación que guardan entre sí las variables que componen el sistema.

En el presente capítulo se desarrollan algunos de los aspectos básicos para efectuar el análisis de independencia de variables a través de tablas de contingencia, por lo que presentamos el concepto de tabla de contingencia y continuando con el desarrollo de "Tablas de una Entrada" que propiamente corresponde a bondad de ajuste, tablas de doble entrada, mencionando el caso en que tenemos muestras pequeñas, generalizando posteriormente con el caso de tablas de K entradas; en seguida presentamos un ejemplo ilustrativo, con información real del propio sistema; quedando estructurado como sigue:

3.1. Tablas de Contingencia.

3.2. Ejemplos de Tablas de Contingencia: doble y triple entrada.

3.3. Casos Especiales de Tablas de Contingencia y Proposición de la Aplicación de la Teoría.

3.1. TABLAS DE CONTINGENCIA.

Una tabla de contingencia es una clasificación múltiple, en donde a cada clasificación se le denomina casilla.

Las tablas de contingencia proporcionan una técnica para probar si -

las variables clasificadas por medio de la tabla son variables aleatorias independientes.

Plantearemos primero el caso de tablas de contingencia de una entrada, que aunque propiamente corresponden a bondad de ajuste nos servirá para presentar la teoría relacionada con tablas de contingencia - para el caso mas sencillo, continuaremos con tablas de contingencia de doble entrada e inmediatamente después plantearemos el caso para tablas de K entradas, generalizando de este modo todo lo relacionado a esta herramienta estadística.

3.1.1. "TABLAS DE UNA ENTRADA."

Como mencionamos anteriormente estas no son propiamente tablas de contingencia, sino que se refiere a pruebas de bondad de ajuste, pero ésta nos servirá para exponer la teoría desde el caso mas simple, haciendo una similitud con una tabla de contingencia y de este modo resultará mas sencillo el planteamiento y comprensión del caso de tablas de K-entradas.

Se cuenta con una tabla de un renglón y C columnas, teniendo en consecuencia $1 \times C$ celdas.

Suponemos que se realiza un experimento cuyos posibles resultados sólo pueden caer en alguna celda de la tabla.

Sea P_j la probabilidad de que el experimento producirá un resultado que caerá en la j-ésima celda. con $j=1, \dots, c$

Sea n_j , $j=1, \dots, c$ el número de veces que el experimento caerá en la j -ésima celda en un total de $n = \sum_{j=1}^c n_j$ experimentos.

Celda #	1	2	3	...	c	
frecuencia observada	n_1	n_2	n_3	...	n_c	$\sum_{j=1}^c n_j = n$
prob. de que el exp. caiga en - cada celda,	P_1	P_2	P_3	...	P_c	$\sum_{j=1}^c P_j = 1$

Pensando en cada celda por separado, debido a que cada vez que el experimento caiga en dicha celda podemos considerar un éxito para dicha celda y al final de efectuar las n repeticiones el experimento solo nos interesa el total de veces que el experimento cayó en la celda; podemos asociar una distribución binomial por cada celda y P_j corresponde entonces a la probabilidad de éxito en un sólo ensayo del experimento y - como se efectúa n veces; el número esperado de éxitos para la j -ésima - celda será la media de la correspondiente variable binomial, np_j .

o_j	n_1	n_2	n_3	...	n_c
e_j	np_1	np_2	np_3	...	np_c

O_j representa las frecuencias observadas.

e_j representa las frecuencias esperadas. en un total de n experimentos.

Queremos demostrar si un conjunto de resultados experimentales es compatible con los resultados esperados sobre la base de las probabilidades que fueron postuladas para las celdas.

La hipótesis queda planteada de la siguiente manera:

$$H_0: P_j = \pi_j, \quad j=1, 2, \dots, c$$

donde las π 's son los valores postulados de las probabilidades de las celdas. Esta es una hipótesis simple; sin embargo y debido a que ordinariamente no tenemos una hipótesis alternativa en mente, emplearemos aquí una prueba de razón de verosimilitudes.

La función de verosimilitud (o densidad conjunta de la muestra) - para una variable aleatoria discreta es la probabilidad de obtener los valores muestrales observados en el orden en el cual fueron obtenidos, por lo tanto:

$$L(\Theta, \underline{x}) = L(\Theta) = P_1^{n_1} P_2^{n_2} \dots P_c^{n_c}$$

\underline{x} representa los n valores muestrales que dieron lugar a las frecuencias n_1, n_2, \dots, n_c . Como $\sum_{j=1}^c P_j = 1$, solamente $c-1$ de los P 's son parámetros independientes; por lo tanto, para encontrar los estimadores máximo verosímiles de las P 's, reemplazaremos P_c por $1 - \sum_{j=1}^{c-1} P_j$ entonces:

$$L(\Theta) = P_1^{n_1} P_2^{n_2} \dots P_{c-1}^{n_{c-1}} \left(1 - \sum_{j=1}^{c-1} P_j\right)^{n_c}$$

$$\begin{aligned} \ln L(\Theta) &= n_1 \ln P_1 + n_2 \ln P_2 + \dots + n_{c-1} \ln P_{c-1} + n_c \ln \left(1 - \sum_{j=1}^{c-1} P_j\right) \\ &= \sum_{j=1}^{c-1} n_j \ln P_j + n_c \ln \left(1 - \sum_{j=1}^{c-1} P_j\right) \end{aligned}$$

$$\frac{\partial \ln L(\Theta)}{\partial P_r} = \frac{n_r}{P_r} - \frac{n_c}{1 - \sum_{j=1}^{c-1} P_j} = \frac{n_r}{P_r} - \frac{n_c}{P_c} \quad \text{donde } r=1, 2, \dots, c-1$$

$$\frac{n_r}{P_r} - \frac{n_c}{P_c} = 0 \Rightarrow P_r = \frac{\hat{P}_c}{n_c} n_r, \quad r=1, \dots, c \Rightarrow \sum_{r=1}^c P_r = \frac{\hat{P}_c}{n_c} \sum_{r=1}^c n_r$$

$$1 = \frac{\hat{P}_c}{n_c} n \Rightarrow \hat{P}_c = \frac{n_c}{n} \quad \therefore \hat{P}_j = \frac{n_j}{n}, \quad j=1, 2, \dots, c.$$

Ahora:

$$L(\omega) = \pi_1^{n_1} \pi_2^{n_2} \dots \pi_c^{n_c}$$

donde no hay parámetros desconocidos, por lo que no se necesitan estimadores aquí.

$$\therefore \lambda = \frac{L(\hat{\omega})}{L(\hat{\Theta})} = \frac{\pi_1^{N_1} \pi_2^{N_2} \dots \pi_c^{N_c}}{\left(\frac{N_1}{n}\right)^{N_1} \left(\frac{N_2}{n}\right)^{N_2} \dots \left(\frac{N_c}{n}\right)^{N_c}} = \left(\frac{n\pi_1}{N_1}\right)^{N_1} \left(\frac{n\pi_2}{N_2}\right)^{N_2} \dots \left(\frac{n\pi_c}{N_c}\right)^{N_c}$$

donde N_j denota la v.a. para la cual n_j es el valor observado.

Existen muchos casos, como el que se presenta aquí, para los cuales resulta casi imposible encontrar la distribución de:

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Theta})}$$

para éstos es aplicable la distribución asintótica de λ , la razón de verosimilitudes, basada en el siguiente teorema:

Teorema.- Sea X_1, X_2, \dots, X_n una muestra aleatoria con densidad conjunta $L(X_1, \dots, X_n, \underline{\Theta})$, donde $\underline{\Theta} = (\theta_1, \theta_2, \dots, \theta_c)$, que se supone satisface las condiciones generales de regularidad. Suponga que el espacio paramétrico (H) es c -dimensional.

En la prueba de hipótesis:

$$H_0: \theta_1 = \theta_1^0, \theta_2 = \theta_2^0, \dots, \theta_r = \theta_r^0, \theta_{r+1}, \dots, \theta_c;$$

donde $\theta_1^0, \theta_2^0, \dots, \theta_r^0$ son conocidos y $\theta_{r+1}, \dots, \theta_c$ no están especificados entonces $-2 \ln \lambda$ se destruye aproximadamente como una Ji-cuadrada con r grados de libertad ($-2 \ln \lambda \sim \chi_{(r)}^2$) cuando H_0 es cierta

y el tamaño de la muestra es grande. (Mood, pag. 440, 3a. edición).

Basandonos en el teorema anterior, utilizaremos la distribución asintótica. Ya que $-2 \ln \lambda$ posee una distribución asintótica Ji-cuadrada con grados de libertad igual a la diferencia en el número de parámetros independientes no especificados bajo H_0 y H_1 , tenemos que el número de grados de libertad aquí es $C-1$. Entonces, nuestra prueba se reduce a considerar

$$-2 \ln \lambda = -2 \sum_{j=1}^c N_j \ln \left(\frac{n \pi_j}{N_j} \right) \dots (*)$$

como una v.a. ji-cuadrada con $C-1$ grados de libertad.

Y la región crítica será entonces determinada por:

$$\lambda \leq \lambda_0 \iff \ln \lambda \leq \ln \lambda_0 \iff -2 \ln \lambda \geq -2 \ln \lambda_0 = c$$

$$\therefore C_0 = \{ \underline{x} \mid -2 \ln \lambda \geq c \} \rightarrow P[-2 \ln \lambda \geq c \mid H_0] = \alpha$$

En general, para probar si un conjunto de frecuencias observadas es compatible con un conjunto de frecuencias esperadas, se considera a la variable aleatoria:

$$\sum_{j=1}^c \frac{(N_j - e_j)^2}{e_j} \dots (**)$$

como una v.a. χ^2 con $c-1$ grados de libertad y se utiliza la cola derecha de la distribución χ^2 con $c-1$ grados de libertad.

Que la v.a. (**) tiene distribución χ^2 con $c-1$ grados de libertad.

se puede demostrar desarrollando en serie de Taylor los logaritmos en (*) de una manera apropiada, demostrando que solamente los términos principales (primeros) son importantes, y finalmente mostrar que cuando esto se hace se obtendrá:

$$-2 \ln \lambda = -2 \sum_{j=1}^c N_j \ln \left(\frac{n \pi_j}{N_j} \right) = \sum_{j=1}^c \frac{(N_j - n \pi_j)^2}{n \pi_j} + R_n$$

donde $P \left[\begin{matrix} R_n \rightarrow 0 \\ n \rightarrow \infty \end{matrix} \right] = 1$

Propiedad.- La prueba de la χ^2 también es aplicable cuando las probabilidades postuladas dependen de parámetros desconocidos, siempre y cuando los parámetros desconocidos sean reemplazados por sus estimadores máximo verosímiles y se resta un grado de libertad por cada uno de los parámetros estimados.

Hoel, Port y Stone pag. 95, 1971.

3.1.2. TABLAS DE DOBLE ENTRADA.

Se cuenta con una tabla de r renglones y c columnas, representada de la siguiente manera:

n_{11}	n_{12}	n_{13}	...	n_{1c-1}	n_{1c}	$n_{1.}$
n_{21}	n_{22}	n_{23}	...	n_{2c-1}	n_{2c}	$n_{2.}$
⋮	⋮	⋮	...	⋮	⋮	⋮
n_{r1}	n_{r2}	n_{r3}	...	n_{rc-1}	n_{rc}	$n_{r.}$
$n_{.1}$	$n_{.2}$	$n_{.3}$		$n_{.c-1}$	$n_{.c}$	$n_{..}$

Se quiere probar si dos variables que se han clasificado por medio de una tabla de doble entrada son variables aleatorias independientes.

Al igual que en las tablas de una entrada, se supone que se realiza un experimento cuyos posibles resultados solo pueden caer en alguna celda de la tabla.

Sea p_{ij} la probabilidad de que un individuo seleccionado al azar de la población bajo consideración caerá en la celda correspondiente al i -ésimo renglón y la j -ésima columna.

Sea $P_i = \sum_{j=1}^c p_{ij}$ la probabilidad de caer en el i -ésimo renglón.

Sea $P_j = \sum_{i=1}^r p_{ij}$ la probabilidad de caer en la j -ésima columna.

El planteamiento de la hipótesis para probar si las dos variables clasificadas son independientes queda de la siguiente manera:

$$H_0 : P_{ij} = P_i \cdot P_j \quad \begin{array}{l} i = 1, 2, \dots, r \\ j = 1, 2, \dots, c \end{array}$$

La función de verosimilitud es como antes pero con una notación un poco mas complicada:

$$L(\underline{x}, \textcircled{H}) = \prod_{i=1}^r \prod_{j=1}^c P_{ij}^{n_{ij}}$$

donde n_{ij} es la frecuencia observada en la celda para la cual P_{ij} es la probabilidad asignada a ella.

$n_{i.}$ representa los totales por renglón y $n_{.j}$ los totales por columna.

$$n_{i.} = \sum_{j=1}^c n_{ij} \quad , \quad n_{.j} = \sum_{i=1}^r n_{ij}$$

Cuando H_0 es verdadera:

$$L(\underline{x}; \Theta_0) = \prod_{i=1}^r \prod_{j=1}^c (P_i \cdot P_j)^{n_{ij}}$$

Como P_i y P_j son parámetros desconocidos, los valores esperados (para cada celda i, j)

$$e_{ij} = n P_{ij} = n P_i \cdot P_j$$

necesarios para la prueba de la χ^2 , son desconocidos.

Entonces los reemplazaremos por sus estimadores máximo verosímiles.

Para encontrar dos estimadores, necesitamos substituir en algún momento P_i y P_j , ya que:

$$\sum_{j=1}^c \sum_{i=1}^r P_{ij} = 1 \quad (\text{condición que se pide})$$

es equivalente a:

$$\sum_{i=1}^r P_i = 1 \quad \text{y} \quad \sum_{j=1}^c P_j = 1$$

entonces:

$$\begin{aligned} L(\underline{x}; \Theta_0) &= \prod_{i=1}^r \prod_{j=1}^c (P_i \cdot P_j)^{n_{ij}} \\ &= \prod_{i=1}^r \prod_{j=1}^c (P_i)^{n_{ij}} (P_j)^{n_{ij}} = \left(\prod_{i=1}^r (P_i)^{n_{i.}} \right) \left(\prod_{j=1}^c (P_j)^{n_{.j}} \right) \end{aligned}$$

$$\ln L(\underline{x}; \Theta_0) = \sum_{i=1}^r n_{i.} \ln P_i + \sum_{j=1}^c n_{.j} \ln P_j$$

Ahora: $\sum_{i=1}^r P_{i.} = 1 \Rightarrow P_{r.} = 1 - \sum_{i=1}^{r-1} P_{i.}$

y $\sum_{j=1}^c P_{.j} = 1 \Rightarrow P_{.c} = 1 - \sum_{j=1}^{c-1} P_{.j}$

$$\begin{aligned} \therefore \ln L(\underline{X}; \Theta_0) &= \sum_{i=1}^{r-1} n_{i.} \ln p_{i.} + n_{r.} \ln p_{r.} + \sum_{j=1}^{c-1} n_{.j} \ln p_{.j} + n_{.c} \ln p_{.c} \\ &= \sum_{i=1}^{r-1} n_{i.} \ln p_{i.} + n_{r.} \ln \left(1 - \sum_{i=1}^{r-1} p_{i.}\right) + \sum_{j=1}^{c-1} n_{.j} \ln p_{.j} + n_{.c} \ln \left(1 - \sum_{j=1}^{c-1} p_{.j}\right) \end{aligned}$$

Derivando e igualando a cero para maximizar tenemos:

$$\frac{\partial \ln L(\underline{X}; \Theta_0)}{\partial p_{q.}} = \frac{n_{q.}}{p_{q.}} + \frac{n_{r.} (-1)}{1 - \sum_{i=1}^{r-1} p_{i.}} = 0$$

$$\frac{\partial \ln L(\underline{X}; \Theta_0)}{\partial p_{.t}} = \frac{n_{.t}}{p_{.t}} + \frac{n_{.c} (-1)}{1 - \sum_{j=1}^{c-1} p_{.j}} = 0$$

Resolviendo las ecuaciones:

$$\frac{n_{q.}}{p_{q.}} = \frac{n_{r.}}{p_{r.}} \Rightarrow \frac{p_{r.}}{n_{r.}} n_{q.} = p_{q.} \Rightarrow \frac{p_{r.}}{n_{r.}} \sum_{q=1}^r n_{q.} = \sum_{q=1}^r p_{q.} = 1$$

$$\Rightarrow \hat{p}_{r.} \frac{n}{n_{r.}} = 1 \Rightarrow \hat{p}_{r.} = \frac{n_{r.}}{n}$$

pero así como se hizo para el r -ésimo término se pudo haber hecho para cualquier $p_{i.}$.

$$\therefore \hat{p}_{i.} = \frac{n_{i.}}{n}, \quad i = 1, \dots, r$$

$$\frac{n_{.t}}{P_{.t}} = \frac{n_{.c}}{1 - \sum_{j=1}^{c-1} P_{.j}} \Rightarrow \frac{n_{.t}}{P_{.t}} = \frac{n_{.c}}{P_{.c}} \Rightarrow \frac{P_{.c}}{n_{.c}} n_{.t} = P_{.t}$$

$$\Rightarrow \frac{P_{.c}}{n_{.c}} \sum_{t=1}^c n_{.t} = \sum_{t=1}^c P_{.t} = 1 \Rightarrow \frac{\hat{P}_{.c}}{n_{.c}} n = 1 \Rightarrow \hat{P}_{.c} = \frac{n_{.c}}{n}$$

de igual manera que en el caso anterior podemos concluir que:

$$\hat{P}_{.j} = \frac{n_{.j}}{n} \quad j = 1, 2, \dots, c$$

$$\therefore \underset{\text{A)}{}}{\text{Sup}} L(\underline{X}; \text{H}_0) = \prod_{i=1}^r \left(\frac{n_{i.}}{n}\right)^{n_{i.}} \prod_{j=1}^c \left(\frac{n_{.j}}{n}\right)^{n_{.j}}$$

Ahora: $\sum_{j=1}^c \sum_{i=1}^r P_{ij} = 1 \Rightarrow P_{rc} = 1 - \left(\sum_{j=1}^{c-1} \sum_{i=1}^{r-1} P_{ij} + \sum_{j=1}^{c-1} P_{rj} \right)$

$$L(\underline{X}; \text{H}_0) = \prod_{i=1}^r \prod_{j=1}^c P_{ij}^{n_{ij}}$$

$$\ln L(\underline{X}; \text{H}_0) = \sum_{i=1}^r \sum_{j=1}^c n_{ij} \ln P_{ij} = \sum_{j=1}^c \sum_{i=1}^r n_{ij} \ln P_{ij}$$

$$= \sum_{j=1}^c \sum_{i=1}^{r-1} n_{ij} \ln P_{ij} + \sum_{j=1}^{c-1} n_{rj} \ln P_{rj} + n_{rc} \ln P_{rc}$$

$$= \sum_{j=1}^c \sum_{i=1}^{r-1} n_{ij} \ln P_{ij} + \sum_{j=1}^{c-1} n_{rj} \ln P_{rj} + n_{rc} \ln \left(1 - \left[\sum_{j=1}^{c-1} \sum_{i=1}^{r-1} P_{ij} + \sum_{j=1}^{c-1} P_{rj} \right] \right)$$

derivando e igualando a cero para maximizar:

$$\frac{\partial \ln L(\underline{X}; \text{H}_0)}{\partial P_{st}} = \frac{n_{st}}{P_{st}} + \frac{n_{rc}(-1)}{1 - \left(\sum_{j=1}^{c-1} \sum_{i=1}^{r-1} P_{ij} + \sum_{j=1}^{c-1} P_{rj} \right)} = 0$$

$$\Rightarrow \frac{n_{st}}{P_{st}} = \frac{n_{rc}}{P_{rc}} \Rightarrow P_{st} = n_{st} \frac{P_{rc}}{n_{rc}} \Rightarrow \sum_{t=1}^c \sum_{s=1}^r P_{st} = \frac{P_{rc}}{n_{rc}} \sum_{t=1}^c \sum_{s=1}^r n_{st} \Rightarrow 1 = \frac{n}{n_{rc}} \hat{P}_{rc}$$

$$\Rightarrow \hat{p}_{rc} = \frac{nr_c}{n} \quad \therefore \hat{p}_{ij} = \frac{n_{ij}}{n} \quad \begin{array}{l} i = 1, \dots, r \\ j = 1, \dots, c \end{array}$$

$$\therefore \text{Sup}_{\Theta} L(\underline{X}; \Theta) = \prod_{i=1}^r \prod_{j=1}^c \left(\frac{n_{ij}}{n} \right)^{n_{ij}}$$

$$\therefore \lambda = \frac{\text{Sup}_{\Theta_0} L(\underline{X}; \Theta_0)}{\text{Sup}_{\Theta} L(\underline{X}; \Theta)} = \frac{\prod_{i=1}^r \left(\frac{n_{i.}}{n} \right)^{n_{i.}} \prod_{j=1}^c \left(\frac{n_{.j}}{n} \right)^{n_{.j}}}{\prod_{i=1}^r \prod_{j=1}^c \left(\frac{n_{ij}}{n} \right)^{n_{ij}}}$$

Donde nuevamente tenemos el problema de saber la distribución de λ , por lo que también usaremos el criterio asintótico $-2 \ln \lambda \sim \chi^2_{(r-1)(c-1)}$ o equivalentemente se usará la estadística $\sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$

$$\text{donde } e_{ij} = n \frac{n_{i.}}{n} \frac{n_{.j}}{n} = \frac{n_{i.} \cdot n_{.j}}{n}$$

entonces la estadística es: $\sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - \frac{n_{i.} \cdot n_{.j}}{n})^2}{\frac{n_{i.} \cdot n_{.j}}{n}} \sim \chi^2_{(r-1)(c-1)}$

y como se tiene la restricción $\sum_{i=1}^r \sum_{j=1}^c p_{ij} = 1$ esta es equivalente a pedir $\sum_{i=1}^r p_{i.} = 1$ y $\sum_{j=1}^c p_{.j} = 1$

de donde sabemos que hay $r + c - 2$ parámetros que deben considerarse independientes y son los que se sustituyen por su estimador máximo verosímil.

Por tanto los grados de libertad de la χ^2 son:

$$\nu = k - 1 = rc - 1$$

de donde hay que restar el número de parámetros que son sustituidos por su estimador máximo verosímil

$$\therefore \nu = rc - 1 - (r + c - 2) = (r-1)(c-1)$$

Por último podemos agregar que los parámetros no son realmente desconocidos, pueden estimarse y al aumentar el tamaño de la muestra, sus estimaciones tienden a los valores verdaderos. Cuando $n \rightarrow \infty$, se conocen los valores de los parámetros y entonces la distribución de λ llega a ser única. Es única porque se elige como valor verdadero del parámetro, el correspondiente a un punto determinado de ω , de modo que se da a las n_{ij} una distribución única.

3.1.2.1. PARA EL CASO DE MUESTRAS PEQUEÑAS.

Sea $f(n_{ij})$ una función de densidad de todas las n_{ij} por $g(n_i, n_j)$ la marginal de todas las n_i y n_j , y por $f(n_{ij} | n_i, n_j) = \frac{f(n_{ij})}{g(n_i, n_j)}$

la condicional de las n_{ij} , dados los totales marginales.

En H_0 , esta distribución condicional es independiente de los parámetros desconocidos, $n_{i\cdot}/n$ y $n_{\cdot j}/n$ constituyen un conjunto suficiente para $p_{i\cdot}$ y $p_{\cdot j}$, lo que nos permite la construcción de la prueba.

La distribución conjunta de las n_{ij} se reduce a la polinomial

$$f(n_{11}, n_{12}, \dots, n_{rs}) = \frac{n!}{\prod_{i,j} n_{ij}!} \prod_{i,j} p_{ij}^{n_{ij}} \quad \text{en } \Omega$$

y en ω (interesa la distribución de λ en H_0) lo que da lugar a :

$$f(n_{11}, n_{12}, \dots, n_{rs}) = \frac{n!}{\prod_{i,j} n_{ij}!} \left(\prod_{i=1}^r p_{i\cdot}^{n_i} \right) \left(\prod_{j=1}^c p_{\cdot j}^{n_j} \right) \dots (A)$$

Distribución Condicional:

1º) Se hallará la distribución de las n_i y n_j , sumando "A"

para todos los conjuntos $n_{ij} \rightarrow \sum_{i=1}^r n_{ij} = n_{.j}, \sum_{j=1}^s n_{ij} = n_{i.} \dots (B)$

en la suma solo interviene $\frac{1}{\prod n_{ij}!} \forall n_{ij} \in "B"$

$$(X_1 + \dots + X_r)^{n_1} (X_1 + X_2 + \dots + X_r)^{n_2} \dots (X_1 + \dots + X_r)^{n_s} = (X_1 + X_2 + X_3 + \dots + X_r)^n$$

en $(X_1 + \dots + X_r)^n$ el coeficiente $\prod X_i^{n_i}$ es

$$\frac{n!}{\prod n_i!}$$

en el 1er hay términos con coeficiente de la forma

$$\frac{n_1!}{\prod n_{i1}!} \frac{n_2!}{\prod n_{i2}!} \dots \frac{n_s!}{\prod n_{is}!} = \frac{\prod n_{.j}!}{\prod n_{ij}!} \dots (C)$$

donde n_{ij} es el exponente de X_i en la polinomial j -ésima.

En "C", las n_{ij} , satisfacen las condiciones "B".

$$\sum \frac{1}{\prod n_{ij}!} = \frac{n!}{\prod n_i! \prod n_{.j}!}$$

la suma que se necesita, ya que evidentemente existe un coeficiente y sólo uno de la forma "C".

luego la distribución de las n_i y n_j es, por tanto

$$g(n_i, n_j) = \frac{(n!)^2}{(\prod n_i!) (\prod n_{.j}!)^2}$$

lo que muestra que las n_i tienen distribuciones independientes de las n_j

La distribución condicional de las n_{ij} . dados lo totales marginales

se obtiene dividiendo:

$$f(n_{11}, n_{12}, \dots, n_{rs} | n_1, n_2, \dots, n_s) = \frac{n! (\prod_i p_i^{n_i}) (\prod_j p_j^{n_j}) (\prod_i n_i!) (\prod_j n_j!)}{\prod_j n_j! (n!)^2 (\prod_i p_i^{n_i}) (\prod_j p_j^{n_j})} = \frac{(\prod_i n_i!) (\prod_j n_j!)}{n! \prod_j n_j!} \dots (D)$$

la que no tiene parámetros desconocidos y por tanto los indicadores son suficientes.

Consideramos la distribución $\mu(\lambda; \theta)$ con θ desconocido, si $\hat{\theta}$ tiene un estimador suficiente $\hat{\theta}$, la distribución conjunta estará dada - por

$$v(\lambda, \hat{\theta}; \theta) = v_1(\lambda | \hat{\theta}) v_2(\hat{\theta}; \theta)$$

y en la distribución condicional de λ , dado $\hat{\theta}$, no intervendrá θ usando la distribución condicional, podremos hallar un número $A(\hat{\theta}) \neq \hat{\theta} \rightarrow$

$$\int_0^{A(\hat{\theta})} v_1(\lambda | \hat{\theta}) d\lambda = 0.05$$

En el plano $\lambda \hat{\theta}$, la curva $\lambda = A(\hat{\theta})$ determina, juntamente con la recta $\lambda = 0$, una región \mathcal{R}

$$\begin{aligned} P[(\lambda, \hat{\theta}) \in \mathcal{R}] &= \int_{-\infty}^{\infty} \int_0^{A(\hat{\theta})} v(\lambda, \hat{\theta}; \theta) d\lambda d\theta \\ &= \int_{-\infty}^{\infty} \left[\int_0^{A(\hat{\theta})} v_1(\lambda | \hat{\theta}) d\lambda \right] v_2(\hat{\theta}; \theta) d\theta \\ &= \int_{-\infty}^{\infty} 0.05 v_2(\hat{\theta}; \theta) d\hat{\theta} = 0.05 \end{aligned}$$

probarémos la hipótesis usando $\hat{\theta}$ en unión de λ .

La región crítica es una región del plano, en vez del intervalo

$$0 < \lambda < A$$

Observamos $\hat{\theta}$ y probamos λ mediante un intervalo $0 < \lambda < A(\hat{\theta})$, utilizando la distribución condicional de λ , dado $\hat{\theta}$.

Si la distribución ("D") se substituye por su aproximación normal multivariante, puede demostrarse que

$$u = \sum \frac{[n_{ij} - (n_i n_j / n)]^2}{n_i n_j / n} \approx \chi^2 \quad \text{con } (r-1)(s-1) \text{ grados de libertad.}$$

3.1.3. TABLAS DE K ENTRADAS.

La extensión del concepto de tablas de contingencia es clara si seguimos el desarrollo de las tablas de doble entrada.

Se cuenta con una tabla de $r_1 \times r_2 \times \dots \times r_k$ elementos. Se quiere probar si las k variables que se han clasificado por medio de la tabla de k entradas son variables aleatorias independientes (independencia total).

Al igual que en las tablas de doble entrada, se supone que se realiza un experimento cuyos posibles resultados sólo pueden caer en alguna celda de la tabla.

Sea $P_{i_1 i_2 \dots i_k}$ las probabilidades asociadas con las casillas individuales.

Sea $n_{i_1 i_2 \dots i_k}$ la frecuencia observada en cada casilla.

Únicamente nos concretaremos a probar la independencia total, o sea que las k variables que se han clasificado por medio de la tabla de k

entradas son variables aleatorias independientes entre sí, lo que queda formalmente planteado de la siguiente manera:

$$H_0: P_{i_1 i_2 \dots i_k} = P_{i_1} \dots P_{i_2} \dots P_{i_3} \dots \underbrace{P_{i_4} \dots i_k}_k$$

donde $i_1 = 1, 2, \dots, r_1$
 $i_2 = 1, 2, \dots, r_2$
 \vdots
 $i_k = 1, 2, \dots, r_k$

La función de verosimilitud es como antes pero con una notación un poco mas complicada:

$$L(\underline{X}; \Theta) = \prod_{i_1=1}^{r_1} \prod_{i_2=1}^{r_2} \dots \prod_{i_k=1}^{r_k} P_{i_1 i_2 \dots i_k}^{n_{i_1 i_2 \dots i_k}}$$

donde $n_{i_1 i_2 \dots i_k}$ es la frecuencia observada en la celda para la cual $P_{i_1 i_2 \dots i_k}$ es la probabilidad asignada a ella.

Sea $n_{i_1 \dots}, n_{i_1 i_2 \dots}, n_{i_1 i_2 i_3 \dots}, \dots, n_{i_1 \dots i_k}$ los totales por arreglo unidimensional.

$$n_{i_1 \dots} = \sum_{i_2=1}^{r_2} \sum_{i_3=1}^{r_3} \dots \sum_{i_k=1}^{r_k} n_{i_1 i_2 \dots i_k}$$

$$n_{i_1 i_2 \dots} = \sum_{i_1=1}^{r_1} \sum_{i_3=1}^{r_3} \dots \sum_{i_k=1}^{r_k} n_{i_1 i_2 \dots i_k}$$

$$\vdots$$

$$n_{i_1 \dots i_k} = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \dots \sum_{i_{k-1}=1}^{r_{k-1}} n_{i_1 i_2 \dots i_k}$$

En adelante usaremos la siguiente notación para cada :

$$n_{i_1 \dots} = m_{i_1}$$

$$n_{i_1 i_2 \dots} = m_{i_2}$$

$$\vdots$$

$$n_{i_1 \dots i_k} = m_{i_k}$$

cuando H_0 es verdadera:

$$L(\underline{X}; H_0) = \prod_{i=1}^{r_1} \prod_{i_2=1}^{r_2} \dots \prod_{i_k=1}^{r_k} (p_{i_1} \dots p_{i_2} \dots p_{i_3} \dots p_{i_4} \dots p_{i_k})$$

como $p_{i_1}, p_{i_2}, \dots, p_{i_3}, \dots, p_{i_k}$ son parámetros desconocidos, los valores esperados (para cada celda i_1, i_2, \dots, i_k), necesarios para la prueba de la χ^2 son los siguientes:

$$E_{i_1, i_2, \dots, i_k} = n p_{i_1, i_2, \dots, i_k} = n p_{i_1} \dots p_{i_2} \dots p_{i_3} \dots p_{i_4} \dots p_{i_k}$$

Entonces los reemplazamos por sus estimadores máximo verosímiles.

Para encontrar los estimadores, necesitamos substituir en algún momento,

$$p_{i_1}, p_{i_2}, \dots, p_{i_k}$$

, ya que:

$$\sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \dots \sum_{i_k=1}^{r_k} p_{i_1, i_2, \dots, i_k} = 1$$

es equivalente a $\sum_{i_1=1}^{r_1} p_{i_1} \dots = 1$

$$\sum_{i_2=1}^{r_2} p_{i_2} \dots = 1$$

⋮

$$\sum_{i_k=1}^{r_k} p_{i_1} \dots i_k = 1$$

En adelante usaremos la siguiente notación para cada p_i

$$p_{i_1} \dots = q_{i_1}$$

$$p_{i_1} i_2 \dots = q_{i_1 i_2}$$

⋮

$$p_{i_1} \dots i_k = q_{i_1 i_k}$$

entonces:

$$L(\underline{x}; \Theta_0) = \prod_{i_1=1}^{r_1} \prod_{i_2=1}^{r_2} \dots \prod_{i_k=1}^{r_k} (q_{i_1} q_{i_2} \dots q_{i_k})^{m_{i_1 i_2 \dots i_k}}$$

$$= \prod_{i_1=1}^{r_1} \prod_{i_2=1}^{r_2} \dots \prod_{i_k=1}^{r_k} (q_{i_1})^{m_{i_1 i_2 \dots i_k}} (q_{i_2})^{m_{i_1 i_2 \dots i_k}} \dots (q_{i_k})^{m_{i_1 i_2 \dots i_k}}$$

$$= \left(\prod_{i_1=1}^{r_1} q_{i_1}^{m_{i_1}} \right) \left(\prod_{i_2=1}^{r_2} q_{i_2}^{m_{i_2}} \right) \dots \left(\prod_{i_k=1}^{r_k} q_{i_k}^{m_{i_k}} \right)$$

$$\ln L(\underline{x}; \Theta_0) = \sum_{i_1=1}^{r_1} m_{i_1} \ln q_{i_1} + \sum_{i_2=1}^{r_2} m_{i_2} \ln q_{i_2} + \dots + \sum_{i_k=1}^{r_k} m_{i_k} \ln q_{i_k}$$

Ahora: $\sum_{i_1=1}^{r_1} q_{i_1} = 1 \Rightarrow \frac{q_{i_1}}{r_1} = 1 - \sum_{i_2=1}^{r_2-1} q_{i_2}$

$$\sum_{i_2=1}^{r_2} q_{i_2} = 1 \Rightarrow q_{r_2} = 1 - \sum_{i_2=1}^{r_2-1} q_{i_2}$$

$$\dots$$

$$\sum_{i_k=1}^{r_k} q_{i_k} = 1 \Rightarrow q_{r_k} = 1 - \sum_{i_k=1}^{r_k-1} q_{i_k}$$

$$\therefore \ln L(\underline{X}; \theta_0) = \sum_{i_1=1}^{r_1-1} m_{i_1} \ln q_{i_1} + m_{r_1} \ln q_{r_1} + \sum_{i_2=1}^{r_2-1} m_{i_2} \ln q_{i_2} + m_{r_2} \ln q_{r_2} + \dots +$$

$$\sum_{i_k=1}^{r_k-1} m_{i_k} \ln q_{i_k} + m_{r_k} \ln q_{r_k}$$

$$= \sum_{i_1=1}^{r_1-1} m_{i_1} \ln q_{i_1} + m_{r_1} \ln \left(1 - \sum_{i_2=1}^{r_2-1} q_{i_2} \right) + \sum_{i_2=1}^{r_2-1} m_{i_2} \ln q_{i_2} +$$

$$m_{r_2} \ln \left(1 - \sum_{i_2=1}^{r_2-1} q_{i_2} \right) + \dots + \sum_{i_k=1}^{r_k-1} m_{i_k} \ln q_{i_k} + m_{r_k} \ln \left(1 - \sum_{i_k=1}^{r_k-1} q_{i_k} \right)$$

Derivando con respecto a cada q_{i_s} , $s=1,2,\dots,k$ e igualando a cero para maximizar tenemos:

$$\frac{\partial \ln L(\underline{X}; \theta_0)}{\partial q_{i_1}} = \frac{m_{i_1}}{q_{i_1}} + \frac{m_{r_1}(-1)}{1 - \sum_{i_2=1}^{r_2-1} q_{i_2}} = 0$$

$$\frac{\partial \ln L(\underline{X}; \theta_0)}{\partial q_{i_2}} = \frac{m_{i_2}}{q_{i_2}} + \frac{m_{r_2}(-1)}{1 - \sum_{i_2=1}^{r_2-1} q_{i_2}} = 0$$

$$\dots$$

$$\frac{\partial \ln L(\underline{X}; \theta_0)}{\partial q_{i_k}} = \frac{m_{i_k}}{q_{i_k}} + \frac{m_{r_k}(-1)}{1 - \sum_{i_k=1}^{r_k-1} q_{i_k}} = 0$$

Resolviendo las k ecuaciones:

$$\frac{m_{i_1}}{q_{i_1}} = \frac{m_{r_1}}{q_{r_1}} \Rightarrow q_{i_1} = \frac{q_{r_1}}{m_{r_1}} m_{i_1} \Rightarrow \sum_{i_1=1}^{r_1} q_{i_1} = \frac{q_{r_1}}{m_{r_1}} \sum_{i_1=1}^{r_1} m_{i_1} = 1$$

$$\Rightarrow \hat{q}_{r_1} \frac{m}{m_{r_1}} = 1 \Rightarrow \hat{q}_{r_1} = \frac{m_{r_1}}{m}$$

pero así como se hizo para el r_1 -ésimo se pudo haber hecho para cualquier q_{i_1}

$$\therefore \hat{q}_{i_1} = \frac{m_{i_1}}{m} \quad i_1 = 1, 2, \dots, r_1$$

de igual manera que en este caso concluimos lo siguiente para las demás i_2, \dots, i_k

$$\hat{q}_{i_2} = \frac{m_{i_2}}{m} \quad i_2 = 1, 2, \dots, r_2$$

⋮

$$\hat{q}_{i_k} = \frac{m_{i_k}}{m} \quad i_k = 1, 2, \dots, r_k$$

$$\therefore \text{Sup } h(\underline{x}; \mathbb{A}) = \prod_{i_1=1}^{r_1} \left(\frac{m_{i_1}}{m}\right)^{m_{i_1}} \prod_{i_2=1}^{r_2} \left(\frac{m_{i_2}}{m}\right)^{m_{i_2}} \dots \prod_{i_k=1}^{r_k} \left(\frac{m_{i_k}}{m}\right)^{m_{i_k}}$$

Ahora:

$$\sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \dots \sum_{i_k=1}^{r_k} q_{i_1 i_2 \dots i_k} = 1 = \sum_{i_{k-1}=1}^{r_{k-1}} \sum_{i_{k-2}=1}^{r_{k-2}} \dots \sum_{i_2=1}^{r_2} \sum_{i_1=1}^{r_1} q_{i_1 i_2 \dots i_k}$$

$$\Rightarrow q_{r_1 r_2 \dots r_k} = 1 - \left(\sum_{i_{k-1}=1}^{r_{k-1}} \sum_{i_{k-2}=1}^{r_{k-2}} \dots \sum_{i_1=1}^{r_1} q_{i_1 i_2 \dots i_k} + \sum_{i_{k-1}=1}^{r_{k-1}} \sum_{i_{k-2}=1}^{r_{k-2}} \dots \sum_{i_2=1}^{r_2} q_{r_1 i_2 \dots i_k} + \dots + \sum_{i_{k-1}=1}^{r_{k-1}} q_{r_1 r_2 \dots r_{k-1} i_k} \right)$$

$$= 1 - A$$

$$L(\underline{x}, \textcircled{H}) = \prod_{i_1=1}^{r_1} \prod_{i_2=1}^{r_2} \dots \prod_{i_k=1}^{r_k} q_{i_1 i_2 \dots i_k}^{m_{i_1 i_2 \dots i_k}}$$

$$\ln L(\underline{x}, \textcircled{H}) = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \dots \sum_{i_k=1}^{r_k} m_{i_1 i_2 \dots i_k} \ln q_{i_1 i_2 \dots i_k}$$

$$= \sum_{i_k=1}^{r_k} \sum_{i_{k-1}=1}^{r_{k-1}} \dots \sum_{i_1=1}^{r_1-1} m_{i_1 i_2 \dots i_k} \ln q_{i_1 i_2 \dots i_k} + \sum_{i_k=1}^{r_k} \sum_{i_{k-1}=1}^{r_{k-1}} \dots \sum_{i_2=1}^{r_2-1} m_{i_1 i_2 \dots i_k}$$

$$\ln q_{i_1 i_2 \dots i_k} + \dots + \sum_{i_2=1}^{r_2-1} m_{r_1 r_2 \dots r_{k-1} i_k} \ln q_{r_1 r_2 \dots r_{k-1} i_k} + m_{r_1 r_2 \dots r_k} \ln(1-A)$$

derivando e igualando a zero para maximizar:

$$\frac{\partial \ln L(\underline{x}; \textcircled{H})}{\partial q_{i_1 i_2 \dots i_k}} = \frac{m_{i_1 i_2 \dots i_k}}{q_{i_1 i_2 \dots i_k}} + \frac{m_{r_1 r_2 \dots r_k} (-1)}{1-A} = 0$$

$$\Rightarrow \frac{m_{i_1 i_2 \dots i_k}}{q_{i_1 i_2 \dots i_k}} = \frac{m_{r_1 r_2 \dots r_k}}{q_{r_1 r_2 \dots r_k}} \Rightarrow q_{i_1 i_2 \dots i_k} = \frac{q_{r_1 r_2 \dots r_k}}{m_{r_1 r_2 \dots r_k}} m_{i_1 i_2 \dots i_k}$$

$$\Rightarrow \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \dots \sum_{i_k=1}^{r_k} q_{i_1 i_2 \dots i_k} = \frac{q_{r_1 r_2 \dots r_k}}{m_{r_1 r_2 \dots r_k}} \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \dots \sum_{i_k=1}^{r_k} m_{i_1 i_2 \dots i_k}$$

$$\Rightarrow 1 = \frac{\hat{q}_{r_1 r_2 \dots r_k}}{m_{r_1 r_2 \dots r_k}} m \Rightarrow \hat{q}_{r_1 r_2 \dots r_k} = \frac{m_{r_1 r_2 \dots r_k}}{m}$$

$$\therefore \hat{q}_{i_1 i_2 \dots i_k} = \frac{m_{i_1 i_2 \dots i_k}}{m} \quad \begin{array}{l} i_1 = 1, \dots, r_1 \\ i_2 = 1, \dots, r_2 \\ \vdots \\ i_k = 1, \dots, r_k \end{array}$$

$$\therefore \text{Sup}_{\textcircled{H}} L(\underline{x}; \textcircled{H}) = \prod_{i_1=1}^{r_1} \prod_{i_2=1}^{r_2} \dots \prod_{i_k=1}^{r_k} \left(\frac{m_{i_1 i_2 \dots i_k}}{m} \right)^{m_{i_1 i_2 \dots i_k}}$$

$$\therefore \lambda = \frac{\sup_{\mathbb{H}_0} L(\underline{X}; \mathbb{H}_0)}{\sup_{\mathbb{H}} L(\underline{X}; \mathbb{H})} = \frac{\prod_{i_1=1}^{r_1} \left(\frac{m_{i_1}}{m}\right)^{m_{i_1} r_2} \prod_{i_2=1}^{r_2} \left(\frac{m_{i_2}}{m}\right)^{m_{i_2}} \dots \prod_{i_k=1}^{r_k} \left(\frac{m_{i_k}}{m}\right)^{m_{i_k}}}{\prod_{i_1=1}^{r_1} \prod_{i_2=1}^{r_2} \dots \prod_{i_k=1}^{r_k} \left(\frac{m_{i_1, i_2, \dots, i_k}}{m}\right)^{m_{i_1, i_2, \dots, i_k}}}$$

Donde nuevamente tenemos el problema de saber la distribución de λ , por lo que también usaremos el criterio asintótico $-2 \ln \lambda \sim \chi^2(\nu)$ ó equivalentemente se usará la estadística $\sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \dots \sum_{i_k=1}^{r_k} \frac{(O_{i_1, i_2, \dots, i_k} - E_{i_1, i_2, \dots, i_k})^2}{E_{i_1, i_2, \dots, i_k}}$

donde

$$E_{i_1, i_2, \dots, i_k} = \frac{m_{i_1}}{m} \frac{m_{i_2}}{m} \dots \frac{m_{i_k}}{m} = \frac{m_{i_1, i_2, \dots, i_k}}{m^{k-1}}$$

entonces la estadística es:

$$\sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \dots \sum_{i_k=1}^{r_k} \frac{(O_{i_1, i_2, \dots, i_k} - \frac{m_{i_1, i_2, \dots, i_k}}{m^{k-1}})^2}{\frac{m_{i_1, i_2, \dots, i_k}}{m^{k-1}}} \sim \chi^2(\nu)$$

y como tenemos la restricción

$$\sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \dots \sum_{i_k=1}^{r_k} q_{i_1, i_2, \dots, i_k} = 1$$

esta es equivalente a pedir:

$$\sum_{i_1=1}^{r_1} q_{i_1} = 1, \sum_{i_2=1}^{r_2} q_{i_2} = 1, \dots, \sum_{i_k=1}^{r_k} q_{i_k} = 1$$

luego los grados de libertad para la $\chi^2(\nu)$ quedan determinados como sigue:

son $K-1$ el # de parámetros estimados.

$$\nu = r_1 \times r_2 \times \dots \times r_k - 1 - \# \text{ parámetros estimados}$$

$$\# \text{ p.e.} = (r_1 - 1) + (r_2 - 1) + \dots + (r_k - 1)$$

$$k - 1 - \# \text{ p.e.} = r_1 \times r_2 \times \dots \times r_k - 1 - [r_1 - 1 + r_2 - 1 + \dots + r_k - 1]$$

$$= r_1 \times r_2 \times \dots \times r_k - 1 - \left[\sum_{i=1}^k r_i - k \right]$$

$$= r_1 \times r_2 \times \dots \times r_k - \sum_{i=1}^k r_i + k - 1 = \nu$$

$$\therefore \chi^2_{(r_1 \times r_2 \times \dots \times r_k - \sum_{i=1}^k r_i + k - 1)}$$

3.2. EJEMPLOS DE TABLAS DE CONTINGENCIA: doble y triple entrada.

A continuación presentamos tres ejemplos que se pretende sean representativos, sobre el sistema de información que se tiene como ejemplo, con los datos reales del mismo:

Ejemplo No. 1

Ho : Se quiere demostrar si existe alguna relación de dependencia entre el tipo de equipo con la severidad de la ocurrencia de siniestros.

Se cuenta con una tabla con las siguientes frecuencias observadas:

SEVERIDAD DE SINIESTROS	TIPO DE EQUIPO								
	100	200	300	400	500	600	900	1000	
NO SINIESTROS	577	108	23	1	13	6	0	7	$n_{1.} = 735$ $n_{2.} = 10$ $n_{3.} = 3$ <hr/> $n_{..} = 748$
PARCIALMENTE	7	0	2	0	0	0	0	1	
TOTALMENTE	1	0	2	0	0	0	0	0	
	585	108	27	1	13	6	0	8	
	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.4}$	$n_{.5}$	$n_{.6}$	$n_{.7}$	$n_{.8}$	

748 = total de observaciones.

sus variables son: i, j.

i : identifica la severidad de la siniestralidad y toma valores de 1 a 3.

j : identifica el tipo de equipo que fue o no siniestrado y toma - valores del 1 al 8.

los totales por renglón y por columna se obtienen de la siguiente forma:

$$n_{1.} = \sum_{j=1}^8 n_{1j} = 735$$

$$n_{2.} = \sum_{j=1}^8 n_{2j} = 10$$

$$n_{3.} = \sum_{j=1}^8 n_{3j} = 3$$

$$n_{.1} = \sum_{i=1}^3 n_{i1} = 585$$

$$n_{.2} = \sum_{i=1}^3 n_{i2} = 108$$

$$n_{.3} = \sum_{i=1}^3 n_{i3} = 27$$

$$n_{.4} = \sum_{i=1}^3 n_{i4} = 1$$

$$n_{.5} = \sum_{i=1}^3 n_{i5} = 13$$

$$n_{.6} = \sum_{i=1}^3 n_{i6} = 6$$

$$n_{.7} = \sum_{i=1}^3 n_{i7} = 0$$

$$n_{.8} = \sum_{i=1}^3 n_{i8} = 8$$

Se obtienen las frecuencias esperadas con la siguiente expresión:

$$e_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

para luego determinar la estadística de prueba.

la tabla de frecuencias esperadas queda como sigue:

SEVERIDAD DE SINIESTROS	100	200	300	400	500	600	900	1000
NO SINIESTROS	575	106	27	1	13	6	0	8
PARCIALMENTE	8	1	0	0	0	0	0	0
TOTALMENTE	2	0	0	0	0	0	0	0

luego la estadística de prueba se determina como sigue:

$$\sum_{i=1}^3 \sum_{j=1}^8 \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \sim \chi^2(\nu)$$

con $\alpha = 0.01$ $4.3873 \sim \chi^2(\nu)$

$$\chi^2_{(3-1)(8-1)} = \chi^2_{(14)} = 29.1 \text{ es el .99 del cuantil de una dist.}$$

luego como $4.39 < 29.1$ entonces \therefore no se rechaza la hipótesis H_0 .

Ejemplo No.2

Ho : Se quiere demostrar la relación que existe entre el deducible aplicado con que el equipo se siniestre o no.

Se cuenta con una tabla con las siguientes frecuencias observadas:

DEDUCIBLE

AFECTION DEL EQUIPO

	SINIESTRADO	NO SINIESTRADO	
0.10	0	4	$n_{1.} = 4$
0.15	0	11	$n_{2.} = 11$
0.20	1	85	$n_{3.} = 86$
0.25	0	72	$n_{4.} = 72$
0.30	2	92	$n_{5.} = 94$
0.35	0	178	$n_{6.} = 178$
0.40	0	121	$n_{7.} = 121$
0.45	0	0	$n_{8.} = 0$
0.50	3	118	$n_{9.} = 121$
1.00	1	88	$n_{10.} = 89$
2.00	0	8	$n_{11.} = 8$
3.00	4	68	$n_{12.} = 72$
10.00	2	6	$n_{13.} = 8$
	$n_{.1} = 13$	$n_{.2} = 851$	

864 = total de observaciones

sus variables son: i, j .

i : identifica el deducible aplicado y toma valores del 1 al 13.

j : identifica la afectabilidad del equipo y toma valores de 1 al 2.

los totales por renglón. como un ejemplo anterior se determina aplicando la expresión

$$n_{i.} = \sum_{j=1}^2 n_{ij} \quad \text{donde } (i=1, \dots, 13), j=1, 2$$

y los totales por columna se obtienen aplicando la siguiente expresión:

$$n_{.j} = \sum_{i=1}^{13} n_{ij} \quad i=1,2,\dots,13, \quad j=1,2$$

luego obtenemos las frecuencias esperadas con la siguiente expresión:

$$e_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

la tabla de frecuencias esperadas queda como sigue:

DEDUCIBLE	AFECTACION DEL EQUIPO	
	SINIESTRADO	NO SINIESTRADO
0.10	0	4
0.15	0	11
0.20	1	85
0.25	1	71
0.30	1	93
0.35	3	175
0.40	2	119
0.45	0	0
0.50	2	119
1.00	1	88
2.00	0	8
3.00	1	71
10.00	0	8

la estadística de prueba se determina como sigue:

$$\sum_{i=1}^{13} \sum_{j=1}^2 \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \sim \chi^2_{(v)}$$

$$17.2347 \sim \chi^2_{(11)}$$

con $\alpha = 0.01$

$\chi^2_{(2-1)(13-1)} = \chi^2_{(12)} = 24.7$ es el .99 del cuartil de una distribución $\chi^2_{(12)}$, como $24.7 > 17.23$ entonces \therefore no se rechaza la hipótesis H_0 .

Ejemplo No.3

Ho: Se quiere determinar si existe alguna relación de dependencia entre el tipo de equipo con el deducible aplicado y la siniestralidad ocurrida.

CLASE DE EQUIPO

DEDUCIBLE	100			200			300			400			
	NO SINIESTRO	SI SINIESTRADO	NO	SI	NO	SI	NO	SI	NO	SI	NO	SI	
0.10-0.30	N 111	N 112	N 121	N 122	-	N 131	23	N 132	9	N 141	1	N 142	1
0.35-0.50	N 211	N 212	N 221	N 222	-	N 231	1	N 232	1	N 241	-	N 242	-
1.00-3.00	N 311	N 312	N 321	N 322	-	N 331	13	N 332	11	N 341	-	N 342	-
10.00	N 411	N 412	N 421	N 422	-	N 431	-	N 432	-	N 441	-	N 442	-
TOTAL	634	16	110	-	-	37	21	1	1				

DEDUCIBLE	500			600			900			1000			
	NO	SI	NO	SI	NO	SI	NO	SI	NO	SI	NO	SI	
0.10-0.30	N 151	N 152	N 161	N 162	-	N 171	2	N 172	-	N 181	1	N 182	-
0.35-0.50	N 251	N 252	N 261	N 262	-	N 271	17	N 272	-	N 281	3	N 282	-
1.00-3.00	N 351	N 352	N 361	N 362	-	N 371	2	N 372	-	N 381	4	N 382	-
10.00	N 451	N 452	N 461	N 462	-	N 471	-	N 472	-	N 481	-	N 482	-
TOTAL	10	-	6	-	-	21	8	-	-				

Tenemos una tabla de 4X8X2 elementos queremos determinar la independencia total de la misma:

T O T A L

DEDUCIBLE	<u>NO SINIESTRADA</u>	<u>SI SINIESTRADA</u>	
0.10-0.30	252	15	$m = 866$ Observaciones
0.35-0.50	414	7	
1.00-3.00	154	15	sus variables son: i, j, k
10.00	7	2	
TOTAL	827	39	

i : es la que identifica el deducible aplicado y toma valores de 1 a 4

j : identifica el tipo de equipo y toma valores del 1 al 8

k : identifica el que una clase de equipo haya sido o no siniestrado y toma el valor 1 y 2.

Se obtendrán los totales por arreglo unidimensional para luego determinar las frecuencias esperadas en cada casilla.

$$\begin{aligned}
 n_{1..} &= \sum_{j=1}^8 \sum_{k=1}^2 n_{1jk} = 252 + 15 = 267 \\
 n_{2..} &= \sum_{j=1}^8 \sum_{k=1}^2 n_{2jk} = 414 + 7 = 421 \\
 n_{3..} &= \sum_{j=1}^8 \sum_{k=1}^2 n_{3jk} = 154 + 15 = 169 \\
 n_{4..} &= \sum_{j=1}^8 \sum_{k=1}^2 n_{4jk} = 7 + 2 = 9 \\
 n_{.1.} &= \sum_{i=1}^4 \sum_{k=1}^2 n_{i.k} = 634 + 16 = 650
 \end{aligned}$$

$$\begin{aligned}n_{.2.} &= \sum_{i=1}^4 \sum_{k=1}^2 n_{izk} = 110 + 0 = 110 \\n_{.3.} &= \sum_{i=1}^4 \sum_{k=1}^2 n_{izk} = 37 + 21 = 58 \\n_{.4.} &= \sum_{i=1}^4 \sum_{k=1}^2 n_{izk} = 1 + 1 = 2 \\n_{.5.} &= \sum_{i=1}^4 \sum_{k=1}^2 n_{isk} = 10 + 0 = 10 \\n_{.6.} &= \sum_{i=1}^4 \sum_{k=1}^2 n_{i6k} = 6 + 0 = 6 \\n_{.7.} &= \sum_{i=1}^4 \sum_{k=1}^2 n_{izk} = 21 + 0 = 21 \\n_{.8.} &= \sum_{i=1}^4 \sum_{k=1}^2 n_{i8k} = 8 + 1 = 9 \\n_{..1} &= \sum_{i=1}^4 \sum_{j=1}^8 n_{ij1} = 634 + 110 + 37 + 1 + 10 + 6 + 21 + 8 + 728 \\n_{..2} &= \sum_{i=1}^4 \sum_{j=1}^8 n_{ij2} = 16 + 0 + 21 + 1 + 0 + 0 + 0 + 1 = 39\end{aligned}$$

Se obtendrán las frecuencias esperadas para después obtener la estadística de prueba.

Sabemos que
$$e_{ijk} = \frac{m}{m} \frac{m_i}{m} \frac{m_j}{m} \frac{m_k}{m}$$
$$= \frac{m_i m_j m_k}{m^2}$$

luego la estadística de prueba queda como sigue:

$$332.65 = \sum_{i=1}^4 \sum_{j=1}^8 \sum_{k=1}^2 \frac{(n_{ijk} - e_{ijk})^2}{e_{ijk}} \sim \chi^2_{(u)}$$

obtendremos ν los grados de libertad de la χ^2

$$\nu = i \times j \times k - (i + j + k) + k - 1$$

sustituyendo valores se tiene

$$\nu = (4)(8)(2) - (4 + 8 + 2) + 2 - 1 = 64 - 14 + 2 = 52$$

$$\therefore \nu = 52$$

con $\alpha = 0.01$

$\chi^2_{(52)} = 77.81$ es el cuantil de una distribución $\chi^2_{(52)}$ como $77.81 < 332.65$ entonces por lo tanto se rechaza H_0 .

3.3. CASOS ESPECIALES DE TABLAS DE CONTINGENCIA Y PROPOSICION DE LA APLICACION DE LA TEORIA.

Como se puede observar en los tres ejemplos presentados existen tablas que presentan dificultad para su manejo por casos como los siguientes:

- a) Clasificación incompleta de datos.
- b) Clasificación de datos en que el valor de las frecuencias en sus celdas es cero.
- c) Dificultad para obtener las frecuencias esperadas por ser tablas de mas de dos entradas.

Son estas algunas de las dificultades que se presentan en el manejo de tablas de contingencia, por lo que se tuvo la necesidad de investigar la existencia de métodos factibles para corregir las deficiencias que se presentan en las tablas de contingencia que se obtienen al trabajar con datos reales.

A continuación mencionaremos brevemente algunos métodos propuestos para tablas con las características mencionadas anteriormente:

3.3.1. TABLAS DE CONTINGENCIA CON CLASIFICACION INCOMPLETA DE DATOS, EN EL CASO DE DOBLE ENTRADA.

Con objeto de ser representativos y no crear un ambiente árido en la exposición de la teoría, únicamente presentaremos lo referente a tablas de doble entrada ya que la extensión de la teoría resulta similar para tablas de k entradas.

En principio con respecto a las tablas con clasificación completa de datos sabemos lo siguiente:

Se tiene una tabla de r renglones y c columnas, denotada $r \times c$.

n_{ij} : denota la frecuencia observada en la celda del i -ésimo renglón y la j -ésima columna.

n : denota la frecuencia total o el total de las observaciones para las celdas (i, j) .

$$n = \sum_{i=1}^r \sum_{j=1}^c n_{ij}$$

P_{ij} = denota la probabilidad de que una observación caiga en la celda del i -ésimo renglón y la j -ésima columna.

con $P_{ij} > 0, \forall i, j$ de tal modo que
$$\sum_{i=1}^r \sum_{j=1}^c P_{ij} = 1$$

luego también existen tablas en las que la clasificación de datos no es completa, ya sea porque existan resultados mezclados o porque sencillamente las observaciones en algunas celdas sean cero, por lo que tenemos:

P_{ij} denota las probabilidades individuales en cada celda.
 $n_{i.}$ denota los totales por renglón y $n_{.j}$ los totales por columna determinados como en el caso de las tablas con clasificación completa de datos:

$$n_{i.} = \sum_{j=1}^c n_{ij}, \quad n_{.j} = \sum_{i=1}^r n_{ij}$$

las probabilidades por columna o renglón quedan como anteriormente:

$P_{i.}, P_{.j}$ de donde:
$$\sum_{i=1}^r P_{i.} = 1, \quad \sum_{j=1}^c P_{.j} = 1$$

por lo que:
$$\sum_{i=1}^r \sum_{j=1}^c P_{ij} = 1$$

con las siguientes condiciones

$$n_{ij} > 0 \quad \forall i, j, \quad P_{ij} > 0 \quad \forall i, j$$

Suponemos que la distribución de probabilidad para todas las observaciones $ij, j = 1, \dots, r$ y $i = 1, \dots, c$ es multinomial y que su función de probabilidad es (nuestra función de verosimilitud $L(X; \theta)$)

$$\frac{n!}{\prod_{i=1}^r \prod_{j=1}^c n_{ij}} \prod_{i=1}^r \prod_{j=1}^c P_{ij}^{n_{ij}} \dots (1)$$

para un desarrollo completo de la distribución multinomial ver "Theory of Probability" de Bernard Harris, pag. 76-78.

Para probar independencia, planteamos la hipótesis tal como lo hicimos con tablas de información completa, de la siguiente manera:

$$H_0: P_{ij} = p_i \cdot p_j \quad \begin{matrix} i=1, 2, \dots, r \\ j=1, 2, \dots, c \end{matrix}$$

Obtenemos los estimadores máximo verosímiles, de la misma manera - que cuando tenemos tablas con clasificación completa de datos.

Se forma el logaritmo natural de (1) teniendo lo siguiente:

$$\begin{aligned} \ln L(\underline{x}; H_0) &= \ln \left(\frac{n!}{\prod_{i=1}^r \prod_{j=1}^c n_{ij}} \right) + \ln \left(\prod_{i=1}^r \prod_{j=1}^c p_{ij}^{n_{ij}} \right) \\ &= \ln n! - \sum_{i=1}^r \sum_{j=1}^c \ln n_{ij} + \sum_{i=1}^r \sum_{j=1}^c n_{ij} \ln p_{ij} \\ &= \ln n! - \sum_{i=1}^r \ln n_{i.} - \sum_{j=1}^c \ln n_{.j} + \sum_{i=1}^r n_{i.} \ln p_{i.} + \sum_{j=1}^c n_{.j} \ln p_{.j} \end{aligned}$$

para encontrar los estimadores necesitamos substituir $p_{r.}$ y $p_{.c}$

$$\begin{aligned} p_{r.} &= 1 - \sum_{i=1}^{r-1} p_{i.} & \text{y} & \quad p_{.c} = 1 - \sum_{j=1}^{c-1} p_{.j} \\ \ln L(\underline{x}; H_0) &= \ln n! - \sum_{i=1}^r \ln n_{i.} - \sum_{j=1}^c \ln n_{.j} + \sum_{i=1}^{r-1} n_{i.} \ln p_{i.} + n_{r.} \ln p_{r.} \\ & \quad + \sum_{j=1}^{c-1} n_{.j} \ln p_{.j} + n_{.c} \ln p_{.c} \end{aligned}$$

derivando con respecto a $p_{i.}$ y $p_{.j}$ obtenemos:

$$\frac{\partial \ln L(\underline{x}; H_0)}{\partial p_{i.}} = \frac{n_{i.}}{p_{i.}} + \frac{n_{r.} (-1)}{1 - \sum_{i=1}^{r-1} p_{i.}} \quad , i=1, 2, \dots, r \dots (2)$$

$$\frac{\partial \ln L(\underline{x}; H_0)}{\partial p_{.j}} = \frac{n_{.j}}{p_{.j}} + \frac{n_{.c} (-1)}{1 - \sum_{j=1}^{c-1} p_{.j}} \quad , j=1, 2, \dots, c \dots (3)$$

igualando a cero

$$\frac{n_{i.}}{p_{i.}} - \frac{n_{r.}}{p_{r.}} = 0 \Rightarrow \frac{n_{i.}}{p_{i.}} = \frac{n_{r.}}{p_{r.}} \Rightarrow \frac{p_{r.}}{n_{r.}} \sum_{i=1}^r n_{i.} = \sum_{i=1}^r p_{i.}$$

$$= \frac{p_{r.}}{n_{r.}} n = 1$$

$$\therefore \hat{p}_{r.} = \frac{n_{r.}}{n}$$

siendo igual para cualquier término $\therefore \hat{p}_{i.} = \frac{n_{i.}}{n} \quad i=1, \dots, r$

de igual modo $\hat{p}_{.j} = \frac{n_{.j}}{n} \quad j=1, \dots, c$

la solución de las ecuaciones (3) y (4) igualadas a cero, nos van a proporcionar los estimadores máximo verosímiles. Existen casos en que no se pueden encontrar los estimadores máximo verosímiles de forma directa, por lo que existe el método iterativo para encontrar las soluciones numéricas, el que se expone a continuación:

Primero.- Se escoge un estimador consistente "arbitrario", como un punto inicial, digamos $p^{(0)}$ que es la probabilidad asignada a la celda elegida arbitrariamente el que debe cumplir con la siguiente definición*)

*) Se dice que una sucesión de estimadores $\{T_n\}$ es consistente si:

$$\lim_{n \rightarrow \infty} E[(T_n - \theta)^2] = 0$$

i.e. T_n tiende a θ en media cuadrática.
observese que

$$\therefore \lim_{n \rightarrow \infty} E[(T_n - \theta)^2] = \lim_{n \rightarrow \infty} [\text{Var}(T_n) + (\text{Sesgo}(T_n))^2]$$

Segundo.- Obtenemos $\{P_{i.}^{(\nu+1)}\}$ de $P^{(\nu)}$ y obtenemos $\{P_{.j}^{(\nu+2)}\}$ de $P^{(\nu+1)}$

donde $P_{i.}^{(-1)} = P_{i.}^{(0)} \quad \forall i$

los estimadores de P_{ij} de $P_{i.}^{(\nu+1)}$ y $P_{.j}^{(\nu+2)}$ están dados por:

$$P_{ij}^{(\nu)} = (P_{i.}^{(\nu-1)})(P_{.j}^{(\nu)}) \quad \text{y} \quad P_{ij}^{(\nu+1)} = (P_{i.}^{(\nu+1)})(P_{.j}^{(\nu)})$$

Este procedimiento convergerá o tenderá a estabilizarse si en alguna etapa $P^{(\nu)}$ es suficientemente cercano a la verdadera solución \hat{P} y si el tamaño de la muestra n es grande, para que nos permita cumplir con la condición de contingencia.

Existen dos casos en que no se pueden obtener los estimadores máximo verosímiles de P_{ij} bajo nuestra hipótesis de independencia y son:

a) Cuando dos o más columnas o renglones están mezclados completamente en forma rectangular.

Por ejemplo: en una tabla de 3×4 , N_{11}, N_{12}, N_{21} y N_{22} están mezclados o también N_{13}, N_{14}, N_{23} y N_{24} están mezclados, es decir que no se pueden determinar las probabilidades de cada celda individualmente donde no podemos separar los elementos de un grupo de $P_{i.}$'s o un grupo de $P_{.j}$'s en la función de verosimilitud y por tanto no podemos estimarlos separadamente.

b) Cuando la disminución de grados de libertad debido a frecuencias

mezcladas es mayor que $(r-1)(c-1)$, ya que de las frecuencias mezcladas perdemos mas grados de libertad que los que la tabla permite.

3.3.2. CASO ESPECIAL DONDE ALGUNAS CELDAS EN LA CLASIFICACION SON APRIORI CERO.

Nos referimos a tablas que contienen en su estructura celdas vacías, distinguiendolas a éstas de las tablas que contienen observaciones con valor cero.

Bajo el modelo de independencia casi completa para los valores esperados de las celdas en la formación incompleta, se tienen las mismas ecuaciones de máxima verosimilitud, que en el caso anterior sólo que el valor de las frecuencias observadas es:

$$n_{ij} = \begin{cases} 0 & \text{si contienen ceros en sus celdas.} \\ 1 & \text{en otro caso.} \end{cases}$$

Teniendo las mismas ecuaciones para los estimadores máximo verosímiles, solo que reduciendo el tamaño de la tabla, sumando renglones o columnas y disminuyendo el número de grados de libertad necesarios, como vemos en el siguiente ejemplo:

Se quiere demostrar la relación existente entre el tipo de equipo siniestrado y el deducible aplicado, o lo que es lo mismo ¿Depende el tipo de equipo de la siniestralidad ocurrida ?

<u>DEDUCIBLE</u>	<u>SINIESTRADA</u>	<u>NO SINIESTRADA</u>
0.10	0	4
0.15	0	11
0.20	1	85
0.25	0	72
0.30	2	92
0.35	0	178
0.40	0	121
0.45	0	0
0.50	3	118
1.00	1	88
2.00	0	8
3.00	4	68
10.00	2	6

TOTALES

Teniendo una tabla de 13X2 con $(2-1) (13-1) = 12$ grados de libertad.

En la que se observa que en su estructura existen muchas celdas con valor cero y la que reducimos teniendo la siguiente tabla:

<u>DEDUCIBLE</u>	<u>SINIESTRADA</u>	<u>NO SINIESTRADA</u>	<u>TOTAL</u>
0.10-0.30	15	252	267
0.35-0.50	7	414	421
1.00-3.00	15	154	169
10.00	2	7	9
TOTAL	39	827	866

teniendo una tabla de 2X4 con $(2-1) (4-1) = 3$ grados de libertad.

3.3.3. METODO DE PARTICIPACIONES DE J1-DUADRADA PARA TABLAS DE CONTINGENCIA DE k ENTRADAS.

A continuación se presenta un ejemplo de triple entrada en el que: Ho: Se quiere determinar si existe alguna relación de dependencia entre el tipo de equipo con el deducible aplicado y la siniestralidad ocurrida.

CLASE DE EQUIPO

DEDUCIBLE	100		200		300		400	
	<u>NO SINIESTRO</u>	<u>SI SINIESTRADO</u>	<u>NO</u>	<u>SI</u>	<u>NO</u>	<u>SI</u>	<u>NO</u>	<u>SI</u>
0.10-0.30	N111 215	N112 5	N121 8	N122 -	N131 23	N132 9	N141 1	N142 1
0.35-0.50	N211 335	N212 6	N221 46	N222 -	N231 1	N232 1	N241 -	N242 -
1.00-3.00	N311 82	N312 4	N321 51	N322 -	N331 13	N332 11	N341 -	N342 -
10.00	N411 2	N412 1	N421 5	N422 -	N431 -	N432 -	N441 -	N442 -
TOTAL	634	16	110	-	37	21	1	1
DEDUCIBLE	500		600		900		1000	
	<u>NO</u>	<u>SI</u>	<u>NO</u>	<u>SI</u>	<u>NO</u>	<u>SI</u>	<u>NO</u>	<u>SI</u>
0.10-0.30	N151 1	N152 -	N161 1	N162 -	N171 2	N172 -	N181 1	N182 -
0.35-0.50	N251 8	N252 -	N261 4	N262 -	N271 17	N272 -	N281 3	N282 -
1.00-3.00	N351 1	N352 -	N361 1	N362 -	N371 2	N372 -	N381 4	N382 -
10.00	N451 -	N452 -	N461 -	N462 -	N471 -	N472 -	N481 -	N482 -
TOTAL	10	-	6	-	21	-	8	-

como se ve al ir aumentado el número de entradas a tabla de contingencia, la determinación de sus estimadores se hace mas compleja, por lo que existe el "Método de Particiones de la Ji-cuadrada para tablas de contingencia de K entradas", el que consiste básicamente en ir tomando participaciones de la tabla de k entradas, reduciendose el problema a considerar subtablas de la tabla original simplificando la prueba, de esta manera. Este método considera también la determinación de frecuencias esperadas - el que consiste en ir determinando las frecuencias marginales de las subtablas que se van obteniendo en las particiones. Para mayor información sobre este método ver Goodman, L.A. (1971). "Portitioning of Chi-Square, Analysis of Marginal Contingency Tables, and Estimation of Expected - Frecuencias in Multidimensional Contingency Tables". Journal of the American Statistical Association, Vol. 66, pag. 339-344.

3.3.4. PRECISION DE UNA APROXIMACION A LA POTENCIA DE LAS PRUEBAS DE BONDAD DE AJUSTE J1-CUADRADA CON PEQUEÑAS PERO IGUALES FRECUENCIAS ESPERADAS.

En un problema de bondad de ajuste con muestra de tamaño N, K grupos, el límite de la distribución nula de la estadística de prueba, cuando $N \rightarrow \infty$ es la distribución χ^2 con $k-1$ grados de libertad. Investigaciones de la extensión de la validez de las pruebas χ^2 para valores pequeños de N usualmente se enfatiza en la magnitud de las frecuencias esperadas en cada grupo mas que en el tamaño de la muestra directamente. Numerosos estudios han demostrado que el nivel de significancia de la prueba permanece igual cuando las frecuencias esperadas son pequeñas pero iguales.

Cuando el tamaño de la muestra se incrementa y la hipótesis alternativa se tiende a la hipótesis nula, la potencia de la prueba Ji-cuadrada puede ser aproximado por una distribución χ^2 no-central. Si nos

referimos a la potencia de la muestra cuando el tamaño es grande como la potencia nominal, entonces podemos comparar la estimación de la potencia de la prueba χ^2 con la potencia nominal, cuando las frecuencias esperadas bajo la hipótesis nula son pequeñas pero iguales. Stakter, M.J. (1968) Accuracy of an Approximation to the Power of the Chi-square Goodness of Fit Test with Small but equal expected Frecuencias. Journal of the American Statistical Association 63, 912-918 (4.5).

3.3.5. SOBRE EL ANALISIS DE TABLAS DE CONTINGENCIA MULTIDIMENSIONAL.

Podemos decir que el análisis de tablas de contingencia consiste en tres pasos básicos que son:

- 1) La formulación de la hipótesis nula, condicionada en lo que necesitamos probar, que en nuestro caso independencia total.
- 2) Una vez determinada la hipótesis nula, el siguiente paso es estimar las frecuencias de las celdas bajo la hipótesis nula usando las frecuencias marginales.
- 3) Por último, una vez determinada la hipótesis y efectuando -- los cálculos de las frecuencias esperadas, es necesario interpretar los resultados en términos de las variables involucradas.

Ku, H.H., Varner, R.N. and Kullback, S. (1971). On the Analysis of Multidimensional Contingency Tables.

Journal of the American Statistical Association, 66. 55-64.

CAPITULO IV

CONCLUSIONES

En el capítulo I después de presentar varias definiciones de sistema, derivamos la que a nosotros nos conviene y en el sentido que nos interesa, diciendo que sistema es un conjunto de elementos coordinados y relacionados entre sí pero interdependientes, elementos que son información estadística, con un fin determinado explícitamente como ser instrumento de apoyo para un tomador de decisiones, planeación, administración, etc...; luego cualquier situación o problema que se nos presenta en la vida diaria podemos relacionarlo con un sistema ya sea abstracto o concreto.

Siempre que se quiera analizar o diseñar un sistema es importante definir e identificar sus objetivos teniendo cuidado de jerarquizarlos, ya que se corre el riesgo de caer en el error de recalcar lo que es obvio.

Es básico identificar el medio ambiente en el que se desarrolla el sistema así como los recursos con que se cuenta y los componentes del mismo, delimitando cada uno de ellos sin confundirlos para tener perfectamente definido cualquier sistema.

En lo que se refiere a la estructura de un sistema podemos decir que es la resultante de la sub-división de sus metas globales en una escala jerárquica de sub-sistemas menos complejos, siendo que cada estructura representa un sistema diferente; siempre que se busque una estructura para el sistema no hay que descuidar los aspectos siguientes: grado de fragmentación, complejidad de las tareas de los sub-sistemas y la comunicación dentro del sistema, seleccionando la estructura que mas se adapte a nuestras necesidades, pensando siempre en nuestros requerimientos de información así como en la concordancia de los objetivos originados con los de mayor nivel.

Para concluir con lo relacionado a sistemas podemos decir que el

sistema ideal es el denominado "integrado" ya que reduce costos, evita duplicidad de datos y actividades y proporciona un acceso fácil a la información y permite cumplir mejor con un complejo procesamiento que se requiere para lograr un planeamiento comprensivo, puesto que está organizado alrededor de una base común de datos.

En el capítulo II, con la exposición de un caso práctico de un sub-sistema de información que actualmente está operando comprobamos la validez de la teoría expuesta en el capítulo anterior, no obstante esto se reafirma al justificar plenamente el sistema a través de la identificación del objetivo del sistema, el medio ambiente del sistema, los recursos con que cuenta el sistema y los componentes del sistema.

El objetivo general del sistema resulta genérico, ya que es aplicable a todos los productos o servicios que ofrece el sector asegurador. Acerca del medio ambiente en el que se desarrolla el sistema, veo dos problemas que son en mi opinión los más importantes: 1°) la falta de conocimiento de la teoría estadística de las personas que generan la información y 2°) la existencia de criterios flexibles para la selección de riesgos y en consecuencia una inapropiada clasificación de la información ya que la ocurrencia de ellos pueden provocar un cambio importante en los objetivos del sistema e inclusive provocar su fracaso.

En lo referente a los componentes, se dice que la única razón para separar el sistema en componentes es para proporcionar al analista el tipo de información que necesita para poder decir si el sistema está operando adecuadamente.

En el capítulo III, al analizar las tablas obtenidas con datos reales al analizar el tipo de información con que se cuenta en la práctica para tener un control sobre el buen funcionamiento de cualquier tipo de

seguro, se pensó en que tipo de análisis estadístico era adecuado en función de este tipo de información por lo que se generó la necesidad de justificación del uso de tablas de contingencia porque se contaba con clasificación incompleta de datos, valor de cero en las frecuencias observadas en las celdas y dificultad para obtener las frecuencias esperadas por ser tablas de más de dos entradas; determinando lo siguiente:

- 1.- Para el caso de tablas de contingencia con clasificación incompleta de datos, caso en el que no se pueden obtener los estimadores máximo verosímiles en forma directa, existe un método iterativo el que se describió detalladamente en el punto 3.3.1.
- 2.- Tablas de contingencia donde algunas celdas en la clasificación son aprioricero, en que se tienen las mismas ecuaciones para los estimadores de máxima verosimilitud y consiste en reducir el tamaño de la tabla, sumando renglones o columnas y disminuyendo el número de grados de libertad necesarios conforme al número de celdas en la tabla reducida.
- 3.- En los casos de tablas de más de dos entradas la obtención de frecuencias esperadas se torna mas complejo conforme van aumentando las entradas, problema que se puede reducir aplicando el "método de Particiones de la Ji-Cuadrada para tablas de contingencia de k-entradas", que consisten en hacer particiones de la tabla de k-entradas para formar sub-tablas e ir obteniendo las frecuencias marginales de las sub-tablas formadas.
- 4.- Sobre el análisis de tablas de contingencia multidimensional, agregamos únicamente el procedimiento básico para su análisis el que consiste en:

- a) Formulación de la hipótesis nula.
- b) Estimar las frecuencias esperadas en cada celda.
- c) Interpretar los resultados obtenidos en términos de las variables involucradas.

A P E N D I C E

TEORIA DE REDES
(CONCEPTOS FUNDAMENTALES)

TEORIA DE REDES.- Ha tenido importantes aplicaciones en toria de la información, cibernética, el estudio de sistemas de transportación y en la planeación y control de proyectos de investigación y desarrollo.

Ciertos aspectos de la toria de redes (comunmente llamado "Teoría - de Flujo de Redes") que se ha convertido en una gran herramienta de la investigación de operaciones.

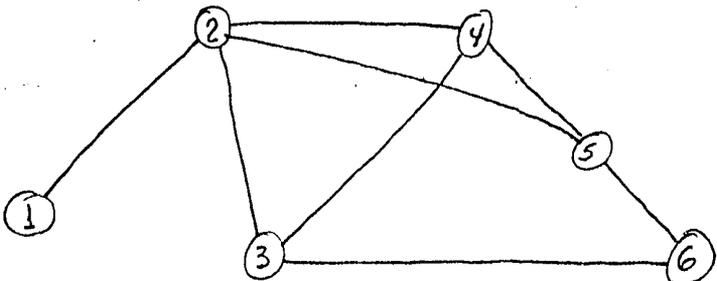
Para nuestro interés en los sistemas de información consiste en escoger un conjunto de conexiones que propocione un flujo o conexión entre dos puntos de una red de tal manera que minimice el total de longitud (ó número de pasos o actividades) de ésas actividades.

La planeación y control de investigación y desarrollo de proyectos son un problema, el cual ha sido atacado involucrando técnicas de análisis de redes, la técnica en este caso es conocida como PERT (Evaluación de Programas y Técnicas de Revisión).

Antes de definir una red se definirá el concepto de gráfica:

Def.- Una "gráfica" es un conjunto de puntos de unión llamados - "nodos", los cuales están unidos por líneas llamadas "ramas", "arcos", "ligas" u "orillas".

Ejemplo:



A continuación se da la definición de red:

Def.- Una "red" es considerar una gráfica con un flujo de algún tipo entre sus arcos.

Ejemplo:

NODOS	ARCOS	FLUJO
punto conmutable puntos de <u>comuni</u> cación.	alambres canales	mensajes (información)

Existe terminología adicional que ha sido desarrollada para describir gráficas. Una "cadena" entre el nodo i y el j es una secuencia de arcos que conectan a ambos nodos.

Ejemplo.- Una cadena entre el nodo (1) y el (6) es la secuencia de arcos (1,2), (2,5), (5,6) o viceversa. Cuando la dirección de sólo el recorrido de la cadena es también especificado, será llamado "camino". Un "ciclo" es una cadena que conecta un nodo consigo mismo.

Ejemplo.- (2,4), (4,3) y (3,2).

Se dice que una gráfica es "gráfica asociada conexa", si existe una cadena que conecte todos los pares de nodos.

Ejemplo.- La gráfica de ejemplo es una gráfica asociada, pero ésta no sería si los arcos (2,3), (3,4) y (5,6) fueran removidos.

Def.- Un "árbol" es una gráfica asociada que no contiene ciclos.

Ejemplo: Sería un "árbol" si fueran removidos los arcos (3,6), (3,4) y (4,5). Un arco de una gráfica se dice está "orientado" ó "dirigido", si existe un sentido de dirección atribuido al arco tal que un nodo es considerado el punto de origen y el otro nodo el punto de destino.

Def.- Una gráfica orientada es aquella en que todos sus arcos están orientados.

Si una gráfica orientada es una red, la orientación de un arco es ficticia o simulada por existir la posibilidad de dirección de flujo por el arco.

La "capacidad" de flujo de un arco en una dirección específica es el límite superior a la posible magnitud de la proporción de flujo (o cantidad total de flujo) en el arco en esa dirección. La capacidad de flujo puede ser cualquier cantidad no negativa, inclusive infinita. Un arco es orientado si la capacidad de flujo es cero en una dirección. Un nodo puede también tener una capacidad de flujo limitada - aunque esa posibilidad será ignorada aquí.

Un nodo es una red, es algunas veces referido como un "origen" si todos sus arcos tienen una orientación tal que el flujo se separe del nodo. De igual modo, es algunas veces llamado un "hundimiento" si cada uno de los arcos está orientado hacia ese nodo.

B I B L I O G R A F I A

- 1.- Diccionario de la Lengua Española.
- 2.- Sistemas y Procedimientos: Un manual para los negocios y la industria.
- 3.- La Administración de Personal en el Sector Público, un enfoque sistemático. Duhalt Krauss, Miguel.
- 4.- Antología de Análisis y Diseño de Sistemas. J.C. Emery.
- 5.- Perspectivas en la Teoría General de Sistemas. Ludwig von Bertalanffy.
- 6.- El Enfoque de Sistemas. C. West Churchman.
- 7.- Introduction to Operations Research. Frederick S. Hillier, Gerald J. Lieberman.
- 8.- Introduction to Statistical Theory. Hoel, P.G., Port, S.C. and Stone, CH.J. Houghton Mifflin Co. 1971.
- 9.- Introducción a la Teoría de la Estadística. Alexander M. Mood y Franklin A. Graybill.
- 10.- Nonparametric Statistical Inference. Jean Dickinson Gibbons.
- 11.- Institución en el Plano Internacional de un Sistema unificado de estadísticas de Seguros.
Conferencia de las Naciones Unidas sobre Comercio y Desarrollo. Grupo de expertos en estadísticas de seguros. 1970.
- 12.- Intercambio de datos de seguro por medio de memorias magnéticas. Un sistema propuesto por la Münchener Rück.
- 13.- Distribution-free test of independence. The Annals of Mathematical - - Statistics, 38, 429-446. (1967). Bell, C.B. and Doksum, K.A.
- 14.- Descriptive Statistics for Nonparametric Modeles. The Annals of Statistics, 3, 1038-1069. (1975). Bickel, P.P. and Lehmann, E.L.
- 15.- The Power of Chi-square test for contingency tables.
Journal of the American Statistical Association, 61, 965-975. (1966). Chapman, D.G. and Meng, R.C.
- 16.- The Analysis of Contingency Tables with Incompletely Classified Data. Biometrics, 32, 133-134. (1976). Chen. T. and Fienberg, S.E.

- 17.- Alternative Analysis of Contingency Tables. Journal of the Royal Statistical Society. (B), 28, 164-179. (1966) Gart, J.J.
- 18.- Partitioning of Chi-square, Analysis of Marginal Contingency Tables and Stimulation of Expected Frequencies in Multidimensional Contingency Tables. Journal of the American Statistical Association, 66, 339-344. (1971). Goodman, L.A.
- 19.- On the Analysis of Multidimensional Contingency Tables. Journal of the American Statistical Association, 66, 55-64. Ku, H.H., Varner, R.N.- and Kallback, S.
- 20.- Accuracy of an Approximation to the Power of the Chi-square Goodness - of fit test with small but equal expected frequencies. Journal of The American Statistical Association, 63, 912-918. (1968). Slakter, M.J.
- 21.- Sistemas de "Desinformación" de Funcionarios, Russell L. Ackoff. Universidad de Pennsylvania. Management Science. Vol.14, No. 4 Dic.1967.
- 22.- Towards a System of Systems Concepts. Russell L. Ackoff. Management Science. Vol. 17, No. 11, July 1971.